



UNIVERSIDADE D  
COIMBRA

André Miguel Pereira Pacheco

**METADADOS PARA A DESCRIÇÃO ARQUIVÍSTICA  
DIGITAL**

PROPOSTA DE UM MODELO PARA A AUTENTICIDADE

Tese no âmbito do Doutoramento em Ciência da Informação orientada pela  
Professora Doutora Maria Cristina Vieira de Freitas e pelo Professor Doutor Carlos  
Guardado da Silva e apresentada ao Departamento de Filosofia, Comunicação e  
Informação da Faculdade de Letras da Universidade de Coimbra

Setembro de 2021

Faculdade de Letras  
da Universidade de Coimbra

# METADADOS PARA A DESCRIÇÃO ARQUIVÍSTICA DIGITAL

## Proposta de um modelo para a autenticidade

André Miguel Pereira Pacheco

Tese no âmbito do Doutoramento em Ciência da Informação orientada pela Professora Doutora Maria Cristina Vieira de Freitas e pelo Professor Doutor Carlos Guardado da Silva e apresentada ao Departamento de Filosofia, Comunicação e Informação da Faculdade de Letras da Universidade de Coimbra

Julho de 2021



UNIVERSIDADE D  
COIMBRA



# Agradecimentos

---

À minha família, que sempre fez tudo por mim, por vezes mais do que aquilo que mereço.

Aos meus amigos por simplesmente serem, com quem passei mais tempo do que a escrever a tese.

Aos meus colegas e amigos de turma no doutoramento, com os quais considero sinceramente ter sido uma sorte coincidir, de tão alegres, simpáticos e bem-dispostos que são.

Aos professores do ciclo de estudos, em particular à professora Maria Manuel Borges, cuja simpatia, apoio, zelo e compromisso para com os estudantes são extraordinários, não tendo eu sido exceção.

Embora haja quem diga que não se deve agradecer aos orientadores, sinto que o devo fazer, tamanha a importância do seu papel na construção deste percurso. Por entre uma ideia de candidatura mudada, um projeto de tese falhado e um quase abandono da esperança em mim, conseguimos os três encontrar um rumo. Além do pessoal, sinto que tive um crescimento científico significativo ao longo destes anos, que não seria possível sem ter tido os orientadores amigos que tive.

Merece também uma menção o professor Daniel Flores, que castigou justamente o projeto de tese, fornecendo assim uma fantástica e muito necessária oportunidade de crescimento.

À Basia, que se aventurou comigo em Coimbra e que me levava bolos do caco enquanto trabalhava na biblioteca, que comíamos alegremente nas escadas.

*Last but not least*, o meu maior agradecimento destina-se à Fundação para a Ciência e Tecnologia por ter financiado os meus estudos durante os quatro anos de estudo no doutoramento, apoio sem o qual provavelmente não teria enveredado por este caminho.



# Resumo

---

No decurso do século 21 têm surgido tecnologias de representação de informação baseadas num modelo de dados em grafo, possibilitando a criação de descrições interoperáveis partilháveis no contexto da *Web Semântica*.

Porém, a descrição arquivística, enquanto forma de representação de informação, tem sido tradicionalmente hierárquica, com uma narrativa imposta pelas instituições, uma visão por vezes insuficiente para representar a natureza dinâmica da informação digital que deve ser recuperada à luz de diversos contextos, sendo o seu sentido construído em cada um desses contextos pelos utilizadores. Além disso, a presunção da autenticidade dos documentos de arquivo em contexto digital não deve continuar a ser vista como residente nos documentos em si, devendo antes ser procurada nos metadados que os acompanham. Porém, a compreensão de quais os elementos que constituem a autenticidade dos recursos digitais permanece uma questão difícil, com respostas díspares.

Perante estes problemas, procede-se a uma revisão de literatura sistemática e crítica sobre a descrição da informação em arquivos, com um enfoque no contexto digital e na garantia da autenticidade e da confiabilidade. Com base numa análise da literatura científica e da literatura técnica, propõe-se ainda um modelo de metadados para a autenticidade. Assim, é empregue uma metodologia mista que combina métodos qualitativos de seleção de fontes e de investigação documental com métodos quantitativos de engenharia de requisitos de forma a extrair da literatura científica e técnica as bases para a construção do modelo.

Conclui-se que as descrições arquivísticas são um dos mecanismos mais importantes para a garantia da autenticidade dos recursos digitais, sendo útil que essa descrição assimile o pensamento pós-moderno e que potencie as capacidades do contexto digital. Também se considera que o conjunto de metadados proposto é adequado e suficiente para essa garantia da autenticidade na maior parte dos contextos.

**PALAVRAS-CHAVE:** representação da informação; descrição arquivística; metadados; modelo; autenticidade

# Abstract

---

*Graph technologies for information representation have appeared throughout the 21<sup>st</sup> century, enabling the creation of interoperable descriptions that can be shared in the framework of the Semantic Web.*

*However, archival description, as a means of information representation, has traditionally been hierarchical, with a narrative imposed by institutions, a view often insufficient to represent the dynamic nature of digital information that should be retrieved according to different contexts. It is in these different contexts that its meaning is created by the users. In addition, the assertion of the authenticity of digital records should not continue to lie in the records themselves but should instead be placed in their metadata. Despite this, the understanding of which metadata elements best assert the authenticity of digital records remains a problematic question without a simple answer.*

*Facing these problems, a critical and systematic literature review about information representation in archives is presented, with a focus in the digital context and in ensuring the authenticity and trustworthiness of records. A metadata model for authenticity is also proposed, grounded in an analysis of the technical and scientific literatures. Therefore, a mixed methodology is employed, combining qualitative methods of documentary research and source selection with quantitative methods of requirements engineering, as a means of extracting from the scientific and the technical literatures key notions to build the model.*

*It is concluded that archival descriptions are one of the most important mechanisms to ensure the authenticity of digital resources, and that such description should assimilate the post-modern thinking, as well as enable the full potential of the digital context. It is also considered that the proposed metadata model is sufficient and suitable to ensure such authenticity in a significant number of contexts.*

**KEYWORDS:** *information representation; archival description; metadata; model; authenticity*

# Sumário

LISTA DE TABELAS.....	XI
LISTA DE FIGURAS .....	XIII
LISTA DE QUADROS .....	XIV
INTRODUÇÃO .....	1
<b>1 METADADOS EM ARQUIVOS.....</b>	<b>7</b>
<b>1.1 CONCEPTUALIZAÇÃO .....</b>	<b>7</b>
1.1.1 A importância dos metadados.....	7
1.1.2 Definição e discussão .....	9
<b>1.2 ETAPAS DO USO DE METADADOS.....</b>	<b>17</b>
1.2.1 Metadados na Antiguidade .....	17
1.2.2 A introdução da imprensa na Europa .....	21
1.2.3 O Século 20.....	28
1.2.4 A era da Internet .....	38
<b>1.3 TIPOS E FUNÇÕES.....</b>	<b>43</b>
1.3.1 Tipos de metadados .....	45
1.3.2 Funções dos metadados.....	52
<b>1.4 FORMAS DE EXPRESSÃO .....</b>	<b>55</b>
1.4.1 Modelo relacional.....	57
1.4.1.1 <i>Dados tabulares</i> .....	57
1.4.1.2 <i>Modelo relacional</i> .....	58
1.4.2 Modelo hierárquico .....	59
1.4.3 Modelo em grafo.....	61
1.4.3.1 <i>A Web Semântica</i> .....	62
1.4.3.2 <i>RDF</i> .....	64
<b>1.5 TERMINOLOGIA DE NORMAS DE METADADOS.....</b>	<b>66</b>
<b>1.6 A IMPORTÂNCIA DA INTEROPERABILIDADE .....</b>	<b>74</b>
<b>2 REPRESENTAÇÃO DA INFORMAÇÃO.....</b>	<b>77</b>
<b>2.1 O CONCEITO DE REPRESENTAÇÃO .....</b>	<b>77</b>
2.1.1 Representação .....	78
2.1.2 Representação da informação .....	83
<b>2.2 ARQUIVOS NA PÓS-MODERNIDADE .....</b>	<b>91</b>
2.2.1 O desrespeito pelos fundos .....	99
<b>2.3 CONCEPTUALIZAÇÃO E OBJETO DA DESCRIÇÃO ARQUIVÍSTICA.....</b>	<b>101</b>
2.3.1 Controlo intelectual em arquivos .....	105
<b>2.4 AUTENTICIDADE E FIDEDIGNIDADE DE E NAS REPRESENTAÇÕES ARQUIVÍSTICAS .....</b>	<b>107</b>
2.4.1 Do vínculo arquivístico para o vínculo informacional .....	111
2.4.2 Autenticidade, confiabilidade e fidedignidade.....	115

2.4.3	Autenticação .....	117
<b>3</b>	<b>METODOLOGIA.....</b>	<b>121</b>
<b>3.1</b>	<b>INTRODUÇÃO .....</b>	<b>121</b>
3.1.1	Motivações e problema .....	121
3.1.2	Objetivos .....	122
3.1.3	Limitações.....	123
<b>3.2</b>	<b>NATUREZA DA INVESTIGAÇÃO .....</b>	<b>124</b>
<b>3.3</b>	<b>MÉTODOS DE RECOLHA DE DADOS .....</b>	<b>135</b>
3.3.1	Investigação documental .....	135
<b>3.4</b>	<b>CORPUS DOCUMENTAL.....</b>	<b>141</b>
3.4.1	Revistas científicas .....	144
3.4.1.1	<i>Library &amp; Information Science Abstracts</i> .....	145
3.4.1.2	<i>Library, Information Science &amp; Technology Abstracts</i> .....	149
3.4.1.3	<i>A escolha da base de dados de referência</i> .....	153
3.4.2	Monografias .....	157
3.4.3	Comunicações científicas.....	157
3.4.4	Teses e dissertações.....	160
3.4.4.1	<i>Open Access Theses and dissertations</i> .....	161
3.4.4.2	<i>Repositório Científico de Acesso Aberto em Portugal</i> .....	162
<b>3.5</b>	<b>ANÁLISE DOS DADOS .....</b>	<b>163</b>
3.5.1	Métodos de análise de dados .....	166
3.5.1.1	<i>Investigação documental</i> .....	166
3.5.1.2	<i>Engenharia de requisitos</i> .....	168
3.5.2	Seleção de artigos científicos .....	169
3.5.2.1	<i>American Archivist</i> .....	170
3.5.2.2	<i>Archival issues</i> .....	171
3.5.2.3	<i>Archival Science</i> .....	171
3.5.2.4	<i>Archivaria</i> .....	173
3.5.2.5	<i>Archives (London)</i> .....	174
3.5.2.6	<i>Archives (Quebec)</i> .....	174
3.5.2.7	<i>Archives and Manuscripts</i> .....	175
3.5.2.8	<i>Archives and Records</i> .....	177
3.5.2.9	<i>Cadernos BAD</i> .....	178
3.5.2.10	<i>Journal of Archival Organization</i> .....	180
3.5.2.11	<i>Provenance</i> .....	182
3.5.2.12	<i>Records Management Journal</i> .....	183
3.5.3	Seleção de monografias .....	186
3.5.3.1	<i>American Archivist</i> .....	187
3.5.3.2	<i>Archival Issues</i> .....	189
3.5.3.3	<i>Archival Science</i> .....	190
3.5.3.4	<i>Archivaria</i> .....	190
3.5.3.5	<i>Archives (London)</i> .....	192
3.5.3.6	<i>Archives (Quebec)</i> .....	192

3.5.3.7	<i>Archives and Manuscripts</i> .....	192
3.5.3.8	<i>Archives and Records</i> .....	196
3.5.3.9	<i>Cadernos BAD</i> .....	198
3.5.3.10	<i>Journal of Archival Organization</i> .....	199
3.5.3.11	<i>Provenance</i> .....	199
3.5.3.12	<i>Records Management Journal</i> .....	200
3.5.3.13	<i>Síntese das monografias identificadas</i> .....	200
3.5.4	Seleção de comunicações científicas.....	205
3.5.5	Seleção de teses e dissertações .....	209
3.5.5.1	<i>Open Access Theses and Dissertations</i> .....	209
3.5.5.2	<i>Repositório Científico de Acesso Aberto em Portugal</i> .....	210
3.5.6	Seleção de normas de metadados.....	212
3.5.7	Estatísticas do <i>corpus</i> documental .....	214
<b>3.6</b>	<b>QUESTÕES ÉTICAS DE INVESTIGAÇÃO</b> .....	<b>217</b>
<b>4</b>	<b>IDENTIFICAÇÃO DE REQUISITOS</b> .....	<b>218</b>
<b>4.1</b>	<b>INTRODUÇÃO</b> .....	<b>218</b>
<b>4.2</b>	<b>REQUISITOS DA LITERATURA CIENTÍFICA</b> .....	<b>220</b>
4.2.1	Representação da informação .....	221
4.2.2	Descrição arquivística .....	226
4.2.3	Gestão da informação .....	231
4.2.4	Metadados.....	235
4.2.5	Normas de metadados .....	239
4.2.6	Autenticidade .....	242
4.2.7	Linked Data .....	250
4.2.8	Sumário.....	254
<b>4.3</b>	<b>REQUISITOS DA LITERATURA TÉCNICA</b> .....	<b>263</b>
4.3.1	ISAD(G) .....	263
4.3.1.1	<i>Contextualização</i> .....	263
4.3.1.2	<i>Estrutura de metadados</i> .....	264
4.3.1.3	<i>Metadados para a autenticidade</i> .....	266
4.3.2	AGRkMS .....	267
4.3.2.1	<i>Contexto histórico</i> .....	267
4.3.2.2	<i>Estrutura de metadados</i> .....	269
4.3.2.3	<i>Metadados para a autenticidade</i> .....	272
4.3.3	EAD .....	273
4.3.3.1	<i>Contexto histórico</i> .....	273
4.3.3.2	<i>Estrutura de metadados</i> .....	274
4.3.3.3	<i>Metadados para a autenticidade</i> .....	275
4.3.4	e-EMGDE.....	276
4.3.4.1	<i>Contexto histórico</i> .....	276
4.3.4.2	<i>Estrutura de metadados</i> .....	277
4.3.4.3	<i>Metadados para a autenticidade</i> .....	281

4.3.5	DACS.....	282
4.3.5.1	<i>Contexto histórico.....</i>	282
4.3.5.2	<i>Estrutura de metadados.....</i>	283
4.3.5.3	<i>Metadados para a autenticidade .....</i>	285
4.3.6	DCMES.....	286
4.3.6.1	<i>Contexto histórico.....</i>	286
4.3.6.2	<i>Estrutura de metadados.....</i>	287
4.3.6.3	<i>Metadados para a autenticidade .....</i>	289
4.3.7	VRA Core.....	289
4.3.7.1	<i>Contexto histórico.....</i>	289
4.3.7.2	<i>Estrutura de metadados.....</i>	290
4.3.7.3	<i>Metadados para a autenticidade .....</i>	292
4.3.8	MODS.....	293
4.3.8.1	<i>Contexto histórico.....</i>	293
4.3.8.2	<i>Estrutura de metadados.....</i>	294
4.3.8.3	<i>Metadados para a autenticidade .....</i>	295
4.3.9	CDWA .....	296
4.3.9.1	<i>Contexto histórico.....</i>	296
4.3.9.2	<i>Estrutura de metadados.....</i>	297
4.3.9.3	<i>Metadados para a autenticidade .....</i>	303
4.3.10	RiC.....	304
4.3.10.1	<i>Contexto histórico.....</i>	304
4.3.10.2	<i>Estrutura de metadados.....</i>	306
4.3.10.3	<i>Metadados para a autenticidade .....</i>	309
4.3.11	Sumário .....	310
<b>5</b>	<b>CONSTRUÇÃO DO MODELO DE REQUISITOS.....</b>	<b>312</b>
5.1	<b>ANÁLISE E DISCUSSÃO DOS REQUISITOS CIENTÍFICOS.....</b>	<b>312</b>
5.2	<b>ANÁLISE E DISCUSSÃO DOS REQUISITOS TÉCNICOS.....</b>	<b>320</b>
5.3	<b>PROPOSTA DO MODELO .....</b>	<b>329</b>
	<b>CONCLUSÃO:.....</b>	<b>337</b>
	<b>BIBLIOGRAFIA.....</b>	<b>341</b>

# Lista de Tabelas

---

TABELA 1 — CLASSES PRINCIPAIS DA CLASSIFICAÇÃO DECIMAL DE DEWEY (CDD), EM INGLÊS. ....	25
TABELA 2 — SÍNTESE DOS TIPOS DE METADADOS MAIS RECORRENTES NA LITERATURA E RESPECTIVAS DEFINIÇÕES.....	46
TABELA 3 — ANÁLISE QUANTITATIVA DA OCORRÊNCIA DA TIPOLOGIA DE METADADOS. ....	51
TABELA 4 — FUNÇÃO DESEMPENHADA POR CADA TIPO DE METADADOS. ....	54
TABELA 5 — TIPOS E FUNÇÕES PRIMÁRIAS DOS METADADOS. ....	54
TABELA 6 — EXEMPLO DA EXPRESSÃO DE METADADOS NO MODELO TABULAR.....	58
TABELA 7 — CARACTERIZAÇÃO DOS TIPOS DE NORMAS.....	70
TABELA 8 — DISTINÇÃO ENTRE NORMA DE METADADOS, ELEMENTO DE METADADOS, ESQUEMA DE METADADOS, MODELO CONCEPTUAL E ONTOLOGIA. ....	73
TABELA 9 — MÉTODOS E METODOLOGIA QUE GUIAM ESTE ESTUDO.....	129
TABELA 10 — IDENTIFICAÇÃO DE REVISTAS POTENCIALMENTE PERTINENTES.....	146
TABELA 11 — REVISTAS EXCLUÍDAS COM BASE NOS CRITÉRIOS DEFINIDOS.....	148
TABELA 12 — IDENTIFICAÇÃO DE REVISTAS DE ARQUIVÍSTICA INDEXADAS NA LISTA. ....	150
TABELA 13 — SELEÇÃO DE REVISTAS IDENTIFICADAS NA LISTA SEGUNDO OS CRITÉRIOS ESTABELECIDOS.....	151
TABELA 14 — LISTA FINAL DE TÍTULOS PERTINENTES OBTIDOS ATRAVÉS DA LISTA. ....	152
TABELA 15 — CONDIÇÕES DE ACESSO AOS TEXTOS INTEGRAIS DE ARTIGOS ALEATORIAMENTE SELECIONADOS. .....	155
TABELA 16 — CARACTERIZAÇÃO DAS REVISTAS SELECIONADAS. ....	156
TABELA 17 — CATEGORIAS DE ANÁLISE EMERGENTES A PARTIR DA REVISÃO DE LITERATURA. ....	164
TABELA 18 — ARTIGOS CIENTÍFICOS RETIRADOS DA <i>AMERICAN ARCHIVIST</i> .....	170
TABELA 19 — ARTIGOS CIENTÍFICOS RETIRADOS DA <i>ARCHIVAL ISSUES</i> .....	171
TABELA 20 — ARTIGOS CIENTÍFICOS RETIRADOS DA <i>ARCHIVAL SCIENCE</i> . ....	172
TABELA 21 — ARTIGOS CIENTÍFICOS RETIRADOS DA <i>ARCHIVARIA</i> .....	173
TABELA 22 — ARTIGOS CIENTÍFICOS RETIRADOS DA <i>ARCHIVES (QUEBEC)</i> .....	174
TABELA 23 — ARTIGOS CIENTÍFICOS RETIRADOS DA <i>ARCHIVES AND MANUSCRIPTS</i> .....	175
TABELA 24 — ARTIGOS CIENTÍFICOS RETIRADOS DA <i>ARCHIVES AND RECORDS</i> . ....	177
TABELA 25 — ARTIGOS CIENTÍFICOS RETIRADOS DOS CADERNOS BAD. ....	178
TABELA 26 — ARTIGOS CIENTÍFICOS RETIRADOS DO <i>JOURNAL OF ARCHIVAL ORGANIZATION</i> . ....	180
TABELA 27 — ARTIGOS CIENTÍFICOS RETIRADOS DA <i>PROVENANCE</i> .....	182
TABELA 28 — ARTIGOS CIENTÍFICOS RETIRADOS DO <i>RECORDS MANAGEMENT JOURNAL</i> . ....	183
TABELA 29 — <i>BOOK REVIEWS</i> RETIRADAS DA <i>AMERICAN ARCHIVIST</i> . ....	188
TABELA 30 — <i>BOOK REVIEWS</i> RETIRADAS DA <i>ARCHIVAL ISSUES</i> . ....	189
TABELA 31 — <i>BOOK REVIEWS</i> RETIRADAS DA <i>ARCHIVARIA</i> . ....	190
TABELA 32 — <i>BOOK REVIEWS</i> RETIRADAS DA <i>ARCHIVES (LONDON)</i> . ....	192
TABELA 33 — <i>BOOK REVIEWS</i> RETIRADAS DA <i>ARCHIVES &amp; MANUSCRIPTS</i> .....	192
TABELA 34 — <i>BOOK REVIEWS</i> RETIRADAS DA <i>ARCHIVES &amp; RECORDS</i> . ....	196
TABELA 35 — <i>BOOK REVIEWS</i> RETIRADAS DA <i>JOURNAL OF ARCHIVAL ORGANIZATION</i> . ....	199

TABELA 36 — <i>BOOK REVIEWS</i> RETIRADAS DA <i>PROVENANCE</i> . .....	200
TABELA 37 — LISTA FINAL DAS MONOGRAFIAS IDENTIFICADAS. ....	201
TABELA 38 — COMUNICAÇÕES CIENTÍFICAS RECUPERADAS EM LÍNGUA INGLESA E ESPANHOLA NA WOS. ..	206
TABELA 39 — TESES E DISSERTAÇÕES RECUPERADAS EM LÍNGUA INGLESA, PORTUGUESA E ESPANHOLA NO OATD. ....	209
TABELA 40 — TESES E DISSERTAÇÕES RECUPERADAS EM LÍNGUA INGLESA, PORTUGUESA E ESPANHOLA NO RCAAP. ....	210
TABELA 41 — NÓRMAS DE METADADOS A ANALISAR. ....	213
TABELA 42 — ENUMERAÇÃO ESTATÍSTICA DO UNIVERSO DE ARTIGO CIENTÍFICOS. ....	214
TABELA 43 — ENUMERAÇÃO ESTATÍSTICA DO UNIVERSO DE MONOGRAFIAS. ....	215
TABELA 44 — ENUMERAÇÃO ESTATÍSTICA DO UNIVERSO DE COMUNICAÇÕES CIENTÍFICAS. ....	216
TABELA 45 — ENUMERAÇÃO ESTATÍSTICA DO UNIVERSO DAS DISSERTAÇÕES E TESES ACADÉMICAS. ....	216
TABELA 46 — SUMÁRIO DOS REQUISITOS IDENTIFICADOS NA LITERATURA CIENTÍFICA, ORGANIZADOS POR TEMA. ....	255
TABELA 47 — ELEMENTOS DE METADADOS NA ISAD(G). ....	265
TABELA 48 — ELEMENTOS DE METADADOS DA AGRKMS. ....	270
TABELA 49 — ELEMENTOS DE METADADOS DA E-EMGE. ....	278
TABELA 50 — ESTRUTURA DE METADADOS DA DACS. ....	284
TABELA 51 — ESTRUTURA DE ELEMENTOS DO DUBLIN CORE, COM ALGUNS <i>REFINEMENTS</i> . ....	288
TABELA 52 — ELEMENTOS DE METADADOS USADOS NA VRA CORE 4.0. ....	291
TABELA 53 — ELEMENTOS E SUBELEMENTOS DA MODS 3.7. ....	294
TABELA 54 — CATEGORIAS E SUBCATEGORIAS PROPOSTAS PELA CDWA. ....	297
TABELA 55 — ENTIDADES E ATRIBUTOS PRESENTES NA RIC v0.2. ....	307
TABELA 56 — SUMÁRIO DOS ELEMENTOS PARA A AUTENTICIDADE IDENTIFICADOS EM CADA NORMA. ....	311
TABELA 57 — REQUISITOS ÚNICOS IDENTIFICADOS NA LITERATURA CIENTÍFICA. ....	313
TABELA 58 — REQUISITOS TEÓRICOS MAIS FREQUENTES. ....	315
TABELA 59 — REQUISITOS ÚNICOS IDENTIFICADOS A PARTIR DA ANÁLISE DA LITERATURA TÉCNICA. ....	322
TABELA 60 — REQUISITOS TÉCNICOS MAIS FREQUENTES. ....	327
TABELA 61 — DEFINIÇÕES DOS METADADOS DO MODELO. ....	333



# Lista de figuras

---

FIGURA 1 — INTERPRETAÇÃO DO SIGNIFICADO DAS FACETAS DE UMA CLASSIFICAÇÃO DE CÓLON. ....	32
FIGURA 2 — PESQUISA FACETADA NA WEB OF SCIENCE. CAPTURA DE ECRÃ EM MAIO DE 2020. ....	33
FIGURA 3 — EXEMPLO DE UM REGISTO MARC DE UM LIVRO. EXTRAÍDO DE LIBRARY OF CONGRESS (2019A) E ANOTADO A VERMELHO PELOS AUTORES. ....	36
FIGURA 4 — SEEING STANDARDS: A VISUALIZATION OF THE METADATA UNIVERSE. EXTRAÍDO DE RILEY (2010). ....	42
FIGURA 5 — COMPARAÇÃO ESQUEMÁTICA DOS TRÊS PRINCIPAIS TIPOS DE MODELOS DE DADOS. BASEADO EM HOOLAND & VERBORGH (2014, P. 13). ....	56
FIGURA 6 — EXEMPLO DAS RELAÇÕES NUM MODELO ENTIDADE-RELAÇÃO. ....	59
FIGURA 7 — ESTRUTURA DE UM TRIPLO. ....	65
FIGURA 8 — EXEMPLIFICAÇÃO DO ENCADEAMENTO DE VÁRIOS TRIPLOS NA EXPRESSÃO DE METADADOS. ....	65
FIGURA 9 — TIPOLOGIA DE NORMAS DE METADADOS. ....	68
FIGURA 10 — ESQUEMA CONCEPTUAL DOS CONCEITOS TEÓRICOS A EXPLORAR. ....	77
FIGURA 11 — “MY WIFE AND MY MOTHER IN LAW”: UMA JOVEM OU UMA IDOSA? FONTE: WIKICOMMONS. ....	97
FIGURA 12 — METODOLOGIA DA ABORDAGEM DEDUTIVA USADA TÍPICAMENTE NA INVESTIGAÇÃO QUANTITATIVA E DA LÓGICA INDUTIVA COMUM NUM ESTUDO QUALITATIVO. ADAPTADO DE CRESWELL (2014, PP. 59 E 66). ....	131
FIGURA 13 — DESENHO METODOLÓGICO DA INVESTIGAÇÃO. ....	132
FIGURA 14 — TRIANGULAÇÃO ENTRE A ANÁLISE DOS DADOS E AS QUESTÕES DE INVESTIGAÇÃO. ....	164
FIGURA 15 — ELEMENTOS DE METADADOS DERIVADOS DA LITERATURA CIENTÍFICA. O NÚMERO INDICA OS REQUISITOS QUE SATISFAZEM. ....	319
FIGURA 16 — ELEMENTOS DE METADADOS IDENTIFICADOS NA LITERATURA TÉCNICA, COM A INDICAÇÃO DO NÚMERO DO REQUISITO. ....	329
FIGURA 17 — ESQUEMA CONCEPTUAL DA CONFIABILIDADE. ....	331
FIGURA 18 — PROPOSTA DE MODELO DE METADADOS PARA A AUTENTICIDADE DAS DESCRIÇÕES ARQUIVÍSTICAS. ....	332

## Lista de quadros

---

QUADRO 1 — LISTA FINAL DE TÍTULOS PERTINENTES OBTIDOS ATRAVÉS DA LISA. ....	149
QUADRO 2 — COMPARAÇÃO DA LISA COM A LISTA. ....	153
QUADRO 3 — ELEMENTOS CONTEMPLADOS PELA EAD3 1.1.1.....	274

# Introdução

---

A cerca de cada dois anos, o número de transístores num circuito integrado duplica. É esta a ideia fundamental da lei de Moore, batizada em honra de Gordon Moore, cofundador da empresa americana Intel, que notou esta duplicação como uma tendência constante no passado e que podia ser prevista para o futuro. Formulada em 1965 como uma duplicação a cada ano e revista em 1975 para uma duplicação a cada dois anos, a lei de Moore, enquanto tendência histórica, tem-se continuado a observar. Acredita-se que estes avanços exponenciais na indústria de semicondutores estejam na base de um crescimento acelerado da tecnologia que se reflete não só na diversidade e no quase-milagre dos instrumentos que temos à disposição, como também na velocidade com que são introduzidos. Recorde-se que entre a imprensa (século 15) e a locomotiva a vapor (século 18) distanciam cerca de três séculos. Porém, entre a aterragem do homem na Lua (1969), a invenção do microprocessador (1969) e do computador pessoal (década 1970), o desenvolvimento do *smartphone* e das redes sociais (década 2000) e o renascimento do carro elétrico (década 2010), passou-se apenas meio século.

Tamanha aceleração do crescimento tecnológico, que talvez seja exponencial, acarreta consigo profundas transformações não só nas tecnologias que os indivíduos usam para comunicar entre si, como também modificam o que é comunicado, quando e como. Nas últimas décadas, tem-se assistido a uma verdadeiramente agitada tempestade tecnológica: das cartas manuscritas para o *email*; da máquina de escrever para o computador; do microfilme para a disquete, para o CD, para o DVD, para a *pen drive*; dos grandes livros enciclopédicos para a Wikipédia; das notas para as criptomonedas. Em suma, do analógico para o digital. Os ventos desta tempestade têm tornado difícil para as áreas que lidam com a informação — como a Arquivística e a Ciência da Informação — encontrar o seu rumo, uma vez que o modo como a informação é produzida, armazenada e comunicada depende diretamente das infraestruturas tecnológicas existentes para o efeito.

Por este motivo, à medida que a sociedade reinventa a transmissão da informação no espaço e no tempo, também estas áreas da informação têm de procurar visitar os seus princípios nucleares e questionar as suas práticas para, no mínimo, tentar perceber se as práticas de uma época e mundividências anteriores continuam a ser relevantes nos novos contextos. Porém, colocar a questão nestes modos “o que eu fazia dantes continua a fazer sentido agora?” é perpetuar a falácia da subjetividade, através da qual se parte para a procura de uma solução com uma hipótese previamente preferida. Uma abordagem mais neutra seria: “qual a prática que mais faz sentido neste determinado contexto?”, e é este caminho que se pretende seguir neste trabalho, tanto quanto possível.

Com esta mentalidade, investiga-se e questiona-se a literatura recente, de modo a procurar perceber, sob uma perspetiva da Arquivística, que se entende ser uma disciplina enquadrada na Ciência da Informação, o que caracteriza o período atual, que Manuel Castells (2002) defendeu ser a Era da Informação, tamanha é a importância da informação nos mais diversos setores da sociedade. Com efeito, tem sido reconhecido tanto por académicos (Park, 2017) como por autores literários (Toffler, 1990) que informação é poder. Os documentos de arquivo exercem esse poder ao registarem atos, testemunhando atividades, fazendo valer direitos, constituindo prova e memória. No entanto, da mesma forma que Foucault (1980) argumenta que o poder não existe num vácuo, podemos também argumentar que a informação, enquanto uma forma de poder, não existe fora do seu contexto. Antes, porque somos seres intrinsecamente sociais, os nossos produtos materiais são necessariamente socializados, mediados pela linguagem e pela consciência e modificados pela atividade social (Yardley, 2002).

Analogamente, a representação da informação arquivística — simplificando, descrição arquivística — entendida como referente aos processos tradicionais de descrição, elaboração de resumos, indexação e classificação, é vista como a *fluid, evolving, and socially constructed practice* (Yakel, 2003, p. 2). Assim, de modo a ser capaz de usar e interpretar os documentos de arquivo, é necessário perceber o contexto que governou a sua criação, respondendo a questões tais como o que são, quem os criou, quando, como, e com que finalidade. No ambiente digital, acredita-se que a melhor forma de explicitar o contexto é através de metadados, assentes na adição de camadas extensíveis de informação e de significado aos documentos.

Para os arquivistas, uma informação crucial a veicular através da descrição arquivística diz respeito à autenticidade dos documentos de arquivo. A importância deste aspeto reflete-se na sua eleição como uma das raízes do código de ética dos arquivistas: a garantia da autenticidade dos documentos sob custódia (International Council on Archives, 1996; Society of American Archivists, 2012). A presunção da autenticidade dos documentos de arquivo assenta em estabelecer a identidade e demonstrar a integridade. A descrição arquivística, entendida enquanto a descrição dos documentos sob custódia arquivística, foi identificada como um dos procedimentos necessários para atestar a autenticidade dos documentos de arquivo digitais (InterPARES 1, 2002), na medida em que explica a classificação dos documentos e as circunstâncias em que foram criados (International Council on Archives, 2016).

O elo entre a autenticidade e a descrição arquivística é visível nas normas de descrição. Por exemplo, na revisão de 2008 das *Rules for Archival Description* do Canadá, é afirmado que um dos propósitos da descrição arquivística é *to establish grounds for presuming the authenticity of archival material by documenting its chain of custody, arrangement, and circumstances of creation and use* (Canadian Council of Archives, 2008, p. xxii). Porém, no que diz respeito às normas internacionais de descrição arquivística, o Conselho Internacional de Arquivos (ICA) tem reconhecido que as atuais normas são inadequadas para a representação e a promoção do acesso a recursos digitais, visto que assentam predominantemente num modelo hierárquico, baseado na

noção de fundo, que contrasta com a abordagem não-direcionada da descoberta de recursos na *web* (International Council on Archives, 2016).

Através da explicação da classificação dos documentos de arquivo e das circunstâncias em que estes foram criados, as descrições arquivísticas comunicam informação relevante ao utilizador para o entendimento sobre a identidade e a integridade dos documentos. Porém, uma parte significativa das normas de descrição arquivística foram redigidas num período predominantemente voltado para uma gestão analógica da documentação, que por vezes se torna difícil de compatibilizar com as novas necessidades de representação. Tais necessidades são criadas pela invenção ou adoção de tecnologias de representação mais recentes, como o *Resource Description Framework*, para relacionar os dados no contexto da *Web Semântica* (Berners-Lee, Hendler, & Lassila, 2001), possibilitando a descoberta automática de recursos e uma interpretação do contexto mais aprofundada.

Assim, notando a inadequação das atuais normas para a representação radial da informação ao mesmo tempo que assistimos à implementação de novas tecnologias para a representação da informação, formula-se as seguintes questões de investigação:

- Tendo em conta o crescente e recente movimento da representação da informação usando modelos de dados em grafo no contexto da *Web Semântica*, (i) quais as vantagens e desvantagens desta forma de descrição quanto à verificação da autenticidade dos documentos de arquivo? (ii) Qual o paradigma de descrição arquivística subjacente a estas formas de representação e acesso à informação? (iii) Será possível construir um modelo teórico que identifique os requisitos para avaliar a autenticidade e determinar a confiabilidade dos documentos de arquivo que possa ser seguido e aplicado no contexto da *Web Semântica*? (iv) Quais os elementos de metadados que esse modelo deve incluir? (v) Quais são os potenciais benefícios deste modelo de metadados, em termos de descoberta da informação e completude das descrições, quando comparado com os atualmente predominantes instrumentos de descrição arquivística?

De modo a responder a estas questões, centra-se a investigação na Ciência da Informação, em particular na Arquivística, ao mesmo tempo que se delimitam como objeto de estudos os metadados usados na descrição arquivística digital, com especial enfoque nos metadados relacionados com a autenticidade e a confiabilidade dos documentos de arquivo. Desta forma, formula-se o seguinte objetivo geral:

- Explorar a literatura científica e técnica de modo a determinar quais os conceitos teóricos sobre os quais assenta a presunção da autenticidade e da confiabilidade dos documentos de arquivo, e quais os elementos de metadados para a autenticidade que uma norma de metadados deve incluir para a representação dinâmica da informação arquivística num contexto *web*, com base num modelo de dados em grafo.

Deste objetivo geral radiam três objetivos mais específicos:

- a) explorar sistemática e exaustivamente a literatura científica mais recente sobre o tópico da representação da informação em arquivos, consubstanciada na descrição arquivística.
- b) comparar a literatura técnica e normas internacionais de gestão de informação e de descrição arquivística de modo a averiguar quais os metadados ou procedimentos recomendados para estabelecer a identidade e determinar a confiabilidade dos documentos de arquivo.
- c) elaborar um modelo de metadados que devem estar presentes para a representação autêntica dos documentos de arquivo em contexto.

A propósito do primeiro objetivo específico, a discussão proposta engloba temas como a compreensão e a discussão de metadados; os paradigmas de descrição arquivística; a importância da autenticidade para a gestão da informação; o papel das descrições arquivísticas na garantia da autenticidade e a crescente adoção de tecnologias para a representação da informação em contexto, como parte da visão da *Web Semântica*. Espera-se que o cumprimento do objetivo geral e específicos desemboque nos seguintes resultados:

- a) a compreensão do papel da descrição arquivística no processo de averiguação da autenticidade dos documentos de arquivo e/ou das próprias representações desses documentos;
- b) o desenvolvimento de um modelo de metadados que inclua os elementos identificados na literatura como desejáveis para a representação da informação arquivística autêntica no contexto da *Web Semântica*.

Neste sentido, a natureza da investigação para responder a estas questões assenta numa metodologia de investigação mista, inserida no paradigma interpretativo, segundo o qual o investigador é um construtor crítico do sentido do texto. A nível de métodos, adota-se a investigação documental enquanto método qualitativo, e a engenharia de requisitos enquanto método quantitativo.

A recolha de dados é guiada pela investigação documental, através do qual se constrói um *corpus* documental com características específicas. A nível de tipologias documental, procuram-se revistas científicas encontradas na base de dados *Library, Information Science & Technology Abstracts, peer-reviewed*, bem como as respetivas monografias cujas resenhas se encontram nessas mesmas revistas. Quanto às comunicações científicas, exploram-se aquelas visíveis na base de dados de referência *Web of Science*. As teses e dissertações são consultadas a partir do *Open Access theses and dissertations* e do *Repositório Científico de Acesso Aberto em Portugal*. Por fim, as normas de metadados são selecionadas com base em parâmetros como tema e comunidade de origem. As obras recolhidas são localizadas temporalmente entre 2009 e 2020.

No caso das revistas científicas e das monografias, a seleção dos textos ocorreu mediante uma análise individual de cada entrada para o período referido, sendo a sua pertinência e a sua adequação

aos temas de investigação decididos de forma individual, com base em critérios subjetivos do investigador.

Acerca das restantes tipologias documentais, a pesquisa incluiu várias palavras-chave, em diferentes línguas, ocorrentes no título, no *abstract* e nas palavras-chave dos textos, de modo a obter os que, à partida, poderão ser mais pertinentes para responder às questões de investigação. A respeito das normas de metadados, procurou-se escolher normas de estrutura, enquadradas nos setores informacionais (arquivos, bibliotecas e museus), redigidas nas línguas eleitas e disponíveis de forma gratuita na Internet. De modo a filtrar ainda mais a seleção, procurou-se escolher um conjunto de normas que refletisse as práticas de diversos continentes e comunidades.

Este trabalho reparte-se por cinco capítulos. No primeiro e segundo capítulos apresenta-se a revisão de literatura, que numa primeira parte aborda o uso de metadados em arquivos, começando pela conceptualização deste termo antes de abordar as várias etapas do uso de metadados ao longo da história. É também realizada uma análise aos vários tipos e funções que os metadados podem assumir, bem como aos diversos modelos de expressão dos metadados. Nesta primeira parte inclui-se ainda uma discussão terminológica do que constitui uma norma e um modelo, e salienta-se a importância da interoperabilidade.

No segundo capítulo, a segunda parte da revisão de literatura, procede-se a uma reflexão sobre o conceito de representação, focando de seguida a representação da informação em arquivos. Daqui parte-se para uma análise do conceito de autenticidade e de confiabilidade das representações arquivísticas, um dos conceitos chave desta investigação.

No terceiro capítulo encontra-se a abordagem metodológica que norteia a investigação. Após a exposição das motivações, dos problemas, dos objetivos e das limitações do trabalho, passa-se para a caracterização da natureza da investigação, elegendo os métodos de recolha de dados e os critérios de constituição do *corpus* documental. Detalham-se ainda os métodos de análise de dados e de seleção das fontes de informação, dentro do *corpus*.

O quarto capítulo pauta-se pela identificação dos requisitos que servem posteriormente de base para a proposta de um modelo para a autenticidade e para a confiabilidade dos recursos, sendo por isso um dos capítulos mais singulares na estrutura do trabalho. Ao longo das páginas deste capítulo analisam-se e interpretam-se em detalhe os requisitos encontrados na literatura científica, organizada de acordo com categorias analíticas de modo a facilitar e a enriquecer a crítica. Também a literatura técnica, composta pelas normas de metadados, é percorrida em detalhe, analisando-se o contexto histórico e a estrutura de cada norma, antes de eleger os metadados que, de acordo com a leitura dos investigadores, mais se adequam à captura da identidade e da integridade dos recursos descritos.

O quinto capítulo surge como um prolongamento natural do quarto capítulo, no sentido em que os requisitos identificados previamente na literatura técnica e na literatura científica são analisados, discutidos e integrados na proposta de um modelo caracterizado por um conjunto de elementos de metadados para a descrição arquivística de recursos autênticos e fidedignos no contexto digital. Cada elemento do modelo é caracterizado através da explicitação da definição do

seu uso, conseguindo assim dar-se resposta aos objetivos de investigação propostos. Na conclusão são recapituladas as principais ideias a reter da revisão de literatura e responde-se às questões de investigação lançadas inicialmente.



# 1 Metadados em arquivos

---

## 1.1 Conceptualização

### 1.1.1 A importância dos metadados

Nos anos recentes, os dados têm sido considerados o ‘novo petróleo’ do século 21. O seu valor monetário e político é enorme, já que cada vez mais as organizações ganham uma vantagem competitiva sobre os seus concorrentes ao possuírem mais dados/informação sobre o mercado e os seus clientes. No entanto, a maior força da metáfora reside no facto de que, tal como o petróleo, é necessário refinar os dados para deles extrair valor: *in its raw form, data is just like crude oil; it needs to be refined and processed in order to generate real value. Data has to be cleaned, transformed, and analyzed to unlock its hidden potential* (TiECON East, 2014). Sem entrar em detalhes, este raciocínio aproxima-se da pirâmide do conhecimento de Ackoff (1989), segundo a qual se transita de dados para informação, desta para conhecimento e de conhecimento para sabedoria, correspondendo a cada nível um maior processamento e assimilação pelo entendimento humano da matéria-prima que constitui o nível anterior.

Os metadados são dados sobre dados, camadas adicionais de informação que contextualizam os dados. A importância dos metadados quando sujeitos a um tratamento e controlo eficazes foi deixada bem clara em 2013, após as denúncias de Edward Snowden sobre as atividades da Agência Nacional de Segurança dos Estados Unidos da América (NSA, *National Security Agency*). Snowden, um antigo funcionário da *Central Intelligence Agency* (CIA) dos Estados Unidos da América (EUA), chocou o mundo ao revelar que a NSA estava a monitorizar as atividades das contas de *email* e dos telefones de milhares de pessoas em todo o planeta, numa rede tão extensa que incluía muitos dos líderes mundiais de outros países. A presidente alemã Angela Merkel foi uma das personalidades envolvidas nesta rede de espionagem, que comparou a NSA à Stasi, a antiga polícia secreta da Alemanha Oriental considerada uma das organizações mais repressivas alguma vez existentes (Sanger & Smale, 2013).

Ao contrário do que possa ser crença comum, na maior parte dos casos a NSA não estava a ouvir o conteúdo das chamadas nem a ler individualmente os *emails* e mensagens de texto, mas antes a recolher dados sobre eles (Joe Rogan Experience, 2019, 1h34m40s). O próprio presidente Obama clarificou publicamente que *nobody is listening to your phone calls* (Obama, 2013, 01m00s). Obama usou o termo *metadata* para se referir a detalhes como o número de telemóvel usado, a localização, o número de chamadas realizadas, a duração das chamadas, os destinatários e a data, tendo desde então o termo entrado na esfera pública. Por esta ocasião, foram tecidas críticas a Snowden com base no argumento de que, como a NSA apenas recolhia metadados, a monitorização não era intrusiva, constituindo uma cedência perfeitamente aceitável em nome da segurança nacional (British Broadcasting Corporation, 2013). A discussão de segurança nacional *versus* privacidade não foi exclusiva dos EUA. No Reino Unido, o *Investigatory Powers Act* (UK

Parliament, 2016) determina que os fornecedores de telecomunicações devem manter metadados sobre as comunicações ocorridas nas redes públicas de modo a facilitar a monitorização e investigações criminais que atentem contra a segurança nacional.

Todavia, os metadados são muito mais do que ‘apenas metadados’. Em abril de 2014, o general Michael Hayden, antigo diretor da NSA e da CIA, assumiu num painel de debate na Universidade Johns Hopkins que *we kill people based on metadata* (Johns Hopkins University, 2014, 18m01s). Esta afirmação surge como resposta a uma citação de Stewart Baker, antigo consultor geral da NSA, que afirmara que *metadata absolutely tells you everything about somebody's life. If you have enough metadata, you don't really need content. It's sort of embarrassing how predictable we are as human beings* (Johns Hopkins University, 2014, 14m55s). Os metadados, quando em contexto, são extremamente reveladores do pensamento e comportamento de um indivíduo. Sendo os humanos animais sociais, e sendo os metadados a pegada das nossas interações sociais, uma análise profunda e detalhada do registo de metadados criados por uma pessoa constitui uma espécie de identificador único de cada indivíduo. Todas as nossas compras na Amazon, todas as pesquisas no Google, todos os nossos levantamentos no multibanco, todas as chamadas telefónicas que fazemos, para quem as fazemos, a duração das chamadas, todos os locais onde estivemos, com quem estivemos, tudo fica registado. No caso das agências de inteligência americanas, se um indivíduo mantém telefonemas com números registados como extremistas e se faz pesquisas no Google sobre temas extremistas, estes metadados constituem informação suficiente para colocar essa pessoa sob escrutínio. Por outras palavras, quando os metadados de diferentes fontes são cruzados, podem ser usados para reconstruir o conteúdo de comunicações com enorme precisão.

Durante o quotidiano, os metadados estão presentes em múltiplas das nossas interações com a sociedade. Além das comunicações, conforme visto, constituem uma parte essencial das aplicações de telemóvel, dos mapas, dos *websites*, dos catálogos, dos fóruns e das transações. Jeffrey Pomerantz expressa o uso de metadados de uma forma elegante: *metadata is a means by which the complexity of an object is represented in a simpler form* (Pomerantz, 2015, p. 11). Por exemplo, os metadados indicam-nos que o autor de *A cidade e as serras* é Eça de Queirós, que se trata de um romance publicado em 1901, mas não discorre sobre as complexidades semânticas e sociais abordadas no livro. Os metadados atuam neste caso como uma simplificação de uma realidade complexa que, desde que cumpra o seu objetivo — neste caso, a identificação do autor de um livro — é perfeitamente legítima.

Porém, enquanto simplificação da realidade, os metadados constituem uma escolha intencional, uma representação seletiva que ignora uma parte do que poderia ser dito em favor de outra considerada de maior pertinência. Por este motivo, é importante realçar que os metadados são *a human construct not found in nature* (Gartner, 2016, p. 4). A sua escolha destina-se a resolver um determinado problema, e a forma que estes assumem depende do contexto do seu criador. Gartner ilustra esta ideia com o exemplo da representação do mundo num globo ou no *Google Maps*. Embora a realidade a representar seja a mesma – a Terra –, o produto final das representações

é extremamente distinto. No caso do globo, são artificialmente acrescentadas linhas longitudinais e latitudinais, são delimitados países, e esses países são denominados numa determinada linguagem. No caso do *Google Maps*, é dada maior importância à representação das vias e dos serviços, uma vez que a sua principal função é auxiliar a navegação, mas os restaurantes que são colocados no mapa podem ser aqueles que pagaram para aparecer. Pode acrescentar-se também que, curiosamente, nas representações do planeta, o continente que figura ao centro é geralmente aquele do qual os autores são provenientes, como mapas criados por europeus tendo a Europa ao centro. Estes exemplos deixam clara a intencionalidade humana na criação de metadados, pelo que estes devem ser geralmente entendidos como uma representação subjetiva e simplificada de uma realidade mais complexa.

O primeiro uso do termo *metadata* remonta a 1968 quando Philip Bagley o usou no contexto da Ciência da Computação, tendo o termo sido rapidamente adotado pelos profissionais desta área. Contudo, o uso de metadados é muito anterior à existência de computadores, sendo observável desde que há listas e catálogos ou, por outras palavras, desde que há a história registada por escrito, e afirmam-se como um dos cimentos unificadores da cultura. A invenção da escrita permitiu superar os limites da memória humana, possibilitando a preservação da memória. Os antropólogos Jack Goody e Ian Watt observaram em 1963 que, pela primeira vez, a escrita permitiu que a cultura fosse transmitida de forma quase intacta entre gerações (Goody & Watt, 1963). Uma vez escrito um testemunho — como um livro —, é necessário um sistema de organização do conhecimento que nos permita saber o que foi escrito e qual a sua localização de modo a poder consultar, um sistema unificador destes ‘fragmentos de informação’. Gartner argumenta que os metadados possibilitam que estes fragmentos sejam *linked together to form knowledge and for this knowledge to be consolidated into what we understand as culture* (Gartner, 2016, p. 9). Porém, antes de explorar o uso de metadados ao longo das culturas, importa primeiro explicar o conceito.

### 1.1.2 Definição e discussão

A explicação da etimologia constitui sempre um bom ponto de partida para a análise de um conceito. No caso de ‘metadados’, o prefixo *μετά-* deriva do grego antigo e assume geralmente a interpretação da proposição ‘sobre’. No *Dicionário Priberam da Língua Portuguesa*, na quarta aceção, é dito que *exprime a noção de reflexão sobre si* (Dicionário Priberam da Língua Portuguesa, 2020a), sendo dado o exemplo de uma metalinguagem como uma linguagem que descreve outra linguagem. Assim, metadados são, na sua aceção mais elementar, dados sobre dados.

Este termo terá sido usado pela primeira vez no âmbito da ciência da computação em 1968, no artigo *Extension of Programming Language Concepts* da autoria de Philip Bagley sob a égide da Força Aérea dos Estados Unidos. Nesse artigo no qual são abordados modelos para criação de linguagens de programação, este autor afirma que uma das componentes essenciais de uma linguagem de programação é:

*the ability to associate explicitly with a data element a second data element which represents data 'about' the first data element. This second data element we might term a 'metadata element'* (Bagley, 1968, p. 26).

A definição de Bagley entende os metadados como uma abstração usada para descrever dados, dando o exemplo de identificadores de dados os conteúdos de um elemento numa tabela ou os próprios parâmetros segundo o qual o elemento deva ser expresso (Zappavigna, 2018, p. 20). Não há dúvidas de que estamos perante o papel de metadados na sua função coeva.

Atualmente, o termo 'metadados' é usado em grande parte das áreas científicas, desde o desenvolvimento de drones à Ciência da Informação, passando pelas áreas da Inteligência Artificial. Dado o seu uso generalizado pelos setores da sociedade, o entendimento deste conceito não é consensual, dependendo da perspetiva e da intenção do seu uso. Por exemplo, Marcia Zeng e Jian Qin (2016, p. 4) defendem que a informação nutricional dos alimentos deve ser considerada metadados porque constitui a formulação de uma propriedade que esclarece o significado de determinado valor de dados (*data value*). Para ilustrar esta ideia, os mesmos autores dão o exemplo de como a propriedade 'calorias' dota de significado o valor de dados '260'. Por sua vez, Haynes (2018, p. 10) discorda por considerar que apenas se pode falar em metadados quando o objeto descrito é um recurso informacional. Segundo este autor, no caso dos rótulos nutricionais, estaríamos perante dados, e não metadados.

O cerne desta discussão diz respeito ao objeto representado, se qualquer dado sobre outro dado, ou apenas se recursos informacionais. De entre estas posições, concorda-se com a visão alargada do objeto conforme formulada por Zeng e Qin, por se partilhar da opinião de que o conceito de metadados não se aplica exclusivamente a recursos informacionais, mas de uma forma generalizada a qualquer dado que expresse uma propriedade sobre outro dado. Em todo o caso, no contexto deste trabalho, a discussão sobre os metadados é centrada no âmbito da Ciência da Informação, pelo que o objeto da representação por metadados são objetos portadores de informação (Gilliland, 2016; Gilliland et al., 2008; Haynes, 2018; Pomerantz, 2015; Zeng & Qin, 2016). Neste caso, os metadados, enquanto representações, podem ser usados para funções como a descoberta, a recuperação dos conteúdos e, como veremos, para a asserção da autenticidade.

Um primeiro passo para a elaboração do conceito de metadados é entendê-lo como *structured data about other data* (Gladney, 2007, p. 129). A noção de estrutura é fundamental, e abundantemente ecoada na literatura (Clobridge, 2010; Cron, 2016; Daniel & Daniel, 2012; Gilliland, 2016; ISO/IEC TR 19583-1, 2019; NISO, 2017; Witten, Bainbridge, & Nichols, 2010). A *Digital Preservation Coalition* aprofunda a simples noção de estrutura acrescentando que *metadata é data about a digital resource that is stored in a structured form suitable for machine processing* (Digital Preservation Coalition, 2015, p. 56). Destaca-se nesta definição a necessidade de essa estrutura ser expressa numa forma processável de forma automática por máquinas, um requisito essencial para a interoperabilidade, partilha e disseminação de dados. Por sua vez, ser processável por máquinas (*machine-actionable*) compreende duas aceções: a interpretação dos metadados contextuais de um objeto e a interpretação do conteúdo (Wilkinson et al., 2016, p. 4).

Assim, central ao conceito de metadados são as noções de dotar os dados de contexto e de estrutura numa forma processável de forma automática. Aliás, conteúdo, contexto e estrutura são características essenciais dos documentos de arquivo (Duranti, 1997; International Council on Archives, 1997; Janes, 2012), pelo que essas características se refletem nos metadados que os representam (Apostolou, 2009; Cron, 2016; Gilliland, 2016; The National Archives of the UK, 2002), entendidas do seguinte modo:

- *Content relates to what the object contains or is about and is intrinsic to an information object.*
- *Context indicates the who, what, why, where, and how aspects associated with the object's creation and subsequent life and is extrinsic to an information object.*
- *Structure relates to the formal set of associations within or among individual information objects and can be intrinsic, extrinsic, or both.*(Gilliland, 2016, p. 2).

Numa definição de 1999, metadados são apontados como sendo:

*any data that aids in the identification, description and location of networked electronic resources ... Another important function provided by metadata is control of the electronic resource, whether through ownership and provenance metadata for validating information and tracking use; rights and permissions metadata for controlling access; or content ratings metadata, a key component of some Web filtering applications* (Hudgins, Agnew, & Brown, 1999, p. 1).

Na introdução à obra *Metadata: a cataloger's primer*, Richard Smiraglia avança com uma definição que inclui a descoberta e a gestão de recursos informacionais:

*metadata are structure, encoded data that describe the characteristics of information-bearing entities to aid in the identification, discovery, assessment and management of the described entities* (Smiraglia, 2005, p. 4).

Contudo, os próprios metadados carecem de uma estrutura onde possam ser registados. Essa estrutura consiste num registo de metadados (*metadata record*), entendido de forma geral como um conjunto de metadados que representam um recurso, que inclui semântica, sintaxe e estrutura (Lubas, Jackson, & Schneider, 2013). Jeffrey Pomerantz salienta que, da mesma forma que um livro é um *container for data but is not data itself ... metadata is data, but metadata cannot exist outsider of a container: a metadata record must exist in some format, be it physical or digital. Likewise a metadata record is itself a container for data about an object* (Pomerantz, 2015, p. 22). Marcia Zeng e Jian Qin denominam estes contentores de metadados de *metadata descriptions* com uma conotação idêntica de *vehicles that carry the information about objects or things* (Zeng & Qin, 2016, p. 28). Reforça-se assim a importância da noção da necessidade de um conteúdo para os metadados, conteúdo esse que lhes confere uma forma na medida em que indica quais os metadados a incluir como representação do objeto informacional.

A ideia de metadados como acrescentando estrutura é central, embora não seja a única característica exclusiva. Além desta, os metadados também desempenham um papel fundamental em dotar os dados de contexto interno e externo. Entende-se o contexto interno como o conjunto de elementos de dados usados para a representação e o seu grau de obrigatoriedade (obrigatório, recomendado ou opcional). O contexto externo corresponde à teia de relações estabelecida por determinado recurso, que inclui os processos administrativos de que decorre, mas que se estende a outros recursos que estabeleçam qualquer relação com o recurso representado.

Esta distinção inspira-se numa perspetiva diplomática, segundo a qual os documentos de arquivo se desdobram em elementos externos — atos, pessoas e procedimentos — e internos — forma documental (Duranti, 1997, p. 214). Assim, o contexto interno corresponde largamente à estrutura interna, isto é, à forma documental, no caso dos metadados entendida como a estrutura dos elementos num registo de metadados. Por sua vez, entende-se a estrutura externa, identificada na teoria arquivística<sup>1</sup> com o vínculo arquivístico (*archival bond*), como o contexto externo dado pelo vínculo informacional (que traduzimos em inglês por *informational bond*), mais abrangente do que o *archival bond* porque não é limitado pelos limiares das relações com outros documentos de arquivo, com as atividades de que resultam e pelos seus produtores.

Duranti (1997, p. 217) defende que o *archival bond* não deve ser confundido com o termo genérico ‘contexto’, uma vez que o contexto é externo aos documentos, ao passo que o *archival bond* é intrínseco. Duranti emprega o termo ‘externo’ numa aceção diplomática, no sentido de não-diretamente presente na redação destes. Porém, como os metadados são por natureza externos ao documento de arquivo, enquanto representações deste, o vínculo informacional atesta não a estrutura externa, mas o contexto externo<sup>2</sup>, que inclui as pessoas, os atos e os procedimentos envolvidos na criação de um registo de metadados, bem como outros recursos que estabeleçam qualquer relação. A ideia de vínculo informacional inclui ainda a noção de que o contexto pode mudar e, com ele, mudam o significado e a interpretação do recurso. Este raciocínio é apoiado pela *American National Archives and Record Administration* (NARA) quando salienta que, além de informação sobre o conteúdo e a estrutura dos documentos de arquivo, *metadata elements also provide contextual information that explains how electronic records were created, used, managed and maintained ... and how they are related to other records* (NARA, 2015, section 2).

A interpretação muda porque o contexto não é fixo. Os metadados são criados por certas pessoas num certo período, com uma parcialidade inerente. A dimensão subjetiva da representação já foi mencionada anteriormente, e que Gartner deixa bem clara em relação aos metadados:

---

<sup>1</sup> Adota-se o entendimento de ‘teoria’ num sentido geral de *any idea used in explaining the world* (Slife & Williams, 1995, p. 16).

<sup>2</sup> Duranti identifica quatro tipos de contexto, do geral para o particular: o contexto jurídico-administrativo, o contexto providencial, o contexto procedimental e o contexto documental (Duranti, 1997, p. 217).

*There is nothing objective about metadata: it always makes a statement about the world, and this statement is subjective in what it includes, what it omits, where it draws its boundaries and in the terms it uses to describe it (Gartner, 2016, p. 4).*

A noção de metadados como uma *statement* (asserção) é não só recorrente na literatura, como também necessária para entender que os metadados não são objetivos porque são uma construção humana e nada que é humano é neutro. Elizabeth Yakel afirma que *through the process of selection of information for inclusion and choice of access points, archivists reveal and conceal, making finding aids political statements* (Yakel, 2003, p. 19), estendendo a subjetividade aos instrumentos de pesquisa, enquanto produtos da representação arquivística. De forma paralela, Michael Cook defende que, ao organizar os documentos, os arquivistas *are providing a statement on meaning and authenticity* (M. Cook, 1999, p. 102).

A ideia de metadados como asserções é também veiculada pelo projeto InterPARES no contexto de metadados para a gestão documental (*recordkeeping metadata*):

*They are assertions about a resource that are both human and machine-readable. An assertion in this context is a statement that contains a reference to the resource, a property of the resource, and a value; and our resource may be of two types: a record or a group of records. Properties are what we want to say about records – what InterPARES calls the attributes of records—and value is the way we say it (InterPARES 2 Project, 2008a, pp. 8-9).*

Esta formulação de *statements* enquanto a expressão das propriedades e do valor de um recurso está também profundamente enraizada na arquitetura do *Resource Description Framework* (RDF), uma das tecnologias mais recentes e versáteis de representação de dados. Segundo este modelo de dados, os metadados são expressos sob a forma de *statements* que incluem um sujeito e as relações que esse sujeito estabelece com um objeto (W3C, 2014b), por exemplo uma relação de autoria ou parentesco.

Assim, pode afirmar-se que *metadata is a statement about a potentially informative object* (Pomerantz, 2015, p. 26). A ideia de objetos portadores de informação é interessante e subscrita por vários autores. Segundo Zeng e Qin, metadados são *the structured, encoded data that describes the characteristics of information-bearing entities (i.e., things)* (Zeng & Qin, 2016, p. 3). Destaca-se o uso de *entities*, um termo deliberadamente genérico que permite a inclusão de entidades como pessoas, que podem constituir informação de acordo com o papel que assumem (enquanto cliente, autor, paciente). Marcia Zeng aborda ainda o uso *information-bearing objects* no contexto de *smart data* (Zeng, 2017). Anne Gilliland também associa metadados a objetos informacionais:

*Perhaps a more useful, ‘big picture’ way of thinking about metadata is as the sum total of what one can say about any information object at any level of aggregation. In this context, an information object is anything that can be addressed and manipulated as a discrete entity by a human being or an information system (Gilliland, 2016 p. 2).*

A definição de metadados usada no contexto do projeto InterPARES aproxima-se da de Gilliland no sentido em que reforça a possibilidade de associação com um sistema ou objeto de informação. De acordo com esta formulação, os metadados são *data associated with either an information system or an information object for purposes of description, administration, legal requirements, technical functionality, use and usage, and preservation* que desempenham *a critical role in ensuring the creation, management, preservation, discovery, use and re-use of trustworthy materials, including records* (Gilliland et al., 2008, p. 1). Esta definição introduz também alguns dos vários propósitos atribuídos aos metadados na literatura. Embora os metadados sejam descritivos por natureza, eles podem desempenhar diversas funções de acordo com o seu uso específico, conforme será discutido na secção 1.3.

Em resumo, corrobora-se com a explicação elucidativa de Lars Marius Garshol:

*Metadata is generally defined as ‘data about data’, which is of course a very broad definition. In computer science this is generally taken to mean information about a set of data in a particular representation, which typically means schema information, administrative information, and so on. However, in content management and information architecture, metadata generally means ‘information about objects’ ... that is, information about a document, an image, a reusable content module ... In general, metadata is best understood as ‘any statement about an information resource’, regardless of what it is being used for, which metadata vocabulary is being used, and how the metadata is represented* (Garshol, 2004, p. 379).

Importa, contudo, tecer algumas observações a propósito da transição de ‘dados sobre dados’ para ‘dados sobre objetos informacionais’. Para tal, é necessário distinguir ‘dado’ de ‘informação, uma distinção que assenta na hierarquia DICS (dado-informação-conhecimento-sabedoria, ou *data-information-knowledge-wisdom*), na versão formulada por Russell Ackoff quando aceitou a presidência da *International Society for General Systems Research* em 1989. Existe uma grande variedade e um notável volume de bibliografia sobre esta temática, que lança profundas raízes na filosofia. Por este motivo, apenas se trazem para esta discussão os conceitos considerados fundamentais.

Assim, de acordo com a proposta de concetualização de Ackoff, *data are symbols that represent the properties of objects and events. Information consists of processed data, the processing directed at increasing its usefulness* (Ackoff, 1989, p. 3). Por conseguinte, a diferença entre dados e informação reside no processamento com um fim que os dote de utilidade. Como tal, se dados são *facts or numbers in a general sense* (Baca, 2016, p. 74), informação consiste em dados processados dotados de uma utilidade, isto é, trata-se do valor que extraímos da informação<sup>3</sup>

---

<sup>3</sup> O conceito de informação é também colocado noutros campos no centro do processo comunicacional entre o emissor e o receptor de uma mensagem (Buckland, 2012; Capurro & Hjørland, 2003; Shannon & Weaver, 1963; A. M. da Silva, 2009; Wersig & Neveling, 1975; Zins, 2007).



(Weinberger, 2010). Esta interpretação é consistente com a apreciação de Ackoff, segundo a qual *the difference between data and information is functional, not structural* (Ackoff, 1989, p. 3).

A hierarquia, frequentemente representada em forma de pirâmide com os dados na base, enquanto elemento mais abundante e menos refinado, inclui também os conceitos de conhecimento (*knowledge*) e sabedoria (*wisdom*), correspondendo a cada transição um maior nível de processamento. Sem entrar em detalhes, *digital codified knowledge* ou, simplificada mente conhecimento, é entendido como *digital information which has been organized, and has relations or links that put it into the correct context* (Kärberg, 2014, p. 127). Segundo esta interpretação, o conhecimento é materialmente intangível porque se trata do *product of a synthesis in the mind of the knowing person* (Zins, 2007, p. 479), ao passo que a sabedoria se caracteriza por um julgamento assente em valores. Ackoff (1989, p. 4) considera ser esta a característica que distingue os homens das máquinas. Importa acima de tudo salientar para a discussão a crescente escala de extração de valor que corresponde a cada componente da hierarquia, sobretudo em considerar a informação como o valor acrescentado atribuído aos dados.

Assim, a definição previamente discutida de *information-bearing objects* pode ser entendida como objetos cujos dados foram processados com um determinado fim e que constituem valor acrescentado quando representados e dispostos diante do entendimento de um indivíduo, ou seja, dados dotados de significado. Convém ainda destacar que se aceita esta noção no âmbito da arquivística, área em que se desenvolve este trabalho. Neste contexto, tais objetos podem corresponder predominantemente a documentos de arquivo, entendidos numa aceção clássica enquanto *recorded information produced or received in the initiation, conduct or completion of an institutional or individual activity and that comprises content, context and structure sufficient to provide evidence of the activity* (International Council on Archives, 1997, p. 21). Contudo, os metadados podem ser usados para descrever qualquer objeto — não apenas documentos de arquivo — que possa constituir informação, isto é, qualquer outra entidade que estabeleça relações com os documentos de arquivo e que são alvo de representação, tais como pessoas, *websites*, cargos, ocupações.

Por exemplo, num sistema de representação de um arquivo pessoal de Eça de Queirós, pode ser expresso que ‘Eça de Queirós’ é o autor de determinado manuscrito. Porém, pode também ser expresso que ‘Eça de Queirós’ tem uma relação de matrimónio com ‘Emília de Castro’, e uma relação de cargo desempenhado com ‘cônsul’. Nestes casos, o manuscrito é o clássico *information-bearing object* em arquivística; contudo, ‘Emília de Castro’ e ‘cônsul’ assumem-se também como objetos portadores de informação, passíveis de representação por metadados. Neste sentido, expande-se o entendimento de *information-bearing object* ou *information object* para qualquer recurso (*resource*), físico ou conceptual, passível de representação num sistema. À medida que as tecnologias de *Linked Data* vão assumindo cada vez mais preponderância, o entendimento de metadados como descrevendo documentos de arquivo deve começar a dar lugar à noção de metadados enquanto *statements* ou triplos (Zeng & Qin, 2016, p. 348). A constituição de um triplo será abordada mais tarde, quando se analisarem as tecnologias em grafo.

Por fim, compete distinguir ainda a ocorrência e o significado do termo ‘metainformação’. Este vocábulo, permutável com ‘metadados’, é por vezes observado na literatura. A nível internacional, por exemplo, foi usado por Tim Berners-Lee, considerado o fundador da *World Wide Web*, em 1997 (Berners-Lee, 1997), ocorrendo também em literatura de modelação de dados (Nart, Degl’Innocenti, Peressotti, & Tasso, 2017) e ao nível da gestão da informação (Eysenbach, Yihune, Lampe, Cross, & Brickley, 2000; Losee, 2014; Middleton, 1999; Oparina, 2012). A nível nacional, no contexto português, o termo é mais comum, tendo sido popularizado por autores da corrente pós-custodial que colocam a ênfase na gestão da informação (Pinto, 2009; A. M. da Silva, Ribeiro, Ramos, & Real, 2009). Também figura em literatura nacional de preservação digital (Barbedo et al., 2007; Corujo & Barbedo, 2012; Ferreira, Saraiva, & Rodrigues, 2012; Cristiana Freitas, 2012). Porém, nenhum destes autores procede à conceptualização do termo ‘metainformação’ em contraste com ‘metadados’. Eysenbach et al. (2000) e Middleton (1999) usam ora metainformação, ora metadados, enquanto na restante literatura científica o termo ‘metainformação’ é usado de forma exclusiva, embora sem justificação explícita. Destaca-se como exceção uma tese de doutoramento recentemente publicada que discute este termo, contrastando-o com ‘metadados’. Corujo (2020, p. 12) começa por caracterizar a metainformação como particular do contexto português, antes de a considerar um termo mais rigoroso face a metadados.

Apesar de se constar uma aparente ausência na literatura sobre a problematização deste termo, salvo a tese de doutoramento de Luís Corujo no contexto português, entende-se que se trata de uma distinção que merece discussão. Em primeiro lugar, este termo esconde uma subtilidade semântica. Se falamos em metainformação, falamos em ‘informação sobre informação’ (Eysenbach et al., 2000, p. 231; Hlava, 2015, p. 30). Tendo em conta que informação são dados com significado assimilados por um indivíduo, como visto anteriormente, falar em ‘metainformação’ é afirmar que os dados não existem por si só, existindo apenas quando um indivíduo os dota de significado. Esta questão aproxima-se largamente do problema filosófico do observador participante (Bates, 2011; Hjørland, 2007, 2009). Se uma árvore cair no meio de uma floresta inabitada por pessoas, é verdade que ela caiu? A resposta pragmática é que a árvore tem uma existência independentemente da observação humana, constituindo a sua queda informação para aves e outros bichos<sup>4</sup>. Da mesma forma, os dados existem mesmo que nenhum indivíduo tenha conhecimento da sua existência. Uma porção dos dados e metadados gerados atualmente provêm de registos automáticos de ferramentas de processamento de enormes volumes de dados, tais como motores de busca (Google) e outras técnicas de *Big Data*. Neste caso, mesmo quando não há humanos envolvidos, constituem informação para outras máquinas porque foram por estas processados e acrescentados de valor.

---

<sup>4</sup> Esta interpretação muda se for postulado que a árvore cai numa floresta sem nenhum ser vivo nas redondezas. Neste caso, a discussão é não apenas da interpretação humana, mas da interpretação como um todo. Se uma árvore cai numa floresta desprovida de vida, não produz som porque o som é uma interpretação neurológica da vibração numa membrana. Produz, antes, apenas vibrações que se propagam no ar que, na ausência de um “cérebro” que as interprete, nada mais são que vibrações desprovidas de significado.

Num segundo argumento, discutivelmente mais importante e menos debatível, prefere-se o uso de ‘metadados’ por uma harmonização terminológica com outros domínios do conhecimento. A título de exemplo, uma área em franca expansão é a da *research data*, ou dados de investigação. Não se conhece nenhum autor que se refira a este conceito enquanto *research information*. A adoção do termo ‘metainformação’ criaria assim uma separação extra entre a gestão da ‘metainformação’ e a gestão da *research data*, o que seria indesejável sobretudo num contexto de crescente integração das áreas científicas. Assim, de forma a não fomentar as divisões ainda existentes ao nível dos tratamentos dos metadados pelas bibliotecas, pelos arquivos e pelos museus, e de modo a favorecer um tratamento integrado aos objetos portadores de informação que constituem as suas coleções, pretere-se o termo ‘metainformação’ em favor do termo ‘metadados’, redutoramente entendidos, numa forma simplificada, como “dados sobre dados”.

A criação de metadados tem sido uma das tarefas primárias dos bibliotecários ao longo dos tempos, que têm criado catálogos (metadados) das coleções (os dados) desde que há história registada. Por este motivo, não é de admirar que muitas das inovações ao longo da história tenham ocorrido no contexto de atuação destes profissionais. Como tal, qualquer tentativa de esboçar uma narrativa do uso de metadados sobrepõe-se em grande medida a uma história da catalogação, como veremos de seguida, ilustrando os principais desafios surgidos nos vários contextos históricos e sociais e destacando algumas das soluções adotadas mais notáveis.

## 1.2 Etapas do uso de metadados

### 1.2.1 Metadados na Antiguidade

A coleção que se pensa ser a mais antiga é a da antiga cidade-estado de Ebla, atual Tell Mardikh no norte da Síria. Escavações começadas em 1964 revelaram, uma década mais tarde, as ruínas de um palácio real contendo um vasto arquivo de milhares de tábuas de argila (*clay tablets*) dos anos 2600-2300 BCE. Já nesta época são observáveis princípios de classificação que assentavam na correspondência da forma das tábuas com o assunto do documento. As tábuas redondas correspondiam a textos económicos e administrativos, enquanto que as tábuas quadradas tratavam de assuntos gerais (Wellisch, 1981, p. 491). Além disso, estavam dispostas nas prateleiras de modo a que a primeira linha de texto e uma breve descrição dos conteúdos se encontrassem visíveis, uma prática não distante das lombadas usadas atualmente, e que pode constituir a primeira tentativa de uso de metadados descritivos. Embora não tenha sobrevivido nenhum catálogo desta coleção, foram encontradas listas sobre deuses, reis, objetos, animais, profissões, ordenadas de acordo com precedência ou importância, que constituem uma primeira tentativa de organização de metadados através da construção de uma visão funcional do mundo.

Cerca de um milénio mais tarde, por volta do século 13 BCE, eram já observáveis práticas mais familiares à catalogação atual. Escavações iniciadas em 1833 por Charles Texier revelaram os destroços do palácio de Hatusa, capital do antigo império Hitita correspondente à moderna Turquia (Texier, 1839). Foram encontradas nos arquivos deste palácio mais de 30.000 tábuas de

argila, constituindo uma coleção vastamente superior à da coleção encontrada em Ebla. Neste arquivo encontram-se já técnicas de catalogação mais avançadas em virtude da sua maior dimensão.

Um dos avanços mais importantes diz respeito ao uso de um cólofon, um termo familiar a bibliotecários hodiernos. Este termo deriva do latim *colophon*, que por sua vez deriva do grego *kolophon*, significando literalmente ‘topo’. Por altura do império Hittite, muitos textos ocupavam mais do que uma tábuas, pelo que era necessário um método de agregação que permitisse relacionar que tábuas pertenciam a que obra. O cólofon servia este propósito fornecendo uma breve descrição dos conteúdos e mantendo em conjunto as tábuas respeitantes à mesma obra, sobretudo pela indicação do número da obra a que determinada tábuas correspondia, num raciocínio similar ao modo como atualmente é indicado o volume de uma obra com vários volumes<sup>5</sup>. Outra inovação pauta-se pela existência de catálogos com informação bibliográfica e descritiva, tais como título, autor, uma descrição dos conteúdos e o número de tábuas ocupadas (Casson, 2001).

Outro exemplo de estrutura e organização foi descoberto durante as escavações na antiga cidade de Ur. Em 1920, o arqueólogo Leonard Woolley descobriu uma divisão de um palácio do sexto século BCE que continha vários objetos mais antigos. Estes objetos encontravam-se ordenados e precediam o resto do palácio em mais de 700 anos. Woolley acabou por concluir tratar-se de um museu criado e mantido pela princesa Bel-Shalti-Nannar, filha de Nabonido, último rei do império neobabilónico. Woolley chegou a esta conclusão devido à descoberta de um conjunto de cilindros de argila com uma particularidade muito curiosa. Cada um consistia num *small drum-shaped clay object on which were four columns of writing; the first three columns were in the Sumerian language ... the fourth column was in the late Semitic speech* (Woolley, 1950, p. 157). Estes objetos que a princesa Bel-Shalti-Nannar colecionou, organizou e catalogou constituem provavelmente um dos exemplos mais primitivos de etiquetas de museu e, conseqüentemente, de metadados para a curadoria de coleções. Em essência, estes metadados têm a função de descrever a coleção e colocar as diferentes partes em contexto de modo a que possam ser compreendidos (Gartner, 2016, p. 15), naquele que é um uso intemporal que ainda hoje persiste.

Um dos avanços mais importantes do mundo antigo ao nível da catalogação ocorreu na lendária Biblioteca de Alexandria. Erigida no terceiro século BCE, não tardou a assumir-se como um uma das maiores e mais prestigiadas bibliotecas da época, e assim se manteve ao longo de mais de 200 anos, até que foi em parte consumida por chamas em 48 BCE, quando Júlio César foi forçado a queimar os seus próprios navios durante o cerco a Alexandria<sup>6</sup> (Haughton, 2011). Fundada com

---

<sup>5</sup> Eis um exemplo de um cólofon: *Second tablet of the Tudhaliyas, the great king, on the oath. End. This tablet was damaged; in the presence of Manhhuzi and Halwalu, I, Duda, restored it* (Casson, 2001, p. 5). Cada cólofon iniciava-se com o número da tábuas em que ocorre. Destaca-se também a indicação do escrivão envolvido no registo do cólofon, na indicação de que esta é a última tábuas da obra através do termo *end*. Embora estas informações constituíssem uma preciosa ajuda para os utilizadores, nem todas as tábuas apresentavam cólofon.

<sup>6</sup> Em inglês, este evento ficou conhecido como *Siege of Alexandria*.

uma ambição imperialista de promoção da cultura helenística, a biblioteca de Alexandria constituía um gigantesco repositório do conhecimento (Macleod, 2004) que atraía vários académicos e intelectuais.

Nesta história dos metadados é importante destacar Calímaco de Cirene (310/305-240 BCE), cuja obra *Pinakes* constitui um dos exemplos mais notáveis e pioneiros de compilação bibliográfica, bem como a primeira coletânea sofisticada de uma lista de autores e respetivas obras (Library of Congress, 2017, p. 14). Convém notar que as coleções encontradas até à fundação da biblioteca de Alexandria constituíam sobretudo arquivos para uso da realeza e para fins burocráticos, não tendo a função pedagógica e erudita para formação de académicos que viriam a assumir (Wellisch, 1981)<sup>7</sup>.

As *Pinakes* (literalmente, ‘tábuas’) eram essencialmente uma lista de autores da Biblioteca de Alexandria e informações sobre as suas vidas e obras. Infelizmente, o nosso conhecimento resume-se a poucos fragmentos que sobreviveram e a referências por outros autores<sup>8</sup>. Pelo que é possível perceber pela observação destes fragmentos, Calímaco compilou e organizou os autores, individualmente dispostos por ordem alfabética, sendo cada nome acompanhado por uma breve nota biográfica (nome do pai, local de nascimento e, por vezes, uma alcunha) e uma lista das suas publicações, também por ordem alfabética. A ordenação alfabética de autores foi uma prática iniciada por Zenódoto de Éfeso (330-260 BCE), primeiro bibliotecário de Alexandria, e continuada por Calímaco, embora ambos usassem apenas a letra inicial (Casson, 2001). Adicionalmente, era também incluída uma nota que indicava o número de linhas que cada texto ocupava, de modo a auxiliar os bibliotecários (Witty, 1958).

Embora as *Pinakes* sejam por alguns autores considerados o primeiro catálogo bibliotecário (Blum, 1991), Gartner sugere que devem antes ser entendidas como obras pioneiras em biografia e bibliografia, uma vez que a sua intenção primária não era auxiliar a localização física de obras (Gartner, 2016, p. 16). A enorme atenção dada a questões biográficas do autor, como detalhes da sua vida, contexto familiar, locais de nascimento e morte, ou até controvérsias associadas à atribuição da autoria (Witty, 1958) realçam o carácter biográfico desta obra.

Ainda assim, não pode ser deixado de mencionar a sua relevância enquanto tentativa de classificação, categorizando os géneros em hierarquias. Calímaco começou por dividir os autores em prosa ou poesia, tendo depois subdividido cada uma destas categorias. A poesia foi desdobrada em poetas dramáticos, que por sua vez se decompunha em poetas trágicos e cómicos, e assim por diante. De entre as categorias incluídas para a prosa contam-se filosofia, oratória, história, medicina e miscelânea (Casson, 2001, p. 40). Estamos perante um dos primeiros exemplos de esquema de classificação hierárquicos, um conceito fundamental em metadados, inclusive no dia de hoje.

---

<sup>7</sup> Carlos Guardado da Silva (2019, p. 58) assinala a distinção por Gaio Vopisco (130 BCE – 87 BCE), escritor romano, entre a produção de documentos de natureza administrativa e de valor literário.

<sup>8</sup> Uma compilação encontra-se na obra em dois volumes de Pfeiffer (1965).

Porém, surgia o problema de como incluir autores que pertencessem a mais do que uma categoria. Os fragmentos que chegaram até nós não nos permitem perceber como Calímaco resolveu esta questão que seria um desafio que vários catalogadores iriam enfrentar ao longo dos próximos séculos.

Os séculos seguintes não trouxeram inovações notáveis ao nível do uso de metadados. Neste intervalo de tempo até ao século 15, a maioria das coleções da Idade Média originava em mosteiros cujos interesses em catalogação eram por vezes parcos. Ainda assim, destaca-se o catálogo da Abadia de Reading, um dos mosteiros beneditinos mais opulentos da Inglaterra medieval, fundado em 1121 pelo rei Henrique I. O catálogo sobrevivente foi compilado no século 12 e constitui sobretudo um inventário das posses da abadia, tais como relíquias e livros. A nível da bibliografia, embora ocupe apenas quatro páginas, lista por assunto cerca de 300 livros dispersos pela biblioteca principal do mosteiro, pelas capelas, pelos quartos dos monges e até pela enfermaria. A sua classificação é rudimentar em comparação com as obras anteriores, na medida em que havia apenas quatro tipos de obras: Bíblias, livros sobre a Bíblia, obras de padres, com destaque para Santo Agostinho, e outros textos litúrgicos para uso diário tais como missais (Westwell, 2012).

Como a localização dos livros só surge no século 14, as funções dos catálogos até então (incluindo o de *Reading*) aproximam-se mais das de um inventário do que de um instrumento de pesquisa para localizar e aceder aos conteúdos (Haynes, 2018). Este tipo de informação talvez não fosse necessário dada a dimensão reduzida das coleções. Contudo, esta realidade viria a mudar drasticamente no século 15, com a invenção da imprensa, cuja adoção massificou o tamanho das coleções, colocou novos desafios à catalogação e, conseqüentemente, motivou a inovações no uso de metadados.

Os exemplos de catálogos da Antiguidade e da Idade Média que temos vindo a ilustrar apresentam um padrão comum: quanto menor a dimensão dos catálogos, menor a necessidade de metadados elaborados para a sua gestão. Desde a relativamente pequena coleção de tábuas de argila de Ebla até à vasta biblioteca de Alexandria, o tamanho das coleções ditou o grau de elaboração dos metadados para a sua gestão. Ainda assim, todos estes casos analisados até agora apresentavam um denominador comum: constituem metadados sobre coleções de manuscritos, um tipo de produção de informação moroso para os nossos padrões atuais de redação digital. O tamanho das coleções era, por conseguinte, afetado pela tecnologia de escrita disponível. Esta realidade viria a mudar quando, em 1450, o ourives alemão Johannes Gutenberg (1400-1468 CE) introduziu na Europa a imprensa usando caracteres móveis<sup>9</sup>.

---

<sup>9</sup> A história da imprensa remonta a 3500 BCE, quando se observa o uso de carimbos cilíndricos como forma de assinatura (Ross, 2014, p. 304). De resto, os primeiros passos no desenvolvimento da imprensa ocorreram na Ásia muitos séculos antes da Europa. A invenção da imprensa usando matrizes de madeira (xilografuras) ocorreu na China antes de 220, inicialmente para impressão de têxteis (Whitfield & Farrer, 1990). Acredita-se que o primeiro livro impresso seja a obra maaiana 'Sutra do Diamante' (*Diamond Sutra*), datado de 868 (British Library, 2020; Soeng, 2011), pelo que a data real do uso de xilografia para impressão de livros

## 1.2.2 A introdução da imprensa na Europa

A adoção da imprensa teve profundos impactos na cultura e na sociedade, uma vez que levou não só a um aumento do volume de informação disponível, mas também da sua variedade, motivo pelo qual é visto como o início da comunicação em massa (Eisenstein, 1980). O aparecimento da imprensa revolucionou todos os passos do consumo da informação, desde a sua produção à disseminação: tornou-se mais barato, rápido e fácil criar documentos, além de permitir uma maior diversidade de tipologias (Gartner, 2016). Esta inédita torrente de informação colocou novas questões e desafios aos esforços de catalogação, embora as inovações neste campo tenham demorado a observar-se, levando mais de um século a cimentar-se. A reação de alguns críticos ao aparecimento relativamente repentino na época foi negativa, que questionaram a autoridade dos clássicos e a qualidade da informação num contexto de tanta diversidade de fontes:

*Dutch social critic Desiderius Erasmus complained, “Is there anywhere exempt from these swarms of new books? Even if taken out, one at a time, they offered something worth knowing, the very mass of them would be a serious impediment to learning. René Descartes quipped that there were “so many useless things and confusion heaped in such large volumes that it would take longer to read those books than we have to live in this life” (Library of Congress, 2017, p. 19).*

Os paralelos são claros com alguns dos receios manifestados em relação à massificação do acesso à informação com a Internet, 550 anos mais tarde. Hoje, é um dado adquirido que o número de conteúdos disponíveis supera largamente o nosso tempo de vida. Paralelamente, a questão da qualidade das fontes no meio de tanto ruído informacional é um tema comum ao período medieval, embora amplificado atualmente algumas ordens de magnitude. Em todo o caso, a Biblioteca Bodleiana (*Bodleian Library*) foi uma das grandes bibliotecas da era pós-Gutenberg. Criada em 1602, esta biblioteca da Universidade de Oxford foi umas das primeiras a criar um catálogo impresso da sua coleção. Este catálogo revela-nos que, por esta altura, as práticas de catalogação ainda não tinham mudado significativamente em comparação com os catálogos anteriores. Ao longo de 264 páginas é apresentada uma lista dos livros em posse da biblioteca, correspondendo cada página a uma prateleira. São indicados o nome do autor, a sua proveniência geográfica, o título da obra, a sua dimensão física, a data e o local de publicação. Este catálogo aproxima-se das *Pinakes* de Calímaco, embora sem o registo biográfico (Philip, 1983). Contudo, ao contrário da obra de Calímaco, cujas entradas se encontravam ordenadas por autor, esta prática encontrava-se

---

deva ser ligeiramente anterior. O desenvolvimento de impressão por caracteres móveis de argila cozida deve-se ao artesão chinês Bi Sheng (990-1051), no século 11 (Needham, 1994, p. 14), ao passo que a invenção de caracteres móveis de metal ocorreu na Coreia, no século 13. Quando Gutenberg introduziu a imprensa por caracteres móveis de metal na Europa, introduziu também inovações ao nível da constituição dos moldes usando uma liga de estanho, chumbo e antimónio, materiais que se tornariam padrão durante mais cinco séculos (Lechêne, 2020).

inicialmente ausente do catálogo bodleiano. Só numa edição mais tardia, de 1620, foi introduzida nesta biblioteca a listagem por autor, como reflexo da crescente importância dada à autoria como ponto de acesso (James, 1986).

Para além do aparecimento da imprensa no século 15, a Revolução Industrial que ocorreu a partir de meados do século 18 e se prolongou pelo século 19 teve igualmente profundos impactos na vida social, cultural e material. A automatização e a expansão da produção de muitos materiais espelharam-se num aumento da criação e da disseminação da informação, pelo que os metadados tiveram de se ajustar uma vez mais a este volume informacional.

O código francês de 1791 é tido como a origem dos catálogos modernos, na medida em que se trata do primeiro código de catalogação nacional com entradas por autor, com cartões de catálogo e regras de aquisição (Haynes, 2018). Contudo, uma das principais inovações desta época foi o reconhecimento de regras de catalogação mais consistentes, face a épocas anteriores. As regras de catalogação foram desenvolvidas por Antonio Panizzi<sup>10</sup> (1797-1879), um bibliotecário italiano em funções no Museu Britânico que, desde 1753, funcionava como a Biblioteca Nacional do Reino Unido. Em 1841, o museu publica as *Rules for the Compilation of the Catalogue*<sup>11</sup>, da autoria de Panizzi em conjunto com os seus colegas Edward Edwards, John Winter Stones, John Humffreys Parry e Thomas Watts, conhecidas como as *91 rules*, que viriam a afirmar-se como as bases das práticas de catalogação das próximas décadas e um ícone da catalogação moderna. Estas regras assentam em três princípios: (1) *the data are derived from the item in hand*, (2) *the title page is the primary source of data*, and (3) *the title is to be transcribed exactly as it appears in the work* (Blake, 2002, p. 6). O objetivo era claro: *to standardize the format of the bibliographic records and to ensure that enough detail was included to differentiate one record from another* (Hufford, 1991, p. 28).

Embora surgidas num período de pleno funcionamento da imprensa, apenas o primeiro volume das 91 regras foi impresso em 1841. Após pressões, Panizzi estimou que, na melhor das hipóteses, apenas uma parte poderia estar impressa em 1895. Na verdade, os catálogos impressos da época eram extremamente difíceis de organizar na medida em que exigiam consideráveis recursos a nível da formatação de texto, revisão e impressão, encontrando-se desatualizados no momento em que eram impressos, tendo as novas entradas de ser adicionadas manualmente (Denton, 2007). Esta demora mereceu-lhe várias críticas que contestavam a sua abordagem à catalogação, críticas essas que conseguiu sempre desviar: *armed with a superior knowledge of libraries and catalogs, trained in law, and possessing an instinct for verbal combat, he subjected his critics' arguments to withering attack. He analyzed their testimony point by point with great*

---

<sup>10</sup> Panizzi era um refugiado italiano, advogado e muito envolvido na política, incluindo em sociedades secretas revolucionárias. A história completa da sua vida pode ser consultada em Miller (1988). A sua próspera mas nem sempre calma carreira no Museu é descrita por Norris (1939).

<sup>11</sup> Vide Panizzi & British Museum (1841).



success, for 'he was a man with the annoying habit of not only being right, but of being able to prove it' (Brault, 1972, pp. 8-9).

Um dos seus adversários era o escritor Thomas Carlyle que entendia a criação de um catálogo como uma tarefa simples de listagem de títulos. Contudo, Panizzi tinha outro entendimento das funções de um catálogo, que se deviam alinhar com as necessidades do utilizador. Esta diferença de posições é ilustrada por Lubetzky:

*Panizzi's response was, in effect: Yes, I require the reader to look in two places for the information he wants, because I want to tell him much more than merely whether or not the library has a particular book; yes, my rules are complicated, but that is because my rules are concerned not only with the book as a single and separate item, but also as a complex of editions and translations of potential interest to an acquiring reader... In Panizzi's own words, a reader may know the work he requires; but he cannot be expected to know all the peculiarities of different editions, and this information he has a right to expect from the catalogues. So here we have two individuals looking at the same object — the book — but seeing different things. Carlyle saw the book as a material object, a separate entity unrelated to any other book in the library, and he did not see why it should not be so represented in the catalog. Panizzi saw the book as an edition of a particular work that is intimately related to the other editions and translations of the work that the library may have, and thought that it should therefore be integrated with them (Lubetzky, 2001a, p. 424).*

Panizzi pretendia colocar uma obra em contexto, uma das funções essenciais dos metadados, atualmente magnificada pelas tecnologias de *linked data*, como veremos mais adiante. Em resumo, pretendia-se a criação de *an adequate catalog, concerned about the actual needs of a reader, must be designed to tell one not only whether the particular book he or she seeks is in the library but also what other editions of the work and what other works of the author the library has. That was the object of Panizzi's rules* (Lubetzky, 2001b, p. 370). Algumas das suas regras ainda se encontram em vigor atualmente nos catálogos contemporâneos. Por exemplo, a regra de que o primeiro autor de uma obra deve ser considerado o principal criador (terceira regra) é ainda aplicada nas Regras de Catalogação Anglo-Americanas<sup>12</sup>, que muitas bibliotecas seguem atualmente (Lehnus, 1972).

As regras de Panizzi são consideradas o berço da catalogação moderna por várias razões: por terem sido desenvolvidas por um grupo; por terem nascido fruto de um intenso debate e justificação; por terem sido aprovadas por entidades governamentais; por serem usadas por uma grande biblioteca e por terem recebido atenção internacional (Denton, 2007). No caso de Panizzi começam já a esboçar-se tendências de normalização que se viriam a intensificar durante o século 20 e que dominam uma parte do discurso da atualidade.

---

<sup>12</sup> Vide (American Library Association, 2020).

Outro bibliotecário famoso do século 19 foi Charles Ammi Cutter<sup>13</sup> (1837-1903), cuja obra de catalogação estende influência até à atualidade. Mais conhecido como autor da Classificação Expansiva que, embora nunca concluída, serve de base à classificação da Biblioteca do Congresso (BC), nos Estados Unidos da América, Cutter tem lugar neste breve apanhado histórico pela publicação das *Rules for a Printed Dictionary Catalogue*, em 1876. A quarta e final edição data de 1904, com o título revisto de *Rules for a Dictionary Catalog* (Cutter, 1904), sendo curioso o facto de o adjetivo ‘impresso’ ter caído do título dada a popularidade das fichas de catálogo (*catalog card*).

Um catálogo em forma de dicionário, isto é, com as entradas organizadas em ordem alfabética, era uma novidade. Além disso, as entradas eram listadas não só por autor, como também por título e assunto. Por este motivo, um catálogo deste tipo reúne tanto obras criadas por uma pessoa, como também obras sobre essa mesma pessoa. Cutter ilustra o seu pensamento sobre as funções de um catálogo no início dessa obra da seguinte maneira:

*Objects:*

1. *To enable a person to find a book of which:*

(A) *the author is known*

(B) *the title is known*

(C) *the subject is known.*

2. *To show what a library has*

(D) *by a given author*

(E) *on a given subject*

(F) *in a given kind of literature.*

3. *To assist in the choice of a book*

(G) *as to its edition (bibliographically)*

(H) *as to its character (literary or topical* (Cutter, 1904, p. 12).

Estes ‘objetos’ constituem o primeiro conjunto de axiomas formulados na catalogação e constituem a base sobre a qual assentam as regras de catalogação usadas por Cutter (Denton, 2007), que se refletem em modelos conceptuais modernos, como as *Functional Requirements for Bibliographic Records* (FRBR)<sup>14</sup>, e que governam os princípios da catalogação. Cutter também dedicou parte da sua vida à normalização e à internacionalização, para as quais a sistematização é indispensável. Ranganathan toca neste ponto ao resumir a importância da obra de Cutter:

*None of the above drafts [as 91 regras de Panizzi e outros códigos germânicos] set forth cataloguing rules in a systematic or exhaustive manner. Nor was there much evidence of their roots stemming from some kind of first principles. Nor again was any of them from a general code not particularly conditioned by the practice of a single library. Rdc [Rules*

---

<sup>13</sup> A sua vida e obra são descritas por Miksa (1977).

<sup>14</sup> Vide International Federation of Library Associations and Institutions (1998).

*for a Dictionary Catalog] was the first code to reach beyond those limitations ... That is why Rdc is whole as an egg* (Ranganathan, 1955, p. 15).

No século 19 surgiu outro método de divisão do conhecimento por assunto, hoje largamente aplicado num número muito significativo de bibliotecas por todo o mundo. Trata-se da Classificação Decimal de Dewey (CDD, em inglês *Dewey Decimal Classification*), publicada em 1876 pelo bibliotecário americano Melvil Dewey (1851-1931), que divide o mundo do conhecimento em 10 categorias (Tabela 1), e se encontra atualmente na 23ª edição.

Tabela 1 — Classes principais da Classificação Decimal de Dewey (CDD), em inglês.

000	Computer Science, information & general works
100	Philosophy & Psychology
200	Religion
300	Social Sciences
400	Language
500	Pure Science
600	Technology
700	Arts & recreation
800	Literature
900	History & geography

FONTE: Elaboração própria.

A estrutura destas regras é indissociável do contexto científico da época. Durante o século 19 notou-se um interesse crescente na taxonomia, particularmente na química — com as discussões em busca de consenso sobre tabela periódica — e da biologia — com a aplicação sistemática do sistema de nomenclatura dos organismos apresentada em 1735 pelo zoologista sueco Carl Linnaeus. Também os bibliotecários sentiram necessidade de explorar os domínios da taxonomia, sendo neste contexto que surge a CDD. Assim, enquanto taxonomia, as 10 categorias principais são subdivididas noutras categorias de acordo com uma hierarquia pré-definida de assuntos. Um dos impactos deste sistema de classificação foi a possibilidade de organização dos livros nas prateleiras de acordo com o assunto, podendo esse assunto ser descrito com um detalhe tão grande quanto o desejo do catalogador. Anteriormente, não havia necessariamente nenhum critério para a colocação dos livros nas estantes, podendo ser feito simplesmente por ordem de aquisição. A partir de então, livros sobre o mesmo assunto e assuntos semelhantes encontram-se agrupados, proporcionando aos utilizadores o prazer da descoberta de livros desconhecidos sobre um assunto específico.

A CDD serviu de inspiração à Classificação Decimal Universal (CDU), desenvolvida nos finais do século 19 pelos bibliotecários belgas Paul Otlet e Henri La Fontaine. Estes agora famosos bibliotecários foram incumbidos com a tarefa de elaboração de um repertório bibliográfico universal pelo Instituto Internacional de Bibliografia, em 1895. No âmbito deste projeto, necessitavam de um modo de ordenar as entradas, tendo-se inspirado na CDD pedindo a Dewey

permissão para a traduzir. Porém, apesar de uma intenção inicial de tradução, estes autores começaram a introduzir algumas inovações. Uma desvantagem da CDD prendia-se com a falta de flexibilidade que impedia a representação de assuntos mais complexos, uma condicionante derivada da sua natureza enumerativa (Simões, 2010, p. 223). Um dos avanços mais radicais prende-se com a capacidade de síntese da codificação. Na CDD, a classificação é enumerativa, estando determinada uma pré-codificação dos assuntos. Na CDU, a classificação é adaptada de forma a permitir a construção de expressões sintáticas que representem assuntos interrelacionados. Esta classificação é familiar a utilizadores de muitas bibliotecas públicas e universitárias, incluindo as bibliotecas da Universidade de Coimbra. Os números e letras que os livros apresentam na lombada correspondem à CDU.

Contudo, os sistemas de classificação decimais apresentam desvantagens inerentes. Gartner (2016) identifica a rigidez da classificação como um problema no caso de obras que versem sobre mais do que um tópico, uma vez que a hierarquia apenas permite a colocação de um número como ponto de acesso; outra condicionante prende-se com a necessária aprendizagem para os utilizadores do significado dos códigos antes que possam descobrir os livros que procuram. Ainda assim, estes inconvenientes não constituíram um entrave a que se tornassem os sistemas de classificação mais adotados a nível mundial.

O século 19 foi também berço da divisão no tratamento diferente de metadados entre arquivos e bibliotecas, no seguimento da enunciação do princípio de respeito pelos fundos, pelo arquivista francês Natalis de Wailly<sup>15</sup>. Este princípio dita que os fundos documentais criados por um organismo devem ser mantidos em conjunto, de modo a dotá-los de sentido e contexto. Desde então, os arquivistas têm-se focado em descrever os fundos e a sua estrutura interna, ao passo que o trabalho dos bibliotecários tem sido marcado pela descrição dos itens individuais que constituem a coleção.

Por este motivo, o equivalente aos catálogos das bibliotecas nos arquivos são os instrumentos de pesquisa (*finding aids*) que, enquanto resultado das atividades de descrição dos arquivistas, distinguem-se dos primeiros por colocar mais enfoque no conjunto documental. Um instrumento de pesquisa pode ser entendido como *any tool that aims to provide users with intellectual and/or physical access to the holdings of archival institutions* (MacNeil, 2012, p. 486). Em essência, trata-se de *a description of a collection as a single entity* (Gartner, 2016, p. 21) cuja

---

<sup>15</sup> A primeira distinção sistemática conhecida entre as diferenças entre material de arquivo e de biblioteca encontra-se na correspondência entre Francesco Bonaini, arquivista do Grão-Ducado da Toscana e Antonio Panizzi, bibliotecário do Museu Britânico e um dos primeiros a estabelecer princípios modernos bibliotecários e regras de catalogação. Duranti (1993, p. 10) assinala que esta correspondência, publicada em 1867, proclamou a separação entre a arquivística e a biblioteconomia e influenciou a sua evolução em separado. A partir de 1980, a arquivística, movida pela disseminação das tecnologias de comunicação eletrónica nas quais o conceito de informação adquire uma posição central, começou a explorar ideias já testadas na biblioteconomia, tanto na Europa como na América.

função consiste em auxiliar um utilizador a localizar material de interesse através de uma descrição coletiva, podendo incluir informação sobre a proveniência, notas biográficas das entidades envolvidas, bem como uma descrição da classificação. A origem dos instrumentos de pesquisa em arquivos data à própria (re)invenção da escrita na antiga Suméria através de registos em tabuinhas de argila que auxiliavam a localização dos documentos (Library and Archives Canada, 2012). Nos séculos 19 e 20, os instrumentos de pesquisa eram redigidos em papel, na forma de listas. No século 21, emergem os instrumentos de pesquisa digitais armazenados enquanto bases de dados. Apesar de tudo, as mais recentes tecnologias de descrição em grafo e o advento da *Web Semântica* começam a desafiar estas antigas separações entre tratamento da informação em arquivos e bibliotecas, conforme será discutido mais tarde.

Por volta do século 19 observa-se igualmente um importante desenvolvimento no suporte físico para armazenamento dos metadados. Até finais do século 18, os catálogos assumiam a forma de tomos encadernados, isto é, de um conjunto de páginas unidas por uma lombada, quer esse conjunto se tratasse de um livro impresso, tal como o catálogo da Abadia de Reading, ou de um livro de guarda, que consistia num livro cujas páginas eram tiras de papel impressas que se podiam acrescentar colando no respetivo local. Ambas abordagens eram muito pouco eficientes e práticas, com um alto custo de cada vez que era necessário fazer uma alteração (Brooks, 2000). No caso dos catálogos impressos, conforme mencionado anteriormente, uma impressão completa do catálogo era um exercício raro, dispendioso e paulatino, que resultava frequentemente em catálogos desatualizados no momento da impressão. Quanto aos livros de guarda, não só era necessário remover e criar novamente as entradas anteriores quando o espaço numa página já se encontrava ocupado, como também as próprias tiras de papel que constituíam a entrada eram de tamanho variável, por vezes sobrepondo-se umas às outras, de modo que *to read the manuscript one must sometimes unfold a paper concertina* (Considine, 2011, p. 218).

Uma solução para a ineficiência dos catálogos teve origem no pós-Revolução Francesa de 1789, quando o governo revolucionário francês ordenou a apropriação das propriedades eclesiásticas, incluindo várias bibliotecas e respetivas coleções. Perante a assimilação de um enorme volume documental, o governo sentiu a necessidade de manter uma lista das obras adquiridas, dando um conjunto de instruções para o efeito, que viria a ser conhecido como o ‘Código de Catalogação Francês de 1791’<sup>16</sup>. Descrito como *a paragon of brevity and practical simplicity* no qual não havia *fuss and bother or philosophizing* (Strout, 1956, p. 267), este instrumento fornecia instruções para facilitar a localização de um livro com o intento de criar um inventário dos livros de cada distrito de modo a preparar um catálogo nacional (Hopkins, 1992). Assim, mais do que a função usual dos catálogos de localização de um volume numa prateleira, este catálogo pretendia recolher informação sobre bibliotecas geograficamente dispersas (Gartner, 2016, p. 23).

---

<sup>16</sup> Vide Massieu (2018).

Todavia, este simples conjunto de regras de catalogação introduziu pela primeira vez o uso de fichas de catálogo, recorrendo a cartas comuns de jogo, enquanto forma de registo de metadados, conforme explicado por Riberette:

*Não será difícil reunir os materiais deste catálogo. Bastará usar cartas de jogo nas quais pode ser escrito o nome da obra, o nome do autor quando for conhecido, o local de impressão e a data ... Todas as cartas serão trazidas para um depósito comum serão selecionadas, colocadas nas divisões e subdivisões bibliográficas e finalmente divididas pelos diferentes copistas cujo trabalho conjunto formará o catálogo (Riberette, 1970, p. 17)<sup>17</sup>.*

A adoção das fichas de catálogo para fins de catalogação, ainda que com um interesse de localização geográfica, constitui um importante passo quanto à possibilidade de pontos de acesso múltiplos e de velocidade de catalogação. Por um lado, até então os catalogadores tinham-se debatido sobre como catalogar uma obra que não pertencesse claramente a um assunto. Com as fichas, tornara-se simples criar mais do que uma entrada através da elaboração de várias fichas. Por outro, atualizar o catálogo consistia simplesmente em acrescentar uma ficha à gaveta, abandonando-se assim o moroso processo de atualização de tomos volumosos (Gartner, 2016, p. 23).

Todavia, as vantagens das fichas catalográficas tardaram a ser reconhecidas. O seu uso apenas se tornou comum nas bibliotecas dos EUA e do Reino Unido a partir da década de 1870, mediante a intervenção ativa de Melvin Dewey (Sharp, 1937). Em 1877, a recentemente criada *American Library Association* (ALA) normalizou as dimensões das fichas de catálogo (Krajewski, 2011), um importante primeiro passo nos esforços de normalização, que se viriam a intensificar ao longo do século 20 e que permitirão a partilha de metadados a uma escala maior.

### **1.2.3 O Século 20**

O século 20 foi um período-chave no desenvolvimento de práticas mais aprofundadas e consistentes ao nível da gestão de metadados. Por um lado, durante as primeiras décadas deste século observou-se uma tentativa de normalização das práticas e procedimentos anteriormente esboçados, de forma a permitir a partilha de metadados a uma maior escala. Por outro, começou a pensar-se nos metadados de uma forma mais abstrata, não apenas em hierarquias como até então, mas explorando novos modelos de organização e classificação que revelariam o seu potencial no ambiente tecnológico.

---

<sup>17</sup> *It suffira d'avoir des cartes à jouer, sur lesquelles on écrira le nom de l'ouvrage, celui l'auteur, quand il s'y trouvera, le lieu d'impression et la date ... Toutes les cartes étant rapportées à un seul dépôt, on en fera le triage, ensuite on procédera aux divisions et subdivisions bibliographiques, on les partagera enfin entre différents copistes dont le travail réuni formera le catalogue.*

Num primeiro momento, destaca-se a iniciativa da Biblioteca do Congresso de, em 1901, começar a vender fichas de catálogo. Reconhecendo a dificuldade em lidar com o crescente volume de informação, a venda de fichas pré-preenchidas, bastando às bibliotecas compradoras completar o resto da informação local, representa um princípio de normalização que permite poupar tempo, reduzir custos e aumentar a qualidade dos catálogos de bibliotecas mais pequenas (Library of Congress, 2017, p. 113). Além disso, nesse mesmo ano, a Biblioteca do Congresso criou pela primeira vez um 'catálogo conjunto' (*union catalog*) que descreve as coleções de mais do que uma biblioteca com o objetivo de discernir os livros detidos por cada biblioteca. Iniciado em 1901, o *National Union Catalog* (NUC) agrega atualmente as coleções de mais de 1100 livrarias nos EUA e Canadá (Library of Congress, 2020c). O volume desta empresa é de tal ordem monumental que teve de ser dividido em duas séries, uma para publicações pré-1956 (conhecido como *Mansell*, o nome da editora), e outra para publicações posteriores a 1955. Só o *Mansell* regista cerca de 13 milhões de entradas, distribuídas por 754 volumes, uma massa documental abismal que se encontra alojada na Biblioteca do Congresso (Beall & Kafadar, 2005). Gartner (2018) destaca a importância da ideia de agregar metadados de recursos geograficamente dispersos enquanto princípio fundamental dos atuais portais eletrónicos, desde os catálogos conjuntos atuais até aos serviços de pesquisa agregada e transversal do Google. O substituto natural do NUC em meio digital é o *WorldCat*<sup>18</sup>, o maior catálogo bibliográfico a nível mundial cujas intenções são semelhantes, mas mais amplas: agregar num só local os catálogos de várias bibliotecas, incluindo registos de livros, música e material audiovisual.

O ano de 1908 foi outro ano importante para a história da catalogação, tendo sido o ano em que a *American Library Association* e a *Library Association of the United Kingdom* publicaram um conjunto de regras de catalogação comuns, conhecidas como as *Anglo-American Cataloging Rules* (AACR). Todavia, devido a algumas divergências, foram publicadas duas versões, uma nos EUA e outra no Reino Unido. Ainda assim, o facto de estas duas entidades terem sentido necessidade de enunciar regras conjuntas revela o interesse na normalização e na internacionalização (Denton, 2007). Michael Gorman denominou este período como a Segunda Era da catalogação, *the era of the committee code, the increasingly loose, baggy monsters* que duraria até 1967, tendo a Primeira Era, *the age of the single-author code*, começado com Panizzi e incluído Cutter (Gorman, 2000, p. 13).

Porém, ao mesmo tempo que estas tentativas de normalização iam ocorrendo, também se iam sucedendo as alternativas disponíveis para as bibliotecas. Havia as regras inglesas, as regras americanas, as fichas de catálogo da Biblioteca do Congresso, e ainda as regras de Cutter. Não havia uma norma única, inequívoca, o que dificultava a partilha de metadados (Denton, 2007, p. 43). A dificuldade em catalogar era tal que, em 1941, Andrew Osborn escreveu um artigo intitulado

---

<sup>18</sup> Catálogo Mundial, acessível em <https://www.worldcat.org/>

*The crisis in Cataloging* no qual descreve várias abordagens à catalogação e ataca o que considera ser um método legalista usado na versão *draft* das regras da ALA da época:

*According to it, there must be rules and definitions to govern every point that arises; there must be an authority to settle questions at issue. So the reviser sits in judgment on the cataloger, and the head cataloger is the supreme court for his particularly library ... Debate, discussion, and decision eat up a surprising amount of time. Hence the demand in some quarters for a cataloging code that will define or rule on all debatable points (Osborn, 1985, pp. 93-94)<sup>19</sup>.*

É neste contexto que se insere parte do enorme contributo de Seymour Lubetzky (1898-2003), um dos maiores catalogadores do século 20. De origem judaica, Lubetzky alternou entre a carreira enquanto docente universitário e bibliotecário da Biblioteca do Congresso, trabalhando arduamente na simplificação das regras de catalogação<sup>20</sup>. Após um longo período de reclusão e estudo da história da catalogação no qual tentou perceber os princípios basilares da catalogação<sup>21</sup>, Lubetzky publicou em 1953 a obra *Cataloging Rules and Principles*. A primeira secção abre com a famosa pergunta: *Is this rule necessary?*, revisitando de seguida Panizzi e admirando *the broad knowledge, keen thinking, and fruitful imagination which the founders of the rules have brought to the profession of cataloging*. Contudo, sobre a abundância de regras que povoavam a sua época, questionava: *one is impelled to ask: Are all these rules necessary? Are all the complexities inevitable? Is there an underlying design which gives our code unity and purpose?* (Lubetzky, 2001d, p. 83).

A procura de Lubetzky por premissas fundamentais sobre as quais assentar a construção de enunciados mais complexos é um exercício admirável de lógica raramente observável, tanto na época, como atualmente. Não raras vezes, há uma preocupação em produzir nova documentação, novas normas e regras, sem um esforço simplista e fundamental de exame da razão de ser das regras, ou de como poderiam ser reduzidas a enunciados mais simples. Por exemplo, Lubetzky analisou as regras de catalogação sobre os nomes das mulheres casadas, e observou que o que estava em questão era uma mudança de nome, não se eram casadas ou não. Como havia já uma regra para mudança de nome, concluiu que não era necessário criar uma regra particular para este caso (Lubetzky, 2001d, p. 91).

Lubetzky defende a necessidade de uma *complete reconstruction* das regras, assente em *deliberately adopted objectives* e *well considered principles*:

---

<sup>19</sup> Originalmente escrito em 1941.

<sup>20</sup> Sobre a vida e obra de Lubetzky, *Vide* a introdução de (Lubetzky, 2001c).

<sup>21</sup> *He moved his residence from his office at the Library of Congress to the stacks, where he spent whole days, from morning to evening, studying the history of cataloging. He wanted to begin at the beginning, to understand the thinking of the visionaries of the golden age of cataloging, Panizzi, Jewett, and Cutter. Lubetzky worked two years on this project (Lubetzky, 2001c, pp. 75-76).*



*The objectives implicit in our rules for entry are two. The first objective is to enable the user of the catalog to determine readily whether or not the library has the book he wants. The catalog is constantly searched by many readers and members of the staff, and the quicker this information can be found, the better the catalog. The second objective is to reveal to the user of the catalog, under one form the author's name, what works the library has by a given author and what editions or translations of a given work (Lubetzky, 2001d, p. 113).*

Esta abordagem axiomática, inspirada nas regras de Cutter, serviu de base para a redação dos *Paris Principles*, nome pelo qual são conhecidos os *Statement of Principles* redigidos na *International Conference on Cataloguing Principles*, organizada em Paris em 1961. Este evento, que contou com a participação de 53 países e 12 organizações internacionais, figura como um marco na história da catalogação pois reúne quatro elementos essenciais para a gestão de metadados como a conhecemos atualmente: uma abordagem axiomática, a consideração pelas necessidades do utilizador, a noção da 'obra' com múltiplas ramificações e a preocupação com a normalização e a internacionalização (Denton, 2007, p. 47). No contexto deste sumário histórico, destacamos sobretudo a tentativa de simplificação de regras como requisito para facilitar a internacionalização e a normalização, uma ideia fundamental que será posteriormente adotada em ambiente digital por normas como o *Dublin Core*, como exemplo.

Ainda ao nível de inovações no domínio das bibliotecas, merece finalmente destaque o famoso bibliotecário e matemático indiano Shiyali Ranganathan (1892-1972). Um dos seus maiores contributos prende-se com a invenção da classificação facetada e a criação da 'Classificação de Cólón', publicada pela primeira vez em 1933 numa tentativa de contornar os problemas criados ao nível da classificação por outros sistemas como a CDD, que atribuíam a um livro um (e um só) lugar numa hierarquia, independentemente da similaridade com outros assuntos (Gartner, 2016, p. 24). Também conhecida por 'Classificação de Cólón' (CC) devido ao uso do símbolo de ':' como pontuação inicial (em inglês, os dois pontos designam-se por *colon*), a classificação facetada representa uma mudança fundamental no tratamento dos metadados, na medida em que permite a atribuição de várias categorias a um item.

Ranganathan desenvolveu a classificação facetada como alternativa aos sistemas enumerativos, que considerava superficiais e desatualizados desde a sua conceção, uma vez que não espelhavam a dinâmica de crescimento constante do conhecimento (Ranganathan, 1951). Conforme salienta Douglas, *hierarchy works well in a stable environment* (cit. por Goddard & Eccles, 2013, p. 135), mas depressa começa a desmoronar-se em ambientes mutáveis e dinâmicos em que a dimensão da realidade conhecida não é fixa. Além disso, ao longo da última década, os avanços tecnológicos têm aumentado sobejamente o volume de informação produzida. Em 1982, Fuller estimou que até 1900 o conhecimento humano duplicara a cada século; em 1945, duplicara a cada 25 anos; e em 1982 estaria a duplicar a cada 12-13 meses (Fuller, 1982). A IBM, empresa norte-americana especializada em tecnologia, projetou em 2006 que, em 2010, a quantidade de informação em circulação no mundo duplicaria a cada 11 horas (IBM, 2006). Outro tipo de cálculos

há que salientam o fluxo contínuo de produção de nova informação: em 2018 foi estimado que se produzissem 2,5 *quintillion bytes* de dados diariamente, com uma previsão de que em 2020 seriam criados 1.7MB de dados por segundo por pessoa (DOMO, 2018). A exatidão destes números é tão discutível quanto inegável é o crescimento exponencial da informação na nossa sociedade atual, em virtude do crescente número de aparelhos conectados a recolher dados e da própria taxa de transmissão dos dados. Depressa se compreende a dificuldade de adoção de um sistema enumerativo a partir do século 20.

O termo ‘facetado’ tem étimo na palavra francesa *facette*, um diminutivo de *face*. Segundo o *Online Etymology Dictionary*, esta ‘pequena face’ referia-se inicialmente à face de um diamante (Harper, 2020). Ranganathan emprega este termo enquanto *a generic term used to denote any component be it a basic subject or an isolate of a Compound Subject* (Ranganathan, 1957, p. 157), uma unidade singular a partir da qual expressões mais complexas podem ser formadas. Ranganathan exemplifica esta ideia com a seguinte classificação: ‘L,45;421:6;253:F.44’N5’, cuja interpretação se traduz em ‘investigação na cura da tuberculose dos pulmões usando raios-X realizada na Índia em 1950’, conforme ilustrado pela Figura 1.

Classe principal	Personalidade	Matéria	Energia			Espaço	Tempo
Medicina	Pulmões	Tuberculose	Tratamento	Investigação	Raios-X	Índia	1950
L ,	45 ;	421 :	6 ,	253 :	f .	44 ‘	N5

Figura 1 — Interpretação do significado das facetas de uma Classificação de Cólón.

FONTE: elaboração própria.

As facetas utilizadas na CC obedecem a cinco categorias principais, *Personality, Matter, Energy, Space e Time*, popularizadas pelo acrónimo *PMEST*, categorias essas que não obedecem a uma hierarquia particular, sendo a sua ordem evocada pelos sinais de pontuação precedentes. Contudo, a noção de hierarquia não se encontra totalmente ausente, apenas simplificada. As classes principais encontram-se organizadas hierarquicamente em classes e subclasses. Ranganathan propôs 31 classes generalistas, tais como Matemática, História, Medicina e Religião, semelhantes às da classificação de Dewey. Todavia, as classes da classificação facetada distinguem-se desta por apenas se desdobrarem em uma única subclasse, ao passo que na classificação de Dewey é possível ir acrescentando subclasses sucessivas.

Apesar do raciocínio inovador de Ranganathan, Gartner (2018, p. 80) argumenta que a classificação facetada também sofre de falta de capacidade de interoperabilidade face a classificações como as de Dewey, cujos valores são universalmente reconhecidos e utilizados, ao passo que o significado das facetas pode variar dependendo da avaliação do catalogador. O resultado final produzia também expressões extremamente longas e de difícil compreensão, uma

desvantagem que sacrificou a sua adoção generalizada nas bibliotecas da época. Contudo, num contexto digital em que as condicionantes físicas não se colocam, a classificação facetada tem-se assumido como a base de muitos sistemas de classificação, conforme exemplificado por uma pesquisa no catálogo da *Web of Science*<sup>22</sup> (Figura 2). Neste caso, as facetas permitem a refinação dos resultados de acordo com a frequência de citação, o ano de publicação, as categorias temáticas na base de dados e tipo de documento, entre outros. Desde bases de dados académicas, a serviços de comércio *online* como a Amazon, e-Bay ou AliExpress, passando pelas bibliotecas universitárias, os sistemas de pesquisa que hoje nos são tão familiares e permitem recuperar recursos com facilidade assentam na combinação de facetas, enquanto formas de metadados.

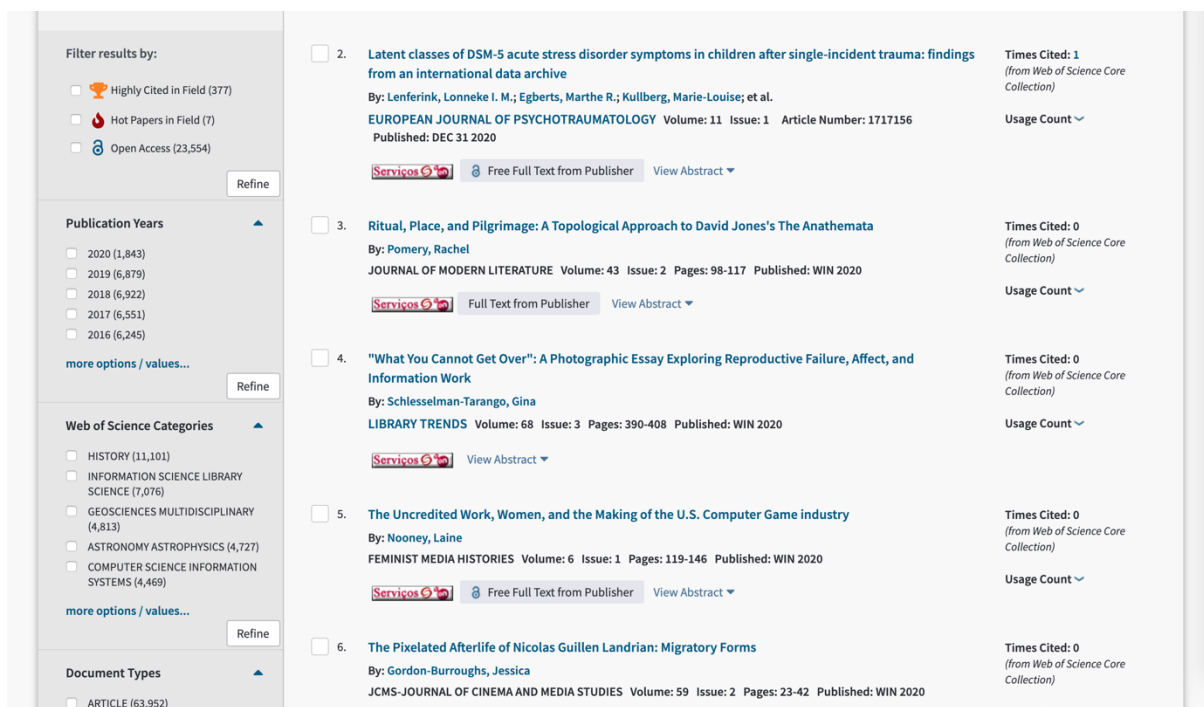


Figura 2 — Pesquisa facetada na Web of Science. As facetas são as seleções de pesquisa, à esquerda.

FONTE: Captura de ecrã em maio de 2020.

No contexto digital, a combinação de facetas é fácil, encontrando-se à distância de um clique, e a verbosidade das expressões encontra-se ausente e simplificada a campos de pesquisa. No exemplo apresentado, é possível refinar a pesquisa com base nos artigos mais citados, nos anos de publicação e no tipo de documento, tornando assim possível reduzir um enorme volume documental a um conjunto interpretável, controlável e pertinente de resultados.

<sup>22</sup> <https://apps.webofknowledge.com>

Todavia, importantes passos ainda tinham de ser dados até chegarmos ao grau de disponibilidade em que os recursos se começam a encontrar atualmente na Internet. Para que isto seja possível, é necessário o cumprimento de dois critérios: 1) a capacidade dos metadados serem interpretados por máquinas; e 2) a sua formulação segundo normas. O primeiro critério é indispensável quando os volumes das coleções são de tal magnitude que seria preciso mais de uma vida para os descrever — se os metadados estivessem num formato apenas compreensível por humanos, não seriam passíveis de automatização, logo todo o *input* teria de ser manual. O segundo critério assegura que os metadados automatizáveis se encontram estruturados de forma previsível, facilitando a sua recolha, interpretação e recontextualização por computadores no momento da pesquisa.

A segunda metade do século 20 viu aparecer os primeiros computadores, que lançaram as bases para o problema da automatização dos metadados. Desde então, os profissionais da informação têm vindo a procurar métodos de expressar os metadados de forma automatizada, passível de ser legível por máquinas, ao mesmo tempo que várias normas de descrição competem por um lugar privilegiado no seio de cada comunidade.

Como tem vindo a ser hábito ao longo da história, os bibliotecários foram pioneiros em encontrar uma solução para os desafios lançados pelo tratamento da informação em ambiente digital. Por volta da década de 1960, a Biblioteca do Congresso<sup>23</sup>, que já havia montado uma tradição de venda de fichas de catálogo normalizadas e de constituição de *union catalogs*, começou a dar os primeiros passos para a automação da partilha eletrónica de dados bibliográficos (Rather & Wiggins, 1989). A resposta encontrada foi a adoção do formato MARC, um dos primeiros formatos para a criação de registos interoperáveis.

O termo ‘MARC’ é um acrónimo de *MAchine Readable Catalog*, estando a sua função perfeitamente espelhada no nome: afirmar-se como um formato para a partilha de dados em catálogos eletrónicos. Como tal, *it was intended to produce catalogue records capable of processing by computer, not only for searching (its primary function) but also for editing, updating, transferring between systems and sharing* (Gartner, 2016, p. 28). Numa palavra, procurava-se a interoperabilidade, que consiste na capacidade automática de comunicação e partilha de dados entre diferentes catálogos e que constitui um dos principais desafios para os profissionais da informação hodiernos.

Embora sendo uma iniciativa coletiva originada no seio da BC, a grande força motora do desenvolvimento do formato MARC foi Henriette Avram (1916-2006), uma das figuras mais importantes do século 20 ao nível do processamento automático de metadados e da normalização em geral. O percurso de Avram é interessante. Com uma primeira formação em programação, numa

---

<sup>23</sup> Uma história pormenorizada do percurso e da constituição da Biblioteca do Congresso por ser consultada em Library of Congress (2017).

época em que as ferramentas disponíveis para programar eram rudimentares<sup>24</sup>, começou a trabalhar para a *National Security Agency*, curiosamente a mesma organização exposta por Snowden que continua tão interessada nos nossos metadados. Em 1965 é contratada pela BC, que começara a dar os primeiros passos na automação dos metadados, onde constrói uma carreira de quase três décadas, até 1991, que revolucionou a catalogação e lhe permitiu afirmar-se como uma *librarian by achievement* (American Libraries, 1997, p. 70).

O projeto MARC teve uma fase piloto iniciada em janeiro de 1966, tornada operacional em novembro de 1966, e lançado como modelo de subscrição para outras bibliotecas em 1969<sup>25</sup>. O projeto consistia no desenvolvimento de um formato de comunicação para os dados bibliográficos. Mais especificamente, o objetivo era *to test the practicability and usefulness of a central source for converting cataloging data to machine-readable form and distributing it to the library community* (Rather & Wiggins, 1989, p. 856). De modo a responder ao desafio, Avram envolveu-se de tal forma na tarefa que colocou numa parede uma larga impressão de uma ficha de catálogo e, em conjunto com os seus colegas, analisou ao detalhe todas as partes constituintes do registo bibliográfico (Pattie, 1998, p. 72). O resultado foi o formato MARC, uma das primeiras aplicações de ferramentas informáticas para o processamento de dados bibliográficos (Rather & Wiggins, 1989), que se tornou influente a nível mundial e que ainda hoje se encontra em uso.

Um registo MARC (*MARC record*) é composto por três elementos, a estrutura do registo, os designadores do conteúdo e o conteúdo. Conforme explica a BC num definição moderna:

*The record structure is an implementation of the international standard Format for Information Exchange (ISO 2709) and its American counterpart, Bibliographic Information Interchange (ANSI/NISO Z39.2). The content designation — the codes and conventions established explicitly to identify and further characterize the data elements within a record and to support the manipulation of that data — is defined by each of the MARC formats. The content of the data elements that comprise a MARC record is usually defined by standards outside the formats. Examples are the International Standard Bibliographic Description (ISBD), Anglo-American Cataloguing Rules, Library of Congress Subject Headings (LCSH), or other cataloging rules, subject thesauri, and classification schedules used by the organization that creates a record (Library of Congress, 2019b).*

Como se percebe, tanto a estrutura como o conteúdo de um registo MARC é fortemente governado por outras normas que foram sendo desenvolvidas ao longo dos anos. No âmbito desta

---

<sup>24</sup> Conforme explica Avram, recordando o início da sua carreira: *You were on your own with far less than perfect tools to learn from. The men —women— were quickly separated from the boys, and the number of people that made it through to become programmers were few indeed. It was an exciting time* (Rather & Wiggins, 1989, p. 860).

<sup>25</sup> Para uma história detalhada destes desenvolvimentos, *vide* Avram (1975).

discussão focamo-nos apenas na *content designation*, isto é, no modo como o formato MARC estrutura os metadados. Para tal, um registo MARC revela três elementos: um *variable data field*, um *indicator position*, e um *subfield code* (Figura 3).

Leader/00-23	*****nam##22*****a#4500
001	<control number>
003	<control number identifier>
005	19920331092212.7
007/00-01	ta
008/00-39	820305s1991###nyu#####001#0#eng##
020	##\$a0845348116 :\$c\$29.95 (£19.50 U.K.)
020	Indicator positions ##\$a0845348205 (pbk.)
040	##\$a[organization code]\$c[organization code]
050	14\$aPN1992.8.S4\$bT47 1991
082	04\$a791.45/75/0973\$d1948-
100	1#\$aTerrace, Vincent,\$d1948- Reminiscências
245	Variable data field 10\$aFifty years of television:\$bba guide to series and pilots, 1937-1988 /\$cVincent Terrace.
246	1#\$a50 years of television
260	##\$aNew York :\$bCornwall Books,\$c1991.
300	##\$a864 p. ;\$c24 cm.
500	##\$aIncludes index.
650	Subfield codes #0\$aTelevision pilot programs\$zUnited States\$vCatalogs.
650	#0\$aTelevision serials\$zUnited States\$vCatalogs.

Figura 3 — Exemplo de um registo MARC de um livro. Extraído de Library of Congress (2019a) e anotado a vermelho pelos autores.

Tomemos por exemplo o campo ‘245’. Sem entrar em detalhes, os primeiros três números correspondem ao *variable data field*, geralmente identificado por uma *tag*, equivalente a um campo bibliográfico tal como autor, título ou publicação. Os dois dígitos seguintes, ‘10’, correspondem aos *indicator positions*, isto é, a instruções para a interpretação e processamento dos dados. Estes valores são independentes, lidos como ‘1’ e como ‘0’, significando neste caso o ‘1’ que o título (campo ‘245’) deve ser tomado como uma forma de entrada adicional, ao passo que o ‘0’ significa que nenhum carater (‘0’) deve ser ignorado quando do processamento<sup>26</sup>. Por fim, a expressão ‘\$a’ corresponde a um *subfield code*, isto é, a dois caracteres que delimitam os *data elements* e que indicam um tratamento em separado (Library of Congress, 2019b). Como se depreende pelo exemplo, refletem subcampos dentro do conteúdo. No exemplo de ‘260’, o código ‘\$a’ indica cidade de publicação, ‘\$b’ a editora e ‘\$c’ a data de publicação.

<sup>26</sup> Esta funcionalidade é especialmente usada quando o título começa com uma proposição, como *The History of Biology*. Neste caso, um indicador ‘X4’ indica que os primeiros 4 espaços devem ser ignorados, contribuindo assim para a homogeneização dos dados.

Apesar de ser um formato para a partilha eletrónica de dados, o formato MARC apresenta um certo grau de semelhança com o contexto bibliográfico de que surgiu. Gartner (2016, p. 30) identifica duas reminiscências com as fichas de catálogo. A primeira é o uso redundante de pontuação como forma de separação de subcampos. Por exemplo, em '260' encontra-se um ':' e ',' antes dos respetivos códigos. Estes caracteres são desnecessários, uma vez que a delimitação dos subcampos é já indicada pelos códigos '\$X'. Contudo, a vírgula e os dois pontos eram os separadores usados nas fichas de catálogo que, por familiaridade, foram mantidos no formato MARC.

A segunda reminiscência diz respeito à criação de campos distintos para entradas principais e adicionais. Por exemplo, o código '100', correspondente ao título, indica que o conteúdo que se segue deve ser usado como entrada principal no registo bibliográfico. Adicionalmente, os indicadores '10' em '245' fornecem instruções para esse campo ser processado como entrada adicional. Num catálogo analógico tradicional existia um registo que servia como entrada principal, equivalente a um dos campos mais usados para a recuperação de uma obra, geralmente o título. Fichas de catálogo adicionais, mais sumárias, poderiam ser adicionadas e redirecionavam para a entrada principal. Como visto anteriormente, esta prática justificava-se pela morosidade do processo de criação de entradas adicionais que, embora simplificado pela adoção das fichas de catálogo, exigia ainda assim a criação manual de cada registo adicional. Em ambiente eletrónico, em que os dados são processados automaticamente, qualquer elemento pode constituir um ponto de acesso, pelo que a distinção entre pontos de acesso primários e secundários se torna artificial.

Esta análise permite realçar, uma vez mais, a natureza parcial dos metadados, enquanto produto humano que reflete a visão de uma comunidade. Ainda assim, estas características não impediram o formato MARC de se afirmar como um formato de referência para a partilha de dados bibliográficos e de constituir um sucesso global. Um dos principais contributos foi possibilitar a criação de catálogos conjuntos automatizados, sucessores dos catálogos conjuntos da BC mas a uma escala significativamente maior (Gartner, 2016). Um exemplo contemporâneo de um catálogo conjunto é o já referido *WorldCat*, o maior catálogo do mundo cuja existência só foi possibilitada pela aplicação do formato MARC e da adoção generalizada de normas.

Avram percebeu desde cedo a importância da normalização para a partilha de registos, revelando um *unyielding commitment to standards* (Rather & Wiggins, 1989, p. 857) desde o momento da conceção do formato MARC. De modo a tornar o formato MARC uma norma de referência internacional no contexto das bibliotecas, colaborou com a *American Library Association* e com o *American National Standards Institute* (ANSI). Como resultado, o formato foi estabelecido como uma norma ANSI em 1971, e uma norma da *International Standards Organization* (ISO) em 1973. Isto não impediu o desenvolvimento de várias cambiantes nacionais do formato MARC, nomeadamente o USMARC, o CAN/MARC ou até o RUSMARC, os formatos MARC dos EUA, do Canadá e da Rússia, respetivamente. Existe ainda o UNIMARC, ou *Universal MARC format*, que foi integrado com as versões norte-americanas e canadianas do MARC

resultando na criação do MARC21, em 1999. Atualmente, o MARC21 a versão dominante, tratando-se de um formato criado a pensar na partilha de informação no século 21<sup>27</sup>.

#### 1.2.4 A era da Internet

Enquanto que o mundo das bibliotecas se encontrava ocupado no desenvolvimento do formato MARC, na década de 1960 ocorreram profundas transformações ao nível da comunicação da informação, tanto no imaginário, como no que toca ao aparecimento de novas tecnologias. A mais marcante foi a Internet, que pode ser entendida como a culminação de tecnologias de comunicação anteriores tais como o telégrafo, a rádio, o telefone e o computador (Leiner et al., 1997) que tem marcado de modo assinalável o período em que vivemos a ponto de o sociólogo Manuel Castells ter caracterizado a nossa sociedade atual como a *network society*, marcada pela organização em rede da economia, da sociedade e da cultura (Castells, 2010). Em 1960, Joseph Licklider (1915-1990), psicólogo, cientista da computação e investigador no *Massachusetts Institute of Technology* (MIT), começa a desenvolver uma série de ideias necessárias para a criação de uma *intergalactic network*<sup>28</sup>, a primeira descrição das interações sociais possibilitadas pela comunicação em rede (Licklider, 1963). A sua visão era a da criação de uma rede de computadores globalmente conectados a partir dos quais as pessoas podiam aceder aos conteúdos, independentemente da localização geográfica.

Licklider foi o primeiro diretor de um programa de investigação em computação na *Defense Advanced Research Projects Agency* (DARPA), uma agência do Departamento de Defesa dos EUA responsável pelo financiamento de projetos de desenvolvimento de tecnologia de ponta para fins militares. Ao longo da sua chefia, inspirou vários colaboradores sobre a importância do conceito de rede. Um desses colaboradores foi Lawrence G. Roberts, investigador do MIT que, em 1967, desenvolve num artigo o conceito de rede computacional, que designou por ARPANET (L. Roberts, 1967), e que viria a tornar-se a base da Internet (Dennis, 2019).

---

<sup>27</sup> O formato MARC21 assenta na norma NISO/ANSI Z39.2 e permite que utilizadores de *softwares* distintos comuniquem e partilhem dados (Joudrey & Taylor, 2018). Pensado de forma a prover a comunidade internacional de uma forma mais acessível de representação bibliográfica, o MARC21 consiste num conjunto de cinco formatos de comunicação. Para maior detalhe, consultar *Library of Congress* (2019b).

<sup>28</sup> Este termo surge pela primeira vez numa circular interna de 1963 escrita por Licklider (1963) aos seus colegas. A visão global de uma rede intergaláctica foi sendo detalhada numa série de notas escritas ao longo de vários anos. A primeira, de 1960, intitulada *Man-Computer Symbiosis*, retrata o seu pensamento sobre o desenvolvimento das interações humano-computador (Licklider, 1960). Na segunda nota, *On-Line Man-Computer Communication*, publicada em 1962, expande com Welden Clark o conceito de interação social através da rede computacional (Licklider & Clark, 1962). Em 1968, em co-autoria com Robert Taylor, publica *The Computer as a Communication Device* onde é discutido o uso de sistemas eletrónicos como um método eficiente de comunicação entre pessoas (Licklider & Taylor, 1968).



Até aos inícios da década de 1960, as tentativas de comunicação entre máquinas recorriam ao uso de circuitos numa linha telefónica, uma opção considerada inadequada (Leiner et al., 1997). Em 1961 Leonard Kleinrock, na proposta da sua tese doutoral, aborda pela primeira vez o método de *packet switching*, que consiste na decomposição dos dados em blocos menores de modo a possibilitar a sua comunicação através de uma rede (Kleinrock, 1971). Descoberta uma técnica eficiente de comunicação de dados, a investigação focou-se na procura de uma forma de colocar os computadores a comunicar em rede.

Para tal, era necessário criar uma linguagem comum para que computadores de diferentes redes pudessem comunicar. Em 1974, Vinton Cerf e Robert Khan formularam o familiar *Transmission Control protocol/Internet Protocol Suite* (TCP/IP) como forma de processamento dos pacotes de dados (Cerf & Kahn, 1974). O protocolo TCP/IP substituiu um anterior *Network Control Protocol* (NCP) que se revelara insuficiente por não ser capaz de processar dados em nós da rede exteriores à ARPANET. Com a transição para o TCP/IP enquanto protocolo padrão de comunicação em rede, tornou-se possível ligar redes, nascendo assim a Internet (Sherry & Brown, 2004, p. 119). Por conseguinte, a Internet era já um terreno fértil em experimentação, com milhares de computadores conectados, décadas antes de Tim Berners-Lee apresentar, em 1989, um conjunto de ideias revolucionárias para a partilha de documentos numa rede, hoje conhecida como a *World Wide Web* (WWW) ou, simplesmente, *Web*.

O desenvolvimento da web estabeleceu os alicerces para a comunicação da informação na Internet tal como é feita atualmente. Nos inícios da década de 1990 sucederam-se diversas iniciativas para a partilha de metadados. Caplan (2000, p. 1) descreveu o movimento de metadados enquanto *a blooming garden traversed by crosswalks, atop a steep and rocky road*. Um jardim florescente devido à proliferação de projetos; e um caminho rochoso pelas dificuldades inerentes a cada implementação, especialmente pelas expectativas de interoperabilidade criadas sobre as instituições. Há uma linha fina que separa o benefício da criação de um novo esquema de metadados e o ruído que introduz no sistema, na medida em que cria alternativa, diversificação e potencial confusão. O cientista da computação Andrew Tanenbaum observou que *the nice thing about standards is that you have so many of them. Furthermore, if you don't like any of them, you can just wait for next year's model* (Tanenbaum, 2003, p. 175). O tom é evidentemente irónico. De forma a haver interoperabilidade na partilha de metadados, é necessário que as instituições sigam modelos comuns, mas se há demasiadas alternativas, as comunidades dispersam-se pelas opções. Ainda assim, durante a década de 1990, cada área juntou-se ao movimento de metadados desenvolvendo os seus modelos.

Nas humanidades e ciências sociais, destacam-se as *Text Encoding Initiative* (TEI) *Guidelines*. Com a primeira versão *draft* distribuída em 1990, e com a primeira versão oficial em 1994, a TEI fornece diretrizes para a codificação de fontes primárias usadas nas humanidades com recurso a linguagens de marcação (TEI, 2020). Para a representação de obras de arte, a *Art Information Task Force* (AITF), financiada pela fundação J. Paul Getty Trust e por uma bolsa de dois anos de montante equivalente pelo *National Endowment for the Humanities*, atribuída ao

*College Art Association*, foi iniciada no início da década de 1990, tendo publicado as *Categories for the Description of Work of Art* (CDWA) em 1996 (Harpring, 2019). Nesse mesmo ano, surgiu a *Visual Resources Association* (VRA) *Core* para a representação de recursos visuais, atualmente mantida pela Biblioteca do Congresso (Visual Resources Association, 2014).

Nos arquivos, um projeto iniciado em 1993 na Universidade Califórnia, em Berkeley, levou à publicação da *Encoding Archival Description* (EAD). O objetivo era o desenvolvimento de uma norma de dados para a descrição de arquivos, de forma análoga à existência do formato MARC para material bibliográfico (Dooley, 1998). Conforme visto anteriormente, o equivalente arquivístico dos catálogos são os instrumentos de pesquisa, pelo que a EAD pretende replicar em ambiente eletrônico os instrumentos de pesquisa impressos. Tal como o seu parente analógico, adota a mesma hierarquia de descrição do geral para o particular, do fundo para o item e inclui elementos muito semelhantes tais como âmbito e conteúdo da coleção, contexto histórico e extensão.

Estes dados podem então ser exportados para um enorme ficheiro digital, como um PDF. Embora a informação apresentada seja completa, o grau de capacidade de automatização destes dados é limitado. Gartner caracteriza os ficheiros resultantes da EAD como *monolithic affairs, less receptive to forming the large union catalogs common in the MARC world* (Gartner, 2016, p. 36). Ainda assim, existem projetos que pretendem agregar coleções de instituições culturais usando esta forma de representação de conteúdos. Um exemplo é o *Online Archive of California* (OAC), que reúne mais de 250.000 imagens digitais e documentos de mais de 200 instituições, incluindo arquivos, bibliotecas e museus (Online Archive of California, 2020). Os guias das coleções descritas apresentam descrições detalhadas dos itens de uma coleção, bem como o seu contexto, estudos históricos e análises efetuadas. No entanto, as possibilidades de pesquisa são muito limitadas, apenas estando disponível uma pesquisa por título da coleção. O conteúdo dos restantes metadados não é facilmente processável uma vez que se encontra em linguagem natural, não em campos facetados de pesquisa que permitiriam um refinamento mais rigoroso e preciso.

Esta falta de automatização não só dificulta a partilha, como também condiciona a descoberta dos recursos descritos. Apesar de tudo, a EAD afirma-se como um importante passo na abertura das coleções na comunidade arquivística ao permitir a criação de instrumentos de pesquisa eletrónicos com uma estrutura baseada na *General International Standard Archival Description* (ISAD(G)) que facilitou a migração para a Internet e reforçou o aparecimento de um consenso na profissão de que *archival description existed to be shared and shared well* (Society of American Archivists, 2019a).

A comunidade bibliotecária foi uma vez mais pioneira a respeito da representação de recursos existentes na *web*. Nos inícios da década de 1990, em particular antes do advento de motores de pesquisa como o Google, a Internet era vista como uma biblioteca gigante cuja catalogação seria feita como anteriormente (Gartner, 2016). Em 1994, houve quem tentasse catalogar a Internet como se fosse um livro, num esforço que hoje nos figura ser, no mínimo, infrutífero (Abbott, 1994). Outros, mais realistas, enveredaram por modelos mais semelhantes à

catalogação eletrónica. A Biblioteca Bodleiana foi pioneira nos desenvolvimentos de metadados ao criar em 1992 uma página de acesso a recursos da sua coleção. Em 1995 o Yahoo foi lançado, inaugurando assim uma tradição de bases de dados muito mais completas e abrangentes. O Yahoo ainda aspirou a criar um índice hierárquico de *websites*, tendo depressa sido suplantado pelo Google, que enveredou por um motor de pesquisa centrado no utilizador (Library of Congress, 2017, p. 19).

A representação de recursos disponibilizados na Internet é particularmente complexa devido à sua enorme diversidade. De entre documentos textuais, páginas *web*, vídeos, imagens, programas, blogs, *tweets*, entre muitos outros, é realmente difícil concordar num conjunto de elementos comuns e suficientes. Este desafio foi respondido por um conjunto bibliotecários, arquivistas e outros especialistas da informação, através de uma *workshop* realizada em Dublin, Ohio, em 1995, patrocinada pelo *Online Computing Library Center* (OCLC) e pelo *National Center for Supercomputing Applications* (NCSA). Era missão da *workshop*:

*to identify the scope of the problem, to achieve consensus on a list of metadata elements that would yield simple descriptions of data in a wide range of subject areas, and to lay the ground-work for achieving further progress in the definition of metadata elements that describe electronic information* (Weibel, Godby, Miller, & Daniel, 1995, Introdução).

À data da realização da *workshop*, existiam dois tipos dominantes de representação de documentos. Por um lado, havia os índices gerados automaticamente pelos motores de pesquisa (por exemplo, Lycos), que continham informação parca. Por outro, havia os registos de catálogo criados por profissionais usando os formatos MARC ou TEI, cujo uso generalizado para a representação de recursos na Internet era difícil devido ao extenso número de elementos usados e ao enorme volume de documentos na Internet. A *workshop* realizada em Dublin pretendia contornar estas duas desvantagens propondo apenas um conjunto essencial de elementos fundamentais para a descoberta dos recursos, sejam eles fichas de catálogo ou páginas *web* (Zeng & Qin, 2016, p. 403). Conforme explicam Weibel *et al.*:

*If only a small amount of human effort were required to create the record, more objects could be described, especially if the author of the resource could be encouraged to create the description. And if the description followed an established standard, only the creation of the record would require human intervention; automated tools could discover the descriptions and collect them into searchable databases* (Weibel et al., 1995, Introduction).

O resultado final das discussões foi a publicação da *Dublin Core Metadata Initiative* (DCMI) com 15 elementos essenciais, designados por *Dublin Core Metadata Element Set* (DCMES). Devido à sua simplicidade, é provavelmente a norma de metadados mais utilizada, tendo o conjunto essencial de elementos sido registado como uma norma internacional (ISO, 2017). Apesar do seu propósito inicial de representação de recursos na Internet, atualmente não é feita qualquer limitação concetual ao que pode ser considerado um recurso.

A conceção de um modelo de dados simplista adequado a todos os propósitos permitiu a adesão generalizada ao movimento dos metadados iniciada na década de 1990. Contudo, a partir do momento em que o conceito de metadados entra em uso a uma escala global, também os recursos por eles representados se diversificam. Conforme visto, inicialmente usados para representar peças de museu, como no museu de Ennigaldi-Nanna, documentos administrativos, como em Ebla, ou de registos bibliográficos como na Biblioteca de Alexandria, os metadados aplicam-se atualmente a tantos recursos quanto aqueles que o génio e a atividade humana produzem. Desde *tweets*, páginas *web*, vídeos, fotografias, documentos de arquivo, objetos de arte, pessoas, relações, propriedades, virtualmente qualquer coisa é passível de representação por metadados. Por conseguinte, atualmente, o conceito de metadados é transversal, não pertencendo exclusivamente à Ciência da Informação e à Arquivística, mas sendo usado em muitas outras áreas, como a Informática, a Geografia e a Biologia.

Por este motivo, constata-se a existência de um mar de normas e modelos de dados passíveis de adoção, que refletem um sem-fim de abordagens possíveis, dependendo do contexto e do uso dos metadados, estando cada uma voltada para o seu domínio específico. Por exemplo, existe a norma ONIX para o comércio digital de livros (EDItEUR, 2009), STEP para produtos industriais (ISO 10303, 2020), MPEG-7 para conteúdo multimédia (ISO/IEC 15938, 2002) ou FOAF para agentes na *web* (Brickley & Miller, 2014). Em 2010, Jenn Riley elaborou um diagrama com as várias normas existentes à época (Figura 4).

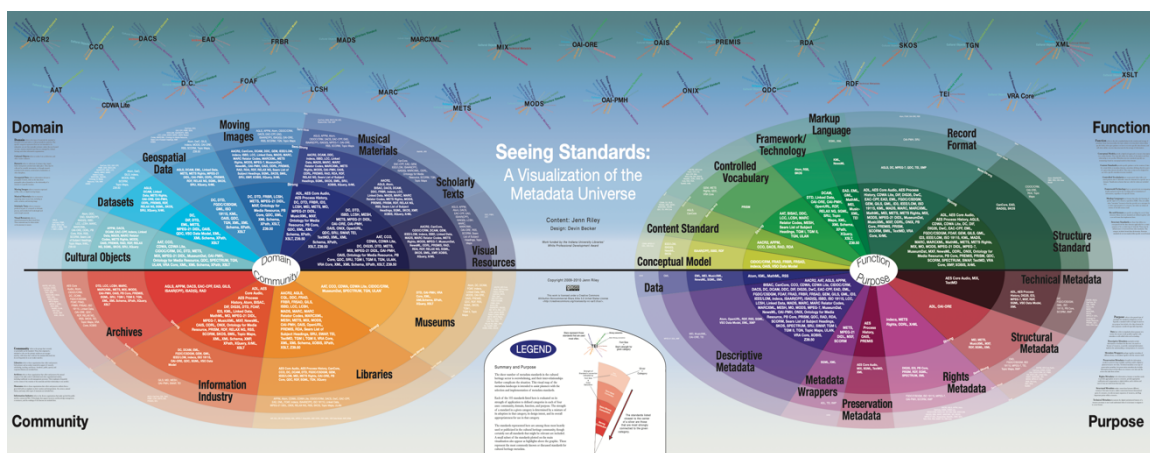


Figura 4 — Seeing standards: a visualization of the metadata universe.

FONTE: extraído de Riley (2010).

A análise imediata da observação da compilação de Riley (2010) é a imensa diversidade de normas. Nota-se também uma divisão por comunidades, repartida entre museus, bibliotecas, indústria da informação e arquivos. De forma não-surpreendente, o movimento de metadados tem vindo ao longo das três últimas décadas a fragmentar comunidades, não a uni-las. Este fenómeno não é de admirar, uma vez que esta divisão já se observava na prática nas instituições, divisão essa criada no século 19 com a formulação do princípio da proveniência e o anterior cisma criado entre

arquivos, bibliotecas e museus. Como a criação de metadados é uma atividade humana, ela reflete as práticas culturais e sociais dos indivíduos que a geram. Como sabiamente observa Gartner (2016, p. 37), *metadata is a human construct and so will always reflect its provenance*. Nos últimos anos, têm felizmente começado a aparecer projetos que pretendem conciliar estes parentes próximos mas artificialmente separados. Um exemplo é o projeto Europeana<sup>29</sup>, que constitui um repositório europeu para objetos do setor cultural, reunindo coleções de arquivos, museus e bibliotecas europeias e fornecendo acesso a uma vasta panóplia de recursos tais como livros, música e obras de arte (Europeana, 2020).

De forma análoga à abordagem seguida pela Europeana, nota-se uma crescente convergência para uma representação comum e integrada dos recursos, independentemente de provirem de coleções de arquivo, museu e biblioteca. O modelo de dados *Records in Context(s)*, publicado pelo *International Council of Archives* em versão rascunho em 2016 (com uma versão 0.2 de 2019) e que, por conseguinte, atua simultaneamente como norma internacional, salienta o entendimento de um documento de arquivo não pode ser feito em isolamento, exigindo antes a representação dos vários contextos em que ele se insere e das relações que estabelece com outros recursos (International Council on Archives, 2016). O grau de sofisticação das tecnologias usadas atualmente para a representação dos recursos, que permite formas mais poderosas de conciliar numa só pesquisa coleções de várias instituições, contribui também para a diluição desta antiga separação institucional, como será discutido em maior pormenor adiante.

Em todo o caso, é inegável que nos encontramos atualmente soterrados de modelos de metadados. Gartner apontou que *we are swimming in a torrent of metadata as resolutely as we have at any point in history: the media that hold it have out of all recognition, human nature has not* (Gartner, 2016, p. 38). Discorda-se que estejamos num momento igual a qualquer outro da história. Nunca antes na história se observou um período em que tão enorme volume de metadados, de tão diferente natureza, de tanta dificuldade de interoperabilidade, e de tamanha pressão em ser capaz de produzir resultados (pesquisas integradas) que funcionem e obedeçam a expectativas crescentes dos utilizadores. Uma solução passa pela adoção de modelos normalizados, que serão discutidas em maior detalhe adiante. Por enquanto, face aos diferentes usos de metadados, discutem-se de seguida as função, tipos e propósitos que estes podem assumir no contexto da representação em arquivística.

### 1.3 Tipos e funções

Embora os bibliotecários e os arquivistas tenham sido os especialistas que tiveram a seu cargo a criação de metadados ao longo dos tempos, este panorama está a mudar devido à prevalência atual do uso de ferramentas digitais em todos os setores de atividade, com a

---

<sup>29</sup> O conteúdo deste tipo de plataformas agregadores são apenas metadados, uma vez que tanto os objetos digitais como as suas representações permanecem nos websites das instituições (Bell, 2013, p. 292).

consequente geração de metadados. Por um lado, o número de criadores de metadados não especializados tem estado a aumentar em virtude dos processos de indexação social<sup>30</sup>. Por outro, o tipo de comunidades que exploram a criação e o processamento de metadados não se cinge unicamente à biblioteconomia e à arquivística. Comunidades e produtores de conteúdos de áreas como as ciências naturais, ciências sociais, comércio digital, gestão, ciências da saúde, biologia, geografia e ciências computacionais estão constantemente envolvidas na procura de formas mais eficientes e interoperáveis de partilha dos seus metadados especializados.

Ainda que este trabalho tenha como enfoque o campo de atuação da Arquivística, o desenvolvimento de metadados especializados noutros setores da sociedade não pode ser ignorado, uma vez que os objetos informacionais produzidos no decurso das atividades destas várias comunidades também constituem o objeto da Arquivística e do modelo de requisitos que se pretende construir. Assim, no contexto da gestão da informação, a crescente diversificação de tipos e funções de metadados tem uma importância incontornável para o nosso estudo.

Um exemplo bastante presente é o do comércio digital. Muitas das transações ocorridas atualmente no contexto digital não geram necessariamente documentos, entendidos como ficheiros geralmente PDF bem-definidos como faturas, confirmações de encomenda ou recibos. A consolidação de grandes retalhistas de comércio *online*, com a consequente adoção de sistemas de informação nos quais a informação respeitante à transação é registada através de números de encomenda e *data logs* das transações bancárias, tem vindo a enfraquecer a importância do documento de arquivo tradicional enquanto uma realidade bem-definida e delimitada por um ficheiro. Por exemplo, a Amazon ainda dispõe da opção de geração de um *invoice* (fatura) com os detalhes de uma encomenda, que atua como recibo e fatura de uma transação. A sua função como evidência é limitada, uma vez que reclamações perante a Amazon são resolvidas tendo por base a informação no sistema informático. Este *invoice* serve apenas para fazer prova perante terceiros, como a alfândega, no caso de bens importados. Porém, outros retalhistas como o AliExpress erradicaram o documento na sua formulação analógica.

Neste caso, a informação relativa a uma transação é igualmente vertida num suporte — o ambiente digital — mas não num PDF, num objeto digital delimitável. Ao invés, esta informação encontra-se embutida nos detalhes de uma encomenda, gerada e autenticada pelo sistema de informação através de metadados que reportam todos os detalhes necessários, tais como detalhes do produto (título, URL), estado da encomenda (entregue, no país de destino, em processamento), opções de reclamação (produto não entregue, defeituoso), bem como informação de localização (*tracking*)<sup>31</sup>.

---

<sup>30</sup> Pense-se na crescente responsabilidade dos utilizadores para colocação de *tags* nos conteúdos que disponibilizam em plataformas como Youtube e Instagram.

<sup>31</sup> Bell (2013) desenvolve mais diferenças entre metadados do setor cultural e comercial.

O advento de obras variáveis no tempo (*time-based media*) e de objetos gerados de forma dinâmica torna a descrição intelectual dos conteúdos e da estrutura destes recursos uma tarefa largamente mais complexa do que a descrição de documentos textuais (Ramesh, Vivekavardhan, & Bharathi, 2015, p. 196), num entendimento dos objetos que recorda o conceito de '*networked information resources*' de Lynch, formulado em 2004<sup>32</sup>. Por conseguinte, a Arquivística vê-se forçada a integrar no seu vocabulário conceitos, usos e práticas de metadados que normalmente não a caracterizavam, expandindo o seu campo de atuação. Num contexto atual de crescente partilha, disseminação e reuso na sociedade dos metadados que representam os objetos digitais, dificilmente se pode afirmar que cada área tem um conjunto específico de metadados com os quais atua. No entanto, podem referir-se tipos e funções predominantes, consoante as áreas e os usos mas, de uma forma geral, deve manter-se abertura para a necessidade de adotar determinado tipo ou função. Assim, qualquer tentativa de caracterização de tipologia e funções de metadados usados em Arquivística não pode ter uma ambição definitiva nem exaustiva. Como tal, a discussão que se segue constitui apenas um exercício de revisão dos tipos mais frequentes na literatura, sem excluir a possibilidade de aparecimento de outros tipos consoante cada uso específico.

### 1.3.1 Tipos de metadados

De um modo geral, os metadados fornecem aos utilizadores *input, control, or information support* (Wilson, 2012, p. 5), razão pela qual assumem uma importância multifacetada. Os metadados são úteis para anotar os recursos, auxiliar os utilizadores a pesquisar, avaliar e encontrar os recursos informacionais (Beyene & Godwin, 2018) e usados enquanto facetas para filtrar os resultados da pesquisa (Holman et al., 2012; X. Niu, Zhang, & Chen, 2014). Por este motivo, podem atuar como *positive search abandonment* (Stamou & Efthimiadis, 2010, p. 1), uma condição segundo a qual as necessidades informacionais dos utilizadores são satisfeitas pela mera contemplação da lista de resultados. A título de exemplo, se um indivíduo interessado em saber qual a prevalência do termo 'metadados' na literatura científica arquivística de determinado período realizar uma pesquisa numa base de dados de referência, e essa pesquisa recuperar apenas três artigos, a informação revelada pelos metadados é suficiente para esse indivíduo saber que a literatura é escassa.

---

<sup>32</sup> Para Lynch (2004, p. 5), *networked information resources* são *objects or interactive services that are available through the network in the broadest sense. These resources include files (which can have semantics as text, images, structured datasets, digital audio or video, or programs); interactive services such as data, audio or video feeds, interactive sessions using protocols such as Telnet; electronic mail based services such as reflectors and List Servers. They also include aggregations of information such as databases, file archives accessible through anonymous FTP, and archives from newsgroups or mailing lists; typically these aggregations support some type of browsing or searching once the aggregate object has been selected. Thus one can view networked information resources as having a wide range of granularities, and some networked information resources as being hierarchically organized.*

Embora os metadados sejam por natureza descritivos, podem ser agrupados em diversas categorias de acordo com o seu tipo, e cumprir uma diversidade de funções. Na verdade, *metadata not only identifies and describes an information object; it also documents how that object behaves, its function and use, its relationship to other information objects, and how it should be and has been managed over time* (Gilliland, 2016, p. 7). Como tal, os metadados são geralmente repartidos por categorias que refletem aspetos da sua funcionalidade<sup>33</sup>. Tais categorias são usualmente designadas por tipos de metadados, embora certos autores usem erroneamente a expressão *metadata types* para se referir antes a elementos de dados (*data elements*), isto é, a elementos de descrição como ‘título’ ou ‘autor’ (Kovačević, Ivanović, Milosavljević, Konjović, & Surla, 2011; Lipinski, Yao, Breitinger, Beel, & Gipp, 2013).

Na Tabela 2 é apresentado um sumário representativo do entendimento que diversos autores fazem da tipologia de metadados. Não há um consenso generalizado sobre a hierarquia dos tipos uma vez que, embora a designação dos tipos seja relativamente uniforme, observa-se uma variação na ênfase dada a determinada categoria quando outras são colocadas como subcategoria dessa. Por exemplo, tanto NISO (2017) como Beyene & Godwin (2018) entendem os *rights metadata*, *technical metadata* e *preservation metadata* como subtipo de *administrative metadata*, ao passo que para Han (2015) *technical metadata* constitui um tipo principal. Uma explicação pode residir no facto de este último autor ter uma perspetiva da preservação digital, segundo a qual os metadados técnicos assumem uma importância central na representação dos objetos.

Neste sentido, entende-se que a hierarquia entre os vários tipos não merece particular importância, uma vez que varia de acordo com o interesse dos autores em relevar certos aspetos dos metadados, de acordo com os contextos em que a sua formação e objetivos de investigação se inserem. Cada tipo de metadados pode assumir maior ou menor destaque consoante a função prioritária para a entidade que recolhe ou gera os metadados. Com esta tabela pretende-se ilustrar os principais tipos de metadados ocorrentes na literatura, bem como exemplos das suas definições que, de um modo global, são consistentes. De entre estes tipos destacam-se os metadados descritivos (*descriptive metadata*), metadados administrativos (*administrative metadata*), metadados estruturais (*structural metadata*), metadados técnicos (*technical metadata*), metadados de preservação (*preservation metadata*), metadados de uso (*use metadata*) e metadados de direitos de utilização (*rights metadata*).

Tabela 2 — Síntese dos tipos de metadados mais recorrentes na literatura científica e respetivas definições.

---

<sup>33</sup> Gilliland (2016), pp. 12-13 refere-se ainda a atributos e características dos metadados.



---

Ramesh *et al.* (2015 pp. 195-196)

**Descriptive metadata:** Includes curatorial information such as name of the file, creator, associated dates, description, summary, locations etc., is the primary source for users to access. This provides content- and-context-related specific information of an information resource useful in its discovery and identification.

**Administrative metadata:** The information pertaining to the management, organisation, provision and method of access ... described for digital collections ... it also provides information about preservation, legal rights, ownerships and usage restrictions.

**Structural metadata:** deals with information relating to the external organisation of digital resources. Ex: chapters in a book, or how pages are ordered to form chapters.

**Technical metadata:** The technical qualities of a digital object are described in technical metadata. Ex: channel number, bit-depth, sampling rate, FPS, unique file identifier.

### Rights metadata

**Preservation metadata:** Provides information pertaining to provenance, authenticity, preservation activities, and other technical details (p. 195)

---

Han (2015, pp. 415-416)

**Descriptive metadata:** is most commonly used to describe a resource for identification and discovery. This type of metadata is the most widely used for resource description and discovery via search engines and local search functions. Typical elements are title, author and keywords.

**Structural metadata:** describes the logical structure and components of content. This type of metadata can be used for page-to-page and semantics navigation in delivering digitization materials to enhance users' access experience.

**Technical metadata:** describes technical attributes of digital objects during digital capture and other processes. It typically comes from digitization equipment such as scanners and digital cameras. Examples are hardware, software to produce the digital object, resolutions, file formats and color profiles.

### Rights metadata

### Preservation metadata

(Lopatin, 2010, p. 717)

**Descriptive metadata:** describes physical and intellectual properties of a resource

**Administrative metadata:** provides information on managing a resource and other technical data

—————> **Technical metadata**

**Structural metadata:** indicates how compound objects are put together, for example, how pages are ordered to form chapters

---

(Gilliland, 2016, p. 10)

**Descriptive metadata:** used to identify, authenticate, and describe collections and related trusted information resources

**Administrative metadata:** used in managing and administering collections and information resources

**Technical metadata:** related to how a system functions or metadata behaves

**Preservation metadata:** a related to the preservation management of collections and information resources.

**Use metadata:** related to the level and type of use of collections and information resources

---

(NISO, 2017, p. 6)

**Descriptive metadata:** information about the content of a resource that aids in finding or understanding it

**Administrative metadata:** an umbrella term referring to the information needed to manage a resource or that relates to its creation

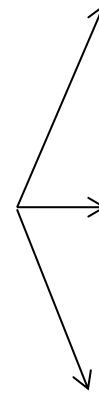
**Structural metadata:** describes the relationships of parts of resources to one another. Ex: pages in a sequence, a table of contents with pointers to the beginnings of milestone sections

**Markup languages:** integrates metadata and flags for other structural or semantic features with content ... these languages mix metadata and content together

**Rights metadata:** details the intellectual property rights attached to the content. Ex: Creative Commons license

**Technical metadata:** information about digital files necessary to decode and render them, such as file type.

**Preservation metadata:** supporting the long-term management and future migration or emulation of digital files, for example, a checksum or hash



---

(Beyene & Godwin, 2018, p. 5)

**Descriptive metadata:** used for resource description, selection and discovery

**Administrative metadata:** helps in management of resources

**Structural metadata:** a “glue” that links digital files belonging to the same set of digital resource. It describes the relationship between parts of a digital resource, such as pages, pages and chapters, chapters and a book, and audio and text transcriptions.

**Rights management metadata**

**Technical metadata**

**Preservation metadata**



---

(ISO/IEC TR 19583-1, 2019)

**Administrative metadata:** the metadata associated with data values, such as metadata describing when and who created the data, who can edit and manage the data, and any other information about the data that is deemed useful, including metadata that describes multimedia data.

**Descriptive metadata:** the metadata used for the discovery and identification of content, such as by librarians, and the metadata that helps to further describe other data, such as metadata that a scientist has observed about continuous or systematically produced data.

**Structural metadata:** the metadata used by those responsible for the management of data in information systems to describe the ‘containers’ of data, for example, the tables and columns in a database managed using the SQL database language.

---

(Cron, 2016)

**Administrative metadata:** used to manage collections of records. Examples include the Transfer Request Number, the Record Group, the name of the person authorized to transfer custody, etc.

**Descriptive metadata:** identifies and describes records. Examples include a photograph’s caption, the title of a book, or the composer of a song.

**Preservation metadata:** the specialized set of information required to preserve and provide access to electronic records. Examples include the file format used to encode a file, the software necessary to view it, or an action taken to maintain it such as the results of a virus scan.

—————> **Technical Metadata** describes aspects of electronic records important to their proper interpretation, rendering, or playback. The type of compression used with a digital image, the audio codec contained in a digital video, or the encryption algorithm used to digitally sign an email are all examples of technical metadata.

**Use metadata:** includes information that describes how records can be accessed or circulated. Metadata identifying copyright status or security classification are examples of use metadata.

FONTE: Elaboração própria, com base nos autores citados.

Conforme mencionado, esta nomenclatura não é exaustiva. Outros autores há que mencionam outros tipos, tais como *accessibility metadata* (Beyene, 2017; Cheetham et al., 2014), *child-driven metadata* (Beak, 2014), *mechanical metadata* (Ramesh et al., 2015) ou *library, archival and museum metadata* (Gilliland, 2016). Num contexto de crescente diversificação e abundância de recursos informacionais, é expectável que esta diversidade apenas tenda a aumentar, uma vez que *the number of metadata types is proliferating as the resources metadata are intended to manage proliferate as well* (Schottlaender, 2003, p. 22).

Contudo, uma análise mais detalhada do uso dos termos por estes autores revela informações interessantes. Embora a amostra seja relativamente diminuta, uma vez que se trata de oito fontes, estes autores são frequentemente citados na literatura, pelo que se acredita que são relativamente representativos de uma tendência no campo científico. Na seguinte Tabela 3 apresenta-se uma leitura destes dados que quantifica a frequência do uso dos tipos de metadados bem como o seu lugar na hierarquia tipológica.

Tabela 3 — Análise quantitativa da ocorrência da tipologia de metadados.

Designação	Total	Principal	Subtipo	Especificação
Descriptive	8	8	-	
Administrative	7	7	-	
Structural	6	6	-	3x Administrative [1, 2, 3] 1x Preservation [4]
Technical	7	3	4	3x Administrative [1, 2, 3] 1x Não-Especificado [5]
Rights	4	-	4	3x Administrative [1, 2, 3] 1x Não-Especificado [5]
Preservation	6	2	4	

NOTA: [1] Laptin (2010) ; [2] Niso (2017) ; [3] Beyene & Goodwin (2018) ; [4] Cron (2016) ; [5] Han (2015).

FONTE: Elaboração própria.

Uma das principais ilações a retirar de imediato da visualização desta tabela diz respeito à ausência de consenso sobre i) as tipologias a incluir e ii) a hierarquia dessas tipologias entre si. Por este motivo, nesta discussão deve-se partir da humildade de que não existe uma resposta definitiva e universal, mas antes que cada posição de cada autor constitui uma interpretação dependente de um contexto mais ou menos específico de aplicação. Por outras palavras, tanto as designações dos tipos de metadados como as suas hierarquias são flexíveis, dependendo da ênfase que cada autor considera — ainda que subconscientemente — relevante colocar para melhor representar a realidade que pretende descrever através dos metadados. O mesmo raciocínio aplica-se às próprias definições dos tipos, que devem igualmente ser vistas como interpretações, pelo que não se considera apropriado neste trabalho entrar na sua análise em maior detalhe.

Ainda assim, é possível observar algumas tendências nesta amostra. Por um lado, os metadados descritivos são incluídos por todos os oito autores desta amostra, e sempre como tipo principal. Também os metadados administrativos e estruturais aparecem muito frequentemente (7 e 6, respetivamente), sempre como principais. Além disso, é possível constatar que os metadados técnicos, de direito de utilização e de preservação são frequentemente entendidos como subtipos, quer dos metadados administrativos (3 ocorrências), quer sem uma hierarquia especificada (1 ocorrência). Também os metadados de uso, ainda que incomuns (2), são sempre considerados como tipo principal. Estes dados apontam para uma convergência em entender os metadados descritivos e administrativos como quase obrigatórios, sendo também muito frequente a necessidade de inclusão de metadados técnicos e de estrutura.

### 1.3.2 Funções dos metadados

Os metadados podem desempenhar várias funções, por vezes designadas na literatura por propósitos (*purposes*) (Day, 2001; Haynes, 2018) ou ainda papéis (*roles*) (International Council on Archives, 2016). Entendem-se estes termos como sinónimos, no sentido em que todos dizem respeito ao fim a que se destina o uso de determinado tipo de metadados.

A função inerente a cada tipo de metadados advém do facto de que a sua criação não é casuística, tendo antes uma finalidade demarcada. Conforme explica David Haynes, *metadata is something that you collect for a particular purpose, rather than being a bunch of data you collect just because it is there or because you have some public duty to collect it* (Haynes, 2018, p. 11). Segundo o Conselho Internacional de Arquivos (ICA), a descrição dos documentos de arquivo tem como finalidade *the management of records, the preservation of records and ongoing use and reuse* (International Council on Archives, 2016, pp. 6-7).

A formulação do ICA é porventura demasiado simplista, embora completa. Anne Gilliland entra em maior detalhe sobre as funções dos metadados, apontando as seguintes:

- *creation, multiversioning, reuse and recontextualization of information objects*
- *organization and description*

- *validation*
- *search and retrieval*
- *utilization and preservation*
- *disposition* (Gilliland, 2016, p. 14).

Henry Gladney também identifica um conjunto de funções genéricas desempenhadas pelos metadados, destacando como uma das suas funções a de documentar a autenticidade e contribuir para a salvaguarda da integridade:

- *identifying items uniquely worldwide*
- *describing collection items (e.g., author, creation date) including their contexts*
- *supporting retrieval and identification*
- *grouping items into collections within a repository*
- *recording authenticity evidence, including historical audit trails*
- *helping protect item integrity against improper change and unintentional corruption*
- *recording access permissions and other digital rights information*
- *facilitating information interchange between autonomous repositories; and*
- *recording technical parameters describing items' representation* (Gladney, 2007, p. 129).

Michael Day realça igualmente o papel dos metadados na preservação do contexto e da autenticidade. Para este autor, os metadados desempenham as seguintes funções:

- *resource description*
- *resource discovery*
- *administration and management of resources*
- *record of intellectual and property rights*
- *documenting software and hardware environments*
- *long-term preservation management of digital resources*
- *preservation of the context and authenticity of resources* (Day, 2001, p. 1).

Contudo, certos tipos de metadados podem estar mais diretamente implicados em determinadas funções. Barbedo et al. (2007, p. 13) apresentam uma categorização simplificada da função principal desempenhada por cada tipo de metadados, fornecendo também o exemplo de uma norma ou modelo de dados especializada nessa função (Tabela 4).

Tabela 4 — Função desempenhada por cada tipo de metadados e norma correspondente.

<b>Tipo de metadados</b>	<b>Função</b>	<b>Exemplo de modelo ou norma</b>
Descritivos	Acesso	EAD
Preservação	Autenticidade	PREMIS
Técnicos	Preservação, pesquisa	NISO Z39.87
Estruturais	Renderização, reconstituição do objeto	METS

FONTE: Elaboração própria, adaptado de Barbedo et al. (2007, p. 13).

A *National Information Standards Organization* (NISO) aprofunda esta categorização, não só incluindo outros tipos de metadados como *rights metadata*, e *markup languages*, como também reconhecendo que os usos não são exclusivos a um tipo (Tabela 5).

Tabela 5 — Tipos e funções primárias dos metadados.

<b>Metadata type</b>	<b>Example Properties</b>	<b>Primary Uses</b>
Descriptive metadata	Title Author Subject Genre Publication date	Discovery Display Interoperability
Technical metadata	File type File size Creation date/time Compression scheme	Interoperability Digital object management Preservation
Preservation metadata	Checksum Preservation event	Interoperability Digital object management Preservation
Rights metadata	Copyright status License terms Rights holder	Interoperability Digital object management
Structural metadata	Sequence Place in hierarchy	Navigation
Markup languages	Paragraph Heading List Name Date	Navigation Interoperability

FONTE: Extraído de NISO (2017, p. 7).

Da análise comparativa das informações avançadas por estes autores é possível observar que, novamente, não há um consenso absoluto ao nível das funções dos metadados, mas antes



emerge a percepção de que *metadata can be used within one application for several different purposes* (Haynes, 2018, p. 17). Por conseguinte, entender que a um tipo de metadados corresponde uma e uma só função é uma perspectiva redutora. Ao invés, da mesma forma que os tipos de metadados não correspondem a categorias nem hierarquias rígidas, também as funções dos metadados se manifestam de forma igualmente líquida em cada elemento de dados. Por exemplo, pode ser dito que a representação do título e autor de determinado recurso tem como função primária a descoberta e interoperabilidade, mas também se pode afirmar que contribui para a presunção de autenticidade ao nos informar sobre a identidade do autor, e da sua relação com a obra criada.

No contexto deste trabalho não compete proceder a uma análise em profundidade das várias funções dos metadados, uma vez que os objetivos de investigação se relacionam especificamente com a autenticidade (*authenticity*), a fidedignidade (*reliability*) e confiabilidade (*trustworthiness*) dos metadados. A capacidade dos metadados em estabelecer inferências sobre estas características é apontada por autores como Gladney (2007), Day (2001), pelo dicionário PREMIS (2015) e pela NISO (2017), estabelecendo assim os fundamentos teóricos para a realização desta investigação. Porém, não é de excluir que estas características os metadados se possam manifestar em mais do que um tipo de metadados, ou até mesmo em todos. Conforme será visto adiante, a presunção destes atributos estabelece-se de forma progressiva, através da recolha de informações que, em conjunto, contribuem para um indivíduo determinar se, na sua interpretação, determinado recurso é confiável. Assim, nesta fase do trabalho não se deve descartar nenhuma destas funções identificadas na literatura científica, uma vez que cada uma pode vir a desempenhar um papel importante na construção posterior do modelo de requisitos com os elementos fundamentais para capturar estas características de um registo de metadados (*metadata record*). Caracterizados em tipologia e função, analisam-se de seguida as formas de expressão dos metadados.

## 1.4 Formas de expressão

Embora a codificação dos dados não esteja diretamente relacionada com os objetivos de investigação, na medida em que o tema diz sobretudo respeito à estrutura, não à sintaxe, considera-se ainda assim pertinente discutir muito brevemente as formas como os metadados podem ser expressos, uma vez que as tecnologias disponíveis para o efeito condicionam a escolha dos metadados a exprimir e os contextos da sua criação e uso. No seguimento da famosa máxima que afirma que *we shape our tools and thereafter they shape us* (Culkin, 1967, p. 70), as tecnologias disponíveis para a manifestação da estrutura dos metadados não pode ser isolada da própria estrutura que os metadados assumem. Nesta parte apenas se explicam os princípios fundamentais subjacentes a cada sistema de modelação de metadados, sem entrar em detalhes como formatos de serialização ou dificuldades técnicas de implementação, uma vez que importa apenas reter para este tema a conceptualização da representação da informação.

Assim, segue-se uma análise dos modelos de dados com maior uso, dando especial destaque ao modelo em grafo, a tecnologia mais recente para expressão de metadados e que

representa uma mudança de paradigma na forma como os recursos informacionais são representados e os respectivos metadados partilhados, disseminados e reutilizados.

A modelação dos dados consiste na definição dos elementos de dados e da relação entre eles com a finalidade de descrever de que forma os dados podem ser eficientemente usados e representados (Host, Jaksic, & Poscic, 2018). Identificam-se três tipos de modelos de dados principais: o relacional, o hierárquico e o modelo em grafo, esquematizados na Figura 5.

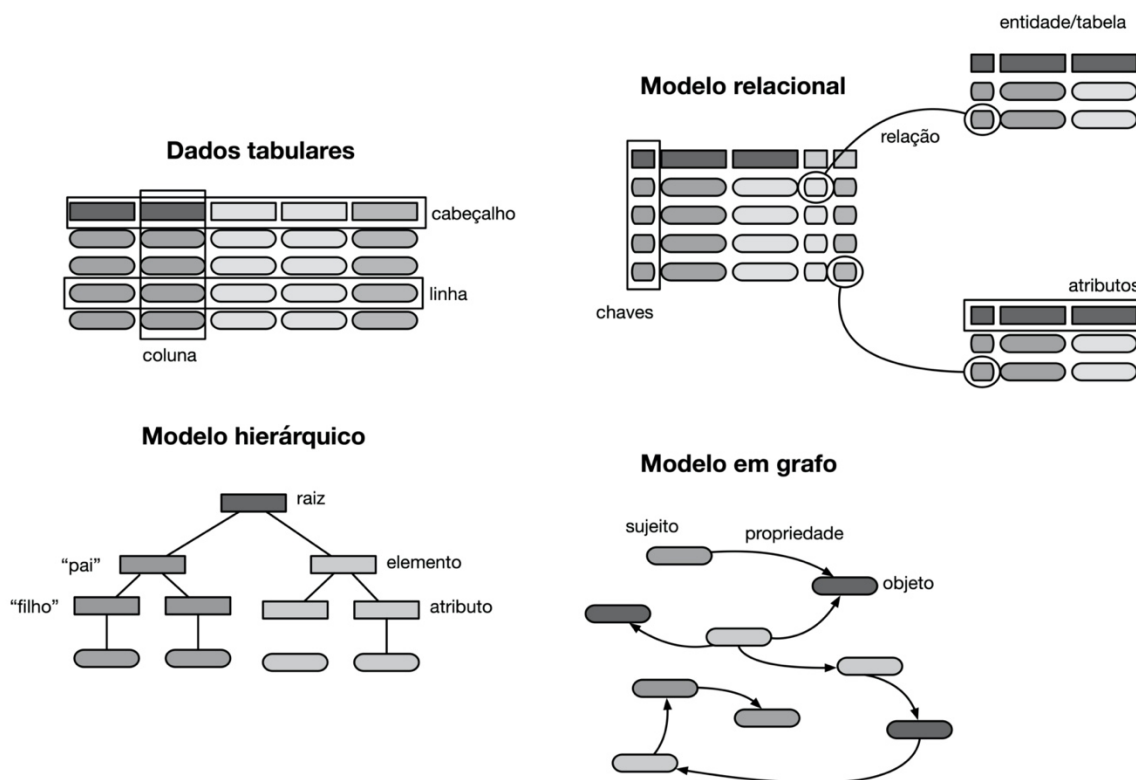


Figura 5 — Comparação esquemática dos três principais tipos de modelos de dados. Baseado em Hooland & Verborgh (2014, p. 13).

As formas arredondadas representam valores de dados individuais, ao passo que formas retangulares indicam mecanismos específicos do modelo para adicionar estrutura. Como se pode constatar, no modelo em grafo são usadas setas, destacando-se a ausência de formas retangulares, o que significa a inexistência de uma hierarquia inerente. Não se pode falar em ausência de estrutura, uma vez que ela existe sob a forma de grafos e setas, mas tão-só ela não é predeterminada numa hierarquia. Este aspeto será crítico, com várias implicações concetuais ao nível da descrição arquivística. Por sua vez, o modelo relacional pode subdividir-se em dois modelos, o modelo tabular e o modelo relacional, podendo ser argumentado que constituem o mesmo modelo de dados, uma vez que os dados são estruturados da mesma forma em colunas e tabelas. Mantém-se a distinção para realçar a dimensão acrescida do estabelecimento de relações entre tabelas na dimensão relacional.

O formato mais comum de seriação do modelo relacional é a SQL (*Structured Query Language*); do modelo hierárquico, destacam-se as linguagens de marcação (*meta-markup*

*languages*), das quais o XML (*eXtensible Markup Language*) é uma das mais representativas; do modelo em grafo, discute-se o RDF (*Resource Description Framework*). Cada modelo representa uma solução com as suas intrínsecas vantagens e desvantagens que, apesar de originarem de contextos tecnológicos distintos, podem ainda hoje encontrar um uso pertinente. Assim, estes modelos *do not supplant one another, but continue to coexist* (Hooland & Verborgh, 2014, p. 14). Cada modelo será brevemente apresentado, não num sentido de sucessão linear, mas num esforço de clarificar as possibilidades e limitações que cada um evidencia face aos demais.

Acredita-se que este exercício é também crucial para o distanciamento do *hype* (entusiasmo) que sucede à adoção de uma nova tecnologia — neste caso, a mais recente é o RDF. Tal entusiasmo é normalmente ilustrado na forma do famoso *hype cycle* de Gartner (Linden & Fenn, 2003, p. 5). Esta representação gráfica da ascensão e declínio da popularidade de uma nova tecnologia alerta para as expectativas exageradas que acompanham a sua introdução. A um *peak of inflated expectations* sucede-se uma desilusão massiva quando se dá a constatação de que, afinal de contas, nem todos os problemas foram solucionados. É só após um período de reflexão que uma tecnologia estabiliza o seu potencial no mercado. A tecnologia de *linked data* constitui, sem dúvida, o *hype* do momento.

Contudo, este entusiasmo desmedido também foi sentido pelo sector do património cultural por volta do ano 2000, aquando da introdução do XML e da *web 2.0*. Apesar da sua ainda atual importância para a criação e gestão de metadados, nenhuma destas tecnologias resolveu os problemas fundamentais subjacentes à interligação de recursos de várias coleções. Analogamente, o RDF é uma forma de representação ainda em estado de desenvolvimento, com desafios e problemas inerentes, embora com um potencial claro. A decisão entre a escolha de cada modelo resume-se à tensão entre as possibilidades de expressão da semântica dos dados e a manipulação tecnológica dos metadados. Por outras palavras, cada tipo de modelo de dados apresenta uma forma própria de representar a informação, com condicionantes inerentes decorrentes da tecnologia que influenciam a própria representação e conseqüente uso.

## **1.4.1 Modelo relacional**

### *1.4.1.1 Dados tabulares*

O modelo tabular é o tipo de modelo mais intuitivo para estruturar dados e tem sido usado em catálogos e índices há séculos, podendo provavelmente ser considerada a forma mais antiga para a representação da informação. Consiste no uso de colunas e linhas. A interseção de uma coluna com uma linha fornece sentido aos dados contidos numa célula em particular. Só existe uma dimensão de modelação, representada pelos campos na primeira linha (*e.g.* título, data). Todas as linhas subsequentes representam diferentes entidades semânticas (*e.g.* documentos de arquivo). Esta disposição facilita uma perceção rápida, quer dos diferentes elementos de metadados usados para descrever os documentos (primeira linha), quer dos diferentes valores exibidos pelos vários documentos (uma coluna). Um exemplo é apresentado na Tabela 6.

Tabela 6 — Exemplo da expressão de metadados no modelo tabular.

<b>Título</b>	<b>Criador</b>	<b>Data</b>	<b>Localização</b>
Mona Lisa	Leonardo Da Vinci	1517	Paris
A última ceia	Da Vinci	1490	Milão
The Monna Lisa (be back in 5 minutes)	Matisse, Sophie	1997	Nova Iorque

FONTE: Elaboração própria.

Além disso, é fácil modificar a base de dados uma vez que podem ser facilmente adicionadas ou retiradas colunas sem que isso cause um impacto significativo na estrutura geral. Contudo, este modelo de dados apenas permite a introdução nas células de tipos de dados em texto, o que pode desencadear uma miríade de inconsistências e dificuldades na pesquisa e recuperação da informação. Por exemplo, na coluna do ‘Criador’ surge tanto ‘Leonardo Da Vinci’ como ‘Da Vinci’ como ‘Matisse, Sophie’. Para além de erros tipográficos que possam surgir pela introdução manual dos dados, a mesma entidade, o nome do criador, pode ser codificada de formas distintas. Uma pesquisa por “Leonardo da Vinci” apenas recuperaria um resultado, ocultando outro. Caso se pretendesse uniformizar o nome do criador, ter-se-ia de atualizar um a um cada campo. Embora este exercício seja fácil quando há apenas três registos, depressa se torna um enorme obstáculo para bases de dados com dezenas de milhar de entradas, ou quando se pretende pesquisar ou uniformizar os dados em mais do que um ficheiro (Hooland & Verborgh, 2014, p. 20).

#### 1.4.1.2 Modelo relacional

O modelo relacional foi desenvolvido nos anos 70 do século 20 num esforço para dar conta das inconsistências dos dados tabulares. Segundo Hooland e Verborgh (2014), foi a abordagem mais bem-sucedida para gerir dados estruturados, e é provável que mantenha esta preponderância ao longo das próximas décadas. O modelo relacional consiste em entidades, relações, atributos e chaves. Uma entidade é *a type of information that varies independently of another* (Ramsay, 2004), capaz de ser identificada de forma única. Os atributos são propriedades de entidades e descrevem-nas. As entidades são ligadas umas às outras através de relações. Cada entidade/tabela contém uma chave (ou outro elemento único) que a identifica, e que pode ser usada por outras para se referir a ela através de uma relação (Figura 6).

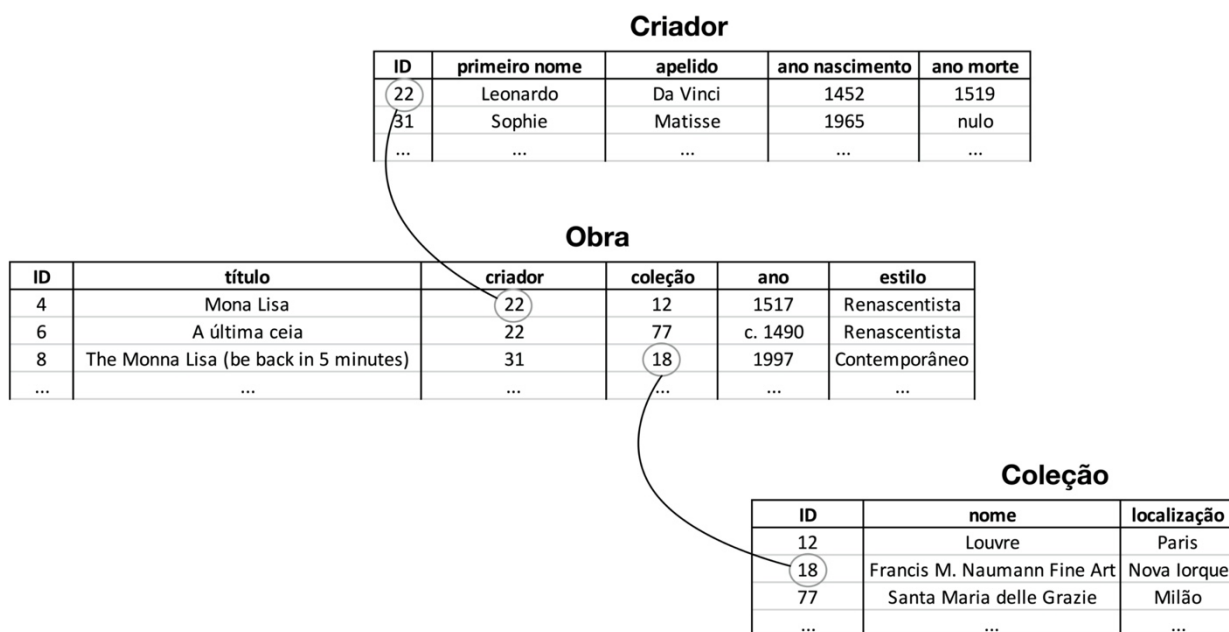


Figura 6 — Exemplo das relações num modelo entidade-relação.

FONTE: elaboração própria.

Este modelo permite o desenvolvimento de descrições complexas dos recursos e das relações entre eles. O que constitui uma entidade ou um atributo depende da importância que se pretende atribuir a determinado aspeto da realidade (Hooland & Verborgh, 2014, p. 24). Porém, qualquer mudança à estrutura implica repensar parte ou todas as tabelas, uma vez que as entidades de uma nova tabela podem sobrepor-se às entidades já existentes e criar conflitos semânticos. Isto implica um enorme obstáculo à mudança, quer em forma de novas adições, quer quanto a uma reestruturação ou recontextualização das relações entre as entidades.

### 1.4.2 Modelo hierárquico

A designação deste modelo advém da sua estrutura hierárquica, a partir do qual um elemento de raiz é sucessivamente desenvolvido em outros ‘filhos’, num esquema em árvore. O conceito de linguagens de marcação começou a ser explorado pelas indústrias farmacêuticas e aeronáuticas a partir da década de 1970 como forma de acrescentar uma estrutura mais flexível a largos volumes de dados não-estruturados. Com o aparecimento da *web* na década de 1990, estas linguagens começaram a ser aplicadas para facilitar a troca de dados partilhados na *web*

acrescentando uma semiestrutura aos dados<sup>34</sup>. O XML é uma das mais significativas implementações, sendo usado para representação de recursos na *web*, como páginas *web*, e de documentos digitais uma vez que os dados estruturados podem ser lidos tanto por humanos como por máquinas.

O XML centra-se na marcação estrutural, ignorando aspetos relacionados com estilo. A este nível, permite que os utilizadores definam os seus elementos de marcação, de onde advém a designação *Extensible*. A possibilidade de cada comunidade escolher os elementos que pretende realçar constitui um ponto forte da sua adoção, exibindo uma plasticidade de acordo com as necessidades de cada contexto de utilização. Além disso, a interpretação de um esquema de XML<sup>35</sup> (*XML schema*), que define o conteúdo, a semântica e a estrutura de um documento (Joudrey & Taylor, 2018, p. 270), encontra-se expressa na sintaxe da XML, resolvendo problemas de incompatibilidade na *web* e fomentando a sua independência face a plataformas ou aplicações (Hooland & Verborgh, 2014, p. 33).

Os documentos XML têm uma estrutura hierárquica que lhes confere uma aparência arbórea. Um elemento de raiz (neste caso, 'Arte') é sucessivamente decomposto em vários ramos que se contêm sucessivamente:

```
<Arte título = Manifestações contemporâneas da arte renascentista>
  <Obra>
    <Título>Mona Lisa</Título>
    <DataCriação>
      <Ano>1517</Ano>
    </DataCriação>
    <Criador>
      <PrimeiroNome>Leonardo</PrimeiroNome>
      <ÚltimoNome>Da Vinci</Último Nome>
    </Criador>
    <Coleção>
      <Nome>Louvre</Nome>
      <Localização>Paris</Localização>
    </Coleção>
  </Obra>
</Arte>
```

---

<sup>34</sup> Dizem-se semiestruturados porque, embora o XML tenha uma estrutura hierárquica própria, as aplicações em muitos casos podem ignorar a árvore em si e apenas listar os dados, ou considerar que além da estrutura hierárquica o conteúdo de cada elemento é geralmente texto livre (sem estrutura adicional).

<sup>35</sup> De forma mais simples, um esquema XML indica que elementos podem ocorrer, onde, em que quantidade na árvore e qual o tipo de dados contidos pelos elementos (se texto ou números, entre outros).

Este modelo é composto por três unidades de representação: entidades, elementos e atributos. Uma das questões nesta forma de representação é a opção entre modelar uma entidade enquanto um elemento ou atributo. Os elementos são serializados como *tags* entre chavetas, sendo os atributos modificadores de uma *tag* ao refinar o significado do elemento sob o qual se encontram. No exemplo, ‘Obra’ é um elemento, ao passo que ‘Título’ é um atributo do elemento ‘Obra’. Ambos são entidades que representam partes da coleção “Manifestações contemporâneas da arte renascentista”.

A bem-definida estrutura de dados permite uma interpretação eficiente dos dados por máquinas ao mesmo tempo que providencia uma estrutura hierárquica de interpretação relativamente flexível. O modelo relacional apresenta uma estrutura forte e facilmente interpretável por máquinas, mas com limitações de uso dada a sua rigidez ao nível da estrutura que limita as aplicações. As bibliotecas, arquivos e museus entusiasmaram-se no início da primeira década deste século com as possibilidades de partilha de metadados em XML através do *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH), partilhando vários materiais como coleções fotográficas e metadados de investigação (NISO, 2017, p. 9). Contudo, a verbosidade da descrição é frequentemente um impeditivo para a legibilidade humana<sup>36</sup>. A expressão de um simples conceito ocupa várias linhas, o que se reflete em documentos muito longos e de interpretação por vezes confusa. Neste exemplo, apenas para expressar a data de criação de uma só obra são necessárias três linhas. Mesmo no caso da interpretação de agentes computacionais, a interpretação do esquema constitui uma barreira para o uso automático dos dados em sistemas de informação na *web* uma vez que, embora a semântica dos dados seja explicada pelo uso dos elementos, diferentes comunidades podem usar os mesmos elementos com diferentes significados. A interpretação dos elementos por terceiros apresenta dificuldades de interoperabilidade, tornando-se necessária a adoção do mesmo esquema de modo a compreender a semântica (Hooland & Verborgh, 2014, p. 43). É neste contexto que, enquanto tecnologia sucessora, surge a principal vantagem do RDF.

### 1.4.3 Modelo em grafo

O percurso feito até ao momento pretende contextualizar o aparecimento das tecnologias de *linked data* em relação ao potencial que constituem para um novo paradigma na representação da informação em arquivos, sem necessidade de uma estrutura demarcada, pré-determinada pelo arquivista. Apesar das potencialidades das bases de dados relacionais e do XML para a gestão de metadados, estes modelos apresentam uma enorme desvantagem na necessidade de compreensão do esquema que descreve a estrutura e a articulação entre os dados. Conforme se analisará de seguida, o RDF tenta contornar esta questão. Porém, é útil primeiro contextualizar o seu papel no âmbito da *Web Semântica*.

---

<sup>36</sup> Linguagens como o JavaScript Object Notation (JSON) têm sido desenvolvidas num esforço de simplificação da verbosidade do XML.

### 1.4.3.1 A Web Semântica

A necessidade de superar estruturas semânticas definidas localmente está relacionada com a visão da *Web Semântica*, que importa introduzir antes de discutir o modelo RDF. O termo *Web Semântica* foi introduzido por Tim Berners-Lee para designar aquilo em que a *World Wide Web will evolve into tomorrow* (Berners-Lee et al., 2001, p. 1). Publicado em 2001, estamos num futuro relativamente adiantado em relação ao momento em que estas palavras foram escritas. Não podemos dizer que convivemos com uma *Web Semântica* plenamente desenvolvida mas, em todo o caso, estamos sem dúvida nela imersos até certo ponto. Por exemplo, se procurarmos “Copenhaga” no Google, dados tais como descrições da cidade, população, hotéis e pontos de interesse nas proximidades constituem uma manifestação do potencial da *Web Semântica*. Ao nível dos recursos representados por instituições detentoras de património cultural, uma pesquisa por palavras-chave por termos como ‘Leonardo Da Vinci’ é trivial. Já uma pesquisa de ‘obras de Leonardo Da Vinci’ é mais problemática, uma vez que a expressão de criação pode variar de repositório para repositório (autor/criador/escritor/pintor), bem como pode variar a forma do nome. Já uma pesquisa como ‘obras influenciadas por Leonardo Da Vinci’ é demarcadamente impossível em sistemas de representação tradicionais.

Esta espécie de exercício não era praticável em 2001 porque as páginas *web* estavam sobretudo desenhadas para que humanos as interpretassem, não para que os computadores as explorassem, pelo que tais inferências não eram automaticamente obtidas<sup>37</sup>. A visão da *Web Semântica* é uma segundo a qual a informação é fornecida num formato legível por máquinas de modo a permitir que agentes computacionais possam autonomamente desempenhar por nós algumas tarefas de descoberta (Berners-Lee et al., 2001, p. 1), tais como automaticamente recuperar pontos de interesse próximos de Copenhaga, o nosso termo de pesquisa inicial. A *Web Semântica* não representa, assim, uma *Web* distinta ou disruptiva, mas antes uma camada de semântica por cima da infraestrutura da atual *Web* (Hooland & Verborgh, 2014, p. 126), na qual a informação é dado um significado preciso e definido que permita a interpretação automática por computadores.

De modo a interligar recursos desta forma, há quatro regras que têm de ser cumpridas:

- *Use URIs as names for things*
- *Use HTTP URIs so that people can look up those names*
- *When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)*

---

<sup>37</sup> Mais especificamente, cada página *web* era desenvolvida independentemente das outras, sem um espírito de interoperabilidade dos elementos que facilitassem uma legibilidade conjunta. É neste sentido que se afirma que o seu desenvolvimento era orientado para o entendimento humano, de satisfação imediata de necessidades de visualização dos utilizadores sem uma preocupação com a interoperabilidade. A *Web Semântica* distingue-se pelo facto de atributos adicionais serem introduzidos nas páginas de forma a facilitar a construção de um grafo de relações.



- *Include links to other URIs, so that they can discover more things* (Berners-Lee, 2006).

Há várias implicações para cada regra mas, por agora, debatamos a quarta, uma vez que é aquela mais diretamente associada com o recente *hot topic* de *linked data*. Embora próximos, *Linked Data* e *Web Semântica* não são sinónimos. A *Web Semântica* depende de significado estruturado na forma de metadados legíveis por máquinas que permitam a descoberta automática por agentes computacionais. *Linked Data* refere-se à ligação harmoniosa entre recursos disponíveis na *Web*, através de informação semanticamente bem-definida em documentos acessíveis na *Web*. Por conseguinte, podemos afirmar que a *Web Semântica* é *a more general term to describe the architecture that allows Linked Data to be represented, connected, shared, searched, and combined* (Gracy, 2015, p. 240).

Empregamos deliberadamente o termo ‘recurso’, e não ‘documento’, uma vez que qualquer coisa pode ser relacionada usando um triplo. Um triplo (*triple*) consiste na explicitação de uma relação entre uma entidade e um objeto, sendo que não só essa relação pode ser de qualquer natureza (autoria, parentesco, localização), como também qualquer entidade pode ser representada (documentos, pessoas, monumentos). Estamos perante o princípio da universalidade, que constitui a propriedade essencial da *World Wide Web* (Berners-Lee et al., 2001, p. 1). O RDF permite a descrição de um recurso, e qualquer coisa pode ser um recurso. Na esfera arquivística, tradicionalmente desejamos que esse recurso seja um documento de arquivo, mas não tem necessariamente de o ser. Por exemplo, além da representação usual de que Eça de Queirós escreveu ‘Os Maias’, pode ao mesmo tempo ser igualmente pertinente representar que casou com Emília de Castro Resende, ou que um documento foi criado por um banco, e que esse banco foi adquirido por outra unidade orgânica, e que há um processo judicial movido contra essa unidade orgânica. No último exemplo, a informação sobre a aquisição é tradicionalmente documentada na história custodial. Contudo, esta informação é apresentada em linguagem natural, em texto, sem significado semântico estruturado, *machine-readable*. Com o RDF, torna-se possível explicitar a aquisição e relacionar esta informação a outros recursos de forma automática e de uma forma mais simples do que noutros modelos. Assim, não é que o RDF permita necessariamente mais possibilidades de representação do que os modelos tabular, relacional e de linguagens de marcação, uma vez que em teoria seria possível expressar as mesmas relações semânticas em qualquer modelo, dependendo da quantidade e qualidade da estrutura. Todavia, o RDF permite expressar relações semânticas complexas mais facilmente do que nos demais modelos.

Assim, a identificação dos recursos (incluindo documentos de arquivo) por um URI (primeira regra) torna-se uma necessidade de modo a permitir a descoberta na *Web* e acrescentar valor à informação contida no património arquivístico. É o uso inesperado da informação através de diferentes sistemas de informação que faz da *Web* um local de valor acrescentado (Berners-Lee, 2006), e que marca a diferença face às tradicionais descrições arquivísticas de narrativa linear. Além disso, a eficácia e o poder desta forma de explorar os dados apenas tende a aumentar à medida

que mais conteúdo legível por máquinas é disponibilizado na *Web*, incrementando assim o verdadeiro potencial sinérgico da *Web Semântica* (Berners-Lee et al., 2001).

Neste contexto, as descrições arquivísticas, entendidas como mais do que a mera descrição de instrumentos de pesquisa, podem ser entendidas como *potential pathways* para relacionar dados arquivísticos a outras fontes de dados na *Web* (Gracy, 2015, p. 246). No entanto, a descrição arquivística pode ser mais do que potenciais vias. A propriedade essencial da arquitetura REST, (*Representational State Transfer*), um formato normalizado de representação de serviços (pedido/resposta) que *has been used to guide the design and development of the architecture for the modern Web* (Fielding, 2000, p. xvii), é a de que os recursos não são transferidos mas, antes, dá-se uma partilha de representações desses recursos (Hooland & Verborgh, 2014, p. 214) ou, por outras palavras, a partilha de metadados. Como vemos, a descrição arquivística é efetivamente a via de partilha e difusão na *Web* de recursos contidos nos arquivos. Num cenário em que a representação da informação é tão central, atrevemo-nos a dizer que é mais pertinente do que nunca estudar a forma como a autenticidade é manifestada nas descrições arquivísticas, um exercício realizado mais adiante no segundo capítulo.

#### 1.4.3.2 RDF

A modelação de dados em grafo é uma tecnologia que pretende colmatar os problemas associados com o uso de esquemas localmente definidos, como é o caso das bases de dados relacionais e do XML. Este exercício é essencial como requisito da ambição da *Web Semântica* que perspetiva que a *web* também seja acessível por agentes de *software*, em vez de conter unicamente informação legível por humanos (Berners-Lee et al., 2001).

O RDF é uma das mais significativas implementações do modelo em grafo. A adoção no RDF de um modelo de dados extremamente simples, consistindo de triplos, torna-o *schema-neutral* (Hooland & Verborgh, 2014, p. 44), permitindo que os dados sejam interpretados por qualquer agente (humano ou máquina) a partir de qualquer local com uma ligação à Internet. Por neutralidade de esquema entende-se a capacidade de agentes computacionais interpretarem a semântica dos dados sem que esta esteja explicitamente formulada. Não significa que o modelo de dados em grafo não possa ter um esquema — pode tê-lo, e é desejável que o tenha quando há a necessidade de definir uma estrutura semântica. Os triplos de RDF são asserções curtas e simples que expressam a relação entre dois recursos. Qualquer coisa pode ser um recurso, incluindo documentos, pessoas, objetos ou até conceitos. Os triplos consistem num sujeito, num predicado e num objeto (Figura 7). De um sujeito (Leonardo Da Vinci) pode ser dito que tem uma relação direcional, expressa na forma de um predicado (criou), com um objeto (Mona Lisa). Esta relação é uma propriedade (Criou/FoiCriadoPor). A simplicidade deste modelo de dados, aliada ao facto da sua neutralidade relativamente a um esquema, constituem dois dos pontos fortes do RDF.

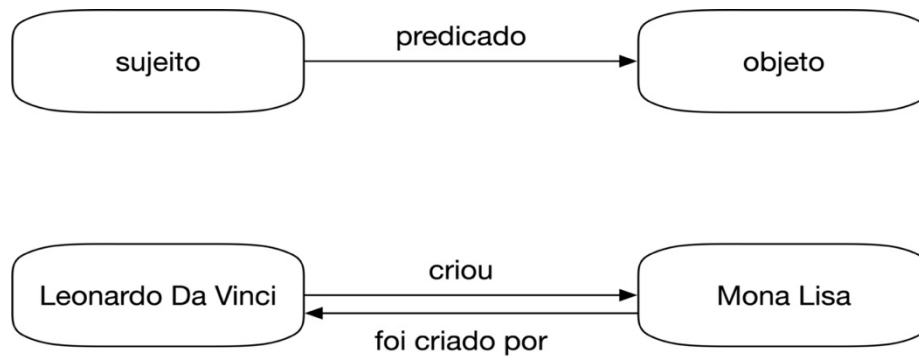


Figura 7 — Estrutura de um triplo.

FONTE: elaboração própria.

Uma porção significativa da força do RDF advém da possibilidade de cada recurso poder ocorrer em vários triplos, quer como sujeito quer como objeto, dando aso à criação de poderosas ligações entre triplos (W3C, 2014b). Além disso, os dados expressos desta forma são atemporais, na medida em que os triplos representam *snapshots* estáticos de informação (W3C, 2014a) e, como tal, adequam-se à preservação no longo termo. As mudanças à estrutura são facilmente introduzíveis, e até incentivadas, na medida em que novos triplos não só não alteram a estrutura subjacente, como também aumentam o contexto e a descoberta dos recursos já existentes. Para tal, basta criar um novo triplo, de forma aditiva, sem implicações para os triplos anteriores. A Figura 8 ilustra como vários triplos podem ser articulados para expressar metadados sobre várias entidades.

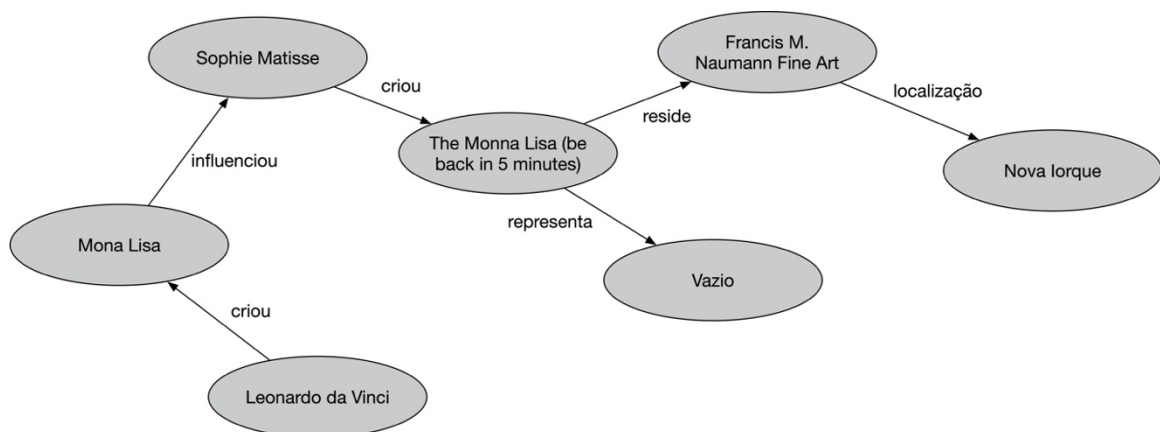


Figura 8 — Exemplificação do encadeamento de vários triplos na expressão de metadados.

FONTE: elaboração própria.

O sujeito de um triplo pode ser o objeto de outro, uma vez que não existe uma estrutura definida. Os triplos são contentores para expressar metadados, mas o grau de destaque dado a cada

entidade representada é definido pelas pesquisas dos utilizadores. Assim, não só tipo este modelo de dados permite uma representação extremamente flexível, duradoura e interoperável dos recursos na *web*, através de metadados facilmente partilháveis e interpretáveis, como também permite a construção de narrativas pessoais ajustadas à perspetiva de cada utilizador e à representação simultânea dos vários contextos de interpretação na qual determinado recurso se insere. Neste exemplo, a Mona Lisa é representada quer como uma obra renascentista do século 16, quer como uma obra que inspira outras produções artísticas no século 21 que nos fazem repensar e reapreciar a obra tal como ela fora produzida cinco séculos antes.

As maiores dificuldades deste modelo prendem-se talvez com um mercado ainda imaturo, que pode não disponibilizar ainda ferramentas eficientes para relacionar os dados e promover a *linked data*, uma situação que deve mudar ao longo dos próximos anos à medida que mais projetos, como o Europeana, crescem em número e influência, ao mesmo tempo que se aguarda o desenvolvimento de mais ferramentas *open source*.

Em resumo, a escolha por um tipo de modelo de dados em particular implica um *trade-off* inevitável entre a complexidade das relações que consegue incorporar e a facilidade com que o resto do mundo pode reutilizar e conectar-se aos dados (Hooland & Verborgh, 2014, p. 50). Assim, não se trata de fazer uma apologia cega do modelo em grafo, o que seria cair na falácia da lei do instrumento, formulada por Abraham Kaplan da seguinte forma: *give a small boy a hammer, and he will find that everything he encounters needs pounding* (Kaplan, 1964, p. 28). Cada um dos modelos de dados tem um lugar a desempenhar na expressão de metadados, dependendo do objetivo e do contexto social e institucional. Por exemplo, na *web* semântica as páginas podem ser codificadas em HTML/XML, com o acréscimo de elementos que expressem as ligações com outras páginas de modo a possibilitar a construção de grafos. No caso das representações em arquivos, é plausível adotar tipos de modelos de dados mistos, como por exemplo dados modelados em XML como forma de estrutura e certos dados em RDF como forma de expressar as relações. Conforme salientado inicialmente, nenhum modelo supera o outro em termos absolutos. Na prática, continuarão a coexistir na medida em que servem propósitos distintos.

Contudo, acredita-se ter demonstrado o potencial do tipo de modelo em grafo para a partilha e descoberta de recursos na *web*. De seguida analisa-se a representação arquivística na pós-modernidade de modo a compreender de que forma a ausência de uma hierarquia pré-estabelecida se articula com a mudança dos princípios da descrição arquivística na era pós-moderna.

## 1.5 Terminologia de normas de metadados

A transversalidade do uso de metadados por praticamente todas as comunidades científicas torna a terminologia diversa, redundante e, por vezes, confusa. Jonathan Furner (2020), só ao nível das normas publicadas pela ISO, identifica 96 normas que definem o termo ‘metadados’, desde áreas como ‘Informação e Documentação’ a ‘Tecnologias de Informação’, passando por ‘Informação Geográfica’ ou ‘sistemas de transporte inteligentes’. Como reflexo deste uso multidisciplinar, não se observa uma uniformidade ou consistência nem relativamente aos termos

usados para denominar os documentos que recomendam os metadados a usar e a forma como estes devem ser usados — que designamos por normas (*standards*) —, nem acerca do conjunto de elementos de metadados usados em tais documentos.

Um bom ponto de partida para a classificação das normas de metadados parece ser por começar por decompor nas suas componentes um registo de metadados (*metadata record*), inspirada na proposta de Lubas *et al.* (2013). Entende-se um registo de metadados como um conjunto de metadados que descreve um recurso, composto por sintaxe, estrutura e semântica:

- A sintaxe é um contendor para a estrutura e para a semântica, atuando como um conjunto de regras segundo as quais os conteúdos devem ser interpretados por um computador (p.e., RDF, XML).
- A estrutura consiste no modelo de dados adotado (p.e., Dublin Core) que governa a escolha dos elementos de metadados apropriados para a representação dos recursos (p.e.: criador, título, assunto).
- A semântica corresponde ao conteúdo e define a forma como a informação nos elementos de metadados deve ser formatada (p.e., 01-03-1990 lê-se 1 de março de 1990 e não 3 de janeiro de 1990).

A categorização das normas de metadados pode ser vista como partindo destas componentes, no sentido em que cada norma visa definir um conjunto de procedimentos para a expressão dos metadados de cada dimensão de um *metadata record*. Importa, antes de mais, esclarecer o uso que adota do termo ‘norma’. No sentido geral, uma norma ou *standard* é *a technical document designed to be used as a rule, guideline or definition. It is a consensus-built, repeatable way of doing something* (European Committee for Standardization, 2020). Existe nesta aceção uma dimensão de obrigatoriedade, vínculo ou de cumprimento de padrão de qualidade de uma indústria cuja implementação não é negociável, sob risco de incorrer sob ilegalidade. Por sua vez, as normas relacionadas com metadados entendem-se como *formal documents that establish uniform criteria, methods, processes, and practices* (Zeng & Qin, 2016, p. 23). Este entendimento de norma é mais genérico, não exclusivo a documentos produzidos por instituições normalizadoras como o ICA, a ISO, a NISO ou entidades governamentais através de legislação. Emprega-se este termo sem uma conotação legal de obrigatoriedade, mas antes como um conjunto de procedimentos entre muitos, que se podem seguir ou não, dependendo das finalidades da representação e das necessidades da comunidade. Por exemplo, um arquivo pode escolher representar a sua informação através da norma Dublin Core ou da norma ISAD(G), recaindo a escolha sobre constrangimentos sociais, comunitários, técnicos e financeiros. Assim, uma norma de metadados é um documento formal que define critérios, métodos, procedimentos e práticas uniformes para a escolha, expressão e/ou partilha de metadados, sem um carácter vinculativo. Nesta aceção, norma de metadados (*metadata standard*) assume-se como um termo genérico que se ramifica em demais designações como modelo ou esquema, consoante será distinguido mais adiante.

Posto isto, com base nas leituras identificam-se quatro tipos mais recorrentes de normas na literatura (Elings & Waibel, 2007; Gilliland, 2016; Research Libraries Group, 2005; Riley, 2010;

Zeng & Qin, 2016): normas para estrutura de dados (*data structure standards*), para conteúdos de dados (*data content standards*), para partilha de dados (*data exchange standards*) e para definição dos valores dos dados (*data value standards*)<sup>38</sup>, conforme expresso pela Figura 9. Steven Miller (2011) menciona ainda normas de preservação (*preservation standards*).

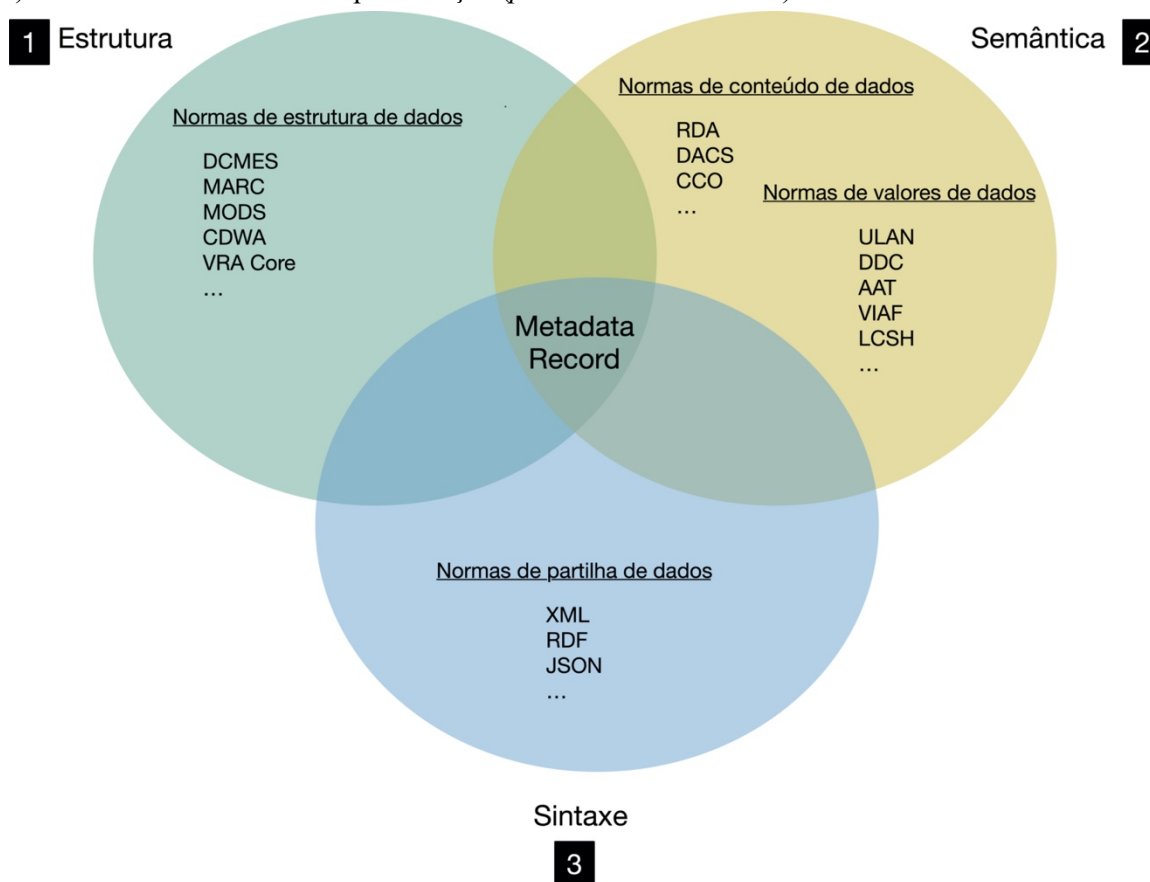


Figura 9 — Tipologia de normas de metadados.

FONTE: elaboração própria.

De forma esquemática, as *data structure standards* abordam a estrutura do *metadata record*; as *data content* e *data value standards* clarificam a semântica; e as *data exchange standards* referem-se à sintaxe.

As normas para estrutura de dados (por vezes designadas de *metadata element sets*, *element sets*, *schemas* ou *metadata vocabularies*) são “*categories*” or “*containers*” of data that make up a record or other information object (Gilliland, 2016, p. 3). Tais categorias correspondem aos *metadata elements*, por vezes designados também de *data elements* ou *data fields*, e figuram como unidades de informação que uma norma indica que devem ser capturados, cuja ordem e organização

<sup>38</sup> Estes tipos referem-se à classificação das normas no contexto da literatura sobre metadados. Outros autores noutros contextos empregam diferentes formas de classificação (Furner, 2020; Timmermans & Epstein, 2010).

hierárquica é definida pela estrutura dos dados (Elings & Waibel, 2007; Research Libraries Group, 2005, p. 3). Há várias normas de estrutura dependendo das necessidades da comunidade. A norma de uso genérico mais conhecida e adotada é a *Dublin Core Metadata Element Set (DCMES)*. As normas da família MARC (MARC, MODS) são principalmente usadas no domínio das bibliotecas; nas descrições arquivísticas, são comuns a EAD e a ISAD(G); as normas CDWA e VRA Core encontram adoção nos museus e nas comunidades com recursos visuais.

As normas para o conteúdo dos dados *are created to guide the practice of metadata generation and cataloging* (Zeng & Qin, 2016, p. 24), fornecendo diretrizes e recomendações sobre como introduzir e formatar a informação em cada campo. Portanto, *data content standards specify the rules that mold data into predictable expressions* (Research Libraries Group, 2005, p. 4). Um exemplo é a já mencionada formatação da data ao nível dos dias, meses e ano. Entre as normas para o conteúdo destacam-se no domínio das bibliotecas as *Anglo-American Cataloguing Rules (AACR2)*, entretanto superadas pela *Resource Description and Access (RDA)*; nos arquivos, a *Describing Archives: A Content Standard (DACS)*; para arte e museus, a *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images (CCO)*.






As normas para os valores dos dados constituem sistemas de organização do conhecimento, sendo por esse motivo por vezes designadas de *value vocabularies* ou *value encoding schemes*. As *data value standards are the thesauri or name authorities providing preestablished terms to populate a data field* (Research Libraries Group, 2005, p. 3), fornecendo um conjunto de termos pré-estabelecidos e as relações semânticas entre eles. De entre os exemplos incluem-se a) listas de termos controlados (*MARC Code List for Relators* e *Union List of Artist Names [ULAN]*); b) esquemas de classificação (*Dewey Decimal Classification [DDC]* e *Universal Decimal Classification [UDC]*); c) thesauri (*Art and Architecture Thesaurus [AAT]*, *Thesaurus of Geographic Names [TGN]*); d) ficheiros de autoridade, ou *authority files (LC Name Authority File [LCNAF]* e *Virtual International Authority File [VIAF]*); e e) listas de cabeçalhos de assunto, ou *lists of subject headings (LC Subject Headings) [LCSH]*.

As normas para a partilha de dados, ou *standards for data exchange*, são por vezes referidas como *format standards* na medida em que se referem aos diferentes formatos dos dados no contexto da partilha de dados e da comunicação, tal como em Gilliland (2016, p. 3), Riley (2010), ou Elings e Waibel (2007). O formato diz respeito à forma como a informação é codificada (*encoded*) num ficheiro, ao passo que a partilha (*exchange*) diz respeito ao protocolo usado para a partilha dos metadados. Quanto ao formato de partilha, o XML é uma das linguagens mais comuns para a expressão dos elementos de metadados, mas desde os finais da década de 2000 outros formatos como o RDF têm adquirido destaque. Como exemplo de protocolo de partilha (*exchange protocol*), destaca-se o *Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)* no contexto dos arquivos.

O *Research Libraries Group* (2005) propõe uma interessante metáfora para o entendimento do papel de cada uma destas categorias. As normas para a estrutura dos dados são equiparadas a uma garrafa de vinho; o conteúdo e o valor refere-se ao líquido que é colocado na garrafa; o formato

é o tipo de caixa que se usa para embalar a garrafa. Elings e Waibel (2007) continuam a metáfora acrescentando que a partilha de dados corresponde ao indivíduo que transporta a caixa e a entrega na morada. Esta metáfora é útil para compreender em termos palpáveis a função de cada categoria das normas. Um resumo é apresentado na Tabela 7 de modo a recapitular as definições e as alusões apresentadas.

Tabela 7 — Caracterização dos tipos de normas.

Tipo	Definição	Metáfora
Normas de <b>estrutura</b> de dados	Indicam a escolha dos elementos de metadados que representam um recurso.	
Normas de <b>conteúdo</b> de dados	Especificam as regras de formatação da informação em cada campo de modo a obter expressões previsíveis e uniformes.	
Normas de <b>valores</b> de dados	Indicam o conjunto de termos controlados previamente estabelecidos para preencher um campo.	
Normas de <b>partilha</b> de dados	Especificam o modo de partilha e comunicação da informação. Dizem respeito ao formato e ao protocolo de comunicação.	
	O formato consiste no tipo de ficheiro escolhido para codificar a informação.	
	O protocolo diz respeito ao método de partilha da informação.	

FONTE: Elaboração própria.

Estas categorias podem parecer rígidas, mas o grau de pertença de uma norma a cada uma não o é. Acontece com frequência que uma norma aborda vários destes aspetos dos metadados, pertencendo a mais do que uma categoria. Por exemplo, Riley (2010) inclui no seu esquema de classificação a ISAD(G) e a RDA tanto como normas para conteúdo e estrutura. A norma EAD tem primariamente uma função de conteúdo, mas exhibe também elementos estruturais. Devido aos



objetivos propostos no âmbito deste trabalho, serão alvo de análise as normas para a estrutura dos dados, uma vez que é a categoria que se enquadra na escolha dos elementos de metadados usados para representar os recursos, independentemente das formas escolhidas para expressar a sua sintaxe e semântica.

Definidos os tipos de normas, urge esclarecer de seguida a terminologia a respeito da designação conceptual das normas, nomeadamente ao nível da distinção entre *metadata standard*, *metadata schema* e *conceptual model*, termos por vezes coocorrentes na literatura, embora com sentidos distintos. Uma *metadata standard*, conforme referido anteriormente, consiste de forma abrangente num documento formal que estabelece critérios, métodos, processos e práticas uniformes. Este termo é por vezes empregue de forma permutável com *metadata schema* como alusão *to an entire element set as well as the encoding of the elements and structure using a standard language* (Zeng & Qin, 2016, p. 27).

Um *metadata schema* refere-se ao formato e à estrutura dos metadados, geralmente descritos num *set of rules for encoding information that supports specific communities of users* (ECLAP, 2011, p. 9). Neste sentido, um esquema assume-se como *a means for defining the structure, content, and, to some extent, the semantics of data* (The Consultative Committee for Space Data Systems, 2013, p. 1-4). Como tal, um esquema de metadados assume geralmente uma dimensão semântica dos dados.

Os referidos elementos de metadados, ou *metadata elements* são itens de metadados, geralmente acompanhados por uma definição que consiste numa *human readable phrase or sentence associated with a data element within a data dictionary that describes the meaning or semantics of a data element* (ECLAP, 2011, p. 9). Estas definições são essenciais em qualquer norma de estrutura ou proposta de elementos de forma a clarificar o sentido, a interpretação e o uso dos elementos de metadados, pelo que serão incluídas no quinto capítulo, onde se apresenta a proposta de um modelo de requisitos de metadados.

Na literatura sobre metadados, modelo conceptual (*conceptual model* ou *conceptual schema* ou *high-level data model* ou *conceptual data model* ou *data model* ou simplesmente *model*) é um mapa dos conceitos e das suas relações. Um modelo concetual descreve de forma abstrata a forma como os dados são representados num determinado sistema (ECLAP, 2011). Trata-se do maior nível de abstração num sistema de representação na medida em que representa a própria descrição arquivística no modo como modela e conceptualiza os recursos (The National Archives of Finland, 2013, p. 8). Por exemplo, se quiser representar um comboio nas suas componentes, alguns conceitos a incluir seriam o de carruagem, assentos, corredor, bar, casas de banho. Carruagens de 1ª classe podem estar relacionadas com bar e casas-de-banho, ao passo que carruagens de 2ª classe podem estar relacionadas apenas com casas-de-banho, e não bar.

Esta aceção aproxima-se do conceito de ontologia, entendida como *a data model that represents the existing knowledge within a domain and is used to reason about the objects in that domain and the relationships between them* (ECLAP 2011, p. 9). As ontologias são usadas como

forma de representação do conhecimento sobre o mundo ou parte dele, descrevendo geralmente sujeitos, atributos e relações, a nível igualmente abstrato.

Contudo, o termo ‘modelo’ assume outra interpretação no seio da Filosofia. A Tabela 8 recapitula de forma esquemática esta terminologia. Segundo Peter Achinstein, os modelos são *a set of assumptions or postulates describing certain physical objects or phenomena* (Achinstein, 1964, p. 330) entendimento na base do qual distingue modelos de teorias:

*To propose something as a model of (an) x is to suggest it as a way of representing x which provides at least some approximation of the actual situation; moreover, it is to admit the possibility of alternative representations useful for different purposes. To propose something as a theory of (an) x, on the other hand, is to suggest that x's are governed by such and such principles, not just that it is useful for certain purposes to represent x's as governed by these principles or that such principles approximate those which actually obtain* (Achinstein, 1965, p. 105).

A proposta de Achinstein sugere que um modelo se distingue de uma teoria no sentido em que o primeiro constitui uma de muitas possíveis representações de qualquer coisa, ao passo que a teoria determina o único conjunto de parâmetros através dos quais essa coisa pode ser representada. A distinção radica-se assim na humildade de um modelo na sua parcialidade, por contraste com a autoridade certa de uma teoria. Contudo, esta posição não é unânime. Wartofsky (1979, p. 2) discorda da distinção entre modelo e teoria por entender que estas entidades partilham uma relação universal de similaridade que as aproxima mais do que as distingue. Em contrapartida, propõe uma tipologia de modelos cujas categorias se distinguem de acordo com o grau de crença e de compromisso existencial colocado em cada modelo enquanto representação (Wartofsky, 1979, pp. 29-33). Neste sentido, este autor postula um modelo enquanto *an abstractive representation of some object or state of affairs* (Wartofsky, 1979, p. 4), uma posição que se adota. Um modelo não requer uma representação completa e fidedigna do que representa, mas tão-só a similitude de certas propriedades que atribuem pertinência ao modelo. Um modelo da forma dos planetas no sistema solar constitui uma representação pertinente quer sejam usadas esferas de metal assentes numa estrutura mecânica que replica as órbitas, quer se usem bolas de ténis dispostas no chão ou se refira por escrito a ocorrência de corpos celestes esféricos. É evidente que tais representações não replicam fidedignamente os planetas do sistema solar mas, para as intenções específicas de um uso, constituem um modelo plausível. Willard McCarty prefere falar em modelação (*modelling*) pois defende que os modelos são um processo iterativo de melhoramento do entendimento da coisa representada, *an attempt to capture the dynamic, experiential aspects of a phenomenon rather than to freeze it into a theoretical abstraction* (McCarty, 2005, p. 293).

Assim, *the limits on modeling are therefore the limits of our conceptions of what respects are relevant to what purposes, and not intrinsic elements which we can foreknow or prescribe for any two entities* (Wartofsky, 1979, p. 6). Os modelos exibem, portanto, uma relação triádica, segundo a qual qualquer coisa pode ser um modelo de qualquer coisa — relação diádica — desde que encontradas propriedades relevantes por um indivíduo — o terceiro elemento na relação que a torna

triádica. Contudo, a asserção da relevância enquadra-se no domínio do discurso humano e das suas intenções inerentes que escolhe que propriedades entre duas coisas são relevantes de forma a que uma possa ser modelo de outra. A adoção e a criação de um modelo está assim inevitavelmente constituída de erro e parcialidade inerente a qualquer atividade humana, pelo que mais do que verdades sobre algo, constituem interpretações que vão sendo construídas e melhoradas sucessivamente segundo o processo de modelação.

Tabela 8 — Distinção entre norma de metadados, elemento de metadados, esquema de metadados, modelo conceptual e ontologia.

<b>Conceito</b>	<b>Definição</b>
Norma de metadados	Um documento formal que estabelece critérios, métodos, processos e práticas uniformes.
Esquema de metadados	O conjunto dos elementos de metadados mais a forma de codificação desses elementos e da estrutura usando determinada linguagem.
Modelo	a) Um mapa dos conceitos e das suas relações. b) Uma representação abstrata de um objeto ou de um estado para fins analíticos.
Ontologia	Um modelo de dados que representa o conhecimento de um domínio, incluindo as relações estabelecidas entre eles.
Elementos de metadados	Itens de metadados, correspondentes a campos de informação, geralmente acompanhados de uma definição que guie o seu uso.

FONTE: Elaboração própria.

Por este motivo, considera-se que o compromisso assumido nesta tese de apresentar uma ‘lista de metadados’ não pode ser designado por esquema de metadados, uma vez que desconsidera o formato usado para os expressar; não é uma norma de metadados, uma vez que não tem pretensões de constituir uma lista exaustiva e completa dos metadados necessários para representar a totalidade de um recurso num determinado contexto e para uma comunidade específica; e muito menos é um modelo ou uma ontologia na aceção da literatura sobre metadados porque não só não lida com os conceitos em abstrato, uma vez que propõe elementos de metadados concretos, como também não inclui as relações entre conceitos ou termos.

Neste sentido, entende-se que o mais adequado é designar a proposta enquanto um modelo de metadados que se foca na autenticidade como propriedade relevante do objeto que se pretende representar. Enquanto modelo, na aceção filosófica, não tem presunções de representar a totalidade do complexo universo das propriedades que constituem os metadados, mas apenas neste propósito específico dos usos de metadados para a asserção da autenticidade, da confiabilidade e da fidedignidade dos recursos, de forma a permitir o seu estudo. Na prática, o modelo materializa-se na proposta de um conjunto de elementos de metadados (*metadata element set*) — simplificando, um conjunto de metadados — que realça a dimensão parcial deste conjunto, que deve ser entendido como um complemento a outros elementos recomendados por outras normas. Como não representam a totalidade das necessidades de uma determinada comunidade, não constituem uma norma autónoma, afirmando-se antes como um conjunto de elementos passíveis de serem integrados noutras normas, complementando-as e enriquecendo-as a respeito da asserção da autenticidade, da fidedignidade e da confiabilidade dos recursos representados. Pode eventualmente contribuir também para a interoperabilidade, caso uma ou várias comunidades decidam seguir adotar em conjunto os metadados propostos.

## 1.6 A importância da interoperabilidade

A interoperabilidade constitui um dos cinco princípios orientadores da conceção dos metadados, segundo Marcia Zeng e Jian Qin (2016, p. 27). A sua proposta baseia-se em literatura científica anterior, nomeadamente em Dempsey e Weibel (1996), Duval, Hodgins, Sutton e Weibel (2002), em Moen (2001) e em NISO (2017). Os cinco princípios propostos são modularidade (*modularity*), extensibilidade (*extensibility*), refinamento (*refinement*), multilinguismo (*multilingualism*) e interoperabilidade (*interoperability*). De entre estes princípios, destaca-se sobretudo para esta investigação a interoperabilidade, uma vez que é discutivelmente o princípio mais importante, na medida em que se encontra implícito nos demais, além de ser o princípio mais diretamente relacionado com o papel do conjunto de metadados que se pretende propor.

A interoperabilidade é um problema recorrente para os gestores da informação, responsáveis pela gestão dos metadados. Este conceito é reconhecido como sendo um principais e mais difíceis desafios na gestão dos metadados. Como tal, há várias tentativas de definições do conceito de interoperabilidade, das quais se destacam as seguintes:

- *Interoperability is the ability of multiple systems with different hardware and software platforms, data structures, and interfaces to exchange data with minimal loss of content and functionality* (NISO, 2004, p. 2)
- *Interoperability is the ability of two or more systems or components to exchange information and use the exchanged information without special effort on either system* (CC:DA, 2000)
- *"Interoperability: The compatibility of two or more systems such that they can exchange information and data and can use the exchanged information and data without any special manipulation* (Joudrey & Taylor, 2018, p. 646).

A ideia mais recorrente é a da compatibilidade entre dois sistemas, entendida como a capacidade de partilha de informação de forma eficiente e automática. Como tal, enquadra-se no âmbito da partilha, disseminação e descoberta dos metadados através da homogeneização dos processos e dos métodos usados para os definir e expressar. A discussão sobre a necessidade de atingir a interoperabilidade coloca-se como resposta ao rápido desenvolvimento nos setores responsáveis pela representação dos recursos, nos quais se observa uma tremenda proliferação de opções e ferramentas para a descrição dos conteúdos. A esmagadora diversidade de normas existentes foi capturada em 2010 por Jen Riley (2010), destacando as diferentes opções adotadas por diferentes comunidades, nomeadamente arquivos, bibliotecas e museus.

Contudo, embora estas comunidades sigam normas distintas para representar os seus recursos, cada vez mais a tendência é para a criação de repositórios integrados, nos quais uma única pesquisa varra os conteúdos apresentados por diferentes instituições. Neste contexto, numa situação ideal, os utilizadores *should be able to discover through one search what digital objects are freely available from a variety of collections, rather than having to search each collection individually* (Tennant, 2001, p. 118). Por conseguinte, *users should not have to know or understand the methods used to describe and represent the contents of the digital collection* (Chan & Zeng, 2006, p. 2). Porém, na prática, tanto os utilizadores como os responsáveis pela gestão dos metadados nas instituições enfrentam tremendos desafios em atingir este desfecho devido à diversidade das normas disponíveis para representação dos recursos. Neste cenário, torna-se fundamental desenvolver abordagens que procurem a interoperabilidade de modo a facilitar a partilha de metadados entre repositórios que usem normas distintas e de forma a possibilitar uma pesquisa transversal através de múltiplas coleções.

A função particular da interoperabilidade é destacada por Gilliland quando refere que os metadados são *governated by community-developed and community-fostered standards and best practices in order to ensure quality, consistency and interoperability* (Gilliland, 2016, p. 2). Porém, de modo a obter uma partilha eficiente de metadados, não basta alcançar a interoperabilidade ao nível sintático, mas também ao nível semântico (Moen, 2001). Não basta comunicar, é preciso saber interpretar, o que aponta para a existência de mais de um nível de interoperabilidade.

A *interoperabilidade sintática* implica a normalização das comunicações entre ambos os sistemas (Schaeffer, Baranski, Foerster, & Brauner, 2012; Sudmanns, Tiede, Lang, & Baraldi, 2018), sendo possível apenas por uma *strict adherence to standard metadata formats and protocols while creating metadata structures and possible only through metadata deployment* (Ramesh et al., 2015, p. 197). Sem a interoperabilidade sintática, *data and information cannot be handled properly with regard to formats, encodings, properties, values, and data types; and, therefore, they can neither be merged nor exchanged* (Koch, 2006, Interoperability; Zeng & Qin, 2016, p. 348).

A *interoperabilidade semântica* pode ser definida como a *capacity for different agents, services, and applications to communicate data, information, and knowledge while ensuring accuracy and preserving the meaning of that data, information, and knowledge* (Zeng & Chan, 2010, p. 4646). Trata-se de um conceito mais complexo na medida em que implica um

entendimento mútuo do significado dos dados e da informação durante a comunicação (Harvey, Kuhn, Pundt, Bishr, & Riedemann, 1999, p. 216), dependendo da correta associação entre os elementos de metadados de sistemas relacionados semanticamente (Ramesh et al., 2015, p. 197). Sem a interoperabilidade semântica, *the meaning of the used language, terminology and metadata values cannot be negotiated or correctly understood* (Koch, 2006, Interoperability).

Além disso, os resultados da interoperabilidade podem observar-se sob vários níveis e perspectivas, nomeadamente ao nível das normas (*schema level*), do *record* (*record level*) ou do repositório (*repository level*) (Chan & Zeng, 2006, p. 5; Zeng & Qin, 2016, p. 351). No contexto deste trabalho, a definição de um conjunto de metadados adequados para a representação dos recursos contribui para facilitar a interoperabilidade semântica, na medida em que ajuda a definir que metadados devem ser comunicados, independentemente dos protocolos ou formatos de comunicação. Por este motivo, o enfoque é colocado ao nível das normas, dando-se prioridade à escolha e adoção dos elementos de metadados independentemente da sua aplicação. De seguida analisam-se algumas normas de metadados colocando-se o enfoque na estrutura dos elementos de metadados. Apresentam-se os critérios de seleção das normas de forma a justificar a escolha das normas usadas no terceiro capítulo para se realizar uma análise mais profunda com vista à identificação dos elementos de dados que cada uma adota para a verificação da confiabilidade, mediante a demonstração da autenticidade e da fidedignidade.

## 2 Representação da informação

---

### 2.1 O conceito de representação

Qualquer esforço para compreender o conceito de representação poderá ser necessariamente difícil e insuficiente, dada a sua origem filosófica que lhe confere uma interpretação aberta. É tão difícil encontrar um consenso nesta área quanto é fácil encontrar contra-argumentos válidos para uma determinada posição. Apesar destes desafios, entende-se que as posições aqui assumidas constituem um ponto de partida útil o suficiente para a compreensão dos objetivos de investigação, ainda que com a consciência de que existem outras perspectivas válidas. De forma a orientar a leitura, apresenta-se um esquema orientador (Figura 10) que pode servir de fio condutor por entre o labirinto terminológico e conceptual das próximas páginas, tal qual como o fio de Ariadne.

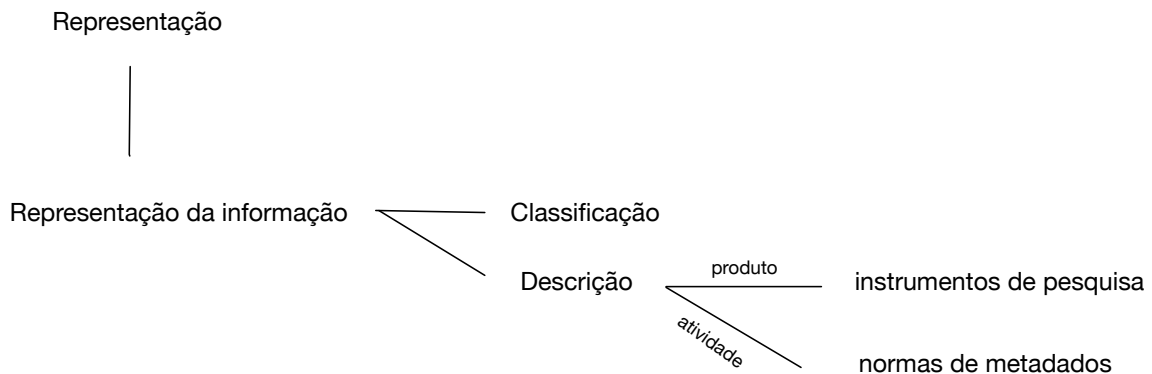


Figura 10 — Esquema conceptual dos conceitos teóricos a explorar.

FONTE: elaboração própria.

Esta figura pretende organizar e estruturar o raciocínio, explicitando as relações entre os conceitos e enquadrando o posicionamento de cada um na revisão de literatura. Neste sentido, no âmbito deste trabalho, a ‘representação’ é entendida como o conceito mais geral, com bases na abstração da filosofia. Quando trazemos o conceito de representação para a área da Ciência da Informação, passamos a lidar com o conceito de ‘representação da informação’, sendo por vezes designada também por ‘representação arquivística’ por alguns autores de Arquivística<sup>39</sup>. Por sua vez, entende-se que a representação da informação engloba a ‘classificação’<sup>40</sup> e a ‘descrição’ dos

---

<sup>39</sup> Vide: Yakel (2003) e Zhang (2012).

<sup>40</sup> O termo original é ‘*arrangement*’. A tradução portuguesa mais próxima será a de ‘classificação’, no sentido de realizar uma classificação dos fundos de carácter definitivo depositados em arquivos históricos ou definitivos, segundo a teoria das três idades de Schellenberg, inicialmente proposta em 1956 (Schellenberg, 2003).

documentos de arquivo. É de notar que não só os documentos de arquivo que podem ser alvo de descrição, podendo esta realizar-se a vários outros níveis, como ao nível da função, do produtor ou do custodiante. Porém, tendo em conta os objetivos da investigação, entende-se que o principal foco de análise devem ser os documentos de arquivo, pelo que se simplifica o esquema ao não incluir outras entidades que pudessem ser descritas.

O conceito de ‘classificação’, bem como uma breve contextualização do seu uso, será apresentado, mas não será analisado em maior detalhe uma vez que excede o âmbito dos objetivos propostos. Já no caso da descrição arquivística, veremos como esta pode ser entendida como i) produto, no sentido de criação de instrumentos de pesquisa que facilitem o acesso aos documentos de arquivo, ou como ii) atividade que se dedica à elaboração de representações dos documentos de arquivo, através da criação de metadados. Por fim, serão alvo particular do nosso estudo os metadados mais diretamente responsáveis pela garantia da autenticidade das descrições arquivísticas.

### 2.1.1 Representação

Rosenberg argumenta que *the essential and characteristic human activity is representation — that is, the production and manipulation of representations* (Rosenberg, 1981, p. 1). O ato de representar é um fenómeno profundamente enraizado no *ethos* do ser humano, com uma tradição consolidada ao longo de largas dezenas de milénios. A representação mais antiga de que temos conhecimento trata-se de um desenho abstrato com 73.000 anos, descoberto na África do Sul (Henshilwood et al., 2018). A primeira pintura data de há pelo menos 40.000 anos, descoberta na floresta de Bornéu, no sudoeste da Ásia (Aubert et al., 2018). Em Portugal, registam-se as pinturas de arte rupestre do Vale do Côa que datam de 22.000 a.C. (Baptista & Gomes, 1995), encontrando-se nesse xisto representações zoomórficas da Natureza inscritas pelo machado de pedra dos nossos antepassados.

O termo ‘representação’ deriva do latim *repraesentātiō*, *-ōnis* significando, numa aceção encontrada no *Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa*, a *ação de tornar presente ou sensível um objeto ausente ou um conceito, por meio de uma imagem, figura ou símbolo* (Academia das Ciências de Lisboa, 2001, p. 3203). No âmbito da ciência, as representações científicas permitem a formulação de hipóteses sobre os sistemas que se encontram representados (Frigg & Nguyen, 2020). Por exemplo, uma ressonância eletromagnética de um cérebro de um paciente fornece informações sobre o estado e funcionamento desse órgão e permite aos médicos definir linhas de ação em caso de anomalias. Swoyer (1991, p. 449) refere-se a este exercício de representação enquanto *surrogate reasoning*, significando que deve ser tido em conta o modo como a interpretação das representações leva a inferências sobre o que está a ser representado. Esta ideia será retomada adiante quando for discutida a representação da informação em arquivos.

Segundo o mesmo Dicionário da língua portuguesa, ‘representar’ tem origem no latim *repraesentāre*, significando re-presentar, isto é, tornar presente de novo, ou *trazer à memória*



(Academia das Ciências de Lisboa, 2001, p. 3204). Em arquivos, é muito interessante o sentido de ‘trazer à memória’. Tomando como pressuposto que os documentos constituem informação vertida num suporte, independentemente da natureza dessa informação (textual, gráfica, áudio) e da materialidade do suporte (pedra, papel, digital), há um elemento comum que preside a toda a *raison d’être* de qualquer registo: o prolongamento da memória, quer por questões de *accountability* ou de apoio à tomada de decisão no contexto de uma organização, quer para a criação de testemunhos históricos para gerações futuras.

Assim, as representações, em arquivo, são como que a memória das memórias, no sentido em que é por meio das representações que trazemos à memória a memória das atividades, ou, por outras palavras, é por meio das representações que tomamos conhecimento da existência, da identidade e da localização de documentos de arquivo. Destacam-se, assim, duas ideias: a dos metadados enquanto representação dos documentos, e a dos documentos enquanto representação dos eventos.

O conceito de representação não é tradicionalmente empregue no domínio dos arquivos para a compreensão dos documentos de arquivo, embora seja muito usado noutras áreas como a arte, o cinema, a linguística, a filosofia e a psicologia. Uma representação é, a nível fundamental, *something that stands, or is believed to stand, for something else* (Yeo, 2018, p. 129). A epopeia ‘Os Lusíadas’, de Luís de Camões, uma resenha histórica de um académico contemporâneo, ou o documentário ‘Os Descobrimentos Portugueses’, do realizador Bento da França, são representações da realidade (as jornadas dos Descobrimentos portugueses), naturalmente parciais.

As línguas naturais e a escrita também podem ser consideradas formas de representação. Na linguística, as línguas naturais são entendidas como um sistema de signos com expressão e conteúdo, formado pela combinação de significantes e significados (Barroso, 1999, p. 27). Alguns cientistas da informação reconhecem o papel desempenhado pelas representações simbólicas. Segundo László Karvalics (2007, p. 8), *systems of information technology operate with information converted into symbols*. Keith Devlin defende uma posição contrária segundo a qual os indivíduos registam dados nos sistemas de informação, tornando-se informação quando integrados num contexto (Devlin, 1999, p. 30) que lhes confira significado. Outros autores argumentam não que os dados sejam factos, mas antes que representam factos<sup>41</sup> (Beynon-Davies, 2009, p. 94; Dervos & Coleman, 2006, p. 52). Numa posição mais prudente, os cientistas da computação Terry Winograd

---

<sup>41</sup> Um facto é entendido como *deeds or acts made into something known ...which have had or do have actual existence, and are true and pertain to objective reality* (McCrank, 2002, p. 627). Porém, no seu cerne, qualquer “facto” é *an assertion by an individual in a context, based on a background of pre-understanding* (Winograd & Flores, 1986, p. 156). Diz-se que contra factos não há argumentos. Todavia, pode e deve haver argumentos contra a génese dos factos e contra a decisão de recolher, organizar e apresentar esses “factos” e não outros que pudessem realçar uma diferente dimensão da realidade representada.

e Fernando Flores sugerem que *computers ... manipulate symbolic representation that some person generated on the belief that they corresponded to facts* (Winograd & Flores, 1986, p. 156).

No entanto, nem todas as representações são exatas, isto é, nem sempre o objeto usado na representação corresponde a um aspeto ou elemento do mundo que se pretende representar. Por exemplo, os modelos geocêntrico e heliocêntrico são dois sistemas de representação, embora o nosso conhecimento atual nos permita afirmar que apenas o modelo heliocêntrico corresponde a uma representação exata da ordem dos corpos celestes no sistema solar. Noutros casos, como saber qual é a representação correta, entendendo por correto a fidedignidade a um elemento da realidade?

No âmbito da filosofia, Morrison (2008, p. 70) relembra que compete à teoria da representação tentar identificar o que constitui uma representação exata. Este exercício pode ser entendido como o problema de definir *standards of accuracy*, devendo cada representação estabelecer padrões que fomentem essa precisão (Frigg & Nguyen, 2020). Deve ser notado que o termo ‘exato’ conforme aqui usado é sinónimo de ‘preciso’. Uma esfera é uma representação exata da forma de um planeta, ao contrário de um quadrado. Isto não quer dizer que uma esfera seja um planeta, embora apresente um elemento comum que, num determinado contexto, a torna válida como representação de um planeta. Pode ser argumentado que nenhuma representação é inteiramente exata porque se o fosse, não seria uma representação mas o objeto representado.

Esta ideia de exatidão relaciona-se com o conceito de similaridade das representações. Segundo esta escola de pensamento, um objeto é similar a outro se apresentar propriedades em comum (Goodman, 1972; Quine, 1969). Por exemplo, uma maçã verde e uma persiana verde são similares no sentido em que ambas partilham uma propriedade comum (a cor verde), ainda que com tons diferentes. Neste sentido, entende-se que as representações, se forem similares em pelo menos uma propriedade, podem ser consideradas exatas e, se forem exatas, podem ser vistas como relativamente corretas.

Se nem todas as representações são exatas, então há representações que não são exatas. A própria reflexão sobre o que torna uma representação exata pressupõe a existência de representações inexatas. Isto levanta o problema da *misrepresentation*, ou falsas representações. Se A não representa B de forma exata, então A é uma representação falsa, não uma não-representação. Devemos portanto estar sempre alertas para a possibilidade de nos encontrarmos perante representações falsas (Contessa, 2007; Frigg, 2002; Suárez, 2003; van Fraassen, 2008).

Contudo, nenhuma representação é capaz de capturar todos os aspetos do fenómeno que pretende representar. Nenhuma representação é imparcial, assumindo-se sempre como uma perspectiva sobre um determinado fenómeno, moldada em função da interpretação subjetiva do representante. Além disso, estas noções de similar, exato ou falso não devem ser vistas em termos absolutos ou binários, mas antes enquanto largamente variáveis, dependendo da perspectiva de análise. Na prática, não se defende que haja descrições arquivísticas absolutamente exatas, porque terão insuficiências. Com este raciocínio apenas se pretende alertar para o facto de que certas descrições podem ser mais similares com o objeto a descrever do que outras. Por exemplo, podem ser mais similares se a descrição for mais completa (isto é, pode fornecer informação sobre mais

aspectos do documento ou apresentá-la com maior qualidade) ou mais verdadeiras se a informação fornecida não for falsa (por exemplo, indicando que a data de criação é de 1900 quando na verdade é 2001).

Além disso, é importante reforçar que estes conceitos dependem do modelo de análise, que é dizer da perspectiva de interpretação. O modelo geocêntrico é uma representação errada da disposição dos planetas no sistema solar, mas pode ser visto como uma representação exata do conhecimento científico de um determinado período da civilização. Um carro não só foi concebido como meio de transporte, como também representa em simultâneo o método da sua manufatura e a atividade de deslocação. Uma representação manifesta não só as intenções do seu autor, como também o próprio contexto que a origina. Acima de tudo, pretende salientar-se o papel das representações enquanto fenómenos socialmente cognoscíveis e contextualizáveis.

A questão das falsidades das representações é particularmente crítica no ambiente digital, uma vez que os documentos digitais existem enquanto objetos físicos (inscrição de signos num suporte), objetos lógicos (objetos reconhecíveis e processáveis por *software*) e objetos conceptuais (conforme reconhecido e compreendido por um indivíduo) (Thibodeau, 2002, p. 6). Na dimensão analógica, as descrições arquivísticas (em bases de dados, catálogos, índices), enquanto representações, constituem meramente a extensão digital de um objeto físico inteligível. Por inteligível pretende-se significar que a informação dos documentos em papel pode ser diretamente interpretada, sem necessidade de recurso a um intermediário tecnológico. Assumindo um bom estado físico do papel, a principal barreira para a compreensão do texto é o conhecimento da língua no qual este foi escrito. Esta abordagem contrasta com a relativa ininteligibilidade da informação em suporte digital. A informação contida num documento que esteja armazenado na memória apenas pode ser interpretado mediante o uso de *software*. Assim, na ausência de um computador ao qual se possa conectar, uma informação numa memória externa (p.e., uma *pen drive*) é ininteligível para os seres humanos.

Consideremos o exemplo de uma carta que Eça de Queirós escreveu a Antero de Quental. Mesmo que no catálogo de um arquivo seja indicado que o autor é 'Eça de Queirós' e o destinatário 'D. Luís de Castro', no momento da leitura desta correspondência — seja em papel, seja mediante uma digitalização que é uma representação exata da informação, se feita com boa resolução e qualidade — é possível identificar o erro e extrair a informação correta do destinatário. Na situação de um documento de arquivo nado-digital, o cenário é diferente, uma vez que o documento não existe numa forma inteligível para a cognição humana.

Retomando o mesmo exemplo, imaginemos que Eça de Queirós escreveu a mesma carta ao mesmo destinatário, mas fê-lo no século 22 e, em vez de papel e caneta, enviou uma mensagem por *e-mail*. A forma física do documento resume-se a uma sequência binária de *bits* (dados), um pacote de informação que circula pelos vários protocolos de comunicação da Internet. O *e-mail*, quando aberto por um programa, permite a leitura e interpretação do seu conteúdo porque o serviço de email opera a representação dos dados numa forma compreensível para os humanos, com linguagem natural disposta num interface gráfico. Este exemplo torna-se ainda mais claro quando

estamos a trabalhar com um documento WORD em .docx, e temos a possibilidade de ‘salvar como’ vários formatos tais como PDF, TXT, RTF, entre outros. Cada formato de exportação corresponde a uma diferente forma de representação dos dados que constituem a forma física do nosso documento. Para o computador, um documento é uma sequência de 0s e 1s, da mesma forma que o é a palavra “palavra”. A representação dessa sequência binária enquanto um ‘p’ e um ‘a’ que possamos compreender é uma gentileza do *software*.

Assim, no caso de documentos nado-digitais, a nossa interpretação da informação resume-se unicamente a representações, uma vez que não temos capacidade de compreender o documento na sua forma física, digital, legível apenas por máquinas. Para o entendermos, temos de criar uma representação familiar ao nosso entendimento. A representação final é uma de muitas possível, mas é mais do que uma representação. Para todos os efeitos, ela é o próprio documento, porque é a nossa única percepção do documento. Como é a única que conhecemos, é verdadeira para nós. Por esta razão, devemos ser muito críticos das falsas representações de documentos digitais, quer se trate da capacidade de reproduzir a informação contida num documento (preservação digital), quer no respeitante a descrições arquivísticas incompletas e/ou incorretas.

Outra característica importante diz respeito à direccionalidade da representação. Conforme destacado por Goodman (1976, p. 5), as representações fazem referência aos seus alvos, mas esses alvos nada contêm sobre as suas representações. Por exemplo, uma fotografia pode representar as janelas num edifício, mas essas janelas não representam a fotografia. Por conseguinte, existe uma direccionalidade inerente às representações, devendo as representações científicas identificar a raiz desta direccionalidade (Frigg & Nguyen, 2020). Aplicado à realidade dos arquivos, inferimos que os instrumentos de pesquisa (catálogos, inventários, bases de dados) têm de fazer referência direta aos documentos que representam, mas os documentos representados não têm qualquer ligação aos instrumentos de pesquisa. Disto decorre a capacidade de cada documento poder ser representado sob diversas formas. Esta ideia é importante porque, conforme será visto mais adiante, a descrição arquivística não tem de ser necessariamente unidireccional e monohierárquica, conforme veiculado pelas primeiras normas de descrição arquivística internacionais.

Por enquanto, pretende-se destacar que, tendo estes aspetos presentes, a construção e o uso de representações constitui um fenómeno transversal que ocorre na mente, verificando-se, por extensão, nos produtos da mente, nas artes, na literatura, na ciência, no desporto, nos sistemas económicos, nos sistemas geográficos, ou em qualquer outro ramo de atividade em que se aventure o pensamento humano. A finalidade da representação varia dependendo do campo. Por exemplo, as ciências cognitivas estudam representações mentais de forma a compreenderem a natureza do pensamento em relação ao mundo; as ciências físico-químicas estudam representações de átomos para conhecerem as reações químicas desencadeadas pela interação dos vários elementos existentes; a ciência da informação estuda a representação da informação de modo a perceberem o seu papel na organização da informação, na promoção do acesso à informação e no incremento da descoberta da informação (Zhang, 2012b, p. 47). Será precisamente sobre este último enfoque que se situa o nosso interesse, e que as próximas páginas desenvolverão.

## 2.1.2 Representação da informação

Nesta secção especifica-se e explora-se a representação da informação em arquivos, desde o papel das tecnologias às formas tradicionais de representação adotadas nos arquivos nos últimos dois séculos. Abordam-se os princípios do respeito pelos fundos e da ordem original, e discute-se a sua pertinência no contexto digital.

As formas de representação da informação no âmbito da Ciência da Informação, no que respeita aos arquivos, museus e bibliotecas, estão diretamente dependentes da disponibilidade da tecnologia para o efeito. Por um lado, as tecnologias de informação existentes num determinado momento da civilização condicionam a produção da informação, tanto em natureza como em volume. Por outro, ditam as potencialidades e o alcance do seu tratamento em sistemas de representação.

Na Suméria, por volta de 3.300 a.C., a produção de informação consistia numa escrita pictórica sobre tabuinhas de argila. A nível de produção e tratamento da informação, era uma tecnologia que não permitia registos muito longos nem numerosos, dado o labor do processo de escrita. Por consequência, também não promovia o aparecimento de instrumentos de descrição, tanto pela dificuldade em os criar, como pela ausência de uma necessidade deles, dado o menor número de documentos disponíveis.

A introdução do papiro, por volta de 2500 a.C., do pergaminho, na viragem do milénio 0, e da imprensa na Europa, por Gutenberg por volta de 1440, foram momentos marcantes no que toca à facilidade de registo de informação, aumentando o número de documentos em circulação e promovendo o aparecimento de instrumentos de pesquisa mais variados como guias, catálogos e inventários em bibliotecas e arquivos. Por fim, o advento do meio digital afirma-se como a grande última revolução nas tecnologias da informação. As novas formas de produção, disseminação, acesso, descoberta, recuperação da informação colocam inéditos desafios à tradição arquivística, modificando inclusivamente a relação que os indivíduos estabelecem com a informação<sup>42</sup>. Em contrapartida, também contamos hoje com a existência de tecnologias de representação cada vez mais poderosas, como é o caso das tecnologias de representação em grafo (RDF, tecnologias semânticas e *Linked Open Data*). As representações geradas têm a função prática de agir como substitutos de coisas (documentos) que, caso contrário, seriam inacessíveis ou ausentes (Ankersmit, 2001; Cummins, 1996).

Geoffrey Yeo (2018, pp. 1370-7) explora a ideia de ‘*records*’ (documentos de arquivo) como representações enquanto *representation of occurrents*. O termo ‘ocorrentes’ é usado sobretudo em filosofia em referência a fenómenos temporais com um fim demarcado. Este autor argumenta que um documento de arquivo representa um ocorrente já terminado, constituindo assim

---

<sup>42</sup> Paul Conway (2010) estuda a forma como a digitalização de fotografias, como modo de facilitar o acesso remoto, constitui um processo de representação que transforma as relações entre os arquivistas, utilizadores e as coleções.

uma extensão temporal dos fenómenos que representam (Yeo, 2007, 2008, p. 136). Como tal, *records are persistent representations because they have the capacity to endure beyond the circumstances that lead to their creation* (Yeo, 2018, p. 130). Esta ideia completa a já mencionada noção de documentos enquanto representação de eventos passados.

Esta perspetiva tem por base a teoria dos atos cujas bases se fundam no pensamento do filósofo inglês John Langshaw Austin, posteriormente desenvolvida pelo seu discípulo e filósofo John Searle. A teoria dos atos de fala analisa a forma como as pessoas usam a linguagem. Austin (1962, pp. 4-7) constatou que, quando comunicamos oralmente ou por escrito, estamos a desempenhar um ato, quer seja dar uma ordem ou a fazer uma declaração, que designou por ‘*performative utterance*’ (enunciados performativos). Tais enunciados são uma característica comum a muitos documentos. Num testamento, a redação ‘Entrego a casa ao meu filho’ constitui um enunciado performativo escrito com a qual os diplomatas estão familiarizados pelo nome de ‘documento dispositivo’<sup>43</sup>. ‘Peço desculpa’ ou ‘prometo’ são outros exemplos usados pelo autor.

Austin (1962, pp. 133-40) argumenta que a declaração de fazer um pedido de desculpas é uma ação em si mesma, a par de ação de pedir desculpa. Assim, as representações são componentes-chave dos atos de fala (Searle, 1986, 2002), e os atos de fala manifestam-se em muitas das comunicações entre as pessoas, mesmo quando tais comunicações aparentam ter somente uma vertente informativa. Se estivermos perante um documento que afirme ‘Paguei a fatura’, há duas camadas de interpretação a considerar: a representação do ato de pagar a fatura, e a representação da ação de declarar que a fatura foi paga (Yeo, 2018, p. 137). Embora interessante, esta discussão afasta-se da linha de investigação proposta, uma vez que diz respeito ao entendimento dos documentos de arquivo como substitutos dos fenómenos que representam. Para fins da investigação salienta-se a relação entre as representações arquivísticas e os objetos por elas representados, constituindo as representações dos documentos de arquivo substitutos desses documentos.

Recuperando o conceito de *surrogate thinking* introduzido aquando da discussão sobre a representação científica, considera-se que as descrições arquivísticas atuam como um mecanismo para tornar presente um fenómeno ausente. Ausente não é usado no sentido físico, mas ao nível da cognição, embora se admite também uma distância física quando se usam descrições de documentos que estão noutra local. O uso de um instrumento de pesquisa de um arquivo, pode ser feito remotamente, fisicamente ausente do local de armazenamento dos documentos, mas também a partir da própria instituição, com o documento numa prateleira próxima. Mesmo neste caso, o uso do instrumento de pesquisa, enquanto produto da descrição arquivística, serve para tornar presente ao nosso entendimento a existência de um documento que, por insciência, se encontra ausente do nosso conhecimento do mundo. Neste sentido, ao longo da tese o enfoque é colocado nas descrições arquivísticas enquanto representações dos documentos, não nos documentos enquanto

---

<sup>43</sup> De acordo com uma perspetiva diplomática, um documento dispositivo é um *retrospective record whose purpose is to put into existence an act, the effects of which are determined by the writing itself; that is, the written form of the record is the essence and substance of the act* (InterPARES 2, 2008c).

representação de ações. Por extensão, os metadados são entendidos como asserções (*statements*) sobre documentos, que por sua vez são asserções de uma determinada ação. Assim, os metadados emergem como formas de representação.

No âmbito da arquivística, o conceito de representação foi adotado no final do século 20 para descrever a estrutura e a função dos instrumentos de pesquisa em arquivos (M. Cook, 1999; Procter & Cook, 2000), tendo desde então sido um conceito debatido e explorado na literatura arquivística. De uma forma geral, os profissionais da informação dependem significativamente da criação e da manipulação de representações. Catálogos de biblioteca, esquemas de classificação, listas de matérias, índices, resumos, instrumentos de pesquisa em arquivo, metadados, inventários museológicos, taxonomias ou ontologias constituem formas de representação de recursos informacionais aos quais pretendem facilitar o acesso (Zhang, 2012b, p. 48). No entanto, tal como as representações noutros campos científicos, as representações da informação são parciais e influenciadas por diversas variáveis (Duff & Harris, 2002, p. 275).

Wendy Duff e Verne Harris consideram que as representações de informação são enviesadas porque refletem uma visão do mundo própria com objetivos específicos. Ainda que não voluntariamente, os contextos culturais e sociais de um indivíduo terão implicações inevitáveis no seu entendimento do mundo e, por extensão, das representações desse mundo por ele construídas. Além disso, as representações são também *a more or less incomplete view of things being represented and something is always lost in the representation process* (Zhang, 2012b, p. 48). Deparamo-nos de novo com o problema previamente discutido das representações incompletas ou falsas representações. Esta preocupação leva ao questionamento das representações, de onde origina a ideia de que qualquer representação se figura como uma recontextualização de um objeto, moldada por um determinado contexto, contendo uma carga de subjetividade que deve ser controlada mediante modelos, princípios e regras de descrição.

Mas qual é, afinal, o propósito da representação de informação? De modo a responder a esta questão, é necessário trazer para a discussão o conceito de recuperação. Segundo David Blair, *the central task of information retrieval is to understand how documents should be represented for effective retrieval* (Blair, 1990, p. vii). Por sua vez, Heating Chu destaca que a informação tem de ser representada antes que possa ser recuperada e que, no ambiente digital, a qualidade da representação da informação tem implicações diretas na capacidade de recuperação da informação<sup>44</sup>:

*we understand that retrieval performance suffers when information representation is not properly done. The importance of having quality information representation cannot be*

---

<sup>44</sup> Sendo os metadados a forma mais preponderante de representação de informação no mundo digital, a qualidade de representação está diretamente associada à própria qualidade dos metadados. No entanto, esta rica e complexa temática (Meshkin, 2010), embora relacionada, constitui uma discussão inteiramente diferente que não será tida aqui em conta.

*overemphasized. We should therefore aim for quality information representation so that we can find what we are looking for in the world of digital information* (Chu, 2010, p. 45).

De forma complementar, Zhang acrescenta que a representação serve *to facilitate effective information discovery and provide adequate archival context* (Zhang, 2012a, p. 46). Considera-se importante a menção ao contexto, uma vez que este é indispensável para a compreensão da informação recuperada. Assim, pode-se afirmar que a informação é representada de modo a facilitar a sua recuperação, e que nessa representação deve estar explicitado o contexto arquivístico, indispensável para a compreensão do sentido e propósito da informação recuperada. Esta perspectiva será muito importante para explorar novas ferramentas para a descoberta de informação em rede, impulsionada pelo aparecimento de tecnologias de informação cada vez mais poderosas, como as tecnologias de representação em grafo que serão discutidas adiante. O termo ‘recuperação da informação’ não deve ser confundido com o termo ‘*archival retrieval*’ (recuperação arquivística), que denota *the process of removing archives from storage* (C. Brown, 2015, p. 82) para fins de divulgação ou exposição.

A representação de informação em arquivos também é por vezes designada de ‘*archival representation*’, nomeadamente no artigo seminal de Elizabeth Yakel intitulado *Archival representation*, publicado em 2003. Desde então, muitos autores têm adotado esta terminologia para discutir estas questões (Frogner, 2015; McQueen, 2018; Sinn, 2015; Zhang & Mauney, 2015). Yakel prefere o termo ‘*archival representation*’ porque acredita ser mais rigoroso a capturar o trabalho do arquivista em organizar, interpretar, criar substitutos e desenvolver arquiteturas para sistemas de representação que possam conter esses substitutos de materiais de arquivo. Assim, a representação arquivística é entendida como:

*a fluid, evolving and socially constructed practice. Representation refers to both the processes of arrangement (respecting or disrespecting order) and description, such as the creation of access tools (guides, inventories, finding aids, bibliographic records) or systems (card catalogs, bibliographic databases, EAD databases) resulting from those activities* (Yakel, 2003, p. 2).

A característica de ser *socially constructed* remete-nos para a questão da parcialidade e da subjetividade da representação previamente discutida, na medida em que uma representação, enquanto produto, depende sempre da mundividência do sujeito que a constrói. Os arquivistas desempenham as suas funções num contexto histórico (cultural, social, material) específico, sendo as suas decisões e julgamentos condicionados por esses mesmos contextos (International Council on Archives, 2016, p. 6). Além disso, é articulado o modo como a representação se desmembra nos processos de *arrangement* e de descrição. Estes dois processos afirmam-se como a tarefa missionária dos arquivistas, através da qual cumprem os fundamentos da sua profissão de tornar os recursos acessíveis. Sem estas atividades, *archival materials would remain lifeless, inaccessible, unknown, and unused; they are ... gifts that often bear fruit in a passion to make accessible the traces of human activity that we call archives and manuscripts* (Prom & Frusciano, 2013, p. 1).



Conforme foi dito antes, em português, o termo ‘*arrangement*’ andarรก mais pr3ximo daquilo que se entende por classifica73o em arquivos definitivos, termo usado no Brasil por ‘arranjo’, se for seguida a tradi73o anglosax3nica. O conceito de arquivos hist3ricos deriva do princ3pio das Tr3s Idades, de Schellenberg (2003), inicialmente proposto em 1956 na obra *Modern Archives: Principles and Techniques*. Segundo esta classifica73o, o ciclo de vida dos documentos 3 composto por tr3s fases: uma ativa, de intenso uso administrativo; uma interm3dia, de consulta moderada; e uma inativa, enquanto dep3sito hist3rico da mem3ria. O processo de classifica73o (*arrangement*) ocorre sobretudo nesta 3ltima fase, quando os documentos de arquivo s3o incorporados nos arquivos de car3cter hist3rico.

O *arrangement* dos documentos de arquivo 3 fundamental porque estes se assumem como *surrogates for memory, preserving evidence of events and ideas for future reference* (Jimerson, 2015, p. 99). Os documentos de arquivo constituem evid3ncia do passado, sendo essa evid3ncia extrapolada de v3rios contextos internos e externos aos documentos, tais como: o local e a 3poca (o tempo) em que o criador dos documentos existiu, e nos quais os documentos originaram; as fun73es e as atividades desenvolvidas pelo criador dos quais os documentos resultam; os sistemas de gest3o documental usados para organizar os documentos; ou ainda pelo pr3prio conte3do dos documentos (Trace, 2015, p. 21). Neste sentido, *arrangement* afirma-se como *a process of identifying relationships, not a process of physically ordering and storing documents* (Eastwood, 2000, p. 93). Neste processo, os arquivistas lidam com os documentos no seu coletivo, enquanto agregados documentais, e representam numa estrutura hier3rquica os v3rios contextos internos e externos aos documentos. Documentar estas rela73es 3 documentar os contextos em que os documentos foram criados e utilizados, garantindo os v3nculos entre os documentos, bem como os v3nculos entre os documentos e as fun73es e os produtores.

Os princ3pios predominantes da classifica73o arquiv3stica foram consolidados no mundo ocidental no s3culo 19. Previamente, as regras de classifica73o correspondiam aos ideias Iluministas de classifica73o do conhecimento consoante a pertin3ncia dos investigadores, dando azo ao desenvolvimento de sistemas de classifica73o por tem3ticas. Contudo, a emerg3ncia dos estados modernos em in3cios do s3culo 19, com o impl3cito aumento dos volumes documentais, exigiu a ado73o de pr3ticas classificativas mais eficientes. Neste contexto, notabilizou-se, em Fran7a, num per3odo p3s-revolucion3rio, o princ3pio de *respect des fonds* (respeito pelos fundos), um novo sistema de classifica73o administrativa descrito por Natalis de Wailly numa circular interna do Minist3rio Franc3s do Interior, a 24 de abril de 1841. Segundo este princ3pio, o primeiro crit3rio de organiza73o do conjunto documental 3 de acordo com o criador, constituindo um fundo<sup>45</sup>.

---

<sup>45</sup> Um fundo (*fonds*) refere-se ao conjunto de documentos de arquivo que uma organiza73o, fam3lia ou indiv3duo criaram e acumularam como resultado de um processo org3nico, refletindo as fun73es do criador (Society of American Archivists, 2019b). Constitui o conceito base do princ3pio de respeito pelos fundos (*respect des fonds*), que dita que os arquivos (documentos de qualquer tipo) criados ou origin3rios de uma administra73o, pessoa ou entidade org3nica devem ser agrupados, sem se misturarem com documentos de

Nas palavras de Wailly, esta forma de organização consiste em *to gather together by fonds, that is to unite all the deeds (i.e., all the documents) which come from a body, an establishment, a family, or an individual, and to arrange the different fonds according to a certain order .... The documents which only make reference to an establishment, a body or a family must not be confused with the fonds of that establishment, body, family* (Duchein, 1983, p. 66). Uns dias mais tarde, respondendo a críticas espoletadas por esta nova ideia, Natalis de Wailly reforçou a sua confiança neste princípio enquanto método de organização documental:

*A classificação pelo fundo é a única forma realmente correta de assegurar a execução devida de uma ordem regular e uniforme ... Se, em vez de este método, que podemos afirmar ser baseado na natureza das coisas, for proposta outra ordem teórica ... os arquivos cairão numa desordem difícil de emendar... Noutra qualquer tipo de classificação que não pelo fundo corre-se o enorme risco de não saber onde encontrar um documento ...* (Desjardins, 1890, p. 30-31).<sup>46</sup>

Em essência, o princípio pragmático de respeito pelos fundos reconhece a individualidade orgânica das instituições ao determinar que os documentos criados por uma entidade no decurso da sua atividade não devem ser misturados com os documentos de outra entidade. Em 1898, os arquivistas holandeses Samuel Muller, Johan Feith e Robert Fruin reforçaram este princípio com contributos teóricos e práticos, no manual *Handleiding voor het Ordenen en Beschrijven van Archieven* (*Manual para a classificação e descrição dos arquivos*)<sup>47</sup>.

Nesta obra, a ideia de proveniência dos documentos, consubstanciada no conceito de fundo, foi desenvolvida para se aplicar aos documentos no contexto do fundo, numa aplicação interna da proveniência. Segundo estes autores, *the system of arrangement must be based on the original organization of the archival collection* (Müller, Feith, & Fruin, 2003, p. 52). Assim, o princípio do respeito pelos fundos desenvolve-se em outros dois princípios, o princípio da proveniência e o princípio da ordem original. O princípio da proveniência dita que os documentos de arquivo criados, acumulados e usados por uma pessoa ou um grupo no decurso da sua vida ou atividade devem ser mantidos juntos, sem se misturarem com outros documentos de uma entidade distinta (J.

---

outros fundos (Duchein, 1983, p. 64). O princípio do respeito pelos fundos tem sido a forma dominante de descrição e classificação arquivísticas, a ponto de ter levado Natalis de Wailly a afirmar ser o único método possível de classificação (Desjardins, 1890, p. 30-31). Atualmente, com o ambiente digital, estes posicionamentos começam a ficar ultrapassados.

<sup>46</sup> *Le classement général par fonds et par matières est le seul vraiment propre à assurer le prompt accomplissement d'un ordre régulier et uniforme ... Si au lieu de cette méthode, qu'on peut dire fondée sur la nature es choses, on propose un ordre théorique ... les archives tomberont dans un désordre auquel il sera difficile de remédier... Dans tout autre classement que celui par fonds, on court grand risque de ne savoir où demander cette pièce ...*

<sup>47</sup> A tradução brasileira existente, de 1960, intitula-se *Manual de arranjo e descrição de arquivos*.

Bailey, 2013). A ordem original refere-se ao agrupamento intelectual e ordenação de documentos no contexto da sua acumulação, e mantê-la é essencial não só para a compreensão das relações partilhadas pelos documentos, como também para a constituição de evidência do modo como eram usados (International Council on Archives, 2016). O princípio da proveniência pode ser entendido como um corolário externo do princípio do respeito pelos fundos, ao passo que o princípio da ordem original se assume como a sua manifestação interna<sup>48</sup>. Em conjugação, estes princípios satisfazem a necessidade essencial de *to know where it [o documento] was created, in the framework of what process, to what end, for whom, when and how it was received by the addressee, and how it came into our hands* (Duchein, 1983, p. 67).

As traduções deste manual holandês em alemão (1905), italiano (1908) e francês (1910) ajudaram a cimentar estas ideias como parte de uma prática arquivística europeia. Em 1910, num congresso internacional de arquivistas e bibliotecários realizado em Bruxelas, delegados de Portugal, França, Alemanha, Rússia, Holanda, Itália, Espanha, Áustria-Hungria, Grã-Bretanha e Estados Unidos da América concordaram em como o princípio da proveniência se assumia como a melhor forma de organizar e descrever os arquivos, *not only from the point of view of a logical classification of pieces but also in the best interest of historical study* (Wosh, 2011, p. 145).

Apesar de os EUA terem marcado presença nesse congresso, o princípio da proveniência para a gestão dos documentos federais só viria a ser adotado mais tarde, em 1940, pelo *National Archives and Records Service* (NARS), o arquivo nacional criado em 1934 pelo presidente Franklin Roosevelt. A nível terminológico, *arrangement* substituiu *classificação*, e foi empregue o termo ‘*record group*’ como equivalente de ‘*fonds*’, embora com um significado distinto de natureza mais prática. O reconhecimento dos esforços europeus foi intensificado com a tradução do manual holandês em inglês no mesmo ano de 1940, numa tradução realizada por Arthur Leavitt, um trabalhador do NARS. Ao longo das próximas décadas, o influente arquivista Theodore Schellenberg ajudou a consolidar estas ideias através de numerosas circulares e publicações académicas.

É também de salientar a influência do modelo de Oliver Wendell Holmes (1964), outro trabalhador do NARS, no qual é formulada de forma explícita a noção de que a classificação arquivística envolve *a breakdown of the whole into its parts* (Schellenberg, 1988, p. 88). O modelo de Holmes fundamenta-se na ideia de fundo indivisível, afirmando inclusive que *what an agency has created in the past no man today can completely tear asunder* (Holmes, 1964, p. 22). Holmes propõe cinco níveis hierárquicos de classificação: *depository, group and subgroup, series, filing unit e document*. A classificação ao nível do *record group* e do *subgroup* era vista como a realização prática do princípio de respeito pelos fundos, ao passo que os níveis da *series* e inferiores

---

<sup>48</sup> A hierarquia entre estes “três” princípios é relativamente discutível. Por exemplo, Horsman (1993, p. 51) defende que uma aplicação interna do princípio da proveniência dá-nos a ordem original, ao passo que uma aplicação externa se manifesta no respeito pelos fundos.

testemunhavam a ordem original (Trace, 2015). A classificação de acordo com estes cinco níveis tinha como propósito determinar *the order or sequence in which records ought to be placed in containers and in which the containers ought to be labeled and shelved* (Holmes, 1964, p. 24). Neste sentido, segundo esta visão da classificação arquivística, a classificação intelectual confundiu-se, de início, com a classificação física, uma vez que esta ordem de organização intelectual corresponde à ordem exata segundo a qual os documentos devem ser fisicamente organizados. Além disso, a classificação figura como um trabalho prévio à descrição, pois os instrumentos de pesquisa têm de referir os níveis de classificação.

O modelo de Holmes ainda é usado atualmente, embora com críticas crescentes motivadas sobretudo pela classificação da informação digital. Cada vez mais, como veremos mais adiante, surgem vozes críticas na literatura da área que levantam problemas quanto às ideias de criador único, de estrutura hierárquica e de ordem original presentes nas tradicionais formulações do princípio de respeito pelos fundos. Yakel afirma que o *arrangement* se refere a uma tentativa de representação *of an intended and well-tended order that probably never existed in the original context* (Yakel, 2003, p. 2). Por sua vez, Kirschenbaum alerta para o facto de que, em ambiente digital, *every contact leaves a trace* (Kirschenbaum, 2008), razão pela qual o ato de interpretar um objeto digital modifica constantemente a sua ordem material.

Além disso, a sobreposição das classificações física e intelectual não tem necessariamente de ocorrer, e afirmamos inclusive que é limitador que ocorra, conforme exposto pelo princípio da direcionalidade da representação anteriormente visto segundo o qual pode haver múltiplos instrumentos de pesquisa para um mesmo documento. Por conseguinte, determinada organização física não invalida que sejam desenvolvidos vários sistemas de representação e recuperação da informação. Observa-se, assim, a emergência de duas noções no que respeita à representação da informação, a ideia de organização física, e a sua descrição intelectual.

Conforme tem vindo a ser notado, *while the management of the storage and intellectual description are commonly closely related, the two activities are distinct, and the relation between them is not essential* (International Council on Archives, 2016, p. 5). Por consequência, no contexto de desenvolvimento deste trabalho, afastamo-nos da interpretação do princípio da proveniência que diz respeito ao modo como os documentos são mantidos e armazenados fisicamente nos arquivos definitivos (classificação), para nos focarmos na componente relacionada com a descrição intelectual do conjunto e das partes do conjunto (descrição).

Assim, o modelo que se propõe no último capítulo constitui uma forma de descrição arquivística, isto é, de realizar o controlo intelectual, independentemente do controlo físico. Porém, o próprio modelo é uma forma de representação, no sentido em que os metadados que compõe o modelo são uma representação dos documentos de arquivo. Além disso, como se pretende propor um modelo para a autenticidade, o modelo é seletivo nas características dos documentos que pretende representar. Como tal, o modelo sob discussão não deve ser visto como uma cópia de um certo estado, mas antes como um modo de ação, isto é, como um artefacto cognitivo que guia uma prática (Wartofsky, 1979, p. xv). Neste contexto, tal prática é a representação da informação que,

mais especificamente, é a descrição arquivística autêntica. Como tal, explora-se o conceito de descrição arquivística enquanto parte integrante do processo de representação da informação em arquivos, procedendo primeiro a uma contextualização da corrente cultural que caracteriza os tempos atuais.

## 2.2 Arquivos na pós-modernidade

O arquivista Terry Eastwood constatou que *one must understand the political, economic, social and cultural milieu of any given society to understand its archives* acrescentando que *the ideas held at any given time about archives are surely but a reflection of wider currents in intellectual history* (Eastwood, 1993, p. 27). Respeitando este raciocínio, uma das principais correntes intelectuais na atualidade é o pós-modernismo, pelo que se torna necessário compreender este movimento para compreender o seu impacto na teoria arquivística e, por extensão, os arquivos na atualidade.

A primeira menção por um arquivista aos impactos da pós-modernidade na teoria arquivística terá sido Terry Cook, no artigo *Electronic records, paper minds: the revolution in Information management and archives in the post-custodial and post-modernist era* (T. Cook, 2007), publicado originalmente em 1994, aprofundado noutro artigo três anos mais tarde (T. Cook, 1997) e republicado em 2007 com novas reflexões. Contudo, ao longo da década de 1990, muita literatura foi escrita sobre as influências da pós-modernidade nos arquivos. Entre outros arquivistas pós-modernos pioneiros destacam-se Brien Brothman (1991, 1993, 1999) e Richard Brown (1991b, 1991a, 1997). Nos anos seguintes, vários arquivistas canadianos expuseram as suas reflexões, nomeadamente Preben Mortensen (1999), Tom Nesmith (1999), Bernadine Dodge (1997), Theresa Rowat (1993), Joan Schwartz (1995), Lilly Koltun (1999), Frank Upward (2000), Sue McKemmish (2002) e Barbara Reed (1997). Na América, contribuíram Margaret Hedstrom (1993), Richard Cox (1997) e James O'Toole (1994). Noutras partes do globo, Eric Ketelaar (1999) e Verne Harris (1997) apresentam também reflexões importantes<sup>49</sup>.

Por sua vez, como o pós-moderno apenas se pode entender face ao moderno, uma vez que surge em reação a este, torna-se necessário compreender os moldes da teoria arquivística no pensamento moderno, que se identifica com os séculos 18, 19 e pelo menos a primeira metade do século 20. Por um lado, a Arquivística era entendida como uma disciplina auxiliar da História, juntamente com a Diplomática e a Paleografia, com uma demarcada componente prática e aplicada (Pacheco, 2017, p. 21). Por outro, a corrente positivista caracterizava a intelectualidade no geral, e a ciência em particular.

---

<sup>49</sup> Esta breve lista tem apenas carácter indicativo, sem ambição de exaustividade. Outros autores há que escreveram sobre temas diversos como Hugh Taylor e David Bearman e outros autores sobre temas mais específicos, como Barbara Craig sobre *appraisal theory*, Catherine Hobbs sobre arquivos pessoais.

O termo positivismo foi introduzido pelo filósofo francês August Comte (1798-1857). O argumento principal de Comte era o de que o conhecimento se encontrava limitado pelo que era possível determinar pela experiência, ou seja, positivamente determinado. Comte insistia que *religious, metaphysical, or philosophical assumptions could not be part of any explanation of natural or social phenomena. Natural and social phenomena should only be explained without reference to anything beyond themselves* (Mortensen, 1999, p. 7).

Com bases no Iluminismo, isto traduz-se numa crença na ciência e na sua objetividade. A ideia positivista da ciência entende que esta está acima de qualquer contexto histórico social, cultural ou político. A ciência é entendida como residente na sua própria verdade objetiva do mundo, comprovada empiricamente pela experiência, logo, alheia ao mundo subjetivo da análise humana. Esta concepção que apenas viria a ser desconstruída por historiadores e filósofos no último quartel do século 20, entendendo-se hoje que nenhuma experiência pode ser completamente objetiva na medida em que é desenhada, executada e interpretada por um indivíduo enquadrado num determinado contexto social, económico e político. Nenhum cientista existe num vácuo, pelo que a sua visão invariavelmente subjetiva do mundo transborda necessariamente para a ciência que produz.

Verne Harris cruza os principais pressupostos do paradigma positivista com o entendimento moderno dos arquivos, resumindo nalguns pontos centrais as suas implicações na teoria arquivística:

- *1. The meaning of the word “archives” is simple, stable, and uncontested. Archives are documents or records, in whatever medium, identified for preservation in archival custody; and archives is the place where such records are preserved or an institution providing such places.*
- *2. Archives, in the sense of archival records, are the organic and innocent product of processes exterior to archivists and reflect, provide an image of, are evidence of, those processes. Stated more crudely, the idea is that archives reflect reality.*
- *3. ... the transformation discourse:*
  - *still defines archival endeavour in terms of custodianship and conceptualizes archives in terms of physical things and places of custody;*
  - *proposes a (narrower) shaping of the record as the carrier of memory rather than a (broader) participation in the process of memory formation; and*
  - *posits the primary archival challenge as being the preservation of a wider and richer reflection of reality* (Harris, 1997, p. 132).

A concepção positivista dos arquivos originou o aparecimento de duas correntes distintas na teoria arquivística, uma que vota a atividade arquivística a uma mera componente prática que dispensa a teoria, e outra que defende que a arquivística pode ter uma componente teórica que a eleva ao patamar de ciência. Conforme explica Mortensen:

*One position denies that there is (or should be) any significant role for theory in the practice or study of archives. Archival work, in this view, is a practical matter, and best left as such. Archivists have no need for theory. ... Contrary to this, others hold that there is such a thing as archival theory and that it has, or can have, the character of a science: archival science (Mortensen, 1999, p. 2).*

Um dos defensores da primeira posição foi John Roberts, que argumenta que os arquivistas apenas carecem de conhecimento dos procedimentos e da tecnologia, da ética da profissão, das expectativas sobre eles, da história, e dos documentos com os quais trabalham (J. Roberts, 1987, p. 74). Roberts valoriza especialmente o conhecimento histórico enquanto razão de ser dos arquivos, ao mesmo tempo que se opõe fortemente à aquisição de uma componente teórica, visto que *archival theory does two things that are profoundly threatening to clarity of thought: it overcomplicates that which is simple and it oversimplifies that which is complicated* (J. Roberts, 1990, p. 111).

Em oposição a esta corrente encontra-se o da ciência arquivística (*archival science*). O Manual Holandês de 1898 foi um ponto de viragem no entendimento científico da arquivística por constituir a primeira tentativa de elaboração de um *scientific archival treatise* considerado *the first real effort to articulate systematically the concepts and methods that find their validity in archival theoretical ideas with internal logic and consistency, rather than in their historical, legal, or cultural context* (Duranti, 1996, p. 5). Para ser devidamente científica, *archival science had to seek its purpose or focus within its own horizon rather than outside, and operate as a self-referential system, fully autonomous from the influences of political, juridical, or cultural conceptions* (Duranti, 1996, p. 5). Assim, a arquivística apenas adquiriria cientificidade quando se abstrairse das grilhetas dos contextos históricos e sociais, na medida em que obteria validade universal e cumpriria três critérios necessários: universalidade, autonomia e consistência e lógica internas.

Tanto uma posição como outra caem em extremos que convidam a argumentos desfavoráveis. Por um lado, a corrente prática é demasiado punitiva para com a componente teórica, uma vez que a prática alimenta a teoria, mas a teoria também é necessária como guia da prática através da crítica. Por outro, a perspectiva científica coloca uma fé excessiva na capacidade de universalidade, na medida em que nenhum sistema de referência científico pode afirmar-se ser plenamente autónomo e isento de influências sociais. Neste sentido, salienta-se a posição intermédia de Mortensen que argumenta que a teoria é indispensável, embora não possa ser inteiramente independente do contexto histórico, social e político. Como tal, entende-se que *theories are developed within archival practice and must be understood as a product of this practice itself. They are developed as solutions to specific problem situations or as deliberations on particular practices* (Mortensen, 1999, p. 3). Acredita-se que existe uma clara necessidade de teoria na Arquivística que deve guiar a prática, caso contrário teríamos uma prática acrítica que poderia facilmente cair na desatualização e na simplificação da realidade. Porém, não compete a esta investigação aprofundar a epistemologia da Arquivística, nomeadamente acerca do seu estatuto de ciência ou de disciplina. Este aspeto é um tema em aberto neste campo científico, observando-se divergências não só entre regiões, como entre autores dessas mesmas regiões. A título de exemplo, em Portugal é geralmente entendido que

a Arquivística constitui uma disciplina da Ciência da Informação, uma posição não subscrita por diversos autores brasileiros que entendem antes a Arquivologia como uma ciência autônoma. Independentemente da hierarquia científica, não parece ser questionável a importância da componente teórica.

Uma obra marcante do período de discussão entre arquivística como teoria ou prática é a ilustrativamente intitulada *Modern Archives: Principles and Techniques*, que o influente arquivista norte-americano Theodore Schellenberg publicou em 1956. Esta obra assinala o epíteto da modernização da cultura arquivística, sob o enfoque do período pós-guerra nos EUA (Ridener, 2009, p. 69). Schellenberg dedica a maior parte da sua atenção à avaliação (*appraisal*), sendo considerado por Terry Cook (1999) o pai desta temática, uma vez que a sua teoria sobre a avaliação de documentos de arquivo foi a mais influente e mais adotada nos EUA (Bantin, 1998, p. 25). Embora não se pretenda discutir a avaliação arquivística, assinala-se esta obra para destacar duas ideias da conceção moderna dos arquivos. Primeiro, a do arquivista como guardião da verdade, defendendo Schellenberg que o contributo do arquivista *to the search for truth lies in making available the evidence that is in his possession* (Schellenberg, 2003, p. 226). Segundo, a da separação das funções intelectuais e profissionais dos arquivistas face aos historiadores, com a consequente separação da profissão *from the exclusively practical work of the archive and reorganized as an independent profession through records management approach to archival theory* (Ridener, 2009, p. 98). Entrando na pós-modernidade, como veremos de seguida, a noção de verdade encontra-se questionada, com a crescente disseminação de um pensamento de pós-verdade. Contudo, a separação da componente prática e teórica e a consequente intelectualização da profissão mantêm-se.

As reflexões pós-modernas em arquivística surgiram na década de 1990, em grande medida como resposta ao advento das tecnologias digitais. Em consequência, o contexto digital traduziu-se na incapacidade da teoria arquivística da época em lidar com o conceito de arquivos num contexto em rápida mudança para algo desconhecido e incerto, e com a insuficiência dos métodos arquivísticos tradicionais para a gestão dos recentemente aparecidos documentos digitais. Estes receios, incertezas e dúvidas foram sentidas por arquivistas de diversas as partes do planeta, conforme exposto anteriormente pela diversidade de autores e nacionalidades que escreveram sobre o tema, o que revela tratar-se de um desafio global para a teoria arquivística.

Em essência, o advento da *Web* e do contexto digital na década de 1990, com a produção e comunicação massificada de documentos digitais, culminou numa mudança de paradigma. A noção de mudanças de paradigma foi proposta pelo filósofo Thomas Kuhn (1922-1996) na obra *The Structure of Scientific Revolutions*, publicada pela primeira vez em 1962 como resultado das suas investigações em história da ciência. Kuhn argumenta que, a determinado momento, a ‘ciência normal’ enfrenta fenómenos que não consegue explicar, desencadeando uma crise que obriga a reformulações profundas com a escolha de outro modelo científico, que se torna por sua vez a ‘ciência normal’. Conforme explica o autor, a transição de um paradigma em crise para uma nova tradição científica não é linear:



*Rather, it is a reconstruction of the field from new fundamentals, a reconstruction that changes some of the field's most elementary theoretical generalizations as well as many of its paradigm methods and applications. During the transition period there will be a large but never complete overlap between the problems that can be solved by the old and by the new paradigm. But there will also be a decisive difference in the modes of solution. When the transition is complete, the profession will have changed its view of the field, its methods, and its goals (Kuhn, 1970, p. 85).*

Por conseguinte, não se observa uma ruptura completa, mas antes um período de transição com sobreposição de ambos modelos teóricos, o antigo e o novo. Da mesma forma, os problemas podem ser solucionados por ambos os paradigmas, embora as soluções sejam radicalmente distintas. A Arquivística incorreu num começo de mudança de paradigma na década de 1990, tendo sido iniciada uma transição de um modelo científico de gestão de documentos em papel, entendidos a nível fundamental como realidades físicas, em direção a um modelo científico de gestão digital, em que os documentos de arquivo e os metadados se tornam objetos digitais geridos enquanto entidades lógicas. Se mudanças radicais ocorrem em determinada interpretação quando as respostas dadas pelo atual modelo para as perguntas de investigação se revelam incapazes de explicar toda a complexidade do fenómeno observável, ou quando as metodologias desse modelo deixam de funcionar, no caso da Arquivística o fenómeno observável foi a produção da informação, nomeadamente nos seus agentes de criação, proveniência e materialidade. A ruptura com a metodologia foi a emergência do documento digital e a quebra nas práticas de gestão de material analógico. Terry Cook sumaria de forma bela a mudança paradigmática na Arquivística:

*At the heart of the new paradigm is a shift away from viewing records as static physical objects, and towards understanding them as dynamic virtual concepts; a shift away from looking at records as the passive products of human or administrative activity and towards considering records as active agents themselves in the formation of human and organizational memory; a shift equally away from seeing the context of human and organizational memory; a shift equally away from seeing the context of records creation resting within stable hierarchical organizations to situating records within fluid horizontal networks of work-flow functionality. For archivists, the paradigm shift requires moving away from identifying themselves as passive guardians of an inherited legacy to celebrating their role in actively shaping collective (or social memory). Stated another way, archival theoretical discourse is shifting from product to process, from structure to function, from archives to archiving, from the record to the recording context, from the "natural" residue or passive by-product of administrative activity to the consciously constructed and actively mediated "archivalisation" of social memory (T. Cook, 2001, p. 4).*

A extensão e a profundidade das mudanças são assinaláveis. A noção fundamental é provavelmente a de que o contexto digital não é estável, no sentido em que os objetos digitais estão em permanente mutação, ao serem enviados, explorados, interpretados e facilmente copiados. Assim, a estabilidade e a identidade de um dos documentos em papel sucumbem perante a incerteza

do ambiente digital, cuja facilidade de criação, reprodução e partilha de documentos acarreta a perda de uma realidade — o documento de arquivo — bem delimitada e facilmente manipulável. De produto para processo, do documento para o contexto, de um resíduo passivo para um ativo construído no tempo, eis o cerne das novas formulações científicas no novo paradigma.

Contudo, muitas outras implicações há. A pós-modernidade surge como uma ruptura contra a modernidade, pelo que muitas outras conceções modernas são igualmente dispensadas. As noções de verdade universal ou de conhecimento objetivo decorrentes da interpretação positivista da ciência são dispensadas como quimeras, adquirindo mais importância o contexto do documento do que o documento em si, enquanto realidade objetiva, ou o seu conteúdo (T. Cook, 2001, p. 7). Ao longo dos últimos anos, várias vozes de intelectuais de vários setores da sociedade têm começado a expressar que habitamos numa era da pós-verdade<sup>50</sup> (*post-truth*), em que as noções de verdade e de factualidade se encontram crescentemente questionadas, mesmo contra toda a evidência (Duranti, 2018; Gilliland, 2018; Yeo, 2018).

Existe uma expressão que afirma que contra factos não há argumentos. Atualmente, num entendimento pós-moderno, não há factos, há interpretações. A Figura 11 é um exemplo clássico desta ambiguidade.

---

<sup>50</sup> Em 2016, a palavra *post-truth* foi declarada pelos dicionários de Oxford como a palavra do ano. Gilliland entende a tendência de pós-verdade como *calling for the production of documents, challenging those documents once produced, refusing to accept all customary guarantees and tests of their trustworthiness, and promulgating some imagined alternative documentary reality* (Gilliland, 2018, p. 3). Gilliland toma como contexto o questionamento das certidões de nascimento de Obama, que foram divulgadas, contestadas e desacreditadas mesmo perante toda a evidência enquanto documentos produzidos e autenticados pelas entidades responsáveis e competentes. O congresso, na voz do seu representante John Boehner, manteve que *the American people have the right to think what they want to think. I can't — it's not my job to tell them* (NBC News, 2011, p. 6). Gilliland também refere a eleição de Trump quando este afirmara que a sua eleição fora a maior cerimónia com mais participantes de sempre. Quando confrontado com estatísticas e análises demográficas que demonstravam como outras tinham tido mais participantes, Trump disse que a Casa Branca se baseava em *alternative facts*.



Figura 11 — “My wife and my mother in law”: Uma jovem ou uma idosa?<sup>51</sup>

FONTE: WikiCommons.

Podemos dizer objetivamente que estamos perante uma mulher idosa? Para alguns assim pode parecer, ao passo que outros irão defender que se trata de uma jovem. Ou é possível até observar-se uma jovem e uma idosa, dependendo da intenção e da perspectiva, embora não ao mesmo tempo. O processo de visualização é distinto do de interpretação. Vemos ou uma mulher jovem ou uma idosa, não ambas em simultâneo. Ver é *to have knowledge of some sorts* (Hanson, 1958, p. 20). Logo, nunca vemos objetivamente, mas antes enquanto determinada manifestação. Mortensen salienta que *observation is dependent on our expectations and on the context in which we observe things* (Mortensen, 1999, p. 10). Da mesma forma, também as interpretações das teorias e dos factos são contextualizáveis nas intenções do intérprete. Para a Arquivística, isto significa que:

*Nothing is neutral. Nothing is impartial. Nothing is objective. Everything is shaped, presented, represented, re-presented, symbolized, signified, signed, constructed by the speaker, photographer, writer, for a set purpose. No text is a mere innocent by-product of action ... but rather a consciously constructed product ... Texts are all a form of narration more concerned with building consistency and harmony for the author ... than they are evidence of acts and facts, or juridical or legal frameworks. And there is not one narrative in a series or collection of records, but many narratives, many stories, serving many purposes for many audiences, across time and space* (T. Cook, 2001, p. 7).

---

<sup>51</sup> Ilustração originalmente criada por um ilustrador anónimo em finais do século 19 na Alemanha para um postal. Versões subsequentes foram elaboradas pela cartoonista inglesa William Hill (1887-1962) e reproduzidas na revista norte-americana *Puck* em 1915. Esta figura foi introduzida na discussão científica em Psicologia em 1930 por Edwin Boring (1930).

Nada é neutro, a imparcialidade medieval e moderna do documento de arquivo está desacreditada. Significa isto que conceitos como a autenticidade (*authenticity*), confiabilidade (*trustworthiness*) e fidedignidade (*reliability*) se tornam supérfluos e dispensáveis, uma vez descartados? Não. Acredita-se que um ambiente de constante desacreditação e questionamento dos factos age no sentido de reforçar a necessidade de demonstrar a autenticidade e fidedignidade dos objetos digitais para fortalecer o controlo intelectual e diminuir a probabilidade de dúvida. Mais do que nunca, a representação arquivística deve esforçar-se por manter estes parâmetros sob controlo, o que se manifesta na escolha e gestão de metadados apropriados para representar estas dimensões (autenticidade, confiabilidade, fidedignidade) dos recursos representados.

Além disso, salienta-se a ideia de que não há uma narrativa única veiculada por determinada coleção, mas várias, dependendo da audiência e do contexto. À medida que os intérpretes dos documentos de arquivo se vão sucedendo no espaço e no tempo, numa diversidade geográfica e cronológica, vão emergindo igualmente várias intenções e interpretações a partir das coleções. Um exemplo do reconhecimento desta multiplicidade é o facto de o Conselho Internacional de Arquivos ter designado a sua última norma internacional de descrição arquivística, publicada em versão rascunho em 2016 — com uma versão 0.2 de Dezembro de 2019 —, de *Record(s) in Context(s)*, enquanto cristalização operada no nome de uma interpretação social. O reconhecimento de contextos múltiplos advém também em grande medida da possibilidade de criação de pesquisas personalizáveis e individualizadas em ambiente digital devido à fácil combinação de metadados através de facetas. Este tipo de novas necessidades dos utilizadores, que não se contentam com a descrição autoritária e estática, condensada num instrumento de pesquisa, induz por sua vez uma mudança de paradigma entre descrições analógicas, com base no papel, em direção a descrições no contexto digital. Anne Gilliland dá conta desta transição:

*In recent years there has been increasing criticism that collection-level, hierarchical metadata as exemplified in archival finding aids, while valuable for retaining context and original order, represents an oversimplified view of the actual complexities of records-creation process and provenance, privileges the scholarly user of the archive ... and necessarily perpetuates a paper-based descriptive paradigm. In the online world, multiple descriptive relationships between objects can be supported simultaneously, and some of these ... may support new types of users and uses in an environment that is not mediated by a reference archivist* (Gilliland, 2016, pp. 5-6).

Em causa está o predomínio moderno em representação dos arquivos do geral para o particular, começando pelo nível da coleção, descrevendo fundos, coleções ou séries, em detrimento de metadados ao nível do item. Contudo, o modo como os utilizadores exploram os recursos em ambiente digital e constroem as suas próprias teias de relações acende a questão da importância entre metadados ao nível da coleção ou ao nível do item. Tradicionalmente, os arquivistas têm colocado o seu esforço no coletivo, movidos pelo *respect des fonds*, mas no contexto digital não tem necessariamente de ser assim.

### 2.2.1 O desrespeito pelos fundos

A representação da informação em arquivos, da qual os instrumentos de pesquisa são um produto que reflete os seus princípios, tem desde meados do século 19 assentado no princípio do respeito pelos fundos. Este princípio, *more than any other principle, has come to provide the archival profession with a unique identity* (Bartlett, 1992, p. 108). O princípio do respeito pelos fundos está presente em outros dois princípios, o princípio da proveniência e o princípio da ordem original.

O princípio da proveniência dita que os documentos de arquivo criados, acumulados e usados por uma pessoa ou um grupo no decurso da sua vida ou atividade devem ser mantidos juntos, sem se misturarem com outros documentos de uma entidade distinta (J. Bailey, 2013). A ordem original refere-se ao agrupamento intelectual e ordenação de documentos no contexto da sua acumulação, e mantê-la é essencial não só para a compreensão das relações partilhadas pelos documentos, como também para a constituição de evidência do modo como eram usados (International Council on Archives, 2016). Em conjugação, estes princípios satisfazem a necessidade essencial de *to know where it [o documento de arquivo] was created, in the framework of what process, to what end, for whom, when and how it was received by the addressee, and how it came into our hands* (Duchein, 1983, p. 67).

Pretende-se focar a discussão no conceito de ordem original, entendida como *the organization and sequence of records established by the creator of the records* (Society of American Archivists, 2005, p. 280). Ao nível da representação da informação, a ordem original tem guiado a forma como a descrição arquivística é conduzida, isto é, tem ditado a ordem segundo a qual informação deve ser apresentada aos utilizadores. As complexidades e limitações deste conceito e deste princípio têm sido discutidas (Meehan, 2010; J. Niu, 2015b), com alguns autores a argumentar ou que a ordem original *is to be respected when it is usable* (Boles, 1982, p. 32), ou que, quando esta é caótica, os documentos devem ser reorganizados numa ordem que reflita o seu uso inicial ou conforme a ordem do arquivista (Schellenberg, 1988, p. 190), ou que deve ser mantida, ainda que imperfeita, por ser demasiado dispendioso reorganizar (Holmes, 1964, p. 22), ou ainda que a ordem por vezes nem é original, uma vez que os documentos já foram reorganizados por outras entidades quando dão entrada na custódia da instituição arquivística (MacNeil, 2008, p. 21). Estas discussões têm em comum a premissa da gestão dos documentos de arquivo em papel, enquanto realidades eminentemente físicas.

Este princípio tem sido veiculado pelas normas internacionais de descrição arquivística que, enquanto guias oficiais das melhores práticas para a profissão, atuam como documentos de referência, tanto de práticas como de teorias. Ao longo das várias normas publicadas pelo ICA desde a década de 1990, a partir da qual os arquivistas têm construído vários instrumentos de pesquisa para acesso aos espólios, a descrição arquivística é apresentada com uma estrutura hierárquica, do geral para o particular, desde o englobante fundo até aos itens individuais. A representação da informação assenta num modelo multinível, segundo o qual a informação dada a cada nível de descrição deve corresponder ao nível de detalhe necessário para a sua compreensão,

e é apresentada do geral para o particular, do fundo, como entidade singular e completa, para as respetivas partes. O enfoque é colocado na *person or group that has accumulated a body of records, and often does not reflect the social and material complexity of the origins of records* (International Council on Archives, 2016, p. 5).

Esta forma de apresentação da informação simboliza uma narrativa linear que tem sido associada à forma como os documentos analógicos são organizados no mundo físico, dispostos por pastas, caixas e papéis, o que acaba por permear a forma como são organizados no espaço intelectual da descrição. David Bearman é claro ao afirmar que tal abordagem *is neither necessary nor desirable for electronic records* (Bearman, 1996, p. 196). De forma similar, Bailey também contraria a ideia do respeito pelos fundos enquanto a forma exclusiva de representação e acesso aos materiais de arquivo em ambiente digital:

*In a database, objects are related but not ordered. The database logic is nonlinear and there is no original order because order is dependent upon query. Search and navigation become algorithm-dependent and not predetermined by a creator, an archivist, or a static finding aid — though all of those organizations are themselves available. Digital objects will have an identifier, yes, but where they “rest” in intellectual space is contingent, mutable. The key point is that, even at the level of representation, arrangement is dynamic: access and representation need not depend on the fonds* (J. Bailey, 2013).

A distinção fundamental assenta na natureza dos objetos a representar. Os objetos digitais são entidades eminentemente lógicas<sup>52</sup>, ao passo que os documentos em papel são entidades fatalmente confinadas às limitações da gestão física. Por conseguinte, emerge a distinção de que *physical things can be stored in only one place and in one observable order; logical things can be physically housed in many places but seen to be together* (Bearman, 1996, p. 196). Os objetos digitais, embora armazenados em partes distintas, ou até mesmo representados por entidades custodiantes distintas, podem ser agrupados e reorganizados de diferentes modos, de acordo com as pesquisas dos utilizadores. Recorde-se a pesquisa facetada, que era demasiado complicada quando Ranganathan a formulou em 1933, mas que explodiu em aplicabilidade com o potencial combinatório do ambiente digital. Nas descrições analógicas, as facetas encontravam-se limitadas pela extensão das expressões, pela complexidade do significado dessas mesmas expressões, e pela capacidade de aprendizagem de cada codificador. Num contexto digital, é fácil combinar

---

<sup>52</sup> Kenneth Thibodeau afirma que todos os objetos digitais são entidades físicas, lógicas e conceptuais (Thibodeau, 2002, p. 6). Concorde-se, uma vez que a informação veiculada pelos documentos digitais pode assumir diversas manifestações ou formas de visualização, mas salienta-se que, quando se postulam os objetos digitais como eminentemente lógicos, se está a destacar a componente lógica como característica definidora da não-linearidade de acesso, em oposição aos documentos analógicos, dos quais a dimensão lógica se encontra ausente.

metadados para refinar pesquisas e é simples porque o papel de cada elemento é indicado na sua designação (o elemento ‘Ano’ filtra por ano). Conforme Bearman explica:

*The properties of logical things are associated with them through formal, defined, logical relations while the properties of physical things are associated with them as material objects with concrete locations, attachments and markings (Bearman, 1996, p. 196).*

A natureza de combinação de metadados em ambiente digital é lógica, pelo que não depende da localização física dos documentos de arquivo. É significativamente mais fácil redefinir termos de pesquisa numa base de dados, escolhendo e recombinando os metadados, do que um arquivista deslocar-se à estante onde estão os documentos e reescrever os códigos da classificação, ou mudar todos os documentos de local para acomodar determinada necessidade de organização, e repetir este exercício para cada utilizador. Assim, no ambiente digital, e numa aceção pós-moderna, a descrição tem potencial para se tornar ela própria relacional, indo além da agregação individual de documentos de arquivo. A lógica da organização física já não é imposta na descrição nem nos subsequentes processos de recuperação da informação. Isto não significa que a ordem original ou a representação hierárquica não sejam importantes — são. Acontece simplesmente que atingimos o entendimento de que já não constituem a forma exclusiva de representação da proveniência e acesso. A proveniência deixa de ser vista apenas como uma relação unidimensional entre o criador e os documentos para passar a ser entendida como uma entidade conceptual e como uma rede de relações contextuais. As implicações para a descrição arquivística são as de que:

*When provenance is conceptualised as a multilateral network of contextual relationships, records creator and provenance are no longer concepts that share the same meaning. Instead of identifying one records creator, analysis and identification of relationships between agents in different roles, as well as relationships between functions and information resources, will be in focus (The National Archives of Finland, 2013, p. 10).*

Como tal, a descrição arquivística, caracterizada pela descrição dos recursos informacionais e da sua proveniência, o produtor deixa de ser o principal foco da descrição, privilegiando-se também os contextos das atividades e os contextos dos vários agentes envolvidos no processo de produção e custódia dos recursos. Importa agora analisar em maior profundidade o desenvolvimento do conceito de descrição arquivística.

## **2.3 Conceptualização e objeto da descrição arquivística**

A representação da informação em arquivos é tradicionalmente designada por ‘descrição arquivística’, pelo que se adota esta terminologia de modo a diminuir a divergência semântica. ‘Descrever’ provém do latim *describere*, com o significado de *fazer a descrição* (Academia das Ciências de Lisboa, 2001, p. 1162). Por sua vez, ‘descrição’ tem origem na expressão latina *descriptio, -ōnis*, que se traduz na *ação de retratar com pormenor; enumeração de características* (Academia das Ciências de Lisboa, 2001, p. 1162).

Por conseguinte, quando agregamos o significado de descrição ao âmbito dos arquivos, deparamo-nos com algo como ‘enumeração das características dos arquivos’ ou, mais precisamente, ‘enumeração das características dos documentos de arquivo’.

A descrição arquivística é normalmente apontada como o principal elo entre os utilizadores e os documentos, no sentido em que ela se destina à criação de instrumentos de pesquisa que identifiquem e expliquem o contexto e o conteúdo dos documentos do arquivo, facilitando o acesso aos mesmos (International Council on Archives, 2012, p. 3). No entanto, Dunia Llanes-Padrón alerta para o facto de este conceito ser frequentemente mal entendido por ainda se julgar que se trata apenas de uma operação com vista a facilitar o acesso aos documentos (Llanes-Padrón, 2011, p. 37). Durante várias décadas, ao longo da segunda metade do século 20, a descrição confundia-se com a produção de instrumentos de pesquisa. Na sua tese de doutoramento, Llanes-Padrón (2011) analisou a evolução do conceito de descrição entre 1960 e 1990, tendo por base a literatura científica e as propostas de prestigiadas instituições arquivísticas como o ICA e a *Society of American Archivists* (SAA). De entre as várias definições tomadas em consideração, concluiu que a todas presidia uma finalidade comum de controlar os fundos documentais, alimentada pela necessidade de proporcionar um acesso mais aberto. Para isso, é fundamental possuir um conhecimento mais generalizado dos fundos de modo a localizar e recuperar a informação, devendo a descrição cumprir a função básica de facilitar a consulta dos documentos. Como resultado, criou-se a tendência para identificar a descrição com o instrumento resultante da atividade descritiva (instrumentos de pesquisa ou de consulta), recebendo várias designações (guias, inventários, catálogos, índices) consoante o nível hierárquico a descrever (Llanes-Padrón, 2011 pp. 37-40).

A partir dos anos 90 do século passado, notam-se dois fenómenos importantes que contribuem para um entendimento mais aprofundado do conceito de descrição: o começo do processo de normalização internacional e as primeiras reflexões pós-modernas. O paradigma pós-moderno define a descrição arquivística como representação, relevando a sua dimensão de um constructo humano face à visão tradicional que entendia a acumulação dos documentos de arquivo como um processo natural, entendimento segundo o qual o material arquivístico constituía o resultado objetivo das atividades e os arquivistas os guardiães imparciais desse material. Porém, conforme explicado no modelo finlandês para a descrição arquivística:

*When archival description is defined as representation, description is not understood as neutral entering of objective facts that describe information resources. Instead, archival description is recognised as an interpretation, which is always produced from a particular point of world view. As a representation, description is understood as a continuum in time, a process which is continuously complemented by agents that produce and manage information resources, by professionals who are responsible for management systems and by the users of information resources* (The National Archives of Finland, 2013).

Para o desenvolvimento do processo de normalização internacional das descrições arquivísticas contribui de forma sobeja a atividade do ICA, publicando várias normas internacionais a partir de 1994. A publicação das normas tinha como objetivos: a) *ensure the creation of*



*consistent, appropriate and self-explanatory descriptions; b) facilitate the retrieval and exchange of information about archival material; c) enable the sharing of authority data; and d) make possible the integration of descriptions from different repositories into a unified information system* (International Council on Archives, 1992, p. 13). Em resumo, trata-se da simplificação de métodos de trabalho, de forma a economizar tempo, garantir melhores resultados no processo de descrição, e facilitar a interoperabilidade entre múltiplas instituições.

Porém, convém fazer uma nota para o facto de que o ICA, com a publicação das várias normas, pretendia estabelecer uma estrutura de dados uniforme de descrição a nível internacional, e não uma estrutura de conteúdos. Assim, compete às comunidades arquivísticas nacionais a elaboração das suas próprias normas de descrição para regular os conteúdos a descrever, consoante a realidade arquivística de cada uma (Llanes Padrón & Fujita, 2017, p. 213). Reino Unido, Estados Unidos da América, Canadá, Portugal, Irlanda, Brasil, Uruguai e Espanha são alguns dos países que desenvolveram normas de descrição arquivísticas nacionais.

A par da publicação das primeiras normas internacionais, as reflexões pós-modernas na década de 1990 também figuram como um estímulo para os arquivistas reposicionarem a sua forma de pensar e trabalhar. Para esta mudança contribuíram ainda o desenvolvimento económico e social dos anos de 1980, a expansão das tecnologias de informação e de comunicação (TIC) e a diversificação dos perfis de utilizadores, com destaque sobretudo para o desenvolvimento da informática. Segundo Tarraubella, a implementação das TIC levou ao aparecimento de um novo conceito de utilizador, que não acede pessoalmente aos arquivos para consultar a documentação, mas que o faz desde casa ou do local de trabalho através do seu computador ou de outro dispositivo (Tarraubella, 1998, p. 201).

A não-presencialidade do acesso exige novas estratégias aos arquivos para o acesso e difusão da informação. Neste sentido, Navarro e Molins defendem que *era necesario contar con unos principios y unas reglas universalmente aceptadas que indicaran cómo describir, qué datos identificativos se debían extraer y cómo estos se debían presentar y organizar* (Navarro, 1998, p. 9). Este apelo fundamenta a necessidade de normalização da descrição arquivística. Assim, o início da normalização acarreta uma reconsideração dos fundamentos da descrição, passando a ser colocada mais ênfase na atividade (representação dos fundos de arquivo) do que no resultado final (criação de instrumentos de descrição, busca, consulta ou investigação).

A mudança do foco da descrição é observável em várias definições começadas a surgir inclusive na segunda metade dos anos 80. Em 1986, o *Bureau Canadien des Archivistes* foi um dos primeiros organismos que começaram a redefinir este conceito, apontando a descrição como a representação mais exata e mais concisa possível de qualquer unidade arquivística, seguindo o princípio da proveniência e da ordem original (Bureau Canadien des Archivistes, 1986, p. 2).

Na mesma linha, em 1989, na segunda edição do *Manual of Archival Description*, Michael Cook e Margaret Procter apontam que a descrição consiste na representação dos documentos utilizando descritores e que, na prática, estes descritores se materializam em sistemas ou instrumentos de pesquisa (Cook & Procter, 1989, p. 6).

A partir de estas propostas e de outros estudos relacionados com normalização da descrição arquivística, o ICA delimitou, numa primeira versão em 1994, a definição de descrição enquanto *the creation of an accurate representation of a unit of description and its component parts, if any, by the process of capturing, collating, analyzing, and organizing any information that serves to identify archival material and explain the context and records systems which produced it* (International Council on Archives, 1994, Glossário).

Numa segunda versão da ISAD(G), publicada em 2000, o ICA estabelece que a descrição consiste na:

*creation of an accurate representation of a unit of description and its component parts, if any, by capturing, analyzing, organizing and recording information that serves to identify, manage, locate and explain archival materials and the context and records systems which produced it. This term also describes the products of the process* (International Council on Archives, 2000, p. 10).

Como se pode observar, apesar de ambas as definições avançadas pelo ICA nas duas versões da ISAD(G) serem quase confundíveis, apresentam algumas diferenças que se considera importante destacar. É interessante notar a ênfase que a segunda edição coloca na identificação, na gestão, na localização e na explicação não só do contexto de produção dos documentos de arquivo, como também dos próprios documentos, num maior destaque da entidade produtora. Além disso, na segunda edição é também acrescentada uma nota que alerta para a possibilidade de se obterem vários instrumentos de consulta como resultado da atividade de descrição.

Llanes-Padrón defende que este novo enfoque na descrição, entendida como a representação de conteúdos com vista à obtenção de diversos instrumentos, constitui uma ampliação do objeto da descrição. Na opinião desta autora, a evolução tecnológica permitiu dissociar o conceito de descrição do conceito de instrumento de pesquisa. A descrição consiste em elaborar uma representação (atividade) que assume diferentes formas de manifestação, enquanto produtos (instrumentos de pesquisa) (Llanes-Padrón, 2011, p. 43).

Esta posição é reforçada pelo próprio ICA quando destaca que a explicação do contexto e do conteúdo dos documentos de arquivo é alcançada *by creating accurate and appropriate representations and by organizing them in accordance with predetermined models* (International Council on Archives, 2000, p. 7). Assim, a partir de uma base de um sistema de representação — uma base de dados — é possível obter diferentes formas de pesquisa dos conteúdos — instrumentos de pesquisa. Em resumo:

*La descripción en los archivos no se debe realizar con la finalidad de crear un determinado instrumento de búsqueda o de consulta. Su objetivo debe estar orientado a representar la información de forma coherente para establecer un sistema general de información archivística que permita representar la información en varios formatos (varios instrumentos de descripción) e integrar datos de varios archivos para facilitar su intercambio* (Llanes-Padrón, 2011, pp. 45-46).

Por fim, devemos destacar que a atividade de descrição se trata de uma prática constante, nunca definitiva (Yakel, 2003). Ketelaar argumenta que o significado dos documentos de arquivo varia com o passar do tempo, consoante o uso que lhes é dado, razão pela qual os arquivistas devem não só reorganizar coleções e visitar descrições insuficientes, como também efetuar novas re-descrições das coleções de forma a refletir os novos significados (Ketelaar, 2001, p. 139).

Muitos arquivos têm já tremenda dificuldade em proceder ao tratamento do volume documental de que dispõem, pelo que exigir-lhes que ainda revisitem antigas descrições parece uma utopia inoportuna. No entanto, em ambiente digital, a necessidade de permanentemente atualizar as descrições parece ser um aspeto incontornável uma vez que metadados são gerados no momento da criação das representações e vão sendo acrescentados ao longo de todo o ciclo de vida (Duff, 1995; Hedstrom, 1993). Neste sentido, *archivists should begin to think less in terms of a single, definitive, static arrangement and description process, but rather in terms of continuous, relative, fluid arrangements and descriptions as on-going representation processes* (Yakel, 2003, p. 4).

Neste enquadramento, a descrição arquivística não se entende como uma atividade singular isolada no tempo, perfilando-se antes como um procedimento contínuo que deve ser revisitado à medida que os documentos de arquivo vão adquirindo novos significados e contextos no decorrer do seu uso, devendo essas novas perspectivas ser constantemente documentadas através da atualização ou acréscimo de metadados. Caso contrário, as descrições tornar-se-iam obsoletas, perdendo rigor na representação da identidade e da integridade dos documentos, e comprometendo a presunção da autenticidade. Sendo esta uma questão central na missão dos arquivistas, deve ser discutido o papel das descrições na garantia da autenticidade tanto dos documentos, como dos metadados que os representam. Porém, importa primeiro debater o grau de controlo intelectual em arquivos recomendável para a representação de recursos em ambiente digital.

### **2.3.1 Controlo intelectual em arquivos**

O controlo intelectual em arquivos pode ser entendido como *the creation of tools such as catalogs, finding aids, or other guides that enable researchers to locate relevant materials relevant to their interests* (Society of American Archivists, 2005, p. 210), desmembrando-se em controlo coletivo (*aggregate control*) e controlo de itens (*item-level control*). Contudo, esta definição parece confundir atividade com produto que, como visto quando discutida a descrição arquivística, são dimensões distintas. Da mesma forma, o controlo intelectual pode ser visto como uma atividade (a representação da informação) relacionada mas independente da criação de vários produtos (instrumentos de pesquisa).

David Bearman salienta que o controlo intelectual *involves two complementary activities: arrangement and description, which take place as steps in the life cycle of records following their accessioning by the archives and preceding their being made available for use* (Bearman, 1989, Chapter III). Focando a análise na descrição, por ser este o âmbito de este trabalho, há duas formas de controlo intelectual ao nível da descrição predominante nos arquivos, o controlo coletivo e o

controlo de itens. O *aggregate control*, ou *collective control* (Bearman, 1996, p. 204), consiste em descrever os documentos de arquivo de forma coletiva, ao nível do fundo, coleção ou séries com a missão de fornecer contexto e de facilitar a pesquisa. Por sua vez, *item-level control* prende-se com a descrição de itens individuais, tais como documentos de arquivo, e apoia a pesquisa e a descoberta dos recursos a título individual (J. Niu, 2015a).

O controlo intelectual dos arquivos ao nível coletivo está de tal forma entranhado no *ethos* dos arquivistas que se afirma como uma das práticas que ditam a identidade da profissão. Conforme declara Jinfang Niu, *this practice has been so entrenched that some archivists believe that aggregate control is unique to archives management, and, without it, the archives management field will no longer exist* (J. Niu, 2015, p. 186). Mark Greene espelha esta ideologia quando afirma que *a relentless focus on the aggregate part is what sets us apart from librarians and museum curators* (M. Greene, 2009, p. 24), considerando assim que o enfoque na descrição coletiva é o traço distintivo entre arquivos, museus e bibliotecas.

Por este motivo, Jane Zhang observa que elevar a descrição de itens a um estatuto de forma de controlo primária de acesso a documentos e instrumentos, relegando o instrumento de pesquisa para segundo plano é um ato que *most archivists would consider unethical because it would turn their collections less or no longer archival* (Zhang, 2012a, p. 336). Como tal, *there would be no archival profession without an appropriate control of archival context* (Zhang, 2012a, p. 338), notando-se assim uma identificação do controlo coletivo com a identidade da profissão.

Todavia, no ambiente digital nota-se uma crescente necessidade de criação de metadados ao nível do item de forma a promover a integração dos recursos com outras plataformas e de permitir aos utilizadores construir a sua própria pesquisa, através de um varrimento transversal dos metadados disponibilizados por diversas bases de dados. Esta abordagem, materializada em projetos como o da Europeia, permite aumentar a visibilidade dos próprios arquivos e do seu património, uma vez que os seus recursos podem aparecer em mais pesquisas do que as tradicionais buscas nos instrumentos de pesquisa divulgados pelas instituições. Bearman argumenta inclusive que *item level information is fundamentally more valuable because it can generate more valid collective level data in addition to serving the needs of item documentation* (Bearman, 1996, p. 205), permitindo também maiores restrições de acesso a documentos com determinado grau de confidencialidade.

Contudo, se uma parte da identidade da profissão residia no controlo coletivo, e se se pede cada vez mais um controlo mais detalhado ao nível dos recursos, então poder-se-á afirmar que essa parte da identidade da profissão está em risco? A resposta imediata é a que sim, no sentido clássico de identidade da profissão. Contudo, isto não significa que a noção de identidade da profissão, em si, desapareça — apenas se vai reorganizar dentro de outra identidade que há de emergir. A respeito da ‘antiga’ identidade, esta mudança reforça o argumento de que cada vez mais é necessário pensar menos em termos daquilo que nos separa (arquivos *versus* bibliotecas *versus* museus) mas antes naquilo que nos aproxima (representação e gestão da informação). É de esperar que o arquivista assuma cada vez mais uma função de gestor de metadados, e nisso dilua um pouco a sua formação

com áreas adjacentes como a biblioteconomia e a informática. Pode argumentar-se que a identidade reside também não na abordagem ao controlo intelectual, mas na natureza dos documentos sujeitos a esse controlo. Isto é, um documento de arquivo resulta de uma atividade, da qual constitui evidência e prova, ou seja, de uma atividade administrativa, ao contrário do material das bibliotecas e museus que resulta do ócio (livros) ou do engenho (ferramentas, arte). O tempo dirá se esta distinção se irá manter, ou qual a robustez da sua fronteira. Torna-se cada vez mais difícil separar claramente estas abordagens. Importa reter para esta discussão a necessidade crescente de representação/controlo não só do coletivo, mas também de itens da coleção.

## 2.4 Autenticidade e fidedignidade de e nas representações arquivísticas

A importância da autenticidade para a profissão arquivística é comprovada pela sua presença como um dos propósitos fundamentais do código de ética dos arquivistas: garantir a autenticidade (*authenticity*) dos documentos de arquivo sob custódia (International Council on Archives, 1996; Society of American Archivists, 2012).

A autenticidade refere-se à *trustworthiness of the record as a record, i.e., the quality of a record that is what it purports to be and is free from tampering or corruption* (InterPARES 2, 2008, p. 8). Para tal, é necessário estabelecer a identidade e demonstrar a integridade ao longo do tempo. A identidade é entendida como *the whole of the characteristics of a document or a record that uniquely identify it and distinguish it from any other document and record*, ao passo que a integridade se refere à *quality of being complete and unaltered in all essential aspects* (InterPARES 2, 2008, pp. 25-26). Questionar a identidade é questionar *was it written by who purports to have written it?*, ao passo que a integridade resume-se à questão *has it been altered in any way since it was first created and, if so, has such alteration changed its essential character?* (Macneil, 2000, p. 53).

O *International Council on Archives* identifica ainda a fidedignidade (*reliability*) e a usabilidade (*usability*) como duas outras características essenciais dos documentos de arquivo. A fidedignidade significa que *a record can be trusted as a full and accurate representation of the transaction(s) to which they attest, and can be depended on in the course of subsequent transactions*, ao passo que a usabilidade pressupõe que este pode ser localizado, recuperado, preservado e interpretado (International Council on Archives, 2008, p. 13).

O interesse em demonstrar que determinado documento de arquivo é autêntico e fidedigno prende-se com verificar a sua confiabilidade (*trustworthiness*) enquanto testemunho das atividades que materializa. A confiabilidade é *the quality of being dependable and reliable* (Society of American Archivists, 2005, p. 388), sendo geralmente definida em relação à fidedignidade e à autenticidade. Como refere Mak, *the purpose of establishing the authenticity of a record is to position it as trustworthy for a particular purpose, and authoritative within a particular framework* (Mak, 2015, p. 122).

No período medieval, nos documentos de arquivo, a confiabilidade coincidia com a autenticidade, uma vez que os altamente controlados processos de criação, manutenção e preservação dos documentos tornava praticamente impossível a sua reprodução, com todos os elementos formais, em locais que não as chancelarias ou os notários (Macneil, 1998, p. 30). Se a identidade dos documentos de arquivo fosse determinada, então presumia-se a sua integridade. No entanto, no caso dos documentos digitais, a identidade e a integridade já não estão diretamente ligadas (Rogers, 2015, p. 21), uma vez que os processos de criação, manutenção e preservação dos documentos já não são altamente controlados como antigamente. Antes, a sua integridade é constantemente colocada em causa pela mutabilidade do ambiente digital no qual os documentos de arquivo são constantemente construídos e reconstruídos.

O contexto tecnológico atual não é o único obstáculo à determinação da confiabilidade dos documentos de arquivo. O próprio contexto social pós-moderno da pós-verdade, caracterizado pela permanente dúvida dos factos e descrença das fontes de informação, subverte muitas tentativas de afirmação de autoridade. Poderemos ainda falar em autenticidade e confiabilidade numa era pós-moderna de crescente questionamento da verdade e da factualidade (Yeo, 2018, p. 177)? Acredita-se que sim. Este constante questionamento apenas vem reforçar a necessidade de um maior controlo sobre os processos de criação e transmissão dos documentos de arquivo e das suas representações, através do controlo intelectual. Se há mais dúvida, torna-se necessária mais transparência. Contudo, isto implica também um reposicionamento por parte dos arquivistas, que devem abandonar a noção medieval de guardiões da verdade, para reconhecerem que nenhum documento revela a verdade, ou a totalidade dos factos, mas apenas uma fração e uma perspetiva. Conforme afirma Terry Eastwood:

*Records themselves do not reveal either all the facts of any matter or speak its truth. We know enough about the dangers of trying to separate fact from value or interpretation to be skeptical that records can render an account of past actions in some straightforward way, but the archival faith is ... a faith in the power of rational discourse resting on evidence of past actions* (Terry Eastwood, 2010, p. 159).

Não podemos deixar de falar em autenticidade e confiabilidade, pois *the road where nothing is trusted ... leads inexorably to chaos* (T. Cook, 1995, p. 80). Pode assumir-se que estes conceitos deixam de ser de uma autoridade absoluta, mas não se pode deixar cair em nome da responsabilização social e institucional e da transparência. Os documentos de arquivo possuem um *crucial fiduciary role* (Moss & Thomas, 2017, p. 62) na medida em que constituem evidência com base na qual governos, instituições e indivíduos podem ser responsabilizados. Assim, conforme destaca Geoffrey Yeo, *in organizational settings, reliance on records underpins the conduct of business and fortifies corporate memory; in the wider society ... it would seem essential that records held in archival institutions are deemed reliable and trustworthy* (Yeo, 2018, p. 177).

Porém, a forma como a autenticidade é demonstrada no contexto digital é significativamente distinta dos procedimentos analógicos. De forma a explorar em maior detalhe esta questão, o projeto InterPARES 1 foi desenvolvido entre 1991 e 2001 pela University of British

Columbia com o objetivo de definir um quadro teórico e metodológico para a preservação autêntica de documentos de arquivo digitais, tendo concluído que:

*the best method of ensuring ongoing authenticity of electronic records is external to the records themselves and involves a tight control on record-making and record keeping procedures and on the flow of metadata into the record's formal elements, rather than digital authentication measures, which have been shown to hamper long-term preservation of authentic records* (Duranti & Blanchette, 2004, p. 4).

De forma convergente, o projeto de Pittsburgh, desenvolvido na Universidade de Pittsburgh entre 1993 e 1996 com objetivos semelhantes, embora com uma metodologia distinta<sup>53</sup>, chegou à mesma conclusão de que *metadata would need to be attached to, and kept with, records over time, and that this metadata would document the content, structure, and context of the records* (Bearman, 2006, p. 18). Duas ilações podem ser retiradas destes projetos: a primeira, a que o movimento da pós-verdade apenas realça a necessidade de intensificar os procedimentos de controlo dos documentos de arquivo de forma a fazer prevalecer a sua confiabilidade; segundo, a de que de esse controlo é feito através de metadados, que por natureza são externos aos documentos. Assim, na gestão de objetos digitais, os elementos sobre os quais baseamos a presunção da autenticidade *are observable not in the document itself but rather in the procedures* de criação, manutenção e preservação dos documentos de arquivo (Terry Eastwood, 1994, p. 127) ou, por outras palavras, metadados.

Sendo as descrições arquivísticas, na prática, um esquema de metadados, elas assumem-se como um dos procedimentos necessários para atestar a autenticidade dos documentos de arquivo digitais (InterPARES 1, 2002) na medida em que clarificam a classificação (*arrangement*) dos documentos e as circunstâncias (contexto) nas quais foram criados (International Council on Archives, 2016; MacNeil, 2005).

A ligação entre a autenticidade e as descrições arquivísticas é visível nas normas de descrição. Por exemplo, a revisão de 2008 das canadianas *Rules for Archival Description* declara que uma das finalidades da descrição arquivística é *to establish grounds for presuming the authenticity of archival material by documenting its chain of custody, arrangement, and circumstances of creation and use* (Canadian Council of Archives, 2008, p. xxii).

Bonnie Mak indicia uma abordagem semelhante quando refere que a integridade é inferida *through the appearance of the record and an unbroken chain of custody* (Mak, 2015, p. 120). O projeto INTERPARES constrói sobre esta ideia ao referir que os documentos digitais carecem de

---

<sup>53</sup> O projeto Pittsburgh adotou uma abordagem *bottom-up* que não partiu de qualquer assunção prévia da tradição arquivística, de modo a derivar uma resposta totalmente pragmática sobre como melhor abordar a gestão documental. Por sua vez, o projeto InterPARES executou uma abordagem *top-down*, baseada nos princípios da tradição arquivística e nas práticas de gestão documental observadas em diversos países ocidentais.

uma *unbroken chain of preservation* que pode envolver um programa ativo e intencional de refrescamento, migração e cópia para se manterem acessíveis e interpretáveis (Duranti, 2005, p. 293), o que se identifica em suma com estratégias de preservação digital. Todavia, Bonnie Mak alerta para o problema da autoridade de quem executa estas estratégias e da consequente necessidade de documentação dessas intervenções. Nas suas palavras, *because these practices are frequently undertaken by agents who may not be invested with public trust, supporting documentation of each transformation is required to aid the determination of the authenticity of the records in their care* (Mak, 2015, p. 120). O grau de completude desta documentação acompanhante, que se manifesta nos metadados adjacentes a cada estratégia de preservação, é colocado sob escrutínio enquanto critério essencial para determinar se essas intervenções foram legítimas, e se o documento de arquivo pode continuar a ser considerado autêntico e confiável. Heather MacNeil atesta esta ideia quando afirma que *verification of the authenticity of electronic copies of authentic electronic records depends on the accuracy of the documentation of the reproduction process, and on the preservation of the documentary and administrative context of the record themselves* (MacNeil, 2000, p. 72).

Importa ainda destacar que a presunção da autenticidade não se trata de uma questão de ser ou não ser, de uma equação binária em que determinado documento de arquivo é autêntico ou não é (Factor et al., 2009). Conforme destaca Will Suvak, *the level of confidence one can have in authentication is related to the strength of the evidence, and should not be considered absolute* (Suvak, 2015, p. 117). Ao invés, observa-se uma gradação da autenticidade (Carta, 2017, p. 195) como se se tratasse de um espectro em que num extremo temos ‘mais autêntico’ e noutra ‘menos autêntico’, sendo o meio habitado por uma série de nuances de acordo com as informações disponíveis que nos permitem inferir se determinado recurso e os seus metadados acompanhantes apresentam um indício mais ou menos forte de serem autênticos. Tal como a autenticidade, também a *reliability is not a characteristic open to scientific measurement* (Yeo, 2018, p. 177), apresentando antes uma dimensão subjetiva e pessoal pois *what makes a business record reliable is, thus, incomplete and biased towards a particular perspective* (MacNeil, 2011, p. 186). Will Suvak aponta ainda para o papel que a comunidade desempenha nesta avaliação:

*the determination of authenticity is specific to and delimited by the needs of a community, and therefore should not be taken as a characteristic that is universally recognized, or one that inheres in the document itself. Rather, what constitutes authenticity in any given context is negotiable, and will change over time* (Mak, 2015, p. 122).

Neste sentido, não há um critério absoluto a partir do qual possa ser determinado o que constitui um documento de arquivo autêntico. Não só este depende da robustez da evidência apresentada, como também das próprias expectativas da comunidade sobre os parâmetros que pretende acompanhar, e das necessidades que pretende ver satisfeitas. Por conseguinte, enquanto o resultado de uma decisão informada, a presunção de autenticidade nunca pode ser absoluta. Trata-se de uma escala de valores relativa segundo a qual a presença ou ausência de determinados elementos nos inclina para afirmar que determinado recurso é ‘provavelmente autêntico’ ou



‘provavelmente não é autêntico’. Os parâmetros nos quais uma comunidade se baseia para verificar a autenticidade não só são subjetivos mas, porque subjetivos, são variáveis no tempo, transformando-se à medida que a própria comunidade na qual se inserem também se transforma, reformula e reposiciona em torno de novos valores e de novas expectativas.

#### 2.4.1 Do vínculo arquivístico para o vínculo informacional

O papel das descrições arquivísticas na asserção de autenticidade também se manifesta na medida em que manifestam o vínculo arquivístico (*archival bond*). O conceito de vínculo arquivístico nasce do entendimento dos arquivos como sendo constituídos por duas componentes: *the records and their interrelationships* (Lodolini, 1998, p. 186).

O termo ‘vínculo arquivístico’ (no original italiano, *vincolo archivistico*) foi sugerido por Giorgio Cencetti (1939) como forma de expressão da interligação que caracteriza todos os documentos de arquivo (Jenkinson, 1937, p. 37), e que Luciana Duranti considera ser o *key archival concept and foundation of the concept of archives* (Duranti, 2015, p. 28). Este conceito foi posteriormente desenvolvido por Luciana Duranti, Terry Cook e Heather MacNeil como parte do projeto de investigação da University of British Columbia em documentos eletrônicos. Partindo de conceitos extraídos da diplomática<sup>54</sup> e da ciência arquivística, procurando conciliar ambos, um dos objetivos do projeto era *to establish what a record is in principle and how it can be recognized in an electronic environment* (Duranti & Eastwood, 1995, p. 4). De modo a responder a este desafio, partiram da perspectiva do criador para determinar quando um documento é criado e se varia no tempo.

A resposta encontrada foi a emancipação do vínculo arquivístico, que Luciana Duranti caracteriza em detalhe enquanto:

*The network of relationships that each record has with the records belonging in the same aggregation. The archival bond is originary, because it comes into existence when a record is created (i.e., when, after being made or received, it is set aside in the fonds of the physical or juridical person who made or received it for action or reference), necessary, because it exists for every record (i.e., a document can be considered a record only if it acquires an archival bond), and determined, because it is qualified by the function of the record in the documentary aggregation in which it belongs. The archival bond first arises when a record is set aside and thereby connected to another in the course of action, but it is incremental, because, as the connective tissue that joins a record to those surrounding it, it is in continuing formation and growth until the aggregation in which the record belongs is no*

---

<sup>54</sup> *Diplomatics is a science originally developed to prove the reliability and authenticity of records. Over the centuries, it has evolved into a very sophisticated system of ideas about the nature of records, their genesis and composition, their relationships with the actions and persons connected to them, and with their organizational, social, and legal context* (Duranti, 1997, p. 213).

*longer subject to expansion, that is, until the activity producing such aggregation is completed (Duranti, 1997, pp. 215-216).*

Desta passagem elucidativa depreende-se que o vínculo arquivístico é originário, necessário, determinado e incremental, até ao ponto em que a atividade do conjunto documental da qual decorre é dada por concluída. A ideia de conjunto documental como resultado de uma atividade parece pressupor a conceção de fundo como unidade coletiva de agregação de documentos de arquivo. Neste caso, o vínculo arquivístico é incremental no sentido em que cada documento adicional produzido por essa entidade no âmbito das suas funções acrescenta ao conjunto documental que constitui o seu fundo, e expande o significado dos documentos produzidos anteriormente. É determinado porque pertence diretamente ao fundo do qual resulta, e é finito porque, extinta a entidade produtora (por exemplo, falência de uma empresa), é dada por concluída a atividade.

No entanto, embora estas noções funcionem de forma relativamente estável em sistemas analógicos, em sistemas digitais os documentos digitais podem pertencer a vários agregados consoante os sistemas de representação, ou até mesmo não pertencer a agregações. O vínculo arquivístico *determines the identity of each record, because the relationships that a record has with other records contribute to its meaning as much as its content and form do* pelo que, neste sentido, *two identical documents existing in different records aggregations are two different records* (Duranti, 2015, p. 28). Duranti defende que um documento (*document*) apenas se torna um documento de arquivo (*record*) quando colocado em relação com outros documentos de arquivo, pelo que o vínculo arquivístico *is what transforms a document into a record* (Duranti, 1997, p. 216).

Um documento de arquivo não existe num vácuo, na medida em que resulta de uma atividade humana localizada no tempo. Por conseguinte, o seu significado decorre da interpretação do contexto que o rodeia, revelado pelo vínculo arquivístico que estabelece com outros documentos, que por sua vez pode ser manifestado pela descrição arquivística. Nesta linha de pensamento, Duranti considera que *in the digital environment, archival description constitutes the only possible collective authentication of an archives* (Duranti, 2015, p. 28). Contudo, o agregado documental não tem necessariamente de se limitar ao conceito de fundo, que dita uma ordem física de organização que agrupa fisicamente os documentos pertencentes ao mesmo fundo. No ambiente digital, *one document may have several classification or identification codes associated with it as metadata: this means that such a document is as many different records as are the codes linked to it* (Duranti, 2015, p. 28).

Isto leva-nos a propor o conceito de vínculo informacional (*informational bond*) como ampliação do conceito de vínculo arquivístico. Emprega-se o vocábulo ‘informacional’ de forma deliberada, porque o vínculo estabelece-se entre recursos já representados por metadados, ou seja, por dados dotados de contexto e de estrutura. Porque possuem estrutura, os metadados são apreendidos por um utilizador enquanto informação, estabelecendo assim a possibilidade de criar relações entre os recursos. O vínculo informacional diz respeito ao uso de metadados e ao modo

como estes documentam as relações ‘causa-efeito’ (Duranti, 1997) ou propriedades (autoria, matrimônio, título) que determinado recurso estabelece com qualquer outro no âmbito de um mesmo sistema de representação. Duranti radica as relações do vínculo arquivístico na Diplomática, que parte da premissa de que o contexto de criação dos documentos de arquivo é manifestado em elementos internos e externos, tais como atos, pessoas, procedimentos e a forma documental (Duranti, 1997, p. 215). Entende-se a forma documental como indicando o contexto interno na medida em que versa sobre a estrutura interna dos documentos, ao passo que o contexto externo é revelado pelas relações que um documento de arquivo estabelece com outros, manifestadas pelo vínculo arquivístico.

Se o contexto interno manifestado pelo vínculo arquivístico é a forma documental, o contexto interno do *informational bond* é o *metadata record*, ou a estrutura de metadados usados para representar um documento de arquivo. Se o contexto externo no âmbito do *archival bond* são os outros documentos de arquivo decorrentes da mesma atividade com os quais apresenta relação, o contexto externo revelado pelo *informational bond* compreende qualquer recurso que estabeleça qualquer relação com o documento de arquivo, seja ele documentos de arquivo ou qualquer outro recurso incluído no sistema de representação. Esta definição apresenta, assim, um maior alcance.

Neste sentido, o vínculo informacional é entendido como a agregação a que cada documento de arquivo pertence num dado momento, reconhecendo que pode pertencer a mais do que uma agregação em simultâneo, acrescida das relações que estabelece com outros recursos, inclusive com recursos externos à agregação de que é originário. A primeira parte desta formulação corresponde em parte ao vínculo arquivístico no sentido em que a agregação resulta de uma atividade, e constitui um conjunto arquivístico. Porém, novas e imprevistas agregações podem também ser constituídas no contexto de uma pesquisa particular por um utilizador, de forma que estas agregações não correspondem necessariamente ao conjunto documental de um fundo. A segunda parte prende-se com a componente informacional deste vínculo, pois qualquer recurso informacional pode estabelecer um novo elo com outros recursos informacionais através de quaisquer propriedades.

A título de exemplo, a criação de um quadro alusivo à ‘Mensagem’ de Fernando Pessoa contribui para a expansão da teia de relações emergentes a partir dessa obra, ao mesmo tempo que magnifica o seu significado porque recontextualiza a sua interpretação. O vínculo informacional, tal como o vínculo arquivístico, é incremental, no sentido em que vão sendo acrescentados documentos de arquivo/recursos que reformulam o seu significado colocando-o em novos contextos interpretativos. Porém, ao contrário do primeiro, que terminada a atividade que produz os documentos, o vínculo informacional é potencialmente infinito dado que a qualquer momento podem surgir novos recursos que se relacionam com os objetos que constituem a agregação inicial. Além disso, através de pesquisas facetadas, um mesmo objeto digital pode ser recuperado por diferentes contextos, com diferentes finalidades e diferentes interpretações por partes dos utilizadores. Como resultado, ao passo que o *archival bond* é determinado, o *informational bond* é indeterminado.

Argumenta-se que este conceito é igualmente central para a demonstração da autenticidade dos documentos de arquivo. Uma vez que a autenticidade compreende a identidade e a integridade, foi já demonstrado como o vínculo informacional contribui para a identidade de cada documento de arquivo, dotando-o de significado. Em respeito à integridade, *the archival bond is also evidence of the integrity of the archives itself, or lack thereof* porque faz com que cada documento de arquivo seja *equally important to the existence of the whole, and the elimination of any record hurts the integrity of archives* (Duranti, 2015, p. 28). Se um documento se torna um documento de arquivo pelas relações que estabelece com outros documentos de arquivo, adquirindo sentido e identidade, é natural que esse sentido seja alterado se alterados/eliminados os documentos de arquivo com os quais se relaciona. Por conseguinte, identidade e integridade são conceitos profundamente entrelaçados, e o vínculo arquivístico é a estrutura que os une. Uma vez que o vínculo arquivístico é manifestado pela representação arquivística, infere-se que a representação arquivística é um dos mais importantes procedimentos para a asserção da autenticidade. Duranti apresenta uma conclusão semelhante quando refere que:

*Archival description, as the means of elucidating the nature of the archival bond in its documentary context, has been traditionally considered the primary way of perpetuating and authenticating the meaning of records and ... is probably the best method for ensuring their authenticity in the long term* (Duranti, 1997, p. 217).

Uma papel semelhante para a descrição arquivística é defendido por Michael Cook (1986, p. 79), sendo uma posição igualmente reforçada por Anne Gilliland quando afirma que *elucidating and preserving context is what assists with identifying and preserving the evidential value of records and artifacts in and over time; it is what facilitates the authentication of those objects, and it is what assists researchers with their analysis and interpretation* (Gilliland, 2016, p. 4).

Em resumo, o informational bond assenta na rede de relações que cada recurso estabelece com outros recursos, independentemente do tipo de relação. É *originário* porque assume existência quando um recurso é criado; é *necessário* porque deve existir para qualquer documento de arquivo; é *dinâmico* porque o seu significado varia consoante a agregação em que é interpretado; e é *ilimitado* porque novos vínculos podem continuar a ser acrescentados indefinidamente ao longo do tempo. Estes recursos podem ser documentos de arquivo produzidos no âmbito de uma mesma atividade, constituindo um agregado documental que se pode identificar com um fundo. Neste caso, sobrepõe-se ao *archival bond*. Porém, não está limitado ao universo dos arquivos, podendo manifestar-se em qualquer recurso descrito em qualquer sistema de representação, incluindo nos repositórios que agregam objetos digitais provenientes de espólios de arquivo e de coleções de bibliotecas e museus e, neste caso, ultrapassa o *archival bond*, no sentido em que alarga o espectro de representação. O vínculo informacional tem, portanto, potencial para se afirmar como um termo abrangente para referir os vários contextos e interpretações em que é possível re-presentar os objetos digitais, de acordo com as relações estabelecidas num sistema de representação através de metadados. Nesse processo, enquanto forma de representação arquivística, contribui para estabelecer a autenticidade dos recursos, na medida em que os contextualiza e enquadra o seu

significado em função de outros recursos com os quais se possam relacionar, ao mesmo tempo que potencia novas interpretações e novas inferências.

#### 2.4.2 Autenticidade, confiabilidade e fidedignidade

A fidedignidade (*reliability*) é, segundo o glossário do InterPARES 2, um dos três conceitos segundo os quais a confiabilidade (*trustworthiness*) de um documento de arquivo é avaliada (InterPARES 2, 2008b, p. 49). A fidedignidade pode ser definida como *the authority and trustworthiness of records as proof and memory of the activity of which they constitute the natural byproduct. A reliable record is essentially one having the capacity to stand for the facts it is about* (Duranti, Eastwood, & MacNeil, 2002, pp. 5-6).

A legitimidade enquanto representação dos factos a que os documentos de arquivo se referem é também destacada por Kevin Owen quando este versa sobre a *reliability* enquanto *the capacity of a record to be trusted as a faithful representation of the juridical fact it speaks of* (Owen, 2015, p. 361), isto é, o grau segundo o qual um documento de arquivo *can be treated as the fact of which it is evidence* (Duranti, 1995, p. 6). A capacidade de agir como evidência, isto é, de ser fidedigno e, por extensão, confiável, depende de duas coisas: *the degree of completeness of a record's form and the degree of control over the documentary procedure in the course of which the record is generated* (Duranti et al., 2002, p. 25).

A completude (*completeness*) refere-se à presença num documento de arquivo de *all the elements required by the creator and juridical system for it to be capable of generating consequences* (InterPARES 2, 2008b, p. 11) ao passo que o procedimento documental (*documentary procedure*) diz respeito ao *body of rules governing the making of an archival document*, sendo que *the more standardized and rigorous the procedure, the more reliable the record is presumed to be* (InterPARES 2, 2008b, p. 18).

A formulação *presumed to be* aponta para a interpretação social desta asserção. Conforme salientado por Wendy Duff, *the legal, administrative, fiscal, or information value of records is dependent upon the degree of trust society places in records as reliable testimony of evidence of the acts they purport to document* (Duff, 1998, p. 88). Owen observa que a confiança social depositada nos documentos de arquivo enquanto representação jurídica dos factos não é uma qualidade inerente, e distingue fidedignidade, entendida como o grau de confiança nos documentos de arquivo *as statements of a fact*, de autenticidade, interpretada como *to trust a record to be what it purports to be, that is, to trust the record as a record* (Owen, 2015, p. 361).

Por conseguinte, enquanto que a autenticidade depende das formas de transmissão e preservação dos documentos, sendo uma responsabilidade partilhada pelo criador e por todos os que tenham tido a custódia dos documentos, a confiabilidade está associada unicamente com o momento de criação, sendo *the sole responsibility of the physical or juridical person making the record* (Duranti et al., 2002, p. 27).

Conforme visto anteriormente, um dos papéis dos metadados assenta na certificação da autenticidade e do grau de completude dos documentos de arquivo (Gilliland, 2016, p. 6). Todavia,

há vários tipos de autenticidade, que correspondem a diferentes perspectivas de análise. Luciana Duranti identifica uma autenticidade histórica, uma diplomática e outra legal<sup>55</sup>, que distingue da seguinte forma:

*Legally authentic documents are those which bear witness on their own because of the intervention, during or after their creation, of a representative of a public authority guaranteeing their genuineness. Diplomatically authentic documents are those which were written according to the practice of the time and place indicated in the text, and signed with the name(s) of the person(s) competent to create them. Historically authentic documents are those which attest to events that actually took place or to information that is true* (Duranti, 1989, p. 17).

As três perspectivas são totalmente independentes entre si. Um documento de arquivo não aceite por um tribunal pode ser histórica e diplomaticamente autêntico, mas não o é legalmente. Um assento paroquial dos séculos 16 e 17 que não termine com a expressão ‘por ser verdade assinei’ pode ser legal e historicamente autêntico, mas não o é diplomaticamente. Um documento aceite num tribunal que contenha a totalidade dos elementos formais, embora se reporte a factos falsos, é legalmente e diplomaticamente autêntico, mas historicamente falso. Assim, um documento de arquivo é autêntico *when it presents all the elements which are designed to provide it with authenticity* e é genuíno *when it is truly what it purports to be* (Duranti, 1989, p. 17).

Uma interpretação análoga pode ser construída a respeito dos metadados. Anne Gilliland pede uma *effective and efficient metadata management* como requisito essencial *to ensure that the metadata we rely on to validate digital resources is itself trustworthy* (Gilliland, 2018, p. 9). Pode lançar-se igualmente a questão da autenticidade dos metadados face à sua genuinidade. A autenticidade dos metadados diz respeito à sua exatidão (*accuracy*), um dos pilares da *trustworthiness*, entendida como *the degree to which data, information, documents or records are precise, correct, truthful, free of error or distortion, or pertinent to the matter* (InterPARES 2, 2008b, p. 2). Se houver uma bula papal assinada pelo Papa João Paulo II (1920-2005) em 2015, os metadados são autênticos se representarem que o autor desse documento é o Papa João Paulo II, independentemente de esta informação ser verdadeira ou não. Ainda que esta informação seja obviamente falsa uma vez que o documento data de 10 anos após a sua morte, os metadados, neste caso, constituem uma representação autêntica e precisa (*accurate*) do recurso.

Por outro lado, a genuinidade dos metadados refere-se à relação entre o conteúdo do recurso informacional e os factos na realidade. Neste exemplo, os metadados são autênticos e fidedignos do objeto que representam, mas são falsos. Pensemos que a data “2015” foi inscrita por engano na bula, quando a data real era 1995. Neste caso, se os metadados corrigirem esta informação para 1995 como data de criação do documento de arquivo, os metadados tornam-se genuínos porque

---

<sup>55</sup> Mais recentemente, largamente inspirado em Duranti, Bonnie Mak (2015, p. 120) também se pronuncia sobre estes três tipos de autenticidade.

representam uma informação verdadeira, mas não são autênticos por não representarem com exatidão o documento de arquivo a que dizem respeito.

A distinção entre estes três tipos de autenticidade alerta para o facto de que se pode lidar simultaneamente com diferentes noções de autenticidade, segundo as quais *a document may be valuable and trustworthy for certain kinds of analyses despite being considered inauthentic by some measures* (Mak, 2015, p. 120). Por este motivo, os critérios usados para determinar o que conta como autêntico *are themselves the product of historical, cultural, and political choices* (MacNeil, 2001, p. 45).

A melhor forma de contornar esta limitação inerente é, uma vez mais, documentando tanto quanto possível os contextos de criação, transmissão, e interpretação dos recursos através de metadados. A importância dos metadados na gestão de coleções para assegurar a integridade intelectual é defendida por Anne Gilliland (2016, p. 7), tendo Paul Conway também aludido a esta capacidade dos metadados a propósito do impacto da digitalização na preservação ao afirmar que *the digital world transforms traditional preservation concepts from protecting the physical integrity of the object to specifying the creation and maintenance of the object whose intellectual integrity is its primary characteristic* (Conway, 1996, p. 1).

No ambiente digital, são múltiplos os desafios e as dificuldades, que revelam a necessidade de um controlo intelectual rigoroso. Um desses problemas prende-se com o facto de os criadores de conteúdos começarem a indexar os seus recursos com termos irrelevantes de modo a aumentar a sua visibilidade nos serviços de indexação na web (Lynch, 1998, 2004), uma constatação que *has led to apprehensions and assertions about the content on the network making it increasingly necessary to verify its correctness, reliability, and trustworthiness* (Ramesh et al., 2015, p. 197). Ocorrências como esta reacendem a necessidade de discussão sobre a confiabilidade, a fidedignidade e a autenticidade dos metadados, sobretudo num contexto *web* em que cada vez mais interagimos unicamente através de metadados no âmbito da descoberta dos recursos. Uma representação menos correta dos recursos pode induzir o engano tanto no que concerne a recuperação indesejada de recursos (recursos a mais, com ruído), a invisibilidade de recursos pertinentes na pesquisa (não-indexação por insuficiência de metadados num formato interoperável) ou ainda a recuperação de recursos que, ainda que pertinentes, não apresentam os metadados desejáveis e suficientes (registo de metadados não completo) para o uso do recurso. Neste contexto, um esforço como o de esta investigação na identificação de metadados adequados para a representação de recursos autênticos e fidedignos na *web* encontra plena justificação e carência.

### **2.4.3 Autenticação**

Outro conceito próximo do de autenticidade, embora distinto, é o de ‘autenticação’. Frequentemente associado com a diplomática e a preservação digital, este termo assume na teoria arquivística dois sentidos: o primeiro sentido, mais coloquial, refere-se ao *process of establishing that a record is what it purports to be* (Suvak, 2015, pp. 116-117). No segundo, legal e diplomático, a autenticação é entendida como uma *declaration of authenticity made by a competent officer, and*

*consists of a statement or an element, such as a seal, a stamp, or a symbol, added to the record after its completion*, o que, neste caso, indicia que uma asserção de autenticidade *only guarantees that a record is authentic at one specific moment in time, when the declaration is made or the authenticating element or entity is affixed* (Duranti, 2010, p. 53).

Tomando ambos os sentidos, especialmente o segundo, é possível lançar a questão se um elemento de metadados que indique quando os metadados foram criados pode funcionar como uma assinatura ou um carimbo. O que é, em essência, uma assinatura? Trata-se de um mecanismo que aplica uma informação num documento que veicula a autoria do documento, na qual confiamos por acreditar que só esse criador é capaz de inscrever essa marca dessa forma específica. Num contexto medieval ou analógico em geral, a capacidade única de inscrever um carimbo prende-se com a posse das ferramentas físicas para criar uma versão autêntica e reconhecida de um determinado carimbo (a forma da estampagem, a tinta, o papel usado), bem como o acesso aos canais de divulgação do documento final.

O que é, em essência, um elemento de metadados que refira “estes metadados foram criados por ‘Fulano’?” Trata-se de um mecanismo que aplica uma informação a um registo de metadados que veicula a autoria da representação do recurso, na qual confiamos por acreditar que só os trabalhadores dessa entidade são capazes de acrescentar esse elemento ao registo, nomeadamente através do acesso à base de dados, competências para edição e publicação dos metadados nesse repositório, e que esses trabalhadores são movidos por um exercício de rigor e ética. O catálogo de um arquivo distrital ou o inventário *online* de um retalhista não se encontram abertos a edições por qualquer cidadão. Apenas têm permissões e condições para o fazer os trabalhadores desse arquivo ou organização que, nesse privilégio, constituem uma fonte de autoridade para a escolha, exatidão e completude dos metadados incluídos.

Tanto um carimbo (‘eu sou o autor’) como um elemento de metadados (‘o autor é’) cumprem a mesma função de veicular a autoria — um dos documentos de arquivo, outro da representação destes. Ao nível das descrições arquivísticas, seguindo o mesmo raciocínio, a assinatura relativamente ao carimbo está implícita no elemento ‘Autor’, sendo esta informação tão confiável e fidedigna quão a nossa confiança na instituição que gera os metadados. Se o Museu Nacional de Arte Antiga declara (recorde-se a ideia dos metadados como *statements*) que o autor dos ‘Painéis de São Vicente’ é Nuno Gonçalves, nós acreditamos. Em essência, o Museu está a colocar uma assinatura nesta informação, e a comprometer-se de que acredita e se responsabiliza pela informação de que Nuno Gonçalves é o autor destas pinturas. Se tivermos uma confiança na competência desta instituição, então acreditamos que, se esta entende que determinada pessoa é autor de um recurso, conforme expresso pelos metadados, então acreditamos que essa pessoa é de facto o autor. Neste sentido, o próprio registo de metadados funciona como um mecanismo de autenticação. A assinatura de Nuno Gonçalves funciona como um mecanismo físico de autenticação entre o pintor e o museu, e os metadados como um mecanismo de autenticação digital entre o museu e os utilizadores.



Isto levanta outra questão: o que é mais importante, a autenticação entre o autor e a instituição custodiante, ou entre a entidade custodiante e os utilizadores? Embora sendo ambas necessárias, há argumentos para destacar a segunda. No contexto da *web*, a descoberta, a partilha e a disseminação de recursos são mediadas por representações destes, isto é, por metadados (Hooland & Verborgh, 2014, 214). A interação através de metadados está embutida na própria filosofia, na conceção e na estrutura das redes digitais nas quais nos movimentamos. Ainda que tenhamos em vista a visualização de determinado objeto digital, a pesquisa, a filtragem e a seleção que nos indica se determinado recurso existe e se é útil consultar para determinado fim desejado é mediada por metadados. Assim, os metadados não só servem para a representação, a descoberta e a recuperação dos recursos como também servem, enquanto agentes primários de interação entre os utilizadores e as instituições detentoras dos recursos, como o primeiro mecanismo de autenticação do conteúdo dos recursos.

Esta formulação da autenticação que aponta para a ideia de proveniência como fonte de autoridade aproxima-se do conceito de *ius archivi*, usado por uma corrente de diplomatas a partir do século 17. Enquanto que os primeiros diplomatas medievais começaram a cultivar a noção de autenticidade nos aspetos materiais dos documentos, surgiu outro movimento que defendia a custódia institucional como principal critério para determinar a confiabilidade dos documentos de arquivo.

O conceito de *ius archivi* surgiu em tratados alemães sobre direito, evidência e litigação no século 17, nos quais a prova por arquivo (*probatio per archivum*) é entendida como um método de verificação da confiabilidade equivalente a métodos mais tradicionais de prova (Head, 2019, p. 299). Em 1658, Johann Myler (1658) incluiu o *ius archivi* como um dos direitos dos territórios imperiais alemães entendendo-o como *a set of procedural rules, resting on sovereignty, that guaranteed the legal validity of the instruments found in princely and urban archives* (Head, 2019, p. 299). Mais tarde, Ahasver Fritsch (1664) chega à mesma conclusão, mas reorganiza algumas ideias de forma a elevar a importância da proveniência do documento acima das suas características físicas ou textuais.

Esta corrente de interpretação diplomática surge como um interessante contrapeso perante a abordagem diplomática clássica de que a autenticidade é inferida pela minuciosa análise dos elementos internos e externos que presidem à elaboração de um documento de arquivo, nomeadamente a sua forma documental. Assim, *whereas the diplomatists eagerly catalogued seals, subscriptions, and hands, the ius archivi actively denied that such material details necessarily mattered for the authority of a document from a sovereign archive* (Head, 2013, p. 929). Segundo a interpretação do *ius archivi*, *archival provenance — that is, the institutional context of the records — supersedes all other considerations in the assessment of authenticity* (Mak, 2015, p. 120).

Tendo por base as considerações acima formuladas sobre a autenticação e a autoridade da voz de quem enuncia as *statements* de metadados, estaremos a regressar a uma certa conceção de *ius archivi*? Numa era de pós-verdade e desinformação, em que é cada vez mais difícil e necessário verificar as fontes e saber quem afirmou o quê, ensurdecidos por tanto ruído, estaremos cada vez

mais dependentes da autoridade de quem fala, mais do que do próprio conteúdo das palavras? Talvez uma resposta passe por uma posição intermédia. Por um lado, temos uma confiança quase cega na autoridade de certas entidades como entidades confiáveis, que nasce por vezes do comodismo em confirmação dos factos, um trabalho laborioso, minuciosos e demorado, que consome um tempo cada vez está menos disponível. Por outro, é prevalente cada vez mais a noção de que há entidades, como certos media, religiosamente comprometidos com uma campanha de desinformação em nome do lucro<sup>56</sup>, em resposta à qual se torna premente um crescente esforço em prol da confirmação dos factos<sup>57</sup>. Ambas as facetas — a autoridade do enunciador e a investigação dos factos — serão sem dúvidas componentes importantes na averiguação da autenticidade e da fidedignidade dos recursos no presente e no futuro próximo. Porém, acredita-se que talvez se venha a favorecer ligeiramente a ideia de proveniência como fator com peso mais imediato.

---

<sup>56</sup> Sobre o modo como muitos meios de comunicação em massa operam segundo um estímulo económico, ainda que isso implique a criação de conteúdos falsos (*fake news*) apenas para ter conteúdos para publicar e obter mais lucros com publicidade, ao invés de procurarem fornecer informações fidedignas, ler Holiday (2017).

<sup>57</sup> De forma interessante e reveladora dos tempos, o canal SIC começou a emitir em abril de 2019 o programa ‘Polígrafo’ no qual procura confirmar a veracidade dos principais “factos” circulados nas notícias da semana anterior.

## 3 Metodologia

---

### 3.1 Introdução

#### 3.1.1 Motivações e problema

Um trabalho de investigação em ciências sociais deve principiar na formulação de uma pergunta de investigação que, como Quivy e Campenhoudt (2013, p. 32) salientam, não é definitiva na sua formulação inicial, constituindo antes um ponto de partida provisório que serve de orientação durante o processo investigativo através do qual o investigador vai adquirindo conhecimento para melhor exprimir o seu objeto de estudo.

Porém, qualquer pergunta de investigação tem por detrás as motivações de um investigador decorrentes da constatação de uma problemática ou de um novo olhar sobre determinada realidade. No caso deste estudo, a ideia para a realização desta investigação nasce da apreciação de que nos inserimos num contexto de acelerada transformação tecnológica, social e cultural. A noção de que os referenciais que constituem os baluartes do nosso pensamento se encontram num processo de reconstrução constitui simultaneamente uma oportunidade entusiasmante de analisar essas transformações, mas também um desafio, uma vez que qualquer período de transição é dominado pela incerteza e por múltiplas possibilidades. As tecnologias de informação e comunicação têm moldado desde a década de 1990 um crescente processo de acesso global à informação, magnificando não só o número de indivíduos conectados, como também a diversidade e quantidade de informação em circulação.

No âmbito da Ciência da Informação e, particularmente, da Arquivística, as revoluções no modo como se produz, transmite e consome informação tem implicações diretas na teoria arquivística porque modifica práticas e obriga a questionar os referenciais tradicionais. Uma das transformações mais profundas, de certa forma subjacente às restantes, reside na transição da forma do documento de arquivo de analógico para digital (Cristina Freitas, 2009), o que tem sido indiciado ter causado uma mudança de paradigma na teoria arquivística (T. Cook, 2007). Novas perceções da teoria arquivística têm consequências no modo como representamos a informação em arquivos, em que o tradicional princípio de respeito pelos fundos (Duchain, 1977) tem vindo a ser questionado no contexto digital (J. Bailey, 2013). Daqui decorre que se observa uma crescente separação entre produto — os instrumentos de pesquisa — e atividade — o ato de representação em si (Llanes-Padrón, 2011). Por sua vez, as modalidades de representação de informação enquadram-se num mais amplo contexto cultural contemporâneo em que dominam as noções de pós-modernidade e pós-verdade (Duranti, 2018; Mortensen, 1999).

De entre este jogo de forças de natureza tecnológica, social e cultural, brota o questionamento do papel que a autenticidade dos documentos de arquivo, ou da informação de natureza arquivística, assume no atual contexto, e de que forma esse papel é materializado. É este o ambiente que serve de premissa e motivação ao desenvolvimento do estudo e à formulação das

primeiras questões de investigação que permitiram dar os primeiros passos metodológicos. Chegados a este ponto, a metodologia do projeto gravita em torno das seguintes questões de investigação:

- Tendo em conta o crescente e recente movimento da representação da informação usando modelos de dados em grafo no contexto da Web Semântica, (i) quais as vantagens e desvantagens desta forma de descrição quanto à verificação da autenticidade dos documentos de arquivo? (ii) Qual o paradigma de descrição arquivística subjacente a estas formas de representação e acesso à informação? (iii) Será possível construir um modelo teórico que identifique os requisitos para avaliar a autenticidade e determinar a confiabilidade dos documentos de arquivo que possa ser seguido e aplicado no contexto da Web Semântica? (iv) Quais os elementos de metadados que esse modelo deve incluir? (v) Quais são os potenciais benefícios deste modelo de metadados, em termos de descoberta da informação e completude das descrições, quando comparado com os atualmente predominantes instrumentos de descrição arquivística?

A formulação destas questões de investigação tem como processo:

- centrar a investigação na área da Ciência da Informação, mais especificamente na Arquivística;
- delimitar o objeto de estudo aos metadados usados na descrição arquivística digital, em especial aqueles relacionados com a autenticidade e confiabilidade dos documentos de arquivo;
- orientar a revisão de literatura em torno de temas como metadados, representação da informação, descrição arquivística, autenticidade e confiabilidade sob uma matriz comum de interpretação no âmbito do contexto digital;
- fornecer um referencial para a seleção das fontes de informação mais adequadas para a elaboração da investigação.

### 3.1.2 Objetivos

Assim, estabelecem-se os objetivos de investigação como forma de responder aos problemas de investigação. A investigação assume o objetivo geral de:

- Explorar a literatura científica e técnica de modo a determinar quais os conceitos teóricos sobre os quais assenta a presunção da autenticidade e da confiabilidade dos documentos de arquivo, e quais os elementos de metadados para a autenticidade que uma norma de metadados deve incluir para a representação dinâmica da informação arquivística num contexto *web*, com base num modelo de dados em grafo.

Este objetivo geral pode ser decomposto em três objetivos específicos:

- d) explorar sistemática e exaustivamente a literatura científica mais recente sobre o tópico da representação da informação em arquivos, consubstanciada na descrição arquivística.

- e) comparar a literatura técnica e normas internacionais de gestão de informação e de descrição arquivística de modo a averiguar quais os metadados ou procedimentos recomendados para estabelecer a identidade e determinar a confiabilidade dos documentos de arquivo.
- f) elaborar um modelo de metadados que devem estar presentes para a representação autêntica dos documentos de arquivo em contexto.

A propósito do primeiro objetivo específico, a discussão proposta englobará temas como compreensão e discussão de metadados; paradigmas de descrição arquivística; importância da autenticidade para a gestão da Informação; o papel das descrições arquivísticas na garantia da autenticidade e a crescente adoção de tecnologias para a representação da informação em contexto, como parte da visão da *Web Semântica*. Espera-se que o cumprimento do objetivo geral e específicos desemboque nos seguintes resultados:

- c) a compreensão do papel da descrição arquivística no processo de averiguação da autenticidade dos documentos de arquivo e/ou das próprias representações desses documentos;
- d) o desenvolvimento de um modelo de metadados que inclua os elementos identificados na literatura como desejáveis para a representação da informação arquivística autêntica no contexto da *Web Semântica*.

### **3.1.3 Limitações**

A abordagem metodológica usada para dar resposta aos desafios levantados pela questão de investigação e pelos subsequentes objetivos será largamente qualitativa, embora também seja considerado o uso de métodos quantitativos. Assim, é adotada uma investigação de métodos mistos (Creswell & Clark, 2013), que se caracteriza pela combinação de abordagens qualitativas e quantitativas na definição da metodologia de investigação (Tashakkori & Teddlie, 1998). Além de enriquecer a análise dos dados e de melhor abordar a realidade sob escrutínio, o uso dos métodos mistos enquadra-se também numa tentativa de esbater as limitações decorrentes do uso único da metodologia qualitativa ou quantitativa.

De uma forma geral, é reconhecido na literatura que os métodos quantitativos permitem um entendimento detalhado de um problema, embora pequem na capacidade de extrapolação ou generalização dos dados para um contexto social mais abrangente. Por oposição, aos métodos qualitativos são geralmente apontadas como limitações a dificuldade de análise de largos volumes de dados, ou de vários participantes, ao mesmo tempo que a investigação se encontra influenciada pelas dimensões subjetivas do investigador. Quanto ao argumento do volume de dados, tentou-se contorná-lo adotando um extenso *corpus* documental que constituísse a base para uma revisão de literatura. A respeito da subjetividade do investigador, esse aspeto é inevitável, constituindo uma limitação da própria revisão de literatura conforme apontado por Mertens (1998, p. 53). Porém, a subjetividade pode ser controlada através da consciência dessa subjetividade. Um certo grau de

parcialidade é inevitável, mas ter presente a inevitabilidade dessa subjetividade é já um primeiro passo para uma tentativa de distanciamento. Neste sentido, procura-se neste estudo fornecer visões completas de várias vozes que se escutam na literatura, destacando os vários pontos de vista sobre uma determinada narrativa, quando existente. As opiniões pessoais dos autores, decorrentes de uma leitura crítica, são apresentadas como as opiniões informadas que são, desprovidas de uma tentativa consciente de fornecer uma única visão do mundo.

Contudo, o próprio *corpus* documental apresenta algumas limitações que importa salientar. A realidade (o conhecimento científico em arquivística) é fatiada num primeiro momento pelas tipologias documentais que servem como fontes de informação. Houve uma tentativa de incluir diversas tipologias (artigos científicos, monografias, comunicações científicas e teses e dissertações) de modo a melhor capturar a diversidade do conhecimento. Num segundo momento, há uma seleção inevitável ao nível da língua dos documentos analisados. A principal língua que alimenta a revisão de literatura deste estudo é o inglês, pelo que é de admitir alguma parcialidade em relação ao contexto anglo-saxónico. Porém, de forma a diluir esta limitação, todas as pesquisas nas fontes documentais foram realizadas também em português e em inglês de forma a representar outros contextos e outras posições que não apenas as defendidas num contexto anglo-saxónico.

### 3.2 Natureza da investigação

Conforme Carlos Guardado da Silva (2020) destaca, a articulação de um dispositivo metodológico implica a identificação, seleção, recolha e verificação dos dados e interpretação da informação. A interpretação, que assenta num processo hermenêutico, é essencial para explicitar o significado, compreender e desenvolver conhecimento empírico (Bowen, 2009, p. 27; Corbin & Strauss, 2008). John Creswell (2014) identifica três abordagens de investigação: qualitativa, quantitativa e métodos mistos, salientando que não constituem categorias rígidas, distintas e opostas. Em vez disso, devem ser entendidas como diferentes extremos de um contínuo (Newman & Ridenour, 1998) no qual um estudo tende a ser mais ou menos quantitativo, sendo frequente apresentar elementos de cada uma das outras abordagens. Wiersma (1995, p. 12) também subscreve a ideia de um contínuo a nível das perspetivas metodológicas qualitativas e quantitativas, segundo a qual os métodos tendem para cada um dos extremos sem neles se diluírem por completo.

Os paradigmas de investigação condicionam as posições teóricas e as perspetivas analíticas ao fornecer uma ampla estrutura de interpretação e análise da investigação. Na opinião de Creswell, os paradigmas facilitam a compreensão dos fenómenos e o enquadramento dos estudos nas suas tendências e perspetivas, atuando como *a general philosophical orientation about the world and the nature of research that a researcher brings to a study* (Creswell, 2014, p. 6). Por este motivo, designa-os de *worldviews* condensadas na ideia de *a basic set of beliefs that guide action* (Guba, 1990, p. 17). Outros autores empregam o termo *paradigms* (Lincoln, Lynham, & Guba, 2017; Mertens, 2019); *epistemologies* e *ontologies* (Crotty, 1998, p. 10) ou *approaches* (Neuman, 2014, p. 96). Creswell identifica na literatura quatro perspetivas sobre o mundo predominantes usadas em

investigação de métodos mistos: *postpositivism*, *constructivism*, *transformative* e *pragmatism* (Creswell & Clark, 2018, p. 36).

De forma sumária, no paradigma pós-positivista, normalmente associado a abordagens quantitativas, os investigadores procuram obter conhecimento através de uma abordagem que assenta num: (1) determinismo ou raciocínio causa-efeito; (2) reducionismo das variáveis; (3) observações e medições detalhadas das variáveis; e (4) teste e refinação continuada das teorias (Slife & Williams, 1995). O construtivismo, tipicamente associado a abordagens qualitativas, opera dentro do pressuposto de que a compreensão de um fenómeno é formada através da interação com os participantes e das suas visões subjetivas. Os testemunhos dos participantes configuram-se no contexto de uma interação social subjacente. Assim, neste paradigma, a investigação surge *bottom-up* desde os modelos mentais consubstanciados em perspetivas individuais a padrões e conclusões mais abrangentes que refletem uma visão do mundo (Talja, Tuominen, & Savolainen, 2005, p. 83).

A perspetiva transformativa do mundo assume como prioridade a questão da justiça social e da salvaguarda dos direitos humanos. Atribui especial importância a comunidades marginalizadas (Mertens, 2009), devendo os investigadores abordar questões de poder, marginalização e hegemonia, entre outras, presentes nestas comunidades (Mertens & Tarsila, 2015; Mertens & Wilson, 2013). O investigador transformativo encontra-se ao serviço do mundo social, devendo contribuir para o melhorar de modo a que os indivíduos se sintam menos marginalizados e mais integrados na sociedade (Creswell & Clark, 2018, p. 37). Por fim, o paradigma pragmático, geralmente identificado como uma filosofia abrangente que inclui vários investigadores em métodos mistos (Tashakkori & Teddlie, 2003) centra-se nas consequências da ação mundo real (Denzin, 2012, p. 83), preocupando-se mais com as questões a explorar do que o método de investigação, numa perspetiva orientada para *what works and real world practice* (Creswell & Clark, 2018, p. 37).

Todavia, estes quatro paradigmas não constituem a única interpretação epistemológica da posição metodológica dos investigadores. No âmbito das Ciências Sociais e Humanas, Clara Coutinho (2019) identifica na literatura três outros paradigmas: o positivista, do qual o pós-positivista é o sucessor, o qualitativo/interpretativo e o sociocrítico/hermenêutico (Bisquerra, 1989; Latorre, Del Rincon, & Arnal, 1996; Morin, 1983).

O paradigma positivista, também denominado quantitativo, empírico-analítico, racionalista ou empiricista (Coutinho, 2019, p. 11) procura adaptar o modelo de investigação das Ciências Naturais às Ciências Sociais através da adoção de uma metodologia de natureza quantitativa (Anderson & Arsenault, 1999; Shaw, 1999; Usher, 1996). Sucedido pelo pós-positivismo (Mertens, 1998), esta abordagem teve origem nos ideais positivistas do século 19, estimuladas pelo filósofo francês Auguste Comte, que defendia a primazia positiva do conhecimento baseado na observação. Este movimento foi de tal forma marcante na história da cultura que o adjetivo ‘positivo’ adquiriu a conotação de algo *afirmativo, decisivo; que se apoia em factos e na experiência* (Dicionário Priberam da Língua Portuguesa, 2020b). Assim, afirmar que “determinada coisa é positivamente *x*” é declarar que se está plenamente convicto de uma propriedade *x* de algo. Com uma ambição

clara e confiante de *how things really are and how things really work* com o fim de *predict and control natural phenomena* (Guba, 1990, p. 19), a investigação assume uma metodologia experimental em que o conhecimento se questiona por hipóteses causais e estatisticamente comprovadas. Segundo Biddle e Anderson:

*Confirmed hypothesis can be used to generalize to populations or contexts similar to the one studied. Simple relationships are assumed among the technical terms and concepts of the investigator, research operations, findings, and conclusions from the research. Above all, social research is regarded as offering the means for generating objective evidence and avoid subjectivity and value judgements* (Biddle & Anderson, 1986, p. 231).

Segundo a interpretação positivista, o papel da teoria é central na medida em que um dos principais objetivos da investigação científica se versa na comprovação dessa mesma teoria (Crotty, 1998; Mertens, 1998; Oldroyd, 1986; Shaw, 1999). As ideias positivistas receberam várias críticas, em particular num contexto de investigação em ciências sociais e humanas em que os objetos de investigação se prendem com atos sociais, logo subjetivos (Bogdan & Biklen, 1994; Crotty, 1998; Flick, 2002). Usher (1996, p. 14) considera que tal metodologia estava necessariamente votada ao insucesso, uma vez que uma perspetiva que concebe a realidade como ordenada e sujeita a leis objetivas e mensuráveis dificilmente se poderia aplicar a uma realidade social aberta e indeterminada. Por sua vez, Cziko (1989) salientou cinco argumentos para a imprevisibilidade do comportamento humano: as diferenças individuais; o caos; o processo evolutivo da aprendizagem e do desenvolvimento; a influência da consciência e do livre arbítrio no comportamento humano; e as implicações da mecânica quântica.

O paradigma interpretativo, também designado por qualitativo ou construtivista (Creswell & Clark, 2018; Lincoln, 1990; Mertens, 1998; Schwandt, 1994) tem origens no século 20 (Bogdan & Biklen, 1994) e identifica-se com a perspetiva fenomenológica patente nos trabalhos dos filósofos Edmund Husserl e Alfred Schutz, com o interacionismo simbólico de Herbert Mead e com a etnometodologia de Harold Garfinkel (Coutinho, 2019, p. 17). Denzin e Lincoln (2005, p. 3) identificaram em 2005 oito momentos históricos de evolução da investigação qualitativa a respeito do contexto norte-americano: o tradicional (1900-1950), o modernista ou idade de ouro (1950-1970), o de géneros esbatidos ou *blurred genres* (1970-1986), o da crise de representação (1986-1990), o pós-moderno, um período experimental (1990-1995), o inquérito pós-experimental (1995-2000), o presente de contestação metodológica (2000-2010) e o futuro (2010-).

O período tradicional terá decorrido entre o início do século 20 e o fim da Segunda Guerra Mundial, mantendo uma relação estreita com a etnografia e a sociologia da Escola de Chicago. Associado com o paradigma positivista, a investigação qualitativa interessa-se pelo outro, o estrangeiro ou estranho. Um interesse mais exacerbado pela investigação qualitativa começou a ser manifestado a partir da década de 1960 por cientistas nos EUA (Crotty, 1998; Mertens, 1998), desencadeando o aparecimento de uma segunda fase, dita modernista, que coincide com a década de 1970 e se caracteriza por um intensificar de estudos que adotam este referencial enquanto reflexo de uma crescente descrença no paradigma positivista. Na fase de géneros esbatidos, as humanidades



assumem um papel mais preponderante na crítica e interpretação da teoria com vários modelos teóricos e métodos a interagirem, como o simbolismo internacional, a etnometodologia a fenomenologia ou a semiótica (Flick, 2002, p. 9). Nestes dois momentos depreende-se também o aparecimento dos argumentos pós-positivistas (Denzin & Lincoln, 2005, p. 3).

Em meados da década de 1980, a crise de representação observadas na inteligência artificial (Winograd & Flores, 1986) e na etnografia (Clifford & Marcus, 1986) atingem a investigação qualitativa, lançando os investigadores numa crise de representação de eles próprios sustentada na conceção de múltiplas versões de interpretação do conhecimento e das conclusões, dando origem a uma diáspora metodológica entre humanistas e cientistas sociais (Denzin & Lincoln, 2005, p. 3). No período pós-moderno, sentido a partir na década de 1990, a realidade e a interpretação da realidade continuaram a ser questionadas, de modo que as narrativas substituíram as teorias, e as teorias passaram a ser entendidas como narrativas específicas de um local e de um tempo (Flick, 2002, p. 10).

O período de contestação metodológica na década de 2000 refere-se provavelmente à consolidação da metodologia mista (Morse, 2003; Tashakkori & Creswell, 2007; Teddlie & Tashakkori, 2009) como uma terceira hipótese conciliadora de abordagem metodológica face às divisões entre metodólogos quantitativos e qualitativos. Por fim, a década de 2010, que identificam com o futuro, corresponderia, segundo Denzin e Lincoln, a um oitavo momento em que as ciências sociais e as humanidades se afirmam como locais privilegiados de discussão de questões fundamentais como a democracia, a raça, o género, a classe, a globalização, a liberdade e a comunidade (Denzin & Lincoln, 2005, p. 3). Com efeito, olhando em 2020 em perspetiva sobre a última década, tem-se assistido à crescente preponderância dos movimentos emancipatórios, caracterizados por uma procura de justiça social, especialmente em questões de cor, raça e de género.

O paradigma interpretativo/qualitativo/construtivista distingue-se do paradigma positivista na medida em que, enquanto que este último encarava a teoria com uma verdade objetiva, sendo o papel do investigador confirmar a existência dessa teoria, o paradigma interpretativo assume uma posição relativista que reconhece que há múltiplas realidades que existem sob a forma de construções mentais social e experiencialmente localizadas, cabendo ao investigador a tarefa subjetiva de construir esse conhecimento (Guba, 1990, p. 25). Segundo Coutinho, trata-se de substituir as noções científicas de *explicação, previsão e controlo* do paradigma positivista pelas de *compreensão, significado e ação* (Coutinho 2019, p. 17).

À luz desta interpretação, o investigador assume-se como um agente ativo na construção do social conhecimento, sendo o conhecimento produzido um produto dos valores dos investigadores (Mertens, 1998, p. 11). Por conseguinte, torna-se necessário penetrar no mundo pessoal dos sujeitos, explorando *cómo interpretan las situaciones, qué significan para ellos, qué intenciones tien* (Latorre et al., 1996, p. 42) de forma a compreender *the complex world of lived experience from the point of view of those who live it* (Schwandt, 1994, p. 221). A ação humana não decorre num vácuo, pelo que se realça a importância de interpretar os contextos sociais de

modo a perceber as motivações e intenções que desencadeiam as ações, e qual o seu significado. Os investigadores consolidam-se como o próprio instrumento da recolha de dados, transpondo para os estudos os seus valores, assunções, crenças e visões parciais do mundo (Mertens, 1998, p. 175). Por este motivo, Anderson (1993, p. 43) rejeita a posição *unbiased, objective* da investigação científica em favor de uma postura que reconheça as influências da própria condição do investigador (raça, classe, género, entre outras) na formulação do conhecimento. Neste sentido, Usher defende que estes significados se constroem na interação humana, pois nas ciências sociais e humanas, ao contrário das ciências naturais, *both the subject (the researcher) and the object (other people) of research have the same characteristic of being interpreters or sense-seekers* (Usher, 1996, p. 19).

Se os significados se constroem na interação humana, e se a interação humana é de interpretação subjetiva, coloca-se a questão de como se opera a construção dos significados. Uma resposta é a de que existem várias vias metodológicas na construção da teoria, da mesma forma que existem múltiplos contextos de atuação dos sujeitos. Na leitura de Gadamer (2013), o conhecimento resultante será necessariamente anterior e parcial uma vez que corresponde a uma perspetiva social contextualizada que constitui uma condição necessária ao nosso entendimento das coisas observadas. Investigar é, portanto, interpretar as ações de quem é intérprete, num processo de ‘dupla hermenêutica’ (Usher, 1996, p. 19), na medida que os sujeitos alvos de investigação são eles mesmos intérpretes das experiências de outrem. As interpretações emergem assim como circulares e parciais, afirmando-se a produção do conhecimento como *um processo circular, iterativo e em espiral* (Coutinho 2019, p. 19).

O paradigma interpretativo foi também alvo de críticas por investigadores inseridos nas Ciências da Educação e da Sociologia que defendem um terceiro referencial designado por sociocrítico (Coutinho 2020) ou emancipatório (Lather, 1992; Mertens, 1998). Mertens identifica esse grupo de investigadores como marxistas, feministas, minorias étnicas, incapacitados, entre outros, argumentando que estes indivíduos se revoltam contra o paradigma interpretativo porque *[it] did change the rules; however, it did not change the nature of the game* (Mertens 1998, p. 15), uma vez um grupo reduzido de investigadores das maiorias étnicas continuava a ser a voz predominante da investigação sobre sujeitos desfavorecidos. Assim, o paradigma emancipatório aborda questões políticas na investigação ao expor a opressão social (Oliver, 1992; Reason, 1994) numa tentativa de questionamento do poder e de dotar os grupos marginalizados de maior controlo (Foster, 1993).

O paradigma emancipatório encontra as suas bases em parte na teoria crítica de Jürgen Habermas. Este filósofo e sociologista alemão defende que a falsa consciência de uma prática tem um efeito retroativo sobre essa mesma prática (Habermas, 2008, p. 255), rejeitando assim a possibilidade de um conhecimento objetivo e de uma postura neutra do investigador, uma vez que o investigador atua num contexto social determinado sendo o conhecimento produzido uma construção social desse contexto. Neste ponto apresenta semelhanças metodológicas com a abordagem interpretativa, mas assume uma componente ideológica mais interventiva, votada para

eixos de atuação no âmbito da educação e da sociologia, dando origem a estudos sobre o currículo (Apple, 1979), administração educativa (Giroux, 1983), formação de professores (Zeichner, 1983) e da metodologia de investigação-ação muito usada nestas áreas (Gómez, Flores, & Jimenez, 1996; Goodson, 2015).

Chegados a este ponto, convém salientar que a metodologia se distingue dos métodos de investigação. Com base na distinção de Harding (1987, p. 2), Lather sugere que por método se devem entender as *techniques for gathering empirical evidence*, ao passo que a metodologia diz respeito à *theory of knowledge and the interpretative framework that guides a particular research project* (Lather, 1992, p. 87). Os métodos representam portanto linhas de ação segundo as quais se aborda o mundo empírico dos dados (Denzin, 1989, p. 235), correspondendo a metodologia à perspectiva filosófica que direciona essas linhas de ação. Posto isto, a investigação condensada nesta tese enquadra-se no paradigma interpretativo, uma vez que rejeita a objetividade do conhecimento veiculada pelo paradigma positivista, e descarta a dimensão interventiva e política do paradigma emancipatório, mais associado às ciências da educação e à sociologia. Adotam-se como métodos<sup>58</sup> a investigação documental e a engenharia de requisitos que dotam a metodologia de investigação de um carácter predominantemente qualitativo, embora apresente também elementos quantitativos, tornando-a mista (Tabela 9).

Tabela 9 — Métodos e metodologia que guiam este estudo.

Metodologia	Métodos
Mista	Investigação documental
Paradigma Interpretativo	Engenharia de Requisitos

FONTE: Elaboração própria.

Conforme mencionado anteriormente, os paradigmas quantitativo e qualitativo não constituem opostos absolutos, mas apenas extremos simbólicos de uma escala de gradientes na qual uma investigação se insere. Além disso, concorda-se com Punch (1998, p. 139) e Hammersley (1992, p. 52) quando referem que o vínculo a um paradigma serve mais para cerrar o debate do que o abrir. Assim, assume-se que, a nível conceptual, o objeto de estudo são não os comportamentos, mas as intenções e situações presentes na literatura, ou seja, trata-se de investigar ideias, de descobrir significados nas ações individuais e nas interações sociais a partir da perspectiva dos atores

---

<sup>58</sup> O termo ‘métodos’, no plural, é empregue mesmo quando se combina um método (ID) com uma técnica (ER). A distinção entre método e técnica não é intrínseca aos conceitos, uma vez que método é geralmente definido como um conjunto de técnicas. Neste sentido, considerar determinada ferramenta de investigação como um método ou uma técnica depende em grande medida do contexto do seu uso no seio de uma investigação, bem como do posicionamento teórico do investigador. A metodologia está, afinal, longe de ser uma ciência exata, pelo que é de esperar divergências teóricas, em certos aspetos significativas.

intervenientes no processo (outros investigadores que produzem textos científicos). Conforme Flick salienta, a investigação qualitativa caracteriza-se essencialmente por procurar *entender, describir y algunas veces explicar fenómenos sociales “desde el interior”* (Flick, 2007, p. 12). A investigação qualitativa tem um carácter eminentemente social que, por isso, assume uma dimensão subjetiva porque qualquer ação humana é intencional e deve ser enquadrada no contexto social, cultural, económico e político do sujeito que a desempenha. Tanto os conceitos como os objetos usados no decurso da investigação constituem ferramentas políticas na medida em que representam o mundo social para um fim específico, politicamente não inocente (Punch, 1998, p. 140).

Neste sentido, nos métodos de investigação qualitativa, dos quais se destaca a investigação documental como método predominante desta investigação, os documentos emergem como formas de constituir processos e artefactos sociais em diálogo (Flick, 2007, p. 13). O texto afirma-se como o material empírico e parte da noção da construção social das realidades submetidas a estudo. Na introdução ao seu manual de investigação qualitativa, Denzin e Lincoln sugerem a seguinte definição “inicial” e “genérica” de investigação qualitativa:

*Qualitative research is a situated activity that locates the observer in the world. Qualitative research consists of a set of interpretive, material practices that make the world visible. These practices transform the world. They turn the world into a series of representations, including fieldnotes, interviews, conversations, photographs, recordings, and memos to the self. At this level, qualitative research involves an interpretive, naturalistic approach to the world. This means that qualitative researchers study things in their natural settings, attempting to make sense of or interpret phenomena in terms of the meanings people bring to them* (Denzin & Lincoln, 2005, p. 3).

Desta forma, os documentos nos quais se baseia esta investigação não constituem uma realidade objetiva, mas afirmam-se como um constructo social que é necessário interpretar e desconstruir. Destaca-se a dimensão interpretativa, segundo a qual o resultado final da investigação é um produto interpretativo, por isso não objetivo, das várias interpretações que constituem o *corpus* documental. A nível metodológico: baseia-se no método indutivo *já que o investigador busca a intenção, o propósito da ação, estudando-a na sua própria posição significativa, isto é, o significado tem um valor enquanto inserido nesse contexto* (J. A. Pacheco, 1993, p. 28), adotando a postura de quem *attempts[s] to make sense of a situation without imposing preexisting expectations on the phenomena under study* (Mertens, 1998, p. 160). Como tal, a construção da teoria brota de modo sistemático e indutivo a partir do terreno à medida que os dados empíricos emergem, num sentido contrário ao da perspetiva quantitativa. Em vez de partir de um conceito claro do objeto de estudo, o investigador deve revelar abertura para que os conceitos se desenvolvem a partir da análise (Flick, 2007, p. 13).

Por sua vez, a investigação quantitativa apoia-se nos referenciais das ciências naturais e de pendor experimental e situa-se filosófica e historicamente no positivismo de Comte e no empirismo de Stuart Mill. De acordo com a perspetiva quantitativa, a investigação é baseada na teoria, consistindo muitas vezes em testar, verificar, comprovar teorias e hipóteses. Do ponto de vista

metodológico, assenta num modelo hipotético-dedutivo com base no qual o investigador parte do pressuposto de que os problemas sociais têm soluções objetivas e que estas se podem estabelecer mediante a utilização de métodos científicos (Carr & Kemmis, 1988, p. 78). O método dedutivo coloca mais ênfase na teoria, na explicação e na abstração, e menos na recolha de dados empíricos através de observação ou experimentação, teoria essa que se pretende posteriormente validar empiricamente (Bisquerra, 1989, p. 62). A teoria assume assim um papel de destaque uma vez que atua como referente teórico-prático do percurso do investigador (J. A. Pacheco, 1993, p. 9), tratando-se uma abordagem invertida comparativamente à abordagem qualitativa. A Figura 12 sintetiza as diferenças entre os processos de investigação entre as abordagens quantitativa e qualitativa.

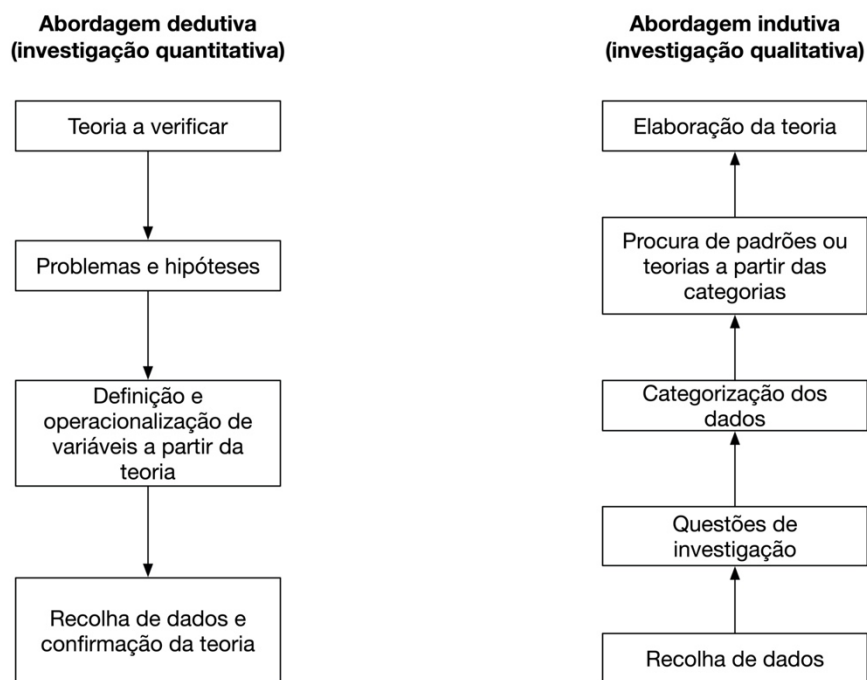


Figura 12 — Metodologia da abordagem dedutiva usada tipicamente na investigação quantitativa e da lógica indutiva comum num estudo qualitativo.

FONTE: Adaptado de Creswell (2014, pp. 59 e 66).

Esta figura 12 ilustra o modo como, na perspetiva qualitativa, o investigador parte da análise dos dados para extrair uma teoria, ao passo que na abordagem quantitativa os investigadores partem da teoria que pretendem ver confirmada nos dados. A presente investigação assume uma índole eminentemente qualitativa na medida em que parte de uma análise dos dados para a construção da teoria. Pretende-se proceder à observação, sem expectativas prévias dentro do possível, de um *corpus* literário que ilumine uma revisão de literatura para esboçar uma teoria arquivística sobre o uso dos metadados para a autenticidade na descrição arquivística digital. Todavia, considera-se pertinente enveredar por uma metodologia de cariz quantitativo que, com base nos referenciais construídos, permita a sua crítica através da formulação de hipóteses obtidas

de modo dedutivo que permitam criticar e analisar a teoria de forma a propor um modelo de metadados mais completo e atualizado (Figura 13).

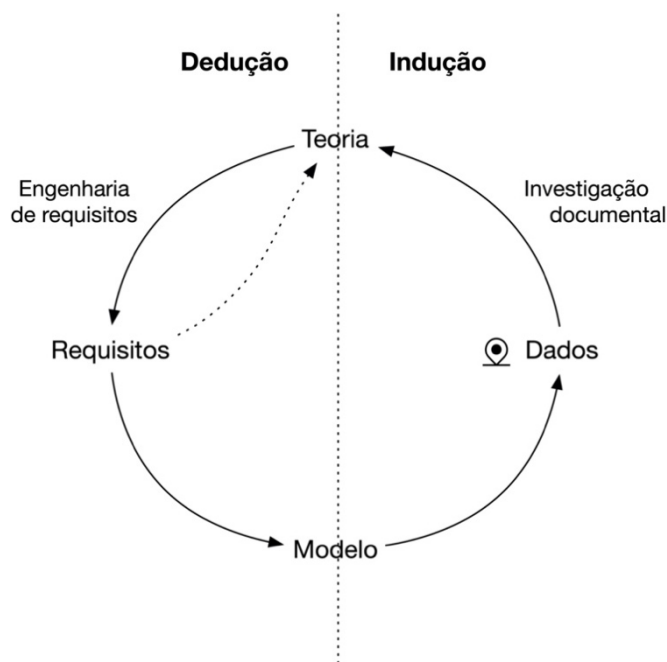


Figura 13 — Desenho metodológico da investigação.

FONTE: elaboração própria.

Assim, a nível metodológico, conforme ilustrado na Figura 13, os dados constituem o ponto de partida, sendo analisados com recurso à investigação documental (ID) de forma a inferir indutivamente uma teoria — a autenticidade dos documentos de arquivo. A teoria, uma vez formulada, será alvo de desconstrução mediante a engenharia de requisitos (ER), com o objetivo de extração de requisitos para a autenticidade. Uma vez deduzidos estes requisitos, estes podem servir como uma base para a crítica da própria teoria da qual foram extraídos em conjunção com uma reflexão crítica pessoal. Porém, pretende-se principalmente usá-los como alicerce para a construção de um modelo que capture de forma completa, no contexto da gestão da informação arquivística, os metadados desejados para a autenticidade dos documentos de arquivo. Por sua vez, este modelo pode ser novamente aplicado como instrumento de crítica e análise dos dados que constituem a literatura científica, alimentando novamente o processo de investigação. O movimento circular, baseado na descrição hermenêutica do círculo de Heidegger (2004), é necessário porque *nothing that needs interpretation can be understood at once* (Gadamer, 2013, p. 197). Um dos principais argumentos da hermenêutica é o de que o significado de uma parte apenas pode ser compreendido em relação com o todo. Como a teoria é construída a partir da interpretação os dados, ela torna-se provisória e dependente dessa mesma interpretação. Torna-se assim necessário que trabalhos futuros usem o modelo, construído a partir da teoria, para reinterpretar os dados e reformular e aperfeiçoar a teoria.

O uso combinado de métodos de investigação qualitativos e quantitativos qualifica esta investigação como *mixed* ou mista (Bryman, 1992; Creswell & Clark, 2018; Hammersley, 1996;

Teddle & Tashakkori, 2009), por vezes também designada de forma relativamente obsoleta por *multimethod research* (Brewer & Hunter, 2006; Morse, 2003). A metodologia mista é entendida como *a third methodological community* (Teddle & Tashakkori, 2015, p. 13) ou *third methodological moment* (Denzin, 2012, p. 82) na medida em que acrescenta uma possibilidade metodológica à tradicional oposição entre métodos qualitativos e quantitativos.

A realidade raramente é simples e facilmente compreensível à primeira vista. Não raras vezes, cada fenómeno ocorre numa complexa e interrelacionada rede de contextos, interpretações e agentes. Por este motivo, quando os investigadores se lançam a investigar tais fenómenos, a opção por uma metodologia de investigação mista testemunha o respeito por essa complexidade, que reconhece que os dados recolhidos por cada método constituem uma de múltiplas perspetivas possíveis (J. Greene, 2007). Por conseguinte, a metodologia mista assenta na combinação de métodos e técnicas distintos no estudo de um fenómeno (Denzin, 1989, p. 234). No âmbito da primeira publicação do *Journal of Mixed Methods Research* em 2007, os editores Abbas Tashakkori e John Creswell avançaram com uma definição abrangente de métodos mistos enquanto uma:

*Research in which the investigator collects and analyzes data, integrates the findings, and draws inferences using both qualitative and quantitative approaches or methods in a single study or a program of inquiry* (Tashakkori & Creswell, 2007, p. 4).

Um estudo que adote uma metodologia mista coloca-se à partida numa posição mais vantajosa quando em comparação com um estudo que siga uma única abordagem metodológica. Um dos argumentos favoráveis mais notáveis, usado há várias décadas, reside no facto de permitir aos investigadores encontrar num método pontos fortes que contrariem as fraquezas de outro método usado de forma isolada (Jick, 1979, p. 602). Creswell e Clark (2018) argumentam que a investigação quantitativa é limitada na interpretação do contexto social em que decorre a investigação, uma vez que os participantes não têm uma voz ativa na investigação, uma fraqueza que pode ser contornada pela adoção adicional de uma metodologia qualitativa. Por sua vez, salientam que a investigação qualitativa não só apresenta uma parcialidade inerente devido às interpretações subjetivas do investigador, como também tem dificuldades em generalizar esse conjunto de conclusões a um grupo mais alargado que não se restrinja apenas ao pequeno número de participantes. Por conseguinte, concluem que *the strengths of one approach make up for the weakness of the other* (Creswell & Clark, 2018, p. 12). Estes autores defendem ainda que o uso de uma metodologia mista permite responder a questões que não podiam ser respondidas de outra forma, dando origem a novas perspetivas de análise além da tradicional separação entre quantitativo e qualitativo, e incentivando o uso de múltiplas perspetivas sobre o objeto científico. Porém, o uso de mais métodos de recolha e análise de dados acarreta necessariamente uma maior exigência ao nível da competência do investigador e dos recursos — sobretudo tempo, mas também potencialmente financeiros. Além disso, Bryman, apesar de defender que *one approach enhances the other*, também alerta que uma investigação que combine métodos quantitativos e qualitativos não é necessariamente preferível a uma de metodologia única, pelo que uma investigação de

metodologia mista tem de ser capaz de justificar claramente as vantagens que proporciona (Bryman, 1992, p. 69).

Na medida em que promove uma abordagem livre e flexível por parte dos investigadores aos problemas sob investigação na escolha dos métodos, não é útil apontar propriamente uma forma exclusiva de execução desta metodologia. Observam-se na literatura diversas formas de combinar a metodologia quantitativa e qualitativa, dependendo em grande medida da natureza do próprio objeto de estudo e da limitação referida dos recursos. Hammersley (1996, pp. 167-168) propõe três formas de vincular a investigação qualitativa e quantitativa:

- 1) *Triangulation. Here, the findings from quantitative and qualitative techniques are used to check one another, on the basis that they are likely to involve different sorts of threat to validity ...*
- 2) *Facilitation. Here one approach acts as a source of hypotheses or as a basis for the development of research strategies in the other ...*
- 3) *Complementarity. Here the two approaches provide different sorts of information which complement one another. Thus, qualitative research is sometimes regarded as better able to provide information about interactional processes and about subjects' perspectives, whereas quantitative method is presumed to be better at documenting frequencies and causal patterns.*

Destaca-se a complementaridade como a perspetiva que mais se parece adequar ao esquema metodológico proposto no âmbito desta investigação. Assim, as abordagens qualitativas e quantitativas adotadas são usadas de forma complementar, de modo a que cada uma ampare as fraquezas da outra, de onde resulta uma investigação mais robusta e com resultados mais pertinentes. Cada método de recolha e análise de dados tem falibilidades inerentes (Eugene Webb, Campbell, Schwartz, & Schrest, 1966, p. 5), pelo que o uso de diversos métodos acrescenta credibilidade ao problema enunciado face a uma investigação que apresente uma excessiva confiança num único método (Lee, 2000, p. 23). Na base desta postura reside a assunção fundamental de que *the combination of qualitative and quantitative approaches provides a more complete understanding of a research problem than either approach alone* (Creswell, 2014, p. 4). Greene também realça esta perspetiva quando se refere à metodologia mista como *multiple ways of seeing and hearing, multiple ways of making sense of the social world, and multiple standpoints on what is important and to be valued and cherished* (J. Greene, 2007, p. 20), salientando como múltiplos olhares sobre o mundo se complementam e enriquecem sem pretensões de estabelecer uma compreensão absoluta e definitiva do fenómeno sob estudo.

Em síntese, de entre as 11 formas que Bryman (1992, pp. 59-61) aponta como formas de integrar a investigação quantitativa e qualitativa, destacam-se duas: o facto de ambas serem passíveis de combinação de forma a proporcionar um quadro mais geral do problema em estudo (hipótese 4); e a maior adequação da investigação qualitativa e quantitativa a diferentes fases do processo de investigação (hipótese 10). No nosso estudo, espera-se que a investigação documental



nos guie a uma compreensão detalhada do problema de como representar a autenticidade dos documentos através de descrições arquivísticas. Contudo, este esforço por si só deixar-nos-ia inscientes sobre como extrapolar as nossas descobertas para um contexto mais geral de análise de vários campos de descrição empregues por um largo número de normas de metadados não só no contexto da arquivística, mas num enquadramento mais amplo de entidades que se dedicam à gestão de informação cultural, como bibliotecas e museus. Por outro lado, o uso solitário da engenharia de requisitos como método facultar-nos-ia uma lista mais ou menos comparativa de cada elemento de metadados usado, mas com limitada massa crítica para os analisar, discutir e enquadrar nas necessidades teóricas do problema.

Portanto, o uso da metodologia mista, enquadrado no paradigma interpretativo, é entendido como sendo rico e desejável enquanto forma de combinar o melhor de ambas abordagens, de modo a orientar o desenvolvimento de um modelo de metadados coerente com os tópicos fundamentais discutidos pelos autores mais influentes, associando-os aos elementos de metadados encontrados em normas de metadados usados sobretudo em arquivística e, inversamente, contextualizando esses elementos de metadados na discussão científica. Reforça-se a necessidade e a vantagem de enveredar por esta abordagem mista devido à complexidade da realidade que pode ser mais bem interpretada se forem aplicados métodos e técnicas complementares para a interpretação dos dados empíricos (Brewer & Hunter, 2006, p. 2), proporcionando uma melhor compreensão do fenómeno sob investigação (Teddlie & Tashakkori, 2009, p. 16).

### **3.3 Métodos de recolha de dados**

Qualquer estudo empírico necessita de se basear num conjunto de dados recolhidos de forma sistemática, transparente e com critérios objetivos. A resposta dada às perguntas de investigação e, conseqüentemente, os resultados de um estudo, dependem diretamente dos dados recolhidos e do que estes revelam, pelo que é também desejável que os métodos e critérios de recolha de dados sejam replicáveis por outros investigadores de forma a comprovar a legitimidade dos dados e a validade dos resultados da investigação. A escolha dos métodos de recolha de dados depende diretamente dos métodos de investigação propostos. No âmbito deste estudo, recorda-se que são dois os métodos que governam a metodologia: investigação documental (ID) e a engenharia de requisitos (ER). Como a ER parte dedutivamente de uma teoria já construída, o momento de aplicação deste método pressupõe que os dados se encontram recolhidos. Neste sentido, nesta parte analisam-se a ID como método que determina o tipo de dados a recolher.

#### **3.3.1 Investigação documental**

Análise documental (Bowen 2009), pesquisa documental (Albarello et al., 2005), investigação de dados documentais ou investigação histórica (Connaway & Powell, 2007) são termos alternativos para designar o método de investigação documental, uma terminologia proposta por Carlos Guardado da Silva por considerar que esta designação é mais abrangente e ilustra melhor a completude do método de investigação enquanto método, e não como uma mera etapa do processo

de investigação (C. G. da Silva, 2021, p. 103). Concorde-se com esta posição, razão pela qual se mantém ‘investigação documental’ como designação deste método.

O método de investigação documental refere-se à análise de documentos que contenham informação relacionada com o fenómeno que constitui o objeto de estudo (K. Bailey, 1994), tendo sido amplamente adotado na investigação em ciências sociais desde a sua formulação (Ahmed, 2010). Inclui a identificação, seleção e análise de documentos e do seu contexto de modo a entrever não só o significado dos seus conteúdos, mas também para compreender o significado do que não está escrito, os silêncios, os vazios que por vezes são igualmente significativos (C. G. da Silva, 2020). Enquanto método de investigação, é adotável por abordagens qualitativas (Ary, L., & Sorensen, 2010; Bowen, 2009; Stake, 1995), quantitativas ou mistas (C. G. da Silva, 2020). Por este motivo, figura como um método fundamental para qualquer projeto de investigação, assumindo um papel determinante numa revisão de literatura ou na caracterização do estado da arte de um determinado tema (Tuckman, 2012), uma vez que os estudos devem basear-se no conhecimento científico prévio (Carmo & Ferreira, 1998).

Como qualquer outro método, apresenta vantagens e desvantagens. Bowen (2009, p. 31) destaca sete vantagens da adoção deste método:

- (1) *eficiência*: a recolha de dados em documentos exige menos tempo do que métodos mais onerosos como o inquérito por entrevista, tornando-se mais eficiente;
- (2) *disponibilidade*: grande parte dos documentos encontra-se no domínio público ou é acessível mediante subscrição das bibliotecas académicas;
- (3) *custo-benefício*: os dados presentes nos documentos já se encontram recolhidos, sendo apenas necessário avaliar a sua qualidade e conteúdo;
- (4) *ausência de intromissão e reatividade*: os dados documentais não são afetados pelo processo de investigação;
- (5) *estabilidade*: a presença do investigador não modifica os documentos;
- (6) *exatidão*: inclusão de detalhes como nomes, referências e detalhes dos eventos;
- (7) *cobertura*: abrangem um longo período de tempo, de eventos e de contextos.

Conforme salientado por Bowen (2009), um dos seus principais pontos fortes diz respeito à não-reatividade da evidência documental (E. Webb, Campbell, Schwarz, & Sechrest, 1984), o que se refere ao facto de os documentos serem geralmente criados para outros fins que não a investigação, pelo que não sofrem de parcialidades inerentes às intenções do investigador. Os dados são preexistentes (Quivy & Campenhoudt, 2013), cabendo apenas ao investigador procurá-los, não criá-los (Sousa, 2005, p. 89). Assim, a posição do investigador não influencia a criação dos documentos, garantindo uma distância imparcial relativamente à produção dos textos, embora a interpretação, conforme visto anteriormente, constitui um exercício intencional, pessoal e subjetivo. Além disso, a recolha de dados é eficiente, relativamente rápida e as fontes são estáveis (Bowen, 2009; Quivy & Campenhoudt, 2013; Yin, 2010).

De entre as desvantagens mais frequentemente mencionadas da investigação documental, destacam-se a disponibilidade limitada de fontes, dificuldades na recuperação devido a informação em falta ou incompleta e tendências inerentes (K. Bailey, 1982, 1994; Hakim, 1993; C. G. da Silva, 2020; Stewart, 1984; Treece & Treece, 1982; E. Webb et al., 1984; While, 1987; Yin, 2010), destacando-se que os documentos são sempre *indiretos e incompletos* (Veyne, 1987, p. 178). As tendências podem dizer respeito quer a parcialidades na seleção dos documentos a serem preservados e publicados, quer a parcialidades do investigador na interpretação das fontes. De um modo geral, mesmo que o investigador seja movido por uma imparcial boa-vontade científica, uma certa parcialidade é inevitável, uma vez que as análises são conduzidas por um indivíduo e um indivíduo é o produto do seu contexto. A respeito das dificuldades de recuperação, as possibilidades atuais de publicação de dados online, que permitem uma pesquisa extensiva e detalhada em repositórios e bases de dados, afirmam-se como uma mais-valia imprescindível que viabiliza a recolha de dados documentais a partir de fontes consideradas pertinentes pela qualidade e pelo conteúdo dos documentos científicos.

A ausência de reatividade dos documentos está na origem da classificação da investigação documental como um método não interferente (*unobtrusive method*) (Lee, 2000). Este termo foi cunhado por Eugene Webb et al. (1966) como alusão a uma técnica de obtenção de dados obtidos que não envolva a recolha direta de informação a partir de sujeitos investigados, exemplificado no caso dos inquiridos por entrevista ou questionário. Os métodos não interferentes caracterizam-se por usarem técnicas de recolha de dados que permitem ao investigador obter informação sobre um fenómeno social sem nele poder intervir ou modificar (Teddlie & Tashakkori, 2009). Lee (2000, p. 17) salienta a fragilidade metodológica das entrevistas e dos questionários porque, por um lado, o investigador intromete-se no contexto social que pretende investigar. Outros autores destacam diferentes modos em como a sua presença constitui elemento estranho que pode levar a respostas falsas (Bishop, Oldendick, Tuchfarber, & Bennett, 1980; Bradburn et al., 1979; Presser & Traugott, 1992). Por outro, o modo como as perguntas são formuladas condiciona o tipo de respostas obtidas (Clark & Schober, 1982). Este fenómeno da reatividade veio a ser conhecido pelo efeito Hawthorne, segundo o qual a simples presença de um investigador no processo de recolha de dados pode distorcer as conclusões obtidas (Lee, 2000, p. 20).

Neste sentido, a particularidade dos métodos não interferentes prende-se com o facto de os investigados não saberem que são alvo de investigação e por conseguinte, agirem naturalmente. A observação é passiva sem qualquer tipo de interferência sobre o desenrolar dos acontecimentos (Teddlie & Tashakkori, 2009), figurando os documentos como fontes não reativas uma vez que a presença do investigador não os altera (Eugene Webb, Campbell, Schwartz, & Schrest, 1981). Por consequência, reforça-se que apresentam as vantagens de não interferir na produção dos dados, anulando os efeitos reativos da observação e da inquirição por parte do investigador. Além disso, podem incidir sobre factos ocorridos no passado, uma vez que os dados estão disponíveis por longos períodos de tempo, adequando-os a estudos longitudinais (Coutinho, 2019, p. 145), legitimando-se na perfeição ao presente estudo. Porém, os documentos podem constituir fontes de informação

incompletas porque nem tudo é registado e arquivado, podendo o acesso ser difícil ou até mesmo vedado. Pretende-se contrariar esta desvantagem através da escolha de bases de dados às quais se possui acesso institucional.

O método de investigação documental implica a constituição de um *corpus* documental e a análise das fontes de informação recuperadas, tópicos que serão detalhados posteriormente. Na sua dimensão de investigação qualitativa, a constituição do *corpus* não obedece a uma seleção da amostra com base em critérios formais de uma população, tais como número de indivíduos do sexo masculino, com uma determinada idade, entre outros. Ao invés, a recolha da amostra apresenta uma intencionalidade subjacente na medida em que seleciona deliberadamente um *corpus* de documentos que atua como exemplo empírico do fenómeno sob estudo (Flick, 2007, p. 50). Os dados de investigação são documentos, quer assumam a forma de textos (Rapley, 2007) ou imagens (Banks, 2007), que constituem o *corpus* documental. Os documentos que neste estudo constituem dados de investigação são a produção académica sob a forma de artigos científicos, monografias, comunicações científicas e teses e dissertações. Isto pressupõe uma seleção do material existente com o objetivo de o analisar, conforme será detalhado em pormenor adiante.

Porém, antes de transitar para o esclarecimento do *corpus* documental, devem-se ainda tecer algumas importantes considerações sobre a recolha de dados. Como se pretende empregar a ID como uma forma de construção de uma teoria que encontre sustentação sistemática nos dados de investigação, destaca-se que uma característica essencial deste método é a de que os referentes teóricos não são formulados no início do estudo. Ao invés, a teoria, enquanto generalizações, emerge dos próprios dados de investigação, ancorando-se diretamente nos métodos de recolha e análise de dados (Mertens, 1998, p. 175). Por este motivo, a teoria desenvolve-se com base nos dados empíricos, cabendo ao investigador aceitar a responsabilidade de interpretação dos dados e de sugestão de uma teoria com base neles. Gómez et al. (1996) salientam o papel desempenhado pelo investigador, que deve iniciar o processo de investigação sem ideias prévias sobre qual a teoria que responde ao seu problema de investigação de forma a construí-la num processo interpretativo, analítico e contínuo entre a análise de dados, a observação de padrões e a recolha de novos dados até obter uma teoria satisfatória. Bravo e Eisman atestam esta relação umbilical entre a teoria e os dados quando afirmam que:

*La teoría se genera a partir de los datos de una realidad concreta, no partiendo de generalizaciones a priori ... la teoría [é] generativa ... inductiva — las teorías se desarrollan desde abajo, a través de la interconexión de evidencias y datos recogidos —, constructiva — las unidades de análisis comienzan a aparecer en el curso de la observación y descripción; no son definidos previamente a la recogida de datos... (Bravo & Eisman, 1998).*

Se a teoria nasce a partir dos dados, a investigação deve iniciar-se pelos dados a investigar, que decorrem por sua vez da formulação do problema de investigação. Segundo esta abordagem — que se adota nesta investigação — a revisão de literatura só deve ocorrer após as diretrizes teóricas presentes nos dados se terem tornado claras (Punch, 1998, p. 168). A revisão de literatura figura

assim como um fim em si mesmo de informar sobre uma prática ou de fornecer um entendimento detalhado sobre o que constitui um tópico (Mertens, 1998, p. 35), e de fornecer diretrizes teóricas que possam ser introduzidas em fases posteriores da análise de dados.

Assim, embora a revisão de literatura surja estruturalmente no primeiro e segundo capítulos, a investigação foi cronologicamente iniciada pela definição do problema de investigação. A partir desta, estabeleceram-se critérios de recolha de dados, que foram subsequentemente repartidos em categorias e analisados. A partir da leitura das fontes de informação construiu-se a revisão de literatura, que serve de base à análise dos dados sob uma perspetiva da engenharia de requisitos, enquanto base de construção do modelo que se pretende propor. Como um dos principais objetivos desta investigação é a proposta de um modelo, e como um modelo é estruturalmente composto por requisitos, há, por isso, a necessidade de definição desses requisitos. No âmbito deste trabalho, os requisitos são definidos por duas vias: ora pela investigação documental, ora pela engenharia de requisitos.

A investigação documental pressupõe a heurística e a hermenêutica dos textos, o que realça uma dimensão interpretativa das fontes de informação, segundo a qual se torna necessário proceder a uma análise e discussão dos dados encontrados nas fontes científicas, de forma a extrair os requisitos encontrados na literatura científica. Como etapa deste processo é realizada uma análise categorial da literatura, organizando-a por temas, que numa segunda fase constituem um ponto de partida para uma segunda análise que visa a identificação de requisitos consoante cada tema. Espera-se que este raciocínio possa ilustrar eventuais similaridades ou diferenças entre determinado tema e o tipo de requisitos que elege para o entendimento da autenticidade e da confiabilidade dos recursos. Por sua vez, a engenharia de requisitos será aplicada na literatura técnica — nas normas de metadados selecionadas — de forma a ilustrar os requisitos emergentes deste *corpus* documental. Desta forma, obtém-se uma complementaridade entre ambos os métodos de investigação, na medida em que ambos contribuem para a definição de requisitos a integrar no modelo proposto, enquanto teoria.

Se entendermos a teoria como uma construção a partir dos dados, a teoria ambicionada nesta investigação prende-se com a forma como as descrições arquivísticas, estudadas enquanto metadados, manifestam a autenticidade dos recursos que descrevem e permitem a um utilizador considerar essas informações como confiáveis. A função da revisão de literatura é a de *fornecer pistas alternativas/complementares que possam ser introduzidas em fases subsequentes da análise de dados, amplificando o seu espectro de intervenção* (Coutinho 2019, p. 346). Conforme referido, uma das técnicas de análise será a engenharia de requisitos que, consistindo na definição de requisitos de um sistema ou, no nosso caso, de uma teoria, pretende operar a passagem dos conceitos teóricos identificados na literatura para os requisitos (elementos de metadados) que um modelo de metadados tenha de incluir para capturar de forma completa a autenticidade dos recursos representados. Detalham-se assim os métodos de investigação com que se pretendem operacionalizar a natureza mista da investigação, ilustrada na Figura 12: a abordagem qualitativa é num primeiro momento iniciada por uma investigação documental de predominantemente de

natureza qualitativa, embora apresente traços quantitativos de quantificação das fontes de informação, que assenta na revisão de literatura no sentido em que se pretende desenvolver a perspectiva teórica sobre o tema em estudo sem condicionantes prévios, dentro das possibilidades da isenção humana. Uma vez constituído o *corpus* de fontes de informação conforme decorrente da adoção dos métodos de investigação documental, as fontes de informação são assimiladas e discutidas através de uma análise consubstanciada na revisão de literatura, a partir da qual emerge uma teoria. Neste ponto importa realçar que a teoria se sustenta na revisão de literatura, o que justifica uma revisão de literatura sistémica, criteriosa e abrangente relativamente às temáticas investigadas, uma vez que, para uma revisão poder ser considerada abrangente, deve incluir no espectro de análise diversos temas, perspectivas e correntes de pensamento. Neste caso, a revisão de literatura não se restringiu unicamente a temas específicos como metadados ou normas de metadados, considerando também textos de áreas afins como a representação da informação, a Diplomática, e a *Web Semântica*.

Quando essa teoria emerge, fundamentada nas leituras, aplica-se uma dimensão tendencialmente mais quantitativa de engenharia de requisitos dessa teoria, de modo a operacionalizar os conceitos em critérios de implementação. A formulação de requisitos pretende não só servir como um instrumento de referência crítica para antigas e futuras teorias mas também, e em especial, como fundamento metodológico para a construção de um modelo que parte da teoria mas que serve também como quadro de referência de aplicação a novos dados que venham a ser produzidos, contribuindo assim para a emergência de uma nova teoria. Este conjunto de interações encerra assim o ciclo hermenêutico no qual opera a investigação científica (Figura 13), através do qual as publicações científicas representam um determinado consenso teórico que serve de base a novas reflexões que eventualmente contribuem para a reformulação desse consenso académico.

No caso desta investigação, a teoria diz respeito à forma como o uso dos metadados nas descrições arquivísticas desempenha um papel na asserção da autenticidade. Os requisitos apresentam uma dimensão teórica ou qualitativa e uma dimensão quantitativa. A nível qualitativo, correspondem aos conceitos que os autores identificados no *corpus* documental entendem ser necessários para manifestar a autenticidade dos recursos representados (identidade, integridade, cadeia de custódia, entre outros). A nível quantitativo, os requisitos traduzem-se nos elementos de metadados concretos que preencherão o modelo de análise (título, identificador único, criador, entre outros).

Em síntese, o método de investigação documental será usado neste projeto como uma base para a revisão da literatura, identificando os autores mais influentes e analisando não só os conceitos e ideias expressas nas suas obras escritas, mas também os silêncios ensurdecedores, situando a discussão no contexto de outras obras semelhantes. Este esforço consiste primariamente nos primeiros dois capítulos, como parte da elaboração de uma revisão de literatura em torno dos conceitos-chave de representação de informação autêntica no âmbito da gestão da informação. Porém, o uso da investigação documental não se limita à revisão de literatura. Como Merriam (1988, p. 118) salienta, *documents of all types can help the researcher uncover meaning, develop*

*understanding, and discover insights relevant to the research problem.* Neste sentido, no quarto capítulo, as normas de metadados constituirão igualmente um *corpus* documental a partir do qual será possível analisar, interpretar e refletir sobre os elementos de metadados estruturais adotados para a descrição dos documentos de arquivo, em particular os metadados mais diretamente relacionados com a verificação da autenticidade e com a demonstração da confiabilidade.

### 3.4 *Corpus* documental

Na abertura do longo e influente manual de investigação redigido por Quivy & Campenhoudt (2013), os autores utilizam a metáfora de uma exploração petrolífera para expressar o modo de investigar em ciências sociais. Da mesma forma que uma perfuração com fins comerciais não é casuística, sendo antes precedida de um estudo cuidadoso dos terrenos e dos locais ótimos, também ao planear uma investigação científica é necessário proceder a um levantamento calculado do terreno, de modo a recuperar os textos científicos mais valiosos.

Num cenário ideal, a revisão de literatura englobaria a totalidade dos registos bibliográficos existentes, de modo que a discussão tivesse como panorama de fundo todo o conhecimento acumulado de todo o génio humano em temas de arquivística. Porém, como a única semelhança entre o ideal e o real reside na rima, tamanha empresa não é possível. Defrontamo-nos assim com o inexorável problema da delimitação de critérios de amostragem que permitam obter uma coletânea de fontes que sejam manuseáveis em número pelos investigadores, ao mesmo tempo que se afirmam como representativas da população total e exemplares na discussão científica por elas veiculada. Na impossibilidade de recolher e analisar todo o conhecimento existente, a observação tem necessariamente de ser *selective and purposive and involves decisions about what should be noticed and should be ignored* (McCall, 1984, p. 270). Por conseguinte, é necessário definir o que se vai estudar, onde e quando (Lee, 2000, p. 73) ou, por outras palavras, há que formular requisitos para a constituição de uma amostra.

A abordagem metodológica adotada segue largamente o percurso metodológico usado por Maria Cristina de Freitas (2009, 2017) e Carlos Guardado da Silva (Cristina de Freitas & Silva, 2016) noutros trabalhos. Neste sentido, adota-se uma amostragem não-probabilística e intencional das fontes de informação. Segundo Garcia Roldán (2003), uma amostragem não-probabilística é aquela em que não é possível determinar as probabilidades de seleção dos ‘indivíduos’ que, no nosso caso, correspondem a textos científicos. A intencionalidade diz respeito aos critérios de seleção do investigador que determina o grau de adequação das unidades de análise aos seus interesses. Assim, a constituição da amostra não se deve à casualidade, mas antes ao enquadramento das fontes de informação nela contidas aos objetivos de investigação pretendidos.

O método de investigação documental implica a seleção do tipo e da natureza das fontes documentais em função do problema de investigação (C. G. da Silva, 2020, p. 8). Existem diversas formas de categorização dos dados documentais (Hammersley & P. Atkinson, 1995, p. 159). Macdonald e Tipton (1996) distinguem entre *public records*, os *media*, *private papers* e *visual documents*. Finnegan (1996) separa as fontes quanto a um uso direto ou indireto. Uma outra

distinção, que se prefere, é a da separação entre fontes de informação primárias e secundárias (Punch, 1998, p. 190). Esta assunção é possível porque, como Carlos Guardado da Silva (2020, p. 4) argumenta, o conceito de documento alargou-se ao de fonte de informação na medida em que os documentos, independentemente da sua forma ou tipo, atuam como recursos sobre os quais as investigações e os estudos são baseados.

As fontes de informação podem ser classificadas em fontes primárias, secundárias ou terciárias. De acordo com o *Dicionário do Livro*, as fontes primárias definem-se enquanto uma *fonte, como livros, publicações em série, relatórios científicos e técnicos, atas de congressos, programas de investigação, teses, etc., que são a primeira consequência da atividade do intelecto; fonte ou publicação em primeira mão; fonte original; documento original sobre um determinado tema* (Faria & Pericão, 2008, p. 565).

Por sua vez, por fonte secundária entende-se *aquela que contém dados e informações sobre fontes primárias* (Faria & Pericão, 2008, p. 565), nomeadamente bibliografias e catálogos de bibliotecas. Por fim, a fonte terciária trata-se de uma *obra especializada que cobre um conjunto de conhecimentos ou explicações concisas relacionadas com temas, autores, trabalhos, associações, recursos, etc.* (Faria & Pericão, 2008, p. 565), tais como monografias de referência, enciclopédias e dicionários.

No âmbito da investigação documental usada neste estudo como parte da elaboração da revisão de literatura, importam sobretudo as fontes primárias de informação, isto é, a produção científica que constituem a base do desenvolvimento de uma teoria sobre a autenticidade dos documentos de arquivo. Porém, torna-se necessário estabelecer critérios de pesquisa e seleção dessas fontes primárias através de fontes secundárias como bases de dados de referência. Neste sentido, explicam-se adiante os critérios de seleção das bases de dados consultadas e de seleção de fontes. O nosso aparato metodológico cinge-se aos espaços de divulgação direta ou indireta da literatura científica em Arquivística (em inglês: *Archival Science* e *Records Management*) ao longo da última década (2009-2019) de modo a incluir, por um lado, o rigor dos veículos de comunicação científica legitimados pelos pares e, por outro, para abarcar a contemporaneidade do objeto de investigação, ao mesmo tempo que se enquadra num período pertinente a análise tendencial da evolução dos conceitos propostos para análise nos objetivos de investigação.

A nível tipológico, o universo de recolha e análise estrutura-se em torno de cinco espécimes de populações de fontes documentais: (i) revistas científicas através das quais se recolhem (ii) artigos científicos e (iii) revisões de monografias; (iv) comunicações científicas em congressos; e (v) teses de doutoramento e dissertações de mestrado. Assumem-se como foco as bases de dados mais preponderantes a nível internacional enquanto canais legitimados pelos pares para a divulgação de conteúdo científico pertinente que será posteriormente analisado com mais detalhe no decorrer do estudo.

De entre estas cinco fontes de informação, as revistas científicas são aquelas que merecem maior atenção dado que se trata de fontes de informação com um potencial informativo bastante relevante, conforme atestado na seguinte passagem do manual de Quivy e Campenhoudt:



*as revistas especializadas ... são particularmente interessantes [como fontes], por duas razões. Primeiro, porque o seu conteúdo traz os conhecimentos mais recentes na matéria ou um olhar crítico sobre os conhecimentos anteriormente adquiridos. Num e noutro caso, os artigos fazem frequentemente o balanço da questão que tratam e, assim, citam publicações a ter em consideração. A segunda razão é que as revistas publicam comentários bibliográficos sobre as obras mais recentes, graças aos quais [se] poderá fazer uma escolha acertada de leituras* (Quivy & Campenhoudt, 2013, p. 55).

Assim, as revistas científicas constituem lugares privilegiados não só como veículos de publicação de artigos científicos que veiculam, debatem e representam o conhecimento mais atual numa determinada área, como também se afirmam como canais que permitem a identificação de monografias recentes e pertinentes. Perante o carácter central que as revistas científicas assumem, figura como indispensável a sua caracterização. Ao longo das próximas secções será construído o perfil de cada revista científica sob uma miríade de perspetivas que, uma vez consideradas em conjunto, poderão elucidar-nos sobre a adequação, ou não, da sua escolha para as finalidades da investigação a que nos propomos. A recolha e o contraste de dados em locais complementares teve como principal foco as páginas *web* das próprias revistas científicas, a base de dados da *International Standard Serial Number*<sup>59</sup> (ISSN), a base de dados *do Information Matrix for the Analysis of Journals*<sup>60</sup> (MIAR), e as listas de títulos da *Library & Information Science Abstracts* (LISA) e da *Library, Information Science and Technology Abstracts*<sup>61</sup> (LISTA). A diversidade de fontes na recolha de dados tem por finalidade aumentar a precisão das informações recolhidas e, conseqüentemente, a fiabilidade da descrição apresentada.

Em relação aos conteúdos, a recolha dos *abstracts* a consultar tem por base uma consulta direta em bases de dados bibliográficas especializadas, de alcance e cobertura internacionais, responsáveis pela indexação e resumo de uma quantidade razoável de publicações influentes no campo da Ciência da Informação — mais especificamente, a LISA e a LISTA. De seguida, os textos completos dos artigos científicos foram recuperados através da consulta dos próprios *websites* das revistas, das editoras ou de repositórios institucionais, mediante o privilégio o acesso aos recursos bibliográficos subscritos pela Universidade de Coimbra.

No que diz respeito ao período de consulta, o principal critério foi o intervalo cronológico de uma década (2009-2019) definido para a investigação. Com base nesta baliza temporal, procurou-se averiguar quais os períodos de atividade das revistas, qual a sua cobertura nas bases de dados consultadas, e qual a disponibilidade de acesso aos textos completos.

---

<sup>59</sup> <https://www.issn.org/>

<sup>60</sup> <http://miar.ub.edu/>

<sup>61</sup> <https://www.ebsco.com/products/research-databases/library-information-science-and-technology-abstracts> [requer credenciais de acesso].

Por fim, as perspectivas destes esforços face aos objetivos primários revestem-se de um ênfase exploratório, ancorado numa pesquisa bibliográfica que pretende retratar uma análise tendencial, na literatura mais recente, dos temas sob discussão. Uma exploração inicial das fontes de informação, aqui não documentada, permitiu estabelecer categorias de análise que auxiliaram na definição de temas e tendências que foram permitindo afinar as questões de investigação e redigir a revisão de literatura.

### 3.4.1 Revistas científicas

A escolha das revistas científicas como base metodológica para a constituição do *corpus* essencial do projeto assenta na crença de que se trata de instrumentos capazes de responder com precisão às questões de investigação, permitindo-nos concretizar os objetivos propostos. Mais especificamente, as revistas têm servido para identificar, selecionar e analisar fontes de informação primárias relevantes para o nosso tema, estabelecendo critérios precisos que assegurem que os títulos recuperados constituem fontes de informação pertinentes no âmbito do enquadramento da investigação.

Conforme previamente mencionado, na base desta escolha subjaz a assunção de que as revistas especializadas, enquanto instrumentos de divulgação do conhecimento, quando veem cumpridos os critérios de qualidade estabelecidos pela comunidade, têm legitimidade suficiente para servir como representação do que melhor se faz numa determinada área científica. Assim, têm o privilégio de servir como instrumento de reconhecimento de temáticas, linhas e tendências contemporâneas de investigação, permitindo enquadrar a nossa investigação no seio das discussões mais recentes e com maior impacto nos canais especializados de divulgação do conhecimento em Arquivística.

A principal tipologia documental veiculada pelas revistas científicas são os artigos científicos. No âmbito deste estudo, entende-se que os artigos científicos são produção científica, vertida predominantemente em texto, divulgados principalmente por revistas científicas com o objetivo de divulgar os resultados de investigação, quer sejam originais ou de revisão<sup>62</sup>, sendo submetidos a uma apreciação crítica por parte dos pares antes da sua publicação. Neste sentido, posicionam-se como instrumentos atualizados do conhecimento numa área científica. Este posicionamento é consistente com o de Borges (2006), para os quais os artigos científicos constituem uma das fontes privilegiadas no mundo científico, na medida em que apresentam dados de pesquisa novos e consistentes, ao mesmo tempo que são sujeitos a um controlo antes da sua publicação, o que nem sempre ocorre noutros canais de comunicação da ciência.

---

<sup>62</sup> De acordo com Borrego Huerta (1999), artigos originais ou de investigação são textos publicados e assinados que apresentam dados empíricos de estudo baseados num emprego de métodos e técnicas. Os artigos de discussão, embora igualmente publicados e assinados, não apresentam necessariamente dados novos ou métodos específicos, no sentido em que têm a finalidade de esclarecer sobre o estado da arte de um determinado tema examinado.

Tendo presentes estes conceitos, identificam-se duas bases de dados de referência predominantes que incluem a produção científica em Arquivística: a *Library & Information Science Abstracts*<sup>63</sup> (LISA) e a *Library, Information Science & Technology Abstracts*<sup>64</sup> (LISTA). Ambas têm vários pontos comuns: (i) apenas agregam *abstracts*, sem acesso ao texto completo; (ii) fornecem metadados descritivos sobre as revistas, bem como sobre os artigos científicos e revisões de livros nelas incluídos; (iii) são bases de dados especializadas que incluem muitas das principais revistas científicas em Arquivística. Além disso, também são ambas detidas pela mesma empresa americana, a ProQuest<sup>65</sup>. Apesar destas semelhanças, também apresentam alguns pontos divergentes. Ao longo das próximas secções faremos um esforço por caracterizar e enquadrar no contexto da nossa investigação cada uma destas bases de dados, tendo por base critérios como os títulos de revistas disponíveis, a sua cobertura, atividade, índice de difusão, entre outros. Por fim, apresentar-se-ão os argumentos que presidiram à escolha da LISTA como sustento metodológico da investigação.

#### 3.4.1.1 *Library & Information Science Abstracts*

Conforme mencionado, a LISA é uma base de dados de referências que disponibiliza para consulta os *abstracts* de várias publicações científicas. Trata-se, portanto, de uma ferramenta internacional especializada de indexação de *abstracts*, com um foco temático nos profissionais da informação, nos quais se incluem os bibliotecários e os arquivistas. Segundo informações oficiais<sup>66</sup> recolhidas em 2019, a LISA reúne cerca de 440 revistas de 68 países, em 20 línguas diferentes, sendo atualizada quinzenalmente.

A abordagem metodológica foi iniciada por uma tentativa de identificação das revistas mais influentes em Arquivística. Para isso, consultou-se em primeiro lugar a *Matriz de Información para el Análisis de Revistas*<sup>67</sup> (MIAR), uma ferramenta colaborativa gerida por investigadores da Universidade de Barcelona, em Espanha, que fornece informações relevantes não só sobre a visibilidade das revistas em bases de dados de indexação e de referência, mas também sobre instrumentos de avaliação, como a presença no *Scimago Journal & Country Rank*<sup>68</sup>. No entanto, o indicador disponibilizado pelo MIAR considerado como mais pertinente para os nossos interesses é o *Índice Compuesto de Difusión Secundaria* (ICDS), que reflete a visibilidade de uma determinada revista científica em diferentes bases de dados científicas de alcance internacional.

---

<sup>63</sup> <https://www.proquest.com/products-services/lisa-set-c.html>

<sup>64</sup> <https://www.ebsco.com/products/research-databases/library-information-science-and-technology-abstracts>

<sup>65</sup> <https://www.proquest.com/about/who-we-are.html>

<sup>66</sup> <https://www.proquest.com/products-services/lisa-set-c.html>

<sup>67</sup> <http://miar.ub.edu/idioma/en>

<sup>68</sup> <https://www.scimagojr.com/>

Assim, um ICDS elevado significa que a revista assinala presença em diferentes fontes de informação relevantes a nível internacional, o que pode reforçar a escolha desse título como uma fonte de informação pertinente.

Neste sentido, o primeiro passo consistiu em efetuar uma pesquisa exploratória na base de dados do MIAR dentro do seguinte critério:

Subject = Library & Information Science

Esta primeira pesquisa, efetuada a 2 de maio de 2019, recuperou um total 1029 resultados. Tratando-se naturalmente de um valor bruto de material incomportável para análise, tornou-se necessário refinar os resultados usando os seguintes critérios:

Title = "archiv\*", "arq\*" ou "records management"

ICDS = 5-11

A opção pela delimitação do título pretende centrar o foco das revistas na Arquivística de modo a não mergulharmos em discussões alheias ao nosso campo científico. A limitação dos termos do título vem reforçar esta intenção, recuperando revistas científicas que digam respeito estritamente à Arquivística. Neste sentido, a expressão truncada 'archiv\*' vota-se a recuperar títulos de revistas em francês e em inglês (*archives*) e em espanhol (*archivos*). A expressão 'arq\*' destina-se a títulos portugueses; e 'records management' tem por vista as revistas de origem anglo-saxónica, contexto no qual a tradição arquivística assume essa designação. Por fim, a delimitação do ICDS entre valores de 5 e 11 prende-se com um critério de qualidade. Quanto mais alto o ICDS, mais pertinentes são consideradas as revistas. Neste sentido, entendeu-se que não seria proveitoso examinar revistas cujo ICDS fosse menor que 6.5, razão pela qual se estabeleceu este limite inferior. Na prática, por imposições de pesquisa pela plataforma, os intervalos selecionados na base de dados foram "9 <= ICDS <= 11", "7 <= ICDS <= 9", e "5 <= ICDS <= 7".

Com base nestes critérios, obteve-se um valor mais útil de 25 revistas científicas especializadas, consideradas de potencial alto interesse para os propósitos da investigação (Tabela 10).

Tabela 10 — Identificação de revistas potencialmente pertinentes

ISSN	Título	ICDS
0795-4778	African Journal of Library, Archives and Information Science	10.9
0360-9081	American Archivist	10.0
1389-0166	Archival Science	9.8
0318-6954	Archivaria	10.0
0003-9535	Archives	10.0
0157-6895	Archives & Manuscripts	10.0
1042-1467	Archives & Museum Informatics	10.0
1533-2748	Journal of Archival Organization	9.7

0016-5522	La Gazette des Archives	10.0
0033-1031	Prologue: Quarterly of the National Archives and Records Administration	10
0956-5698	Records Management Journal	10.0
0303-7940	Archifacts	6.5
0066-6327	Archiv für Geschichte des Buchwesens	6.5
0003-9233	Archiv für Kulturgeschichte	6.5
1067-4993	Archival Issues	6.5
0003-9748	Archives et Bibliothèques de Belgique	6.5
0044-9423	Archives (quebec)	6.5
0004-0398	Archivni Casopis	6.5
0210-4067	Archivo Hispalense	6.5
0004-0703	Archivum Linguisticum; A Review Of Comparative Philology And General Linguistics	6.5
0037-9816	Journal of the Society of Archivists	6.5
0196-0075	Library & Archival Security	6.5
0873-5670	Páginas a&b: arquivos e bibliotecas	6.3
1050-2343	Records Management Quarterly	6.5
1012-2796	S.A. Archives Journal	6.5
<b>Total</b>		<b>25</b>

FONTE: elaboração própria, com base em informações recolhidas no MIAR.

A partir deste conjunto exploratório, determinaram-se requisitos adicionais de exclusão que ajudaram a filtrar ainda mais o universo de unidades a analisar:

- a. *ICDS < 6.5;*
- b. *Não ser peer-reviewed;*
- c. *Não estar indexada na lista de títulos da LISA, pesquisando pelo ISSN;*
- d. *Não apresentar informação descritiva para consulta;*
- e. *Não estar em atividade;*
- f. *Não ter um idioma legível por nós;*
- g. *Ser considerada fora do escopo.*

Os resultados da aplicação destes critérios são apresentados graficamente na Tabela 11. Concretamente, foi eliminada por ter uma classificação demasiado baixa a revista portuguesa *Páginas a&b: arquivos e bibliotecas*. Pelo terceiro critério, foram excluídas as seguintes nove revistas: *African Journal of Library, Archives and Information Science, Archives & Museum Informatics, La Gazette des Archives, Prologue: Quarterly, Archival Issues, Archives et bibliothèques de Belgique, Archivo Hispalense, Records Management Quarterly* e *S.A. Archives*

*Journal*. Por estar inativa, foi afastada a *Archifacts*. Por não lermos alemão nem checo, foram excluídas as *Archiv für Geschichte des Buchwesens*, *Archiv für Kulturgeschichte* e *Archivni Casopis*. Por fim, por ser considerada fora do âmbito do estudo, foram excluídas as *Archivum Linguisticum* e *A review of comparative Philology and General Linguistics*.

Tabela 11 — Revistas excluídas com base nos critérios definidos.

ISSN	Título	ICDS
0795-4778	<del>African Journal Of Library, Archives And Information Science</del>	10.9
0360-9081	American Archivist	10.0
1389-0166	Archival Science	9.8
0318-6954	Archivaria	10.0
0003-9535	Archives	10.0
0157-6895	Archives & Manuscripts	10.0
1042-1467	<del>Archives &amp; Museum Informaties</del>	10.0
2325-7962	Archives and Records	11.0
1533-2748	Journal of Archival Organization	9.7
0016-5522	<del>La Gazette des Archives</del>	10.0
0033-1031	<del>Prologue: Quarterly of the National Archives and Records Administration</del>	10
0956-5698	Records Management Journal	10.0
0303-7940	<del>Archifacts</del>	6.5
0066-6327	<del>Archiv Für Geschichte Des Buchwesens</del>	6.5
0003-9233	<del>Archiv Für Kulturgeschichte</del>	6.5
1067-4993	<del>Archival Issues</del>	6.5
0003-9748	<del>Archives Et Bibliothèques De Belgique</del>	6.5
0044-9423	Archives (Quebec)	6.5
0004-0398	<del>Archivni Casopis</del>	6.5
0210-4067	<del>Archivo Hispalense</del>	6.5
0004-0703	<del>Archivum Linguisticum; A Review Of Comparative Philology And General Linguistics</del>	6.5
0196-0075	Library & Archival Security	6.5
0873-5670	<del>Páginas A&amp;B: Arquivos E Bibliotecas</del>	6.3
1050-2343	<del>Records Management Quarterly</del>	6.5
1012-2796	<del>S.A. Archives Journal</del>	6.5
<b>Total</b>		<b>10</b>

FONTE: elaboração própria, com base em informações recolhidas a partir do MIAR, da LISA e dos *websites* das revistas.

A partir desta seleção obtemos uma lista final de nove títulos considerados muito pertinentes para a nossa investigação (Quadro 1), considerados incontornáveis para a abordagem

metodológica, caso se opte por usar a LISA como base de dados de referência para as pesquisas. É de referir que se apresentam nove títulos, em vez dos 10 ‘sobreviventes’ da Tabela 11, uma vez que a *Journal of Archival Organization* (2002-presente) incorpora a *Library & Archival Security* (1978-2014). Embora tanto no MIAR como na LISA os títulos destas revistas surjam em separado, parece-nos mais adequado para fins de simplificação incluir apenas o título mais recente. Numa segunda nota, a revista *Archives & Museum Informatics* (1987-2000) foi também incorporada na *Archival Science* (2000-presente).

Quadro 1 — Lista final de títulos pertinentes obtidos através da LISA.

American Archivist
Archival Science
Archivaria
Archives (London)
Archives & Manuscripts
Archives And Records
Journal of Archival Organization
Records Management Journal
Archives (Quebec)
<b>Total: 9</b>

FONTE: Elaboração própria.

#### 3.4.1.2 *Library, Information Science & Technology Abstracts*

À semelhança da LISA, a LISTA é uma base de dados internacional de referência que providencia o acesso a *abstracts*. Especializa-se igualmente em Ciência da Informação e Biblioteconomia, indexando 560 revistas consideradas fundamentais, bem como revisões de livros e *research reports*<sup>69</sup>. Além disso, a cobertura da base de dados estende-se desde meados da década de 1960, embora este aspeto seja de menor relevância uma vez que o intervalo temporal de análise da bibliografia estipulado pelos objetivos se situa nos últimos 10 anos (2009-2019).

Uma vez que se possui acesso a esta base de dados via subscrição pela Universidade de Coimbra, desta vez o primeiro passo consistiu em obter uma seleção prévia de revistas científicas versadas diretamente em Arquivística. Para isso, efetuou-se uma pesquisa exploratória com o seguinte critério anteriormente adotado:

---

<sup>69</sup> Mais informações podem ser consultadas em: [http://support.ebsco.com/help/?int=ehost&lang=en&feature\\_id=Databases&TOC\\_ID=Always&SI=0&BU=0&GU=1&PS=0&ver=live&dbs=lxhjnh,lxh](http://support.ebsco.com/help/?int=ehost&lang=en&feature_id=Databases&TOC_ID=Always&SI=0&BU=0&GU=1&PS=0&ver=live&dbs=lxhjnh,lxh).

Title = "arch\*" ou "arq\*" ou "records management"

Esta pesquisa inicial permitiu recuperar um total de 20 títulos, apresentados na Tabela 12. Também se recolheram informações acerca da cobertura das revistas na LISTA, bem como se as revistas são sujeitas a revisão por pares e qual a sua frequência anual de publicação, uma vez que estes dados serão úteis na hora de tomada de decisão sobre a inclusão, ou não, das revistas no *corpus* do estudo.

Tabela 12 — Identificação de revistas de Arquivística indexadas na LISTA.

ISSN	Título	Cobertura	Peer-reviewed	Frequência
0360-9081	American Archivist	2004-presente	S	2
0303-7940	Archifacts	2004-presente	N	2
1067-4993	Archival Issues: Journal of the Midwest Archives Conference	2001-presente	S	2
1389-0166	Archival science	2003-presente	S	4
2199-9252	Archivar	2008-presente	S	4
0318-6954	Archivaria	2003-presente	S	2
0003-9535	Archives [London]	2004-presente	S	2
0044-9423	Archives [Quebec]	2004-2013	S	2
0157-6895	Archives & manuscripts	2004-presente	S	2
2325-7962	Archives & Records	1967-presente	S	2
1121-2462	Archivi & Computer	2008-2013	N	2
0005-3856	Baessler-archiv (ger)	2008-presente	S	1
0007-9421	Cadernos de Biblioteconomia, Arquivística e Documentação	2006-presente	S	2
1945-8533	Journal for the Society of North Carolina Archivists	2007-presente	S	2
1533-2748	Journal of Archival Organization	2002-presente	S	4
2033-639X	META: Tijdschrift voor Bibliotheek & Archief	2011-presente	N	9
0739-4241	Provenance: The Journal of the Society of Georgia Archivists	2009-presente	S	1
0956-5698	Records Management Journal	2001-presente	S	3
1746-9465	Information & Records Management Society Bulletin	2010-presente	N	6
1012-2796	S. A. Archives Journal	1990-2003	S	1
<b>Total</b>				<b>20</b>

FONTE: elaboração própria, segundo informações recolhidas a partir da LISTA e dos *websites* das revistas.



De seguida, à semelhança dos passos tomados a respeito da LISA, revelou-se necessário proceder a uma seleção destes títulos segundo critérios formais de qualidade:

- a. ICDS => 6.5
- b. ser peer-reviewed
- c. estar em atividade
- d. estar redigida numa língua legível pelos investigadores (português, espanhol, inglês, francês)
- e. estar enquadrada no âmbito temático da investigação
- f. disponibilizar informação descritiva para consulta

Aplicando estes critérios, foram excluídas sete revistas científicas, conforme demonstra a Tabela 13. De entre as revistas excluídas, indica-se que não foi incluído o *Journal for the Society of North Carolina Archivists* por apresentar um ICDS inferior a 6.5. Por não estar sujeita à revisão por pares, foram excluídas a *Archifacts* e a *Information & Records Management Society Bulletin*. Por estar inativa, a *Archivi & Computer*. Finalmente, por se encontrarem em idiomas fora do nosso entendimento, optou-se por não incluir a alemã *Baessler-archiv* e a belga *META: Tijdschrift voor Bibliotheek & Archief*.

Tabela 13 — Seleção de revistas identificadas na LISTA segundo os critérios estabelecidos.

ISSN	Título	Cobertura	Peer-reviewed	Frequência	ICDS
0360-9081	American Archivist	2004-presente	S	2	10
<del>0303-7940</del>	<del>Archifacts</del>	<del>2004-presente</del>	<del>N</del>	<del>2</del>	<del>6.5</del>
1067-4993	Archival Issues: Journal of the Midwest Archives Conference	2001-presente	S	2	6.5
1389-0166	Archival science	2003-presente	S	4	9.8
<del>0003-9500</del>	<del>Archivar</del>	<del>2008-presente</del>	<del>S</del>	<del>4</del>	<del>6.5</del>
0318-6954	Archivaria	2003-presente	S	2	10
0003-9535	Archives [London]	2004-presente	S	2	10
0044-9423	Archives [Quebec]	2004-2013	S	2	6.5
0157-6895	Archives & Manuscripts	2004-presente	S	2	10

2325-7962	Archives & Records	1967- presente	S	2	11
<del>1121-2462</del>	<del>Archivi &amp; Computer</del>	<del>2008-2013</del>	<del>N</del>	<del>2</del>	<del>4.4</del>
<del>0005-3856</del>	<del>Baessler Archiv</del>	<del>2008- presente</del>	<del>S</del>	<del>1</del>	<del>6.5</del>
0007-9421	Cadernos de Biblioteconomia, Arquivística e Documentação	2006- presente	S	2	6.5
<del>1945-8533</del>	<del>Journal for the Society of North Carolina Archivists</del>	<del>2007- presente</del>	<del>S</del>	<del>2</del>	<del>3.0</del>
1533-2748	Journal of Archival Organization	2002- presente	S	4	9.7
<del>2033-639X</del>	<del>META: Tijdschrift voor Bibliotheek &amp; Archief</del>	<del>2011- presente</del>	<del>N</del>	<del>9</del>	<del>6.5</del>
0956-5698	Records Management Journal	2001- presente	S	3	10
<del>1746-9465</del>	<del>Information &amp; Records Management Society Bulletin</del>	<del>2010- presente.</del>	<del>N</del>	<del>6</del>	<del>6.5</del>
0739-4241	Provenance: The Journal of the Society of Georgia Archivists	2009- presente	S	1	6.5
<del>1012-2796</del>	<del>S. A. Archives Journal</del>	<del>1990-2003</del>	<del>S</del>	<del>1</del>	<del>6.5</del>
<b>Total</b>					<b>12</b>

FONTE: elaboração própria, segundo informações recolhidas a partir da LISTA e dos *websites* das revistas.

Além disso, também foi excluída a revista *S.A. Archives Journal*. Embora ativa, o período de cobertura desta revista na LISTA insere-se entre 1990 e 2003, ficando de fora do intervalo de análise do nosso estudo (2009-2019). Como tal, não nos pareceu pertinente a sua inclusão na lista de revistas a analisar. Neste sentido, a identificação e seleção de revistas científicas mediante a LISTA chegou-se a um conjunto final de 12 títulos considerados muito pertinentes para os nossos objetivos de investigação (Tabela 14).

Tabela 14 — Lista final de títulos pertinentes obtidos através da LISTA.

ISSN	Título	ICDS
0360-9081	American Archivist	10

1067-4993	Archival Issues: Journal of the Midwest Archives Conference	6.5
1389-0166	Archival Science	9.8
0318-6954	Archivaria	10
0003-9535	Archives [London]	10
0044-9423	Archives [Quebec]	6.5
0157-6895	Archives & Manuscripts	10
2325-7962	Archives & Records	11
0007-9421	Cadernos de Biblioteconomia, Arquivística e Documentação	6.5
1533-2748	Journal of Archival Organization	9.7
0739-4241	Provenance: The Journal of The Society of Georgia Archivists	6.5
0956-5698	Records Management Journal	10
<b>Total</b>		<b>12</b>

FONTE: elaboração própria.

De seguida serão apresentadas as vantagens e desvantagens do uso de cada uma das bases de dados de referência de forma a argumentar de forma lógica e consistente a nossa escolha da LISTA como base de dados de referência para a consulta de revistas científica com vista à recuperação de artigos científicos e de revisões bibliográficas representativos das discussões mais recentes em Arquivística.

#### 3.4.1.3 A escolha da base de dados de referência

Com a discussão que se segue não se pretende entrar numa comparação acérrima da individualidade de cada base de dados, mas tão-só na comparação das suas principais características, e a interpretação da sua adequação, ou não, aos nossos objetivos de investigação e ao nosso contexto geral enquanto investigadores.

Por uma questão de clareza de raciocínio, apresenta-se de forma visualmente confortável as principais vantagens e desvantagens que se consideram que cada base de dados de referência apresenta (Quadro 2).

Quadro 2 — Comparação da LISA com a LISTA.

Vantagens	Desvantagens
<b>LISA</b>	
+ mais prestigiada base de dados de referência em <i>Library &amp; Information Science</i>	- menos revistas elegíveis para consulta após aplicados os critérios de seleção (9)  - consulta dificultada por falta de credenciais de acesso

---

## LISTA

---

+ acesso imediato e fácil via Universidade de Coimbra	- potencialmente menos prestígio do que a LISA
+ mais revistas indexadas, depois de aplicados os critérios de seleção (12)	
+ enfoque temático mais alargado, incluindo <i>Technology</i>	

---

FONTE: Elaboração própria.

A primeira intenção de consulta centrou-se na LISA, uma vez que esta base de dados é reconhecida como sendo a base de dados de referência de maior prestígio em Ciência da Informação, onde se espera encontrar indexadas algumas das revistas mais influentes nesta área. No entanto, a pesquisa exploratória recuperou apenas nove títulos relevantes para consulta, uma vez aplicados os critérios de pesquisa e filtrada a amostra. Além disso, também se constatou que, em verdade, não dispúnhamos da possibilidade de consulta dessa base de dados, uma vez que não é um recurso subscrito nem pela Universidade de Coimbra, nem por outras universidades portuguesas que outrora frequentadas (Universidade do Minho e Universidade de Lisboa).

Devido a estas dificuldades, considerou-se a hipótese de consulta da LISTA. Após aplicados os mesmos critérios de seleção formulados para a pesquisa na LISA, concluiu-se que a LISTA apresenta 12 títulos relevantes para a nossa investigação, o que nos parece ser um volume aceitável. Além disso, esta base de dados tem a característica de também incluir textos da área da tecnologia, o que parece ser uma vantagem e uma oportunidade devido à ramificação tecnológica do tema de investigação — não é plausível falar em representação da informação de uma forma alheia e desinteressada da tecnologia existente, uma vez que a disponibilidade desta influencia de sobremaneira a primeira. Por fim, também constitui um aspeto determinante o facto de a Universidade de Coimbra subscrever este recurso, possibilitando uma consulta fácil e direta a esta base de dados.

Dentro dos nossos critérios de pesquisa, ficou demonstrado que a LISTA não só inclui todas as revistas que a LISA recuperou, como ainda acrescenta mais três. É certo que quantidade não significa qualidade, mas a qualidade dos títulos recuperados é assegurada pelo escrutínio dos critérios de pesquisa adotados. Por este motivo, entende-se que é vantajoso para uma análise tendencial da literatura nos últimos 10 anos, como é a presente, que se inclua o maior volume possível de revistas científicas especializadas com publicação ativa neste período.

Um último aspeto que era importante esclarecer antes da tomada de decisão dizia respeito a dois aspetos cruciais para o futuro da investigação: 1) a cobertura das revistas selecionadas na LISTA; e 2) a disponibilidade de acesso em texto completo aos artigos científicos e revisões publicadas nessas mesmas revistas. Na Tabela 13 é possível verificar que o período de cobertura

nas revistas sobreviventes é sempre anterior a 2009, e decorre até ao presente. Isto enquadra-se perfeitamente no intervalo temporal definido para a análise, compreendido entre 2009 e 2019.

De modo a verificar a disponibilidade em texto integral dos recursos selecionados, foi realizado um teste por amostragem aleatória. O teste consistiu no seguinte: usando o *interface* de pesquisa da LISTA, foi selecionado como único termo de pesquisa o ISSN de uma determinada revista, recuperando todos os resultados compreendidos entre 2009 e 2019. A coleção obtida representa todos os artigos publicados por essa revista entre 2009 e 2019, que se encontram indexados na LISTA. Trata-se, portanto, de uma fração do universo de análise deste estudo. De seguida, foram escolhidos aleatoriamente dois artigos científicos, cujos textos completos se tentaram obter, assinalando-se a respetiva via de acesso e consulta. Este exercício foi repetido para cada uma das 12 revistas selecionadas (Tabela 15).

Tabela 15 – Condições de acesso aos textos integrais de artigos aleatoriamente selecionados.

ISSN	Title	Acesso
0360-9081	American Archivist	Texto completo via website da revista
1067-4993	Archival Issues: Journal of the Midwest Archives Conference	Texto completo via repositórios institucionais (University of Dayton eCommons e Illinois State University)
1389-0166	Archival Science	Texto completo via editora (Springer)
0318-6954	Archivaria	Texto completo via website da revista
0003-9535	Archives [London]	Texto completo via repositório institucional (Liverpool university press online)
0044-9423	Archives [Quebec]	Texto completo via website da revista
0157-6895	Archives & Manuscripts	Texto completo via editora (Tandfonline)
2325-7962	Archives & Records	Texto completo via editora (Tandfonline)
0007-9421	Cadernos de Biblioteconomia, Arquivística e Documentação	Texto completo via website da revista
1533-2748	Journal of Archival Organization	Texto completo via editora (Tandfonline)

0739-4241	Provenance: The Journal of the Society of Georgia Archivists	Texto completo via repositório institucional (digitalcommons.kennesaw.edu)
0956-5698	Records Management Journal	Texto completo via editora (Emerald Insight)
<b>Total</b>		<b>12</b>

FONTE: elaboração própria.

A partir destes resultados, é possível verificar que se obtiveram os textos integrais de todos os artigos selecionados, o que significa que todas as revistas selecionadas são fontes de informação viáveis e realistas para o desenvolvimento da investigação. É de destacar que a pesquisa ocorreu procurando os títulos dos artigos na *web*, tendo o acesso sido facilitado principalmente pela editora (5 itens ou 41%), seguido pelos próprios sites das revistas científicas (4 itens ou 33%) e por repositórios institucionais (3 itens ou 25%).

Por todas estes motivos, pesando as vantagens e desvantagens de cada uma, a LISTA foi eleita como a base de dados de referência para a identificação e seleção de revistas científicas. A partir dos títulos identificados, procedeu-se a uma análise de cada revista de modo a extrair artigos científicos e revisões de livros considerados pertinentes para responder aos objetivos de investigação. A metodologia de identificação e seleção destas fontes será explicada adiante.

Por último, parece-nos útil apresentar uma breve caracterização de cada uma das revistas selecionadas (Tabela 16), de forma a compreender a sua procedência, periodicidade de publicação, tipo de arbitragem científica a que estão sujeitas, a sua taxa de pervivência, enquanto sinónimo de prestígio e estabilidade e, por fim, os indicadores métricos em que as publicações estão incluídas.

Tabela 16 — Caracterização das revistas selecionadas.

ISSN	Título	Procedência	Periodicidade	Tipo de arbitragem	de Pervivência/ antiguidade	Inclusão
0360-9081	American Archivist	USA	Semestral	Double blind	+1.5/81/1938	Scimago, Scopus
1389-0166	Archival science	Holanda	Trimestral	Peer reviewed	+1.3/19/2000	Scimago, Scopus
0318-6954	Archivaria	Canadá	Semestral	Double blind	+1.5/44/1975	Scimago, Scopus
0003-9535	Archives (london)	Reino Unido	Semestral	Peer reviewed	+1.5/70/1949	Scimago, Scopus
0157-6895	Archives & manuscripts	Austrália	Trimestral	Double blind	+1.5/64/1955	Scimago, scopus

2325-7962	Archives and records: The Journal of the Archives and Records Association	Reino Unido	Trimestral	Double blind	+1.5/64/1955	Scimago, Scopus
1533-2748	Journal of archival organization	USA	Quadrimestral	Double-blind	+1.2/17/2002	Scimago, Scopus
0956-5698	Records management journal	Reino Unido	Trimestral	Double blind	+1.5/30/1989	Scimago, Scopus
0044-9423	Archives (quebec)	Canadá	Semestral	Double blind	+1.5/50/1969	-

FONTE: elaboração própria com base em informações recolhidas a partir do MIAR e das páginas das revistas. Dados válidos para maio de 2019.

### 3.4.2 Monografias

Segundo Peter Mann (1980), *monographs are conventionally books which report specialized research work to a restricted readership. The publishing of such books requires special publishing skills of both an editorial and financial type* (Mann, 1980, p. 1). Por outras palavras, as monografias veiculam conhecimento científico altamente especializado para um público igualmente especializado, motivo pelo qual consideramos a sua inclusão incontornável no processo de recolha de fontes de informação.

As monografias foram identificadas através de resenhas publicadas nas revistas científicas previamente identificadas. A metodologia consistiu em pesquisar no MIAR cada revista selecionada pelo seu ISSN, estabelecer como intervalo cronológico 2009-2019, e escolher *Book review* como tipo de fonte. Uma vez estabelecida uma lista de resenhas a monografias presentes em cada uma das 12 revistas, estas foram analisadas e escrutinadas de modo a avaliar a sua pertinência para os objetivos de investigação.

### 3.4.3 Comunicações científicas

Houve um período em que as atas dos congressos eram geralmente consideradas veículos de comunicação de ciência de menor prestígio em comparação com os artigos publicados em revistas científicas ou em monografias (Mueller, Campello, & Dias, 1996), uma vez que, em favor de uma maior facilidade de comunicação oral, abdicariam da qualidade e rigor normalmente exigidos aos outros canais de comunicação. Por este motivo, foram no passados frequentemente consideradas como *fontes de informação de natureza parcial e inacabada* (Freitas, 2009, p. 123).

Porém, esta perspetiva tem vindo a alterar-se ao longo dos últimos anos. Num inquérito a 1832 investigadores da Universidade de Coimbra pertencentes a 39 centros de investigação de diversas áreas avaliados pela Fundação para a Ciência e Tecnologia, Maria Manuel Borges identificou que as atas de reunião científica figuravam como o segundo formato de publicação mais frequente (56%) pelos inquiridos, atrás apenas da publicação em revistas científicas internacionais arbitradas (85%) (Borges, 2006, p. 245). Mais recentemente, esta tendência é confirmada por Meho (2019) que salienta o papel cada vez mais crucial desempenhado pelas conferências em determinadas áreas científicas, nomeadamente na Ciência da Computação, na qual mais de 60% dos resultados de investigação são publicados em atas de conferência. Por sua vez, Vrettas e Sanderson (2015) assinalam que um número reduzido de conferências de elite apresentam, em média, maior número de citações do que revistas científicas de topo. Estes resultados não são completamente inesperados, uma vez que uma das vertentes mais importantes da ciência é a comunicação, e a partilha de ideias de forma presencial é uma das formas mais eficazes de comunicação. Por estes motivos, considera-se de certa forma natural que as publicações em atas de conferência tenham vindo a adquirir maior legitimidade e relevância enquanto veículos de ciência, assumindo atualmente um estatuto que não deve ser ignorado no âmbito de uma revisão de literatura sistemática como é a do presente trabalho.

Devido a estas razões, optamos por incluir esta tipologia documental no universo de recolha de fontes, explicando-se os critérios para a identificação e seleção deste tipo de fontes. A metodologia assenta na consulta da base de dados *Conference Proceedings Citation Index*<sup>70</sup>, que integra a *core collection* da base de dados da referência *Web of Science* (WoS). Trata-se de um serviço de indexação de referências bibliográficas originalmente criado pelo *Institute for Scientific Information* (ISI), tendo vindo a ser posteriormente gerido pela Clarivate Analytics. A escolha justifica-se pela completude e alcance internacional deste recurso, além de cumprir a cobertura pretendida<sup>71</sup>.

Uma primeira pesquisa foi efetuada com os seguintes critérios:

SU = "Information Science & Library Science"

Type: Proceedings Paper

Date: 2009–2019

Language: English OR Portuguese OR Spanish

Databases:

- Conference Proceedings Citation Index (CPCI-S) 1990–presente

---

<sup>70</sup> <http://apps.webofknowledge.com/>

<sup>71</sup> De acordo com informações disponibilizadas pela Clarivate Analytics, trata-se de um índice multidisciplinar que agrega mais de 180 000 *conference proceedings*, desde 1990 até ao presente. Informações disponíveis em <<https://clarivate.com/products/web-of-science/databases/>>.



- Conference Proceedings Citation Index- Social Science & Humanities  
(CPCI-SSH)

A partir desta pesquisa inicial realizada no dia 18 de novembro de 2019 foram obtidos 12 517 artigos, que constitui o número total de comunicações científicas nesta base de dados, nesta área, nestas línguas e neste período. De forma a refinar este universo, os resultados foram filtrados pesquisando no ‘topic’<sup>72</sup> as seguintes palavras-chave:

Operador booleano comum a todos os termos: **OR**

Archival description

Descripción archivística

Information representation

Representación de la información

Metadata

Metadatos

Trustworthiness

Confiabilidad

Authenticity

Autenticidad

Linked data

AND archiv\*

A escolha destas palavras-chave tem por premissa dois raciocínios: recuperar comunicações científicas (1) em temas relacionados com os objetivos de investigação; e (2) nas línguas inglesa e espanhola. Por este motivo, os conceitos nascem das questões de investigação, ao passo que a forma linguística que os conceitos tomam revela o código linguístico do inglês e do espanhol. Foi ainda estabelecido como requisito que todas as comunicações recuperadas contivessem ‘archiv\*’ no ‘topic’ de forma a centrar a discussão na arquivística. O asterisco permite recuperar qualquer variante de termos que contenham ‘archiv’, podendo recuperar termos em inglês tais como ‘archival’, ‘archive’ ou ‘archives’, bem como termos em espanhol como ‘archivística’, ‘archivo’ ou ‘archivos’.

De forma recuperar comunicações em língua portuguesa, realizou-se uma pesquisa com as seguintes palavras-chave:

Operador booleano comum a todos os termos: **OR**

Autenticidade

Metainformação

arquiv\*

<sup>72</sup> Na WoS, o ‘topic’ inclui: ‘title’, ‘abstract’, ‘author keywords’, ‘keywords plus’.

Metadados

Linked data

AND

Representação da informação

Confiabilidade

Descrição arquivística

A necessidade de uma pesquisa em separado para a língua portuguesa justifica-se pelo facto de o termo ‘arquivo’ apresentar um étimo único ao português, considerando o universo destas três línguas. Assim, a escolha da expressão ‘arquiv\*’ pretende igualmente centrar o tema na Arquivística, recuperando tanto ‘arquivística’ como ‘arquivo’ ou ‘arquivos’

Em síntese, a pesquisa foi executada mediante a seguinte expressão:

(SU= (Information Science & Library Science)) AND LANGUAGE: (English OR Portuguese OR Spanish) AND DOCUMENT TYPES: (Proceedings Paper)

AND

(TS=(archival description OR descripción archivística OR information representation OR representación de la información OR metadata OR metadatos OR trustworthiness OR confiabilidad OR authenticity OR autenticidad OR linked data) AND TS=archiv\*) AND LANGUAGE: (English OR Portuguese OR Spanish) AND DOCUMENT TYPES: (Proceedings Paper)

O conjunto de documentos recuperados é apresentado numa secção adiante. Importa destacar que a WoS desempenha o papel de agregador de metadados, pelo que não disponibiliza os textos integrais das comunicações. O acesso aos textos é realizado por sites externos, como as páginas das editoras, pelo que a livre consulta de todas as fontes de informação não é garantida.

#### 3.4.4 Teses e dissertações

Para efeitos de definição, consideram-se teses e dissertações como escritos que tendem a assumir o carácter monográfico de um teor altamente especializado, apresentadas publicamente perante um júri arguente, com vista à obtenção do grau de “doutor” ou “mestre”, respetivamente. A sua aprovação, armazenamento e distribuição são competências da instituição académica de ensino superior onde o grau decorreu (Freitas, 2009, p. 126).

De forma a localizar teses e dissertações de interesse, usaram-se os recursos *Open Access theses and dissertations*<sup>73</sup> e do *Repositório Científico de Acesso Aberto em Portugal (RCAAP)*<sup>74</sup>. Em relação ao primeiro, segundo informações disponibilizadas pela própria plataforma recolhidas em maio de 2019, agrega mais de 3,5 milhões de teses e dissertações em formato eletrónico,

---

<sup>73</sup> <https://oatd.org/>

<sup>74</sup> <https://www.rcaap.pt>

disponíveis em *open access*, de cerca de 600 instituições de ensino superior<sup>75</sup>. Por conseguinte, a escolha deste recurso garante um alcance e cobertura significativos, ao mesmo tempo que assegura que todas as teses e dissertações encontradas poderão ser livremente consultadas.

O RCAAP, por sua vez, afirma-se como um ponto de acesso único a milhares de teses e dissertações em acesso aberto contidas numa pluralidade de repositórios portugueses<sup>76</sup>, agregando também artigos científicos e artigos publicados em revistas científicas. Neste sentido, com a escolha deste agregador pretende-se cobrir a análise das teses mais pertinentes no enquadramento do nosso estudo, publicadas em Portugal. O facto de ser um repositório em acesso aberto garante a liberdade de acesso aos resultados encontrados. Estes dados podem ser comparados com os dados fornecidos pelo Renates<sup>77</sup>, a base de dados oficial de registo de teses e dissertações realizadas em Portugal.

Foi ainda considerada a hipótese de usar a base de dados do ProQuest<sup>78</sup> como terceira base de dados de consulta de teses e dissertações. Porém, as pesquisas realizadas adotando critérios similares ao do RCAAP e da OATD recuperaram apenas 19 documentos, dos quais nenhum foi considerado pertinente para os fins da investigação. Por este motivo, exclui-se a apresentação destes dados do aparato metodológico.

#### 3.4.4.1 *Open Access Theses and dissertations*

Tal como nas tipologias documentais anteriores, procurou-se recuperar teses e dissertações redigidas em inglês, português e espanhol, publicadas entre 2009 e 2019 no domínio da arquivística. No caso da língua inglesa, os parâmetros de pesquisa foram:

```
Any of these words: archive; archives  
All of these words: information science  
Written in: English  
Published in: any country  
Earliest date: 2009  
Latest date: 2020
```

A escolha da data como 2020 justifica-se pelo facto de esta base de dados não permitir a escolha do mês. Assim, embora a pesquisa tenha sido realizada em novembro 2019, inseriu-se 2020 de forma a abarcar todas as teses e dissertações publicadas em 2019. A seguinte expressão sintetiza a pesquisa:

---

<sup>75</sup> A lista de fontes de instituições que alimentam a base de dados pode ser consultada em <<https://oatd.org/oatd-publishers.html>>.

<sup>76</sup> A lista de fontes pode ser consultada em <<https://www.rcaap.pt/directory.jsp>>.

<sup>77</sup> <https://renates2.dgeec.mec.pt/>

<sup>78</sup> <https://search.proquest.com/>

(abstract:(archive; OR archives) AND discipline:(Information AND Science)  
AND language:(en OR eng OR english) AND pub\_dt:[2009-01-01T00:00:00Z TO  
2020-01-01T00:00:00Z]

No caso da língua portuguesa, os critérios de pesquisa foram:

Any of these words: arquivo; arquivos; metadados; autenticidade;  
confiabilidade

Written in: Portuguese

Published in: any country

Earliest date: 2009

Latest date 2020

(abstract:(arquivo; OR arquivos) AND abstract:(metadado; OR  
autenticidade; OR confiabilidade) AND language:(pt OR por OR portuguese)  
AND pub\_dt:[2009-01-01T00:00:00Z TO 2020-01-01T00:00:00Z])

Para a língua espanhola adotaram-se os seguintes critérios:

Any of these words: archivo; archivos; metadatos; autenticidad;  
confiabilidad

Written in: Spanish

Published in: any country

Earliest date: 2009

Latest date: 2020

(abstract:(archivo; OR archivos) AND abstract:(metadatos; OR  
autenticidad; OR confiabilidad) AND language:(es OR esp OR espanol OR  
spanish) AND pub\_dt:[2009-01-01T00:00:00Z TO 2020-01-01T00:00:00Z])

A principal limitação desta plataforma prende-se com o facto de o motor de busca não reconhecer caracteres acentuados (‘ç’, ‘ã’, ‘ó’) o que inviabiliza a pesquisa na área da ‘Ciência da Informação’ nas línguas portuguesa e espanhola. Como alternativa, definiram-se para estas línguas palavras-chave adicionais com que filtrar os resultados. Não foram usados os termos ‘descrição arquivística’ e ‘representação’ por conterem caracteres acentuados.

#### 3.4.4.2 *Repositório Científico de Acesso Aberto em Portugal*

Tal como nas tipologias documentais anteriores, procurou-se recuperar teses e dissertações redigidas em inglês, português e espanhol, publicadas entre 2009 e 2019 no domínio da arquivística. No caso da língua inglesa, os parâmetros de pesquisa foram:

Document type: doctoral thesis + master thesis

Date: 2009-2020

Subject: information science

Language: English

Em língua espanhola:

Document type: doctoral thesis + master thesis

Date: 2009-2020

Subject: ciencia de la información

Language: Spanish

E em língua portuguesa:

Document type: doctoral thesis + master thesis

Date: 2009-2020

Subject: ciência da informação

Language: portuguese

Title: archiv\*

No caso do português, foi necessário restringir os resultados às obras com o termo ‘arquiv’ no título devido à vasta amostra obtida, num total de 1219. Em comparação com os nove resultados para a língua inglesa e os 18 resultados para a língua espanhola, nota-se que, pelo menos no domínio da ciência da informação, o RCAAP apresenta um claro pendor lusófono.

### 3.5 Análise dos dados

Nesta secção discute-se como se procedeu à análise do *corpus* documental identificado no ponto prévio, nomeadamente no que diz respeito à identificação das fontes primárias (artigos científicos, monografias, comunicações, teses e dissertações) contidas nas fontes secundárias (bases de dados). Uma vez identificadas as fontes de informação, explicam-se os critérios para a sua seleção enquanto documentos pertinentes para integração na revisão de literatura.

O papel da análise dos dados pode ser ilustrado recorrendo à metáfora de uma pirâmide, segundo a qual as questões de investigação e os objetivos presidem ao vértice superior, sustentados na base pelos dados empíricos e pelos dados teóricos (Figura 14). Por outras palavras, procura-se perceber de que modo os dados teóricos e empíricos respondem às questões de investigação, cumprindo os objetivos propostos.

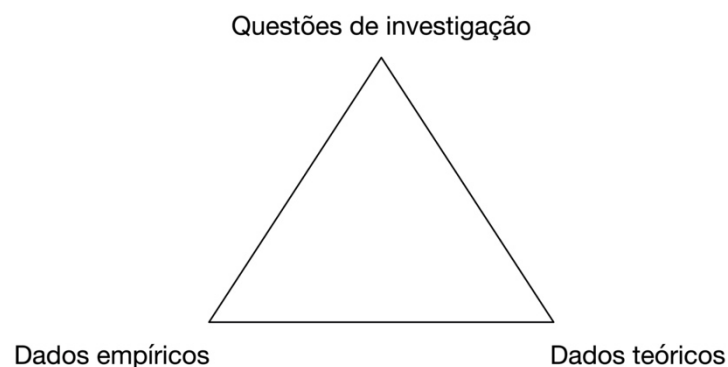


Figura 14 — Triangulação entre a análise dos dados e as questões de investigação.

FONTE: elaboração própria.

Esta metáfora não se entende de forma estática, mas enquanto um processo contínuo através do qual os as questões de investigação e os dados se vão moldando mutuamente, adequando-se aos novos contextos e significados que vão surgindo durante a investigação. Recorda-se a circularidade do processo interpretativo veiculado na hermenêutica, segundo a qual as partes devem ser compreendidas em relação a um todo, e o todo só pode ser compreendido em função das partes, porque é formado por elas. Como resolução desta aparente contradição, o círculo é transformado em espiral (Radnitzky, 1970, p. 23) de forma a que se, começando por uma parte, se tente relacioná-la com o todo, adquirindo um novo entendimento, voltando depois para o estudo da parte, repetindo este processo. Neste caso, o processo circular ou em espiral traduz-se no facto de as questões de investigação determinarem um primeiro momento o tipo de dados empíricos a recolher. Porém, uma vez recolhidos esses dados, a melhor compreensão e perceção da realidade a investigar pode levar ao reposicionamento das próprias perguntas de investigação. Neste caso, acrescentam-se os dados teóricos obtidos a partir da revisão de literatura como outro eixo de influência para a definição da problemática de investigação (Punch 1998).

No que concerne a análise dos dados, empregam-se fundamentalmente dois métodos: a investigação documental e a engenharia de requisitos, sendo que ambas apresentam componentes empíricas e teóricas. A ID apresenta igualmente uma natureza qualitativa e quantitativa, na medida em que a definição do tipo de dados a analisar representa uma dimensão qualitativa, ao passo que a extração de contagens estatísticas se enquadra na abordagem quantitativa. Como parte da revisão de literatura constituem-se sete categorias de análise que emergem a partir da análise dos dados: obras sobre (1) representação da informação; (2) descrição arquivística; (3) epistemologia da gestão da informação (por vezes gestão documental); (4) normas de metadados; (5) epistemologia metadados; (6) autenticidade e confiabilidade; (7) e *Linked data*. Estes temas presentes nos dados empíricos constituem os pilares temáticos da revisão de literatura nos quais é possível categorizar a grande maioria dos autores identificados na seleção bibliográfica. A Tabela 17 ilustra não só estas categorias, como também alguns dos autores e das obras de destaque para o respetivo tema, independentemente do tipo de fonte documental.

Tabela 17 — Categorias de análise emergentes a partir da revisão de literatura.

Categoria	Autor e obra
<b>Representação da Informação</b>	(Duff, Yakel, & Tibbo, 2013; J. Niu, 2015a; Serewicz, 2010; I. Silva, 2012; Verborgh et al., 2015; Yakel, 2003; Yeo, 2012; Zhang, 2012b, 2012a)

<b>Descrição arquivística</b>	(Anchor, 2013; Douglas, 2016; Gracy & Lambert, 2014; Hedstrom, 1993; Llanes-Padrón, 2019; MacNeil, 2009, 2012; Meehan, 2009; Moyano Collado, 2013; J. Niu, 2013; Zhang, 2012c)
<b>Gestão da Informação</b>	(Acker, 2017; J. Bailey, 2013; C. Brown, 2019; Cumming, 2010; Duranti & Franks, 2015; Kallberg, 2012; MacNeil, 2017; McLeod, 2014; Ridener, 2009; Tough, 2016; Wright, 2014; Yeo, 2011)
<b>Metadados</b>	(Apostolou, 2009; Baca, 2016; Beyene & Godwin, 2018; Chen, Wen, Chen, Lin, & Sum, 2011; Gartner, 2016; Gilliland, 2016; Gladney, 2009; Haynes, 2018; Li & Sugimoto, 2017; NISO, 2017; Pomerantz, 2015; Yeo, 2018; Zeng & Qin, 2016)
<b>Normas de metadados</b>	(Andrade, Berti Júnior, Pereira, & Baptista, 2014; Baños-Moreno, Valentín Ruiz, & Blázquez Martín de las Mulas, 2019; Botão, 2011; Bunn, 2013; Dow, 2009; Dryden, 2009; Ducheveva & Pennington, 2019; Henttonen, 2009; L. Linden, 2017; Llanes-Padrón & Moro-Cabero, 2017; Llanes-Padrón & Pastor-Sánchez, 2017; Moro-Cabero, Martín-Pozuelo, & Zazo, 2011; Riley, 2010; Rolan, 2017; Woodley, 2016; Youn, 2015)
<b>Autenticidade</b>	(Bhatia & Wright de Hernandez, 2019; Bountouri, Gratz, & Sanmartin, 2017; Dryden, 2011; Duncan, 2009; Duranti et al., 2019; Engvall, 2019; Gladney, 2009; Hofman, Lemieux, Joo, & Batista, 2019; Jansen, 2015; Lemieux, 2016; Mak, 2012; McLeod & Gormly, 2017; Price & Smith, 2011; Rogers, 2015; Rogers & Tennis, 2013; Yeo, 2013)
<b>Linked Data</b>	(Gartner, 2015; Hooland & Verborgh, 2014; Jones, 2018; Machado, Souza, & Graça Simões, 2019; J. Niu, 2016; Rolan, 2015; Rolan et al., 2019; Samouelian, 2009)

FONTE: elaboração própria.

A análise dos dados procura responder às questões de investigação e aos objetivos com base nos dados empíricos e teóricos. Os dados teóricos correspondem aos primeiros dois capítulos de revisão de literatura, elaborada a partir da leitura atenta e crítica de estes autores e de outros autores que se enquadram nestes eixos temáticos. No quarto capítulo, estas categorias serão revisitadas de forma a aferir se o posicionamento temático de partida tem influência nos requisitos veiculados por cada área temática. Uma lista detalhada das obras recuperadas para servir de fonte

de informação é apresentada adiante. Por agora importa destacar que os dados teóricos respondem adequadamente às questões de investigação, enquadrando-se a metodologia nos objetivos propostos.

### 3.5.1 Métodos de análise de dados

#### 3.5.1.1 *Investigação documental*

A investigação documental pressupõe a análise das fontes de informação através da aplicação de processos técnicos precisos que permitam a apreensão do significado dos dados. A análise é essencial *for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use* (Krippendorff, 2004 p. 18), uma definição que assume que o conteúdo emerge no processo através do qual um investigador analisa um texto em relação a um contexto. Com base nessa análise, o investigador constitui de categorias de análise relacionadas com as questões de investigação (Bowen, 2009, p. 32). Carlos Guardado da Silva reforça que a análise das fontes possibilita não só a descoberta de informação relevante num estudo, mas também *distorções ou esterótipos* (C. G. da Silva, 2021, p. 103). Este ponto é particularmente pertinente porque os textos veiculam não só as ideias dos seus autores, como também as ideias que os seus autores não têm. Deve-se olhar não só para o que está escrito, como também para o que não foi escrito e porque não foi escrito. Os silêncios dos textos comunicam pela sua ausência.

Este processo interpretativo implica uma hermenêutica textual, conforme abordado previamente. A hermêutica pode entende-se tanto como um ramo da filosofia como um modelo de análise de dados qualitativos (Bleicher, 1980). Enquanto modelo de análise, a dimensão que interessa trazer para este estudo, afirma-se como uma forma de *buscar significados latentes nos textos escritos* (Coutinho, 2019, p. 228) assente na lógica circular de interpretação e reinterpretação contínua que envolve o texto e as suas componentes num movimento do todo para a parte e da parte para o todo (Gadamer, 2013). Neste sentido, a hermenêutica afirma-se como uma técnica usada pelos investigadores para tecer uma apreciação crítica do material de investigação (Alvesson & Skoldberg, 2010, p. 115) servindo-se de um *inquérito rigoroso das fontes, a sua imaginação e, quando possível, a sua maturidade* (C. G. da Silva, 2020, p. 104). Relembra-se uma vez mais a analogia da interpretação circular segundo um texto apenas pode ser compreendido na relação com os outros textos que emergem do mesmo contexto social (Alvesson & Skoldberg, 2010, p. 92). É com esta intenção que emerge a revisão de literatura, enquanto uma teia textual de obras relacionadas entre si com a finalidade de contextualizar o problema da autenticidade nas descrições arquivísticas, colocando os textos em diálogo e, com as nossas reflexões, proceder a uma crítica construtiva da própria autenticidade.

Quivy e Campenhoudt (2013, p. 227) identificam duas categorias de análise: os métodos quantitativos e os qualitativos. No decurso da análise da metodologia empregue nesta investigação aplicam-se ambas as categorias. De um modo geral, a análise das fontes de informação neste estudo pressupõe um critério quantitativo. A seleção de revistas e de monografias obedece em primeiro



lugar a um critério quantitativo que determinou a realização de uma seleção de revistas com a ocorrência de um determinado termo no título (archiv\* e variantes), filtrada por outros critérios quantitativos tais como o ICDS e a data.

Todavia, a seleção dos artigos científicos e das monografias presentes em cada revista selecionada, no intervalo definido, ocorreu segundo um critério qualitativo porque não se procedeu à recuperação de artigos com base em termos de pesquisa, efetuando-se antes uma leitura atenta e individual dos resumos de todos os artigos publicados pelas revistas entre 2009 e 2019. A determinação da pertinência é assumidamente subjetiva e qualitativa, tentando-se perceber pelos títulos dos artigos científicos, pelos resumos e, quando necessário, por uma leitura do próprio artigo, quais os artigos que melhor abordam as questões de investigação e que deviam ser selecionados para integrar a revisão de literatura. Um exercício semelhante foi realizado a nível das monografias. As resenhas críticas de cada monografia encontradas nas revistas selecionadas, entre 2009 e 2019, foram alvo de análise e, através delas, depreenderam-se se as linhas temáticas desenvolvidas na monografia se enquadravam na discussão proposta pelas questões de investigação. Por sua vez, a recuperação de comunicações científicas e de teses e dissertações académicas obedeceu a um método mais estritamente quantitativo na medida em que a análise decorreu com base na ocorrência de termos definidos em palavras-chave. Porém, a seleção destas fontes para os fins da investigação pautou-se por um semelhante critério qualitativo e subjetivo de pertinência.

A análise pressupõe a constituição de categorias ou unidades de análise, entendidas como *secções de texto de natureza e dimensão variáveis, que devem afigurar-se como elementos detentores de um sentido completo e com pertinência para o objeto de estudo* (Coutinho, 2019, p. 219). Identificam-se três categorias de análise: (i) as unidades de contexto, (ii) as unidades de registo e (iii) as unidades de enumeração. Para Laurence Bardin, a unidade de contexto *serve de unidade de compreensão para codificar a unidade de registo e corresponde ao segmento da mensagem, cujas dimensões (superiores às da unidade de registo) são ótimas para que se possa compreender a significação exata da unidade de registo* (Bardin, 2009, p. 133). Trata-se de segmentos latos de onde são retiradas as unidades de registo. A unidade de contexto é delimitada pelo objeto de investigação, correspondendo à identificação de temáticas em função das quais são analisados os dados. No âmbito desta investigação, as unidades de contexto identificam-se com a tipologia documental (revistas, artigos, monografias, comunicações e teses e dissertações) na qual se enquadram as fontes de informação primárias.

Por sua vez, as unidades de registo definem-se como *a unidade de significação a codificar ao segmento de conteúdo a considerar como unidade base, visando a categorização e a contagem frequencial* (Bardin, 2009, p. 130). Embora exista uma certa ambiguidade a respeito dos critérios de distinção das unidades de registo, entende-se neste estudo que correspondem aos documentos que se extraem de cada fonte secundária, bem como o próprio método de extração. Conforme referido, a identificação de artigos e de monografias obedeceu a um critério qualitativo e subjetivo de averiguação da pertinência através da leitura do título, do resumo e do texto integral, quando necessário. Lee (2000, p. 158) alerta para as falácias de uma contagem cega da ocorrência dos

termos que pode levar a interpretações erradas, uma vez que podem ser usados pronomes, o significado dos termos pode variar, ou pode ser necessário fazer uma série de iterações de pesquisas até obter os dados relevantes para a pesquisa. Weber (1985, p. 158) também nota que as frequências das palavras podem ser enganadoras quando o contexto de ocorrência desses termos é ignorado. Por estes motivos, optou-se por uma leitura atenta e individualizada das unidades de registo (os documentos) no caso dos artigos científicos e das monografias.

Por último, identificam-se ainda as unidades de enumeração enquanto representação estatística das outras unidades de análise. As unidades de enumeração compreendem a conversão estatística dos dados em função de indicadores e percentagens, permitindo uma avaliação mais direta do *corpus* documental. Se de uma revista que contém 100 artigos forem selecionados 10, ‘10’ atua como o indicador, ao passo que 10% constitui a percentagem. A discriminação das várias unidades de análise é apresentada mais adiante.

### 3.5.1.2 Engenharia de requisitos

A segunda técnica de análise de dados a ser empregue como parte da metodologia mista é a engenharia de requisitos. Esta abordagem é normalmente usada no desenvolvimento de sistemas de informação (Jackson, 1995; Laplante, 2013; Robertson & Robertson, 2012; Wiegers & Beatty, 2013) não só como parte do processo de compreensão das necessidades do utilizador, como também na definição das estruturas e propriedades dos sistemas que devem ser adaptadas de forma a ir ao encontro das necessidades institucionais, disposições legais ou normas obrigatórias (Pohl, 2010).

Pamela Zave propõe a seguinte definição de engenharia de requisitos:

*Requirements engineering is the branch of software engineering concerned with the real-world goals for, functions of, and constraints on software systems. It is also concerned with the relationship of these factors to precise specifications of software behaviour, and to their evolution over time and across software families* (Zave, 1997, p. 315).

A definição de Zave salienta o modo como um determinado *software* é desenvolvido com o fim de solucionar um problema concreto no mundo, pelo que a compreensão do problema e do respetivo contexto figura como um ponto central na engenharia de requisitos. Neste sentido, Bennaceur e colegas destacam a necessidade de compreender não só os objetivos do sistema, como também os constrangimentos do ambiente no qual o sistema opera (Bennaceur, Tun, Yu, & Nuseibeh, 2019, p. 2). É evidente que esta tese não se insere no domínio do desenvolvimento de aplicações de sistemas de informação, pelo que não se deve tomar esta definição de engenharia de requisitos de um modo literal. Antes, pretende-se retirar deste contexto eminentemente tecnológico o raciocínio teórico de desenvolvimento de um instrumento que aborde determinado problema do mundo real, devendo esse instrumento ter em consideração um entendimento do problema e do seu contexto.

Mais especificamente, a engenharia de requisitos refere-se ao processo sistemático de desenvolvimento de requisitos de sistemas (Pinheiro, 2003). Inclui a identificação dos objetivos do

sistema e a descrição detalhada de cada propriedade, em conjunto com as condicionantes do seu desenvolvimento (Santos, 2015). Este método pode ser dividido em duas grandes áreas interdependentes de atividades que devem sempre acontecer de forma simultânea: i) análise do problema; e ii) especificação de requisitos (Pinheiro, 2003). No decurso desta tese, os princípios e técnicas da engenharia de requisitos serão particularmente úteis para o desenvolvimento do modelo de metadados. Assim, completando a metáfora referida na análise dos dados, o problema a abordar identifica-se com a representação da autenticidade dos documentos de arquivo através das suas descrições arquivísticas. Os requisitos a desenvolver constituem os critérios sobre os quais a autenticidade assenta, extraídos da teoria arquivística. Por conseguinte, espera-se que o método de engenharia de requisitos auxilie na identificação e na caracterização dos elementos de metadados necessários para verificar a autenticidade dos documentos de arquivo e comprovar a sua confiabilidade. Por fim, o sistema ou instrumento a desenvolver é simbolizado pelo modelo de metadados, o *software* que visa dar uma resposta concreta e contextualizada ao problema encontrado no mundo real.

A necessidade de enveredar pela construção de um modelo justifica-se pelo facto de os dados teóricos e empíricos encontrados, evidenciados na revisão de literatura, não responderem plenamente às questões de investigação. Em conjugação com a análise detalhada dos metadados estruturais das normas de metadados, procura-se confirmar a terceira hipótese de investigação que as normas de metadados refletem parcialmente aspetos relacionados com a autenticidade, não capturando a totalidade dos requisitos necessários para a verificação da autenticidade e da confiabilidade dos documentos de arquivo. Avança-se então para a elaboração de uma proposta de um modelo de metadados para a autenticidade no quarto capítulo.

### **3.5.2 Seleção de artigos científicos**

A escolha dos artigos científicos mais pertinentes para o estudo partiu da análise atenta das 12 revistas identificadas no Quadro 1. Para cada revista, consultaram-se todos os números publicados entre 1 de janeiro 2009 e 31 de dezembro de 2019 e, para cada número, foram escrutinados todos os artigos publicados. A escolha dos textos, conforme mencionado anteriormente, assentou num critério qualitativo através do qual se analisou o título, o resumo e de modo a determinar se o tema do artigo seria potencialmente relevante para responder às questões de investigação. Em síntese, os critérios foram:

Date: 01-2009 a 12-2019  
Document type: article  
Language: all  
ISSN: [o de cada revista]

Neste sentido, procuraram-se artigos que abordassem, independentemente do ângulo, pelo menos um dos seguintes temas: a representação da informação, a descrição arquivística, os metadados, e a autenticidade. De seguida apresentam-se todos os artigos científicos extraídos das revistas científicas, e considerados relevantes para potencialmente integrar a revisão de literatura.

A recolha decorreu entre finais de setembro de 2019 e inícios de outubro 2019 na base de dados LISTA. Para cada artigo atribui-se: (1) um identificador único; (2) indica-se o ano de publicação; (3) os autores e (4) o título.

### 3.5.2.1 *American Archivist*

Os seguintes dados da revista *American Archivist* (ISSN: 3609081) foram recolhidos a 23 de setembro de 2019. De um universo de 204 artigos publicados entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 17 títulos constados na Tabela 18, num total de 8%.

Tabela 18 — Artigos científicos retirados da *American Archivist*.

Item	Ano	Autor(es)	Título
AA1	2009	Mary Samouelian	Embracing Web 2.0: Archives and the Newest Generation of Web Applications.
AA2	2009	Meehan Jennifer	Making the Leap from Parts to Whole: Evidence and Inference in Archival Arrangement and Description.
AA3	2009	Jenn Riley & Kelcy Sheperd	A Brave New World: Archivists and Shareable Descriptive Metadata.
AA4	2009	Nancy Kunde	Getting It Done-Collaboration and Development of the Digital Records Conversion Standard.
AA5	2009	Henry Gladney	Long-Term Preservation of Digital Records: Trustworthy Digital Objects.
AA6	2010	Morgan Daniels & Elizabeth Yakel	Seek and You May Find: Successful Search in Online Finding Aid Systems.
AA7	2011	Emily Monks-Leeson	Archives on the Internet: Representing Contexts and Provenance from Repository to Website.
AA8	2011	Kate Theimer	What Is the Meaning of Archives 2.0?
AA9	2011	Joshua Sternfeld	Archival Theory and Digital Historiography: Selection, Search, and Metadata as Archival Processes for Assessing Historical Contextualization.
AA10	2012	Axley DeRidder	Leveraging Encoded Archival Description for Access to Digital Content: A Cost and Usability Analysis.
AA11	2013	Jane Zhang & Dayne Mauney	When Archival Description Meets Digital Object Metadata: A Typological Study of Digital Archival Representation.
AA12	2013	Katherine Wisser and Jackie Dean	EAD Tag Usage: Community Analysis of the Use of Encoded Archival Description Elements.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
AA13	2013	Gretchen Gueguen, Vitor Manoel Marques da Fonseca, Daniel Pitti, & Claire Sibille-de Grimouard	Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives' Experts Group on Archival Description.
AA14	2014	Karen Gracy and Frank Lambert	Who's Ready to Surf the Next Wave? A Study of Perceived Challenges to Implementing New and Revised Standards for Archival Description.
AA15	2015	Wendy Duff & Jessica Haskell	New Uses for Old Records: A Rhizomatic Approach to Archival Access.
AA16	2016	Jennifer Douglas	Toward More Honest Description.
AA17	2016	Ine Fintland	Archival Descriptions through the Looking Glass: Paratexts in Wonderland.

FONTE: elaboração própria.

### 3.5.2.2 *Archival issues*

Os seguintes dados da revista *Archival Issues* (ISSN: 10674993) foram recolhidos a 25 de setembro de 2019. De um universo de 58 artigos publicados entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 2 títulos constados na Tabela 19, num total de 3%.

Tabela 19 — Artigos científicos retirados da *Archival Issues*.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
AI1	2011	Katie Rudolph	Separated At Appraisal: Maintaining the Archival Bond Between Archives Collections and Museum Objects.
AI2	2011	Rachel Howard, Heather Fox & Caroline Daniels	The Born-Digital Deluge: Documenting Twenty-First Century Events.

FONTE: elaboração própria.

### 3.5.2.3 *Archival Science*

Os seguintes dados da revista *Archival Science* (ISSN: 13890166) foram recolhidos a 25 de setembro de 2019. De um universo de 207 artigos publicados entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 18 títulos constados na Tabela 20, num total de 9%.

Tabela 20 — Artigos científicos retirados da *Archival Science*.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
AS1	2009	Jennifer Meehan	The archival nexus: rethinking the interplay of archival ideas about the nature, value, and use of records.
AS2	2010	Geoffrey Yeo	'Nothing is the same as something else': significant properties and notions of identity and originality.
AS3	2010	Fiorella Foscarini	Understanding the context of records creation and use: 'Hard' versus 'soft' approaches to records management.
AS4	2011	Dara Prince & Johanna Smith	The trust continuum in the information age: a Canadian perspective.
AS5	2011	Anneli Sundqvist	Documentation practices and recordkeeping: a matter of trust or distrust?
AS6	2011	Paul Conway	Archival quality and long-term preservation: a research framework for validating the usefulness of digital surrogates.
AS7	2011	Gillian Oliver, Brenda Chawner & Hai Ping Liu	Implementing digital archives: issues of trust.
AS8	2011	Dharma Akmon, Ann Zimmerman, Morgan Daniels & Margaret Hedstrom	The application of archival concepts to a data-intensive environment: working with scientists to understand data management and preservation needs.
AS9	2011	Luciana Duranti	Educating for trust.
AS10	2012	Wendy Duff	Contexts built and found: a pilot study on the process of archival meaning-making
AS11	2012	Kathleen Fear & Devan Ray Donaldson	Provenance and credibility in scientific data repositories.
AS12	2012	Heather MacNeil	What finding aids do: archival description as rhetorical genre in traditional and web-based environments.
AS13	2012	Ciaran Trace	The evolution of the finding aid in the United States: from physical to digital document genre.
AS14	2013	Devan Ray Donaldson & Elizabeth Yakel	Secondary adoption of technology standards: The case of PREMIS.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
AS15	2016	Greg Bak	Trusted by whom? TDRs, standards culture and the nature of trust.
AS16	2016	Ricardo Eito-Brun	Remote access to EAC-CPF context and authority records for metadata indexing: a solution based on open information retrieval standards.
AS17	2017	Julie Mcleod & Brianna Gormly	Using the cloud for records storage: issues of trust.
AS18	2018	Laurent Romary, Charles Riondet	EAD ODD: a solution for project-specific EAD schemes.

FONTE: elaboração própria.

### 3.5.2.4 *Archivaria*

Os seguintes dados da revista *Archivaria* (ISSN: 03186954) foram recolhidos a 30 de setembro de 2019. De um universo de 171 artigos publicados entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 19 títulos constados na Tabela 21, num total de 11%.

Tabela 21 — Artigos científicos retirados da *Archivaria*.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
Archivaria1	2009	Lisa Darms	The Archival Object: A Memoir of Disintegration.
Archivaria2	2009	Kenneth Hawkins	A Pattern Language for Electronic Records.
Archivaria3	2009	Katherine Timms	New Partnerships for Old Sibling Rivals: The Development of Integrated Access Systems for the Holdings of Archives, Libraries, and Museums.
Archivaria4	2009	Chris Duncan	Counterpoint: authenticity or bust
Archivaria5	2010	Jennifer Meehan	Rethinking Original Order and Personal Records.
Archivaria6	2011	Ciaran Trace	Beyond the Magic to the Mechanism: Computers, Materiality, and What It Means for Records to Be "Born Digital".
Archivaria7	2012	Bonnie Mak	On the Uses of Authenticity.
Archivaria8	2012	Geoffrey Yeo	The Conceptual Fonds and the Physical Collection.
Archivaria9	2012	Richard Dancy	RAD Past, Present, and Future
Archivaria10	2012	Geoffrey Yeo	Bringing Things Together: Aggregate Records in a Digital Age.

Item	Ano	Autor(es)	Título
Archivaria11	2012	Andrew Janes	Of Maps and Meta-Records: Eighty-Five Years of Map Cataloguing at The National Archives of the United Kingdom.
Archivaria12	2012	Jane Zhang	Original Order in Digital Archives.
Archivaria13	2014	Dancy Richard	Developing Archival Standards.
Archivaria14	2016	Jinfang Niu	Linked Data for Archives.
Archivaria15	2018	Christopher Becker	Metaphors We Work By: Reframing Digital Objects, Significant Properties, and the Design of Digital Preservation Systems.
Archivaria16	2018	Alyssa Hamer	Ethics of Archival Practice: New Considerations in the Digital Age.
Archivaria17	2018	Michael Moss, David Thomas & Tim Gollins	The Reconfiguration of the Archive as Data to Be Mined.
Archivaria18	2018	Jinfang Niu	Integrated Online Access to Objects and Archives.
Archivaria19	2019	Devon Mordell	Critical Questions for Archives as (Big) Data.

FONTE: elaboração própria.

### 3.5.2.5 *Archives (London)*

Os seguintes dados da revista *Archives (London)* (ISSN: 00039535) foram recolhidos a 30 de setembro de 2019. De um universo de 80 artigos publicados entre 1 de janeiro 2009 e 31 de dezembro de 2019, nenhum título foi considerado pertinente, num total de 0%. Os artigos encontrados dizem respeito ao tratamento de manuscritos, de história ou de estudos de caso de âmbito exterior à investigação (religião, período medieval, antigas colónias). Por este motivo, nenhum foi considerado relevante para os objetivos propostos.

### 3.5.2.6 *Archives (Quebec)*

Os seguintes dados da revista *Archives (Quebec)* (ISSN: 0044-9423) foram recolhidos a 2 de outubro de 2019. De um universo de 98 artigos publicados entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 4 títulos constados na Tabela 22, num total de 4%. Também nesta revista se encontram vários artigos fora do contexto de investigação, com uns a versar sobre arte, outros sobre estudos de caso específicos.

Tabela 22 — Artigos científicos retirados da *Archives (Quebec)*.



<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
Arch.(QBC)1	2009	Nahuet Robert	L'arch ivistique contemporaine a l'âge adulte: pertinence et actualité du respect desfonds.
Arch.(QBC)2	2010	David Rajotte	La réflexion archivistique à l'ère du document numérique: un bilan bistorique.
Arch.(QBC)3	2013	Diane Baillargeon	Les archives universitaires sontelles solubles dans le Web 2.0?
Arch.(QBC)4	2013	Christine Dufour & Stéphanie Cadieux	Web 2.0 et milieux documentaires : qu'en retenir?

FONTE: elaboração própria.

### 3.5.2.7 *Archives and Manuscripts*

Os seguintes dados da revista *Archives and Manuscripts* (ISSN: 0157-6895) foram recolhidos a 2 de outubro de 2019. De um universo de 228 artigos publicados entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 33 títulos constados na Tabela 23, num total de 15%.

Tabela 23 — Artigos científicos retirados da *Archives and Manuscripts*.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
A&M1	2009	Joanne Evans, Susan Marilyn, McKemmish & Barbara Reed	Making Metadata Matter: Outcomes from the Clever Recordkeeping Metadata Project.
A&M2	2010	Bob Pymm	Archives and Web 2.0: the example of the September 11 digital archive.
A&M3	2011	Alex Byrne	WikiLeaks and Web 2.0: privacy, security and other things that keep me awake ....
A&M4	2013	Robert VanderBerg	Converging libraries, archives and museums: overcoming distinctions, but for what gain?
A&M5	2013	Blinda Battley	Finding aids in context: using Records Continuum and Diffusion of Innovations models to interpret descriptive choices
A&M6	2013	Jiufang Niu	Recordkeeping metadata and archival description: a revisit.
A&M7	2014	Joanne Evans	Designing dynamic descriptive frameworks.
A&M8	2014	Shigeo Sugimoto	Digital archives and metadata as critical infrastructure to keep Community memory safe for the future - lessons from Japanese activities.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
A&M9	2014	Barbara Reed	Reinventing access.
A&M10	2014	Kate Cumming	Reinventing Archival Methods.
A&M11	2014	Belinda Battley, Elizabeth Daniels, Gregory Rolan	Archives as multifaceted narratives: linking the 'touchstones' of community memory.
A&M12	2014	Mark Crookston	Archives as multifaceted narratives: linking the 'touchstones' of community memory.
A&M13	2014	Joanne Evans	Reflections on the promise and pitfalls in reinventing recordkeeping metadata.
A&M14	2014	Cassie Findlay	Full docs or it didn't happen.
A&M15	2014	Michael Jones	Contrapuntal archival methods.
A&M16	2014	Charlotte Maday	Records management for scientific data.
A&M17	2014	Julie Mcleod	Reinventing archival methods: reconceptualising electronic records management as a wicked problem.
A&M18	2014	Adelaide Parr	In an interconnected world - why do we think in functions?
A&M19	2014	Barbara Reed	Rethinking approaches to recordkeeping metadata.
A&M20	2014	Tim Sheratt	Contexts, connections, access: the glorious possibilities of getting it all wrong.
A&M21	2014	Kirsten Wright	Broadening the record and expanding the archives.
A&M22	2015	Gregory Rolan	Towards Archive 2.0: issues in archival systems interoperability.
A&M23	2015	Jinfang Niu	Original order in the digital world.
A&M24	2015	Julia Kastenhofer	The logic of archival authenticity: ISO 15489 and the varieties of forgeries in archives.
A&M25	2015	Gilliand Oliver & Fiorella Foscarini	The value of international standards for records management: perspectives from education and training.
A&M26	2016	Jinfang Niu	Organisation and description of datasets.
A&M27	2017	Laura Millar	On the crest of a wave: transforming the archival future.
A&M28	2017	Ross Spencer	Binary trees? Automatically identifying the links between born-digital records.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
A&M29	2018	Morgan Currie & Britt Paris	Back-ups for the future: archival practices for data activism.
A&M30	2018	Cassie Findlay	Crunch time: the revised ISO 15489 and the future of recordkeeping.
A&M31	2019	Gregory Rolan, Glen Humphries, Lisa Jeffrey, Evanthia Samaras, Tatiana Antsoupova & Katharine Stuart	More human than human? Artificial intelligence in the archive.
A&M32	2019	Millicent Weber & Rachel Buchanan	Metadata as a machine for feeling in Germaine Greer's archive.
A&M33	2019	Jane Winters & Andrew Prescott	Negotiating the born-digital: a problem of search.

FONTE: elaboração própria.

### 3.5.2.8 *Archives and Records*

Os seguintes dados da revista *Archives and Records* (ISSN: 2325-7962) foram recolhidos a 4 de outubro de 2019. De um universo de 88 artigos publicados entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 17 títulos constados na Tabela 24, num total de 19%.

Tabela 24 — Artigos científicos retirados da *Archives and Records*.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
A&R1	2013	Margaret Procter, Elizabeth Sheperd	Writing the record office.
A&R2	2013	Geoffrey Yeo	Trust and context in cyberspace.
A&R3	2013	Rachel Anchor	'More product, less process': method, madness or practice?
A&R4	2013	Jennifer Bunn	Developing descriptive standards: a renewed call to action.
A&R5	2014	Tarvo Kärberg	Digital preservation and knowledge in the public archives: for whom?
A&R6	2015	Eunha Youn	Adoption of ISAD(G) in practice: a close look at the standardization process of ISAD(G) in a manuscript archives of Korea.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
A&R7	2015	Ann-Sofie Klareld	'Isn't it information assets we're really talking about?' A discourse analysis of a panel discussion on digital archives.
A&R8	2016	Dorothy Waugh, Elizabeth Russey Roke & Erika Farr	Flexible processing and diverse collections: a tiered approach to delivering born digital archives.
A&R9	2016	John Langdon	Describing the digital: the archival cataloguing of born-digital personal papers.
A&R10	2016	Jinfang Niu	Aggregate control of scientific data.
A&R11	2016	Jennifer Bunn	Archival description and automation: a brief history of going digital.
A&R12	2016	Alistair Tough	Thinking about and working with archives and records: a personal reflection on theory and practice.
A&R13	2017	Margaret Procter	Protecting rights, asserting professional identity.
A&R14	2018	Michael Jones	From catalogues to contextual networks: reconfiguring collection documentation in museums.
A&R15	2018	Peter Lester	Of mind and matter: the archive as object.
A&R16	2018	Anne Gilliland	To what lengths the 'Physical and Moral Defence of the Record' in times of conflict and exigency? * .
A&R17	2018	John Alexander McEwan	The past, present and future of sigillography: towards a new structural standard for seal catalogues.

FONTE: elaboração própria.

### 3.5.2.9 *Cadernos BAD*

Os seguintes dados da revista *Cadernos BAD* (ISSN: 0007-9421) foram recolhidos a 4 de outubro de 2019. De um universo de 201 artigos publicados entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 16 títulos constados na Tabela 25, num total de 8%.

Tabela 25 — Artigos científicos retirados dos *Cadernos BAD*.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
BAD1	2011	Olivier Le Deuff	Library 2.0 and the Culture of information: new paradigms?
BAD2	2014	Pedro Príncipe & Eloy Rodrigues	Infraestrutura OpenAIRE: serviço para implementação do Acesso Aberto no Horizonte 2020.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
BAD3	2014	Maria Amante, Susana Lopes, Bruno Marçal & Teresa Segurado	A interoperabilidade entre o Repositório e um sistema CRIS: o caso do ISCTE-IUL.
BAD4	2014	Diego Macêdo, Milton Shintaku, Washington de Carvalho Segundo & Ronnie de Brito	Biblioteca Digital Brasileira de Teses e Dissertações: ações para melhoria na qualidade dos dados.
BAD5	2014	Fátima dos Santos	Interoperabilidade semântica entre repositórios de saúde pública.
BAD6	2014	Maria da Graça de Melo Simões, M. Cristina V. de Freitas, Blanca Rodríguez Bravo & Sofia Gomes	OPAC e repositórios institucionais: uma análise preliminar sob a perspetiva da interoperabilidade.
BAD7	2014	Morgana Andrade, Décio Junior, Gleice Pereira & Ana Baptista	Esquemas de metadados utilizados por repositórios digitais científicos e de objetos educacionais: estudo preliminar.
BAD8	2015	Henrique dos Santos	Políticas de preservação digital para documentos arquivísticos.
BAD9	2015	Juliete Ju, Oswaldo Oswaldo	O poder da informação na sociedade da informação e nas organizações empresariais.
BAD10	2016	Lucas Paganine & Michelli Costa	Repositórios de dados de pesquisa para as ciências da saúde.
BAD11	2016	Ricarddo Amorim & Fábio Silva	Integração de plataformas institucionais: o caso EUDAT.
BAD12	2016	Yulia Karimova João Aguiar Castro	Vocabulários controlados na descrição de dados de investigação no Dendro.
BAD13	2016	Cristina Ribeiro, João Rocha da Silva, João Aguiar Castro, Ricardo Carvalho Amorim & João Correia Lopes	Projeto TAIL - Gestão de dados de investigação da produção ao depósito e à partilha (resultados preliminares).

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
BAD14	2018		Proposta de Elementos de Metadados para Representação e Recuperação de Memória Técnica: o caso da Rede Ufes-Rio Doce.
BAD15	2018	Morgana Andrade, Milton Shintaku & Patrícia de Barros	Acervo de preservação de crônicas brasileiras: um caso de interoperabilidade da informação entre sistemas.
BAD16	2018	Júlia dos Santos	Organização da informação na saúde: uma análise dos padrões de metadados brasileiros.

FONTE: elaboração própria.

### 3.5.2.10 *Journal of Archival Organization*

Os seguintes dados da revista *Journal of Archival Organization* (ISSN: 1533-2748) foram recolhidos a 7 de outubro de 2019. De um universo de 141 artigos publicados entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 37 títulos constados na Tabela 26, num total de 26%.

Tabela 26 — Artigos científicos retirados do *Journal of Archival Organization*.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
JAO1	2009	Bem Primer	Resources for Archives: Developing Collections, Constituents, Colleagues, and Capital.
JAO2	2009	Deborah Kaplan	Choosing a Digital Asset Management System That's Right for You.
JAO3	2009	Thomas Frusciano	Continuing the Discussion on Descriptive Standards.
JAO4	2009	Heather MacNeil	Trusting Description: Authenticity, Accountability, and Archival Description Standards.
JAO5	2009	Elizabeth Dow	Encoded Archival Description as a Halfway Technology.
JAO6	2009	Jean Dryden	A Review of “Two New ICA Descriptive Standards: ISDF and ISDIAH”.
JAO7	2009	Jean Dryden	The Open Archival Information System Reference Model.
JAO8	2010	Joyce Chapman	Observing Users: An Empirical Analysis of User Interaction with Online Finding Aids.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
JAO9	2010	Jason Speck	Protecting Public Trust: An Archival Wake-Up Call.
JAO10	2010	Stephanie Crowe & Karen Spilman	MPLP @ 5: More Access, Less Backlog?
JAO11	2010	Robert Cox	Maximal Processing, or, Archivist on a Pale Horse.
JAO12	2010	Jean Dryden	A Structure Standard for Archival Context: EAC-CPF is Here.
JAO13	2010	Dennis Meissner & Mark Greene	More Application while Less Appreciation: The Adopters and Antagonists of MPLP.
JAO14	2010	Cory Nimer	RDA and Archives.
JAO15	2010	Jean Dryden	Standards: News, Progress Reports, and Reviews.
JAO16	2011	Gordon Daines III & Cory Nimer	Re-Imagining Archival Display: Creating User-Friendly Finding Aids.
JAO17	2011	Shannon Maier	MPLP and the Catalog Record as a Finding Aid.
JAO18	2011	Jean Dryden	Measuring Trust: Standards for Trusted Digital Repositories.
JAO19	2011	Lina Bountouri & Manolis Gergatsoulis	The Semantic Mapping of Archival Metadata to the CIDOC CRM Ontology.
JAO20	2011	Michelle Mascaro	Controlled Access Headings in EAD Finding Aids: Current Practices in Number of and Types of Headings Assigned
JAO21	2012	Jane Zhang	Archival Representation in the Digital Age.
JAO22	2012	Patrick le Boeuf	Modeling Rare and Unique Documents: Using FRBR OO/CIDOC CRM.
JAO23	2012	Lars-Erik Hansen & Anneli Sundqvist	To Make Archives Available Online: Transcending Boundaries or Building Walls?
JAO24	2013	Matt Gorzalski	Examining User-Created Description in the Archival Profession.
JAO25	2015	Geoffrey Yeo	Contexts, Original Orders, and Item-Level Orientation: Responding Creatively to Users' Needs and Technological Change.
JAO26	2015	Jiufang Niu	Archival Intellectual Control in the Digital Age.
JAO27	2015	Christian James & Ricardo Punzalan	Legacy Matters: Describing Subject-Based Digital Historical Collections.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
JAO28	2015	Jean Dryden	Standards Don't Stand Still.
JAO29	2015	Richard Collier & Mary Samouelian	Playing in the Sandbox: EAC-CPF Projects at Duke University.
JAO30	2015	Ricardo Eítuo-Brun	A Metadata Infrastructure for a Repository of Civil Engineering Records: EAC-CPF as a Cornerstone for Content Publishing.
JAO31	2015	Daniel Pitti, Rachel Hu, Ray Larson, Brian Tingle & Adrian Turner	Social Networks and Archival Context: From Project to Cooperative Archival Program.
JAO32	2015	Florence Clavaud	Building a Knowledge Base on Archival Creators at the National Archives of France: Issues, Methods, and Prospects.
JAO33	2015	Gavan McCarthy, Ailie Smith & Michael Jones	Looking Beyond the Archive: Utilizing Encoded Archival Context in a Broader Societal Context.
JAO34	2016	Greg Bak	Not Meta Just Data: Redefining Content and Metadata in Archival Theory and Practice.
JAO35	2016	Ruth Tilman	Opportunities for Encoding EAD for Linked Data Extraction and Publication.
JAO36	2017	Rachel Searcy	Beyond Control: Accessioning Practices for Extensible Archival Management.
JAO37	2019	Sharmila Bhatia & Wright de Hernandez	Blockchain Is Already Here. What Does That Mean for Records Management and Archives?

FONTE: elaboração própria.

### 3.5.2.11 Provenance

Os seguintes dados da revista *Provenance* (ISSN: 0739-4241) foram recolhidos a 7 de outubro de 2019. De um universo de 23 artigos publicados entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 3 títulos constados na Tabela 27, num total de 13%.

Tabela 27 — Artigos científicos retirados da *Provenance*.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
PROV1	2010	Kathleen Roe	Let's Give Them Something to Talk About: Advocating for Archives.



<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
PROV2	2010	Richard Pearse-Moses	Through to Cyberspace: And What Janus Found There.
PROV3	2010	Christine de Catanzaro	Over, Under, Around, and Through: Getting Around Barriers to EAD Implementation.

FONTE: elaboração própria.

### 3.5.2.12 *Records Management Journal*

Os seguintes dados da revista *Records Management Journal* (ISSN: 0956-5698) foram recolhidos a 7 de outubro de 2019. De um universo de 189 artigos publicados entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 47 títulos constados na Tabela 28, num total de 25%.

Tabela 28 — Artigos científicos retirados do *Records Management Journal*.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
RMJ1	2009	Nick Baratt	From memory to digital record: Personal heritage and archive use in the twenty-first century.
RMJ2	2009	Pekka Henttonen	A comparison of MoReq and SÄHKE metadata and functional requirements.
RMJ3	2009	Steve Bailey	Forget electronic records management, it's automated records management that we desperately need.
RMJ4	2009	Philipp Wilhelm	An evaluation of MoReq2 in the context of national EDRMS standard developments in the UK and Europe.
RMJ5	2009	Aliza Ismail & Adnan Jamaludin	Towards establishing a framework for managing trusted records in the electronic environment.
RMJ6	2010	Kate Cumming	Ways of seeing: contextualising the continuum.
RMJ7	2010	John McDonald	Records management and data management: closing the gap.
RMJ8	2010	David Best	The future of information management.
RMJ9	2010	Luciana Duranti	Concepts and principles for the management of electronic records, or records management theory is archival diplomatics.
RMJ10	2010	Susan Healy	ISO 15489 Records Management: its development and significance.
RMJ11	2010	David Ryan	What is the essence of records management?
RMJ12	2010	Rick Barry	Opinion piece – electronic records: now and then.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
RMJ13	2010	Lawrence Serewicz	Do we need bigger buckets or better search engines? The challenge of unlimited storage and semantic web search for records management.
RMJ14	2010	Katharine Stuart & David Bromage	Current state of play: records management and the cloud.
RMJ15	2010	Kate Cumming & Cassie Findlay	Digital recordkeeping: are we at a tipping point?
RMJ16	2010	Steve Bailey & Jay Vidyarthi	Human-computer interaction: the missing piece of the records management puzzle?
RMJ17	2011	Geoffrey Yeo	Rising to the level of a record? Some thoughts on records and documents.
RMJ18	2011	Maria-Manuela Cabero, Maria-Paz Martín-Pozuelo & José Bonal Zazo	ISO 15489 and other standardized management systems: analogies and synergies.
RMJ19	2011	Jan Ashkhoz, Shigeo Sugimoto & Mitsuharu Nagamori	Preserving records in the cloud.
RMJ20	2012	Pauline Joseph, Shelda Debowski, Peter Goldschmidt	Paradigm shifts in recordkeeping responsibilities: implications for ISO 15489's implementation.
RMJ21	2012	Julie McLeod	Thoughts on the opportunities for records professionals of the open access, open data agenda.
RMJ22	2012	Maria Kallberg	Archivists 2.0: redefining the archivist's profession in the digital age.
RMJ23	2013	Frank Upward, Barbara Reed, Gillian Oliver & Joanne Evans	Recordkeeping informatics: re-figuring a discipline in crisis with a single minded approach.
RMJ24	2013	Julie McLeod & Sue Childs	A strategic approach to making sense of the "wicked" problem of ERM.
RMJ25	2013	Sue Childs & Julie McLeod	Tackling the wicked problem of ERM: using the Cynefin framework as a lens.
RMJ26	2014	Gillian Oliver	International records management standards: the challenges of achieving consensus.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
RMJ27	2014	Sue Childs, Julie McLeod, Elizabeth Lomas & Glenda Cook	Opening research data: issues and opportunities.
RMJ28	2014	Erik Borglund & Tove Engvall	Open data? Data, information, document or record?
RMJ29	2015	Corinne Rogers	Diplomatics of born digital documents - considering documentary form in a digital environment.
RMJ30	2015	Frederick Cohen	Digital diplomatics and forensics: going forward on a global basis.
RMJ31	2015	Basma Shabou	Digital diplomatics and measurement of electronic public data qualities: What lessons should be learned?
RMJ32	2015	Jane Zhang	Correspondence as a documentary form, its persistent representation, and email management, preservation, and access.
RMJ33	2015	Ruben Verborgh, Seth van Hooland, Aaron Cope, Sebastian Chan, Erik Mannens, Rik Van de Walle	The fallacy of the multi-API culture: Conceptual and practical benefits of Representational State Transfer (REST).
RMJ34	2015	Silvio Peroni, Alexander Dutton, Tanya Gray & David Shotton	Setting our bibliographic references free: towards open citation data.
RMJ35	2015	Anika Ludwig & Mary Marshall	Using crime data in academic research: issues of comparability and integrity.
RMJ36	2016	Victoria Lemieux	Trusting records: is Blockchain technology the answer?
RMJ37	2016	Wei Guo, Yun Fang, Weimei Pan & Dekun Li	Archives as a trusted third party in maintaining and preserving digital records in the cloud environment.
RMJ38	2017	Laurence Maroye, Seth Van Hooland, Fiona Celorrio, Sébastien Soye, Bénédicte Losdyck, Odile Vanreck,	Managing electronic records across organizational boundaries: The experience of the Belgian federal government in automating investigation processes.

<b>Item</b>	<b>Ano</b>	<b>Autor(es)</b>	<b>Título</b>
		Cécile de Terwangne	
RMJ39	2017	Asen Ivanov	Practice theory: a new approach for archival and recordkeeping research.
RMJ40	2017	Gregory Rolan	Towards interoperable recordkeeping systems.
RMJ41	2017	Rebecca Grant	Recordkeeping and research data management: a review of perspectives.
RMJ42	2017	Jaana Pinnick	Exploring digital preservation requirements: A case study from the National Geoscience Data Centre (NGDC).
RMJ43	2017	Giovanni Carta	Metadata and video games emulation: an effective bond to achieve authentic preservation?
RMJ44	2018	Natália Nascimento, María Moro Cabero, Marta Valentim	The adoption of ISO standards in Brazil, Iberian Peninsula and United Kingdom in information and documentation.
RMJ45	2019	Elizabeth Sheperd, Jenny Bunn, Andrew Flinn, Elizabeth Lomas, Anna Sexton, Sara Brimble, Katherin Chorley, Emma Harrison, James Lowry & Jessica Page	Open government data: critical information management perspectives.
RMJ46	2019	Frank Upward	The monistic diversity of continuum informatics: A method for analysing the relationships between recordkeeping informatics, ethics and information governance.
RMJ47	2019	Tove Engvall	Situating trust challenges of online trade.

FONTE: elaboração própria.

### **3.5.3 Seleção de monografias**

A seleção das monografias a analisar partiu das resenhas críticas publicadas em cada uma das 12 revistas identificadas no Quadro 1. Para cada revista, consultaram-se todos os números publicados entre 1 de janeiro 2009 e 31 de dezembro de 2019 e, para cada número, foram escrutinadas todas as resenhas críticas publicadas. Identificam-se ao longo das próximas tabelas todas as resenhas encontradas nas revistas, no período selecionado, de acordo com os seguintes critérios de pesquisa:

Date: 01-2009 a 12-2019  
Document type: book review  
Language: all  
ISSN: [o de cada revista]

A recolha decorreu entre finais de setembro de 2019 e inícios de outubro 2019 na base de dados LISTA. Para cada resenha indica-se: (1) um identificador único; (2) o ano de publicação da resenha; (3) os autores da resenha; (4) o ano de publicação do livro na sua edição mais recente, indicada entre parênteses; (5) o autor do livro; (6) o título do livro; (7) a prioridade de consulta e (8) as notas. Ao contrário dos artigos científicos, que apenas são publicados uma vez em cada revista, podem encontrar-se resenhas acerca de uma mesma monografia em várias revistas, uma vez que não apresentam carácter de unicidade. Diferentes autores em diferentes anos podem formular e publicar a sua opinião sobre determinada monografia. Assim, no caso desta fonte documental, foi acrescentado um campo ‘Notas’ que assinala se a monografia em questão já se encontra identificada num momento anterior na pesquisa. Recordar-se que o interesse desta recolha é identificar títulos de monografias, e não tanto em ler e analisar as diferentes resenhas. Assim, a partir do momento em que uma monografia surge numa revista A, quando esta surge numa revista seguinte B é considerada uma entrada duplicada. No final apresenta-se uma lista com a totalidade das monografias identificadas, eliminados os duplicados. A escolha das obras assentou num critério qualitativo através do qual se analisou o título e a resenha de modo a determinar se o tema do artigo seria potencialmente relevante para responder às questões de investigação.

Além das notas, decidiu-se também classificar cada monografia quanto ao nível de prioridade de consulta, de acordo com três graus: *high* (H), *medium* (M) e *low* (L). A prioridade é entendida como a adequação do tema às questões de investigação e aos objetivos propostos, isto é, procuraram-se monografias que abordassem, independentemente do ângulo, pelo menos um dos seguintes temas: a representação da informação, a descrição arquivística, os metadados, e a autenticidade. A classificação tem por base a consulta do título da obra, da resenha crítica e, por vezes, da tabela de conteúdos caso se encontre disponível na Internet. Considera-se importante atribuir um grau de prioridade de consulta às monografias porque, ao contrário dos artigos científicos para os quais foi assegurada a possibilidade de acesso, não é garantido que se tenha acesso a uma cópia dada a dificuldade acrescida de acesso integral a monografias em virtude do seu custo e a menor disponibilidade em bibliotecas. O principal método de consulta foi a disponibilidade em bibliotecas universitárias. As obras de baixa prioridade não existentes em bibliotecas tenderam a não ser consultadas. Quanto a obras de prioridade média e alta, muitas foram adquiridas a título pessoal de forma a garantir o acesso.

### 3.5.3.1 *American Archivist*

Os seguintes dados da revista *American Archivist* (ISSN: 3609081) foram recolhidos a 24 de setembro de 2019. De um universo de 151 *book reviews* publicadas entre 1 de janeiro 2009 e 31

de dezembro de 2019, foram considerados pertinentes os 11 títulos constados na Tabela 29, num total de 7%.

Tabela 29 — *Book reviews* retiradas da *American Archivist*.

<b>Item</b>	<b>Ano <i>review</i></b>	<b>Autor <i>review</i></b>	<b>Ano <i>livro</i></b>	<b>Autor(es) livro</b>	<b>Título</b>	<b>Prio.</b>	<b>Nota</b>
AA- BR1	2009	Greg Bak	2008	Steve Bailey	Managing the Crowd: Rethinking Records Management for the Web 2.0 World.	M	
AA- BR2	2009	David Susan	2008 (3°)	Jackie Bettington et al. (eds)	Keeping Archives.	L	
AA- BR3	2009	Harvey Ross	2008	Ziming Liu	Paper to Digital: Documents in the Digital Age.	M	
AA- BR4	2010	Michelle Belden	2010	Kate Theimer	Web 2.0 Tools and Strategies for Archives and Local History Collections.	L	
AA- BR5	2011	Elizabeth Yakel	2017 (2°)	Terry Eastwood & Heather MacNeil	Currents of Archival Thinking.	M	
AA- BR6	2012	Joseph Turrini	2010	Laura Millar	Archives: Principles and Practices.	M	
AA- BR7	2012	Lori Birrell	2011	Kate Theimer	A Different Kind of Web: New Connections between Archives and Our Users.	L	
AA- BR8	2013	Pam Hackbart- Dean	2013	Thomas Frusciano & Christopher Prom (eds)	Archival Arrangement and Description.	H	
AA- BR9	2015	Paul Conway	2014	Anne Gilliland	Conceptualizing 21st-Century Archives.	H	
AA- BR10	2017	Olga Virakhovskaya	2014	Kate Theimer	Description: Innovative Practices for Archives and Special Collections.	M	
AA- BR11	2017	Creighton Barrett	2016	Victoria Lemieux	Building Trust in Information: Perspectives on the Frontiers of Provenance	H	
AA- BR12	2018	Geoffrey Yeo	2016	Anne J. Gilliland, Sue McKemmish, Andrew J. Lau (eds)	Research in the archival multiverse	H	
AA- BR13	2018	Carly Deaborn	2016	Angela Dappert, Rebecca Squire	Digital Preservation Metadata for Practitioners: Implementing PREMIS.	L	

<b>Item</b>	<b>Ano</b> <i>review</i>	<b>Autor</b> <i>review</i>	<b>Ano</b> <b>livro</b>	<b>Autor(es) livro</b>	<b>Título</b>	<b>Prio.</b>	<b>Nota</b>
				Guenther & Sébastien Peyrard			

FONTE: elaboração própria.

### 3.5.3.2 *Archival Issues*

Os seguintes dados da revista *Archival Issues* (ISSN: 1067-4993) foram recolhidos a 25 de setembro de 2019. De um universo de 125 *book reviews* publicadas entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 16 títulos constados na Tabela 30, num total de 13%.

Tabela 30 — *Book reviews* retiradas da *Archival Issues*.

<b>Item</b>	<b>Ano</b> <i>review</i>	<b>Autor</b> <i>review</i>	<b>Ano</b> <b>livro</b>	<b>Autor(es) livro</b>	<b>Título</b>	<b>Prio.</b>	<b>Nota</b>
AI- BR1	2010	Colleen McFarland	2009	Gordon Daines & Cory Nimer	The Interactive Archivist: Case Studies in Utilizing Web 2.0 to Improve the Archival Experience.	M	
AI- BR2	2011	Nicole Garrett	2010	Kate Theimer	Web 2.0 Tools and Strategies for Archives and Local History Collections	-	Dup. (AA)
AI- BR3	2011	Jean Dryden	2010	Laura Millar	Archives: Principles and Practices.	-	Dup. (AA)
AI- BR4	2012	Eric Fritzler	2011	Steven Miller	Metadata for Digital Collections: A How-To-Do-It Manual.	H	
AI- BR5	2012	Joshua Zimmerman	2011	Kate Theimer	A Different Kind of Web: New Connections Between Archives and Our Users.	-	Dup. (AA)
AI- BR6	2014	Eric Fritzler	2013	Thomas Frusciano & Christopher Prom (eds)	Archival Arrangement and Description.	-	Dup. (AA)
AI- BR7	2015	Susan Tucker	2014	Anne Gilliland	Conceptualizing 21st-Century Archives	-	Dup. (AA)
AI- BR8	2015	Erin Wolfe	2014	Ruben Verborgh & Seth Van Hooland	Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata.	H	
AI- BR9	2015	Genya O’Gara	2014	Edward Corrado & Moulaison Sandy	Digital Preservation for Libraries, Archives, and Museums	M	

<b>Item</b>	<b>Ano review</b>	<b>Autor review</b>	<b>Ano livro</b>	<b>Autor(es) livro</b>	<b>Título</b>	<b>Prio.</b>	<b>Nota</b>
AI- BR10	2015	Lisa Sjoberg	2015	Ian Ruthven & Gobinda Chowdhury (eds)	Cultural Heritage Information: Access and management	H	
AI- BR11	2017	Lucas Arel	2015	Eric Mitchell	Metadata Standards and Web Services in Libraries, Archives, and Museums.	M	
AI- BR12	2017	Gregory Wiedeman	2008	Marcia Lei Zeng & Jian Qin	Metadata	H	
AI- BR13	2017	Erin Wolfe	2016	Philip Bantin	Building Trustworthy Digital Repositories: Theory and Implementation	M	
AI- BR14	2018	Ryan Leimkuehler	2017	Lois Hamil	Archival Arrangement and Description: Analog to Digital.	H	
AI- BR15	2018	Nathan Saunders	2017	Laura Millar	Archives: Principles and Practices.	-	Dup. (AA)
AI- BR16	2018	Tamar Zeffren	2017	Kris Kiesling & Christopher Prom	Putting Descriptive Standards to Work.	H	

FONTE: elaboração própria.

### 3.5.3.3 *Archival Science*

Os seguintes dados da revista *Archival Science* (ISSN: 013890166) foram recolhidos a 25 de setembro de 2019. Para o intervalo de pesquisa em 1 de janeiro 2009 e 31 de dezembro de 2019 não foi encontrada nenhuma resenha.

### 3.5.3.4 *Archivaria*

Os seguintes dados da revista *Archivaria* (ISSN: 03186954) foram recolhidos a 30 de setembro de 2019. De um universo de 87 *book reviews* publicadas entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 14 títulos constados na Tabela 31, num total de 16%.

Tabela 31 — *Book reviews* retiradas da *Archivaria*.

<b>Item</b>	<b>Ano review</b>	<b>Autor review</b>	<b>Ano livro</b>	<b>Autor(es) livro</b>	<b>Título</b>	<b>Prio.</b>	<b>Nota</b>
Archivaria- BR1	2011	Caroline Williams	2017 (2°)	Terry Eastwood & Heather MacNeil	Currents of Archival Thinking.	-	Dup. (AA)



<b>Item</b>	<b>Ano review</b>	<b>Autor review</b>	<b>Ano livro</b>	<b>Autor(es) livro</b>	<b>Título</b>	<b>Prio.</b>	<b>Nota</b>
Archivaria- BR2	2011	Jim Suderman	2009	Graham Underwood & Jonathan Penner	Electronic Evidence in Canada.	L	
Archivaria- BR3	2011	Karine Burger	2010	Kate Theimer	Web 2.0 Tools and Strategies for Archives and Local History Collections.	-	Dup. (AA e AI)
Archivaria- BR4	2012	Terry Cook	2011	Jennie Hill	The Future of Archives and Recordkeeping: A Reader.	M	
Archivaria- BR5	2012	Jennifer Whitlock		Steven Miller	Metadata for Digital Collections: A How- To-Do-It Manual.	-	Dup. (AI)
Archivaria- BR6	2012	Kathryn Harvey		Kate Theimer	A Different Kind of Web: New Connections between Archives and Our Users.	-	Dup. (AA e AI)
Archivaria- BR7	2013	Kathryn Harvey	2012	Tomas Lidman	Libraries and Archives: A Comparative Study.	L	
Archivaria- BR8	2015	Donald Force	2014	Edward Corrado & Moulaison Sandy	Digital Preservation for Libraries, Archives, and Museums.	-	Dup. (AI)
Archivaria- BR9	2015	Elizabeth Sheperd	2014	Anne Gilliland	Conceptualizing 21st-Century Archives.	-	Dup. (AA e AI)
Archivaria- BR10	2016	Juan Ilerbaig	2015	Luciana Duranti & Patricia Franks (eds)	Encyclopedia of Archival Science.	H	
Archivaria- BR11	2017	Lisa Snider	2015	Steve Marks	Module 8: Becoming a Trusted Digital Repository.	L	
Archivaria- BR12	2017	Jonathan Dorey	2017 (2º)	Terry Eastwood & Heather MacNeil	Currents of Archival Thinking.	-	Dup. (AA)
Archivaria- BR13	2017	Rebecka Sheffield	2016	Anne Gilliland, Sue McKemmish & Andrew Lau (eds)	Research in the archival multiverse.	-	Dup. (AA)
Archivaria- BR14	2018	Eric Stoykovich	2018	Paul Delsalle & Margaret Procter	A History of Archival Practice.	L	

FONTE: elaboração própria.

### 3.5.3.5 Archives (London)

Os seguintes dados da revista *Archives (London)* (ISSN: 00039535) foram recolhidos a 30 de setembro de 2019. De um universo de 259 *book reviews* publicadas entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 3 títulos constados na Tabela 32, num total de 1%.

Tabela 32 — *Book reviews* retiradas da *Archives (London)*.

<b>Item</b>	<b>Ano review</b>	<b>Autor review</b>	<b>Ano livro</b>	<b>Autor(es) livro</b>	<b>Título</b>	<b>Prio.</b>	<b>Nota</b>
Arch. (LDN)- BR1	2015	Hazel Anderson	2015	Steve Marks	Module 8: Becoming a Trusted Digital Repository.	-	Dup. (Archivaria)
Arch. (LDN)- BR2	2015	Robin Darwall- Smith	2015	Jeannette Bastian, Megan Sniffin- Marinoff & Donna Webber	Archives in libraries: What librarians and archivists need to know to work together.	L	
Arch. (LDN)- BR3	2017	Geoffrey Yeo	2016	Fiorella Foscarini, Heather MacNeil, Bonnie Mak & Gillian Oliver	Engaging with Records and Archives: Histories and theories	L	

FONTE: elaboração própria.

### 3.5.3.6 Archives (Quebec)

Os seguintes dados da revista *Archives (Quebec)* (ISSN: 0044-9423) foram recolhidos a 2 de outubro de 2019. Para o intervalo de pesquisa em 1 de janeiro 2009 e 31 de dezembro de 2019 não foi encontrada nenhuma resenha.

### 3.5.3.7 Archives and Manuscripts

Os seguintes dados da revista *Archives and Manuscripts* (ISSN: 0157-6895) foram recolhidos a 3 de outubro de 2019. De um universo de 166 *book reviews* publicadas entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 34 títulos constados na Tabela 33, num total de 20%.

Tabela 33 — *Book reviews* retiradas da *Archives & Manuscripts*.

<b>Item</b>	<b>Ano review</b>	<b>Autor review</b>	<b>Ano livro</b>	<b>Autor(es) livro</b>	<b>Título</b>	<b>Prio.</b>	<b>Nota</b>
A&M- BR1	2010	Michal Piggot	2009	John Ridener	From Polders to Postmodernism: A Concise History of Archival Theory.	H	

<b>Item</b>	<b>Ano <i>review</i></b>	<b>Autor <i>review</i></b>	<b>Ano <i>livro</i></b>	<b>Autor(es) <i>livro</i></b>	<b>Título</b>	<b>Prio.</b>	<b>Nota</b>
A&M- BR2	2011	Peter Horseman	2010	Adrian Cunningham (ed)	The arrangement and description of archives amid administration and technological change: essays and reflections by and about peter J scott.	M	
A&M- BR3	2011	Kate Cummng	2010	Adrian Cunningham (ed)	The arrangement and description of archives amid administration and technological change: essays and reflections by and about peter J scott.	-	Dup. (A&M)
A&M- BR4	2011	Sue McKemmish	2017 (2º)	Terry Eastwood & Heather MacNeil	Currents of Archival Thinking.	-	Dup. (AA)
A&M- BR5	2011	Sigrid McClausland	2010	Laura Millar	Archives: Principles and Practices.	-	Dup. (AA e AI)
A&M- BR6	2011	Anne-Marie Coné	2011	Francis X. Blouin Jr & William G. Rosenberg	Processing the past: contesting authority in history and the archives.	L	
A&M- BR7	2011	Jackie Bettington	2011	Jennie Hill	The Future of Archives and Recordkeeping: A Reader.	-	Dup. (Archivaria)
A&M- BR8	2012	Nea Fitzgerald	2011	James Gleick	The Information: A History, a Theory, a Flood	M	
A&M- BR9	2012	Andrew Wilson	2011	David Giaretta	Advanced Digital Preservation.	L	
A&M- BR10	2013	Daniel Wilkish	2011	Kate Theimer	A Different Kind of Web: New Connections between Archives and Our Users.	-	Dup. (AA, AI e Archivaria)
A&M- BR11	2013	Maggie Shapley	2013	Thomas Frusciano & Christopher Prom (eds)	Archival Arrangement and Description.	-	Dup. (AA e AI)
A&M- BR12	2014	Julie Shanks	2013	Patricia Franks	Records and Information Management.	L	
A&M- BR13	2014	Maureen Sullivan	2012	Tomas Lidman	Libraries and Archives: A Comparative Study.	-	Dup. (Archivaria)

<b>Item</b>	<b>Ano review</b>	<b>Autor review</b>	<b>Ano livro</b>	<b>Autor(es) livro</b>	<b>Título</b>	<b>Prio.</b>	<b>Nota</b>
A&M- BR14	2015	Amie Baron	2014	Anne Gilliland	Conceptualizing 21st-Century Archives.	-	Dup. (AA, AI e Archivaria)
A&M- BR15	2015	Avid Fricker	2014	Gillian Oliver & Fiorella Foscarini	Records Management and Information Culture: Tackling the People Problem.	L	
A&M- BR16	2015	Rowena Loo	2013	Caroline Brown	Records Management and Information Culture: Tackling the People Problem.	M	
A&M- BR17	2016	Christine Holmes	2014	Ross Harvey & Martha Mahard	The Preservation Management Handbook: A 21st-Century Guide for Libraries, Archives and Museums.	M	
A&M- BR18	2016	Richard Lehane	2015	Steve Marks	Module 8: Becoming a Trusted Digital Repository.	-	Dup. (Archivaria)
A&M- BR19	2016	Lise Summers	2015	Ian Ruthven & G. G. Chowdhury (eds)	Cultural Heritage Information: Access and management.	-	Dup. (AI)
A&M- BR20	2016	Michel Piggot	2015	Luciana Duranti & Patricia Franks (eds)	Encyclopedia of Archival Science.	-	Dup. (Archivaria)
A&M- BR21	2016	David Roberts	2014	Kate Theimer	Description: Innovative Practices for Archives and Special Collections.	-	Dup. (AA)
A&M- BR22	2016	Meg Travers	2012	Michèle Valerie Cloonan	Preserving our heritage: perspectives from antiquity to the digital age.	L	
A&M- BR23	2016	David Bromage	2014	Ruben Verborgh & Seth Van Hooland	Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata.	-	Dup. (AI)
A&M- BR24	2016	Barbara Reed	2015	Wolfgang Ernst	Stirrings in the archives: order from disorder.	L	
A&M- BR25	2016	Mike Summerell	2015	Michael Moss & Barbara Endicott-Popovsky	Is digital different? How information creation, capture and discovery are being transformed.	M	

<b>Item</b>	<b>Ano <i>review</i></b>	<b>Autor <i>review</i></b>	<b>Ano <i>livro</i></b>	<b>Autor(es) <i>livro</i></b>	<b>Título</b>	<b>Prio.</b>	<b>Nota</b>
A&M- BR26	2016	Allie Smith	2015	Jeannette A. Bastian, Megan Sniffin-Marinoff & Donna Webber	Archives in Libraries: What Librarians and Archivists Need to Know to Work Together.	-	Dup. (Arch [LDN])
A&M- BR27	2017	Karen Horsefall	2016	Philip Bantin	Building Trustworthy Digital Repositories: Theory and Implementation.	-	Dup. (AI)
A&M- BR28	2017	Simon Wall	2015	Kristin Briney	Data Management for Researchers: Organize, Maintain and Share Your Data for Research Success.	M	
A&M- BR29	2017	John Machin	2016	David Stuart	Practical Ontologies for Information Professionals.	H	
A&M- BR30	2017	Karen Anderson	2016	Anne J. Gilliland, Sue McKemmish & Andrew J. Lau (eds)	Research in the archival multiverse.	-	Dup. (AA e Archivaria)
DA&M- BR31	2017	Christopher Colwell	2016	Ed Jones & Michele Seikel	Linked data for cultural heritage.	H	
A&M- BR32	2017	Jaye Weatherburn	2016	Allen Foster & Pauline Rafferty	Managing digital cultural objects: analysis, discovery and retrieval.	L	
A&M- BR33	2018	Aliisa McCulloh	2016	Ed Jones & Michele Seikel	Linked data for cultural heritage.	-	Dup. (AA e Archivaria)
A&M- BR34	2018	Jackie Bettington	2016	Victoria Lemieux (ed.)	Building Trust in Information: Perspectives on the Frontiers of Provenance.	-	Dup. (AA)
A&M- BR34	2018	Kristen Wright	2016	Fiorella Foscarini & Heather MacNeil	Engaging with Records and Archives: Histories and Theories.	-	Dup. (Arch [LDN])

FONTE: elaboração própria.

### 3.5.3.8 Archives and Records

Os seguintes dados da revista *Archives and Records* (ISSN: 2325-7962) foram recolhidos a 4 de outubro de 2019. De um universo de 146 *book reviews* publicadas entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 31 títulos constados na Tabela 34, num total de 21%.

Tabela 34 — *Book reviews* retiradas da *Archives & Records*.

<b>Item</b>	<b>Ano review</b>	<b>Autor review</b>	<b>Ano livro</b>	<b>Autor(es) livro</b>	<b>Título</b>	<b>Prio.</b>	<b>Nota</b>
A&R- BR1	2013	Margaret Procter	2011	Kate Theimer	A Different Kind of Web: New Connections between Archives and Our Users.	-	Dup. (AA, AI, Archivaria e A&M)
A&R- BR2	2013	Margaret Procter	2013	Thomas Frusciano & Christopher Prom (eds)	Archival Arrangement and Description.	-	Dup. (AA, AI e A&M)
A&R- BR3	2014	Birthe Christensen	2013	Helen Ford & Jonathan Rhys- Lewis	Preserving archives (2nd ed.).	L	
A&R- BR4	2014	Susan Graham	2012	Charles Oppenheim	The no-nonsense guide to legal issues in Web 2.0 and cloud computing.	M	
A&R- BR5	2014	Matthew Stephenson	2014	Gillian Oliver & Fiorella Foscarini	Records Management and Information Culture: Tackling the People Problem.	-	Dup. (A&M)
A&R- BR6	2014	Sarah Millard & Margherita Orlando	2013	Caroline Brown	Archives and Recordkeeping: Theory into Practice.	-	Dup. (A&M)
A&R- BR7	2015	Lee Pretlove	2013	Patricia Franks	Records and Information Management.	-	Dup. (A&M)
A&R- BR8	2015	Timothy Gollins	2014	Louise Corti, Veerle Van den Eynden, Libby Bishop & Matthew Woollard	Managing and sharing research data, a guide to good practice.	M	

<b>Item</b>	<b>Ano review</b>	<b>Autor review</b>	<b>Ano livro</b>	<b>Autor(es) livro</b>	<b>Título</b>	<b>Prio.</b>	<b>Nota</b>
A&R- BR9	2015	Jane Maxwell	2013	Woollard Matthew	Archives and archivists 2: current trends, new voices.	M	
A&R- BR10	2015	Elizabeth Sheperd	2012	Thomas Lidman	Libraries and Archives: A Comparative Study.	-	Dup. (Archivaria e A&M)
A&R- BR11	2015	Andrew James	2014	Ruben Verborgh & Seth Van Hooland	Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata.	-	Dup. (AI e A&M)
A&R- BR12	2015	Michael Moss	2014	Anne Gilliland	Conceptualizing 21st-Century Archives.	-	Dup. (AA, AI, Archivaria e A&M)
A&R- BR13	2016	James Lowry	2015	Luciana Duranti & Patricia Franks (eds)	Encyclopedia of Archival Science.	-	Dup. (Archivaria e A&M)
A&R- BR14	2016	Birthe Christensen	2012	Michèle Valerie Cloonan	Preserving our heritage: perspectives from antiquity to the digital age	-	Dup. (A&M)
A&R- BR15	2016	Simon Wilson	2014	Edward Corrado & Heather Sandy	Digital Preservation for Libraries, Archives, and Museums.	-	Dup. (AI e Archivaria)
A&R- BR16	2016	Jenny Mitcham	2015	Steve Marks	Module 8: Becoming a Trusted Digital Repository.	-	Dup. (Archivaria, Arch[LDN] e A&M)
A&R- BR17	2016	Tola Dabiri	2015	Jeannette Bastian, Megan Sniffin-Marinoff & Donna Webber	Archives in Libraries: What Librarians and Archivists Need to Know to Work Together.	-	Dup. (Arch[LDN] e A&M)
A&R- BR18	2016	Anthea Seles	2015	Michael Moss, Barbara Endicott- Propovsky & Marc Dupuis	Is digital different? How information creation, capture, preservation and discovery are being transformed.	-	Dup. (A&M)
A&R- BR19	2016	Alexandrina Buchanan	2015	Wolfgang Ernst	Stirrings in the archives: order from disorder.	-	Dup. (A&M)

<b>Item</b>	<b>Ano review</b>	<b>Autor review</b>	<b>Ano livro</b>	<b>Autor(es) livro</b>	<b>Título</b>	<b>Prio.</b>	<b>Nota</b>
A&R- BR20	2017	Jenny Mitcham	2016	Philip Bantin	Building Trustworthy Digital Repositories: Theory and Implementation	-	Dup. (AI e A&M)
A&R- BR21	2017	Adrian Brown	2016	Allen Foster & Pauline Rafferty (eds)	Managing digital cultural objects: analysis, discovery and retrieval.	H	
A&R- BR22	2017	Ellen O'Flaherty	2016 (3º)	Murtha Baca (ed.)	Introduction to metadata.	H	
A&R- BR23	2017	Simon Wilson	2016	Christopher Prom (ed.)	Digital preservation essentials.	M	
A&R- BR24	2018	Ellen O'Flaherty	2014	Edward Corrado & Heather Sandy	Digital Preservation for Libraries, Archives, and Museums.	-	Dup. (AI, Archivaria e A&R)
A&R- BR25	2018	Valerie Johnson	2016	Anne Gilliland, Sue McKemmis & Andrew Lau (eds)	Research in the archival multiverse.	-	Dup. (AA, Archivaria e A&M)
A&R- BR26	2018	Elizabeth Sheperd	2018	Paul Delsalle & Margaret Procter	A History of Archival Practice.	-	Dup. (Archivaria)
A&R- BR27	2018	Jone Garmendia	2017	Lois Hamil	Archival Arrangement and Description: Analog to Digital.	-	Dup. (AI)
A&R- BR28	2019	Alexandrina Buchanan	2017	Laura Millar	Archives: Principles and Practices.	-	Dup. (AA, AI e A&M)
A&R- BR29	2019	James Lowry	2017	Frans Smit, Arnoud Claudemans & Rienk Jonker (eds)	Archives in liquid times.	H	
A&R- BR30	2019	Caroline Williams	2018	Caroline Brown (ed.)	Archival futures.	H	
A&R- BR31	2019	Emily Chen	2018	Heather Ryan & Walker Sampson	The no-nonsense guide to born-digital content.	L	

FONTE: elaboração própria.



### 3.5.3.9 Cadernos BAD

Os seguintes dados da revista *Cadernos BAD* (ISSN: 0007-9421) foram recolhidos a 7 de outubro de 2019. Para o intervalo de pesquisa em 1 de janeiro 2009 e 31 de dezembro de 2019 não foi encontrada nenhuma resenha.

### 3.5.3.10 Journal of Archival Organization

Os seguintes dados da revista *Journal of Archival Organization* (ISSN: 1533-2748) foram recolhidos a 7 de outubro de 2019. De um universo de 66 *book reviews* publicadas entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 6 títulos constados na Tabela 35, num total de 9%.

Tabela 35 — *Book reviews* retiradas da *Journal of Archival Organization*.

Item	Ano <i>review</i>	Autor <i>review</i>	Ano livro	Autor(es) livro	Título	Prio.	Nota
JAO- BR1	2013	Joanne Kaczmarek	2011	Kate Theimer	A Different Kind of Web: New Connections between Archives and Our Users.	-	Dup. (AA, AI, Archivaria, A&M e A&R)
JAO- BR2	2013	Audra Yun	2013	Thomas Frusciano & Christopher Prom (eds)	Archival Arrangement and Description.	-	Dup. (AA, AI, A&M e A&R)
JAO- BR3	2016	Sibyl Schaefer	2016	Christopher Prom (ed.)	Digital preservation essentials.	-	Dup. (A&R)
JAO- BR4	2107	Lori Lindberg	2016	Philip Bantin	Building Trustworthy Digital Repositories: Theory and Implementation	-	Dup. (AI, A&M e A&R)
JAO- BR5	2017	Rosalyn Metz	2015	Steve Marks	Module 8: Becoming a Trusted Digital Repository.	-	Dup. (Archivaria, Arch[LDN], A&M e A&R)
JAO- BR6	2019	Celeste Brewer	2018	Trevor Owens	The Theory and Craft of Digital Preservation.	M	

FONTE: elaboração própria.

### 3.5.3.11 Provenance

Os seguintes dados da revista *Provenance* (ISSN: 0739-4241) foram recolhidos a 7 de outubro de 2019. De um universo de 36 *book reviews* publicadas entre 1 de janeiro 2009 e 31 de dezembro de 2019, foram considerados pertinentes os 6 títulos constados na Tabela 36, num total de 17%.

Tabela 36 — *Book reviews* retiradas da *Provenance*.

Item	Ano <i>review</i>	Autor <i>review</i>	Ano <i>livro</i>	Autor(es) livro	Título	Prio.	Nota
PROV- BR1	2010	Debra March	2010	Kate Theimer	Web 2.0 Tools and Strategies for Archives and Local History Collections	-	Dup. (AA, A&I e Archivaria)
PROV- BR2	2010	Robert Richards	2010	Randall Jimerson	Archives Power: Memory, Accountability, and Social Justice.	L	
PROV- BR3	2010	Jordon Steele	2010	Randall Jimerson	Archives Power: Memory, Accountability, and Social Justice.	-	Dup. (PROV)
PROV- BR4	2012	Angela Flenner	2011	Kate Theimer	A Different Kind of Web: New Connections between Archives and Our Users.	-	Dup. (AA, A&I, Archivaria, A&M, A&R e JAO)
PROV- BR5	2018	Tommy Brown	2018	Paul Delsalle & Margaret Procter	A History of Archival Practice.	-	Dup. (Archivaria e A&R)
PROV- BR6	2018	Amanda Hawk	2017	Lois Hamil	Archival Arrangement and Description: Analog to Digital.	-	Dup. (AI e A&R)

FONTE: elaboração própria.

### 3.5.3.12 Records Management Journal

Os seguintes dados da revista *Records Management Journal* (ISSN: 0956-5698) foram recolhidos a 7 de outubro de 2019. Para o intervalo de pesquisa em 1 de janeiro 2009 e 31 de dezembro de 2019 não foi encontrada nenhuma resenha.

### 3.5.3.13 Síntese das monografias identificadas

De forma a facilitar a compreensão das monografias que foram identificadas e selecionadas uma vez processados os dados de todas revistas científicas, na Tabela 37 apresenta-se uma lista

final de todos os 56 títulos únicos encontrados na sua primeira ordem de ocorrência, eliminando-se os duplicados.

Tabela 37 — Lista final das monografias identificadas.

<b>Item</b>	<b>Código</b>	<b>Ano</b> <b>livro</b>	<b>Autor(es)</b> <b>livro</b>	<b>Título</b>	<b>Prio.</b>
1	AA-BR1	2008	Ziming Liu	Paper to Digital: Documents in the Digital Age.	M
2	AA-BR2	2010	Kate Theimer	Web 2.0 Tools and Strategies for Archives and Local History Collections.	L
3	AA-BR3	2017 (2º)	Terry Eastwood & Heather MacNeil	Currents of Archival Thinking.	M
4	AA-BR4	2017	Laura Millar	Archives: Principles and Practices.	L
5	AA-BR5	2011	Kate Theimer	A Different Kind of Web: New Connections between Archives and Our Users.	M
6	AA-BR6	2013	Thomas Frusciano & Christopher Prom (eds)	Archival Arrangement and Description.	M
7	AA-BR7	2014	Anne Gilliland	Conceptualizing 21st-Century Archives.	L
8	AA-BR8	2014	Kate Theimer	Description: Innovative Practices for Archives and Special Collections.	H
9	AA-BR9	2016	Victoria Lemieux (ed).	Building Trust in Information: Perspectives on the Frontiers of Provenance.	H
10	AA-BR10	2016	Anne Gilliland, Sue McKemmish & Andrew J. Lau (eds)	Research in the archival multiverse.	M
11	AA-BR11	2016	Angela Dappert, Rebecca Squire Guenther & Sébastien Peyrard	Digital Preservation Metadata for Practitioners: Implementing PREMIS.	H
12	AI-BR1	2009	Gordon Daines & Cory Nimer	The Interactive Archivist: Case Studies in Utilizing Web 2.0 to Improve the Archival Experience.	M

<b>Item</b>	<b>Código</b>	<b>Ano livro</b>	<b>Autor(es) livro</b>	<b>Título</b>	<b>Prio.</b>
13	AI-BR4	2011	Steven Miller	Metadata for Digital Collections: A How-To-Do-It Manual.	H
14	AI-BR8	2014	Ruben Verborgh & Seth Van Hooland	Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata.	H
15	AI-BR9	2014	Edward Corrado & Heather Sandy	Digital Preservation for Libraries, Archives, and Museums.	M
16	AI-BR10	2015	Ian Ruthven & G. G. Chowdhury	Cultural Heritage Information: Access and management.	H
17	AI-BR11	2015	Erik Mitchell	Metadata Standards and Web Services in Libraries, Archives, and Museums.	M
18	AI-BR12	2016 (3º)	Marcia Lei Zeng & Jian Qin	Metadata.	H
19	AI-BR13	2016	Philip Bantin	Building Trustworthy Digital Repositories: Theory and Implementation.	M
20	AI-BR14	2017	Lois Hamil	Archival Arrangement and Description: Analog to Digital.	H
21	AI-BR16	2017	Nimer Wisser & Sheperd Rubinstein	Putting Descriptive Standards to Work.	H
22	Arquivaria- BR2	2009	Graham Underwood & Jonathan Penner	Electronic Evidence in Canada.	L
23	Arquivaria- BR4	2011	Jennie Hill	The Future of Archives and Recordkeeping: A Reader.	M
24	Arquivaria- BR7	2012	Thomas Lidman	Libraries and Archives: A Comparative Study.	L
25	Arquivaria- BR10	2015	Luciana Duranti & Patricia Franks (eds)	Encyclopedia of Archival Science.	H
26	Arquivaria- BR11	2015	Steve Marks	Module 8: Becoming a Trusted Digital Repository.	L

<b>Item</b>	<b>Código</b>	<b>Ano</b> <b>livro</b>	<b>Autor(es)</b> <b>livro</b>	<b>Título</b>	<b>Prio.</b>
27	Archivaria- BR14	2018	Paul Delsalle & Margaret Procter	A History of Archival Practice.	L
28	Arch. (LDN)-BR2	2015	Jeannette A. Bastian, Megan Sniffin-Marinoff & Donna Webber	Archives in libraries: What librarians and archivists need to know to work together.	L
29	Arch. (LDN)- BR3	2016	Fiorella Foscarini & Heather MacNeil	Engaging with Records and Archives: Histories and theories.	L
30	A&M-BR1	2009	John Ridener	From Polders to Postmodernism: A Concise History of Archival Theory.	H
31	A&M-BR2	2010	Adrian Cunningham (ed.)	The arrangement and description of archives amid administration and technological change: essays and reflections by and about peter J scott.	M
32	A&M-BR6	2011	Francis X. Blouin, Jr & William G. Rosenberg	Processing the past: contesting authority in history and the archives.	L
33	A&M-BR8	2011	James Gleick	The Information: A History, a Theory, a Flood.	M
34	A&M-BR9	2011	David Giaretta	Advanced Digital Preservation.	L
35	A&M- BR12	2013	Patricia Franks	Records and Information Management.	L
36	A&M- BR15	2014	Gillian Oliver & Fiorella Foscarini	Records Management and Information Culture: Tackling the People Problem.	L
37	A&M- BR16	2013	Caroline Brown	Archives and Recordkeeping: Theory into Practice.	M
L38	A&M- BR17	2014	Ross Harvey & Martha Mahard	The Preservation Management Handbook: A 21st-Century Guide for Libraries, Archives and Museums.	M
39	A&M- BR22	2012	Michèle Valerie Cloonan	Preserving our heritage: perspectives from antiquity to the digital age.	L

<b>Item</b>	<b>Código</b>	<b>Ano livro</b>	<b>Autor(es) livro</b>	<b>Título</b>	<b>Prio.</b>
40	A&M- BR24	2015	Wolfgang Ernst	Stirrings in the archives: order from disorder.	L
41	A&M- BR25	2015	Michael Moss & Barbara Endicott- Popovsky	Is digital different? How information creation, capture and discovery are being transformed.	M
42	A&M- BR28	2015	Kristin Briney	Data Management for Researchers: Organize, Maintain and Share Your Data for Research Success.	M
43	A&M- BR29	2016	David Stuart	Practical Ontologies for Information Professionals.	H
44	A&M- BR31	2016	Ed Jones & Michele Seikel	Linked data for cultural heritage.	H
45	A&M- BR32	2016	Allen Foster & Pauline Rafferty	Managing digital cultural objects: analysis, discovery and retrieval.	L
46	A&R-BR3	2013	Helen Ford & Jonathan Rhys- Lewis	Preserving archives (2nd ed.).	L
47	A&R-BR4	2012	Charles Oppenheim	The no-nonsense guide to legal issues in Web 2.0 and cloud computing.	M
48	A&R-BR8	2014	Louise Corti, Veerle Van den Eynden, Libby Bishop & Matthew Woollard	Managing and sharing research data, a guide to good practice.	M
49	A&R- BR21	2016	Allan Foster & Pauline Rafferty (eds)	Managing digital cultural objects: analysis, discovery and retrieval.	H
50	A&R- BR22	2016 (3º)	Murtha Baca (ed.)	Introduction to metadata.	H
51	A&R- BR23	2016	Christopher Prom	Digital preservation essentials.	M

Item	Código	Ano livro	Autor(es) livro	Título	Prio.
52	A&R- BR39	2017	Frans Smit, Arnoud Claudemans & Rienk Jonker (eds)	Archives in liquid times.	H
53	A&R- BR30	2018	Caroline Brown (ed.)	Archival futures.	H
54	A&R-BR	2018	Heather Ryan & Walker Sampson	The no-nonsense guide to born-digital content.	L
55	JAO-BR6	2018	Trevor Owens	The Theory and Craft of Digital Preservation.	M
56	PROV- BR2	2010	Randall Jimerson	Archives Power: Memory, Accountability, and Social Justice.	L

FONTE: elaboração própria.

### 3.5.4 Seleção de comunicações científicas

A seleção das comunicações científicas em atas de congresso teve por base os critérios de pesquisa definidos anteriormente para a base de dados *Web of Science*. A pesquisa foi realizada no dia 18 de novembro de 2019. Para cada comunicação científica indica-se: (1) um identificador único; (2) o ano da realização; (3) o(s) autor(es); (4) a designação da conferência em que foi apresentada; (5) o título; (6) a disponibilidade; e (7) a prioridade. Considera-se importante registrar a disponibilidade das comunicações científicas uma vez que a WoS é apenas um agregador de metadados que não disponibiliza acesso aos textos integrais. Por este motivo, a consulta das obras está dependente de recursos externos, nomeadamente páginas das editoras, redes sociais académicas (como ResearchGate), Biblioteca do Conhecimento Online (B-On) e pedidos pessoais endereçados aos autores. Casos há em que a única via de acesso é mediante pagamento. Nestas situações, os textos não foram consultados. A disponibilidade do texto integral de forma gratuita para os investigadores, independentemente da via de acesso, é indicada através do sinal '+’.

Devido a estes constrangimentos, atribuiu-se igualmente uma classificação às comunicações que espelha a sua prioridade de consulta, de acordo com três graus: *high* (H), *medium* (M) e *low* (L). O objetivo da prioridade é atribuir solicitar e procurar em primeiro lugar as comunicações consideradas mais pertinentes para a revisão de literatura na medida em que que abordem, independentemente do ângulo, pelo menos um dos seguintes temas: a representação da informação, a descrição arquivística, os metadados, e a autenticidade. A classificação tem por base a consulta do título e do resumo. A escolha das obras assentou num critério qualitativo através do qual se analisou o título e o resumo e de modo a determinar se o tema do artigo seria potencialmente relevante para responder às questões de investigação.

Os seguintes dados foram recolhidos a 18 de novembro de 2019. De um universo de 157 comunicações científicas publicadas em atas de conferência entre 1 de janeiro 2009 e 31 de dezembro de 2019, em inglês ou em espanhol, foram considerados pertinentes os 24 títulos constados na Tabela 38, num total de 15%.

Tabela 38 — Comunicações científicas recuperadas em língua inglesa e espanhola na WoS.

Item	Ano	Autor(es)	Título	Conferência	Prio.	Disponi.
WoS1	2009	Nicola Ferro & Gianmaria Silvello	The NESTOR Framework: How to Handle Hierarchical Data Structures.	International Conference on Theory and Practice of Digital Libraries	M	+
WoS2	2010	András Holl	Small Data Archives and Libraries.	Library and Information Services in Astronomy	M	+
WoS3	2010	Klaus Lepsky, Thomas Mueller & Jens Wille	Metadata improvement for image information retrieval.	11th International ISKO Conference	L	+
WoS4	2010	Thiago Barros & João Batista de Moraes	From archives to archival science: Elements for a discursive construction.	11th International ISKO Conference	H	+
WoS5	2010	Shirley Lim & Chern kliew	GLAM metadata interoperability.	International Conference on Asian Digital Libraries	L	Pago
WoS6	2011	Jan Askhoj, Shigeo Sugimoto & Mitsuharu Nagamori	A Metadata Framework for Cloud-Based Digital Archives Using METS with PREMIS.	13th International Conference on Asia-Pacific Digital Libraries	H	Pago
WoS7	2011	Thierry Bouche, Claude Goutorbe, Jean-Paul Jorda & Michael Jost	The EuDML Metadata Schema	Conference on Towards a Digital Mathematics Library	H	+
WoS8	2011	Riitta Autere & Mikael Vakkari	Towards Cross-Organizational Interoperability: The LIDO XML Schema as a National Level Integration Tool for the National Digital Library of Finland.	15th International Conference on Theory and Practice of Digital Libraries	M	Pago



Item	Ano	Autor(es)	Título	Conferência	Prio.	Disponi.
WoS9	2011	Steffen Hennicke	Leveraging EAD in a Semantic Web Environment to Enhance the Discovery Experience for the User in Digital Archives.	15th International Conference on Theory and Practice of Digital Libraries	M	+
WoS10	2011	Ya-Ning Chen, Chun-Ya Wen, Hui-Pin Chen, Yen-Hung Lin & Hon-Chung Sum	Metrics for Metadata Quality Assurance and Their Implications for Digital Libraries.	13th International Conference on Asia-Pacific Digital Libraries	H	Pago
WoS11	2012	Natália Tognoli & José Augusto Guimarães	Challenges of Knowledge Representation in Contemporary Archival Science.	12th International ISKO Conference	H	+
WoS12	2012	Yibeltal Tafere Bayih	Application of Preservation Metadata for Long-Term Accessibility of Digital Objects.	New Trends in Qualitative and Quantitative Methods in Libraries	M	Pago
WoS13	2012	Courtney Mumma	Building On OAIS: Processing Born-Digital Records Using Archivematica.	9th Imaging Science & Technology Archiving Conference		Pago
WoS14	2013	Ji Ha Lee & Andrew Perti	User-centered Approach in Creating a Metadata Schema for Video Games and Interactive Media.	13th ACM/IEEE-CS Joint Conference on Digital Libraries	M	Pago
WoS15	2013	Ciro Gonano, Francesca Tomasi, Francesca Mambelli, Fabio Vitali & Silvio Peroni	Zeri e LODE. Extracting the Zeri photo archive to linked open data: Formalizing the conceptual model.	14th IEEE/ACM Joint Conference on Digital Libraries	H	+
WoS16	2014	Mikko Lampi & Olli Alm	Flexible data model for linked objects in digital archives.	Imaging Science & Technology Archiving Conference	H	Pago
WoS17	2014	Anders Sode-Pedersen	Crowdsourcing: An integrated part of archival description.	Imaging Science & Technology Archiving Conference	L	Pago

Item	Ano	Autor(es)	Título	Conferência	Prio.	Disponi.
WoS18	2014	Angela Di Iorio, Marco Schaerf, Maria Guercio, Silvia Ortolani & Matteo Bertazzo	A Digital Infrastructure for Trustworthiness: The Sapienza Digital Library Experience.	9th Italian Research Conference on Digital Libraries	M	Pago
WoS19	2014	Dydimus Zengenene, Vittore Casarosa & Carlo Meghini	Towards a Methodology for Publishing Library Linked Data.	9th Italian Research Conference on Digital Libraries	M	Pago
WoS20	2015	Anna Jordanous	Enhancing Information Retrieval and Resource Discovery from Data using the Semantic Web	4th International Symposium on Emerging Trends and Technologies in Libraries and Information Services	M	+
WoS21	2015	Adam Jansen	Using Archival Diplomatics and Object-Oriented Programming to Build Preservation Objects in the Cloud.	Digital Heritage International Congress	M	+
WoS22	2015	Joseph Tennis	Archival Metadata for Digital Cultural Heritage Conceptual Provenance, Contextual Forensics, and the Authority of the Found Digital Object.	Digital Heritage International Congress	H	+
WoS23	2017	Lina Bountouri, Patrick Gratz & Fulgencio Sanmartin	Digital Preservation: How to Be Trustworthy.	Digital Cultural Heritage	H	+
WoS24	2018	Lucy McKenna, Christophe Debruyne & Declan O'Sullivan	Understanding the Position of Information Professionals with regards to Linked Data: A Survey of Libraries, Archives and Museums.	18th ACM/IEEE-CS Joint Conference on Digital Libraries	H	+

FONTE: elaboração própria.

Uma pesquisa semelhante, mas desta vez usando as palavras-chave em português definidas anteriormente, foi realizada no dia 18 de novembro de 2019. Porém, não foi encontrado qualquer

título usando os critérios da língua portuguesa. De modo a averiguar se se tratava de um problema das palavras-chave ou uma limitação da base de dados de referência, realizou-se uma pesquisa experimental com o objetivo de determinar se a WoS indexa algum documento em português.

Document type: Proceedings Paper

Year range: 1917–2019

Language: Portuguese

TS= a

Estes critérios pretendem recuperar comunicações científicas publicadas entre 1917 e 2019 (os extremos cronológicos da WoS à data da pesquisa), em língua portuguesa, que contenham a letra ‘a’ em qualquer parte do texto. Constatou-se que, embora os autores sejam de língua portuguesa (portugueses e brasileiros), com conferências realizadas em países lusófonos, os títulos das comunicações aparecem sempre em inglês. Acredita-se que, por este motivo, os termos de pesquisa em português não recuperam nenhum resultado. De acordo com este raciocínio, as comunicações relevantes foram recuperadas aquando da pesquisa com os termos em inglês e em espanhol.

### 3.5.5 Seleção de teses e dissertações

A seleção teses e das dissertações teve por base os critérios de pesquisa definidos anteriormente para as duas bases de dados escolhidas, a OATD e o RCAAP. Para cada tese ou dissertação indica-se: (1) um identificador único; (2) o ano da realização; (3) o autor (4) o tipo de obra, se mestrado ou doutoramento; (5) o título; e (6) a prioridade. Dada a extensão e especificidade destes estudos, atribuiu-se uma classificação que espelha a sua prioridade de consulta, de acordo com três graus: *high* (H), *medium* (M) e *low* (L). O objetivo da prioridade é direcionar a leitura num primeiro momento para as teses de doutoramento (PhD), dissertações de mestrado (MD) e dissertações de licenciatura (BA) consideradas mais pertinentes para a revisão de literatura na medida em que abordem, independentemente do ângulo, pelo menos um dos seguintes temas: a representação da informação, a descrição arquivística, os metadados, e a autenticidade. A classificação tem por base a consulta do título e do resumo. A escolha das obras assentou num critério qualitativo através do qual se analisou o título e o resumo e de modo a determinar se o tema do artigo seria potencialmente relevante para responder às questões de investigação.

#### 3.5.5.1 Open Access Theses and Dissertations

Os seguintes dados foram recolhidos a 6 de novembro de 2019 utilizando os critérios de pesquisa definidos anteriormente. De um universo de 42 teses e dissertações concluídas 1 de janeiro 2009 e 31 de dezembro de 2019, em inglês, português e espanhol, foram considerados pertinentes os 7 títulos constados na Tabela 39, num total de 17%.

Tabela 39 — Teses e dissertações recuperadas em língua inglesa, portuguesa e espanhola no OATD.

Item	Ano	Tipo	Autor(es)	Título	Prio.
OATD-1	2014	MD	Sean Petiya	Building a Semantic Web of Comics: Publishing Linked Data in HTML/RDFa Using a Comic Book Ontology and Metadata Application Profiles.	H
OATD-2	2015	PhD	Tiffany Chao	Methods metadata: curating scientific research data for reuse.	M
OATD-3	2016	PhD	Rhiannon Bettivia	Encoding power: the scripting of archival structures in digital spaces using the Open Archival Information System (OAIS) Reference Model.	M
OATD-4	2017	MD	Alex Kinnaman	Digital preservation auditing metrics as design tools for digital repositories.	M
OATD-5	2014	PhD	Newton de Mello	Os documentos fiscais eletrônicos no Brasil e na América Latina e seus impactos na automação de processos e no intercâmbio eletrônico de informações na cadeia logística.	L
OATD-6	2012	BA	Juan Zolotoochin	Implementación de una red semántica de archivos.	L
OATD-7	2013	BA	Marcelo Chiarandía	Un enfoque colaborativo para generación de meta-información de objetos de aprendizaje.	L

FONTE: elaboração própria.

### 3.5.5.2 Repositório Científico de Acesso Aberto em Portugal

Os seguintes dados foram recolhidos a 6 de novembro de 2019 utilizando os critérios de pesquisa definidos anteriormente. De um universo de 93 teses e dissertações concluídas 1 de janeiro 2009 e 31 de dezembro de 2019, em inglês, português e espanhol, foram considerados pertinentes os 13 títulos constados na Tabela 40, num total de 14%.

A este respeito, importa referir que a pesquisa em inglês recuperou 9 títulos, dos quais nenhum foi considerado pertinente (0%). No caso do espanhol foram identificados 18 títulos, embora nenhum tenha sido selecionado (0%). Assim, os 13 títulos selecionados provêm todos de uma pesquisa em português que identificou um total de 93 teses e dissertações (20% de pertinência em língua portuguesa).

Tabela 40 — Teses e dissertações recuperadas em língua inglesa, portuguesa e espanhola no RCAAP.

<b>Item</b>	<b>Ano</b>	<b>Tipo</b>	<b>Autor(es)</b>	<b>Título</b>	<b>Prio.</b>
RCAAP-1	2011	MD	Fernando Gama	As contribuições das linguagens de marcação para a gestão da informação arquivística digital.	M
RCAAP-2	2011	MD	António Botão	Recuperação da informação digital: a Norma Brasileira de Descrição Arquivística (NOBRADE) na descrição de material imagético.	M
RCAAP-3	2011	MD	Sandra Rodrigues	A web 2.0 nos arquivos portugueses.	M
RCAAP-4	2011	MD	Ana Castro de Amorim	Arquivos e informação: os caminhos da Gestão de Documentos na Universidade Federal da Paraíba.	M
RCAAP-5	2012	PhD	Irisneide Silva	A organização e a representação do conhecimento no domínio da arquivística.	M
RCAAP-6	2013	MD	Andressa Aguiar	Representação da Informação Arquivística: Diálogos e conexões interdisciplinares.	M
RCAAP-7	2013	MD	Jean Xavier	O uso de Metadados no Arquivo da Marinha: Implementação do Software ICA-AToM.	M
RCAAP-8	2015	MD	Célia Dantas	Representação da informação arquivística: uma proposta para o Arquivo Histórico Waldemar Duarte.	L
RCAAP-9	2016	MD	Ivínia Kuroki	Demarcações conceituais dos princípios científicos da arquivologia e da ciência da informação.	L
RCAAP-10	2017	MD	Theresa Alves	Representação descritiva da informação arquivística: proposta de descrição de medalhas no Acervo José Simeão Leal.	L
RCAAP-11	2017	MD	Leolíbia Linden	O tratamento temático da informação em instrumentos normativos de descrição arquivística.	H
RCAAP-12	2017	PhD	Luiane Vital	Representação temática de documento arquivístico: as contribuições do modelo conceitual FRASAD.	L
RCAAP-13	2018	MD	Humberto Moraes	Records in Contexts - a conceptual model for archival description (RiC-CM): análise da proposta de um padrão	H

Item	Ano	Tipo	Autor(es)	Título	Prio.
				internacional integrado de descrição arquivística.	

FONTE: elaboração própria.

### 3.5.6 Seleção de normas de metadados

De entre o vasto universo de normas, pretende-se aprofundar o estudo das normas de metadados de estrutura, ou da dimensão da estrutura de normas que abordem outros aspetos, pois é esta a componente que mais se enquadra nos objetivos da investigação. Escolhe-se um conjunto de normas uma vez que nenhuma norma, por si só, se adequa à descrição de todos os tipos de coleções ou de recursos de todas as comunidades (Apostolou, 2009, p. 38; Gilliland, 2016, p. 5). Ao invés, a diversidade de metadados potencialmente necessários encontra-se mais bem representada se consideradas várias normas de diferentes comunidades, tanto a nível da área científica, como de diferentes línguas e de diferentes regiões do globo.

Neste sentido, constroem-se os seguintes critérios que orientam a seleção das normas a analisar:

- Norma de estrutura de dados;
- Enquadramento de atividade nos setores informacionais (arquivos, bibliotecas, museus);
- Linguagem de redação em inglês, português ou espanhol;
- Disponibilização de forma gratuita na Internet (acesso aberto).

A dimensão da estrutura já foi discutida. A opção por incluir normas de outros setores de atividade além do contexto estritamente arquivístico, como bibliotecas e museus, justifica-se por uma intenção de potencial enriquecimento dos metadados encontrados com os contributos de outras áreas, que podem revelar abordagens pertinentes para a asserção da autenticidade e demonstração da confiabilidade que estejam a ser ignorados ou desconsiderados nos arquivos. Além disso, também se pretende promover tanto quanto possível uma maior consistência estrutural com as práticas descritivas de outros domínios. A restrição linguística deve-se ao facto de serem estas as línguas nas quais se tem competência. Além disso, apenas se consideram normas gratuitamente disponíveis de forma a garantir o acesso às mesmas, e à sua viabilidade enquanto fontes de informação sem barreiras monetárias.

Tabela 41 — Normas de metadados a analisar.

<b>Título</b>	<b>Ano</b>	<b>Foco</b>	<b>Contexto</b>
<i>General Standard for Archival Description</i> (ISAD-[G])	2011	Documentos de arquivo	Internacional
<i>Australian Recordkeeping Metadata Schema</i> (AGRkMS)	2015	Documentos de arquivo	Austrália
<i>Encoded Archival Description (EAD)</i>	2015	Instrumentos de pesquisa	Internacional
<i>Esquema de Metadados para la Gestión del Documento Electrónico (e-EMGDE)</i>	2016	Documentos de arquivo	Espanha
<i>Describing Archives: A Content Standard</i> (DACS)	2019	Documentos de arquivo	Internacional
<i>Dublin Core Metadata Element Set</i> (DCMES)	2012	Vários propósitos	Internacional
<i>Visual Resources Association Core</i> (VRA Core)	2007	Recursos visuais e imagens	Internacional
<i>Metadata Object Description Schema</i> (MODS)	2009	Material bibliográfico	Internacional
<i>Categories for the Description of Works of Art</i> (CDWA)	2019	Obras de arte e outro material cultural em museus	Internacional
<i>Records in Contexts</i> (RiC)	2019	Documentos de arquivo	Internacional

FONTE: elaboração própria.

Neste sentido, escolheram-se 10 normas de metadados, um número considerado adequado para os objetivos propostos (

Tabela 41). Destaca-se também uma preocupação em incluir na amostra normas numa língua que não só o inglês, como representação de comunidades linguísticas distintas, como também de

continentes distintos. Assim, esta seleção inclui, além das normas de âmbito internacional, uma norma australiana (AGRkMS). Ponderou-se a inclusão da norma canadiana *Rules for Archival Description* (RAD), opção abandonada por a estrutura dos elementos proposta seguir *ipsis verbis* a ISAD(G). Por este motivo não foram também incluídas as normas em língua portuguesa *Orientações para a Descrição Arquivística* (ODA), aplicada em Portugal, e a *Norma Brasileira de Descrição Arquivística* (NOBRADE). Esta última poderia também ter atuado como representativa do continente sul-americano. Incluem-se as normas *Visual Resources Association Core* (VRA Core) por se focar na descrição de obras gráficas da cultura e de imagens associadas; a *Metadata Object Description Schema* (MODS) por versar sobre material bibliográfico; e a *Categories for the Description of Works of Art* (CDWA) por se especializar na representação de obras de arte num contexto museológico. Inclui-se ainda a RiC, embora ainda não exista uma versão definitiva à data de início da análise, em junho de 2020, estando apenas ainda disponíveis as duas versões *draft* v0.1 de 2016 e v0.2 de dezembro de 2019<sup>79</sup>. Pese embora o seu carácter provisório, considera-se que a importância que a RiC assume no contexto da descrição normativa em arquivos justifica a sua inclusão. Além disso, inclui-se a ISAD(G) como a ainda norma internacional completa para a definição dos elementos de metadados. As restantes normas enquadram-se no contexto da descrição arquivística, constituindo a porção mais significativa da amostra.

Esta lista será o ponto de partida para o início da abordagem quantitativa, onde se aplicará uma engenharia de requisitos para a determinação dos metadados essenciais para a determinação da autenticidade nas descrições.

### 3.5.7 Estatísticas do *corpus* documental

Nesta secção apresenta-se de forma esquemática o balanço final das estatísticas relativas à análise das fontes de informação, salientando o número e a percentagem de registos encontrados e selecionados, de acordo com cada uma da tipologia documental (artigos científicos, monografias, comunicações científicas e teses e dissertações académicas. A Tabela 42 sintetiza as unidades de enumeração relativamente aos artigos científicos.

Tabela 42 — Enumeração estatística do universo de artigo científicos.

Revista	Recuperados	Selecionados	Percent.
American Archivist	204	17	8%
Archival Issues	58	2	3%

<sup>79</sup> A edição completa da v0.2 foi publicada a 22 de Julho de 2021, quando este trabalho já se encontrava em processo de submissão. Por este motivo, a presente análise incide sobre o *draft* da v0.2.



<b>Revista</b>	<b>Recuperados</b>	<b>Selecionados</b>	<b>Percent.</b>
Archival Science	207	18	9%
Archivaria	171	19	11%
Archives (London)	80	0	0%
Archives (Quebec)	98	4	4%
Archives & Manuscripts	228	33	14%
Archives & Records	88	17	19%
Cadernos BAD	201	16	8%
Journal of Archival Organization	141	37	26%
Provenance	23	3	13%
Records Management Journal	189	47	25%
<b>Total:</b>	<b>1688</b>	<b>213</b>	<b>13%</b>

FONTE: elaboração própria.

Por sua vez, a Tabela 43 sumariza os registos relativos a monografias.

Tabela 43 — Enumeração estatística do universo de monografias.

<b>Revista</b>	<b>Recuperados</b>	<b>Selecionados</b>	<b>Percent.</b>
American Archivist	151	11	7%
Archival Issues	125	16	13%
Archival Science	0	0	0%
Archivaria	87	14	16%
Archives (London)	259	3	1%
Archives (Quebec)	0	0	0%
Archives & Manuscripts	166	34	20%
Archives & Records	146	31	21%

<b>Revista</b>	<b>Recuperados</b>	<b>Selecionados</b>	<b>Percent.</b>
Cadernos BAD	0	0	0%
Journal of Archival Organization	66	6	9%
Provenance	36	6	17%
Records Management Journal	0	0	0%
<b>Total:</b>	<b>1036</b>	<b>121</b>	<b>12%</b>
	Após eliminação de duplicados	<b>55</b>	<b>5%</b>

FONTE: elaboração própria.

A amostragem das comunicações científicas é apresentada na seguinte Tabela 44.

Tabela 44 — Enumeração estatística do universo de comunicações científicas.

<b>Base de dados</b>	<b>Recuperados</b>	<b>Selecionados</b>	<b>Percent.</b>
<i>Web of Science</i>	157	24	15%

FONTE: elaboração própria.

Por fim, a enumeração estatística das dissertações e teses académicas encontra-se representada no Tabela 45.

Tabela 45 — Enumeração estatística do universo das dissertações e teses académicas.

<b>Base de dados</b>	<b>Recuperados</b>	<b>Selecionados</b>	<b>Percent.</b>
OATD	42	7	17%
RCAAP	93	13	14%
<b>Total:</b>	<b>135</b>	<b>20</b>	<b>15%</b>

FONTE: elaboração própria.

No somatório destas várias unidades de registo ou fontes documentais, o total é constituído por 3016 títulos recuperados mediante as pesquisas, dos quais 312 (10%) títulos únicos foram selecionados após aplicados os critérios quantitativos e qualitativos para filtragem. Considera-se que este universo é representativo da produção académica nesta área e nestas temáticas, assegurando a legitimidade do *corpus* documental como ponto de partida para a revisão de literatura.

### **3.6 Questões éticas de investigação**

Declara-se que este estudo decorre dentro do código de ética definido pela Universidade de Coimbra, não havendo conflito de interesses na investigação. Além disso, os dados recolhidos e apresentados foram verificados minuciosamente pelo que se assegura a sua precisão e rigor, não tendo sido omitidos dados de forma a acomodar a narrativa científica do estudo. Por fim, o trabalho é original, fruto do trabalho do doutorando e dos seus orientadores.

## 4 Identificação de requisitos

---

### 4.1 Introdução

Este capítulo tem como principal objetivo identificar os requisitos de metadados encontrados na revisão de literatura, tanto a respeito da literatura científica, como da técnica. Relativamente à primeira, os vários autores das obras científicas consultadas foram categorizados como parte de uma análise categorial da revisão da literatura (Tabela 17), de modo a constituir um ponto de partida para a perceção da relação entre os requisitos encontrados e o posicionamento temático dos autores. É de destacar que na literatura científica nem sempre se encontram propostas de elementos de metadados em específico, mas antes ideias ou recomendações. Como tal, os requisitos identificados nestas fontes de informação assumem um carácter mais genérico, aproximando-se da definição de princípios cuja implementação pode obedecer a variações na designação dos elementos de metadados especificamente definidos pelas instituições.

Por sua vez, a literatura técnica consiste numa seleção de normas de metadados de diferentes áreas, origem geográfica e âmbito, previamente discutidas na

Tabela 41. Em contraste com a literatura científica, as normas de metadados escolhidas são de normas de estrutura, pelo que apontam concretamente quais os elementos de metadados que devem ser escolhidos para descrever os recursos. Por este motivo, a interpretação dos dados é mais direta, cabendo aos investigadores o exercício de interpretação de quais os elementos de metadados mais diretamente relacionados com a presunção da autenticidade.

Desta forma, pretende-se que estas duas abordagens complementares — a da literatura científica e a técnica — possam ilustrar de forma abrangente um conjunto de requisitos para a definição de um modelo para a autenticidade das descrições arquivísticas, enquadradas num contexto de representação digital. É particularmente importante considerar um vasto *corpus* de fontes de informação, isto é, de considerar várias perspetivas, porque, como salientam Ramesh et al., *there is no unique metadata standard sufficient to describe all the documents emerging in various kinds of formats, helpful in efficient information retrieval* (Ramesh et al., 2015, p. 198). Esta ideia é igualmente ecoada por Anne Gilliland quando afirma que *there is no single metadata standard or set of standards that is adequate for describing all types of collections and materials* (Gilliland, 2016, p. 5). Assim, considerar diferentes normas de metadados de diferentes

comunidades afirma-se como uma abordagem mais sensata, não só para capturar a diversidade de perspectivas sobre a questão da autenticidade e da confiabilidade dos recursos, mas também porque, ao analisar, comparar e discutir os metadados que diferentes comunidades escolhem para representar os recursos, está-se a criar uma ponte comum de entendimento que pode facilitar a interoperabilidade.

É ainda importante lembrar, conforme discutido na revisão de literatura, que a autenticidade não se trata de um conceito imutável. Por um lado, este conceito depende em larga medida do contexto de interpretação, na medida em que a presunção de autenticidade é uma decisão de um sujeito observador que decide se considera determinado recurso ou descrição como autêntica, dependendo da informação disponível. Neste sentido, afirma-se como uma decisão subjetiva. Por outro, a autenticidade é um conceito variável no tempo, na medida em que a demonstração da integridade dos recursos deve ser um processo permanente, ininterrompível. Consequentemente, o conjunto de metadados escolhidos por cada comunidade como apropriados depende em larga medida do contexto de aplicação. No âmbito desta investigação, uma vez que se pretende propor um modelo abrangente, procurar-se-á identificar e propor um conjunto de requisitos que seja geral o suficiente para ser universalmente válido, sem ser minucioso ao exagerado até um ponto em que, por tão extenso e detalhado, se tornaria impraticável.

Ainda assim, alguns projetos pautam-se por definir um conjunto mínimo, essencial ou fundamental de metadados. Não se considera que haja um adjetivo apropriado, uma vez que não é possível obter um consenso generalizado sobre o que constitui essencial ou fundamental. De entre estes três termos, prefere-se o adjetivo ‘mínimo’ porque representa com mais exatidão a noção de um conjunto reduzido tanto quanto possível, mas sem a conotação de presunção que pauta os outros adjetivos. De entre esses projetos destaca-se o Dublin Core<sup>80</sup>, que visa a representação de recursos na *Web*, o InterPARES, focado na preservação dos documentos de arquivo<sup>81</sup>, e o TRIPLE<sup>82</sup>, do domínio da representação de recursos no âmbito da investigação nas ciências sociais e humanas. Estes modelos de metadados definem um conjunto mínimo de metadados para a representação porque constituem modelos desenvolvidos para serem aplicados num contexto específico e explicitado. Neste caso, uma vez que o seu enquadramento está bem definido, e pretendem servir uma comunidade específica, assumem autoridade para propor um conjunto mínimo de elementos de metadados a adotar.

Porém, o mesmo raciocínio não pode ser aplicado aos modelos conceptuais. Nestes casos, como é o caso presente neste trabalho, os modelos são desenvolvidos de forma abstrata, independentemente do contexto de aplicação. Este aspeto é particularmente sensível no caso da

---

<sup>80</sup> A lista de termos proposta pela Dublin Core Metadata Initiative encontra-se em <<https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>>.

<sup>81</sup> A tentativa de definir metadados essenciais é explicitada na quarta questão de investigação (InterPARES 2, 2008a, p. 191).

<sup>82</sup> A estrutura de dados considerada mínima é designada por *TRIPLE data model* (TRIPLE, 2020, pp. 25-27).

autenticidade uma vez que, conforme discutido previamente, se trata de um conceito subordinado à interpretação subjetiva do sujeito que determina, perante o conjunto de informações disponíveis, se estão reunidas as condições para considerar determinado recurso ou serviço autêntico e fidedigno, e assim depositar a sua confiança. Como consequência, um modelo de metadados para a autenticidade não pode ter presunções de definir um conjunto “mínimo” de elementos de metadados, uma vez que a noção do que constitui mínimo varia de contexto para contexto, de indivíduo para indivíduo. Similarmente, também não pode incorrer na falácia de abarcar um conjunto “completo” ou universal de metadados, porque esta noção também depende da subjetividade. Perante este cenário, parece apropriado que a ambição do modelo aqui proposta seja a de incluir um conjunto de elementos de metadados que se acredita que, quando incluídos, constituem de um modo geral — e reforça-se o termo geral — uma forte base de presunção da autenticidade dos recursos e, conseqüentemente, da confiabilidade da sua informação.

O modelo de dados proposto, além de completo, deve ser simples o suficiente para fomentar a interoperabilidade, uma vez que é necessário encontrar um equilíbrio entre simplicidade, interoperabilidade e riqueza semântica (Ramesh et al., 2015, p. 197). Neste processo, para a proposta de um modelo, é crítico definir o objetivo e o nível de granularidade a incluir relativamente a cada tipo de metadados, isto é, o grau de detalhe a que se deve aprofundar cada elemento de metadados estrutura do recurso. Conforme Gilliland avisa, *there will likely always be important tradeoffs between the costs of developing and managing metadata that can be capitalized on for the future, often unanticipated uses* (Gilliland, 2016, p. 20). Por outras palavras, num mundo ideal, procurar-se-ia incluir o máximo de metadados possível, de modo a capturar todos os aspetos expectáveis da representação do recurso, de forma a tentar satisfazer qualquer tipo de uso, previsto e imprevisto. Na realidade, os custos de tal abordagem, tanto económicos como de tempo, seriam incomportáveis. Quantos mais metadados incluídos como obrigatórios, mais completa e confiável é a descrição. Porém, também mais custosa se torna para as instituições que pretendem implementar o modelo. É necessário um meio termo, um ponto de equilíbrio em que é suficiente completo mas suficientemente simples de descrever, de modo que não seja demasiado caro nem moroso de implementar.

Assim, no âmbito desta investigação, o modelo a propor tentará, pelo menos à partida, ser o mais completo possível, uma vez que se foca num subtipo de metadados, os metadados considerados pertinentes para a presunção da autenticidade e da confiabilidade. Quanto a este conjunto, procurar-se-á explorar a literatura científica e técnica sobre os requisitos emergentes, sem expectativas iniciais, pelo que o grau de granularidade a definir dependerá dos dados empíricos, que se analisam de seguida.

## 4.2 Requisitos da literatura científica

A literatura científica foi subdividida em categorias emergentes que se analisam ao longo das próximas páginas. Para cada categoria, exploram-se os autores e procura-se identificar no texto ideias que possam contribuir para a definição dos requisitos de autenticidade dos recursos. Uma

análise sistematizadora e mais crítica terá lugar no quinto capítulo, quando os dados recolhidos neste capítulo forem considerados em conjunto como base teórica para a proposta do modelo.

#### **4.2.1 Representação da informação**

Antes da análise dos trabalhos ocorrentes neste tema, considera-se importante incluir uma nota terminológica. Entende-se que a representação da informação pode ser entendida como tendo dois sentidos, um abrangente e um restrito. No sentido restrito, identifica-se com a indexação e com a classificação. No sentido abrangente, inclui a descrição arquivística. Ainda assim, opta-se pela separação dos dois temas de modo a poder incluir na categoria da representação da informação autores que não sejam da área arquivística.

Na categoria da representação da informação, os artigos têm frequentemente um foco demarcado na questão de mediação da informação entre o utilizador dos arquivos e os arquivistas que organizam e descrevem as coleções. Entre algumas das questões consideradas nestes estudos incluem-se a facilidade de uso dos instrumentos de pesquisa, relação entre a competência do utilizador e o modo como realizam as pesquisas, níveis de controlo da informação, ou as implicações do armazenamento ilimitado para a representação dos recursos.

A relação entre os utilizadores e o material de arquivo, através de uma mediação por um arquivista, foi analisada por Duff et al. (2013). Através da realização de entrevistas e de questionários a arquivistas e a utilizadores dos arquivos, este artigo salienta:

- i. a ideia de que é desejável que os utilizadores sejam ensinados a usar os instrumentos de pesquisa de modo a compreenderem os sistemas de acesso, dominando conhecimento relativo à recuperação da informação, como conhecimento da base de dados, técnicas de pesquisa e estrutura do instrumento de pesquisa. Relativamente a este ponto, foi identificado que o arquivista deve ser capaz de explicar as regras e os procedimentos arquivísticos de descrição dos recursos.
- ii. a necessidade de introduzir conhecimento contextual, expresso através de informação sobre o criador dos documentos de arquivo. Informação sobre a proveniência e a história dos documentos de arquivo emergem como essenciais para a interpretação dos documentos. Este aspeto foi também apontado como requisito fundamental para presumir a autenticidade dos documentos de arquivo.
- iii. a importância de os documentos de arquivo estabelecerem relações não só com outros documentos da coleção, mas também com documentos existentes em outros repositórios, de forma a enriquecer as necessidades informacionais dos utilizadores.
- iv. a ideia de que é essencial compreender as necessidades dos utilizadores e de estes desenvolverem boas relações pessoais com os arquivistas de modo a serem mais ajudados.

Deve ser realçado que este estudo assenta sobretudo num entendimento da relação presencial e física entre os utilizadores e um arquivista, que se senta com os utilizadores numa conversa e

apoio pessoal. No contexto da representação de conteúdos na *Web Semântica*, o arquivista pessoal desaparece e é substituído pelo sistema de informação, que passa a desempenhar as funções apontadas ao arquivista em termos de mediação. Por exemplo, a ideia de cair nas graças dos arquivistas pode ser extrapolável para o contexto digital, como no desenvolvimento de interfaces *user-friendly*.

Por sua vez, discussões sobre o nível de controlo das representações da informação são promovidas por Jinfang Niu (2015a). Neste artigo, a autora defende as seguintes ideias:

- i. o controlo intelectual em arquivos, isto é, a descrição arquivística, deve realizar-se ao nível do item em adição ao controlo coletivo das coleções. Adicionalmente, o controlo intelectual deve estender-se ainda a um nível inferior, ao nível do conteúdo, que pode incluir descrições de variáveis nos *datasets* e anotação de conteúdo do texto. O *content control* destaca-se pela função de auxiliar os investigadores a localizar informação relevante nos conteúdos dos recursos. Esta ideia expande a prática mais tradicional, espelhada pelos instrumentos de pesquisa analógicos, segundo a qual a descrição multinível das coleções arquivísticas atuava como a principal forma de controlo coletivo das descrições arquivísticas.
- ii. as coleções digitais suportam o desenvolvimento de estruturas polihierárquicas que permitem que um item pertença a mais do que uma coleção, e que uma coleção pertença a múltiplas comunidades. Esta ideia baseia-se na capacidade em explicitar relações que fornecem informações sobre a semântica, a proveniência e os contextos administrativos dos recursos.
- iii. a necessidade de produção de metadados em formato *linked data* e de tornar as descrições arquivísticas aptas para processamento *web* e para integração com outros recursos expressos em *linked data*.
- iv. o uso de instrumentos de pesquisa em estilo *website* como forma de auxiliar a navegação por parte dos utilizadores, e de usar termos genéricos em vez de termos técnicos de forma a facilitar a compreensão da informação. Um exemplo é a substituição de “âmbito e conteúdo” por “sobre a coleção”.
- v. a defesa do uso de registos de metadados independentes mas interligados de forma a possibilitar uma pesquisa com base em vários nós da hierarquia. Quando vários instrumentos de pesquisa monolíticos, como os gerados pela EAD, são reunidos por um agregador, as pesquisas nesses agregadores apenas permitem uma pesquisa em cada um dos instrumentos de pesquisa, sem os considerar como um conjunto. Os autores sugerem que gerar *metadata records* para cada nível da hierarquia (grupo, série, ficheiro) é essencial para navegar pelos níveis individuais das coleções. Esta abordagem teria também a vantagem de explicitar a representação das relações hierárquicas entre os componentes das coleções, capturadas pelos metadados de relação como ‘contém’ ou ‘deriva de’. Trata-se



de explicitar através de metadados de relação a estrutura que normalmente se encontra implícita.

- vi. os metadados devem ser capturados e acumulados ao longo do ciclo de vida dos documentos de arquivo. Os metadados também podem ser extraídos a partir de outras instituições ou dos próprios criadores, o que pressupõe a interoperabilidade entre as estruturas de metadados.

Uma linha discursiva não muito distante da de Niu, no que diz respeito a diferenças nos níveis de controlo, é apontada por Lawrence Serewicz (2010). Neste artigo, o autor debate as implicações de um maior volume documental derivado de evoluções na capacidade de armazenamento, destacando que:

- i. na *Web Semântica*, também referida como *Web 2.0*, a abordagem da representação deve ser *bottom-up*, centrada nas necessidades das comunidades e dos utilizadores, cabendo a estes definir as suas estruturas de pesquisa individualizadas.
- ii. é imperativo que os instrumentos de pesquisa estejam capazes de estabelecer *links* para outros recursos, evocando implicitamente a ideia de ter os metadados em *Linked data*, num formato partilhável e interoperável.
- iii. existe pressão social e crescentes expectativas para maior abertura e transparência de processos de representação, flexibilizando as estruturas de metadados. Serewicz aponta para a perda de importância das estruturas de representação definidas pelas instituições, em favor das pesquisas pessoais de cada utilizador. O autor refere-se a estes procedimentos como sistemas baseados em semântica para a captura do significado (*semantic-based systems to capture meaning*), o que permitirá facilitar o acesso, pesquisa e exploração dos documentos de arquivo, face às tradicionais estruturas de representação criadas pelas organizações.

Na sua tese de doutoramento, Irisneide Silva (2012) revisita algumas das transformações recentes no entendimento teórico da arquivística no que diz respeito à organização da informação. A autora aponta para a necessidade de reavaliação pela arquivística dos conceitos de proveniência, ordem original e de documento, de acordo com as transformações operadas pelo contexto digital. Também nota que as organizações tendem a não se encontrar estruturadas de acordo com um modelo hierárquico de organização, pelo que os sistemas de representação se devem tornar eles próprios mais flexíveis e transversais. Além disso, introduz uma “nova” função da identificação, que assinala ter sido formulada pela comunidade arquivística espanhola, referente à identificação do fundo, do sujeito produtor e da documentação produzida e acumulada. Esta autora pretende que atue como uma sistematização do conhecimento sobre uma instituição que produziu os documentos, enquanto parte essencial do processo de classificação e de descrição dos documentos de arquivo.

Numa abordagem mais próxima do domínio das tecnologias, Verborgh et al. (2015) exploram as implicações da arquitetura REST (*Representational State Transfer*) para a representação do recursos na *Web*. Na verdade, mais do que uma tecnologia, a REST é um estilo para arquiteturas de sistema que ditam constrangimentos na arquitetura. Um dos exemplos mais notáveis é a *World Wide Web*, cujo protocolo HTTP é governado por princípios REST. Neste contexto, os autores assinalam que, no contexto *Web*, cada recurso deve ter o seu próprio URL, uma vez que a unidade de informação essencial das arquiteturas REST é o recurso, uma unidade conceptual que deve ter um identificador único. Representados desta forma, os recursos permitem estabelecer mapas conceptuais a entidades relacionadas, e podem ser feitas modificações às entidades ou aos recursos sem que o significado das relações não se perca. Os autores também destacam que, nas arquiteturas REST, os recursos não são transferidos, mas antes é transferida uma representação do recurso entre o cliente e o servidor. Um recurso pode ser representado em HTML e em JSON para consumo por software, o que realça a necessidade de ter os metadados num formato *machine-readable* de forma a que possam ser processados automaticamente e a fomentar a interoperabilidade e a descoberta dos recursos.

Outra perspetiva mais teórica, mas igualmente pertinente, é seguida por Yaker (2003) num artigo sobre representação da informação em arquivos. Esta autora explora o próprio conceito de representação, e as implicações que as escolhas de representação têm no modo como os utilizadores acedem aos recursos. Assim, Yaker:

- i. sugere que a representação arquivística é uma prática fluída e socialmente construída. Esta ideia implica, por um lado, que determinada classificação se afirma como uma de várias representações possíveis. Por outro, pressupõe a noção de que as descrições arquivísticas têm uma carga subjetiva e parcial inerente, razão pela qual devem ser questionadas. A ordem original é um dos conceitos sob exame, que Yaker aponta que muitas vezes se trata mais de uma ordem intelectual arquivística idealizada, do que uma ordem bem-definida que provavelmente nunca existiu no contexto de origem.
- ii. assinala que a função da representação arquivística é permanente, pelo que o significado dessas representações varia no tempo à medida que os documentos são submetidos a novos usos. Por este motivo, as descrições devem ser revisitadas pelos arquivistas, e consideradas como um processo fluído e contínuo. Aponta ainda que a descrição arquivística principia na criação dos documentos e se desenvolve à medida que vão sendo acrescentados metadados como registo de versões ou de acesso.
- iii. relembra que, como as descrições arquivísticas constituem interpretações, e como as interpretações se sucedem no tempo à medida que novas formas de descrever e classificar os recursos são criadas, é útil preservar as descrições arquivísticas usadas no passado como forma de testemunho histórico. Neste caso, permite-se a coexistência de múltiplas narrativas de representação dos recursos, cada uma contando a sua história, como reflexo da evolução do significado dos sistemas de representação escolhidos.

- iv. alerta para a ideia de que os instrumentos de pesquisa devem ser vistos não como um documento rígido em si, mas como um agregado de dados que podem ser reutilizáveis e reconfiguráveis. Enquanto fontes de informação sobre uma coleção, os instrumentos de pesquisa desempenham papéis de autoria e de autoridade, motivos pelos quais o autor dos próprios instrumentos de pesquisa deve ser identificado. Também sugere a identificação da indicação da versão do instrumento de pesquisa.
- v. salienta que cada vez é mais importante privilegiar a ordem intelectual do que a ordem física, tornando-se essencial fornecer informação contextual detalhada e abrangente de forma a esclarecer a proveniência e a estabelecer a autoridade da fonte. A descrição deve estar radicada em normas e boas práticas partilhadas.

Jane Zhang (2012b) também introduz importantes reflexões sobre a representação arquivística, em particular na era digital, que considera ser uma nova era de consumo de informação em arquivos porque tanto os utilizadores como os materiais se libertam de condicionantes impostas pela organização física. A autora assinala que a representação em arquivos deve capturar a proveniência (identificação de materiais decorrentes da mesma fonte), a ordem (estrutura interna de organização) e o conteúdo (acesso ao conteúdo informacional) dos documentos. A proveniência é apresentada principalmente pelo nome do criador dos documentos, e caracterizada na história administrativa e na história custodial, e é fundamental para preservar, perpetuar e autenticar o significado dos documentos ao longo do tempo, protegendo a integridade da informação. Porém, deve ser vista como um de múltiplos possíveis pontos de acesso à informação. A representação da ordem arquivística também se afirma como uma asserção da autenticidade pelos arquivistas, na medida em que explicita o significado dos documentos e testemunha os processos que os geraram. Finalmente, a autora destaca que a representação do conteúdo é crítica para a descoberta e para o acesso à informação no contexto digital, enquanto complemento à informação sobre a proveniência.

Como desenvolvimento desta ideia, Zhang (2012a) sugere que metadados relativos ao conteúdo de itens individuais devem ser facilmente exploráveis e acessíveis, sem necessidade estarem escondidos atrás de uma estrutura hierárquica que seria necessário consultar em primeiro lugar. Este modelo permite fornecer descrições mais específicas como ‘título’, ‘sujeito’, ‘descrição’, ‘autor’, ‘criador’, ‘recipiente’, ‘contribuidor’, ‘data’, ‘lugar’ e ‘tipo de material’. Por sua vez, este exercício promove a normalização, facilitando a interoperabilidade entre coleções a um nível mais granular. Assim, Zhang sugere que metadados ao nível dos itens devem ser incorporados nas normas, desde que a informação seja exata e autêntica.

Se esta abordagem for seguida, Geoffrey Yeo (2012) defende a possível coexistência de múltiplas agregações documentais, podendo estas corresponder a diferentes necessidades de conceptualizações dos próprios agregados, passíveis de serem formados por diferentes critérios a partir de diversas fontes. Essencialmente, o autor defende que os recursos podem e devem ser representados de diversas formas, cada uma resultante de uma narrativa, refletindo diversas ordens originais, diversas interpretações do contexto, e diversas necessidades dos utilizadores.

## 4.2.2 Descrição arquivística

Embora no título se opte pelo termo “descrição arquivística”, deve ser notado que este termo se enquadra na representação da informação, conforme discutido previamente. Porém, prefere-se este termo para obter uma maior homonímia e consistência terminológica com as normas de descrição arquivística.

Neste tema encontram-se textos que debatem a abordagem *more product, less process* (MPLP), que questionam a influência ativa e subjetiva dos arquivistas na descrição dos recursos e na recriação do contexto, que procuram revisitar as funções essenciais a descrição arquivística, ou que exploram a relação entre a descrição arquivística e a curadoria digital.

Rachel Anchor (2013) reflete sobre a descrição arquivística com base na metodologia MPLP, proposta em 2005 por Mark Green e Dennis Meissner e caracterizada por uma abordagem mínima à classificação e descrição dos recursos em virtude da explosão documental observada no século 20<sup>83</sup>. O objetivo: aumentar a produtividade dos arquivistas. A ideia essencial é a de que é melhor ter alguma descrição do que nenhuma, pelo que os processos de descrição devem ser vistos como flexíveis, tornando-se importante definir para cada coleção um conjunto de metadados mínimos que possam descrever a coleção de forma simples mas suficientemente completa, dentro dos constrangimentos orçamentais e temporais, e com a descrição centrada nas necessidades do utilizador. Neste caso, é vantajoso que haja normas a indicar os metadados usados para descrever as coleções de forma sistemática.

Em contraste com a abordagem minimalista do MPLP, Jennifer Douglas (2016) entende que se deve ir na direção contrária e incluir nos instrumentos de pesquisa ainda mais informação — ou talvez um tipo diferente de informação — de modo a tornar a descrição mais honesta. Eis algumas das principais ideias introduzida por esta autora:

- i. a descrição é mais honesta se reconhecer os diferentes processos transformativos que influenciam a constituição dos arquivos ao longo do tempo. Propõe que se admita um papel mais ativo ao criador primário dos arquivos, por exemplo através de citações ou de notas do próprio criador. Outra sugestão é a de facultar acesso aos investigadores de ficheiros de aquisição (*accession files*) e de processamento, uma vez que contêm informação sobre a proveniência e a história custodial, fornecendo assim maior contextualização de um fundo e das negociações a partir das quais assume forma.
- ii. para aumentar a transparência e *accountability*, devem ser acrescentadas aos instrumentos de pesquisa notas de rodapés que testemunhem decisões no processamento e que reconheçam o papel interventivo do arquivista. Devem também serem explicitadas as

---

<sup>83</sup> Green e Meissner (2005) propõem que se reduzam as hierarquias descritivas e que a descrição se concentre em níveis mais gerais do fundo e da série. A abordagem deve-se basear no *golden minimum*: respeito pelos fundos e pela ordem original e pela descrição como dependente da classificação.

- políticas e as práticas institucionais relativas à descrição e dedicar mais atenção à reconstituição da história de um fundo, e deve ser dado maior destaque à história custodial.
- iii. a sugestão, baseada em MacNeil, de incluir a história arquivística como um campo autónomo da descrição, subdivisível em três elementos: ‘nome(s) do(s) custodiante(s)’, ‘história custodial’ e ‘história da classificação’ dos documentos de arquivo e de instrumentos de pesquisa relacionados.
  - iv. a ideia de que a ordem original é um constructo arquivístico, uma inferência do arquivista com base num conhecimento limitado do contexto, pelo que sugere que a ordem original deveria ser antes denominada de ordem arquivística e de reconhecer o papel do arquivista em a determinar. Neste sentido, entende que devem ser incluídos metadados que explicitem informação sobre o processamento arquivístico, preferencialmente colocados nas zonas de estrutura ou de conteúdo.
  - v. a inclusão de uma discussão sobre as ordens físicas e arquivística. Recupera uma sugestão de Gabrielle Dean que propõe a inclusão de textos paralelos que abordem o ‘porquê’ e o ‘como’ das fontes, além do ‘como’ e ‘onde’.
  - vi. a proposta de criar oportunidades para os investigadores contribuírem com o seu conhecimento para os instrumentos de pesquisa através de comentários, *tags* e/ou anotações, operando uma mudança de um modelo de descrição autoritária para outro de autoridade e colaboração partilhada.

A subjetividade do arquivista é igualmente realçada por um estudo de MacNeil (2012) que examina a natureza socialmente construída e mediada da descrição arquivística, enquanto um género retórico. Neste sentido, os instrumentos de pesquisa podem assumir-se como veículos de ações sociais como tornar as coleções acessíveis, confirmar a coesão e autenticidade do conjunto documental, e de validar a autoridade dos arquivistas e das instituições arquivísticas enquanto entidades aptas a preservar e disponibilizar os recursos.

A inclusão de informação detalhada e completa sobre a proveniência e o contexto de criação, enquanto pilares da compreensão do significado dos documentos de arquivo, é também veiculada por Margaret Hedstrom (1993). A autora afirma que este tipo de informação é essencial para que os utilizadores sejam capazes de interpretar o conteúdo e de estabelecer a autenticidade dos documentos de arquivo. Neste sentido, propõe que na descrição arquivística se inclua:

- i. o nome do ficheiro, a localização no sistema de armazenamento e especificações técnicas de *hardware* e *software* necessárias para a recuperação da informação. Dado o antigo contexto tecnológico em que foi escrito este artigo, pode atualmente interpretar-se a localização física do ficheiro como um URL que permita a identificação única num contexto *Web*.
- ii. informação que documente a relação entre o contexto e o conteúdo dos documentos de arquivo, como requisito para perceber o significado destes.

- iii. informação suficiente sobre a gestão dos documentos ao longo do seu ciclo de vida de modo a determinar a natureza, o motivo e a autenticidade das transações. Como componente essencial da autenticidade, Hedstrom sustenta que devem ser cumpridos procedimentos de acesso e que devem existir mecanismos de verificação (*audit trails*) que permitam testar a integridade do sistema.
- iv. uma descrição mais detalhada sobre as características técnicas dos sistemas de informação.
- v. o entendimento de que uma única estrutura hierárquica não é capaz de capturar de forma suficiente a teia de relações entre a criação e o uso dos documentos e o seu conteúdo e contexto, uma vez que os documentos digitais são sucessivamente moldados por vários criadores e utilizadores ao longo do tempo. Por este motivo, sugere a descrição dos recursos independentemente dos contextos em que são criados e usados.

Por sua vez, Llanes-Padrón (2019) explora a relação íntima entre a descrição arquivística e a curadoria digital, que visa garantir a fidedignidade, a autenticidade e a disponibilidade de acesso dos recursos digitais. A autora parte de um pressuposto de que a representação deve ser entendida como um contexto pluridimensional, em contraste com as tradicionais abordagens unidimensionais centradas no documento. Destaca a necessidade de sistematizar a representação dos conteúdos e da sua codificação, a partir de normas e de modelos conceptuais, de modo a facilitar a interoperabilidade entre sistemas de informação, e subscreve alguns metadados sugeridos pela RiC-CM, nomeadamente ‘nota de autenticidade e integridade’, ‘qualidade da informação’, ‘formato de codificação’ e ‘material físico em que se representa a informação’. Esta autora também defende a documentação de vertentes da curadoria digital, nomeadamente criação/receção, acesso e uso em aplicações informáticas, seleção e ingestão ou captura em repositório como informações essenciais para assegurar a autenticidade, a confiabilidade, a integridade e a usabilidade dos recursos.

Uma reflexão mais profunda acerca da relação entre a descrição arquivística e a autenticidade é encetada por Heather MacNeil (2009). Uma das interessantes ideias discutidas neste artigo é a de que a autenticidade pode focar-se mais na própria descrição arquivística do que nos documentos a serem descritos, uma vez que determinada descrição pode ser considerada uma representação autêntica do modo como um arquivista num dado período entende a estrutura de um agregado documental. Esta sugestão levanta necessariamente a questão da responsabilização dos arquivistas no tipo de representações que constroem, e da influência que as condicionantes dessas representações têm na presunção da autenticidade por parte dos utilizadores. A autora deixa várias outras recomendações, nomeadamente:

- i. a necessidade de documentar mudanças físicas dos documentos, uma vez que as características físicas não são meramente uma condição que afeta o acesso e o uso, mas veiculam parte do significado dos documentos e podem condicionar a identidade.
- ii. a inclusão de uma declaração pública de autoria e responsabilidade pelo arquivista que elaborou a descrição, a partir da qual os utilizadores possam pressupor a autenticidade do conjunto documental, e enquanto um requisito fundamental para promoção da

transparência e da responsabilização (*accountability*). Um passo nesse sentido poderia ser através da inclusão de um colofão no final da descrição indicando as escolhas tomadas pelo arquivista no processo de avaliação, classificação e descrição dos documentos, bem como informação sobre o próprio arquivista, tal como nome e *curriculum vitae* (CV), enquanto testemunho da consciência do papel do arquivista na modelação da comunicação de uma coleção.

- iii. a criação de um registo detalhado da história dos documentos de arquivo, documentando tanto quanto possível a jornada dos documentos até chegarem à instituição arquivística. Como parte desta história deve ser incluída informação sobre a história de conservação, que testemunhe as mudanças físicas dos documentos (como condição atual ou intervenções ao longo do tempo) e os consequentes efeitos no sentido e na autenticidade destes.
- iv. com base na proposta de Laura Millar que sugere a inclusão de três componentes: ‘história do criador’ (a história de quem criou, acumulou ou usou os documentos), ‘história dos documentos de arquivo’ (a história da gestão física dos documentos), e ‘história custodial’ (a explicação da transferência da custódia de um criador ou custodiante para a instituição arquivística), MacNeil entende que deve existir uma ‘*Creator History Area*’ com o nome do criador e a história administrativa, a que se acrescenta uma ‘*Archival History Area*’ com o nome do custodiante, a história custodial e a história da classificação dos documentos de arquivo e dos respetivos instrumentos de pesquisa.

A ideia dos arquivistas como utilizadores ativos dos arquivos, na medida em que os usam para constituir evidência de ações passadas, é também realçada por Jennifer Meehan (2009). Deste artigo sobressai o entendimento de que o contexto, embora formado a partir da análise das fontes existentes, adquire um sentido como um todo que supera a soma das partes. Para o recriar, o arquivista deve consultar várias fontes no decurso da análise arquivística, como relatórios, CVs, gráficos ou obras de referência que lhe permitam inteirar-se de funções, atividades, datas, nomes ou lugares. O contexto torna-se assim um conhecimento inferível a partir das fontes existentes, embora não exista nelas diretamente. A autora defende a recolha dos seguintes dados para a reconstituição do contexto:

- i. quem criou os documentos, como e porquê? Qual é a função específica ou atividade com que os documentos se relacionam? Qual é o procedimento específico ou processo que lhes deu origem? Como é que os documentos foram mantidos pelo criador e/ou custodiante? Como é que foram usados e transmitidos no espaço e no tempo? Quais são os formatos dos documentos? Que funções é que esses formatos desempenharam no contexto da ação que os documentos desempenharam?
- ii. para os arquivos pessoais doados, deve ser incluído quem doou os documentos (‘nomes’ e ‘data’), como foram doados e em que momento (‘descrições do material’ e ‘datas’), quem são os criadores ou custodiantes (‘nomes’ e ‘datas’), se há conjuntos de documentos dentro da própria coleção (‘nomes’, ‘datas’, ‘funções’ e ‘atividades’).

- iii. no caso de documentos de uma instituição governativa, aponta a importância de recolher informação sobre a autoridade ou esfera de responsabilidades da instituição ('funções'), aspetos fundamentais da administração ('operações básicas' e 'procedimentos de gestão documental') e da história administrativa ('eventos', 'datas', 'partes envolvidas'). Esta informação serviria para estabelecer uma relação entre o(s) criador(es) e os documentos.
- iv. detalhes sobre o raciocínio implícito na criação de determinada descrição, nomeadamente acrescentando um colofão aos instrumentos de pesquisa, que explique o procedimento e de descrição e o seu autor, como forma de explicitar os processos arquivísticos.
- v. inclusão de notas de rodapé ou de citações nos campos descritivos de modo a abordar preocupações com a falta de informação das fontes existentes, e de realçar partes de interpretações dúbias ou decisões problemáticas.

Além disso, Julián Moyano Collado (2013) aponta a importância da descrição arquivística, enquanto uma das tarefas mais críticas dos arquivos, na medida em que o uso e a recuperação dos documentos depende da sua devida execução. O autor destaca a necessidade de os elementos que constituem os recursos serem descritos de forma independente entre si, de modo a adquirir uma estrutura flexível para representação na *web*, apontando o potencial do RDF.

Porém, Jinfang Niu (2013), partindo da perspectiva de que os metadados embarcam num processo evolutivo desde a criação à descrição arquivística, aborda um raramente discutido aspeto da extinção dos metadados durante a descrição arquivística. A autora defende que a informação sobre destruição dos documentos ajuda a averiguar a autenticidade, e que este tipo de metadados não deve ser de carácter temporário, mas deve ser incluído na descrição. Embora o principal foco do artigo seja uma discussão do tipo de metadados que podem ser herdados entre modelos de gestão documental e de descrição arquivística, propondo para isso um modelo conceptual com três entidades primárias (*agents*, *records* e *business*) e duas entidades adicionais (*mandates* e *relations*), Niu salienta que os *audit trails* são importantes para garantir a autenticidade e a integridade dos documentos de arquivo no contexto da gestão documental. Neste sentido, lança o repto para novos estudos explorarem se a inclusão de alguns destes *audit trails* enquanto elementos da descrição arquivística seria pertinente, nomeadamente nos elementos 'história administrativa', 'avaliação', 'destruição e eliminação planeadas'. De um modo geral, a autora alerta para a necessidade de considerar a avaliação (*appraisal*) de metadados não só antes da aquisição dos documentos, mas também após.

Por fim, Jane Zhang (2012c) tece algumas considerações a respeito do conceito da ordem original em arquivos digitais. Um dos pressupostos mencionados é o de que a associação da ordem original com a organização física deixa de ser válida para os documentos digitais. Zhang sugere que os arquivistas não devem descrever os conjuntos documentais num e num só contexto. Antes, emergem múltiplas relações lógicas entre os documentos que refletem diversos usos. Porém, observa que em certas organizações os documentos podem encontrar-se organizados de acordo com um sistema de ficheiros que reflete a sua organização nas pastas dos computadores. Nestes casos,



herdeiros da organização analógica, uma estrutura única em ficheiros e pastas oferece parca acessibilidade. Em contraste, em sistemas de gestão documental orientados para a criação de metadados, tendem a emergir diversas representações da ordem original, enquanto reflexo das complexas relações estabelecidas entre os documentos mantidos pelo sistema. No artigo destaca-se a ideia de que o conceito de ordem original se refere ao contexto nos quais os documentos são criados, estruturados, acedidos e utilizados. Porém, como os documentos não se encontram organizados em uma ordem específica, os arquivistas preservam o contexto e, dentro do contexto, os documentos são governados por relações lógicas expressas através dos metadados que permitem que estes sejam organizados, recuperados e interpretados.

Além disso, a autora destaca o potencial de metadados ao nível do item para promover o acesso direto à informação presente nos documentos de arquivo. Os próprios mecanismos de preservação digital devem passar a centrar-se ao nível do item, uma vez que os objetos digitais requerem um controlo a este nível, não ao nível do agregado documental. Finalmente, Zhang destaca ainda que este tipo de metadados pode ser incluído automaticamente nas descrições arquivísticas.

### 4.2.3 Gestão da informação

Nesta categoria analítica os textos tendem a focar mais especificamente os procedimentos de gestão e organização da informação em arquivos. Discutem-se o conceito de fundo, a abordagem australiana do *continuum* para a gestão documental, a relação entre os documentos de arquivo e o sistema de informação do qual emanam, o próprio conceito de documento de arquivo e os paradigmas arquivísticos.

Um dos principais argumentos do artigo de Jefferson Bailey (2013) é o de que, no ambiente digital, o conceito de fundo deixa de ser a forma exclusiva de representação e organização dos conjuntos documentais. Este autor defende a criação de descrições ao nível do item, automatizáveis quanto possível, como forma de acelerar e enriquecer a informação veiculada pelas descrições. Segundo este entendimento, a classificação deixa de ser um processo governado por hierarquias intelectuais. Em termos de representação arquivística, as implicações são as que aos metadados dos recursos podem ser dados novos usos e sentidos, aumentando a sua rede de relações contextuais. Trata-se de facultar ao utilizador ferramentas para construir os seus próprios padrões e hierarquias de descoberta, recriando contextos e significados dos recursos. É de realçar que o autor defende que o conceito de fundo ou a utilidade da ordem original não são abandonadas, mas antes deve ser revisitado e questionado o lugar privilegiado que têm ocupado na descrição arquivística.

Largamente inspirada pelas ideias de Sue McKemmish e Frank Upward, Kate Cumming (2010) subscreve a ideia de um *continuum* documental, uma perspetiva australiana que surgiu como integração da divisão norte-americana entre o *records management*, ou fase ativa e intermédia, e os *archives*, correspondentes à fase definitiva dos arquivos. A ideia do *continuum* é essencialmente um modo de entender os arquivos enquanto uma unidade coerente formada por diversas componentes, com o objetivo de capturar como os documentos são criados, mantidos e usados. A gestão da informação é entendida como um contínuo de atividades que pretendem garantir que o

significado, o contexto, o acesso e a capacidade de constituir evidência são capturados e mantidos ao longo do tempo. Porém, o contexto e o significado são entendidos como variáveis, dependendo da perspectiva dos utilizadores. O contexto é visto como chave, e admite-se a coexistência de múltiplas realidades. Uma das implicações desta abordagem é a de que a descrição dos documentos digitais deve ser feita desde a sua criação, não podendo ser retardada até à entrada sob a custódia arquivística. A autora defende que esta perspectiva, por sua vez, deve transvasar para as normas de metadados, através da inclusão de metadados que explicitem práticas de gestão documental, evidência (contexto jurídico, administrativo, histórico ou social), transações e identidade.

Num capítulo do vasto livro *Research in the Archival Multiverse*, Amelia Acker (2017) introduz reflexões sobre a necessidade de considerar a infraestrutura que preside à criação dos documentos digitais, e de procurar entender como estes percorrem a infraestruturas nos pontos de criação, transmissão e receção, pontos geralmente tidos como críticos para manter a autenticidade dos documentos de arquivo. A autora apela sobretudo à consideração da materialidade dos documentos de arquivo digitais como um requisito fundamental para se compreender o contexto, o conteúdo e a proveniência. Para a materialidade deve considerar-se não só a forma física, mas também os sistemas, as práticas e as instituições sociais baseadas nos artefactos. No fundo, trata-se de entender os documentos digitais enquanto objetos físicos com uma história material que é representada pelos sistemas de informação. Acker sugere que uma forma de abordar os processos envolvidos na transmissão dos documentos eletrónicos seria através da abordagem do *continuum* para a gestão da informação, de forma a permitir conhecer as relações entre os níveis físico, lógico e conceptual dos documentos. Deste entendimento parece resultar a noção de que se torna essencial perceber de que modo a infraestrutura (as estruturas de dados que governam a produção documental, e que resultam de decisões intencionais que criam e moldam constituição de evidência) condicionam o significado, a autenticidade e a integridade dos documentos de arquivo digitais, enquanto uma dimensão dos múltiplos contextos de criação.

Noutro capítulo do mesmo livro, MacNeil (2017) reitera a ideia de que o tradicional entendimento de ordem original assenta num conjunto de pressupostos cuja neutralidade é difícil de justificar (que os documentos que sobrevivem espelham o conjunto, que personificam o criador e permitem entrar na consciência deste), afirmando-se antes como uma construção socio-histórica imposta pelos arquivistas. Ao invés, as ordens intelectuais e físicas dos documentos vão sendo moldados à medida em que os documentos são recontextualizados por diferentes agentes, numa re-interpretação constante do seu significado ao longo do tempo, que deve ser celebrada como uma parte da história dos documentos, contribuindo para uma maior erosão do conceito de fundo como tendo um criador único, mas antes uma comunidade de criadores. Como resultado, a descrição arquivística deve abarcar a complexidade das circunstâncias históricas e dos contextos de criação dos documentos, não se limitando a uma representação linear e redutora.

Julie McLeod (2014) alerta contra perspectivas de gestão em arquivos demasiado ambiciosas, focadas no futuro, recordando que a gestão documental deve estar alinhada aos sistemas de informação e de negócio e à governança (*governance*), incluindo segurança e gestão de risco, o que

se considera acertado, uma vez que estas dimensões constituem o presente imediato e a razão de ser dos arquivos. Além disso, destaca como a formulação do problema da gestão documental é o próprio problema, uma vez que depende do contexto, das perspetivas e dos requisitos do sistema, que por sua vez dependem do *stakeholder*, isto é, da entidade que justifica a existência dos arquivos. Estes aspetos tornam os problemas de cada organização únicos, pelo que não se observa uma solução mágica válida para todos os cenários, e os impactos de cada solução apenas podem ser observados ao longo do tempo uma vez confrontados com as consequências, não de forma imediata.

Num artigo reflexivo de Alistair Tough (2016) parecem encontrar-se alguns ecos das ideias de McLeod. Neste artigo, o autor defende que para compreender o uso dado aos arquivos é fundamental compreender o contexto no qual foram criados. Neste sentido, destaca que a função primária dos documentos é a de constituir um produto de uma determinada atividade, não a de alimentar os arquivos. Por extensão, a função primária dos sistemas de gestão documental é igualmente a de administrar as práticas de negócio de uma instituição. As implicações para o presente estudo são as de que os metadados devem, em primeiro lugar, refletir os contextos de negócio dos quais os documentos de arquivos são um subproduto. Assim, se os metadados refletem o contexto, e se entender o contexto é essencial para compreender o uso dos arquivos, esta abordagem parece criar uma coerência racional entre a escolha dos metadados e condições adequadas para perceber o uso.

O mesmo repto é lançado por Kirsten Wright (2014) quando critica a ênfase colocada por Barbara Reed no *record-as-document*, como reflexo da noção de um documento-objeto que desconsidera ou ignora a transação que gerou o documento de arquivo. A autora considera que este aspeto é particularmente crítico nos documentos digitais, uma vez que, como o sistema de negócio fornece o contexto para a transação, se não houver metadados suficientes que documentem a transação, ou se estes metadados se encontrarem separados do documento, a compreensão do uso fica comprometida. Neste sentido, Wright defende que representar um documento é a única forma de facultar o acesso. Porém, como a representação é um processo complexo e mediado, sugere que o papel do criador da representação deve ser explicitado. Além disso, Wright menciona ainda o crescente papel dos arquivistas em facilitar o acesso a materiais normalmente não considerados documentos de arquivo, o que evoca a necessidade crítica de estabelecer relações (*links*) entre os recursos como forma de enriquecer as coleções, apontando como exemplo o uso de *linked data*. A autora sugere ainda uma expansão da definição do que deve ser considerado um *record*, enquanto intenção de alargar as fronteiras entre o material com que o arquivista trabalha. Curiosamente, esta expansão parece concretizar-se nesta tese através do uso do termo ‘recurso’ para designar os materiais a serem geridos e representados pelos sistemas de representação arquivísticos, e que não se limitam aos tradicionais documentos de arquivo.

A própria fronteira entre o que constitui um *record* (documento de arquivo) e um *document* (documento) tem sido questionada por Yeo (2011). Este artigo de discussão terminológica parte da ideia comum nos gestores de informação de que os documentos se tornam *records* consoante uma declaração de alguém, interpretação segundo a qual todos os *records* seriam documentos, mas nem

todos os documentos seriam *records*. O autor começa por salientar que os *records* não existem apenas em formato textual, podendo incluir áudio, vídeo ou mesmo artefactos 3D, como marcações em ossos enquanto registo de atividades. Um ponto forte da sua argumentação é o de que, com a revolução digital, deu-se uma separação entre *records* e um formato documental, na medida em que os *records* não estão necessariamente confinados a uma estrutura documental fixa e facilmente reconhecível, podendo também existir em sistemas de dados que contenham dados. Neste caso, um registo de movimentos bancários, apesar de serem dados, atuam como *records*. A ideia-chave é a de que se observam *records* em formatos não-documentais, de formas distintas dos protótipos abstratos em que estamos habituados a pensar.

Outra conceção que Yeo questiona é a de que os *records* devem ser materiais designados para preservação ou formalmente capturados num sistema de gestão documental, uma postura frequentemente tida pelos gestores documentais. Esta perspetiva implica a noção de que há um momento único no tempo que congela um documento e o cristaliza enquanto *record*. Porém, Yeo considera-a uma ideia muito frágil, uma vez que depende de decisões por pessoas, e as decisões humanas são necessariamente parciais e subjetivas. O autor argumenta que um *record* não muda de natureza porque alguém decidiu que é tempo de lhe dar uma preservação formal. Antes, entende-o como uma representação de evidência e um testemunho informacional, enquanto ocorrente das atividades. Assim, uma lição importante a retirar deste artigo parece ser a de que a fronteira tradicional entre *record* e documento é em parte artificialmente criada pelo que, mais do que estabelecer escolhas de preservação ou captura com base nesta terminologia, um sistema de informação deve procurar registar a informação que constitui evidência das atividades.

Outra interessante análise geral é fornecida por John Ridener (2009), que no capítulo sete da monografia *From Polders to Postmodernism: A Concise History of Archival Theory* oferece uma discussão dos paradigmas arquivísticos. O autor identifica quatro momentos. O primeiro período de ‘consolidação’, marcado pelo Manual Holandês de Muller, Feith e Fruin, identifica os arquivistas como organizadores dos documentos de arquivo. Sucedido pelo período de ‘reforço’, no qual Hilary Jenkinson parte dos fundamentos teóricos propostos pelo manual holandês, reforça os conceitos e posiciona o arquivista como um profissional. O terceiro período, a era moderna, ter-se-á sucedido após a Segunda Guerra Mundial, período em que Theodore Schellenberg batalha com a gestão das enormes massas documentais, decidindo organizá-las com base na sua natureza orgânica, e aprofunda a necessidade de avaliação (*appraisal*) dos documentos de arquivo. Por fim, Ridener identifica um quarto paradigma que caracteriza como ‘questionamento’, emergente a partir das décadas de 1960 e 1970, como fruto das crescentes exigências não só para maior abertura dos arquivos no acesso aos documentos, mas também para que se preservem os documentos de grupos marginais, tradicionalmente sub-representados. Trata-se sobretudo de um paradigma social e cultural que entende os como um meio de comunicação, que podem ser interpretados e questionados, de onde resulta que os arquivistas se tornam eles próprios questionadores dos documentos de arquivo. Um resultado é a abordagem pós-custodial — ela própria identificada com um paradigma — segundo a qual os arquivistas não são meros guardiães passivos dos documentos,

mas antes construtores ativos do seu significado e da forma como estes são representados. Será este último momento o paradigma que tendencialmente caracteriza as expectativas sociais no presente, e com o qual esta investigação se identifica.

#### 4.2.4 Metadados

Nesta secção procura-se abordar a literatura científica mais próxima do próprio conceito de metadados, com uma tendência preferencial para a perspectiva arquivística, embora alguns dos autores aqui incluídos não sejam arquivistas ou investigadores da arquivística como disciplina aplicada da CI, o que se acredita acrescentar uma salutar diversidade de abordagens. Assim, de entre os temas discutidos encontram-se o papel dos metadados na gestão do conhecimento, princípios para a criação de metadados, métricas para a qualidade dos metadados, desafios sobre o uso de metadados e metadados para a proveniência.

Na sua tese de doutoramento, Christina Apostolou (2009) comparou duas instituições ao nível das práticas no uso de metadados no âmbito de sistemas de gestão do conhecimento (*knowledge management systems*). De entre algumas das principais ilações retiradas da sua investigação, destacam-se:

- i. a identificação da posse (*ownership*) e das mudanças efetuadas aos documentos ao longo do tempo (*timeliness*) como aspetos importantes para fomentar a confiança no conteúdo. As avaliações de outros utilizadores e anotações podem também contribuir para acrescentar valor ao conteúdo dos documentos, bem como informação contextual (informação geográfica, funções específicas, área de investigação). De entre os metadados considerados mais necessários no desenvolver da atividade das empresas destacam-se: ‘controlo de versão’, ‘data de criação e modificação’, ‘título’, ‘descrição’, ‘localização primária’, ‘criador’.
- ii. o uso de taxonomias para os elementos de metadados que podem exibir valores pré-definidos, tais como regiões, facilitam de forma significativa a criação de metadados, além de enriquecer a sua qualidade.
- iii. o reconhecimento de que cada indústria tem diferentes requisitos, o que incentiva a autora a propor não um esquema de metadados, mas um quadro de referência (*framework*) de requisitos que pode fornecer orientações para a criação de esquemas de metadados adaptados às necessidades de cada organização.

No final do livro *Introduction to Metadata*, organizado por Murtha Baca (2016), encontra-se enumerado um conjunto de princípios práticos para a criação e a gestão de metadados. De entre estes, considera-se mais pertinente para esta investigação destacar que:

- i. a criação de metadados de qualidade deve ser entendida como uma das atividades fundamentais das instituições de memória, a par da preservação, do acesso e da disseminação das coleções.

- ii. a criação de metadados é um processo incremental e permanente que deve ser visto como uma responsabilidade partilhada entre as instituições e os utilizadores.
- iii. as instituições devem procurar automatizar os processos de criação de metadados de forma a reduzir o tempo gasto e a dificuldade do trabalho, sendo uma parte deste processo o foco em poucos elementos de metadados considerados essenciais.
- iv. a criação de metadados partilháveis e reutilizáveis deve ser incluída nas atividades diárias das instituições.

Num artigo de carácter inclusivo, no sentido em que considera as necessidades de utilizadores normalmente marginalizados, Wondwossen Beyene e Thomas Goodwin (2018) discutem os metadados adequados para qualquer tipo de utilizador, incluindo aqueles com deficiências que afetam o manuseamento de informação impressa, tal como cegueira, dislexia, pouca visão ou deficiências motoras (em inglês, *print disability*). Quando se arquetam os metadados de um sistema, se a aspiração é o acesso universal, deve ser tida em conta a diversidade de potenciais utilizadores em termos de idade, género, cultura e capacidades físicas e mentais, entre outros fatores. Uma vez que a organização dos resultados de pesquisa é uma das principais dificuldades sentidas, fornecer os tipos adequados de metadados é crucial para o melhoramento deste aspeto. Neste sentido, Beyene e Godwin recomendam:

- i. o uso de metadados facetados para flexibilizar o acesso e reforçar um *positive search abandonment*, isto é, a satisfação informacional através da consulta dos resultados da pesquisa, sem necessidade de analisar os documentos recuperados em maior pormenor.
- ii. a tolerância a erros de gramática pela ferramenta de pesquisa. Outros autores sugerem também preenchimento automático e ferramentas de sugestão de correções gramaticais.
- iii. para além de disponibilizar conteúdos em formatos alternativos, por exemplo em áudio em alternativa ao tradicional texto, devem ser incluídos metadados de acessibilidade que expressem a origem dessa voz (se humana ou gerada artificialmente).

Chen et al. (2011) enveredaram por um esforço semelhante de crítica dos metadados presentes em determinado repositório, desta vez com o objetivo de estabelecer métricas comuns para a qualidade dos metadados ao nível dos elementos de metadados. Entendendo a qualidade como *fitness for purpose*, os autores procuraram assegurar a qualidade dos metadados de formatos e fontes heterogéneas para facilitar a descoberta, a representação e a reutilização futura dos recursos. Para esta investigação destacam-se os parâmetros da proveniência e da completude dos metadados usados para a descrição dos recursos, uma vez que a proveniência está relacionada com a autoridade do criador, que por sua vez depende da asserção da sua identidade. Por seu turno, a completude e a precisão (*accuracy*) dos metadados são importantes para garantir que o objeto digital se encontra adequadamente descrito. Assim, os autores entendem que para o estabelecimento da proveniência são essenciais os elementos ‘Título’, ‘Identificador’, ‘Formato’, ‘Assunto’, ‘Direitos’ e ‘Editor’ (*Publisher*). Para a completude, devem ser indicados tantos metadados quanto possível,

salientando ‘Criador’, ‘Contribuidor’, ‘Formato’, ‘Editor’, ‘Fonte’ e ‘Cobertura’. Chen et al. também tecem importantes recomendações sobre a acessibilidade, indicando que, no caso de coleções que agreguem descrições de várias instituições, devem incluir-se nas descrições dos recursos URLs que remetam para as descrições nos repositórios de origem, bem como mecanismos de verificação da unicidade desses URLs, de modo a que cada *metadata record* se refira a um objeto digital.

Na monografia *Metadata: shaping knowledge from antiquity to the semantic web*, Gartner (2016) discute uma miríade de aspetos relacionados com o uso de metadados. Porém, dada a natureza da obra, o discurso assume frequentemente um carácter mais generalista, teórico, não propondo implementações específicas de elementos de metadados num determinado contexto. Ainda assim, considera-se pertinente realçar a defesa de Gartner da necessidade de reduzir a rigidez das hierarquias tradicionais usadas na representação dos recursos, adotando antes classificações facetadas como forma de flexibilização do acesso à informação. Outra forma de flexibilização é através do uso de ontologias para representação da informação, que potencia a capacidade de relacionar conceitos em redes de ideias e de relações tão flexíveis quanto as necessidades do utilizador, numa apresentação semelhante à de uma teia de aranha. O autor defende também que esta forma de representação parece também ser uma aproximação mais natural às estruturas de conhecimento como os humanos as percebem, com um certo caos. Gartner realça também as enormes conquistas do uso de triplos para a acessibilidade e reutilização da informação no contexto da *Web Semântica*, na qual a *Internet* é tratada como uma única coleção gigante, mas alerta para o possível problema de determinar a proveniência e a confiabilidade, uma vez que a origem dos recursos pode ser difícil de determinar. Embora não apresente requisitos específicos de solução, o autor esboça que as potenciais soluções podem residir na combinação das possibilidades técnicas de *linked open data* com mudanças culturais na gestão dos metadados.

Em outra monografia, Henry Gladney (2007) aborda a preservação de objetos digitais, recorrendo em pormenor sobre a confiabilidade dos objetos digitais. Porém, o discurso torna-se por vezes muito técnico, focando-se em especificações de sistemas tecnológicos, em questões de *hardware* ou até mesmo em fórmulas matemáticas. Apesar disso, podem ser retiradas algumas ilações úteis para esta investigação, embora generalistas. Este autor realça que cada objeto digital deve ser preservado com provas da sua própria proveniência e integridade. Tal certificação pode acontecer pelo uso de chaves criptográficas que permitem aos utilizadores testarem a autenticidade dos objetos digitais, ou através de uma asserção credível de como um repositório recebe os metadados do criador ou de outro repositório, de forma a certificar a proveniência da informação. Além disso, Gladney recorda que deve ser sempre considerado quem decide confiar e quando a decisão é tomada, realçando que se trata de uma decisão individual de cada utilizador. O máximo que o criador dos recursos pode fazer é encorajar essa confiança através de informação sobre a proveniência e integridade. Os metadados auxiliam na preservação da evidência da autenticidade, por exemplo através do registo de *audit trails*. Porém, realça que o que é considerado essencial para cada recurso depende em larga medida do contexto.

Os desafios no uso de metadados para a gestão da informação são abordados por David Haynes (2018). De entre as várias temáticas discutidas nesta obra segundo uma perspectiva teórica, sem compromisso com sugestões específicas de elementos de metadados, é possível retirar algumas observações sobre a autenticidade dos recursos, e de alguns metadados a usar. Haynes refere-se aos elementos de metadados em designações genéricas, uma vez que a forma específica do nome adotado por cada norma dependerá das necessidades específicas da comunidade em que se insere. Ainda assim, este autor salienta que:

- i. a autenticidade é reforçada por uma descrição apropriada dos recursos. Como parte dessa descrição, os metadados devem expressar-se, de uma forma geral, sobre o ‘Identificador’, o ‘Título’, o ‘Criador’, a ‘Data de criação’, o ‘Formato’, e a ‘Descrição’ do recurso.
- ii. os metadados constituem uma parte indispensável de políticas de avaliação. De entre estas, salienta dois tipos de metadados: os de intenção e os de ação. Os metadados de intenção estão associados a documentos com um tempo de vida previsto, ou a documentos já selecionados para eliminação. Nestes casos, os metadados podem ser usados para criar listas de documentos a eliminar, e quando. Quando a eliminação ocorrer, os metadados podem fornecer *audit trails*. Neste sentido, devem ser implementados alguns metadados relacionados com a eliminação, que devem expressar as noções de ‘Ação de eliminação’, ‘Período temporal da eliminação’, ‘Data prevista para a eliminação’, e ‘Eliminação autorizada por’.
- iii. estabelecer a proveniência dos metadados constitui um aspeto importante da qualidade. A proveniência pode ser atestada recorrendo a metadados que descrevam os metadados, respondendo a questões como: Quem criou os metadados? Quando foram criados ou modificados? Quais as circunstâncias na origem da criação dos metadados?

O uso dos metadados na gestão da informação é também discutido por Marcia Zeng e Jian Qin (2016) na vasta obra simplisticamente intitulada *Metadata*. Embora as autoras não dediquem nenhuma secção explicitamente à discussão de aspetos sobre a autenticidade ou a confiabilidade, destacam-se as seguintes ideias que podem atuar como requisitos:

- i. os metadados devem ser usados para descrever recursos segundo o princípio 1:1, que dita que deve existir uma e uma só descrição de metadados para um recurso. Por exemplo, a peça ‘António e Cleópatra’ constitui um recurso. ‘William Shakespeare’ é outro recurso, e ‘As obras completas de William Shakespeare’ são um terceiro recurso. Um registo de metadados relativo a esta peça contém um *description set* que contém três descrições (peça, autor, coleção). Cada descrição contém uma *statements*, que descreve uma propriedade (título) e um valor (António e Cleópatra).
- ii. descrições ao nível do item são mais adequadas para recursos cujas características individuais são de maior importância para os utilizadores, tais como cartas e fotografias.
- iii. informações ao nível da coleção constituem uma componente essencial da proveniência, da qualidade e da confiabilidade dos metadados. Este tipo de informação contextual deve



expressar as origens, a custódia e a propriedade de um item ou de uma coleção, podendo incluir elementos como título, entidade responsável, âmbito e finalidade, frequência de atualização.

- iv. metadados de autenticação para avaliação da integridade dos recursos. Alguns dos elementos a incluir para esta finalidade podem ser ‘fonte’, ‘relação’, ‘versão/edição’, ‘assinatura digital’.

Chunqiu Li e Shigeo Sugimoto (2017) pronunciam-se especificamente sobre como descrever a proveniência dos recursos usando metadados, e mantendo-os de forma consistente. Reconhecendo as potenciais inconsistências geradas a partir de mudanças nas designações dos termos de metadados e nas suas definições ao longo do tempo, os autores propõem um modelo simplificado de relações primitivas para descrever a proveniência, essencial para fazer julgamentos com base nas informações disponíveis de modo a determinar se é de confiança. Li e Sugimoto defendem que é necessário manter um histórico de revisões efetuadas aos termos de metadados que descrevem as entidades e as atividades, tanto ao nível das suas designações como das suas definições. Mais especificamente, propõem que sejam descritas adições, eliminações e substituições dos termos e das suas definições como requisito para estabelecer a proveniência das informações.

#### **4.2.5 Normas de metadados**

Nesta secção temática da revisão de literatura encontram-se assuntos tais como problemas existentes no mapeamento de metadados entre normas, dificuldades na implementação de normas, diferentes prioridades na conceção das normas, e requisitos das normas de metadados para adaptação à *Web Semântica*.

Mary Woodley (2016) alerta para vários problemas que podem surgir no processo de mapeamento dos metadados entre duas ou mais normas. De entre os desafios, a autora identifica a falta de alinhamento entre os metadados, incluindo, entre outros:

- i. equivalências imperfeitas: um conceito num esquema não tem um equivalente perfeito no segundo esquema.
- ii. um elemento para muitos: os dados que existem num elemento no esquema A podem existir em elementos separados no esquema B.
- iii. muitos elementos para um: os dados existentes em elementos separados no esquema A podem encontrar-se num único elemento no esquema B.
- iv. falta de correspondência: não existe nenhum elemento equivalente no esquema B.
- v. as normas evoluem ao longo do tempo, podendo passar a incluir novos elementos em diferentes contextos.

Este problema assume uma dimensão importante com os resultados do estudo de Andrade et al. (2014), que atesta a pluralidade de normas de metadados adotadas pelos repositórios digitais para a descrição dos seus conteúdos. Por outras palavras, sabendo que os repositórios tendem a

adotar duas ou mais normas, e conhecendo os problemas do mapeamento entre várias normas, torna-se fundamental tentar articular com clareza a forma como essas normas são cruzadas e integradas.

Porém, as normas não apresentam todas o mesmo enfoque. Jennifer Bunn (2013) alerta para uma distinção entre as normas para a descrição arquivística e as normas para a partilha da informação. Segundo a autora, as primeiras caracterizam-se por um olhar interno que se concentra no cerne da atividade dos arquivistas, ao passo que as normas de partilha de informação, baseadas na interoperabilidade dos metadados, adotam um olhar externo cujo foco é facilitar a partilha de informação sobre o material arquivístico. A autora entende que a descrição arquivística supera a mera partilha de informação, no sentido em que também se preocupa em garantir o significado e a autenticidade dos documentos de arquivo ao longo do tempo e através de vários domínios, e de comunicar esse significado e os problemas decorrentes da temporalidade e da multiplicidade de domínios dos documentos. Considera que as normas publicadas pelo ICA (ISAD[G], EAD) não cumprem os objetivos originais da descrição arquivística — abordar os problemas decorrentes da manutenção da autenticidade e do significado dos documentos através de comunidades e do tempo — atuando antes como coletes de forças (*sraitjackets*) para a partilha da informação. Com base neste raciocínio, Bunn apela a que se faça uma distinção entre um enfoque na partilha de informação ou na descrição arquivística e que se observem as relações entre normas de diferentes áreas, como as provenientes da preservação digital (p.e. MoReq) e da descoberta dos recursos (p.e. MODS, Dublin Core).

Num artigo explicativo do *Open Archival Information System* (OAIS), Jean Dryden (2009) destaca alguns aspetos que contribuem para o sucesso deste modelo conceptual, nomeadamente (i) o facto de ser *open*, tanto no sentido colaborativo do seu desenvolvimento como na sua disponibilização gratuita, e (ii) a clareza da redação, que explica claramente o propósito da norma, orienta o leitor para a estrutura que deve ler de acordo com as suas necessidades, e adota uma linguagem compreensível que permite que a norma seja facilmente compreensível por quem não é familiar com a área. Estes aspetos que Dryden louva no OAIS parecem ser requisitos importantes para a redação de qualquer outra norma, não sendo exceção o modelo de requisitos que será proposto por esta investigação.

Dunia Llanes-Padrón e Juan Pastor-Sánchez (2017) também se expressam sobre alguns requisitos das normas de metadados. Pensando num contexto de *Web Semântica*, caracterizado pela democratização da informação, os autores estabelecem como premissa para a descrição arquivística a procura de soluções teóricas e tecnológicas que facilitem a pesquisa, a recuperação e o uso da informação na *web*, enquanto plataforma universal de integração. É salientado que os modelos conceptuais para a descrição arquivística devem facilitar a interoperabilidade entre os sistemas de informação de diversas comunidades, tais como arquivos, bibliotecas, museus. Llanes-Padrón e Pastor-Sánchez centram o discurso na análise da primeira versão *draft* da norma *Records in Contexts* (RiC) como uma plataforma para defender uma maior interoperabilidade semântica das descrições arquivísticas com a finalidade última de promover a reutilização, a fiabilidade e a

acessibilidade dos recursos num contexto *web*. Como parte deste processo, a RiC atribui grande importância ao contexto de produção dos documentos, pronunciando-se com clareza sobre os propósitos dos documentos e sobre as entidades que os produziram, reconhecendo a pluralidade de contextos e de propósitos que a descrição arquivística deve capturar.

Numa linha de pensamento muito próxima, Llanes-Padrón e Manuela Moro-Cabero (2017) recuperam o exemplo da RiC como um veículo de reflexão para as vantagens de uma descrição pluridimensional que contenha as representações e as relações entre várias entidades arquivísticas, não só baseada no princípio da proveniência, mas também centrada na separação e inter-relação dos principais componentes da descrição arquivística a fim de proporcionar uma base de interoperabilidade entre diferentes sistemas. As autoras destacam ainda o modo como a descrição de entidades arquivísticas mediante metadados viabiliza a manutenção da autenticidade e da integridade dos documentos através da documentação do contexto.

Num raciocínio similar, Gregory Rolan (2017) propõe um meta-modelo para metadados de gestão da informação, enquanto uma tentativa de apoiar a interoperabilidade entre sistemas diferentes, através de um enfoque na descrição das atividades, por contraste com uma perspetiva demasiado focada nos artefactos produzidos pelos agentes. Embora interessantes, estas abordagens colocam-se naquilo que Bunn diria serem normas ou modelos para a partilha de informação, não necessariamente para a descrição arquivística no sentido de um olhar interno sobre a natureza dos recursos, motivo pelo qual a arquitetura destes modelos para a interoperabilidade não é explorada em maior profundidade.

Num artigo que discute a norma *Australian Recordkeeping Metadata Schema* (RKMS), Sue McKemmish et al. (1999) apontam a necessidade de tornar os metadados explícitos, e de os associar de forma persistente com os documentos, um esforço fundamental para quando os documentos transitam entre comunidades, no sentido em que o significado implícito de uma comunidade deve ser explicitado para que esse significado seja perceptível por outros domínios. Este exercício é particularmente importante na *web*, dada a possibilidade de um utilizador poder pertencer a qualquer domínio. Além disso, os autores posicionam a descrição como um processo iterativo de captura e relacionamento de metadados com documentos criados no contexto de atividades sociais, cuja finalidade é a de fornecer um controlo intelectual que permita que documentos autênticos, confiáveis, com significado e acessíveis possam ser mantidos no tempo e através de limiares organizacionais por tanto tempo quanto necessário para os múltiplos propósitos a que se destinam. De entre os vários elementos de metadados descritos no (RKMS), destaca-se, ao nível dos documentos de arquivo, a inclusão de metadados para a avaliação, o controlo, a preservação, a recuperação, o acesso e o uso, embora sem definições que pudessem facilitar a interpretação. Um histórico de auditoria dos processos sobre os documentos é também proposto pelo elemento ‘Event History’.

Num caso de estudo sobre a adoção da ISAD(G) num arquivo da Coreia do Sul, Eunha Youn (2015) identifica, através das perceções dos funcionários, que a inclusão de demasiados elementos de metadados numa norma contribui para resistências na sua implementação, uma vez que a

aplicação de todos os elementos se torna um esforço demasiado intensivo para os funcionários. A criação de descrições multiníveis, como sugerido pela ISAD(G), é outro fator que contribui para o aumento da carga de trabalho, da própria dificuldade do trabalho, e do tempo despendido. Além disso, a autora também identifica a limitada educação arquivística entre os técnicos como um obstáculo à interpretação e implementação da ISAD(G) no arquivo sob estudo.

#### 4.2.6 Autenticidade

Os trabalhos classificados como pertencentes a esta secção têm uma intenção explícita, geralmente no título, de discutir a autenticidade e/ou a confiabilidade de recursos, documentos ou informação. Alguns artigos são de carácter teórico, ao passo que outros reportam requisitos de projetos, de carácter mais aplicado. De entre os tópicos abordados destaca-se a confiabilidade sob a perspetiva da preservação digital, o uso da tecnologia *blockchain*, a confiança em recursos alojados na *web*, e abordagens diplomáticas à autenticidade.

Lina Bountouri et al. (2017) apresentam o projeto iniciado pelo *Publics Officer of the European Union* para a preservação das publicações digitais no seu repositório digital, procurando garantir que o repositório digital arquivístico pode ser confiável, isto é, que um objeto digital se mantém o mesmo em relação ao momento da sua criação, não tendo sido alterado durante o seu ciclo de vida, nem antes nem após a sua ingestão no repositório. O projeto adota a ISO 14721:2012, que descreve o *Open Archival Information System* (OAIS), e a ISO 16363:2012 para verificar a confiabilidade do repositório arquivístico digital. De entre os vários requisitos discutidos, destacam-se os seguintes como relacionados com a descrição:

- i. a necessidade de documentar a invariabilidade (*fixity*), uma propriedade dos objetos digitais que indica que estes não foram alterados entre dois momentos no tempo. Os autores indicam que a invariabilidade deve ser sempre integrada como parte dos metadados de proveniência enquanto contributo para a prova da autenticidade e da integridade de um objeto digital ao longo do tempo. Pode assumir a forma de códigos de verificação (*checksums*<sup>84</sup>) ou de verificação de tamanho de ficheiros, entre outros.
- ii. os metadados de proveniência são identificados como uma das fontes de informação mais importantes para documentar o ciclo de vida de um recurso, descrevendo todos os eventos que possam ter produzido uma mudança durante o ciclo de vida. Os autores subscrevem a noção de que a identidade reside só não na estrutura interna e no conteúdo de um objeto digital, mas também no complexo sistema de relações que estabelece com outros objetos digitais, no sentido em que uma alteração a um objeto afeta outros objetos, reforçando a

---

<sup>84</sup> Um *checksum* é um número calculado agregando toda a informação de um ficheiro, pelo que se o conteúdo for alterado, o número obtido é distinto. Este método é frequentemente usado como forma de averiguar se um documento é idêntico ao original, uma vez que a probabilidade de dois ficheiros diferentes terem o mesmo código é residual.

ideia da importância da informação de proveniência para a presunção da autenticidade. Esta informação deve ser recolhida desde a criação dos objetos.

Chris Duncan (2009) explora em profundidade o conceito de autenticidade dos documentos de arquivo sob uma perspectiva filosófica, através da lente dos arquivistas. O autor começa por debater várias interpretações do termo, desde Hillary Jenkinson a Luciana Duranti, do glossário da *Society of American Archivists* à ISO15489, como justificação da complexidade da negociação deste termo, implicando uma luta considerável entre perspectivas subjetivas e evolutivas do que são considerados documentos autênticos. O significado do termo é, afinal, dependente das perceções e práticas de quem o emprega. Neste sentido, o autor afirma que qualquer juízo apenas pode ser feito mediante a análise do contexto do qual o sentido emerge, juízo esse que comporta graus variáveis de mais ou menos autêntico de modo que se distingam os documentos autênticos dos não-autênticos dos provavelmente autênticos. Da continuação da discussão da autenticidade, que Duncan defende dever ser referido mais como um conceito do que como um termo, emergem os seguintes requisitos:

- i. a memória deve ser uma das principais questões a considerar quando se proclama a autenticidade devido à sua veracidade enquanto um repositório fidedigno da experiência passada. Além disso, a subjetividade emerge a partir da perda de memória, na medida em que é devido à perda de memória que se criam documentos enquanto fragmentos de informação do passado.
- ii. uma segunda noção é a de prova. Se a memória é uma forma de evidência histórica, a prova é o juiz dessa evidência. O autor compara a memória dos documentos aos genes, ao passo que a prova seria o teste de ADN. Isto significa que para um documento ser considerado autêntico, deve fundamentar-se numa memória autêntica certificada por uma prova. Relacionar memória e evidência apenas se torna possível mediante o escrutínio de várias camadas de informação através das quais se colocam questões sobre a prova e, se essas informações forem encontradas, o documento torna-se crescentemente autêntico.
- iii. a sugestão de que a autenticidade deve ser vista mais como um conceito do que como um termo, por virtude da complexa série de interpretações subjetivas em jogo. Defende que a noção de autenticidade não deve ser formulada numa forma abstrata: ‘é autêntico?’, mas com base num referente: ‘é autêntico em relação a quê?’. Assim, retira-se a ilação de que os arquivistas devem questionar-se sobre os objetivos da autenticidade e questionar a sua importância para as coleções, bem como perguntar como as coleções podem beneficiar da autenticidade. Entender a autenticidade como um conceito relativo significa localizar o ponto de referência do ‘outro’ mediante o qual se mede o grau de autenticidade, e assenta largamente em identificar atributos essenciais que podem ser usados para expressar o valor dos documentos.

Mais focado nas questões da confiança nos documentos no contexto digital, em particular nos serviços de *cloud storage*, surge o artigo de Luciana Duranti et al. (2019). Os autores estabelecem

uma premissa de crise de confiança generalizada devido à dificuldade em conhecer ou avaliar as fontes de informação cada vez mais dispersas pela *web*, ao mesmo tempo que se observa uma crescente dependência dos serviços *online*, a par de uma quebra histórica da confiança em organizações públicas e privadas. Porém, qualquer interação com serviços digitais gera uma marca, dados que as organizações recolhem sobre nós que levantam necessariamente questões sobre a capacidade de confiar nessas organizações para usarem esses dados em nome dos interesses dos utilizadores. Deste artigo destacam-se as seguintes ideias:

- i. a capacidade de confiar nos documentos e nos dados digitais não pode mais assentar sobre os mesmos pressupostos que governaram a confiança nos documentos em papel e noutros suportes analógicos (o precedente legal e o conhecimento acumulado ao longo dos séculos). A natureza das relações de confiança na Internet é fértil em riscos, falhas, e insuficiências dos recursos digitais.
- ii. diferentes sistemas para avaliar a confiança são necessários em diferentes contextos, uma vez que os parâmetros usados para avaliar a confiança podem variar de acordo com o contexto cultural.
- iii. a confiança pode ser vista como um processo empírico segundo o qual ela é construída ao longo do tempo através da aplicação de certos padrões em certas situações. Neste contexto, a construção da confiança é incremental e informada pela experiência pessoal.
- iv. perante os usos e abusos das organizações com os dados que são necessariamente gerados durante as atividades na *web* (através do comércio digital, redes sociais, *e-mail*, utilização de motores de pesquisa, entre outros), a confiança nos dados e nos documentos digitais, bem como nas organizações que os recolhem e gerem, depende de aspetos críticos tais como transparência e responsabilização (*accountability*). A questão da transparência aplica-se também aos metadados, devendo ser indicado quem os gere. Por sua vez, a *accountability* garante que uma organização deve ser capaz de responder pelas suas ações legal, administrativa e historicamente, ao longo do tempo. Os autores defendem que a conjugação da transparência e da *accountability* garantem aos utilizadores a segurança de uma solução em caso de falha, promovendo assim a confiança.
- v. A transparência estende-se ao conhecimento da origem, proveniência e cadeia de custódia dos documentos e dos dados, assumindo-se como uma forma de objetividade.
- vi. os metadados, registos (*logs*) e auditoria (*audit trails*) emergem também como meios mais diretos de estabelecer a confiança, na medida em que fornecem mecanismos para estabelecer a autenticidade dos documentos e dos dados, nomeadamente a identidade e a integridade. Os metadados podem ser gerados por humanos ou automaticamente, ao passo que os *logs* e *audit trails* são gerados automaticamente pelo sistema de informação.

De igual forma focadas no tópico da confiança nos serviços *cloud* como armazenamento de documentos de arquivo encontram-se Julie McLeod e Brianna Gormly (2017). Os serviços de armazenamento na *cloud* têm-se tornado omnipresentes na rotina das organizações, por vezes

constituindo até o próprio padrão de atividade. Porém, a dependência de fornecedores de serviços para o armazenamento de coleções digitais levanta questões importantes sobre a confiança nesses serviços. As autoras apresentam os resultados de um estudo anterior em que participaram no qual foi elaborado um inquérito a elementos da comunidade arquivística, tendo sucessivamente sido entrevistados profissionais na área da gestão documental. Um dos achados importantes deste estudo é o da percepção da necessidade de longevidade dos serviços como base de confiança, que no caso da descrição arquivística se pode traduzir no uso de identificadores permanentes. Este artigo também ilustra a subjetividade da percepção da confiança, com vários respondentes a eleger diferentes critérios, dependendo da sua experiência ou função na organização (funcionário ou diretor). De um modo geral, as autoras identificam três preocupações que afetam a confiança: a sustentabilidade do serviço, a sua viabilidade económica e a capacidade em responder às necessidades dos documentos.

A questão da confiança dos documentos digitais é também discutida por Henry Gladney (2009), neste caso sob a lente da preservação digital. O autor assume uma posição crítica da tecnologia de gestão de conteúdos (*content management*), uma vez que considera que não é adequada para a preservação da autenticidade e da inteligibilidade dos documentos digitais a longo prazo. Como alternativa, sugere a arquitetura do *Trustworthy Digital Object* (TDO) enquanto veículo para os arquivistas gerirem repositórios digitais que procurem preservar a integridade intelectual dos documentos digitais de forma confiável no longo termo. De entre os requisitos identificados destacam-se os seguintes:

- i. a informação de proveniência deve estar associada ao documento, que por sua vez deve estar protegido contra alterações indevidas. Gladney classifica este tipo de metadados como autoritários (*authoritative metadata*).
- ii. um objeto é tornado *self-describing* fornecendo por exemplo provas sobre a autenticidade e sobre informação contextual.
- iii. a integridade é assegurada por uma autenticação criptográfica. A informação essencial a veicular é um carimbo temporal, o algoritmo de identificação da assinatura, o identificador da assinatura da autoridade, e a chave pública da autoridade assinante. Outros mecanismos que contribuem para estabelecer provas de integridade e de proveniência são marcas de água.
- iv. uso de um identificador único para cada documento, ou até mesmo para cada versão de cada documento.
- v. tornar explícita informação que permita fazer a auditoria de mudanças efetuadas sobre os documentos (*audit trail*).

A confiança sob a perspetiva de um *continuum* na era da informação é discutida por Dara Price e Johanna Smith (2011). Apoiando-se no contexto dos documentos federais do governo canadiano, as autoras refletem sobre o uso de um modelo de confiança, que assenta na análise das relações entre o criador, os arquivos e o utilizador. Uma das observações notadas é a de que a única forma

de garantir que os arquivos refletem a sociedade que os produziu é integrando as coleções de várias instituições, criando uma rede de instituições de memória que contribuam para enriquecer o contexto. As autoras reforçam também a ideia de que um repositório digital, para ser confiável, carece de metadados suficientes sobre o objeto que demonstrem a integridade e a autenticidade, e que documentem todas as alterações, migrações, ou intervenções sobre os objetos, posicionando assim os requisitos ao nível dos metadados como um aspeto crítico da presunção da confiança.

Tove Engvall (2019) também aborda a questão da confiança no ambiente digital, mais especificamente no contexto do negócio recorrendo a estudo etnográfico do comércio *online*. Uma ideia importante expressa por Engvall é a do equilíbrio entre *accountability* e confiança. A *accountability* ajuda a controlar os comportamentos, enquanto a confiança assenta numa crença no comportamento dos outros, que se espera que sejam previsíveis, controláveis e de boa-fé. O autor subscreve o modelo de confiança de Kelton et al. (2008), que identifica vários níveis de confiança (individual, interpessoal, relacional e ao nível da sociedade). A nível individual, este modelo entende a confiança enquanto um traço psicológico que se reflete nas expectativas que um indivíduo tem de outrem, e que depende da sua personalidade, das suas experiências e do seu contexto social. Além disso, é referido que a confiança apenas emerge sob as condições de incerteza e de vulnerabilidade. Por um lado, a confiança pretende reduzir a incerteza. Por outro, a confiança apenas importa se um dos participantes se colocar numa posição de vulnerabilidade que pressupõe que a entidade confiada detém o potencial para satisfazer as necessidades do confiante.

A questão da confiança no contexto digital é também explorada por Geoffrey Yeo (2013), subscrevendo a ideia de que a documentação da proveniência e do contexto constituem as bases para realçar a transparência dos procedimentos documentais e para reforçar a confiança neles. Porém, o autor identifica uma crescente descrença na tradicional ideia de objetividade dos arquivistas, enquanto guardiães neutros nos quais se possa confiar para proteger os documentos e os manter incorruptíveis. Além disso, identifica também a perda de intermediação resultante de acessos *online* como um fator que contribui para a cada vez menor interação com os arquivistas ou com a dimensão física de uma instituição. De forma a combater esta emergente consciência da falta de objetividade, Yeo recupera a ideia de que *transparency is the new objectivity* (Weinberger, 2009) no sentido em que a transparência expõe o contexto de criação de um recurso, que por sua vez fornece as bases para a confiança. Emerge assim uma relação umbilical entre a proveniência e a confiabilidade, sobretudo no contexto da *Web Semântica*, uma ideia recorrentemente ecoada ao longo do artigo. É ainda destacado que, de forma a obter metadados descritivos em número suficiente sem sobrecarregar irrealisticamente os intervenientes, é necessário adotar mecanismos automáticos de criação de metadados, por exemplo através de inteligências artificiais que analisem a forma, o conteúdo e o contexto dos objetos para criar descrições apropriadas. Outro aspeto mencionado é o da infinidade do contexto, de onde deriva que os arquivistas e as instituições devem tomar uma decisão sobre que contexto se deve realçar em detrimento de outro, pelo que as representações do contexto são parciais e imperfeitas. Yeo realça ainda que depositar a confiança num determinado recurso é uma decisão pessoal, variável de indivíduo para indivíduo, destacando



igualmente que as próprias representações a que se chama informação/metadados de proveniência constituem um documento no sentido em que são registos de asserções que alguém fez sobre eventos passados. Segundo o autor, tanto os recursos como os metadados que os descrevem carecem de autenticidade e de fidedignidade.

O entendimento de discussões em torno da autenticidade como mecanismo de negociação de ordem social é discutido por Bonnie Mak (2012). A autora argumenta que o questionamento da autenticidade das fontes digitais pode servir como uma tentativa de instaurar mudança e de negociar *status* político e social, à semelhança de negociações que ocorreram em séculos passados. Questões sobre a autenticidade das fontes são assim vistas como uma forma de proclamar uma visão alternativa do mundo, enquanto ponto de contestação. No século 21, alguns dos pontos de negociação podem ser o uso dos recursos digitais à luz da teoria arquivística e a própria a autoridade/*status* dos intervenientes (criador, custodiante).

Outros autores há que abordam a autenticidade sob uma perspetiva diplomática. Um desses autores é Adam Jansen (2015), que propõe uma metodologia centrada na diplomática dos objetos digitais como instrumento para a elaboração de documentos de arquivo autênticos ao longo do tempo, mesmo quando removidos dos seus sistemas originais de criação. Para isso, o autor procura identificar quais os elementos de metadados considerados essenciais para a presunção da autenticidade, procurando captá-los através de *software*, normalizá-los e prever mecanismos de extensão de mais elementos consoante as necessidades de cada tipologia documental. Eis alguns dos requisitos apontados:

- i. deve ser procurada uma autodescrição dos processos de génese, constituição interna e transmissão, identidade dos criadores e dos custodiantes, e da relação dos documentos com as atividades em que participaram, de forma que os objetos digitais atuem como contentores de todos os elementos descritivos necessários para a sua interpretação, independentemente de sistemas terceiros.
- ii. Devem ser considerados os seguintes elementos de metadados nucleares de um conceito abstrato de um documento de arquivo: ‘action’, ‘bond’, ‘addressee’, ‘writer’, ‘author’, ‘adminContext’, ‘digitalObject’, ‘ingestDate’. Defende que estes elementos devem ser definidos *a priori* de modo a estabelecer um controlo sobre os atributos da descrição e sobre os valores de cada atributo, isto é, a forma como esses elementos são preenchidos.
- iii. definido o modelo ideal de documento, Jansen propõe três características que devem ser observadas na representação arquivística: herança, encapsulamento e polimorfismo. A herança consiste num mecanismo segundo o qual novas classes de objetos podem ser derivadas de classes existentes, permitindo uma construção incremental dos documentos desde o nível mais alto de abstração, uma vez que cada camada sucedente herda os atributos do seu parente. O conceito de encapsulamento significa aqui a cristalização intencional de certos atributos, não permitindo alterações, como tentativa de estabelecer a imutabilidade de metadados de integridade e de identidade. Por fim, o polimorfismo pressupõe a possibilidade de criação de dois objetos do mesmo tipo, embora com elementos de

metadados ligeiramente diferentes. Estas abordagens diplomáticas podem ser difíceis de justificar no contexto da descrição arquivística digital. Por um lado, a herança pressupõe a existência de uma estrutura hierárquica que frequentemente não existe no âmbito da *Web Semântica*. Por outro, o contexto dificilmente pode ser congelado, uma vez que este é constantemente reconstruído e reconfigurado ao longo da vida dos recursos. De igual forma, a identidade e a integridade devem ser constantemente demonstradas ao longo do tempo, uma vez que não parece ser plausível a cristalização de um objeto digital sem necessidade de intervenções futuras.

Corinne Rogers (2015) também explora na sua tese de doutoramento a autenticidade dos documentos digitais sob uma perspetiva diplomática. Notando um desfasamento entre as sugestões de projetos e casos de estudo que procuravam estabelecer mecanismos de determinação e proteção da autenticidade (nomeadamente o InterPARES), e entre as práticas de profissionais de gestão documental, a autora realizou uma série de entrevistas a arquivistas e gestores de documentos com o fim de compreender a perceção da autenticidade por estes profissionais, e que procedimentos empregam nesse sentido. Deste estudo destacam-se as seguintes ideias:

- i. a noção de que o conceito de autenticidade é um constructo social dependente do contexto ou da disciplina que o define ou interpreta.
- ii. os procedimentos de verificação de autenticidade com recurso à análise dos mecanismos físicos de inscrição de uma mensagem num suporte físico, e assinando-o, não são plausíveis em contexto digital, uma vez que não é possível avaliar a autenticidade por inspeção visual. Com os documentos digitais, os metadados tornam-se fontes de informação sobre a autenticidade dos documentos na medida em que permitem a identificação das suas características.
- iii. o entendimento da descrição arquivística como um indicador social, isto é, enquanto um instrumento desenvolvido por uma organização para apoiar a criação, gestão e preservação dos recursos. O aspeto social parece remeter para a subjetividade das representações. De entre outros indicadores da autenticidade são referidos: informação sobre o *software* que gerou os documentos, informações sobre ações exercidas sobre os documentos, agendas de retenção e eliminação, registos de auditorias, técnicas criptográficas de validação e metadados normalizados.

Adotando uma postura diferente, Hofman et al. (2019) analisam a tecnologia disruptiva *blockchain*<sup>85</sup>. Embora as suas preocupações sejam em como conciliar a imutabilidade da

---

<sup>85</sup> Um *blockchain* público é um registo (*ledger*) distribuído, semelhante a uma base de dados, mas com a particularidade de, ao invés de ser controlado por uma autoridade central, o registo é disperso por vários computadores (um nó) de qualquer parte do mundo detidos por qualquer indivíduo com ligação à Internet. Novos registos apenas são acrescentados em tempo real por um mecanismo de consenso aceite pela maioria

informação registada no *ledger*<sup>86</sup> com os requisitos de proteção de dados no âmbito do Regulamento Geral de Proteção de Dados (RGPD), os autores referem que, com o uso desta tecnologia, se torna extremamente difícil classificar documentos de arquivo porque o vínculo arquivístico não é manifestado no *blockchain*. Como solução, a identificação da identidade dos documentos é mediada pela criação detalhada de metadados. Este raciocínio parece realçar o papel crítico que os metadados assumem também nesta tecnologia na caracterização da identidade dos documentos. No *blockchain*, a integridade reside na própria força dos mecanismos matemáticos e criptográficos nos quais o cálculo das transações assenta, sem intermediários humanos. Assim, o sistema afirma-se como *trustless*, no sentido em que não existe um elemento humano que possa servir como autoridade, com a vantagem de garantir que esse elemento humano também não introduz corrupção nem más intenções no sistema. Hofman et al. reconhecem o papel que a tecnologia de *blockchain* pode assumir em garantir que as organizações possuam documentos de arquivo facilmente auditáveis e resistentes à corrupção.

A tecnologia de *blockchain* é também explorada de forma conceptual por Victoria Lemieux (2016), desta vez explicitamente enquanto uma solução para a criação de documentos de arquivo digitais confiáveis. Depois de estabelecer uma base de presunção da confiança a partir dos requisitos de várias normas de gestão documental (ISO 15489, ISO 14721 e ISO 16363), para as quais os conceitos de fidedignidade, autenticidade, identidade, integridade e proveniência são caros, a autora argumenta que a tecnologia *blockchain* não contribui para a fidedignidade (*reliability*), uma vez que não tem forma de garantir a fidedignidade dos documentos antes da sua criação, isto é, antes da sua entrada na cadeia enquanto registo. Porém, esta tecnologia pode prestar fortes contributos para a autenticidade através da preservação da integridade dos documentos usando chaves criptográficas e fornecendo também informações detalhadas sobre transações que podem servir para auditoria. Porém, é de destacar que a autora baseia as suas reflexões numa implementação específica da tecnologia no contexto das Honduras pela empresa Factom, que por sua vez assenta num entendimento da tecnologia não só focado numa perspetiva de criptomoedas, como também, dentro desse enfoque financeiro, considera apenas Bitcoin. A tecnologia tem evoluído bastante nos últimos quatro anos, pelo que algumas das dificuldades notadas pela autora podem já ter sido ultrapassadas.

---

dos nós através de métodos criptográficos. Além disso, devido à arquitetura do sistema, os registos, uma vez criados, não podem ser removidos. Assim, como características essenciais do *blockchain*, destacam-se: descentralização (não há intermediários), distribuição (o registo é partilhado por toda a rede), encriptação, anonimidade (os identificadores são apenas números), e integridade dos dados (os dados não podem ser removidos).

<sup>86</sup> O termo '*ledger*' denota como que um livro de conta-corrente no qual se registem informações. O *blockchain* assenta num *ledger* público e de livre acesso que regista matematicamente, de forma automática e virtualmente incorruptível, todas as transações ocorridas, incluindo intervenientes e datas.

Sharmila Bhatia e Wright de Hernandez (2019) também abordam a tecnologia de *blockchain*, embora de uma perspectiva vantajosa em relação a Lemieux porque o seu artigo é mais recente, e já inclui os desenvolvimentos no *blockchain* do Ethereum, que permite a criação de *smart contracts*<sup>87</sup> e aplicações descentralizadas. A possibilidade de programação de *smart contracts* executáveis apresenta uma oportunidade para a gestão documental no âmbito dos arquivos. Um dos requisitos apontados é a possibilidade de fornecer informações sobre a proveniência dos documentos através da identificação da custódia (*ownership*) dos materiais ao longo do tempo, mediante um registo constante validado por assinaturas. A confiança é também reforçada pela manutenção de um *ledger* público estabelecido de forma distribuída, sem um controlo central que possa ser facilmente manipulado. Porém, os autores partilham as preocupações de Lemieux quando referem que o *blockchain* apenas garante a integridade e autenticidade dos documentos após a sua introdução no sistema, ou seja, coloca-se a questão da fidedignidade no momento da criação dos registos.

#### 4.2.7 Linked Data

Esta secção abarca obras de investigação que refletem sobre os arquivos à luz de conceitos como *Web Semântica* (WS) e *linked data* (LD), portanto geralmente focadas na interoperabilidade, na partilha e no reuso das descrições no contexto digital.

Um dos artigos que discute formas de promover a interoperabilidade semântica mediante descrições arquivísticas é o de Richard Gartner (2015), que propõe um esquema em XML construído a partir da EAD, mas focado mais numa descrição de elementos como se fosse uma base de dados, em vez de seguir uma abordagem mais documental. A codificação dos dados não é o objeto de estudo desta tese de doutoramento, mas o autor apresenta primeiro algumas ideias sobre as vantagens da interoperabilidade e dos requisitos de dados necessários para a alcançar:

- i. o reconhecimento de que, apesar dos sucessos da EAD em fomentar um catálogo coletivo de arquivos, da mesma forma que as normas MARC facilitaram a criação de largos catálogos coletivos no setor das bibliotecas, a EAD não é apropriada para criar o grau de interoperabilidade das descrições arquivísticas desejado no contexto digital. De entre as razões são apontadas a flexibilidade da sua arquitetura, que permite que conceitos semelhantes sejam registados de formas diferentes, e a sua orientação mais para a descrição de documentos, centrada em aspetos textuais, do que para a descrição atomística de dados, a múltiplos níveis de granularidade.
- ii. a interoperabilidade, entendida como a capacidade de partilhar e pesquisar informação em simultâneo em bases de dados diferentes, como se fosse numa só, é mais fácil de alcançar quanto mais pequenas e precisas forem as unidades de dados que formam os elos entre os

---

<sup>87</sup> A nível fundamental, *a smart contract is a self-executing contract with the terms of the agreement between buyer and seller being directly written into lines of code. The code and the agreements contained therein exist across a distributed, decentralized blockchain network. The code controls the execution, and transactions are trackable and irreversible* (Frankfield, 2019).

recursos, constituindo assim um argumento para o uso de triplos no âmbito da *Web Semântica*.

- iii. porém, aponta também três argumentos contra o uso de RDF como um formato para a descrição arquivística: (i) o vasto número de triplos necessário para codificar mesmo uma simples descrição; (ii) os potenciais problemas relativos aos direitos intelectuais que podem derivar da perda de controlo sobre os dados em ambientes RDF; e (iii) problemas de curadoria entre os ambientes fluídos de dados em RDF e os contentores mais rígidos exigidos por modelos de preservação como o OAIS. Ainda assim, o autor destaca que devem ser usados identificadores únicos como URIs para definir conceitos e as suas relações como um mecanismo de fomentar a interoperabilidade semântica na *web*.
- iv. uma descrição arquivística rica deve ser capaz de conjugar as nuances das descrições textuais com os requisitos mais atomísticos da interoperabilidade.
- v. de entre as facetas (elementos de dados) identificadas pelo projeto *Collaborative European Digital Archive Architecture* (CENDARI), destacam-se: descrição da coleção (identificadores, títulos), descrições dos componentes ('localização física', 'descrição material', 'descrição do conteúdo'), instituições detentoras, datas, relações, informações sobre direitos, impedimentos de uso, linguagem e assunto. É de notar que este projeto tem como missão integrar coleções de múltiplos arquivos e bibliotecas nos temas de história medieval e moderna, pelo que a escolha de alguns destes elementos não pode ser dissociada do seu contexto.

Num completo livro que aborda o uso de LD em arquivos, bibliotecas e museus, Seth van Hooland e Ruben Verborgh (2014) apresentam várias formas de modelação de dados (o modelo tabular, o relacional, o hierárquico e o em grafo), destacando que não existe uma solução universalmente melhor do que outra, mas que todas têm o seu uso dependendo do contexto e da finalidade. Porém, destacam que o recorrer a LD, que assenta no uso de triplos para expressar metadados, é o tipo de modelo de dados que oferece maior potencial e flexibilidade para a expressão de metadados na *web*, contribuindo para uma interoperabilidade semântica. Os autores destacam que é fundamental os metadados serem expressos de uma forma que permita o processamento automático dos metadados por agentes computacionais, de forma a possibilitar a integração automática de várias coleções, que demasiadas vezes atuam como 'ilhas' de informação, sem comunicação entre si. Também referem que a unidade de informação essencial no sistema de arquiteturas sobre o qual a Internet é construída, a arquitetura REST, é o recurso, uma entidade concetual que deve ser identificada de forma única. Adicionalmente, toda a partilha de dados nas arquiteturas REST, que é o mesmo que dizer na Internet, ocorre mediante a partilha de representações dos recursos, isto é, metadados.

Além disso, Hooland e Verborgh alertam para a qualidade dos metadados, para a qual adotam o entendimento da ISO 9000, que pode ser simplificado na expressão *fitness for purpose*. Porém, alertam que a finalidade dos metadados depende em larga medida do contexto em que é usada, no

sentido em que o que é mais adequado para um contexto pode não o ser noutra. Desta forma, emerge uma tensão entre profundidade e extensão do propósito (*purpose-depth* e *purpose-scope*) que dita que, na melhor das hipóteses, metadados de alta qualidade são otimizados para o propósito específico para o qual foram desenvolvidos (*purpose-depth*), ao mesmo tempo que podem ser facilmente reutilizados para outras finalidades (*purpose-scope*).

Jinfang Niu (2016) também reflete sobre a utilização de LD em arquivos, enquanto fundamento da WS, nomeadamente no seu impacto na descoberta da informação arquivística e na descrição arquivística, destacando o seu potencial, embora lembrando que ainda nos encontramos numa fase inicial deste processo. Porém, a autora realça o potencial da LD para enriquecer as descrições arquivísticas, tornando-as mais interoperáveis e granulares, e possibilitando a resposta a questões de utilizadores mais complexas. O mecanismo de pesquisa com LD permite aos utilizadores formular pesquisas mais complexas não possíveis em sistemas de pesquisa tradicionais. A abertura torna as descrições arquivísticas (metadados) disponível para serem utilizadas de formas imprevistas, e a natureza interligada permite que as descrições sejam utilizadas de forma descentralizada pelas instituições e posteriormente agregadas via vínculos semânticos. Também possibilita a descrição de entidades que não sejam apenas documentos, potenciando a riqueza e a complexidade de relações que é possível estabelecer.

Uma das consequências apontadas por Niu da crescente adoção de LD em arquivos é a necessidade de maior granularidade na descrição dos componentes individuais dentro dos agregados, tornando cada componente passível de ser descoberto de forma individual. Em termos ilustrativos, a agregação documental assume a forma de um grafo RDF, e cada nó desse grafo é um elemento individualmente passível de descoberta. Além disso, refere que os *user-interfaces* devem ser fáceis de utilizar por qualquer utilizador, sem necessidade de este ter qualquer conhecimento de LD. De entre algumas das características de um bom interface, a autora menciona: permitir pesquisar múltiplas entidades, sugestão automática de termos de pesquisa, visualização de dados e múltiplas formas de navegação pelos dados.

Michael Jones (2018) parte da ideia de que o conhecimento se constrói na interconexão das partes para defender a necessidade de diferentes comunidades, neste caso arquivistas e profissionais de museu, cooperarem no desenvolvimento de coleções conjuntas através de um enfoque em descrições não só ao nível coletivo, como também ao nível do item. A intenção é a construção de redes estruturadas — mas não de forma hierárquica — que atuem como agregados de recursos interconectados. Jones entende que este esforço seria essencial para reforçar o aspeto de prova dos recursos, mas também de os valorizar e de estimular pesquisa futura. Assim, o autor identifica os seguintes requisitos para a descrição no século 21:

- i. o foco futuro das descrições deve ser na relacionalidade das redes enquanto aglomerados de coisas interligadas. As descrições devem ser não-hierárquicas, por oposição a uma abordagem centrada nos objetos/documentos/coleções/pessoas/organizações.
- ii. as descrições, que no artigo são referidas como *collection documentation*, devem ser vistas como um processo de gestão de conhecimento através do qual os recursos devem ser

- explicitamente relacionados com outros recursos, ao invés de carecer de conhecimento implícito do *staff* para navegar entre as coleções.
- iii. a natureza do conhecimento é complexa, inter-relacional e frequentemente multidisciplinar, pelo que defende a necessidade de integração das coleções de várias comunidades como forma de construção de *thick descriptions*, contextualizadas e culturalmente ricas.
  - iv. as descrições e os dados não devem ser considerados neutralmente históricos, mas historizados, uma vez que os observadores nunca são objetivos. O conhecimento emerge e é documentado num contexto socio-temporal que varia ao longo do tempo. O autor defende que deve ser possível aos utilizadores não só descobrir e aceder aos recursos, como também poderem citá-los, como forma de constituir prova.

A interconexão dos recursos é uma das principais características da *Web Semântica* (WS), que Luís Machado et al. (2019) procuram caracterizar, distinguindo entre *Semantic Web*, *Web of Data*, *Linked Data* e *Web of Linked Data*. Estes autores entendem a WS como uma construção ambiciosa que tem como difícil missão transformar a *web* numa base de dados global, permitindo não só a virtualmente ilimitada possibilidade de criação de relações entre os dados, como também a capacidade de estabelecer essas relações de forma automática, por agentes computacionais que capitalizam na interoperabilidade semântica. Além disso, notam uma evolução gradual da sociedade em direção à crescente operacionalização da WS.

Porém, é notada na literatura uma enorme dificuldade em implementar a interoperabilidade em contexto *web*. Gregory Rolan (2015) dá conta de alguns dos obstáculos encontrados que se traduzem num ainda prevaletente entendimento dos sistemas arquivísticos enquanto constructos monolíticos orientados para uma audiência muito específica, e que veicula uma narrativa privilegiada. Partindo da ideia de que os arquivos se têm tornado um multiverso fragmentado e contestado, formado a partir de uma mistura heterogénea de repositórios institucionais, empresariais e pessoais, o autor posiciona a interoperabilidade como um aspeto crítico para possibilitar o acesso aos recursos, bem como forma de expressão das vozes que subvertem a narrativa oficial. A WS emerge assim como um mecanismo que agrega sistemas distribuídos, heterogéneos e descentralizados, atuando como uma plataforma para colaboração e partilha que, no contexto arquivístico, se pode traduzir numa forma de superar divisões culturais e organizacionais, de realizar uma pesquisa e acesso integrados aos recursos arquivísticos, e de facilitar a emergência de diversas memórias narrativas a partir de uma pluralidade de perspetivas. O autor também critica das normas do ICA, de entre as quais a ISAD(G), por não incentivarem diretamente a criação de metadados que testem a pluralidade de contextos nos quais os documentos foram criados, sugerindo a inclusão de metadados que declarem explicitamente os contextos da origem dos documentos.

A crescente influência da WS na mundividência arquivística é ainda discutida por Mary Samouelian (2009), que já notava esta tendência há mais de uma década nos projetos de

digitalização desenvolvidos em arquivos. A autora caracteriza este espaço como um ambiente partilhado que convida à participação coletiva, e possibilita que previamente passivos recipientes de conteúdo possam agora contribuir para a criação, combinação e partilha de informação de novas formas imprevisíveis. De entre alguns dos mecanismos, Samouelian entende que os arquivos devem integrar nos seus *websites* ferramentas que estimulem a participação dos utilizadores, nomeadamente blogs, *wikis*, avaliações e revisões, *podcasts* e *bookmarks*.

Tomando outra abordagem, Gregory Rolan et al. (2019) discutem o uso da inteligência artificial (AI) em arquivos como um auxílio da gestão documental, uma vertente que não pode ser ignorada e que com toda a probabilidade terá um profundo impacto nas práticas arquivísticas. De entre as aplicações destacam-se apoio na avaliação, na classificação e na eliminação de documentos, não havendo ainda muita literatura nem casos de sucesso dado o carácter recente do tema.

#### **4.2.8 Sumário**

De forma a sistematizar os requisitos identificados em cada área científica e a facilitar a leitura, apresenta-se a Tabela 46 que consolida as informações encontradas. Esta tabela recapitula e agrega os requisitos extraídos únicos para cada tema, reunindo-se sob uma única entrada a repetição entre autores, dentro da mesma categoria. Porém, é de notar que ainda não se inclui uma componente crítica na forma como estes poderão ser traduzidos em metadados, que tomará lugar no próximo capítulo. Analisada em pormenor a literatura científica, passa-se na próxima secção a uma exploração da literatura técnica que consiste nas normas de metadados (de estrutura) seleccionadas previamente na metodologia.



Tabela 46 — Sumário dos requisitos identificados na literatura científica, organizados por tema.

Categoria	Requisitos
Representação da informação	<ul style="list-style-type: none"> <li>• o arquivista deve ser capaz de explicar as regras e os procedimentos arquivísticos de descrição dos recursos (Duff et al., 2013), sendo mais transparente (Serewicz, 2010).</li> <li>• a necessidade de incluir informação contextual, nomeadamente sobre o criador, a história e a proveniência dos documentos de arquivo (Duff et al., 2013).</li> <li>• os documentos estabelecem relações não só com membros do seu agregado mas com documentos de outras coleções e repositórios (Duff et al., 2013).</li> <li>• deve ser fomentado um bom relacionamento entre utilizadores e arquivistas de forma a facilitar a pesquisa (Duff et al., 2013).</li> <li>• a descrição arquivística deve realizar-se não só ao nível coletivo das coleções, mas também nível do item (J. Niu, 2015a). Alguns metadados possíveis são ‘título’, ‘sujeito’, ‘descrição’, ‘autor’, ‘criador’, ‘recipiente’, ‘contribuidor’, ‘data’, ‘lugar’ e ‘tipo de material’ (Zhang, 2012b).             <ul style="list-style-type: none"> <li>○ o controlo intelectual pode estender-se ao nível do conteúdo, mediante a descrição de <i>datasets</i>, figuras, tabelas, ou anotação de texto (J. Niu, 2015a; Zhang, 2012b).</li> <li>○ o meio digital permite que os recursos pertençam a mais do que uma coleção, e que uma coleção pertença a múltiplas comunidades (J. Niu, 2015a).</li> <li>○ os instrumentos de pesquisa podem ser independentes mas interligados num único agregador (Yeo, 2012). Devem ser incluídos metadados de relação que atestem a origem das descrições, tais como ‘contém’ ou ‘deriva de’. (J. Niu, 2015a), bem como a proveniência ao nível do ‘criador’, com uma caracterização na ‘história administrativa’ (Zhang, 2012b).</li> <li>○ os instrumentos de pesquisa têm de estabelecer links para outros recursos mediante metadados em <i>linked data</i> (Serewicz, 2010).</li> </ul> </li> <li>• os instrumentos de pesquisa devem ser fáceis de usar, sendo vantajoso que estejam construídos em estilo <i>website</i>, de forma a serem familiares aos utilizadores (J. Niu, 2015a).</li> </ul>

Categoria	Requisitos
	<ul style="list-style-type: none"> <li>• os metadados devem ser capturados e acumulados ao longo do ciclo de vida dos recursos, pelo criador ou por outras organizações (J. Niu, 2015a).</li> <li>• a representação deve ser <i>bottom-up</i>, construída a partir das necessidades dos utilizadores (Serewicz, 2010).</li> <li>• os sistemas de representação devem tornar-se mais flexíveis e transversais, como reflexo da complexidade das organizações (I. Silva, 2012).</li> <li>• os recursos devem ter um identificador único, legível por máquinas, de forma a facilitar o processamento automático dos metadados e de promover a descoberta e o acesso aos recursos (Verborgh et al., 2015).</li> <li>• as representações arquivísticas são práticas subjetivas, socialmente construídas, que podem ser questionadas. Trata-se também de um processo evolutivo no tempo, que deve ser revisitado e reinterpretado à luz de novos contextos (Yakel, 2003).</li> <li>• as descrições arquivísticas do passado devem ser mantidas como testemunho histórico e reflexo da evolução das narrativas de representação (Yakel, 2003).</li> <li>• devem ser identificados os autores e a versão de cada instrumento de pesquisa (Yakel, 2003).</li> <li>• é cada vez mais importante privilegiar a ordem intelectual do que a ordem física (Yakel, 2003; Zhang, 2012b).</li> </ul>
<p><b>Descrição arquivística</b></p>	<ul style="list-style-type: none"> <li>• com base na metodologia MPLP, é proposto que os processos de descrição devem ser simplificados aos metadados essenciais, de modo a viabilizar os custos de descrição de maiores volumes documentais (Anchor, 2013).</li> <li>• a descrição deve estar centrada nas necessidades do utilizador (Anchor, 2013).</li> <li>• a descrição deve ser mais honesta, reconhecendo os processos transformativos que influenciam a constituição dos arquivos ao longo do tempo. Devem-se incluir notas e/ou citações do próprio criador da descrição, indicar a sua autoria (Douglas, 2016) (Meehan, 2009) e até o seu CV (MacNeil, 2009).</li> <li>• as descrições devem ser mais transparentes, podendo ser acrescentadas notas de rodapé que testemunhem decisões dos arquivistas, bem como serem explicitadas as práticas de descrição (Douglas, 2016; MacNeil, 2009; Meehan, 2009).</li> <li>• deve ser dado um enfoque significativo à história custodial dos documentos. Sugestão de um campo autónomo de descrição, subdivisível em três elementos: 'nome(s) do(s) custodiante(s)', 'história custodial' e 'história da classificação' dos documentos de arquivo e de instrumentos</li> </ul>

Categoria	Requisitos
	<p>de pesquisa relacionados' (Douglas, 2016), nos quais seja incluído um registo detalhado da história dos documentos de arquivo até chegarem à instituição arquivística, incluindo 'história da conservação' (MacNeil, 2009).</p> <ul style="list-style-type: none"> <li>• possibilidade de criação de uma área própria para a história do criador com 'nome do criador' e 'história administrativa' e de uma área para a história arquivística com os campos 'nome do custodiante', 'história custodial', e 'história da classificação dos documentos de arquivo' e 'história da classificação dos instrumentos de pesquisa' (MacNeil, 2009).</li> <li>• as descrições arquivísticas são constructos subjetivos por arquivistas inseridos num determinado contexto (Douglas, 2016; MacNeil, 2009, 2012).</li> <li>• sugestão da possibilidade de utilizadores contribuírem para a descrição arquivística mediante comentários, <i>tags</i>, anotações ou revisões (Douglas, 2016).</li> <li>• a descrição arquivística deve incluir: <ul style="list-style-type: none"> <li>○ o nome do ficheiro, localização no sistema de armazenamento e especificações técnicas de hardware e software necessárias para a recuperação da informação (Hedstrom, 1993). De entre os metadados podem incluir-se 'formato de codificação' e 'material físico' (Llanes-Padrón, 2019). Devem também ser documentadas as mudanças físicas aos documentos (MacNeil, 2009).</li> <li>○ informação que ateste a relação entre o contexto e o conteúdo dos documentos (Hedstrom, 1993; Meehan, 2009).</li> <li>○ a documentação da destruição de documentos. Estas informações podem ser incluídas nos elementos 'história administrativa', 'avaliação', 'destruição e eliminação planeadas' (J. Niu, 2013).</li> <li>○ informação sobre a gestão dos documentos ao longo do seu ciclo de vida, incluindo <i>audit trails</i> (Hedstrom, 1993; J. Niu, 2013).</li> </ul> </li> <li>• a incapacidade de uma única estrutura hierárquica capturar de forma suficiente a complexidade das relações que os documentos estabelecem durante o seu uso, podendo os documentos serem interpretados à luz de vários contextos (Hedstrom, 1993; Llanes-Padrón, 2019; Meehan, 2009; Zhang, 2012c).</li> <li>• os recursos devem ser descritos de forma independente, ao nível do item, de modo a permitir a construção de vários agregados pelos utilizadores na <i>web</i> (Moyano Collado, 2013; Zhang, 2012c).</li> <li>• os metadados devem ser acrescentados de forma progressiva e gradual ao longo do ciclo de vida dos documentos (J. Niu, 2013).</li> </ul>

Categoria	Requisitos	
<b>Gestão da informação</b>	<ul style="list-style-type: none"> <li>• a representação deve ser realizada ao nível do item de modo a possibilitar a recontextualização dos recursos mediante as pesquisas individuais dos utilizadores (J. Bailey, 2013).</li> <li>• existem múltiplas interpretações dos documentos de arquivo, sendo o significado e o contexto variáveis no tempo (J. Bailey, 2013; Cumming, 2010; MacNeil, 2017).</li> <li>• a descrição arquivística deve ser um processo permanente, ao longo de todo o ciclo de vida dos documentos (Cumming, 2010).</li> <li>• as condições físicas dos materiais devem ser documentadas, como uma parte crítica da perceção do significado, do contexto e da autenticidade (Acker, 2017).</li> <li>• os desafios da gestão documental são próprios de cada organização, não se registando uma solução mágica, única e universal (McLeod, 2014).</li> <li>• os metadados devem refletir, em primeiro lugar, o contexto das atividades das quais os documentos de arquivo são um subproduto (McLeod, 2014; Tough, 2016; Wright, 2014).</li> <li>• o criador da representação deve ser identificado (Wright, 2014).</li> <li>• o objeto de representação em arquivo deve ser alargado (Yeo, 2011), permitindo que sejam criadas relações semânticas com outros recursos de outras coleções (Wright, 2014).</li> <li>• os arquivistas são construtores ativos das representações e dos significados que estas assumem (Ridener, 2009).</li> </ul>	
	<b>Metadados</b>	<ul style="list-style-type: none"> <li>• a identificação da posse dos metadados, bem como de mudanças efetuadas aos documentos ao longo de tempo como base para a confiança (Apostolou, 2009).</li> <li>• anotações, comentários e avaliações de utilizadores podem ser incluídos como reforço do valor do conteúdo e contexto (Apostolou, 2009).</li> <li>• o reconhecimento de que cada comunidade tem diferentes requisitos e necessidades, pelo que não existe um esquema universal de metadados, nem um entendimento comum do que possa ser considerado “essencial” (Apostolou, 2009; Gladney, 2007; Haynes, 2018).</li> <li>• a criação de metadados deve ser vista como um processo incremental, de responsabilidades partilhadas (Baca, 2016).</li> <li>• de forma a simplificar e a agilizar a gestão de metadados, a sua criação deve ser tão automática quanto possível (Baca, 2016).</li> </ul>

Categoria	Requisitos
	<ul style="list-style-type: none"> <li>• os metadados devem ser criados pelas organizações de forma que possam ser facilmente partilháveis e reutilizáveis (Baca, 2016; Gartner, 2016).</li> <li>• o uso de metadados facetados para flexibilização do acesso (Gartner, 2016) e como capacidade de satisfação do motivo da pesquisa mediante a simples consulta da lista de resultados (Beyene &amp; Godwin, 2018).</li> <li>• é destacada a relação entre a proveniência e a autoridade do criador, assente no conhecimento da sua identidade: <ul style="list-style-type: none"> <li>○ para explicitação da proveniência, são considerados importantes os elementos ‘Título’, ‘Identificador’, ‘Formato’, ‘Assunto’, ‘Direitos’, ‘Editor’ (Chen et al., 2011), bem como ‘Data de criação’ e ‘Descrição’ (Haynes, 2018).</li> <li>○ para a completude das descrições, é indicado que devem ser incluídos tantos elementos quanto possível, nomeadamente ‘Criador’, ‘Contribuidor’, ‘Formato’, ‘Editor’, ‘Fonte’ e ‘Cobertura’ (Chen et al., 2011), além de ‘entidade responsável’ e ‘frequência de atualização’ (Zeng &amp; Qin, 2016)</li> <li>○ para a avaliação da integridade dos recursos, é sugerida a inclusão de metadados como ‘fonte’, ‘relação’, ‘versão/edição’, ‘assinatura digital’ (Zeng &amp; Qin, 2016).</li> </ul> </li> <li>• devem ser criados metadados que possam servir como <i>audit trails</i> a fim de encorajar a demonstração da proveniência e da integridade dos documentos, reforçando a confiança nas instituições (Gladney, 2007; Li &amp; Sugimoto, 2017).</li> <li>• a proveniência pode também ser reforçada com metadados que respondam às seguintes questões: Quem criou os metadados? Quando foram criados ou modificados? Quais as circunstâncias na origem da criação dos metadados? (Haynes, 2018). As próprias adições, eliminações e substituições dos termos de metadados e das suas definições devem ser abarcadas por estas questões (Li &amp; Sugimoto, 2017).</li> <li>• devem ser incluídos metadados referentes a políticas de retenção/disposição dos documentos, tais como ‘Ação de eliminação’, ‘Período temporal da eliminação’, ‘Data prevista para a eliminação’, e ‘Eliminação autorizada por’ (Haynes, 2018).</li> <li>• cada recurso deve ter uma e só descrição, segundo o princípio 1:1 (Zeng &amp; Qin, 2016).</li> </ul>
<b>Normas de metadados</b>	<ul style="list-style-type: none"> <li>• o mapeamento de metadados entre normas não é linear, encontrando-se várias inconsistências e correspondências imperfeitas ou inexistentes (Andrade et al., 2014; Woodley, 2016).</li> </ul>

Categoria	Requisitos
	<ul style="list-style-type: none"> <li>• a distinção entre normas de descrição arquivística e normas de partilha de informação, sendo que as devem procurar garantir a autenticidade dos documentos ao longo do tempo e através dos vários contextos e domínios, ao passo que as últimas atuam como coletes de forças que governam a partilha de informação (Bunn, 2013).</li> <li>• os modelos bem sucedidos tendem a ser gratuitos e obedecem a uma redação clara que assegura a sua simples leitura e interpretação (Dryden, 2009).</li> <li>• o contexto de produção dos documentos deve ser capturado, e as descrições devem ser partilháveis e reutilizáveis entre várias comunidades presentes na <i>web</i> (Llanes-Padrón &amp; Moro-Cabero, 2017; Llanes-Padrón &amp; Pastor-Sánchez, 2017).</li> <li>• a criação de metadados é vista como um processo iterativo ao longo do ciclo de vida dos documentos. Como parte do controlo intelectual, devem ser incluídos metadados para a avaliação, o controlo, a preservação, a recuperação, o acesso e o uso, bem como o histórico de auditorias (Mckemmish et al., 1999).</li> <li>• a inclusão de demasiados elementos de metadados por uma norma contribui para dificuldades na sua implementação, gerando custos temporais e financeiros (Youn, 2015).</li> </ul>
Autenticidade	<ul style="list-style-type: none"> <li>• a inclusão de mecanismos para documentar a invariabilidade dos objetos digitais, fixando a sua integridade física, ao nível de metadados de proveniência. Algumas possibilidades são <i>checksums</i> ou verificação de tamanho de ficheiros (Bountouri et al., 2017). A integridade de um objeto pode também ser assegurada mediante uma autenticação criptográfica (Gladney, 2009; Rogers, 2015).</li> <li>• as informações sobre a proveniência, essenciais para documentar o ciclo de vida de um recurso, devem ser recolhidas desde a criação dos objetos (Bountouri et al., 2017).</li> <li>• a ideia de que a asserção da autenticidade e a confiança são atos subjetivos moldados por diversas interpretações (Duncan, 2009; Duranti et al., 2019; Engvall, 2019; McLeod &amp; Gormly, 2017; Rogers, 2015; Yeo, 2013). Um ponto de partida pode ser estabelecido através da localização do ponto de referência do “outro” mediante o qual se mede o grau de autenticidade (Duncan, 2009).</li> <li>• os critérios para a confiança nos documentos digitais são necessariamente distintos dos pressupostos que governam a presunção da confiança nos documentos em suportes analógicos. No ambiente digital, é necessária mais transparência e <i>accountability</i> nos processos de criação e de gestão dos documentos e dos metadados (Duranti et al., 2019; Engvall, 2019; Rogers, 2015; Yeo, 2013).</li> <li>• a explicitação da origem, da proveniência, da cadeia de custódia, a manutenção de registos e a realização de auditorias são apontados como aspetos indispensáveis para a transparência e para estabelecer uma base de confiança (Duranti et al., 2019; Gladney, 2009; Price &amp; Smith, 2011).</li> </ul>

Categoria	Requisitos
	<ul style="list-style-type: none"> <li>• a informação de proveniência deve estar associada ao documento, tornando os recursos auto-descritivos (Gladney, 2009). Alguns dos elementos de metadados nucleares de um conceito abstrato de um documento de arquivo, baseado numa perspectiva diplomática, podem ser: ‘action’, ‘bond’, ‘addressee’, ‘writer’, ‘author’, ‘adminContext’, ‘digitalObject’, ‘ingestDate’ (Jansen, 2015)</li> <li>• cada documento, ou até mesmo cada versão de cada documento, deve ter um identificador único (Gladney, 2009).</li> <li>• as coleções de várias instituições devem ser agregadas de forma a enriquecer o contexto e de melhor refletir a sociedade (Price &amp; Smith, 2011).</li> <li>• a criação de metadados deve ser automatizada tanto quanto possível de modo a poupar custos e a viabilizar o próprio procedimento (Yeo, 2013).</li> <li>• enquanto indicadores da autenticidade, devem ser indicadas informações sobre o <i>software</i> que gerou os documentos, informações sobre ações exercidas sobre os documentos, agendas de retenção e eliminação, registros de auditorias e técnicas criptográficas de validação (Rogers, 2015).</li> <li>• a viabilidade da tecnologia <i>blockchain</i> para atuar ao serviço da gestão documental enquanto baluarte da integridade e da identidade dos documentos, documentando de forma perene a proveniência (Bhatia &amp; Wright de Hernandez, 2019; Hofman et al., 2019; Lemieux, 2016).</li> </ul>
<i>Linked data</i>	<ul style="list-style-type: none"> <li>• a interoperabilidade é facilitada por uma maior granularidade das descrições (Gartner, 2015; Jones, 2018; J. Niu, 2016).</li> <li>• a importância do uso de identificadores únicos para definir entidades e as suas relações (Gartner, 2015).</li> <li>• alguns elementos metadados potencialmente úteis para agregadores de múltiplos repositórios são descrição da coleção (identificadores, títulos), descrições dos componentes (‘localização física’, ‘descrição material’, ‘descrição do conteúdo’), instituições detentoras, datas, relações, informações sobre direitos, impedimentos de uso, língua e assunto (Gartner, 2015).</li> <li>• a lição de que cada tipo modelo de representação (tabular, relacional, hierárquico, em grafo) apresenta vantagens e desvantagens inerentes, sendo que cada um tem legitimidade em ser implementado, dependendo do contexto (Hooland &amp; Verborgh, 2014). <ul style="list-style-type: none"> <li>○ o uso de RDF como um formato para a partilha dos metadados apresenta desvantagens, nomeadamente problemas derivados do elevado número de triplos necessários para a descrição, problemas ao nível do controlo de direitos intelectuais e de curadoria (Gartner, 2015).</li> </ul> </li> </ul>

Categoria	Requisitos
	<ul style="list-style-type: none"> <li>○ ainda assim, é vantajoso procurar flexibilizar as descrições, devendo procurar-se expressar os metadados de forma a que possam ser processados automaticamente por agentes computacionais, por exemplo através de <i>linked data</i>, facilitando a integração de várias coleções (Gartner, 2015; Hooland &amp; Verborgh, 2014; Jones, 2018; Machado et al., 2019; J. Niu, 2016).</li> <li>• o foco das descrições deve ser na relacionalidade das redes, não em hierarquias rígidas centradas em documentos, coleções ou organizações (Jones, 2018).</li> <li>• o conhecimento deve ser explicitado em metadados de forma facilitar a apreensão do significado por parte dos utilizadores, que não devem ter de procurar o <i>staff</i> dos arquivos para obter informações (Jones, 2018). Os utilizadores também devem ser capazes de construir as suas próprias narrativas e interpretações (Rolan, 2015; Samouelian, 2009).</li> <li>• as descrições enquanto produtos históricos de um determinado contexto socio-temporal (Jones, 2018; Rolan, 2015).</li> <li>• a sugestão de os arquivos integrarem ferramentas que convidem a participação dos utilizadores, nomeadamente blogs, <i>wikis</i>, avaliações e revisões, <i>podcasts</i> e <i>bookmarks</i> (Samouelian, 2009).</li> <li>• a inteligência artificial poderá vir a ser um aliado precioso na gestão da informação (Rolan et al., 2019).</li> </ul>

FONTE: elaboração própria.



## 4.3 Requisitos da literatura técnica

Ao longo das próximas páginas procede-se a uma caracterização das várias normas de metadados, previamente selecionadas na metodologia, que formam o *corpus* documental da literatura técnica.

Para cada norma, descreve-se de forma breve o seu contexto histórico, indicando aspetos como o criador, a origem, as datas das versões, o âmbito de aplicação, o propósito, as regras e os níveis de descrição. De seguida, apresenta-se a estrutura de elementos de metadados prevista por cada norma, de acordo com os vários campos de descrição, salientando os aspetos que os criadores determinaram como obrigatórios. Por fim, segue-se uma interpretação desses elementos de forma a extrair quais aqueles mais diretamente relacionados com a asserção da autenticidade, isto é, quais os elementos que informam sobre a autenticidade e sobre a integridade dos recursos e/ou das próprias descrições.

### 4.3.1 ISAD(G)

#### 4.3.1.1 Contextualização

A ISAD(G) é uma norma internacional que pretende orientar a descrição arquivística, podendo atuar como base para o desenvolvimento de normas nacionais. Criada pelo *International Council on Archives* (ICA) numa primeira versão em 1994, com uma segunda versão de 2000, a ISAD(G) foi a primeira norma internacional de descrição arquivística, representando assim um primeiro esforço de normalização das descrições em arquivos. Neste contexto, foi pensada de forma a facilitar a criação de descrições arquivísticas que *identify and explain the context and content of archival material in order to promote its accessibility* (International Council on Archives, 2000, p. 10). Embora esta norma também seja aplicada em arquivos tradicionalmente entendidos como históricos — i.e., recebendo a custódia de documentos quando o criador entende já não possuírem valor corrente — as regras da ISAD(G) destinam-se a qualquer fase do ciclo de vida dos documentos, nomeadamente deste o início ou até antes da criação dos documentos de arquivo, de forma a estabelecer o controlo intelectual dos documentos e a garantir uma descrição confiável, autêntica e acessível. Porém, as regras não fornecem instruções para a descrição de materiais especiais como selos, gravações de som, ou mapas. Nestes casos, devem ser consultados outros manuais já existentes.

A descrição arquivística na ISAD(G) baseia-se em quatro princípios teóricos:

- Descrição do geral para o específico.
- Informação relevante a cada nível de descrição.
- Ligação entre as descrições.

- Não-repetição da informação.

O nível de descrição mais abrangente adotado nesta norma é o fundo, cujas partes formam níveis subsequentes que apenas devem ser descritos se a sua descrição se justificar no contexto da descrição do fundo. Neste sentido, a ISAD(G) admite descrições ao nível do fundo, da série, do processo (*file*) e/ou do item, podendo ainda ocorrer níveis intermédios como subfundo ou subsérie. Assume-se que a informação dada a um nível mais abrangente é transferida para os níveis inferiores, devendo por isso a ligação orgânica entre os níveis ser clara. Apenas deve ser dada informação pertinente ao nível de descrição, sem repetições. Estes princípios encontram-se espelhados noutras normas de descrição arquivística compatíveis com a ISAD(G), tais como a DACS nos Estados Unidos, a RAD no Canadá, e a NOBRADE no Brasil.

#### 4.3.1.2 Estrutura de metadados

Existem 26 elementos de metadados identificados na ISAD(G). Para cada um, as regras explicitam que deve ser incluído (i) o nome do elemento, (ii) uma declaração do propósito do elemento para a descrição, (iii) as regras que se aplicam a esse elemento e (iv) exemplos ilustrativos, quando aplicável. Os elementos encontram-se distribuídos por sete áreas de descrição:

- **Identity Statement Area:** informação essencial para identificar a unidade de descrição;
- **Context Area:** informação sobre a origem e custódia da unidade de descrição;
- **Content and Structure Area:** informação sobre o conteúdo e a classificação (*arrangement*) da unidade de descrição;
- **Condition of Access and Use Area:** informação sobre a disponibilidade da unidade de descrição;
- **Allied Materials Area:** informação sobre os materiais pertinentes para a unidade de descrição;
- **Note Area:** informação especializada e informação que não pode ser incluída nas outras áreas;
- **Description Control Area:** informação sobre como, quando e quem preparou a descrição.

Todavia, de todos os 26 elementos usados, apenas seis são considerados essenciais para a partilha internacional de informação descritiva:

- reference code;*
- title;*
- creator;*
- date(s);*
- extent of unit description;* e
- level of description.*

De entre estes seis elementos considerados essenciais pelo ICA, é interessante notar que cinco derivam da área de identificação, com apenas o elemento *creator* a originar na área de contexto. Este aspeto parece salientar a importância que o ICA coloca na identificação dos recursos como um dos principais objetivos da descrição arquivística. A inclusão de mais elementos no processo descrição dependerá da natureza do recurso a ser descrito e do seu contexto. A Tabela 47 apresenta a lista completa de elementos de metadados presentes na ISAD(G), estando os elementos obrigatórios assinalados a negrito. A designação da função é integralmente retirada do texto da norma, abreviando-se *unit of description* para UD.

Tabela 47 — Elementos de metadados na ISAD(G).

Área	Elemento	Função
1. Identity Area	1.1 <b>Reference code(s)</b>	To identify uniquely the unit of description (UD) and to provide a link to the description it represents.
	1.2 <b>Title</b>	To name the UD.
	1.3 <b>Date(s)</b>	To identify and record the date(s) of the UD.
	1.4 <b>Level of description</b>	To identify the level of arrangement of the UD.
	1.5 <b>Extent and medium of the UD (quantity, bulk, or size)</b>	To identify and describe a) the physical or logical extent and b) the medium of the UD.
2. Context Area	2.1 <b>Name of creator(s)</b>	To identify the creator (or creators) of the unit of description.
	2.2 Administrative / Biographical history	To provide an administrative history of, or biographical details on, the creator (or creators) of the UD to place the material in context and make it better understood.
	2.3 Archival history	To provide information on the history of the UD that is significant for its authenticity, integrity and interpretation.
	2.4 Immediate source of acquisition or transfer	To identify the immediate source of acquisition or transfer.
3. Content and Structure Area	3.1 Scope and content	To enable users to judge the potential relevance of the UD.
	3.2 Appraisal, destruction and scheduling information	To provide information on any appraisal, destruction and scheduling action.
	3.3 Accruals	To inform the user of foreseen additions to the UD.
	3.4 System of arrangement	To provide information on the internal structure, the order and/or the system of classification of the UD.

Área	Elemento	Função
4. Conditions of Access and Use Area	4.1 Conditions governing access	To provide information on the legal status or other regulations that restrict or affect access to the UD.
	4.2 Conditions governing reproduction	To identify any restrictions on reproduction of the UD.
	4.3 Language/scripts of material	To identify the language(s), script(s) and symbol systems employed in the UD.
	4.4 Physical characteristics and technical requirements	To provide information about any important physical characteristics or technical requirements that affect use of the UD.
	4.5 Finding aids	To identify any finding aids to the UD.
5. Allied Materials Area	5.1 Existence and location of originals	To indicate the existence, location, availability and/or destruction of originals where the UD consists of copies.
	5.2 Existence and location of copies	To indicate the existence, location and availability of copies of the UD.
	5.3 Related units of description	To identify related units of description.
	5.4 Publication note	To identify any publications that are about or are based on the use, study, or analysis of the UD.
6. Notes Area	6.1 Note	To provide information that cannot be accommodated in any of the other areas.
7. Description Control Area	7.1 Archivist's note	To explain how the description was prepared and by whom.
	7.2 Rules or conventions	To identify the protocols on which the description is based.
	7.3 Date(s) of descriptions	To indicate when this description was prepared and/or revised.

FONTE: International Council on Archives (2000).

#### 4.3.1.3 Metadados para a autenticidade

Conforme visto anteriormente, a autenticidade abarca a demonstração da identidade e a verificação da integridade, pelo que nesta secção se procura identificar os elementos de metadados que potencialmente satisfazem estes requisitos.

Embora a norma apresente “apenas” seis elementos como obrigatórios, entende-se que se podem eleger nove elementos que permitem estabelecer a identidade de um agregado documental: *reference code, dates, extent and medium, level of description, system of arrangement, scope and content, title, name of creator* e *administrative/biographical history*. O código de referência, as datas, a extensão e o suporte ajudam a definir a dimensão física do conjunto documental. O nível de descrição, o sistema de classificação e o âmbito e conteúdo auxiliam na descrição da estrutura

do conjunto e na compreensão das atividades das quais os documentos resultam. Por fim, o título, o nome do criador e a história administrativa/biográfica contribuem para estabelecer a proveniência dos documentos.

A respeito da integridade, identificam-se na ISAD(G) oito elementos que ajudam a documentar as alterações realizadas ao conjunto documental ao longo do tempo: *existence and location of originals, existence and location of copies, accruals, immediate source of acquisition, physical characteristics and technical requirements, related units of description, archival history e appraisal, scheduling and destruction information*. A localização dos originais e de cópias pode não só apontar para quem detém a custódia, como também informar sobre documentos descritos que já foram eliminados. A indicação de unidades de descrição relacionadas pode apontar para outros documentos que pertençam ao mesmo agregado, mas que se encontrem espalhados por outras organizações, coleções ou repositórios. A história arquivística é fundamental como uma oportunidade para fornecer informações sobre transferências de custódia e sobre a história de classificações adotadas, documentando assim aspetos relacionados com a transmissão e com a sua representação. Estes campos podem ser complementados pelas fontes de aquisição e de adições ao fundo. A indicação das características técnicas é também pertinente para documentação da integridade dos documentos no momento da descrição. Por último, informação sobre destruição e avaliação ajuda a explicitar intervenções dos arquivistas que tenham alterado a composição do agregado documental.

Além de elementos de metadados sobre a autenticidade dos documentos, observa-se também a indicação de elementos de metadados que podem contribuir para a asserção da autenticidade da própria descrição arquivística. Estes três elementos encontram-se na área de controlo da descrição: *archivist's notes, rules or conventions e date of descriptions*. As notas do arquivista podem servir para indicar decisões do arquivista, contribuindo para uma maior transparência e responsabilização. A explicitação das regras seguidas pode ajudar os utilizadores a compreender as decisões tomadas, além de elucidar sobre potenciais sistemas de representação adotados e a justificação das opções adotadas. A data das descrições é importante para contextualizar a representação no tempo.

Assim, é de notar que, de entre os 16 elementos identificados como potenciais contributos para a asserção da autenticidade, seis desses se encontram incluídos como obrigatórios. Por outras palavras, todos os elementos considerados obrigatórios pela ISAD(G) estão diretamente relacionados com a autenticidade. Porém, parece claro que a norma foi concebida para documentos analógicos, dada a prevalência às dimensões físicas dos objetos, por contraste com a ausência de mecanismos de controlo da autenticidade dos documentos digitais.

### **4.3.2 AGRkMS**

#### **4.3.2.1 Contexto histórico**

A *Australian Government Recordkeeping Metadata Standard (AGRkMS)* é uma norma desenvolvida pelo *National Archives of Australia (NAA)* com o objetivo de definir um conjunto de

metadados passíveis de aplicação em sistemas eletrônicos de negócio e de gestão documental das agências do governo australiano. Embora formulada para o contexto digital, alguns dos conceitos podem ser aplicados em contextos analógicos tradicionais ou em contextos de gestão documental híbrida.

A primeira versão 1.0 foi publicada em 1999 como fruto da necessidade de um quadro de referência para a especificação, mapeamento e normalização de metadados de gestão documental. O NAA participou no *SPIRIT Recordkeeping Metadata Project*, liderado pela Universidade de Monash, que desenvolveu a primeira abordagem de entidades múltiplas, tendo publicado os resultados em 2000. Desde então, o NAA procurou desenvolver o modelo de entidades, que versões sucessivas tentam incorporar. Assim, foi lançada em 2008 a versão 2.0 que efetua uma revisão completa da primeira versão, seguida de uma versão 2.1 em 2011 que atualiza algumas das normas relacionadas, sendo a versão mais recente a 2.2, de 2015, que encetou uma revisão mais a fundo das propriedades dos metadados.

O modelo de entidades múltiplas usado na atual versão 2.2 é derivado da *Australian Standard Information and Documentation – Records Management Processes – Metadata for Records* (AS ISO 23081). Esta família de normas, em conjunto com a AGRkMS, procura fomentar a convergência de abordagens para o controlo de metadados de gestão documental do governo australiano, entendidos como *structured or semi-structured information that enables the creation, management and use of records through time and across domains* (National Archives of Australia, 2015, p. 8). A norma pretende assim descrever os documentos e os contextos nos quais estes são capturados e usados, através da sua identificação e autenticação das entidades intervenientes (agentes, negócios e mandatos) e das relações entre os documentos de arquivo. A AGRkMS foi concebida para ser utilizada como um instrumento de referência por organizações, gestores e documentos e desenvolvedores de *software* que procurem implementar sistemas eletrônicos de gestão documental, tendo por base contextos administrativos. Constitui uma fundação básica que as organizações podem expandir de acordo com os seus critérios específicos.

De entre as suas principais características, destacam-se:

- Flexibilidade: as agências têm liberdade para definir o número de entidades, além de que os requisitos mínimos de metadados indicados pela norma fornecem um guia prático que pode ser adaptado às necessidades específicas.
- Repetibilidade: as propriedades e subpropriedades podem ser aplicadas a múltiplas instâncias das entidades.
- Extensibilidade: as agências podem acrescentar novas propriedades ao conjunto inicial de metadados.
- Reuso: alguns dos elementos definidos para fins de gestão documental podem ser reutilizados para outros propósitos, diminuindo assim a necessidade de criação de metadados.
- Interdependência: algumas propriedades e subpropriedades encontram-se relacionadas de modo que mudanças numa impliquem alterações noutra.

Relativamente aos níveis de descrição, a norma assenta num modelo de descrição de cinco entidades: *Record*, *Agent*, *Business*, *Mandate* e *Relationship* que podem ser descritas por 27 elementos de metadados, definidos enquanto propriedades, e 44 subpropriedades. As implementações podem incluir um número variável de entidades, nomeadamente uma (implementação de entidade única), duas (*Record* e *Agent*), três (*Record*, *Agent* e *Business* ou *Reccord*, *Agent* e *Relationship*) ou a totalidade das cinco. Na implementação de uma entidade única, a abordagem mais simples, todas as entidades que não são documentos são descritas como se o fossem. Numa abordagem de entidades múltiplas, cada entidade é descrita de forma independente e detalhada.

#### 4.3.2.2 Estrutura de metadados

Ao contrário da ISAD(G), nem os 27 elementos de metadados nem as potenciais 44 subpropriedades se encontram agrupados por zonas de descrição. A norma limita-se a apresentar e a caracterizar os metadados, que por sua vez podem ser usados para descrever as entidades, consoante as escolhas da organização responsável pela implementação. Cada propriedade é acompanhada de uma indicação de: (i) uma definição; (ii) propósito; (iii) continente (*container*); (iv) obrigatoriedade; (v) condições de uso; (vi) aplicabilidade; (vii) repetibilidade; (viii) valor padrão; (ix) esquema e (x) comentários. São indicados como obrigatórias sete propriedades:

- a. *Category*;
- b. *Identifier*;
- c. *Name*;
- d. *Date Range*;
- e. *Related Entity*;
- f. *Disposal*;
- g. *Extent*;

É de notar que as primeiras quatro propriedades se aplicam a todas as entidades, ao passo que a propriedade *Related Entity* diz apenas respeito à entidade *Relationships*. As duas últimas propriedades devem ser aplicadas na descrição dos documentos de arquivo. A Tabela 48 apresenta a lista detalhada de propriedades e de subpropriedades sugerida pela AGRkMS, com os elementos obrigatórios assinalados a negrito. A obrigatoriedade das subpropriedades é também indicada na terceira coluna, consoante os metadados forem *mandatory* (M), *conditional* (C) ou *optional* (O). A aplicação de propriedades condicionais depende de circunstâncias específicas da organização, ao passo que as propriedades opcionais reforçam as descrições, mas podem não se aplicar em todos os contextos.

Tabela 48 — Elementos de metadados da AGRkMS.

<b>Propriedade</b>	<b>Obrigatoriedade</b>	<b>Entidade a que se aplica</b>	<b>Subpropriedades obrigatórias</b>	<b>e</b>
0. Entity Type	C  M for multiple entity implementations	Todas	-	
<b>1. Category</b>	M	Todas	-	
<b>2. Identifier</b>	M	Todas	2.1 Identifier String [M] 2.2 Identifier Scheme [C]	
<b>3. Name</b>	M	Todas	3.1 Name Words [M] 3.2 Name Scheme [C]	
<b>4. Date Range</b>	M	Todas	4.1 Start Date [M] 4.2 End Date [C]	
5. Description	O	Todas	-	
<b>6. Related Entity</b>	M	Relationship	6.1 Assigned Entity ID [M] 6.2 Assigned Entity ID Scheme [C] 6.3 Relationship Role [M]	
<b>7. Change History</b>	C	Relationship	7.1 Property Name [M] 7.2 Prior Value [M] 7.3 Relationship ID [C]	
<b>8. Jurisdiction</b>	O	Record, Agent, Business, Mandate	-	
<b>9. Security Classification</b>	C for Record, Mandate  O for Business	Record, Business, Mandate	-	
<b>10. Security Caveat</b>	C	Record, Mandate	10.1 Caveat Text [M] 10.2 Caveat Category [C]	
<b>11. Permissions</b>	C	Agent, Business	11.1 Permission Text [M] 11.2 Permission Type [M]	
<b>12. Rights</b>	C	Record	12.1 Rights Statement [M] 12.2 Rights Type [M] 12.3 Rights Status [C]	



<b>Propriedade</b>	<b>Obrigatoriedade</b>	<b>Entidade a que se aplica</b>	<b>Subpropriedades obrigatoriedade</b>	<b>e</b>
13. Contact	C	Agent	13.1 Contact Details [M] 13.2 Contact Type [M]	
14. Position	O	Agent	-	
15. Language	C for Record O for Agent	Record, Agent	-	
16. Coverage	O	Record, Mandate	16.1 Jurisdictional Coverage [O] 16.2 Temporal Coverage [O] 16.3 Spatial Coverage [O]	
17. Keyword	C	Record	17.1 Keyword Term [M] 17.2 Keyword ID [O] 17.3 Keyword Scheme [C] 17.4 Keyword Scheme Type [C]	
18. Disposal	M	Record	18.1 Records Authority [M] 18.2 Disposal Class ID [C] 18.3 Disposal Action [C] 18.4 Disposal Trigger Date [C] 18.5 Disposal Action Due [C]	
19. Format	C	Record	19.1 Format Name [C] 19.2 Format Version [C] 19.3 Creating Application Name [C] 19.4 Creating Application Version [C] 19.5 Format Registry [C]	
<b>20. Extent</b>	M	Record	20.1 Physical Dimensions [C] 20.2 Logical Size [C] 20.3 Quantity [C] 20.4 Units [M]	
21. Medium	C	Record	-	
22. Integrity Check	C	Record	22.1 Hash Function Name [M] 22.2 Message Digest [M]	

Propriedade	Obrigatoriedade	Entidade a que se aplica	Subpropriedades obrigatoriedade	e
23. Location	O	Record	-	
24. Document Form	O	Record	-	
25. Precedence	O	Record	-	
26. Dissemination Limiting Markers	C for Mandate O for Business	Record, Mandate	Record, Business,	-

FONTE: National Archives of Australia (2015).

#### 4.3.2.3 Metadados para a autenticidade

Não obstante os sete elementos obrigatórios, identificam-se 12 elementos que podem auxiliar na identificação da identidade dos recursos e dos criadores (no modelo, entendidos como agentes): *name, identifier, date range, extent, medium, document form, category, description, coverage, related entity, change history* e *name identifier*. O nome, o identificador, as datas, o suporte e a dimensão, bem como a forma documental, possibilitam uma descrição da dimensão física dos documentos de arquivo. A categoria, a descrição e a cobertura dizem respeito à natureza dos documentos e explicitam as atividades das quais estes resultam. As entidades relacionadas aplicam-se apenas a outras propriedades, sendo fundamentais para estabelecer o contexto no qual as atividades ocorrem. A história das modificações introduz informações sobre a proveniência, que podem ser complementadas pelo uso do nome e de um identificador dos próprios criadores.

A nível da integridade, destacam-se três elementos que podem auxiliar na verificação de alterações sofridas pelos documentos ao longo do tempo: *disposal, location* e *integrity check*. O conhecimento da localização dos documentos pode informar sobre a custódia ou posse, ao passo que a inclusão de informação sobre a eliminação se torna fundamental enquanto testemunho das alterações sofridas pelo conjunto documental. Possuir mecanismos de verificação da integridade é também uma propriedade útil como reforço forma da facilitação de deteção de alterações indevidas aos documentos que pudessem condicionar o seu significado.

Embora a AGRkMS não possua uma área de identificação de elementos para a própria descrição arquivística, destaca-se o elemento *contact*, que se aplica unicamente a agentes, que pode servir para promover uma via de contacto com os criadores dos documentos ou das próprias descrições. Este aspeto talvez seja importante como forma de promoção da transparência. Como complemento, talvez possam ser reutilizados os elementos *name, identifier* e *date range*, aplicando-os desta vez aos próprios agentes, como forma de contribuir para a sua identificação e para documentar as datas nas quais a descrição foi realizada.

### 4.3.3 EAD

#### 4.3.3.1 Contexto histórico

A *Encoded Archival Description* (EAD) é uma norma internacional de metadados de transmissão para a descrição hierárquica de documentos de arquivo. Tem vindo a ser desenvolvida pelo *EAD Working Group of the Society of American Archivists* (SAA), tratando-se de um formato XML usado por arquivistas de todo o mundo.

A primeira versão foi publicada em 1998, seguindo-se da versão EAD2002, publicada enquanto um *Document Type Definition* (DTD) em 2002 e enquanto um esquema da W3C em 2007. Estas revisões tiveram por objetivo fomentar um maior alinhamento internacional com outras normas, possibilitando a criação de instrumentos de pesquisa eletrónicos regidos por uma estrutura de dados compatível com a ISAD(G). A terceira e mais atual versão é a EAD3, lançada em 2015 e atualizada para a versão EAD3 1.1.1 em 2019. A revisão para a terceira edição foi iniciada em 2010 como fruto da constatação da necessidade de atualizar a EAD para um mundo em que a descrição arquivística tinha sofrido intensas mudanças nos últimos 10 anos, destacando-se a emergência de *linked data* como uma possibilidade de criação de descrições arquivísticas semanticamente ricas na *web*. A EAD era vista pelos próprios criadores como *too complex, too forgiving, and too flexible for its own good* (Society of American Archivists, 2019a, p. 5).

Os quatro princípios orientadores das mudanças realizadas na atualização para a EAD3 são elucidados no prefácio da *EAD3 Tag Library*:

- *Achieving greater conceptual and semantic consistency in the use of EAD.*
- *Exploring mechanisms whereby EAD-encoded information might more seamlessly and effectively connect with, exchange, or incorporate data maintained according to other protocols.*
- *Improving the functionality of EAD for representing descriptive information created in international and particularly in multilingual environments.*
- *Being mindful that a new version will affect current users* (Society of American Archivists, 2019a, p. 6).

Estas alterações tiveram como objetivo principal tornar a EAD mais consistente e alinhada com outras normas internacionais de modo a viabilizar o seu uso e a facilitar a sua implementação pelos utilizadores. É de destacar que durante o processo de revisão, o comité técnico rejeitou sugestões de redução do número de elementos, decidindo que a consistência e a claridade semântica seriam melhores manifestações de simplicidade do que o número de elementos no esquema. Além disso, devido à sua estrutura em XML, a informação descritiva é herdada por cada subcampo da hierarquia.

#### 4.3.3.2 Estrutura de metadados

A estrutura da EAD é composta por 165 elementos de metadados e 85 atributos. Os atributos encontram-se associados aos elementos e refletem propriedades destes, podendo assumir diferentes valores em função do contexto em que surgem. Por este motivo, a análise incidirá unicamente nos elementos de metadados elegíveis para a descrição arquivística dos documentos.

Ao contrário da ISAD(G) e da AGRkMS, não são indicados elementos obrigatórios. Na EAD, os elementos apenas são exigidos no contexto de ramos filhos de um atributo de um nível superior na hierarquia. Depreende-se que compete a cada organização escolher os metadados que considera pertinentes para a descrição arquivística. O único elemento obrigatório é a raiz <ead>, que atua como um contentor de todos os elementos contidos no instrumento de pesquisa. O elemento <ead> deve conter <control> seguido de <archdesc>. Os elementos a incluir na descrição arquivística devem ser aqueles que as organizações consideram adequados para a descrição do conteúdo, contexto e das dimensões físicas dos documentos, bem como fornecer informações administrativas que facilitem o uso dos recursos.

Relativamente à estrutura, é interessante notar que a SAA mantém no documento da revisão mais recente da EAD um apêndice com os elementos e os atributos que se tornaram obsoletos, ou foram descontinuados. Em outro apêndice inclui ainda uma história da revisão, na qual explicita as mudanças efetuadas aos elementos em cada versão. Conforme defendido durante o estudo dos requisitos científicos, estes dois esforços são muito importantes para a transparência e para a responsabilização, considerando-se positiva a sua inclusão.

A presente estrutura de 165 elementos proposta pela EAD é reproduzida no Quadro 3. É de notar que, enquanto formato XML, os elementos não se encontram organizados em secções de descrição. Antes, consistem simplesmente numa lista não-estruturada de metadados. Cada elemento é definido por sete subcampos: *summary*, *may contain*, *may occur within*, *attributes*, *description and usage*, *availability* e *example*. Por motivos de simplificação, não se reproduzem aqui estes subcampos.

Quadro 3 — Elementos contemplados pela EAD3 1.1.1.

abbr abstract accessrestrict accruals acqinfo address addressline agencycode agencyname agent agenttype altformavail appraisal archdesc archref arrangement author bibliography bibref bioghist blockquote c c01 c02 c03 c04 c05 c06 c07 c08 c09 c10 c11 c12
chronitem chronitemset chronlist citation colspec container control controlaccess controlnote conventiondeclaration corpname custodhist dao daoset date daterange dateset datesingle defitem descriptivenote did didnote dimensions dsc
ead edition editionstmt emph entry event eventdatetime eventdescription eventtype expan famname filedesc fileplan footnote foreign fromdate function genreform geogname geographiccoordinates head head01 head02 head03
index indexentry item label langmaterial language languagedeclaration languageset lb legalstatus list listhead localcontrol localtypedeclaration maintenanceagency maintenanceevent maintenancehistory maintenancestatus materialspec name namegrp notestmt num objectxmlwrap occupation
odd originalsloc origination otheragencycode otherfindaid otherrecordid p

part persname physdesc physdescset physdescstructured physfacet physloc phystech prefercite processinfo ptr ptrgrp publicationstatus publicationstmt publisher quantity quote recordid ref relatedmaterial relation relationentry relations repository representation rightsdeclaration row scopecontent script separatedmaterial seriesstmt source sourceentry sources sponsor subject  subtitle table tbody term tgroup thead title titleproper titlestmt todate unitdate unitdatestructured unitid unittitle unittype userrestrict
---

FONTE: Adaptado da Society of American Archivists (2019a).

#### 4.3.3.3 *Metadados para a autenticidade*

De entre os elementos possíveis, destacam-se 29 elementos relacionados com a autenticidade dos recursos e das descrições. De entre estes, identificam-se 13 elementos que podem mais diretamente contribuir para a explicitação da identidade: <agencycode>, <recordid>, <unitid>, <unitdate>, <unitdatestructured>, <physdesc>, <physdescstructured>, <archdesc>, <c>, <scopecontent>, <unittitle>, <origination> e <bioghist>. O identificador do documento, a data e a descrição física são elementos que caracterizam o conjunto de documentos enquanto um agregado físico. Os elementos da descrição arquivística, incluindo o subcampo <c> que atua como um contentor lógico para outros elementos que explicitem o contexto e o conteúdo, bem como a classificação e o âmbito e conteúdo, cumprem a função de descrever a estrutura da agregação, de manifestar as relações entre os documentos, e de caracterizar as atividades das quais os documentos são reflexo. Adicionalmente, a descrição da entidade originária, o título e a história biográfica auxiliam na fixação da proveniência dos recursos.

No que diz respeito à integridade, identificam-se nove elementos que podem contribuir para a demonstração de alterações sofridas pelos documentos e pelo agregado ao longo do tempo: <custodhist>, <acqinfo>, <accruals>, <appraisal>, <phystech>, <originalsloc>, <altformavail>, <relatedmaterial>, <separatedmaterial>. A indicação do local onde se encontram os originais e as cópias, a par das suas características físicas e eventuais requisitos para reprodução, é útil para esclarecer sobre quem detém a custódia dos documentos, além de constituir um possível vestígio de documentos eliminados. A história custodial e a informação de aquisição é importante para documentar a história da transferência de custódia dos documentos, clarificando sobre a sua transmissão. Informação sobre avaliação contém elementos que indiquem a destruição de documentos, sendo por isso um aspeto crítico para o estabelecimento da integridade. Os elementos sobre os materiais relacionados podem fornecer dados sobre relações com outros documentos que pertençam ao agregado, contribuindo para a integridade do conjunto documental.

Além dos elementos da autenticidade dos recursos, identificam-se sete elementos adicionais referentes à autenticidade das próprias descrições: <processinfo>, <conventiondeclaration>, <maintenanceevent> <eventdatetime>, <maintenanceagency>, <maintenancehistory> e <maintenancestatus>. A agência indica o criador da descrição, enquanto a história e o estado atual servem como um registo das alterações efetuadas e do grau de completude da versão atual. A EAD prevê também a indicação da data da descrição e de notas do arquivista, além da explicitação das regras ou convenções adotadas no processo de representação. Em

conjunto, consistem em elementos-chave para a documentação da história da representação, contribuindo assim para a transparência e responsabilização dos arquivistas.

#### 4.3.4 e-EMGDE

##### 4.3.4.1 Contexto histórico

O *Esquema de Metadatos para la Gestión del Documento Electrónico* (e-EMGDE) atua como documentação complementar à *Norma Técnica de Interoperabilidad de Política de gestión de documentos electrónicos*. Foi elaborado pela *Dirección de Tecnologías de la Información y las Comunicaciones* (DTIC), uma entidade sob a administração do governo espanhol, com a finalidade de gerir a documentação de arquivo digital criada no decurso das atividades governamentais, contribuindo para a autenticidade, fiabilidade, integridade, disponibilidade e interoperabilidade dos documentos ao longo do tempo. A primeira versão foi publicada em 2012, sendo mais recente a versão 2.0, de 2016.

Trata-se de um modelo entidade-relação, tal como o AGRkMS. O e-EMGDE estabelece o modelo conceptual no qual se apoia a criação de metadados, na medida em que define as suas propriedades. Também fornece orientações sobre os valores que o conteúdo dos elementos deve assumir, e referencia normas relacionadas usadas como fonte. Baseia-se na norma AGRkMS v.2.2, adaptando os elementos de metadados à realidade jurídica espanhola. Esta proximidade parece ser coerente, uma vez que ambas as normas têm por missão primária o controlo intelectual dos metadados usados para a descrição de documento digitais criados e mantidos no decurso das atividades de cada um dos respetivos governos. Conforme analisado anteriormente, a primeira versão da AGRkMS data de 1999, pelo que se assume como uma influência natural para o desenvolvimento da e-EMGDE.

Assim, não é surpreendente constatar que, à semelhança da sua predecessora, a e-EMGDE adote um modelo de cinco entidades: *Documento*, *Agente*, *Actividad*, *Regulación* e *Relación*. Cada entidade é sucessivamente decomposta em categorias de descrição. Por exemplo, ao nível do documento as categorias são *Grupo de fondos*, *Fondo*, *Serie*, *Agregación*, *Expediente* e *Documento Simple*. A adoção de um modelo de múltiplas entidades pretende reconhecer a importância de representar diversas relações que enriquecem o contexto dos documentos de arquivo, contribuindo para a perceção do seu valor. Assim, a norma reconhece que os documentos não são mais do que uma de múltiplas entidades que é necessário contemplar na descrição arquivística, uma vez que *un documento sin contexto no es un documento en el que asegure su valor probatorio o su fiabilidad como evidencia de las actividades y procedimientos* (Espanha, 2016, p. 13). Porém, contempla que a descrição arquivística possa ser feita unicamente usando a entidade documento, consoante as práticas de cada organização.

Destacam-se as seguintes características deste modelo:

- Implementação modular: uma organização pode decidir o número de entidades a adotar;

- Independência da tecnologia: não são especificados os procedimentos específicos de implementação tecnológica dos metadados, pedindo apenas que os valores sejam preenchidos de forma automática sempre que possível;
- Flexibilidade: permite adequação a diferentes contextos, desde estruturas hierárquicas a descrições lineares;
- Repetibilidade: os elementos podem ser usados mais do que uma vez para descrever a mesma entidade;
- Extensibilidade: os metadados podem ser complementados com outras implementações específicas;
- Compatibilidade: adequação ao uso paralelo com outras normas nacionais e internacionais de descrição arquivística;
- Reutilização: os elementos de metadados podem ser reaproveitados para outras finalidades.

#### 4.3.4.2 Estrutura de metadados

A estrutura de metadados da e-EMGDE é composta por 30 elementos, dos quais 16 são obrigatórios, 7 condicionais e 7 opcionais. O uso de elementos condicionais depende do tipo da entidade a ser descrita e do seu contexto. Os opcionais podem ser usados em situações que se considere necessário fornecer uma descrição mais detalhada.

Cada elemento é definido por 13 subcampos: (i) *nombre formal*, (ii) *sub-elemento de*, (iii) *definición*, (iv) *aplicabilidad*, (v) *obligación*, (vi) *automatizable*, (vii) *repetible*; (viii) *sub-elementos*, (ix) *valores*, (x) *compatibilidad*, (xi) *finalidad*, (xii) *comentarios* e (xiii) *ejemplos*. Os 16 elementos obrigatórios são:

- |                                      |                                    |
|--------------------------------------|------------------------------------|
| 1) <i>Categoría</i> ;                | 9) <i>Tipo documental</i> ;        |
| 2) <i>Identificador</i> ;            | 10) <i>Estado de elaboración</i> ; |
| 3) <i>Nombre</i> ;                   | 11) <i>Clasificación</i> ;         |
| 4) <i>Fechas</i> ;                   | 12) <i>Versión NTI</i> ;           |
| 5) <i>Entidad relacionada</i> ;      | 13) <i>Órgano</i> ;                |
| 6) <i>Calificación</i> ;             | 14) <i>Origen del documento</i> ;  |
| 7) <i>Características técnicas</i> ; | 15) <i>Estado del expediente</i> ; |
| 8) <i>Firma</i> ;                    | 16) <i>Interesado</i> .            |

Os elementos podem ser subdivididos em 56 subelementos, que por sua vez são extensíveis em outros 29 subsubelementos. Existem ainda dois sub-sub-subelementos. Independentemente do grau de obrigatoriedade de cada elemento, os subelementos apenas podem ser aplicados quando o elemento que lhes é hierarquicamente superior é utilizado. Similarmente, a norma define que cada elemento de metadados não deve apresentar o conteúdo por si só, devendo ser decomposto nos subelementos hierarquicamente inferiores que o compõem. A Tabela 49 reproduz os elementos de metadados propostos por esta norma, sendo indicada a obrigatoriedade do uso, as entidades a que

se aplicam e as suas subpropriedades, a par do grau de obrigatoriedade de cada subpropriedade. Os elementos obrigatórios são também assinalados a negrito. De forma a simplificar a leitura, apenas se incluem os elementos e os subelementos. De forma a manter consistência com as abreviaturas usadas para a AGRkMS, indicam-se as letras ‘M’, ‘C’ e ‘O’ para elementos obrigatórios, condicionais e opcionais, respetivamente.

Tabela 49 — Elementos de metadados da e-EMGE.

<b>Elemento</b>	<b>Obrigatoriedade</b>	<b>Entidade a que se aplica</b>	<b>Subpropriedades obrigatoriedade</b>	<b>e</b>
0. Tipo de Entidad	C	Documento, Actividad, Relación	Agente, Regulación,	-
<b>1. Categoría</b>	M	Documento, Actividad, Relación	Agente, Regulación,	-
<b>2. Identificador</b>	M	Documento, Actividad, Relación	Agente, Regulación,	2.1 Secuencia de identificador [M] 2.2 Esquema de identificador [C]
<b>3. Nombre</b>	M	Documento, Actividad, Relación	Agente, Regulación,	3.1 Nombre Natural [M] 3.2 Nombre del fichero [O]
<b>4. Fechas</b>	M	Documento, Actividad, Relación	Agente, Regulación,	4.1 Fecha inicio [M] 4.2 Fecha fin [C]
5. Descripción	O	Documento, Actividad, Relación	Agente, Regulación,	
<b>6. Entidad Relacionada</b>	M	Relación		6.1 ID de entidad relacionada [M] 6.2 Esquema de ID de entidad relacionada [C] 6.3 Rol de la relación [M]
7. Ambito	O	Documento, Actividad, Regulación	Agente,	
8. Seguridad	C	Documento, Regulación	Actividad,	8.1 Nivel de Seguridad [C] 8.2 Advertencia de seguridad [C] 8.3 Permisos [C] 8.4 Sensibilidad datos de carácter personal [C] 8.5 Clasificación ENS [C]



<b>Elemento</b>	<b>Obrigatorie dade</b>	<b>Entidade a que se aplica</b>	<b>Subpropiedades obligatoriedade</b>	<b>e</b>
			8.6 Nivel de confidencialidad de la información [M]	
9. Derechos de acceso, uso y reutilización	C	Documento	9.1 Tipo de acceso [C] 9.2 Código de la causa de limitación [C] 9.3 Causa legal/normativa de limitación [C] 9.4 Condiciones de reutilización [C]	
10. Contacto	C	Agente	10.1 Tipo de contacto [C] 10.2 Dato de contacto [C] 10.3 Puesto [C]	
11. Idioma	C	Documento, Agente	-	
12. Puntos de acceso	O	Documento, Regulación	12.1 Término punto de acceso [O] 12.2 ID de punto de acceso [MC] 12.3 Esquema [C]	
<b>13. Calificación</b>	M	Documento	13.1 Valoración [M] 13.2 Dictamen [M] 13.3 Transferencia [M]	
<b>14. Características técnicas</b>	M	Documento	14.1 Formato [M] 14.2 Versión de formato [O] 14.3 Resolución [O] 14.4 Tamaño [C] 14.5 Profundidad de color [O]	
15. Ubicación	O	Documento	15.1 Soporte [C] 15.2 Localización [O]	
16. Verificación de integridad	C	Documento	16.1 Algoritmo [O] 16.2 Valor [O]	
<b>17. Firma</b>	M	Documento	17.1 Tipo de Firma [M] 17.2 Rol de firma [C] 17.3 Valor del CSV [C]	

Elemento	Obrigatorie dade	Entidade a que se aplica	Subpropriedades obrigatoriedade	e
			17.4 Definición generación CSV [C]	
			17.5 Firmante [O]	
<b>18. Tipo documental</b>	M	Documento	-	
19. Prioridad	O	Documento, Actividad	-	
<b>20. Estado de elaboración</b>	M	Documento	-	
21. Trazabilidad	O	Documento, Actividad, Relación	Agente, Regulación,	21.1 Acción [O] 21.2 Motivo reglado [O] 21.3 Usuario de la acción [C] 21.4 Descripción [C] 21.5 Modificación de los metadatos [C] 21.6 Historia del cambio [C]
<b>22. Clasificación</b>	M	Documento	22.1 Código de clasificación [M] 22.2 Denominación de clase [M] 22.3 Tipo de clasificación (SIA/Funcional) [M]	
<b>23. Versión NTI</b>	M	Documento	-	
<b>24. Órgano</b>	M	Documento	-	
<b>25. Origen del documento</b>	M	Documento	-	
26. Identificación del documento origen	C	Documento	-	
<b>27. Estado del Expediente</b>	M	Documento	-	
<b>28. Interesado</b>	M	Documento	-	
29. Asiento Registral	O	Documento	29.1 Tipo de asiento registral [O] 29.2 Código de la oficina de registro [O] 29.3 Fecha del asiento registral [O] 29.4 Número de asiento registral [O]	

Elemento	Obrigatoriedade	Entidade a que se aplica	Subpropriedades obrigatorias	e
----------	-----------------	--------------------------	------------------------------	---

FONTE: Adaptado de Espanha (2016).

#### 4.3.4.3 Metadados para a autenticidade

A partir da estrutura de metadados, identificam-se 21 propriedades ou subpropriedades que se entendem auxiliar na asserção da autenticidade dos recursos. Destes, 15 dizem respeito à autenticidade dos documentos: *identificador*, *nombre* (documento), *fechas*, *tamaño* (14.4), *soporte* (15.1), *tipo documental*, *categoría*, *descripción*, *ámbito*, *entidad relacionada*, *clasificación*, *historia del cambio*, *nombre* (agente), *órgano e origen del documento*. O identificador, o nome do documento, as datas, a extensão física, o suporte e o tipo documental caracterizam os documentos enquanto objetos físicos. A categoria, a descrição, o âmbito, as entidades relacionadas e o código de classificação dizem respeito à estrutura da agregação, à natureza dos documentos no contexto dessa agregação, e indicam as atividades das quais os documentos constituem a manifestação. A história de alterações, o nome do agente, o órgão e a origem do documento informam sobre a proveniência dos documentos.

Acerca da integridade, destacam-se seis elementos: *dictamen* (13.2), *características técnicas*, *localización* (15.2), *verificación de integridad*, *firma e estado de elaboración*. A avaliação (*dictamen*) informa sobre as intervenções do arquivista na gestão dos documentos. As características técnicas ajudam a esclarecer as especificidades tecnológicas do suporte do documento que, em conjunto com o seu estado de elaboração, são úteis para a compreensão da sua transmissão. A verificação da integridade e a assinatura constituem mecanismos explícitos de manifestação da integridade, enquanto que a localização pode informar sobre a entidade responsável pela custódia e por eventuais alterações aos documentos.

A respeito da autenticidade das próprias descrições, observa-se uma notável ausência de metadados que possam prestar informações sobre este aspeto. Ainda assim indicam-se dois elementos que poderiam adequar-se: *contacto* e *historia del cambio*. O contacto do agente pode ser útil como uma forma de transparência e de acessibilidade ao criador das descrições. Porém, este elemento não se refere explicitamente ao criador dos metadados. Embora o elemento da história das mudanças pareça estar concebido para documentar alterações ao conteúdo dos metadados, talvez se possa incluir uma história das alterações sofridas pelos próprios metadados.

É ainda de referir que não foram considerados como importantes para a autenticidade três dos elementos obrigatórios previstos pela norma — *versión NTI*, *estado del expediente* e *interesado* — uma vez que parecem estar diretamente dependentes da realidade jurídica espanhola. Assim, como a proposta de modelo pretende ter uma aplicabilidade relativamente universal, opta-se pela sua exclusão. Além disso, a respeito da autenticidade dos documentos, regista-se que parecem redundantes os termos ‘*nombre*’ (aplicado ao Agente), ‘*órgano*’ e ‘*origen del documento*’, na

medida em que os três identificam o produtor, os dois primeiro indicando o seu nome e o último indicando se se trata de um cidadão ou uma administração, o que deve ser facilmente deduzível a partir do nome.

### 4.3.5 DACS

#### 4.3.5.1 Contexto histórico

A *Describing Archives: A Content Standard* (DACS) é a norma de conteúdo oficial da comunidade arquivística norte-americana, adotada pela SAA em 2005, data da sua primeira publicação. A norma foi concebida para a descrição arquivística e para a produção de instrumentos de pesquisa. A DACS substituiu a *Archives, Personal Papers, and Manuscripts* (APPM), que tinha servido a comunidade arquivística norte-americana durante mais de duas décadas. Era uma norma de conteúdo desenvolvida por Steve Hensen enquanto funcionário da Biblioteca do Congresso, que por sua vez se baseava nas AACR2, uma norma para a descrição de conteúdos bibliográficos.

Uma segunda edição da norma foi publicada em 2013, atestando uma significativa revisão com base no *feedback* da comunidade. Dois dos aspetos sob revisão foram a crescente convergência entre as normas descritivas entre arquivos, museus e bibliotecas — destacando-se a publicação da RDA em 2010 —, bem como a necessidade sentida de melhor adequação da norma à representação de recursos digitais. Outra motivação da revisão, iniciada em 2010, prende-se com uma intenção de alinhar a DACS com as normas desenvolvidas pelo ICA, de onde resulta que a DACS se baseia na ISAD(G), podendo ser utilizada para criar instrumentos de pesquisa com vários níveis de descrição. A norma é composta por duas partes, sendo a primeira um espelho da ISAD(G), ao passo que a segunda parte reflete a ISAAR(CPF). As funções — abordadas pela ISDF — e as instituições arquivísticas — descritas pela ISIADH — não são atualmente consideradas na DACS.

A terceira e atual versão da norma foi publicada em 2020, na sequência de uma revisão realizada em 2018 à declaração de princípios da descrição arquivística. Os princípios veiculados pela edição de 2013 refletem uma assinalável familiaridade com os princípios de descrição prescritos pela ISAD(G), destacando-se a descrição multinível e hierárquica, centrada mais na criação dos instrumentos de pesquisa como um fim do que nas necessidades dos utilizadores, como o eixo concetual da representação. Porém, a publicação da versão *draft* da RiC pelo ICA em 2016 incentivou a SAA a rever a DACS numa linha de evolução convergente com a RiC, contextualizando assim uma profunda revisão aos princípios de descrição. Atualmente, a DACS fundamenta-se nos seguintes princípios para a descrição arquivística:

- **Princípio 1:** *Archival description expresses professional ethics and values.*
- **Princípio 2:** *Users are the fundamental reason for archival description.*
- **Princípio 3:** *Because archival description privileges intellectual control in context, descriptive rules apply to all records, regardless of format or carrier type.*
- **Princípio 4:** *Records, agents, activities, and the relationships between them are four fundamental concepts that constitute archival description.*

- **Princípio 4.1:** *records must be described in aggregate and may be described in parts.*
- **Princípio 4.2:** *Record creators and other agents must be described sufficiently to understand the meaning of records.*
- **Princípio 4.3:** *Activities that are essential to understanding records must be described.*
- **Princípio 4.4:** *The relationship among records, agents, and activities are essential to understanding archives and must be described.*
- **Princípio 5:** *Archival description must be clear about what archivists know, what they don't know, and how they know it.*
- **Princípio 6:** *Archivists must document and make discoverable the actions they take on records.*
- **Princípio 7:** *Archival description is accessible.*
- **Princípio 8:** *Archival description should be easy to use, re-use, and share.*
- **Princípio 9:** *Each collection within a repository must have an archival description.*
- **Princípio 10:** *Archivists must have a user-driven reason to enhance existing archival description.*
- **Princípio 11:** *Archival description is a continuous intellectual endeavor (Society of American Archivists, 2020, pp. xiii-xvii).*

A atualização dos princípios reflete claramente um esforço de adequação às práticas contemporâneas de descrição arquivísticas, com destaque para a crescente consciencialização da subjetividade do arquivista e da necessidade de promover a transparência dessa própria subjetividade, conforme expresso pelo princípio 5.

A DACS prevê vários níveis de descrição, podendo os instrumentos de pesquisa possuir um ou mais níveis de descrição. A escolha do número de níveis a adotar deve ser tomada pelos arquivistas de acordo com as prescrições das suas instituições. A norma também fornece liberdade para expressar os elementos sob qualquer ordem de apresentação. Além disso, a DACS pretende ser de uma implementação neutra quanto à definição do formato para expressão do conteúdo, no sentido em que não prescreve nenhuma forma específica de *output* dos metadados, revelando assim abertura para ser codificada em MARC21, RDF, XML ou qualquer outro formato.

#### 4.3.5.2 Estrutura de metadados

A DACS prevê o uso de 25 elementos, um conjunto desenvolvido em proximidade com os 26 elementos da ISAD(G). Cada elemento é caracterizado de acordo com o seu *purpose and scope*, *exclusions*, *sources of information* e *general rules*. A primeira parte da norma, focada na descrição dos materiais arquivísticos, é composta por 7 áreas:

2. *Identity elements*
3. *Content and structure elements*

4. *Conditions of access and use elements*
5. *Acquisition and appraisal elements*
6. *Related materials elements*
7. *Notes elements*
8. *Description control element*

A numeração aqui apresentada começa intencionalmente no 2, de modo a replicar a numeração seguida na norma, evitando confusões. Os números correspondem aos capítulos que abordam cada secção, sendo o primeiro capítulo dedicado aos níveis de descrição. A DACS prescreve o uso de 9 elementos obrigatórios e de dois considerados ótimos, além de prever a adição de qualquer informação que possa acrescentar valor. Os elementos ótimos são aqueles que, embora não sendo obrigatórios, constituem um extra desejado. Existem ainda as descrições com valor acrescentado que, incluem todos os elementos obrigatórios e ótimos. É dada liberdade a cada organização de definir o que pode acrescentar valor. Os elementos obrigatórios são:

- |   |  |
|---|--|
| a. <i>reference code</i>                  | f. <i>name of creator[s]</i>                   |
| b. <i>name and location of repository</i> | g. <i>scope and content</i>                    |
| c. <i>title</i>                           | h. <i>conditions governing access</i>          |
| d. <i>date</i>                            | i. <i>language and scripts of the material</i> |
| e. <i>extent</i>                          |  |

Estes elementos podem ser complementados com dois elementos ótimos: *administrative/biographical history* e *access points*. Tanto os elementos obrigatórios como os ótimos são comuns tanto a descrições com um só nível, como a descrições multinível. Porém, nas descrições multinível acrescenta-se a necessidade de identificação da relação da parte com o todo, indicada no nível mais alto da hierarquia e remetendo para hierarquias subsequentes. Além disso, todos os níveis inferiores devem incluir novamente os elementos usados pelo nível parente, exceto se a informação for idêntica. A lista completa de elementos contemplados pela norma é reproduzida na Tabela 50, bem como o seu grau de obrigatoriedade. Os elementos obrigatórios são também reforçados a negrito.

Tabela 50 — Estrutura de metadados da DACS.

Área	Elemento	Obrigatoriedade
2. Identity	<b>2.1 Reference code</b>	Required
	<b>2.2 Name and location of repository</b>	Required
	<b>2.3 Title</b>	Required
	<b>2.4 Date</b>	Required
	<b>2.5 Extent</b>	Required
	<b>2.6 Name of creator(s)</b>	Required, if known

Área	Elemento	Obrigatoriedade
	2.7 Administrative/biographical history	Optimum
3. Content and Structure	<b>3.1 Scope and content</b>	Required
	3.2 System of arrangement	Added Value
4. Conditions of Access and Use	<b>4.1 Conditions Governing Access</b>	Required
	4.2 Physical access	Added Value
	4.3 Technical access	Added Value
	4.4 Conditions governing reproduction and use	Added Value
	<b>4.5 Languages and scripts of the material</b>	Required
	4.6 Finding aids	Added Value
5. Acquisition and Appraisal elements	5.1 Custodian history	Added Value
	5.2 Immediate source of acquisition	Added Value
	5.3 Appraisal, destruction, and scheduling information	Added Value
	5.4 Accruals	Added Value
6. Related Materials	6.1 Existence and location of originals	Added Value
	6.2 Existence and location of copies	Added Value
	6.3 Related archival materials	Added Value
	6.4 Publication note	Added Value
7. Notes	7.1 Notes	Added Value
8. Description Control	8.1 Description control	Added Value

FONTE: Society of American Archivists (2020).

#### 4.3.5.3 Metadados para a autenticidade

A partir da leitura desta estrutura de metadados, identificam-se 16 elementos relacionados com a autenticidade dos documentos. De entre este grupo, os seguintes nove informam sobre a identidade dos documentos: *reference code*, *title*, *date*, *extent*, *scope and content*, *system of arrangement*, *name of creator(s)*, *name and location of repository* e *administrative/biographical history*. O código de referência, o título, a data e a extensão descrevem as características físicas dos documentos. O âmbito e conteúdo e o sistema de classificação clarificam a relação entre os documentos e as atividades que governaram a sua criação. Por fim, o nome do(s) criador(es), a identificação e a localização do repositório, bem como a história administrativa ou biográfica contribuem para estabelecer a proveniência dos documentos.

A respeito da integridade, destacam-se sete elementos: *custodian history*, *immediate source of acquisition*, *appraisal*, *destruction*, *and scheduling information*, *accruals*, *existence and location of originals*, *existence and location of copies* e *related archival materials*. As fontes de aquisição e a indicação da localização dos originais e das cópias assistem na demonstração da custódia dos documentos ao longo do tempo. Os campos sobre a avaliação, a destruição e adições também contribuem para documentar as modificações sofridas pelos agregados, complementando a história da transmissão dos documentos e das intervenções dos arquivistas nessa história. A indicação de materiais relacionados reforça as relações com outros documentos do mesmo agregado.

Por último, salientam-se apenas dois campos que possam auxiliar na identificação da autenticidade dos metadados: *notes* e *description control*. As notas constituem um espaço de manifestação da voz do arquivista, podendo incluir virtualmente qualquer informação, como declarações de responsabilidade ou assinaturas. O controlo da descrição, embora não subdividido, tem como objetivo geral documentar a criação e a revisão das descrições arquivísticas, mediante indicação do nome dos responsáveis pela descrição/revisão, das datas e das convenções seguidas.

### 4.3.6 DCMES

#### 4.3.6.1 Contexto histórico

O *Dublin Core Metadata Element Set* (DCMES), conhecido de forma simplificada por “Dublin Core”, resulta de uma *workshop* sobre semântica em metadados realizada em Dublin, Ohio, em 1995, de onde deriva o seu nome. Os contornos deste momento já foram brevemente mencionados em 1.2.4, quando se assinalou que mais de 50 profissionais da informação, entre os quais bibliotecários e arquivistas, se reuniram para discutir se poderia ser criado um conjunto essencial de atributos semânticos para a descrição de recursos na *web*, o que facilitaria extraordinariamente a pesquisa e a recuperação numa *web* então largamente desestruturada.

O resultado foi a criação de um conjunto de elementos de metadados que reuniu o consenso internacional para a criação e para a disseminação de recursos na *web*. Neste sentido, os elementos não são específicos a nenhum domínio, afirmando-se antes como uma lista generalista de termos que podem ser aplicados em qualquer contexto. É favorecida a simplicidade da descrição, como forma de promover a capacidade de descoberta dos recursos, em detrimento de uma descrição extensiva. Esta abordagem tem sido extremamente bem-sucedida a nível internacional, tendo o DCMES sido aprovado como norma ANSI/NISO Z39.85 em 2001, e como norma ISO 15836 em 2003 (Dublin Core Metadata Initiative, 2021b).

A lista de termos é mantida pela *Dublin Core Metadata Initiative* (DCMI), que desde 2002 se tem afirmado como uma agência normativa na medida em que gere e atualiza a documentação sobre os *DCMI Metadata Terms*. Os termos consistem num conjunto de vocabulários formado pelo DCMES, a que se acrescentam dezenas de propriedades, classes, tipos de dados e esquemas de codificação. Estes termos podem ser utilizados em conjunto com outros metadados de outros vocabulários em cenários específicos de implementação através da construção de *application*



*profiles* que respondam às necessidades de um determinado domínio de aplicação (Dublin Core Metadata Initiative, 2021a).

Os metadados do DCMI são preferencialmente expressos em RDF para uso no âmbito de *linked data*. Porém, os metadados também podem ser usados como XML, JSON, UML ou bases de dados relacionais, bastando apenas para isso desconsiderar os identificadores globais e as implicações específicas das implementações RDF. Ao nível da descrição, Joudrey e Taylor (2018, p. 335) identificam seis princípios gerais pelos quais o DCMES se rege:

- o conjunto essencial de elementos pode ser expandido de acordo com as necessidades das comunidades;
- todos os elementos são opcionais;
- todos os elementos são repetíveis;
- cada recurso tem uma descrição;
- qualquer elemento pode ser refinado por qualificadores, mas
- os qualificadores não podem mudar o significado do elemento.

A ideia de que cada descrição se refere a um e um só recurso é uma manifestação do princípio 1:1 (*one-to-one*). Por outras palavras, cada elemento deve focar-se no recurso a ser descrito, não em entidades associadas com esse recurso. Contudo, é de notar que o DCMES não prescreve a forma de preenchimento do conteúdo de cada elemento, referindo apenas boas práticas que as organizações podem decidir seguir ou não. Ainda assim, o DCMES tem sido usado de forma global por diversas comunidades, desde catálogos a bases de dados. Um dos exemplos de implementação é na base de dados do WorldCat, a biblioteca global mencionada no primeiro capítulo.

#### 4.3.6.2 Estrutura de metadados

O DCMES consiste num conjunto essencial de 15 elementos, dos quais nenhum é obrigatório. Cada termo é caracterizado de acordo com quatro atributos mínimos que auxiliam a sua interpretação: (i) *label*; (ii) *URI*; (iii) *definition* e (iv) *type of term*.

Estes 15 elementos podem ser expandidos através de subelementos denominados *refinements*, que podem pertencer a duas classes. Estes podem consistir em qualificadores que especificam o significado de um elemento, ou ser esquemas de codificação que identificam um vocabulário controlado que auxilia na interpretação do valor de um elemento (referidos como *Vocabulary Encoding Schemes*, *Syntax Encoding Schemes* ou *Data Types*). No âmbito da análise desta investigação, consideram-se sobretudo os primeiros qualificadores, uma vez que representam elementos adicionais de estrutura que focam a descrição arquivística, ao passo que a segunda classe entra mais no domínio da partilha da informação.

A estrutura dos elementos base proposta pelo DCMES é apresentada na Tabela 51, juntamente com as suas definições e com alguns dos *refinements* aplicáveis. São ainda destacados

alguns termos que, embora não associados a nenhum dos 15 elementos essenciais, podem ser usados na descrição de coleções.

Tabela 51 — Estrutura de elementos do Dublin Core, com alguns *refinements*.

<b>Elemento</b>	<b>Refinamento</b>	<b>Definição</b>
1. Contributor	-	An entity responsible for making contributions to the resource.
2. Coverage	Spatial; temporal	The spatial or temporal topic of the resource, spatial applicability of the resource, or jurisdiction under which the resource is relevant.
3. Creator	-	An entity primarily responsible for making the resource.
4. Date	available; created; dateAccepted; dateCopyrighted; dateSubmitted; issued; modified; valid	A point or period of time associated with an event in the lifecycle of the resource.
5. Description	abstract; tableOfContents	An account of the resource.
6. Format	extent; medium	The file format, physical medium, or dimensions of the resource.
7. Identifier	bibliographicCitation	An unambiguous reference to the resource within a given context.
8. Language	-	A language of the resource.
9. Publisher	-	An entity responsible for making the resource available.
10. Relation	conformsTo; hasFormat; hasPart; hasVersion; isFormatOf; isPartOf; isReferencedBy; isReplacedBy; isRequiredBy; isVersionOf; references; replaces; requires	A related resource.
11. Rights	accessRights; license	Information about rights held in and over the resource.
12. Source	-	A related resource from which the described resource is derived.
13. Subject	-	The topic of the resource.
14. Title	alternative	A name given to the resource.
15. Type	-	The nature or genre of the resource.
Termos associados à descrição de coleções	accrualMethod; accrualPeriodicity;	

Elemento	Refinamento	Definição
	accrualPolicy; provenance	

FONTE: Dublin Core Metadata Initiative (2021a).

#### 4.3.6.3 Metadados para a autenticidade

Uma análise dos elementos acima apresentados permite identificar 10 elementos e quatro termos para a autenticidade dos recursos, um número relativamente reduzido, que não é de admirar dado o caráter generalista da norma.

Acerca da identidade dos recursos, mencionam-se oito elementos de metadados: *coverage*, *date*, *format*, *identifier*, *subject*, *creator*, *contributor*, *title*, bem como o termo *provenance*. A cobertura, a data, o formato e o identificador descrevem o recurso nas suas características físicas. O assunto presta informações sobre a natureza do recurso, ao passo que o criador, o contribuidor, o título e a declaração de proveniência ajudam a estabelecer a proveniência dos recursos, incluindo potenciais descrições de mudanças de custódia.

A respeito da integridade, assinalam-se apenas dois elementos, *relation* e *source*, a par dos termos *accrualMethod*, *accrualPeriodicity* e *accrualPolicy*. A descrição de recursos relacionados, incluindo da fonte a partir de qual a descrição é derivada, ajuda a manifestar as relações que os recursos possam ter com outros recursos, independentemente do agregado. Além disso, informações sobre os métodos, periodicidade e políticas de adição são importantes para testemunhar uma parte da história das mudanças sofridas pelo conjunto documental. É importante notar uma aparente ausência de um histórico de eliminações e de outras informações decorrentes da avaliação.

Relativamente ao controlo das descrições, indica-se um único elemento, *publisher*, enquanto explicitação da entidade responsável pela disponibilização do recurso, que se depreende referir ao recurso enquanto descrição.

### 4.3.7 VRA Core

#### 4.3.7.1 Contexto histórico

A *Visual Resources Association (VRA) Core* é uma norma de metadados *for the description of works of visual culture as well as the images that document them* (Visual Resources Association, 2014). De entre tais obras incluem-se pinturas, esculturas, arquitetura, fotografias, livros e artes performativas. Trata-se de uma norma internacional que pode ser usada quer por si própria, quer como uma extensão à norma *Metadata Transmission and Encoding Schema (METS)*. A VRA Core usada não só em escolas de arte e de arquitetura, como também em bibliotecas,

arquivos e museus quando é necessário fornecer acesso a obras de artes culturais e às respectivas imagens<sup>88</sup>. A norma é expressa segundo um esquema XML de forma a garantir a interoperabilidade.

A primeira versão (1.0) da VRA Core remonta a 1996, tendo sido largamente baseada na CDWA (Visual Resources Association, 1996). A versão 2.0 foi publicada em 1999, no qual é introduzido o conceito de ‘*work*’ por substituição de ‘*object*’ (Visual Resources Association, 1999). Na versão 3.0, de 2002, adotou-se um único *element set* que seria aplicado quantas vezes necessário para a descrição das obras, de acordo com o princípio 1:1 (Visual Resources Association, 2002). A versão 4.0, atualmente a mais recente, foi publicada em 2007 e atualizada em 2014. Nesta versão é empregue o uso de três entidades, *collection*, *work* e *images*, sendo as descrições construídas a partir da explicitação das relações entre estas entidades.

Nesta versão adota-se o princípio *1-to-many*, segundo o qual podem estabelecer-se relações entre diversas descrições e uma única entidade. Uma única imagem pode relacionar-se com uma ou mais obras, da mesma forma que uma coleção pode agregar diversas obras ou imagens. Ainda assim, o principal foco da descrição da VRA Core é a obra, que pode ser associado a uma ou mais imagens através do elemento *relation*. Esta noção atesta que o significado de uma imagem é derivado da explicitação da sua relação com a obra da qual deriva. A natureza da relação é definida pelo elemento *type*.

#### 4.3.7.2 Estrutura de metadados

A documentação que caracteriza os elementos a descrever na VRA Core 4.0 assinala o possível uso de 19 categorias de elementos, sendo cada elemento caracterizado por uma definição, por recomendações de uso de vocabulários controlados, se o elemento é obrigatório ou repetível, bem como mapeamentos para outras normas, nomeadamente as versões 2.0 e 3.0 da VRA Core, a CDWA, a Dublin Core e a *Cataloging Cultural Objects (CCO)*.

Apenas o elemento *work*, *collection* or *image*, que especifica o tipo de recurso a descrever, é obrigatório e não-repetível. Todos os restantes elementos são opcionais e repetíveis, uma abordagem condizente com o princípio *1-to-many*. Neste sentido, cabe às organizações escolher os elementos de metadados que mais se adequam para a descrição das obras e das imagens e das coleções associadas.

A norma prevê ainda o uso de nove atributos globais que podem ser adicionados a qualquer elemento ou subelemento: *dataDate*, *extent*, *href*, *pref*, *refid*, *rules*, *source*, *vocab* e *xml:lang*. Cada elemento pode ainda ser descrito mais especificamente por subelementos, também opcionais (Visual Resources Association, 2007b). A lista integral dos elementos é apresentada na Tabela 52, juntamente com uma definição e subelementos possíveis. O único elemento obrigatório encontra-se assinalado a negrito.

---

<sup>88</sup> Uma lista de organizações que empregam a VRA Core é mantida em <[http://core.vraweb.org/vracore\\_registry.html](http://core.vraweb.org/vracore_registry.html)>.

Tabela 52 — Elementos de metadados usados na VRA Core 4.0.

<b>Elemento</b>	<b>Subelementos</b>	<b>Definição</b>
1. Work, collection or image	-	A choice of one of three elements, WORK, COLLECTION, or IMAGE, defines a VRA 4.0 record as describing a WORK (a built or created object), a COLLECTION (an aggregate of such objects), or an IMAGE (a visual surrogate of such objects).
2. Agent	attribution, culture, name, role, dates,	The names, appellations, or other identifiers assigned to an individual, group, or corporate body that has contributed to the design, creation, production, manufacture, or alteration of the work or image.
3. CulturalContext	-	The name of the culture, people (ethnonym), or adjectival form of a country name from which a Work, Collection, or Image originates, or the cultural context with which the Work, Collection, or Image has been associated.
4. Date	earliestDate, latestDate	Date or range of dates associated with the creation, design, production, presentation, performance, construction, or alteration, etc. of the work or image.
5. Description	-	A free-text note about the Work, Collection, or Image, including comments, description, or interpretation, that gives additional information not recorded in other categories.
6. Inscription	author, position, text	All marks or written words added to the object at the time of production or in its subsequent history, including signatures, dates, dedications, texts, and colophons, as well as marks, such as the stamps of silversmiths, publishers, or printers.
7. Location	name, refid	The geographic location and/or name of the repository, building, site, or other entity whose boundaries include the Work or Image.
8. Material	-	The substance of which a work or an image is composed.
9. Measurements	-	The physical size, shape, scale, dimensions, or format of the Work or Image.
10. Relation	-	Terms or phrases describing the identity of the related work and the relationship between the work being cataloged and the related work or image.
11. Rights	notes, rightsHolder, text	Information about the copyright status and the rights holder for a work, collection, or image.
12. Source	name, refid	A reference to the source of the information recorded about the work or the image.

<b>Elemento</b>	<b>Subelementos</b>	<b>Definição</b>
13. State Edition	description, name	The identifying number and/or name assigned to the state or edition of a work that exists in more than one form and the placement of that work in the context of prior or later issuances of multiples of the same work.
14. Style Period	-	A defined style, historical period, group, school, dynasty, movement, etc. whose characteristics are represented in the Work or Image.
15. Subject	term	Terms or phrases that describe, identify, or interpret the Work or Image and what it depicts or expresses.
16. Technique	-	The production or manufacturing processes, techniques, and methods incorporated in the fabrication or alteration of the work or image.
17. Textref	name, refid	Contains the name of a related textual reference and any type of unique identifier that text assigns to a Work or Collection that is independent of any repository.
18. Title	-	The title or identifying phrase given to a Work or an Image.
19. WorkType	-	Identifies the specific type of WORK, COLLECTION, or IMAGE being described in the record.

FONTE: Visual Resources Association (2007a, 2007b).

#### 4.3.7.3 Metadados para a autenticidade

Na VRA Core 3.0 identificam-se 17 elementos relacionados com a asserção da autenticidade, o que parece revelar uma elevada preocupação com a caracterização da identidade e da integridade das obras de arte. Ao nível da identidade, destacam-se 11 elementos: *date*, *material*, *measurements*, *description*, *subject*, *culturalContext*, *stylePeriod*, *title*, *workType*, *agent* e *source*. A indicação das datas, dos materiais e das medidas físicas pretendem descrever as obras/imagens na sua dimensão física. Informações sobre a descrição, o assunto, o contexto cultural e a corrente estilística ajudam a caracterizar as atividades das quais os recursos resultam, contribuindo para a perceção do contexto. A indicação do título, do tipo de obra, do agente e da fonte estabelecem a proveniência dos recursos.

A respeito da integridade, destacam-se os seguintes seis elementos: *location*, *inscription*, *relation*, *textRef*, *stateEdition* e *technique*. A localização pode prestar informações geográficas sobre entidades que detêm a custódia dos recursos. A indicação de inscrições nas obras, bem como das técnicas usadas para as criar, caracterizam as alterações físicas sofridas pelos recursos ao longo do tempo. A edição, as relações e as referências com outras obras estabelecem conexões entre

recursos relacionados na mesma ou em outras coleções. É ainda de assinalar que não se constatou o uso de metadados para controlo das descrições.

### 4.3.8 MODS

#### 4.3.8.1 Contexto histórico

A *Metadata Object Description Schema* (MODS) consiste numa estrutura XML para a representação de recursos bibliográficos, sendo por isso principalmente aplicado em bibliotecas. Foi derivada a partir do formato MARC21, do qual inclui um conjunto de subcampos MARC, que usam *tags* linguísticas em vez de *tags* numéricas (Library of Congress, 2020a), conforme visto no primeiro capítulo. Este aspeto facilita a interpretação por falantes da língua inglesa, mas dificulta a criação de *tags* por falantes de outras línguas. Além disso, uma vez que muitos dos elementos MARC foram agregados sob um mesmo elemento no MODS, os documentos que forem convertidos do formato MARC para o MODS não podem ser reconvertidos sem perdas de informação.

Este esquema foi desenvolvido em 2002 pela *Library of Congress Network Development* e pelo *MARC Standards Office* em colaboração com outros especialistas (Library of Congress, 2016) com o objetivo de complementar outras normas de metadados e de fornecer uma alternativa *between a very simple metadata format with a minimum of fields and no or little substructure (for example, Dublin Core), and a very detailed format with many data elements having various structural complexities such as MARC21* (Guenther, 2003, p. 139). Por outras palavras, é mais complexo que o DC mas mais simples que o MARC21. A versão atual, de 2020, é a 3.7<sup>89</sup>. À data da escrita, em inícios de 2021, a Library of Congress (2021) enumera 39 projetos que utilizam esta norma.

A Biblioteca do Congresso assinala 18 princípios orientadores da conceção desta norma:

1. *Support localization and customization needs;*
2. *Accommodate widely adopted descriptive practices;*
3. *Maintain a relatively small number of elements and attributes to reduce training, application, and implementation costs;*
4. *Support the communication of resource and authority descriptions;*
5. *Support validation of the encoding;*
6. *Allow use of MODS/MADS elements by other standards and in application profiles;*
7. *Maintain continuity of structure and content;*
8. *Maintain a single way to encode a piece of information;*
9. *Accommodate indexing of data in the description;*
10. *Accommodate presentation of data in the description;*

---

<sup>89</sup> Uma lista de alterações pode ser consultada em <<https://www.loc.gov/standards/mods/changes-3-7.html>>.

11. *Make element and attribute names as intelligible as possible to a general audience;*
12. *Allow for extensibility to include data from richer element sets;*
13. *Accommodate information about the metadata and record itself;*
14. *Accommodate conversion to and from other commonly used resource and authority description encodings (such as Dublin Core, MARC, VRA Core);*
15. *Accommodate controlled vocabularies that are commonly used in resource and authority description;*
16. *Allow full description of whole-to-part and similar types of relationships;*
17. *Support encoding a description for any type of resource;*
18. *Support encoding the relationship of an agent to a resource (Library of Congress, 2009).*

Dada a sua estrutura em XML, a MODS adota uma descrição hierárquica segundo a qual um elemento é sucessivamente decomposto em subelementos que herdam as propriedades semânticas. Os elementos raiz são <mods> (para um único documento MODS) e <modsCollection> (para coleções de documentos MODS).

#### 4.3.8.2 Estrutura de metadados

A MODS 3.7 regista 20 elementos, que podem ser especificados mediante atributos ou subelementos. Todos os elementos de topo, bem como todos os atributos, são opcionais. O único requisito é que cada documento MODS tenha pelo menos um elemento. Os subelementos são geralmente opcionais, embora alguns elementos não possam ser contentores vazios. A Tabela 53 consolida os elementos e os subelementos desta norma.

Tabela 53 — Elementos e subelementos da MODS 3.7.

Elemento	Subelementos
1. titleInfo	title, subtitle, partNumber, partName, nonSort
2. name	namePart, nameIdentifier, displayForm, affiliation, role, description
3. typeOfResource	-
4. genre	-
5. originInfo	place, publisher, dateIssued, dateCreated, dateCaptured, dateValid, dateModified, copyrightDate, dateOther, edition, issuance, frequency
6. language	languageTerm, scriptTerm
7. physicalDescription	form, reformattingQuality, InternetMediaType, extent, digitalOrigin, note
8. abstract	-
9. tableOfContents	-



Elemento	Subelementos
10. targetAudience	-
11. note	-
12. subject	topic, geographic, temporal, titleInfo, name, geographicCode, genre, hierarchicalGeographic, cartographics, occupation
13. classification	-
14. relatedItem	Qualquer elemento MODS.
15. identifier	-
16. location	physicalLocation, shelfLocator, url, holdingSimple, holdingExternal
17. accessCondition	-
18. part	detail, date, text
19. extension	-
20. recordInfo	recordContentSource, recordCreationDate, recordChangeDate, recordIdentifier, recordOrigin, recordInfoNote, languageOfCataloging, descriptionStandard

FONTE: (Library of Congress, 2020b).

#### 4.3.8.3 Metadados para a autenticidade

Embora a MODS seja uma norma desenvolvida segundo uma perspectiva bibliotecária, podendo contrastar com a perspectiva arquivística da autenticidade, é ainda assim possível identificar 12 elementos de metadados e quatro subelementos diretamente relacionados com a asserção da autenticidade.

Ao nível da identidade, selecionam-se 10 elementos: *physicalDescription*, *identifier*, *classification*, *genre*, *targetAudience*, *abstract*, *subject*, *titleInfo*, *name*, *typeOfReserve* e três subelementos de *originInfo*: *dateIssued*, *dateCreated* e *dateCaptured*. A indicação da data de emissão, criação ou captura, bem como a descrição das características físicas e de possíveis identificadores contribuem para a caracterização dos recursos na sua dimensão de objetos físicos a gerir. A classificação, o género, a audiência-alvo, o resumo e o assunto descrevem a estrutura e a finalidade do agregado, além de caracterizarem o conteúdo dos recursos enquanto produto desse agregado. A indicação do título e de nomes envolvidos na autoria do recurso, bem como a indicação do tipo de recurso, são úteis para estabelecer a proveniência.

Em relação à integridade, assinalam-se dois elementos (*relatedItem* e *location*) e um subelemento *dateModified* (de *originInfo*). A indicação da localização dos recursos pode indiciar sobre os detentores da custódia, ao passo que a clarificação de recursos relacionados reforça a explicitação dos elos que os recursos estabelecem com outros, do mesmo agregado ou externos.

Por último, a data de modificação pode servir como um registo de histórico de intervenções aos documentos. É ainda de mencionar o potencial do elemento *recordInfo* para servir como controlo das descrições, indicando as fontes, a data de criação, datas de modificação e permitindo aos criadores das descrições acrescentar notas.

### 4.3.9 CDWA

#### 4.3.9.1 Contexto histórico

A *Categories for the Description of Works of Art* (CDWA) é, à semelhança da VRA Core, uma norma de metadados usada para a descrição de obras de arte, de arquitetura, de coleções de obras e das suas imagens relacionadas. Mais especificamente, trata-se de um quadro de referência conceptual que pode ser usado como um conjunto de orientações de melhores práticas para a catalogação e descrição de recursos culturais, permitindo às organizações a construção de modelos de dados e de bases de dados a partir das recomendações da CDWA (Harpring, 2019).

A CDWA é o resultado de uma iniciativa da *Art Information Task Force* (AITF) que encorajou o diálogo entre a comunidade artística (historiadores da arte, curadores de museus, profissionais de recursos visuais, repositórios de arte) e a comunidade da informação (gestores de informação, bibliotecários de arte, especialistas técnicos) com vista à ao desenvolvimento de orientações comuns para a descrição de obras de arte. A atividade da AITF foi financiada pela J. Paul Getty Trust através de uma bolsa de dois anos do *National Endowment for the Humanities* (NEH) ao *College Art Association* (CAA). A primeira versão da CDWA foi publicada em 1996, tendo a última revisão sido efetuada em 2019.

Enquanto um conjunto de boas práticas, a AITF tem estabelecido uma série de mapeamentos entre a CDWA e outras normas do setor cultural, tais como a *Cultural Objects Name Authority* (CONA), que por sua vez está mapeada para vários vocabulários controlados como o AAT, TGN, ULAN e o *Getty Iconography Authority* (IA). A CDWA também está umbilicalmente relacionada com a CCO, na medida em que esta é composta por regras e exemplos de subconjuntos derivados da CDWA e da VRA Core. Houve também no passado uma derivação XML denominada *CDWA Lite*, que consistia num esquema XML para a descrição de obras de arte e de materiais culturais cujos elementos constituíam uma versão simplificada da CDWA e da CCO. O objetivo principal era enriquecer as descrições no âmbito dos catálogos conjuntos e de outros repositórios que pudessem agregar a informação mediante o protocolo *Open Archives Initiative* (OAI). Em 2010, o CDWA Lite foi expandido e integrado no *Lightweight Information Description Objects* (LIDO).

Originalmente, a CDWA foi recomendada enquanto uma estrutura de dados relacional, segundo a qual os documentos se encontram relacionados através de relações hierárquicas. Porém, as orientações mais recentes admitem a possibilidade de aplicação dos elementos de metadados a modelações de dados em grafo, refletindo assim as tendências mais recentes na representação da informação usando *linked data*. Uma recomendação é o uso do *Conceptual Reference Model*

(CRM) do CIDOC, na medida em que fornece uma ontologia extensa para conceitos e informação do setor cultural e museológico. Outra recomendação é o uso de vocabulários controlados. A par destas recomendações, a CDWA fornece mais esclarecimentos sobre algumas dúvidas comuns durante a descrição dos recursos, nomeadamente sobre a especificidade e a exaustividade, sobre a incerteza e a ambiguidade, sobre o desconhecido e o indeterminado, sobre a informação possível ou impossível, sobre contradições na fonte, sobre indexações de informações importantes, e sobre fontes.

#### 4.3.9.2 Estrutura de metadados

Na sua versão mais recente, a CDWA fornece um vasto leque de possibilidades para a descrição dos recursos, incluindo cerca de 540 categorias e subcategorias de informação. Deste largo conjunto, 28 categorias são consideradas pela norma essenciais e mínimas para qualquer descrição, que se encontram distribuídas por cinco conjuntos fundamentais:

- *For the object, architecture or group;*
- *For the person/corporate body authority;*
- *For the place/location authority;*
- *For the generic concept authority;*
- *For the subject authority* (Harpring, 2019, p. 7).

Porém, a seleção deste conjunto não significa que as organizações não possam escolher outros elementos mínimos, de acordo com o contexto específico de determinada implementação, considerando variáveis como as necessidades dos utilizadores que o sistema de informação de arte pretende servir ou a missão da instituição.

Dada a extensão da estrutura de metadados, não se reproduz aqui a totalidade da norma. Opta-se por incluir todas as 31 categorias principais, independentemente do seu grau de obrigatoriedade, acrescentando apenas as subcategorias, sem detalhe das subsubcategorias. De entre as 31 categorias principais, 27 dizem respeito à descrição dos objetos, da arquitetura ou do grupo, enquanto que as restantes quatro descrevem as autoridades. A interpretação de cada elemento é auxiliada pela indicação de uma definição. Assim, a Tabela 54 consolida as categorias e subcategorias previstas pela CDWA, estando assinaladas a negrito as 49 categorias que a norma identifica como sendo *core* a qualquer implementação. Apenas se incluem as subsubcategorias quando estas pertencem ao conjunto de elementos *core*. Além disso, todas as categorias incluem os subelementos *remarks* e *citations*, pelo que não se incluem de modo a abreviar a dimensão da tabela.

Tabela 54 — Categorias e subcategorias propostas pela CDWA.

Categoria	Subcategoria
<b>1. Object/Work</b>	<b>1.1 Catalog Level</b>
	<b>1.2 Object/Work Type</b>

<b>Categoria</b>	<b>Subcategoria</b>
	1.3 Object/Work Type Date
	1.4 Components/Parts
<b>2. Classification</b>	<b>2.1 Classification Term</b>
<b>3. Titles or Names</b>	<b>3.1 Title Text</b>
	3.2 Title Type
	3.3 Preference
	3.4 Title Language
	3.5 Title Date
<b>4. Creation</b>	<b>4.1 Creator Description</b>
	<b>4.1.3 Creator Identity</b>
	<b>4.1.4 Creator Role</b>
	<b>4.2 Creation Date</b>
	<b>4.2.1 Earliest Date</b>
	<b>4.2.2. Latest Date</b>
	4.3 Creation Place/Original Location
	4.4 Object/Work Culture
	4.5 Commissioner
	4.6 Creation Numbers
5. Styles/Periods/Groups/Movements	5.1 Styles/Periods Description
	5.2 Styles/Periods Indexing Terms
<b>6. Measurements</b>	<b>6.1 Dimensions Description</b>
	6.2 Dimensions Type
	6.3 Dimensions Value
	6.4 Dimensions Unit
	6.5 Dimensions Extent
	6.6 Scale Type
	6.7 Dimensions Qualifier
	6.8 Dimensions Date
	6.9 Shape
	6.10 Format/Size
<b>7. Materials/Techniques</b>	<b>7.1 Materials/Techniques Description</b>

<b>Categoria</b>	<b>Subcategoria</b>
	7.2 Materials/Techniques Flag
	7.3 Materials/Techniques Extent
	7.4 Materials/Techniques Role
	7.5 Materials/Techniques Name
	7.6 Material Color
	7.7 Material Source Place
	7.8 Watermarks
	7.9 Performance Actions
8. Inscriptions/Marks	8.1 Inscription Transcription or Description
	8.2 Inscription Type
	8.3 Inscription Author
	8.4 Inscription Location
	8.5 Inscription Language
	8.6 Typeface/Letterform
	8.7 Mark Identification
	8.8 Inscription Date
9. State	9.1 State Description
	9.2 State Identification
	9.3 Known States
10. Edition	10.1 Edition Description
	10.2 Edition Number or Name
	10.3 Impression Number
	10.4 Edition Size
11. Facture	11.1 Facture Description
12. Orientation/Arrangement	12.1 Orientation/Arrangement Description
	12.2 Orientation Indexing Terms
13. Physical Description	13.1 Physical Appearance
	13.2 Physical Description Indexing Terms
14. Condition/Examination History	14.1 Condition/Examination Description
	14.2. Examination Type
	14.3 Examination Agent

<b>Categoria</b>	<b>Subcategoria</b>
	14.4 Examination Date
	14.5 Examination Place
15. Conservation/Treatment History	15.1 Conservation/Treatment Description
	15.2 Treatment Type
	15.3 Treatment Agent
	15.4 Treatment Date
	15.5 Treatment Place
<b>16. Subject Matter</b>	16.1 Subject Display
	<b>16.2 General Subject Terms</b>
	16.3 Specific Subject Terms
	16.4 Outside Iconography Terms
	16.5 Subject Interpretive History
17. Context	17.1 Historical/Cultural Events
	17.2 Architectural Context
	17.3 Archeological Context
	17.4 Historical Location Context
18. Descriptive Note	18.1 Descriptive Note Text
19. Critical Responses	19.1 Critical Comment
	19.2 Comment Document Type
	19.3 Comment Author
	19.4 Comment Date
	19.5 Comment Circumstances
20. Related Works	20.1 Related Work Label/Identification
	20.2 Work Broader Context
	20.3 Relationship Number
<b>21. Current Location</b>	21.1 Current Location Description
	<b>21.2 Repository/Geographic Location</b>
	<b>21.2.3 Repository Numbers</b>
22. Copyright/Restrictions	22.1 Copyright Statement
	22.2 Copyright Holder Name
	22.3 Copyright Place

<b>Categoria</b>	<b>Subcategoria</b>
	22.4 Copyright Date
23. Ownership/Collecting History	23.1 Provenance Description
	23.2 Transfer Mode/Method of Acquisition
	23.3 Cost or Value
	23.4 Legal Status
	23.5 Owner/Agent
	23.6 Ownership Place
	23.7 Ownership Date
	23.8 Owner's Numbers
	23.9 Owner's Credit Line
24. Exhibition/Loan History	24.1 Exhibition/Loan Description
	24.2 Exhibition Title or Name
	24.3 Exhibition Type
	24.4 Exhibition Curator
	24.5 Exhibition Organizer
	24.6 Exhibition Sponsor
	24.7 Exhibition Venue
	24.8 Exhibition Object Number
	24.9 Exhibition Object/Work/Label/identification
25. Cataloging History	25.1 Cataloging Institution
	25.2 Cataloger Name
	25.3 Cataloger Action
	25.4 Area of Record Affected
	25.5 Cataloging Date
	25.6 Object/WORK ID
	25.7 Cataloging Language
26. Related Visual Documentation	26.1 Image References
	26.2 Image Label/Identification
<b>27. Related Textual References</b>	27.1 Citations for Sources
	<b>27.2 Brief Citation</b>

<b>Categoria</b>	<b>Subcategoria</b>
	<b>27.7.2 Source Full Citation</b>
<b>28. Person/Corporate Body Authority</b>	28.1 Person Authority Record Type <b>28.2 Person/Corporate Body Name</b> <b>28.2.8 Name Source</b> <b>28.3 Display Biography</b> <b>28.4 Birth Date</b> <b>28.5 Death Date</b> 28.6 Birth Place 28.7 Death Place <b>28.8 Person Nationality/Culture/Race</b> 28.9 Gender <b>28.10 Life Roles</b> 28.11 Person/Corporate Body Event 28.12 Related Person/Corporate Body 28.13 Person/Corporate Body Broader Context 28.14 Person/Corporate Body Label/Identification 28.15 Person/Corporate Body Descriptive Note
<b>29. Place/Location Authority</b>	29.1 Place Authority Record Type <b>29.2 Place Name</b> <b>29.2.8 Place Name Source</b> 29.3 Geographic Coordinates <b>29.4 Place Types</b> 29.5 Related Places <b>29.6 Place Broader Context</b> 29.7 Place/Location Label/Identification 29.8 Place/Location Descriptive Note
<b>30. Generic Concept Authority</b>	30.1 Concept Authority Record Type <b>30.2 Generic Concept Term</b> <b>30.2.8 Term Source</b> 30.3 Related Generic Concepts



<b>Categoria</b>	<b>Subcategoria</b>
	<b>30.4 Concept Broader Context</b>
	30.5 Generic Concept Label/Identification
	<b>30.6 Concept Scope Note</b>
	<b>30.6.1 Note Source</b>
<b>31. Subject Authority</b>	31.1 Subject Authority Record Type
	<b>31.2 Subject Name</b>
	<b>31.2.8 Name Source</b>
	31.3 Subject Date
	31.4 Subject Roles/Attributes
	31.5 Related Subject
	<b>31.6 Subject Broader Context</b>
	31.7 Related Place/Location
	31.8 Related Person/Corporate Body
	31.9 Related Generic Concept
	31.10 Subject Label/Identification
	31.11 Subject Descriptive Note
	31.12 Subject Authority Record ID

FONTE: Harpring (2019).

#### 4.3.9.3 Metadados para a autenticidade

A partir das categorias apresentadas na tabela anterior, identificam-se 51 elementos e subelementos que podem contribuir para a autenticidade dos recursos, dos quais 20 se prendem com a identidade, e 31 com a integridade. No que concerne à identidade, elegem-se nove categorias: *physical description, place/location authority, measurements, styles/periods/groups/movements, orientation/arrangement, subject matter, context, descriptive note* e *person/corporate body authority*, às quais se acrescentam 11 subcategorias: *creation date, dimensions description, materials/techniques description, catalog level, object/work type, classification term, general subject terms, title text, creator description, creator identity* e *creator role*. A data de criação, as medidas (incluindo a descrição das dimensões), os materiais e as técnicas usadas na concepção dos objetos, a descrição física e a indicação do local de criação são aspetos que contribuem para caracterizar os recursos enquanto objetos físicos localizados no espaço e no tempo. O nível de catalogação, o tipo de obra, o termo de classificação, a caracterização do sistema de classificação, dos movimentos artísticos bem como do contexto cultural e do assunto, e ainda uma nota descritiva que aborde os problemas derivados da obra, são elementos que ajudam a descrever a estrutura da

agregação, a caracterizar a natureza dos objetos no seio dessa agregação e a contextualizar os movimentos intelectuais dos quais os objetos resultam. Por fim, o título da obra, a identidade, o papel e a descrição do criador contribuem para estabelecer a proveniência das obras.

A respeito da integridade, salientam-se 11 categorias: *materials/techniques, inscriptions/marks, state, facture, condition/examination history, conservation/treatment history, related works, related visual documentation, current location, related textual references e ownership/collecting history*. A indicação da localização atual é importante para explicitar a entidade responsável pela custódia, bem como onde essa custódia se realiza. A menção de obras relacionadas, quer sejam visuais ou textuais, é útil para estabelecer elos entre os recursos que reforcem a integridade do agregado. A indicação dos materiais e das técnicas, de inscrições ou marcas, do estado de transmissão da obra e das características de execução (*facture*) remetem para a asserção de alterações físicas sofridas pelas obras ao longo do tempo, que podem condicionar a sua interpretação. A história de eventuais intervenções dos curadores é ainda detalhada nas histórias da examinação e da conservação. Destaca-se ainda a história da custódia como um lugar de destaque para descrever a história da transmissão dos objetos ao longo do tempo.

Por último, a respeito do controlo dos metadados, assinala-se a categoria *cataloging history* nos quais podem ser indicadas informações como a instituição responsável pela criação dos metadados, a ação (criação, revisão, etc), a data, a linguagem e observações.

## 4.3.10 RiC

### 4.3.10.1 Contexto histórico

Por fim, após revistas várias normas de metadados e adquirida uma visão de conjunto dos elementos de metadados mais comumente usados para a asserção da autenticidade, considera-se pertinente proceder a uma análise da versão *draft* da norma internacional *Records in Contexts* (RiC), de modo a incluir nessa análise uma dimensão crítica que se considera porventura ainda mais importante quando uma norma se encontra em fase de revisão e, como tal, numa fase em que os comentários são mais oportunos.

A RiC é uma norma internacional em desenvolvimento pelo ICA com o objetivo de descrever os documentos de arquivo de acordo com os princípios arquivísticos. O ICA constituiu em finais de 2012 um *Experts Group on Archival Description* (EGAD), incumbindo-o, entre 2012-2016, com a tarefa de *developing a comprehensive descriptive standard that reconciles, integrates, and builds on the four existing standards: ISAD(G), ISAAR(CPF), ISDF e ISDIAH* (International Council on Archives, 2016, p. 1). De modo a responder ao difícil desafio de desenvolver uma norma de aplicação internacional, foi composta uma equipa de 21 membros de 13 países de forma a representar a diversidade de teorias e de práticas arquivísticas.

O resultado chegado foi um rascunho de norma em duas partes, um modelo concetual para a descrição arquivística (RiC-CM) e uma ontologia (RiC-O). No âmbito desta análise, o enfoque é

colocado unicamente na RiC-CM, uma vez que se prende mais diretamente com a descrição arquivística no sentido em que procura identificar e definir as entidades descritivas primárias e as suas inter-relações.

Uma importante característica da RiC é a de abordar a descrição dos documentos digitais, sendo desenvolvida tendo presente os vários contextos que os documentos digitais assumem durante o seu ciclo de vida, cuja ideia se encontra cristalizada na sua própria denominação: *records in contexts*, intencionalmente no plural. A RiC serve como afirmação internacional da ideia de que a identidade da informação (arquivística, mas não só) reside na compreensão do contexto, sendo fundamental garantir o acesso e a capacidade de reconstrução desse contexto no longo prazo. Numa época em que os tradicionais princípios arquivísticos foram postos em causa<sup>90</sup>, a captura do contexto — ou, melhor dizendo, dos vários contextos nos quais os documentos ocorrem — parece encontrar-se num processo de consolidação como um dos baluartes da descrição arquivística. Dado que a garantia da autenticidade da informação se trata de componente fundamental da preservação do contexto, entende-se que a perspetiva adotada nesta tese é reforçada pela abordagem que a RiC sugere.

O modelo concetual ainda se encontra em fase de desenvolvimento, estando disponível uma primeira versão *draft* 0.1 de 2016, e a versão 0.2, incompleta, de 2019. Dada a natureza provisória dos documentos, a análise é necessariamente ela mesma provisória. Ainda assim, considerando a importância do ICA enquanto instituição normalizadora, e da missão da RiC de se afirmar como uma norma internacional abrangente no contexto digital, considera-se pertinente incluir esta norma na análise, ainda que os elementos de metadados possam vir a sofrer alterações durante a redação da versão final. Na versão 0.2 ainda se encontra ausente a introdução, diagramas e apêndices. Porém, é apresentado o modelo conceptual atualizado face à versão 0.1, que serve de base à análise da estrutura de metadados. Além disso, a introdução da versão 0.1 serve como fonte para a contextualização histórica e para a interpretação dos princípios orientadores da norma.

O modelo concetual da RiC é um modelo entidade-relações, estando na versão 0.2 previstas 22 entidades e 41 propriedades. Em termos gerais, informação adicional é escassa porque se encontra ainda em *draft*. Perante estes dados, a leitura não pode ser feita nem de forma completa, nem de forma consolidada, faltando ainda informação sobre quais os elementos obrigatórios ou opcionais, e sobre regras de descrição. Porém, uma interpretação da introdução da versão 0.1 permite enumerar algumas das principais características teóricas da RiC face ao processo de descrição:

- **aproximação entre comunidades:** *in the world of analogue records, cooperation and collaboration between records managers and archivists was highly desired ... The*

---

<sup>90</sup> São, entre outros, os princípios do respeito pelo fundo e o princípio da ordem original. A anterior revisão de literatura contribuiu para começar a esboçar uma certa incompatibilidade destes princípios com a descrição arquivística em contexto digital.

*explosive growth of electronic records makes such collaboration and cooperation not merely desirable but an absolute necessity.*

- **expansão do conceito de proveniência:** *provenance is much more complex than the long established understanding of fonds: a fonds does not exist in isolation, but within layers of interconnected contexts, past, present, and future.*
- **a consciencialização do papel subjetivo do arquivista enquanto um interveniente parcial no processo de construção da memória:** *archivists observe that the archival perspective is one among many possible ... that they themselves are performing their jobs in a particular historical (cultural, social, material) context, that their judgements and acts are shaped and informed by the contexts within which they live and work.*
- **descrição multidimensional, em grafo:** *The multidimensional model thus enables the description of the fonds, but also sees the fonds existing in a broader context, in relation to other fonds (International Council on Archives, 2016, pp. 2-10).*

#### 4.3.10.2 Estrutura de metadados

Na versão 0.2, são indicadas 22 entidades e 41 atributos. As entidades são: *Thing, Record Resource, Record Set, Record, Record Part, Instatiation, Agent, Person, Group, Family, Corporate Body, Position, Mechanism, Event, Activity, Rule, Mandate, Date, Single Date, Date Range, Date Set e Place*. Cada uma destas entidades é caracterizada de acordo com (i) *ID*; (ii) *name*; (iii) *definition*; (iv) *scope notes*; (v) *examples* e (vi) *comments*.

Por sua vez, o uso de cada um dos 41 atributos é clarificado através da indicação de (i) *ID*; (ii) *name*; (iii) *definition*; (iv) *grouped in*; (v) *domain* e (vi) *specifications*; (vii) *extensibility*; (viii) *repeatability*; (ix) *value schema*; (x) *scope notes* e (xi) *examples*.

Recorda-se que, nesta versão provisória de dezembro de 2019, ainda não existe indicação de quais elementos são obrigatórios ou opcionais. Porém, é possível observar que existe uma hierarquia entre entidades, segundo a qual as entidades de níveis inferiores herdam os elementos de metadados do nível superior. Por exemplo, considerem-se as entidades *Thing, Record Resource, Record Set*, de níveis um, dois e três, respetivamente. Os atributos de *Thing* são herdados pelo *Record Resource*, cujos atributos são por sua vez herdados pelo *Record Set*. Assim, o terceiro nível incorpora todos os atributos definidos nos primeiros e nos segundos níveis, aos quais se acrescentam eventuais atributos que possam ser considerados pertinentes indicar num terceiro nível.

Como tal, os três atributos da entidade *Thing* são, por este motivo, partilhados por todas as entidades: *Descriptive Note, Identifier* e *Name*. Uma vez que estes atributos ocorrem em todos os níveis de descrição, parece que, no mínimo, serão elementos obrigatórios. Considera-se importante realçar que a RiC v0.2 parece indicar que o uso de uma designação, de um identificador e de uma nota descritiva são elementos inegociáveis para qualquer descrição.

Para além destes atributos comuns a todas as entidades, para cada nível hierárquico das entidades é indicado um conjunto de atributos partilhado pelo respetivo subnível. A estrutura hierárquica das entidades pode ser consultada em International Council on Archives (2019, p. 21). Uma lista das entidades bem como dos atributos que as compõem é apresentada na Tabela 55. De forma a evitar repetições, omite-se a indicação em cada herança dos atributos da entidade *Thing*, uma vez que estes três atributos (A16, A22 e A28) se aplicam a todas as entidades.

Tabela 55 — Entidades e atributos presentes na RiC v0.2.

Entidade	Herança	Atributos
E01 Thing		A16 Descriptive Note A22 Identifier A28 Name
E02 Record Resource		A03 Authenticity Note A07 Classification A08 Conditions of Access A09 Conditions of Use A10 Content Type A21 History A24 Integrity A25 Language A35 Record Resource Extent A38 Scope and Content A39 State A40 Structure A26 Legal status
E03 Record Set	Todos da E02 mais:	A01 Accrual A36 Record Set Type
E04 Record	Todos da E02 mais:	A17 Documentary Form Type
E05 Record Part	Todos da E02 mais:	A17 Documentary Form Type
E06 Instantiation		A03 Authenticity Note A04 Carrier Extent A0 Carrier Type A08 Conditions of Access

<b>Entidade</b>	<b>Herança</b>	<b>Atributos</b>
		A09 Conditions of Use A21 History A23 Instantiation Extent A24 Integrity A31 Physical Characteristics A33 Production Technique A34 Quality of Representation A37 Representation Type A40 Structure
E07 Agent		A21 History A25 Language A26 Status
E08 Person	Todos da E07 mais:	A15 Demographic Group A30 Occupation Type
E09 Group	Todos da E07.	
E10 Family	Todos da E07 mais:	A20 Family Type
E11 Corporate Body	Todos da E07 mais:	Corporate Body Type
E12 Position	Todos da E07.	
E13 Mechanism	Todos da E07 mais:	A41 Technical Characteristics
E14 Event		A06 Certainty A18 Event Type A21 History
E15 Activity	Todos da E14 mais:	A02 Activity Type
E16 Rule	A21 History	-
E17 Mandate	A21 History	-
E18 Date		A06 Certainty A13 Date Qualifier A14 Date Standard A19 Expressed Date

Entidade	Herança	Atributos
		A29 Normalized Date
E19 Single Date	Todos da E18.	-
E20 Date Range	Todos da E18.	-
E21 Date Set	Todos da E18.	-
E22 Place		A11 Coordinates A21 History A27 Location A32 Place Type

FONTE: Adaptado de International Council on Archives (2019).

#### 4.3.10.3 Metadados para a autenticidade

Uma análise da tabela anterior permite apontar 18 entidades e atributos que mais diretamente podem servir para estabelecer a autenticidade dos recursos, enquadrando-se 11 na asserção da identidade e 7 na verificação da integridade.

No que diz respeito à identidade dos recursos, salientam-se duas entidades: *Date* e *Agent* e nove atributos: *identifier*, *name*, *record resource extent*, *documentary form type*, *record set type*, *classification*, *scope and content*, *structure* e *history*. O identificador, o nome, as datas, a dimensão e o tipo documental afirmam-se como elementos fundamentais para a caracterização dos documentos enquanto uma entidade física. O tipo do agregado documental, a classificação, o âmbito e o conteúdo e a estrutura são úteis para a descrição da estrutura da agregação, da sua composição e da natureza da classificação, ao mesmo tempo que clarifica a natureza dos documentos no contexto dessa agregação. Por fim, o agente e a história podem servir para estabelecer a proveniência ao identificar o criador e eventos relacionados com a proveniência do agregado.

Relativamente à integridade, é de destacar a entidade *place* e seis atributos: *integrity*, *state*, *accrual*, *physical characteristics*, *production technique* e *authenticity note*. A localização pode informar sobre a custódia dos recursos, ao passo que as adições revelam intervenções dos arquivistas ao agregado documental. Além disso, as notas sobre a integridade e a autenticidade, juntamente como o estado de reprodução e as características físicas e técnicas contribuem para descrever a transmissão dos recursos no tempo e no espaço.

É de notar a ausência metadados que possam ser aplicados diretamente ao controlo das descrições. Este aspeto é de certa forma surpreendente, dado o reconhecimento pela RiC da subjetividade do arquivista enquanto autor de uma de múltiplas perspetivas possíveis sobre a descrição dos recursos. Esta surpresa é realçada pelo facto de este tipo de metadados ocorrer com frequência nas restantes normas analisadas, incluindo na própria ISAD(G), que é uma das

antecessoras da RiC. Talvez este tipo de informações sobre os próprios metadados — tais como autor da descrição, data da descrição, regras ou convenções adotadas — possam ser incluídas através da entidade *Thing*, que é intencionalmente generalista. Porém, mesmo que seja este o caso, não é incluído nos exemplos de uso desta entidade informações sobre os meta-metadados, o que parece indicar que o ICA não previu explicitamente a entidade *Thing* como lugar para o controle das descrições.

Além disso, destaca-se a entidade *Instantiation* como um sinal claro de que a RiC foi redigida em contexto digital, uma vez que esta entidade reconhece as diferentes manifestações que a informação pode assumir quando é comunicada no espaço e no tempo, por exemplo quando um documento é impresso, gravado como docx ou pdf.

Contudo, os restantes metadados sobre a identidade e a integridade dos recursos enquadram-se na linha das anteriores normas, com a exceção já assinalada de uma aparente ausência de metadados sobre o controle das descrições. Esta observação parece reforçar que o principal contributo da RiC se prende não tanto com uma eventual inovação ao nível dos elementos de metadados propostos para a descrição dos recursos, mas sobretudo através do modo como esses elementos de metadados podem ser repetíveis em diferentes entidades de modo a espelhar diferentes perspetivas conforme a entidade que descrevem, contribuindo assim para a multidimensionalidade da representação da informação, por oposição aos anteriores modelos hierárquicos, multinível, usados em normas como a ISAD(G).

#### **4.3.11 Sumário**

De forma a resumir os principais achados das últimas secções no que diz respeito à identificação dos requisitos a partir da literatura técnica, agregam-se na Tabela 56 os elementos que em cada norma se selecionaram como úteis para a asserção da autenticidade dos recursos e das próprias descrições. Assim, esta tabela recupera as informações que, para cada norma, se incluíram na secção “Metadados para a autenticidade”. Para cada norma, os elementos são estruturados em três conjuntos, separados por um espaço. O primeiro conjunto diz respeito aos elementos para a identidade dos recursos, o segundo agrupa os elementos relacionados com a integridade dos recursos, enquanto o último grupo se refere aos elementos relacionados com o controle das descrições.



Tabela 56 — Sumário dos elementos para a autenticidade identificados em cada norma.

ISAD(G)	AGRkMS	EAD	e-EMGDE	DACS	DCMES	VRA	MODS	CDWA	RiC
Reference code	Name	agencycode	Identificador	Reference code	Coverage	Date	dateissued	Creation Date	Identifier
Dates	Identifier	recorded	Nombre	Title	Date	Material	dateCreated	Measurements	Name
Extent and medium	Date range	unitid	Fechas	Date	Format	Measurement	dateCaptured	Dimensions	Date
Level of description	Extent	unitdate	Tamaño	Extent	Identifier	s	physicalDescript	description	Record Resource Ext.
System of arrangement	Medium	unitdatestructured	Soporte	Scope and content	Subject	Description	ion	Materials/techniques	Documentary Form
Document	Document	physdesc	Tipo	System of arrangement	Creator	Subject	Identifier	description	Record Set Type
Scope and content	Form	physdescstructured	documental	arrangement	Contributor	CulturalConte	Classification	Physical description	Classification
Title	Category	archdesc	Categoría	Name of creator(s)	Provenance	xt	targetAudience	Place/location	Scope and Content
Name of Creator	Description	c	Descripción	Name and location	Title	StylePeriod	abstract	authority	Structure
Administrative/biographical history	Coverage	scopecontent	Ámbito	of repository	accrualMethod	Title	genre	Catalog level	Agent
Related entity (relationship)	Related entity (relationship)	unittitle	Entidad	Administrative/biographical history	accrualPeriodicity	WorkType	subject	Object/Work Type	History
Existence and location of originals	Change history	origination	Relacionada	Clasificación	accrualPeriodicity	Agent	titleInfo	Classification term	Integrity
Existence and location of copies	Name	bioghist	Historia del	Custodian history	accrualPolicy	Source	name	Styles/Periods...	Integrity
Physical characteristics and technical requirements	Disposal	custodhist	cambio	Immediate source	Relation	Location	typeofResource	Orientation/Arrangement	State
Related unit of description	Location	acqinfo	Nombre (Agente)	Appraisal, destruction, and scheduling information	Source	Inscription	dateModified	Subject matter	Accrual
Source of acquisition	Integrity check	appraisal	Órgano	Accruals	Publisher	Relation	relatedItem	General subject terms	Physical characteris.
Archival history	Integrity check	phystech	Origen del documento	Existence and location of originals	StateEdition	Technique	location	Context	Production technique
Appraisal, scheduling and destruction information	Integrity check	originalsloc	Dictamen	Existence and location of originals	Technique	textref	recordInfo	descriptive note	Place
Archivist's note	Contact (Agent)	altformavail	Verificación de integridad	Related archival materials	textref	-	-	Title text	Authenticity Note
Rules or conventions	Contact (Agent)	relatedmaterial	Firma	Notes	-	-	-	Creator identity	-
Dates of descriptions	Contact (Agent)	separatedmaterial	Estado de elaboración	Description control	-	-	-	Creator role	-
		processinfo	Localización					Creator description	
		conventiondeclaration	Verificación de integridad					Person/corporate body authority	
		maintenanceevent	Firma					Materials/techniques	
		eventdatetime	Estado de elaboración					Inscriptions/marks	
		maintenanceagency	Estado de elaboración					State	
		maintenancehistory	Contacto					Facture	
		maintenancestatus	Contacto					Condition/examination history	
								Conservation/treatment history	
								Related works	
								Related visual documentation	
								Related textual references	
								Current location	
								Ownership/collecting history	
								Cataloging history	

## 5 Construção do modelo de requisitos

---

Neste capítulo procura-se sistematizar os resultados encontrados no capítulo anterior a respeito da identificação de requisitos. O objetivo é, com base nos dados empírico recolhidos, promover uma discussão crítica que precede a proposta de um modelo para a autenticidade fundamentado nos dados de investigação. Assim, num primeiro momento analisam-se e discutem-se os requisitos recolhidos a partir da literatura científica, depois os requisitos recolhidos a partir das normas de metadados para, por fim, se proceder à proposta de um modelo.

### 5.1 Análise e discussão dos requisitos científicos

Esta secção assenta em grande medida nos dados teóricos recolhidos na secção 4.2, com um sumário na Tabela 46. Porém, enquanto nessa tabela os requisitos encontrados foram agrupados por área, independentemente das suas repetições ao longo dos diversos autores, de ora em diante pretende-se sistematizar a informação, agrupando os diferentes autores que defendem o mesmo argumento sob um único requisito. Assim, constrói-se a Tabela 57 que consolida todos os requisitos científicos únicos identificados, que se entendem diretamente pertinentes para a construção de um modelo de elementos de metadados de estrutura para a autenticidade e a confiabilidade dos recursos.

São identificados um total de 32 requisitos, sendo indicado, para cada um, as obras científicas das quais originam. Além disso, é também identificada a categoria analítica sob a qual podem ser classificados, correspondendo a constituição das categorias à análise da revisão da literatura, discutida previamente na metodologia. Sete são as categorias: a Representação da Informação (RI), Descrição Arquivística (DA), Gestão da Informação (GI), Metadados (MD), Normas de Metadados (NMD), Autenticidade (AUT) e Linked Data (LD).

Uma análise desta tabela permite realizar algumas observações. Uma das linhas de análise prende-se com a frequência de ocorrência dos requisitos. Conforme demonstrado pela Tabela 58, os requisitos mais frequentes são o primeiro e o sétimo, com 5 ocorrências cada um (71%), seguindo-se dos requisitos 2, 4, 10 e 17, cada um com 4 ocorrências (57%). Estes resultados parecem indicar que autores de várias orientações teóricas reconhecem a necessidade de maior transparência e responsabilização dos autores das descrições (requisito 1), enquanto reflexo da crescente consciencialização da subjetividade do processo de representação (requisito 10). Como tal, as agregações intelectuais que os arquivistas constroem podem e devem ser questionadas, havendo a possibilidade de os recursos pertencerem simultaneamente a múltiplos agregados e contextos dependendo da sua perspectiva de interpretação (requisito 4). Além disso, é consensual pela maioria que os metadados devem ser recolhidos ao longo do ciclo de vida dos recursos

(requisito 7), que devem incluir informação contextual em abundância e pormenor, como forma de documentação da proveniência (requisito 2) e de reforço da verificação da integridade dos recursos (requisito 17).

Tabela 57 — Requisitos únicos identificados na literatura científica.

#	Requisitos	Tema
1	Maior transparência e <i>accountability</i> nos procedimentos de descrição, devendo o criador dos metadados ser identificado e fornecer explicações, quando necessário (Douglas, 2016; Duff et al., 2013; Duranti et al., 2019; Engvall, 2019; Haynes, 2018; Li & Sugimoto, 2017; MacNeil, 2009; Meehan, 2009; Rogers, 2015; Serewicz, 2010; Wright, 2014; Yeo, 2013).	RI, DA, GI, MD, AUT
2	Inclusão de informação contextual sobre o criador, a história e a proveniência dos documentos de arquivo, esclarecendo a história custodial (Chen et al., 2011; Douglas, 2016; Duff et al., 2013; Duranti et al., 2019; Gladney, 2009; Haynes, 2018; MacNeil, 2009; J. Niu, 2015a; Price & Smith, 2011; Zeng & Qin, 2016; Zhang, 2012b).	RI, DA, MD, AUT
3	Os documentos estabelecem relações não só com membros do seu agregado mas com documentos de outras coleções e repositórios (Duff et al., 2013; Jones, 2018; Price & Smith, 2011; Serewicz, 2010; I. Silva, 2012; Yeo, 2012).	RI, AUT, LD
4	Os documentos podem pertencer a múltiplos agregados, e ser interpretados à luz de diversos contextos (J. Bailey, 2013; Cumming, 2010; Jones, 2018; MacNeil, 2017; Moyano Collado, 2013; J. Niu, 2015a; Rolan, 2015; Samouelian, 2009; Zhang, 2012c).	RI, DA, GI, LD
5	A descrição arquivística deve realizar-se não só ao nível coletivo das coleções, mas também nível do item (J. Bailey, 2013; Moyano Collado, 2013; J. Niu, 2015a; Zhang, 2012b, 2012c).	RI, DA, GI
6	Os instrumentos de pesquisa devem ser <i>user-friendly</i> , construídos como se fossem <i>websites</i> (J. Niu, 2015a).	RI
7	Os metadados devem ser capturados ao longo do ciclo de vida dos recursos, sendo a descrição um processo permanente, incremental, que carece de constante atualização (Baca, 2016; Cumming, 2010; Mckemmish et al., 1999; J. Niu, 2013, 2015a; Yakel, 2003).	RI, DA, GI, MD, NMD
8	As representações devem ser construídas a partir das necessidades dos utilizadores (Anchor, 2013; Serewicz, 2010).	RI, DA
9	Devem ser usados identificadores únicos (Gartner, 2015; Gladney, 2009; Verborgh et al., 2015).	RI, AUT, LD
10	Considerar que as representações arquivísticas são práticas subjetivas, socialmente construídas, e passíveis de questionamento (Douglas, 2016; Duncan, 2009; Duranti et al., 2019; Engvall, 2019; Jones, 2018; MacNeil, 2009, 2012; McLeod & Gormly, 2017; Ridener, 2009; Rogers, 2015; Rolan, 2015; Yakel, 2003; Yeo, 2013).	RI, DA, AUT, LD
11	Manter as descrições do passado como testemunho histórico das narrativas (Yakel, 2003).	RI
12	Devem ser identificados os autores e a versão de cada instrumento de pesquisa, bem como mudanças que os metadados sofreram ao longo do tempo, nomeadamente adições, modificações ou eliminações dos elementos e/ou das suas definições (Apostolou, 2009; Li & Sugimoto, 2017; Yakel, 2003).	RI, MD
13	No contexto digital, deve privilegiar-se a ordem intelectual face à ordem física (Yakel, 2003; Zhang, 2012b).	RI
14	A descrição deve ser completa mas simples o suficiente a ponto de ser viável (Anchor, 2013; Youn, 2015).	DA, NMD

#	Requisitos	Tema
15	Os utilizadores devem poder contribuir para a descrição arquivística mediante mecanismos como comentários, <i>tags</i> , anotações ou revisões (Apostolou, 2009; Douglas, 2016; Samouelian, 2009).	DA, MD, LD
16	Inclusão de informações sobre a avaliação dos documentos, abrangendo planos de eliminação (Haynes, 2018; J. Niu, 2013; Rogers, 2015).	DA, MD
17	Mecanismos de verificação da integridade e da proveniência, nomeadamente <i>audit trails</i> , registo de versões, assinaturas digitais ou <i>checksums</i> (Bountouri et al., 2017; Duranti et al., 2019; Gladney, 2007, 2009; Hedstrom, 1993; Li & Sugimoto, 2017; Mckemmish et al., 1999; J. Niu, 2013; Price & Smith, 2011; Rogers, 2015; Zeng & Qin, 2016).	DA, MD, NMD, AUT
18	A descrição arquivística deve esclarecer sobre características materiais e modificações aos documentos (Acker, 2017; Hedstrom, 1993; Llanes-Padrón, 2019; MacNeil, 2009).	DA, GI
19	Entender que não existe uma estrutura universal para a representação dos recursos, dependendo do contexto de aplicação e das necessidades de cada comunidade (Apostolou, 2009; Gladney, 2007; Haynes, 2018; McLeod, 2014).	GI, MD
20	A principal finalidade dos metadados deve ser refletir o contexto das atividades das quais os documentos são um subproduto (McLeod, 2014; Tough, 2016; Wright, 2014).	GI
21	O objeto da representação deve ser alargado além do tradicional entendimento de documento de arquivo (Wright, 2014; Yeo, 2011).	GI
22	A criação dos metadados deve ser automatizada na medida do possível (Baca, 2016; Yeo, 2013).	MD, AUT
23	Os metadados devem ser criados de forma a que possam ser facilmente partilháveis e reutilizáveis (Baca, 2016; Gartner, 2015, 2016; Hooland & Verborgh, 2014; Jones, 2018; Llanes-Padrón & Moro-Cabero, 2017; Llanes-Padrón & Pastor-Sánchez, 2017; Machado et al., 2019; J. Niu, 2016).	MD, NMD, LD
24	O uso de facetas nas pesquisas é vantajoso (Beyene & Godwin, 2018; Gartner, 2016).	MD
25	Cada recurso deve ter uma e só descrição, segundo o princípio 1:1 (Zeng & Qin, 2016).	MD
26	O mapeamento entre as normas é frequentemente inconsistente (Andrade et al., 2014; Woodley, 2016).	NMD
27	É vantajoso que os modelos sejam de acesso gratuito e apresentem uma redação simples em linguagem compreensível (Dryden, 2009).	NMD
28	As informações sobre a proveniência devem ser recolhidas desde o momento de criação dos recursos (Bountouri et al., 2017).	AUT
29	Os recursos devem ter associada informação suficiente de modo a que sejam auto-descritíveis (Gladney, 2009; Jansen, 2015).	AUT
30	A possibilidade de uso da tecnologia <i>blockchain</i> (Bhatia & Wright de Hernandez, 2019; Hofman et al., 2019; Lemieux, 2016).	AUT
31	A possibilidade de uso de inteligência artificial (Rolan et al., 2019; Yeo, 2013).	AUT, LD
32	Quanto maior a granularidade das descrições, mais facilitada é a interoperabilidade (Gartner, 2015; Jones, 2018; J. Niu, 2016).	LD

FONTE: elaboração própria.

Apesar de se destacarem estes requisitos por serem os que se encontram veiculados pela maioria, isto é, por quatro ou mais temas, isso não significa que os restantes requisitos menos frequentemente mencionados sejam de menor importância. Entende-se que todas as perspectivas devem ser consideradas como um potencial contributo, ainda que haja um só autor a propô-la. Aliás, pode ser argumentado que as sugestões mais raras podem ser potencialmente as mais ricas, uma vez que podem corresponder a sugestões originais e perspectivas inovadoras sobre um problema. Neste sentido, na discussão dos resultados e na proposta do modelo, considera-se a totalidade 32 requisitos encontrados.

Tabela 58 — Requisitos teóricos mais frequentes.

<b>Requisito</b>	<b>Ocorrências</b>	<b>Percentagem</b>
1	5	71%
7	5	71%
2	4	57%
4	4	57%
10	4	57%
17	4	57%

FONTE: elaboração própria.

Posto isto, é necessária uma discussão dos dados, que deve começar por duas ressalvas. A primeira, quanto à própria constituição das categorias. A designação das categorias obedece a generalizações com base nos dados recolhidos durante a revisão de literatura, e com base na identificação subjetiva das principais ideias veiculadas pelas obras científicas no contexto dos objetivos da investigação. As observações que se seguem não devem ser interpretadas de um modo absoluto, com uma presunção de conclusão definitiva sobre as áreas referidas. Conforme mencionado, a própria constituição das áreas é subjetiva, sendo as designações formulações dos investigadores, que podem ser insuficientes para capturar a complexidade dos temas sob outras investigações com outros objetivos. Ademais, há casos em que o mesmo autor apresenta artigos diferentes classificados pelos investigadores em categorias distintas. Embora os artigos se posicionem em categorias diferenciadas, o facto de um mesmo indivíduo ser comum a várias categorias pode contribuir para a repetição dos requisitos em diferentes temas, uma vez que a forma de pensar é idêntica.

A segunda ressalva diz respeito à própria subjetividade de seleção dos requisitos. Frequentemente longos, complexos e ricos artigos foram simplificados, tendo-se procurado extrair as ideias (requisitos) que se entendem como possíveis contributos para a construção de um eventual modelo para a autenticidade e confiabilidade dos recursos, no âmbito da descrição arquivística. Consequentemente, foi necessário selecionar informação, ocultando certas ideias para dar destaque

a outras. Uma primeira seleção ocorreu no capítulo anterior, quando se analisaram em pormenor, a nível individual, as obras científicas, tendo uma segunda seleção ocorrido na constituição da Tabela 57, referida acima. Em ambos os momentos, procurou-se realçar os requisitos considerados potencialmente úteis para a elaboração do modelo, de acordo com os objetivos de investigação.

Apesar de estas reservas, entende-se que uma discussão dos resultados obtidos é merecida. Uma leitura global dos tipos de requisitos identificados por cada tema permite observar algumas tendências específicas a cada categoria. No âmbito da literatura sobre a representação da informação, é possível notar uma aguda consciência da subjetividade das representações, e da possibilidade de construção de várias formas de representação, isto é, da constituição de vários controlos intelectuais que não têm necessariamente de replicar o controlo físico dos documentos. Esta observação parece ser natural, uma vez que autores que trabalham com representações conhecem geralmente que as representações são uma de múltiplas possíveis formas de apresentação. Aplicando o mesmo raciocínio ao conhecimento arquivístico, se as descrições arquivísticas são formas de representação, parece natural entender determinado instrumento de pesquisa como uma de várias possibilidades.

Esta subjetividade é também observada na área da descrição arquivística, não fosse esta também uma forma de representação da informação, embora aqui se note um maior apelo à explicitação do papel do arquivista na construção das descrições, talvez porque os autores estão mais diretamente inseridos no contexto arquivístico. Constata-se também que estes autores procuram em maior pormenor particularizar em concreto os elementos de metadados através dos quais a descrição arquivística deve ser operacionalizada, incluindo designações, definições e exemplos dos elementos em uso. Os autores também tendem a estabelecer a separação entre controlo intelectual e controlo físico dos documentos, à semelhança dos autores da representação da informação. Estes aspetos parecem refletir que quem trabalha com a descrição arquivística tende a sugerir propostas concretas de metadados que possam ser aplicadas imediatamente na construção de instrumentos de pesquisa. Embora seja importante para a discussão conhecer em pormenor, sem ambiguidades, os elementos de metadados e as suas definições, tais propostas não devem ser imediatamente implementadas sem um espírito crítico, uma vez que as descrições arquivísticas são geralmente criadas de forma a corresponder às necessidades dos utilizadores de um determinado contexto, pelo que se torna essencial compreender esse contexto para determinar quais os elementos de metadados mais pertinentes.

A respeito da gestão da informação, nota-se uma tendência para realçar a importância de manifestar as relações entre os documentos e as atividades das quais resultam, documentando a influência dos sistemas de informação na criação dos documentos. Além disso, também se destaca neste tema o questionamento entre o que constitui um documento de arquivo (*record*), registando-se uma defesa da expansão do termo. Ao contrário da descrição arquivística, os autores geralmente não procedem a propostas concretas de elementos de metadados, preferindo a formulação de orientações abstratas e teóricas de como os recursos/objetos/documentos devem ser geridos ao longo do seu ciclo de vida. Este tipo de discurso pode revelar-se mais difícil para as organizações

acompanharem, uma vez que se torna necessário um profundo domínio das teorias de forma a esboçar uma tentativa de implementação.

No âmbito da literatura em metadados, volta a notar-se uma tendência para a procura da especificidade das orientações, através da sugestão de elementos de metadados concretos para a implementação. Nota-se assim uma coincidência entre esta abordagem e a da descrição arquivística, o que não surpreende uma vez que a descrição arquivística se baseia, na prática, pela definição de uma estrutura de metadados. Também se observa neste tema um destaque da importância dos metadados enquanto forma de estabelecer a proveniência dos documentos, procurando assim legitimar-se a investigação. Ocasionalmente, a literatura também discute outras dimensões que não apenas a estrutura dos metadados, pronunciando-se por vezes sobre a sua semântica e sintaxe. Estas vertentes dos metadados foram abordadas na revisão da literatura, posicionando os objetivos de investigação sobretudo em relação à estrutura dos metadados.

A tendência de realçar a semântica, enquanto requisito para interoperabilidade, é particularmente evidente na literatura sobre normas de metadados. Os autores tendem a procurar explicitamente definir requisitos para a interoperabilidade e de construir mapeamentos entre os elementos usados por várias normas focadas na partilha da informação. Assim, estes autores geralmente procuram conciliar as normas existentes, estabelecendo elos, paralelos, e criando pontes entre as diferentes normas de metadados e de como estas devem ser interpretadas.

Relativamente à autenticidade, enquanto nas outras áreas se encontram afirmações genéricas sobre a importância de documentar a proveniência como parte da presunção da autenticidade, os textos classificados sob esta categoria refletem explicitamente sobre a autenticidade como conceito, problematizando-o e questionando as práticas pelas quais se interpreta que algo é autêntico. A autenticidade não é vista como um dado adquirido, nem como um termo com uma definição consensual nem bem consolidada, mas antes como um conceito que merece discussão e interpretação. Por isso, os autores tendem a revelar uma preocupação explícita em estabelecer bases sobre as quais a confiança possa ser estabelecida, que carece da demonstração da autenticidade, que por sua vez assenta na demonstração da identidade e da integridade dos recursos, dos metadados, e dos próprios criadores e custodiantes, quando aplicável. Além disso, os autores tendem a realçar a subjetividade da confiança, que reconhecem não se tratar de um conceito binário (é ou não é), mas antes de uma decisão individual com base nas informações disponíveis. Nota-se ainda que os autores deste tema apresentam geralmente uma formação base ou experiência na Diplomática, o que não surpreende dada a relação umbilical desta área com a necessidade de demonstração da autenticidade dos documentos de arquivo desde o período medieval. Porém, os tradicionais procedimentos diplomáticos de análise da forma documental enquanto metodologia para asserção da autenticidade podem ser raciocínios traiçoeiros no ambiente digital, uma vez que se tem vindo a perceber que os mecanismos sobre os quais se presume a autenticidade dos documentos digitais não pode mais corresponder aos mesmos princípios seguidos para os documentos analógicos. Um exemplo é dado pelo facto de os documentos digitais não apresentarem necessariamente a forma documental facilmente identificável que a Diplomática tradicionalmente

requer. Estas ressalvas procuram aconselhar uma leitura prudente de alguns autores que, inspirados ainda por uma veia diplomática, procuram encapsular certos objetos digitais em formas documentais fixas e estáveis, um esforço por vezes difícil no contexto digital.

Por fim, uma das principais leituras a retirar do tema *linked data* é a de que o conhecimento se constrói em rede e que para construir a rede é necessária a criação de metadados ao nível do item de forma que as descrições de recursos de diferentes comunidades possam ser agregadas. Começam a encontrar-se nesta categoria vários autores que, embora escrevam sobre a partilha de informação em arquivos, não são necessariamente arquivistas, trabalhando por vezes em museus, bibliotecas ou noutras instituições. O ponto comum é a ideia da necessidade da integração das coleções de forma a promover a construção em rede do conhecimento de uma forma que melhor reflita a forma como este se origina na sociedade. A literatura inserta nesta categoria é em larga medida consensual em apelar, direta ou indiretamente, a uma maior integração das coleções de diferentes instituições culturais, não só como reconhecimento de que os recursos podem pertencer a múltiplos contextos, mas também como reflexo do entendimento que, quanto mais interligadas forem as relações semânticas entre os recursos, maior informação contextual se manifesta, e mais ricas e diversas se tornam as interpretações possíveis.

Com base na interpretação dos aspetos comuns e das especificidades manifestadas por cada categoria da análise da literatura científica, torna-se possível proceder a uma identificação provisória de requisitos para a autenticidade, apresentados na Figura 15, que inclui algumas orientações teóricas para a representação, a par de elementos de metadados concretos. Esta figura pretende sistematizar os requisitos que, a partir da leitura e análise da literatura científica, se adequam aos objetivos de investigação.

#### Orientações para a representação

- usar taxonomias para preenchimento automático..... 6
- além de descrições ao nível da coleção, realizar também descrição ao nível do item..... 3, 4, 5, 32
- criação de metadados enquanto *linked data*..... 3, 4, 23
- automatizar sempre que possível a criação de metadados..... 2, 4, 22, 23
- assemelhar os instrumentos de pesquisa a *websites*, de preferência com múltiplas formas de visualização dos dados..... 6
- uso de metadados facetados..... 4
- tolerância nas pesquisas a erros gramaticais..... 6
- realizar a descrição segundo o princípio 1:1..... 4, 9, 25
- usar linguagem compreensível, não-técnica..... 6, 27
- não incluir demasiados elementos, ser realista e prático..... 14
- a descrição deve ser vista como um processo incremental, permanente..... 7
- possibilidade de gerar citações automáticas sobre os recursos..... 6
- criar um espaço para *input* dos utilizadores..... 8, 15
- criar as representações em função das necessidades dos utilizadores..... 8, 13, 19, 20
- privilegiar a ordem intelectual face ao controlo físico..... 13
- o objeto de representação deve ir além dos documentos de arquivo..... 21
- as informações de proveniência devem ser recolhidas pelo menos desde a criação dos recursos..... 1, 2, 28

#### Recurso

- Título..... 2
- Identificador (único)..... 9
- Data de criação..... 2
- Criador(es)..... 2
- Descrição/assunto..... 2
- Formato..... 2
- Localização..... 2
- Direitos de uso..... 2
- Versão..... 12, 17, 18
- História custodial..... 2
- Nome(s) do(s) custodiante(s)..... 2
- História da classificação..... 2
- Descrição material..... 18
- Avaliação..... 16
- Língua..... 2

#### Criador

- Nome..... 2
- História administrativa..... 2

#### Controlo das representações

- Autor/criador dos metadados..... 1
- Data de criação ou de última modificação dos metadados..... 2
- Histórico de alterações aos elementos de metadados..... 12
- (adições, eliminações e/ou substituições)..... 11, 12, 18
- Notas do criador da descrição..... 1
- *Curriculum Vitae* do criador da descrição..... 10



Figura 15 — Elementos de metadados e orientações derivados da literatura científica. O número indica os requisitos que satisfazem.

FONTE: elaboração própria.

Os requisitos apontados previamente parecem manifestar-se em quatro dimensões: em metadados relativos ao recurso, em metadados sobre o criador dos recursos, em metadados referentes ao controlo das representações, e naquilo que se consideram ser orientações gerais para a representação. Emprega-se o termo ‘recurso’ devido à sua universalidade, podendo capturar qualquer conceito ou objeto que possa ser alvo de representação. O número indicado na figura ao lado de cada elemento de metadados corresponde ao requisito da Tabela 57 com que se identifica. Procurou-se que cada um dos 32 requisitos seleccionados se traduza numa proposta concreta para implementação, de modo a capturar de forma completa o universo de análise.

Não se procederá agora a uma definição detalhada dos elementos de metadados, uma vez que as designações são relativamente autoexplicativas, reservando-se esse exercício para a proposta final do modelo. Porém, é necessário fazer algumas observações quanto às orientações sugeridas, de forma a contextualizá-las. Em primeiro lugar, embora o uso de taxonomias diga mais diretamente respeito à partilha da informação, considera-se pertinente incluir esta sugestão porque, na prática, este aspeto será fundamental para a interoperabilidade das descrições arquivísticas. Similarmente, a tolerância a erros gramaticais na realização das pesquisas, bem como a recomendação para que os instrumentos de pesquisa estejam estruturados enquanto *websites* com *interfaces* simples e intuitivos, têm como objetivos promover uma experiência positiva do uso dos instrumentos de pesquisa. Uma experiência dos utilizadores positiva é um aspeto crítico da utilização de qualquer serviço, devendo as próprias descrições arquivísticas ser construídas de uma forma que facilitem e incentivem a interação. O uso de uma linguagem natural, facilmente compreensível ausente de jargão técnico, é outro aspeto que se considera fundamental para não criar barreiras na interpretação das descrições.

Ainda de modo a facilitar uma experiência positiva dos utilizadores, entende-se que o uso de metadados facetados contribui para um *positive search abandonment*, o que significa que os resultados da pesquisa devem, por si só, serem suficientes para os utilizadores decidirem se as suas necessidades de informações iniciais foram satisfeitas, se precisam de realizar uma nova pesquisa ou se o repositório possui os documentos ou a informação que procurava. Além disso, considera-se útil que os *websites* que manifestam os instrumentos de pesquisa incluam um espaço para *feedback* dos utilizadores, quer sob a forma de comentários, avaliações ou com outra informação que explicita quais os recursos mais consultados, ou relações entre recursos (como por exemplo, sugestões que indiquem que quem consulta X também consulta Y, inferências que retalhistas como a Amazon inovaram).

Nota-se ainda que, apesar da tentação de incluir o máximo de metadados possível de modo a satisfazer a incerteza dos criadores das descrições, ter demasiados elementos de metadados é prejudicial porque inviabiliza os esforços de descrição, que passam a requerer demasiado investimento por parte das organizações. Assim, recomenda-se que deve ser procurado um meio-

termo entre simplicidade e completude, devendo-se procurar incluir metadados suficientes para a documentação da proveniência e da autenticidade, mas não mais do que seja realisticamente possível às organizações descrever. Este processo é facilitado se a criação de metadados for automatizada sempre que possível, sendo ainda um bônus se esses metadados forem criados enquanto *linked data*, isto é, num formato interoperável e *machine-readable* que permita a sua reutilização por outros agregadores para a integração das coleções.

Esta figura constitui um ponto de partida para a construção do modelo de requisitos, representando uma metade do universo de análise. Como tal, incluíram-se todas as potenciais propostas para a constituição da estrutura de metadados. Mais adiante procura-se realizar um raciocínio semelhante para as normas de metadados, de onde se espera que emerjam os requisitos técnicos. Ambas as propostas provisórias serão então comparadas e integradas na proposta final do modelo, que procurará conciliar a economia dos elementos com a sua completude.

## 5.2 Análise e discussão dos requisitos técnicos

A análise e a discussão que se seguem têm predominantemente como base os dados empíricos recolhidos na secção 4.3, cujo sumário é apresentado na Tabela 56. Esta tabela sistematiza os metadados que, a partir da análise de cada uma das nove normas selecionadas, podem contribuir para a asserção da autenticidade.

Como seria de esperar, as normas estudadas apresentam alguns pontos convergentes e divergentes. De modo a sistematizar os requisitos encontrados, identificam-se na Tabela 59 todos os requisitos identificados a partir da análise da estrutura das normas. Na primeira coluna é indicado um código de identificação, correspondendo o ‘T’ à natureza técnica do requisito, enquanto na segunda coluna se utiliza uma etiqueta que funciona como título autoexplicativo do requisito identificado. As colunas seguintes indicam não só as normas a partir das quais os requisitos foram extraídos, como também especificam os elementos de metadados que serviram de base a essa extração.

Todavia, é necessário estabelecer duas ressalvas. Primeiro, os requisitos foram construídos unicamente a partir dos elementos identificados pelos investigadores como sendo apropriados para a asserção da autenticidade dos recursos e das descrições, pelo que houve uma filtragem prévia e subjetiva da amostra. Em segundo lugar, a própria identificação dos elementos considerados adequados para a autenticidade baseou-se na análise dos elementos e subelementos presentes na estrutura das normas. Por outras palavras, não se efetuou uma análise aprofundada de todas as possibilidades de descrição previstas pelas normas. Neste sentido, a Tabela 59 que se segue resulta de duas camadas de interpretação: 1) análise das categorias e das subcategorias de descrição; e 2) de entre essas categorias e subcategorias, análise de quais os elementos para a autenticidade. Como tal, não deve ser entendida como uma representação exhaustiva de todos os requisitos possíveis a retirar de cada norma, mas tão-só como uma interpretação dos metadados para a autenticidade enquadrada no contexto dos objetivos da presente investigação.

Apesar destas condicionantes, uma análise da distribuição dos elementos de metadados nesta tabela permite constatar algumas tendências entre certas comunidades, uma vez que cada norma de metadados é desenvolvida no contexto das necessidades específicas dos utilizadores que, em princípio, a vão utilizar. Tendo presente a diversidade de comunidades (arquivos, bibliotecas, museus) que se procurou representar através da constituição deste *corpus* técnico, notam-se características próprias de certas normas na descrição dos recursos.

A respeito da ISAD(G), trata-se de uma norma internacional do ICA para aplicação específica na comunidade arquivística, pelo que não surpreende a inclusão de vários elementos que refletem a mundividência da gestão da informação em arquivos. A estrutura de elementos proposta reflete o período (décadas de 1990-2000) em que a norma foi redigida, notando-se um pendor para as dimensões físicas dos documentos analógicos e uma preferência para uma estrutura hierárquica, multinível. Os requisitos de representação das características físicas dos documentos continuarão a ser importantes pelo futuro indeterminado porque, mesmo que amanhã a documentação passasse a ser toda digital, seria ainda necessário descrever os documentos analógicos herdados do passado. Neste sentido, entende-se que o modelo para a autenticidade não se deve centrar apenas na informação digital, prevendo também a descrição de documentos analógicos. Uma leitura semelhante pode ser feita para a DACS, na medida em que manifesta uma grande aproximação à estrutura de elementos da ISAD(G). Um aspeto que distingue a DACS é a inclusão de um elemento para o nome e a localização do repositório que aloja os recursos digitais.

A AGRkMS, enquanto uma norma para a gestão documental de uma administração pública, também revela uma familiaridade com algumas práticas arquivísticas, de que são exemplo a indicação da forma documental, da história custodial e da eliminação de documentos. Além disso, relativamente à ISAD(G), nota-se uma maior proximidade à informação digital, como revelam os elementos de identificação e de mecanismos digitais de verificação da integridade. Porém, destaca-se nesta norma uma saliente ausência de elementos de metadados para o controlo das descrições.

Tabela 59 — Requisitos únicos identificados a partir da análise da literatura técnica.

#	Requisito	ISAD(G)	AGRkMS	EAD	e-EMGDE	DACS	DCMES	VRA	MODS	CDWA	RiC
T1	<b>Identificador</b>	Reference code	Identifier	unitid, agencycode, recordid	Identificador	Reference code	Identifier	-	Identifier	-	Identifier
T2	<b>Título</b>	Title	-	unittitle	Nombre	Title	Title	Title	titleInfo	Title text	Name
T3	<b>Criador</b>	Name of creator	Name	origination	Organo	Name of creator(s)	Creator, Contributor	Agent	-	Creator identity Creator role Creator description Person/corporate body authority	Agent
T4	<b>Data</b>	Dates	Date range	unitdate, unitdatestructured	Fechas	Date	Date	Date	dateissued dateCreated dateCaptured	Creation Date	Date
T5	<b>Extensão</b>	Extent	Extent	physdesc, physdescstructured	Tamaño	Extent	Format	Measurements	-	Measurements Dimensions description	Record resource extent
T6	<b>Suporte</b>	Medium	Medium	-	Soporte	-	Format	-	-	Materials/techniques description	-
T7	<b>Formato</b>	-	-	-	Formato	-	Format	-	-	-	-
T8	<b>Características físicas</b>	Physical characteristics and technical requirements	-	phystech	Características técnicas	-	-	Material Technique	physical Description	Physical description Facture	Physical characteristics Production technique
T9	<b>Audiência-alvo</b>	-	-	-	-	-	-	-	targetAudience	-	-
T10	<b>Nome e localização do repositório</b>	-	-	-	-	Name and location of repository	Coverage	-	name	Place/location authority	-
T11	<b>Forma documental</b>	-	Document form	-	Tipo documental	-	-	WorkType	typeofResource	Object/Work Type	Documentary form type
T12	<b>Orientação/disposição</b>	-	-	-	-	-	-	-	-	Orientation/Arrangement	Structure
T13	<b>Nível de Descrição</b>	Level of description	Category	archdesc, c	Categoria	-	-	-	-	Catalog level	Record set type

#	Requisito	ISAD(G)	AGRkMS	EAD	e-EMGDE	DACS	DCMES	VRA	MODS	CDWA	RiC
T14	Sistema de classificação	System of arrangement	-	-	Clasificación	System of arrangement	-	-	Classification	Classification term	Classification
T15	Âmbito e conteúdo	Scope and content	Description	scopecontent	Descripción	Scope and content	Subject	Subject Description	abstract subject	Subject matter subject terms Context descriptive note	Scope and content
T16	História administrativa/biográfica	Administrative/biographical history	Coverage	bioghist	Âmbito	Administrative/biographical history	-	CulturalContent StylePeriod	genre	Styles/Periods...	History
T17	Origem do recurso	-	-	-	Origen del documento	-	-	-	-	-	-
T18	Estado de elaboração	-	-	-	Estado de elaboración	-	-	-	-	State	State
T19	Localização	existence and location of originals, ... of copies	Location	originalsloc, altformavail	Localización	Existence and location of copies, existence and location of originals	-	Location	location	Current location	Place
T20	Recursos relacionados	Related unit of description	Related entity	relatedmaterial separatedmaterial	Entidad Relacionada	Related archival materials	Relation	Relation StateEdition textref	relatedItem	Related works Related visual documentation Related textual references	-
T21	História custodial	Archival history	Change history	custodhist	Historia del cambio	Custodian history	Provenance	-	dateModified	Ownership/collecting history	-
T22	Avaliação	Appraisal, destruction and scheduling information	Disposal	appraisal	Dictamen	Appraisal, destruction, and scheduling information	-	-	-	-	-
T23	Adições	Accruals	-	accruals	-	Accruals	accrualMethod, accrualPeriodicity, accrualPolicy	-	-	-	Accrual
T24	Fonte de aquisição	-	-	acqinfo	-	Immediate source of Acquisition	-	-	-	-	-

#	Requisito	ISAD(G)	AGRkMS	EAD	e-EMGDE	DACS	DCMES	VRA	MODS	CDWA	RiC
T25	Declarações de integridade	-	Integrity check	-	Verificación de integridad, Firma	-	-	Inscription	-	Inscriptions/marks	Integrity Authenticity note
T26	História de intervenções físicas	-	-	-	-	-	-	-	-	Condition/examination history Conservation/treatment history	-
T27	Criador da descrição	-	-	maintenanceagency	-	Description control	Publisher	-	recordInfo	Cataloging history	-
T28	Data da descrição	Dates of description	-	maintenanceevent, eventdatetime	-	Description control	-	-	recordInfo	Cataloging history	-
T29	Histórico da descrição	-	-	maintenance history, maintenance status	-	-	-	-	-	-	-
T30	Regras ou convenções	Rules or conventions	-	conventiondeclaration	-	Description control	-	-	-	-	-
T31	Fontes para a descrição	-	-	-	-	Description control	Source	Source	recordInfo	-	-
T32	Linguagem da descrição	-	-	-	-	-	-	-	recordInfo	-	-
T33	Notas	Archivist's note	-	processinfo	Notes	-	-	-	-	Cataloging history	-
T34	Contacto	-	Contact (Agent)	-	Contacto	-	-	-	-	-	-

FONTE: elaboração própria.

A natureza da EAD, enquanto norma para a codificação de descrições arquivísticas, é claramente expressa pela estrutura de metadados que adota. Nesta estrutura é patente a criação de metadados usados na gestão da informação em arquivos, como comprovam elementos como a descrição arquivística, a história biográfica, a história custodial e a avaliação. Neste sentido, aproxima-se de outras normas de descrição usadas em arquivos (ISAD[G], AGRkMS). Todavia, em comparação com as restantes normas, a EAD fornece particular realce à transparência da autoria das próprias descrições, documentando os seus autores, datas, e alterações efetuadas às descrições ao longo do tempo. Esta preocupação, que se considera importante, talvez esteja relacionada com o facto de a EAD servir para a criação de instrumentos de pesquisa, tendo por isso a SAA procurado responsabilizar os criadores das representações.

A e-EMGDE é um produto do contexto jurídico e normativo que governa a criação de documentos no âmbito da administração pública espanhola. Observa-se a inclusão dos tradicionais elementos usados na descrição arquivística, como extensão e suporte, classificação, história custodial e avaliação. Além disso, nota-se a distinção entre documentos públicos e privados, exemplificada pela necessidade de esclarecimento sobre a origem dos documentos e de vários outros campos que atestam o contexto legal espanhol. A norma também revela outras características mais próprias da gestão corrente dos documentos, como a indicação do estado de elaboração e das assinaturas aplicadas. No entanto, salienta-se a ausência de metadados para controlo das descrições, o que demonstra que a norma dedica mais atenção às entidades envolvidas na criação dos documentos, sem considerar propriamente as entidades envolvidas na própria criação dos metadados.

A abordagem simplista do DCMES reflete-se nas designações gerais da estrutura de metadados, que não dedica particular atenção à descrição de propriedades tradicionalmente usadas em arquivos, como a forma documental, o sistema de classificação, a história administrativa ou informação sobre destruição de documentos. Antes, caracteriza-se por definir um conjunto essencial de elementos necessários à identificação básica dos recursos, destacando elementos como a história custodial, o assunto, o criador das descrições e as fontes usadas, além dos incontornáveis identificadores dos objetos (título, criador, data, dimensões). Assim, os elementos identificados a partir da análise desta norma parecem constituir um mínimo indispensável, cuja não-inclusão condicionaria a perceção da autenticidade dos recursos. Este conjunto poderá certamente ser complementado com base em requisitos identificados noutras fontes.

Sendo a VRA Core uma norma desenvolvida para a representação de obras de arte, não é de estranhar a ausência de elementos de metadados do contexto arquivístico, tais como história administrativa, avaliação ou sistemas de classificação. Esta norma concentra mais as atenções sobre os elementos que se esperaria encontrar em obras de arte, tais como o contexto cultural, a corrente artística de que resultam, e as condições técnicas dos materiais usados para a produção das obras. Constata-se ainda uma ausência quase absoluta de metadados para o controlo das descrições, sendo apenas prevista a indicação da fonte das informações. Apesar de ter sido desenvolvida para uma

comunidade distinta, entende-se que a indicação da marcas e inscrições aos recursos possa ser um elemento interessante para adotar enquanto um critério de verificação da integridade.

A MODS é outra norma que não se enquadra propriamente na comunidade arquivística, inserindo-se preferencialmente num contexto de representação de recursos bibliográficos, conforme o uso em bibliotecas. Assim, o raciocínio exposto para a ausência de certos elementos de gestão documental na VRA Core é extensível à estrutura de elementos da MODS. Porém, salienta-se que na MODS se observa um importante destaque dado ao controlo das descrições, prevendo elementos para o autor, a linguagem, as datas, as fontes e notas sobre os metadados. Além disso, ao passo que na VRA Core não é indicado qualquer sistema de classificação, na MODS já se constata o uso deste campo, o que não é de admirar dada a sua significação na gestão de coleções bibliográficas. Um elemento aparentemente único usado na MODS é o da audiência alvo, que pode ser difícil de justificar. Conforme tem sido discutido ao longo desta investigação, num contexto de partilha das representações dos recursos na *web*, os contextos em que as representações vão ser consultadas — que condicionam o seu significado — é imprevisível, não devendo os criadores das descrições tentar prever qual a audiência-alvo, uma vez que não se deve fechar a porta a qualquer utilizador. Também se pode entender este elemento apenas como indicativo, e não como uma manifestação exclusiva. Ainda assim, talvez seja preferível não o incluir.

A CDWA elege como foco as obras de arte e outros objetos artísticos como arquitetura. Esta abordagem é clara numa elevada preocupação com a documentação das características físicas das obras, desde técnicas, marcas, instrumentos e métodos de construção à história de conservação e à história da examinação sob mecanismos como luz ultravioleta. Por contraste, considerando a aparente ausência de elementos de metadados específicos da gestão digital, como identificadores únicos, características de *hardware* e *software* ou controlo de versões, não parece dedicar muita atenção à representação de recursos digitais. Ainda assim, uma interessante lição observada prende-se com uma preocupação recorrente na norma de explicitar as fontes nas quais se baseiam as informações incluídas. Outro requisito interessante é a inclusão de um elemento para a descrição da história da conservação das obras de arte, que pode ser generalizado para uma história das alterações dos custodiantes à dimensão física dos documentos.

A RiC, embora em desenvolvimento por uma associação internacional de arquivistas, pretende afirmar-se como uma norma internacional que consolida a representação da informação em comunidades próximas, mas que nem sempre dialogam diretamente, tais como arquivística, a gestão de documentos e o setor do património cultural. Como tal, os metadados apresentados não refletem em pormenor nenhuma destas áreas, procurando antes introduzir um modelo concetual modular e flexível que permita aos gestores da informação selecionar os elementos que considerem mais apropriados. É uma norma pensada e desenvolvida para o contexto digital, de que é exemplo a modularidade dos atributos e a diversidade de entidades que podem servir como categorias primárias de descrição. Porém, salienta-se a ausência de elementos de metadados explícitos para o controlo da descrição, o que seria útil para qualquer comunidade.



Apesar destas divergências, é possível também observar uma extrema coincidência nalguns dos requisitos, conforme demonstrado pela Tabela 60, que inclui os requisitos encontrados em metade ou na maioria ( $\Rightarrow 50\%$ ) das normas, ou seja, com 5 ou mais ocorrências. De entre os elementos presentes na estrutura principal de todas as normas analisadas, destaca-se a indicação da data de criação dos recursos (T4), uma descrição do seu conteúdo (T15). Além disso, com uma presença em 9 das 10 normas analisadas (90%), destaca-se a indicação do título do recurso (T2), do seu criador (T3), das suas dimensões físicas (T5), da história administrativa (T16), da sua localização atual (T19) e os recursos relacionados (T20). O uso de um identificador para a descrição dos recursos (T1) é também extremamente frequente, ocorrendo em 8 normas (88%), a mesma ocorrência da história custodial (T21). A elevada frequência destes requisitos parece sugerir que são elementos incontornáveis na descrição dos recursos, independentemente das comunidades.

Um pouco menos frequente, mas ainda assim presente na maioria dos casos (70%, ou 7 em 10), destaca-se também a descrição das características físicas dos recursos (T8). Nesta tabela indicam-se ainda os requisitos observados em seis das normas (60%), nomeadamente a forma documental ou tipo de recurso (T11), o nível de descrição (T13) e o sistema de classificação adotado (T14).

Ocorrendo em metade das normas, ou seja, em 5 de 10 (50%), salienta-se ainda o suporte físico dos recursos (T6), informação sobre a avaliação dos recursos, incluindo a destruição (T22), adições ao agregado documental (T23), a indicação do autor das descrições (T27) e a data das descrições (T28). Embora relativamente menos abundantes, também se entende que estes elementos constituem informações prioritárias na descrição dos recursos digitais.

Tabela 60 — Requisitos técnicos mais frequentes.

<b>Requisito</b>	<b>Ocorrências</b>	<b>Percentagem</b>
T4	10	100%
T15	10	100%
T2	9	90%
T3	9	90%
T5	9	90%
T16	9	90%
T19	9	90%
T20	9	90%
T1	8	80%
T21	8	80%
T8	7	70%
T11	6	60%

T13	6	60%
T14	6	60%
T6	5	50%
T22	5	50%
T23	5	50%
T27	5	50%
T28	5	50%

---

FONTE: elaboração própria.

Porém, o facto de os restantes requisitos ocorrerem com menor frequência não significa que merecem menos consideração. Existe um possível argumento de que, quanto mais raros, mais originais são certos campos de descrição, merecendo ainda uma atenção redobrada. Como tal, todos os requisitos são considerados na construção do modelo para a autenticidade. Quando muito, a leitura da frequência de certos metadados pode ajudar na decisão de incluir ou não certo requisito, uma vez que, se diversas normas de distintas comunidades incluem os mesmos campos de descrição é porque, à partida, esses elementos são importantes em qualquer contexto.

À semelhança da abordagem efetuada na discussão dos requisitos da literatura científica, também agora se estruturam os requisitos identificados na literatura técnica de acordo quatro categorias. Neste caso, identifica-se que os metadados podem referir-se: 1) ao recurso; 2) ao agregado; 3) ao criador; ou 4) ao controlo das representações. A distribuição dos metadados por cada categoria é apresentada na Figura 16.

Nesta figura estão incluídos e organizados os 34 requisitos derivados da análise das normas. Conforme mencionado na discussão dos requisitos científicos, uma discussão do significado exato de cada elemento será reservada para a proposta final do modelo. Estas designações são relativamente autoexplicativas, por isso suficientes para um entendimento básico das informações a incluir nestes campos de descrição. Todavia, entende-se que devem ser feitas algumas observações quanto às escolhas de incorporar certos elementos sob o ‘recurso’, e não sob o ‘agregado’.

O facto de a categoria ‘agregado’ apenas incluir um elemento não significa que outras informações não sejam úteis ao nível do agregado. Por exemplo, a audiência alvo (T9), o sistema de classificação (T14), a história administrativa (T16), a história custodial (T21), a avaliação (T22), a fonte de aquisição (T24), e até mesmo o âmbito e conteúdo (T15), poderiam facilmente ser metadados incluídos ao nível do agregado. Na verdade, acredita-se que estas informações referentes ao agregado podem e devem ser dadas em qualquer descrição arquivística. Contudo, também se entende que, além de presentes ao nível do agregado, estes metadados devem estender-se à própria descrição mais individualizada dos recursos.

<b>Recurso</b>	<b>Agregado</b>
<ul style="list-style-type: none"> <li>• Identificador..... T1</li> <li>• Título..... T2</li> <li>• Data..... T4</li> <li>• Extensão.....T5</li> <li>• Suporte..... T6</li> <li>• Formato.....T7</li> <li>• Características físicas.....T8</li> <li>• Audiência-alvo.....T9</li> <li>• Nome e localização do repositório... T10</li> <li>• Forma documental..... T11</li> <li>• Orientação/disposição..... T12</li> <li>• Nível de descrição.....T13</li> <li>• Sistema de classificação.....T14</li> <li>• Âmbito e conteúdo..... T15</li> <li>• História administrativa.....T16</li> <li>• Origem do recurso.....T17</li> <li>• Estado de elaboração.....T18</li> <li>• Localização.....T19</li> <li>• Recursos relacionados..... T20</li> <li>• História custodial..... T21</li> <li>• Avaliação..... T22</li> <li>• Fonte de aquisição.....T24</li> <li>• Declarações de integridade.....T25</li> <li>• História de intervenções físicas.....T26</li> </ul>	<ul style="list-style-type: none"> <li>• Adições..... T23</li> </ul> <p><b>Criador</b></p> <ul style="list-style-type: none"> <li>• Nome..... T3</li> <li>• História biográfica.....T16</li> </ul> <p><b>Controlo das representações</b></p> <ul style="list-style-type: none"> <li>• Criador da descrição..... T27</li> <li>• Histórico da descrição..... T29</li> <li>• Regras ou convenções..... T30</li> <li>• Fontes para a descrição..... T31</li> <li>• Linguagem da descrição..... T32</li> <li>• Notas..... T33</li> <li>• Contacto..... T34</li> </ul>

Figura 16 — Elementos de metadados identificados na literatura técnica, com a indicação do número do requisito.

FONTE: elaboração própria.

Por este motivo, opta-se por os incluir sob o recurso, como defesa de uma descrição ao nível do item, sem prejuízo para que essas informações sejam também prestadas ao nível da coleção. Na prática, as descrições podem realizar-se num primeiro momento ao nível do agregado, a partir do qual os recursos, enquanto subcategorias desse agregado, herdam as informações dos níveis superiores. Esta implementação é característica das estruturas hierárquicas (como o XML), encontrando-se em abundância nas descrições em arquivos. O argumento central é o de que estas informações devem estar incluídas ao nível do recurso, independentemente da forma de codificação dos metadados, e independentemente da sua repetição noutros níveis da descrição.

### 5.3 Proposta do modelo

No final desta secção é feita a proposta do modelo para a autenticidade, com base nos dados recolhidos ao longo de toda a investigação. Mais concretamente, para o modelo consideram-se como fontes principais os requisitos identificados previamente na literatura científica (Figura 15) e na literatura técnica (Figura 16).

Uma comparação dos requisitos encontrados revela algumas convergências entre ambas as abordagens. Por exemplo, entende-se que em cada categoria de dados se encontram argumentos

para organizar os requisitos segundo aqueles que dizem respeito aos recursos, aos criadores, ou ao controlo das representações. Ao nível dos criadores, tanto os dados da literatura técnica como da científica apontam para a indicação do seu nome e da sua história biográfica. A respeito dos metadados para a gestão das coleções, por vezes também denominados de meta-metadados (Joudrey & Taylor, 2018, p. 196; Zeng & Qin, 2016, p. 20), no sentido em que se trata de metadados que descrevem metadados, também se observa uma elevada coincidência, com ambas as fontes remetendo para a descrição do criador da descrição, da data, de um histórico de alterações e de notas do criador.

É ao nível dos metadados para a descrição dos recursos que se começa a observar uma maior divergência. A literatura científica tende a ser mais sucinta, geralmente repetindo um conjunto de elementos indispensáveis, que se traduz num menor número de metadados do que aqueles observados na literatura técnica. Como resultado, relativamente aos metadados sobre os recursos, a literatura técnica inclui praticamente todos os metadados veiculados pelos dados teóricos, e mais alguns. Dito de outra forma, é mais completa. Em contrapartida, a literatura científica dedica mais alguma atenção à formulação de considerações no processo de representação, dedicando assim mais esforços na teorização dos objetivos da descrição e de algumas orientações de implementação. Algumas normas de metadados também veiculam ideias por vezes semelhantes quando explicitam — geralmente na introdução ou no prefácio — algumas regras de descrição.

Em termos de proposta de modelo, o passo mais fácil seria simplesmente combinar os metadados encontrados em ambas as abordagens, fornecendo assim uma lista extensiva. Apesar de tentador, entende-se que estes dados devem ser interpretados com um espírito crítico, questionando o propósito da sua inclusão. Além disso, em qualquer modelo de metadados existe um *trade-off* implícito entre completude e dificuldade de implementação, isto é, quantos mais elementos forem incluídos mais bem representados ficam os recursos, e mais oneroso se torna para as organizações realizar as descrições. Por este motivo, a tentação de incluir todos os metadados possível deve ser contrabalançada com a justificação da sua inclusão. No caso da descrição arquivística, mais nem sempre é melhor. O próprio surgimento de normas simplistas como o DCMES ou da própria abordagem MPLP deve servir como recordação da necessidade que certas comunidades sentem em simplificar procedimentos.

Aproveita-se também para esclarecer as necessidades teóricas que o modelo pretende satisfazer. Conforme discutido na revisão da literatura, a autenticidade é uma etapa necessária para inferir a confiabilidade. Por outras palavras, demonstra-se a autenticidade como fundação para estabelecer a confiança. Mais especificamente, além da autenticidade, devem também serem demonstradas a fidedignidade (*reliability*) e a precisão (*accuracy*) das descrições (InterPARES 2, 2008a, p. 335). Guiada por uma abordagem diplomática, a investigação do projeto InterPARES 2 concluiu que a fidedignidade consiste na examinação da completude e do procedimento de criação dos documentos, ao passo que a precisão diz respeito ao grau segundo o qual os documentos são precisos, corretos, verdadeiros e pertinentes. No presente modelo não se consideram explicitamente a fidedignidade e a confiabilidade de modo a não dispersar a investigação. Também poderia ser

argumentado que estes aspetos se focam mais no processo de criação documental, e não tanto na descrição arquivística. Das conclusões obtidas pelo projeto InterPARES pretende-se salientar como a autenticidade implica duas componentes: a demonstração da identidade e a verificação da integridade.

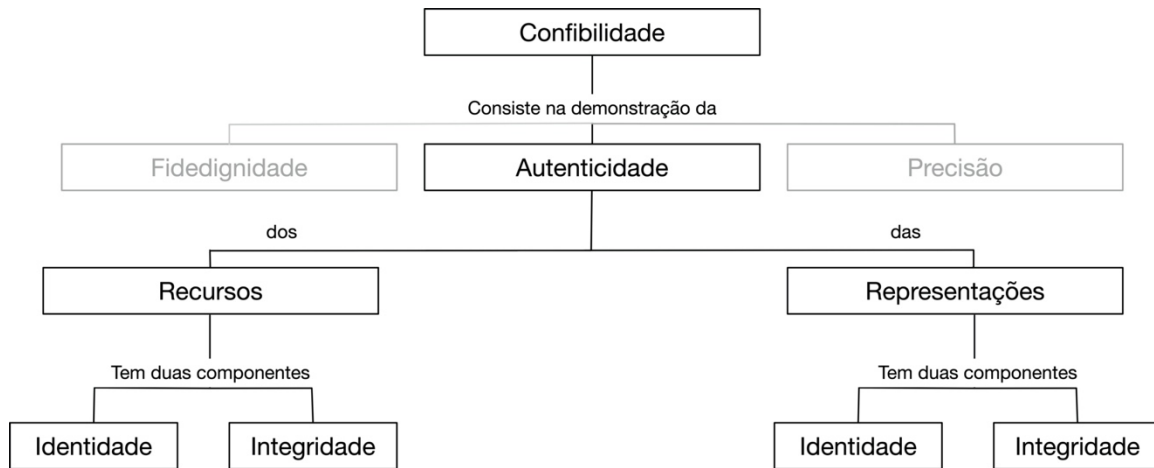


Figura 17 — Esquema conceitual da confiabilidade.

FONTE: elaboração própria.

Tendo por base estas premissas, este raciocínio pode ser expandido além dos documentos de arquivo, o objeto de análise do projeto InterPARES. A presunção da autenticidade pode ela própria colocar-se ao nível das descrições arquivísticas que, conforme visto na revisão de literatura, devem ser questionadas enquanto práticas subjetivas. Assim, da mesma forma que se entende a autenticidade dos recursos como consistindo na identidade e na integridade destes, também se pode sugerir que a autenticidade das descrições arquivísticas consiste na demonstração da sua identidade e da sua integridade. Estas ideias são esquematizadas pela Figura 17, que pretende ilustrar os critérios que guiam a discussão dos elementos de metadados a incluir na proposta do modelo para a autenticidade. Assim, quando se refere autenticidade, subentende-se simultaneamente a autenticidade dos recursos e das próprias descrições arquivísticas, numa dupla vertente. Um raciocínio semelhante poderia ser realizado para os conceitos de fidedignidade e de precisão, alargando-os também ao próprio âmbito das descrições arquivísticas. A questão da fidedignidade das descrições, embora não sendo o foco da investigação, tem sido abordada de certa forma aquando das discussões da completude da estrutura de metadados. Porém, conforme assinalado acima, não constitui o foco da investigação.

Posto isto, os metadados que se entendem que devem figurar num modelo para a autenticidade são apresentados na Figura 18. Na Tabela 61 é indicada com maior detalhe uma definição de cada elemento, esclarecendo que tipo de informações se esperam. Salienta-se que estes metadados constituem apenas uma recomendação, devendo as organizações decidir complementá-los ou usar um número mais reduzido. Além disso, também se admite que a estrutura de

apresentação dos metadados não seja necessariamente a estrutura prevista pelo modelo. Desde que a informação correta seja veiculada, a ordem dos metadados não é inquestionável.

## **1. Recurso**

- 1.1 Identificador
- 1.2 Título
- 1.3 Data
  - 1.3.1 Data de criação
  - 1.3.2 Data da última modificação
  - 1.3.3 Versão
  - 1.3.4 Estado de elaboração
- 1.4 Criador
- 1.5 Características físicas
  - 1.5.1 Dimensões e suporte
  - 1.5.2 Formato
  - 1.5.3 Software e hardware
- 1.6 Nível de descrição
- 1.7 Assunto
- 1.8 Língua
- 1.9 Localização
  - 1.9.1 Nome do repositório
  - 1.9.2 Nome dos custodiantes
  - 1.9.3 Morada
- 1.10 Forma de representação
- 1.11 História custodial
- 1.12 História da classificação
- 1.13 Avaliação
  - 1.13.1 Eliminação
  - 1.13.2 Planos de retenção
- 1.14 Recursos relacionados
- 1.15 Declarações de integridade
- 1.16 Histórico de intervenções físicas

## **2. Criador**

- 2.1 Identificador
- 2.2 Nome
- 2.3 História biográfica

## **3. Agregado**

- 3.1 Adições

## **4. Meta-metadados**

- 4.1 Criador dos metadados
- 4.2 Data dos metadados
  - 4.2.1 Data de criação dos metadados
  - 4.2.2 Data de última modificação aos metadados
- 4.3 Histórico de alterações aos metadados
- 4.4 Regras/convenções
- 4.5 Fontes
- 4.6 Língua dos metadados
- 4.7 Notas
- 4.8 Contacto

Figura 18 — Proposta de modelo de metadados para a autenticidade das descrições arquivísticas.

FONTE: elaboração própria.

Acredita-se que este modelo aborda de forma completa a necessidade de demonstrar a identidade e a integridade dos recursos e das representações. Relativamente aos recursos, a indicação de um identificador, do título, das datas, do nome do criador e das características físicas são fundamentais para delimitar os recursos enquanto objetos físicos a gerir. O nível da descrição, o assunto e a forma de representação contribuem para descrever a estrutura da agregação, as atividades que se refletem nos recursos, e a natureza dos recursos nessa agregação.

Além disso, a descrição da localização dos recursos e da história custodial é essencial para esclarecer sobre a transferência de custódia ao longo do tempo. As declarações de integridade também ajudam a esclarecer a história de transmissão dos documentos, desta vez com um enfoque nas manipulações físicas que estes sofreram.

Uma descrição das intervenções dos arquivistas na gestão dos recursos é possibilitada pelos elementos da história da classificação, da avaliação e das intervenções físicas. Ao nível da proveniência, a língua pode também ser uma informação importante. Os metadados sobre o criador

contribuem para reforçar o conhecimento da proveniência, indicando um identificador, o seu nome e a sua história biográfica.

Ainda ao nível dos recursos, a inclusão de informação sobre recursos relacionados é importante para estabelecer relações com recursos de outros agregados, contextualizando assim a própria interpretação e o significado dos recursos. Em relação ao agregado, a descrição de adições pode também ajudar a esclarecer sobre intervenções dos arquivistas/gestores. Muitas outras informações podem ser prestadas ao nível do agregado, tais como características físicas, níveis de descrição, assunto, línguas, localização, formas de representação, história custodial, história da classificação e recursos relacionados. Embora estes metadados surjam no modelo como subjacentes aos recursos, também se admite que estas informações possam ser descritas ao nível do agregado e posteriormente herdadas de forma automática pelos recursos.

Porém, o facto de terem sido incluídas sob os recursos pretende realçar que este tipo de informações deve ser dado explicitamente para cada recurso, de forma a promover uma maior granularidade das descrições. Assim, enquanto um agregado pode conter recursos que digam respeito a cinco assuntos diferentes, um recurso em princípio dirá respeito a um assunto. Desta forma, estes metadados, quando referentes ao agregado, podem ser entendidos mais como uma soma dos valores dos metadados dos recursos que constituem esse agregado.

Por fim, entende-se que todos os metadados propostos para controlo das descrições — ou meta-metadados — importantes para atestar a autenticidade das representações, esclarecendo sobre a proveniência destas e promovendo a transparência de procedimentos e a responsabilização dos criadores das descrições.

Tabela 61 — Definições dos metadados do modelo.

Referente	Elemento	Definição
1. Recurso	1.1 Identificador	Um identificador único do recurso, como um URI.
	1.2 Título	O nome do recurso.
	1.3 Data	Informação cronológica útil para a contextualização.
	▪ 1.3.1 Data de criação	A data de criação dos recursos.
	▪ 1.3.2 Data da última modificação	Data da última modificação aos recursos.
	▪ 1.3.3 Versão	Indicação da versão do documento.
	▪ 1.3.4 Estado de elaboração	Informação sobre o estado de transmissão (cópia, original, cópia autenticada).
	1.4 Criador	Nome do responsável pela criação dos recursos.
	1.5 Características físicas	Informação física útil para o manuseamento.
	▪ 1.5.1 Dimensões e suporte	Informação sobre a extensão e materiais dos documentos.
▪ 1.5.2 Formato	Indicação da extensão dos ficheiros.	
▪ 1.5.3 Software e hardware		

Referente	Elemento	Definição
		Informação sobre as características técnicas da execução, que auxiliem a interpretação.
	1.6 Nível de Descrição	Identificação do nível de classificação.
	1.7 Assunto	Uma descrição do conteúdo.
	1.8 Língua	Indicação da(s) língua(s) utilizadas.
	1.9 Localização	Informação geográfica que caracteriza a custódia.
	<ul style="list-style-type: none"> <li>▪ 1.9.1 Nome do repositório</li> <li>▪ 1.9.2 Nome dos custodiantes</li> <li>▪ 1.9.3 Morada</li> </ul>	Designação oficial do repositório. Nome formal do indivíduo ou organização que detém a custódia. Localização geográfica ou digital (URI) do repositório.
	1.10 Forma de representação	Informação sobre a estrutura e a ordem internas, o sistema de classificação ou sobre a forma de visualização da informação.
	1.11 História custodial	Informação sobre transferências de custódia.
	1.12 História da classificação	Esclarecimento das diferentes classificações usadas ao longo do tempo.
	1.13 Avaliação	Informação sobre intervenções dos arquivistas.
	<ul style="list-style-type: none"> <li>▪ 1.13.1 Eliminação</li> <li>▪ 1.13.2 Planos de retenção</li> </ul>	Indicação de documentos eliminados, e quando. Informações sobre eliminações previstas ou de transferências de custódia.
	1.14 Recursos relacionados	Identificação de recursos relacionados.
	1.15 Declarações de integridade	Mecanismos de verificação da integridade, tais como <i>checksums</i> , assinaturas, selos, carimbos ou marcas.
	1.16 História de intervenções físicas	Indicações de modificações físicas aos documentos pelos custodiantes, tais como atos de conservação e restauro ou ações de preservação digital, como migrações.
2. Criador	2.1 Identificador	Um identificador único do criador dos recursos.
	2.2 Nome	O nome oficial do criador dos recursos.
	2.3 História biográfica	Detalhes biográficos do criador que contextualizam os recursos.
3. Agregado	3.1 Adições	Informação sobre adições antecipadas.
4. Meta-metadados	4.1 Criador dos metadados	Nome do criador dos metadados.
	4.2 Data dos metadados	Informação cronológica útil para a contextualização das representações.
	<ul style="list-style-type: none"> <li>▪ 4.2.1 Data de criação dos metadados</li> <li>▪ 4.2.2 Data de última modificação aos metadados</li> </ul>	Indicação da data de criação da descrição. Indicação da data de modificação à estrutura da descrição.



Referente	Elemento	Definição
	4.3 Histórico de alterações aos metadados	Documentação de alterações à estrutura das descrições.
	4.4 Regras/convenções	Explicitação de princípios nos quais a descrição se baseia.
	4.5 Fontes	Fontes de informação usadas na descrição.
	4.6 Língua dos metadados	Língua da descrição.
	4.7 Notas	Qualquer informação adicional pelos criadores da descrição.
	4.8 Contacto	Forma de contacto dos responsáveis pela descrição.

FONTE: elaboração própria.

Merecem ainda uma palavra os metadados que se optou por não incluir no modelo. De entre os requisitos identificados na literatura científica, foram excluídos os requisitos que apontavam para a indicação dos direitos de uso dos recursos e do CV dos criadores das descrições. Considera-se que as licenças de uso não estão tão relacionadas com a demonstração da identidade ou da integridade, mas antes à transmissão e ao uso dos recursos. Por sua vez, a inclusão do currículo do criador pode ser enganadora. Por um lado, admite-se a participação de múltiplos intervenientes na criação dos metadados, o que significa muitos currículos que, por sua vez, representaria porventura demasiada informação para um utilizador realisticamente consultar. Além disso, um currículo, enquanto documentação da educação formal, não informa necessariamente sobre a subjetividade dos indivíduos. Acredita-se que os objetivos que guiaram a literatura a sugerir a inclusão dos CVs se encontram cumpridos no elemento sobre as regras/convenções das representações.

A respeito da literatura técnica, também se deliberou não incluir os requisitos referentes à audiência-alvo, à forma documental, à orientação/disposição de visualização e à origem dos recursos. A indicação da audiência parece ser uma previsão subjetiva e desnecessária, não obstante a consciência que as organizações devem ter durante a descrição das necessidades dos utilizadores de uma comunidade primária. Admite-se que a forma documental poderia ser útil para a descrição de documentos analógicos, mas pode ser difícil de justificar para os objetos digitais. Informações sobre a orientação dos recursos parecem ser mais próprias das obras de arte, sendo a sua necessidade questionável para a descrição arquivística. Por fim, a origem do recurso advém de uma norma espanhola para a gestão documental na administração pública, que sentiu por isso a necessidade de distinguir entre documentos de origem pública e privada. De um modo geral, entende-se que esta distinção, apesar de porventura útil nalguns casos, não constitui um requisito fundamental para a demonstração da autenticidade dos recursos.

Para concluir, os metadados incluídos neste modelo pretendem transmitir que, quando presentes, estas informações tornam a descrição dos recursos e das próprias descrições completas, ajudando a estabelecer a identidade e a fornecer informações sobre a manutenção da integridade dos recursos, das coleções e das descrições ao longo do tempo. Também reconhecem a subjetividade dos arquivistas (ou de qualquer outro profissional da informação) enquanto atores

participantes no processo de construção das representações, e estabelecem as condições para que os recursos possam ser descobertos de forma individual pelos utilizadores, em contextos imprevisíveis formados a partir de pesquisas que capitalizem nas relações semânticas entre os recursos. Neste sentido, contribuem para a transparência do processo de descrição e para a responsabilização dos seus autores. Este modelo não fornece guias para a codificação. Porém, algumas orientações de representação foram sugeridas na discussão dos requisitos da literatura científica. De resto, caberá às organizações a decisão de quais os metadados a incluir. Este modelo é, afinal, uma proposta.

## Conclusão:

---

Considera-se que a investigação documental foi um método apropriado como resposta a alguns dos objetivos propostos, na medida em que permitiu explorar criticamente um *corpus* documental rico e diverso o suficiente para servir de base a uma revisão da literatura que se acredita ser transversal, atualizada e esclarecedora, dentro do tema escolhido.

Neste sentido, recapitulando os principais pontos mencionados durante a revisão da literatura, da importância dos metadados retira-se que ter metadados em abundância suficientes e com contexto equivale a ter conteúdo, sendo estes uma representação simplificada, parcial de uma realidade mais complexa. Quanto à sua definição, chegou-se a um entendimento de metadados enquanto dados que dotam os recursos de estrutura, contexto e significado.

Das secções sobre as etapas do uso de metadados ao longo dos tempos, salienta-se que os metadados não são apenas digitais, existindo há milénios desde que há registos escritos. Porém, o digital vem potenciar formas de combinação e seleção de metadados que dotam as pesquisas, a partilha e o reuso de um poder sem paralelo na história anterior.

Relativamente aos tipos e às funções, destaca-se que os metadados podem ser categorizados de acordo com vários tipos, e que esses tipos não correspondem a uma hierarquia estabelecida, dependendo antes da ênfase preferencial de cada comunidade, que por sua vez se reflete numa diversidade de posições defendidas pelos autores. Os tipos podem estar associados a funções, mas essas funções não são exclusivas, permeando vários tipos de metadados.

Sobre as formas de expressão, nota-se que o RDF parece ser o modelo de dados com maior potencial para a expressão, a partilha e o reuso de metadados no contexto *web*, no sentido em que apresenta uma neutralidade de esquema por virtude da extrema simplificação das asserções de metadados, o que potencia a reutilização e a construção de pesquisas mais ricas e adaptadas às necessidades dos utilizadores.

Da discussão da terminologia de normas de metadados aponta-se o registo de metadados (*metadata record*) como um conjunto de metadados que descreve um recurso, composto por sintaxe, estrutura e semântica, focalizando a análise deste trabalho em normas de metadados de estrutura e nos elementos de metadados (*metadata elements*) que as constituem. Esclarece-se que se entende uma norma de metadados como um documento formal que define critérios, métodos, procedimentos e práticas uniformes para a escolha, expressão e/ou partilha de metadados, sem um carácter vinculativo, enquanto um termo genérico que se ramifica noutros como modelo ou esquema. Apresentam-se também os argumentos segundo os quais se justifica a proposta de um modelo, nomeadamente porque a proposta se foca na autenticidade como propriedade relevante do objeto que se pretende representar. Discute-se ainda brevemente a interoperabilidade, contextualizando o modelo proposto como forma de potenciar a interoperabilidade semântica entre as organizações que eventualmente adotem o modelo proposto, ou partes deste. Uma das vantagens da adoção por uma comunidade deste conjunto de metadados prende-se com a melhoria da interoperabilidade ao *schema level*, de forma a facilitar a pesquisa integrada, sem esforços de descrição adicionais, em

vários repositórios em simultâneo, ao mesmo tempo que reforçaria a confiabilidade dos utilizadores nos metadados disponibilizados pelas instituições.

No segundo capítulo reflete-se sobre o conceito de representação, salientando que qualquer representação é uma simplificação da realidade e que, como tal, a sua construção é necessariamente humana, subjetiva e parcial. Esta carga subjetiva é inevitável uma vez que, se é uma simplificação, há elementos da realidade que não são incorporados na representação, o que implica uma seleção subjetiva, influenciável pelo contexto temporal, geográfico e social do autor da representação. Esclarece-se também como se entende que a representação se desdobra noutros dois conceitos: num sentido alargado, relaciona-se com a indexação e, num sentido restrito, identifica-se com a descrição arquivística.

A partir daqui, reflete-se sobre a representação da informação em arquivos, procurando discernir o papel da tecnologia e das correntes sociais no processo de representação. Depois de estabelecer o pressuposto de que a representação — e os seus produtos, como os instrumentos de pesquisa — é um meio de tornar presente algo ausente, defende-se que a informação é representada de modo a facilitar a sua recuperação, e que nessa representação deve estar explicitado o(s) contexto(s) arquivístico(s), indispensável(is) para a compreensão do sentido e do propósito da informação recuperada. Recordam-se os princípios que governaram a descrição arquivística nos últimos dois séculos — o princípio da proveniência e da ordem original —, procedendo a uma análise crítica segundo a qual se identifica que estes princípios têm dificuldade em manter-se relevantes no contexto digital, dada a plasticidade e a fluidez da produção e da comunicação da informação, que questiona o conceito de fundo e de ordem original. Nota-se que a descrição é uma atividade distinta dos instrumentos de pesquisa, que são os produtos dessa atividade e que, no contexto digital, o controlo intelectual pode e deve ser distinto do controlo físico, de modo a poder criar pesquisas personalizadas às necessidades dos utilizadores, promovendo assim a descoberta e o uso dos recursos.

Acerca das correntes de representação arquivística, começa-se por enquadrar a teoria arquivística contemporânea na corrente intelectual do pós-modernismo, caracterizado por uma descrença na imparcialidade e em verdades absolutas em favor da ideia de que qualquer ilação é um ato interpretativo e subjetivo. Para os arquivos isto significa, por um lado, que os documentos de arquivo não representam factos nem verdades encerradas em si, mas antes constituem uma forma de evidência, socialmente condicionada e contextualizada, um testemunho subjetivo de determinada atividade da qual surge como produto. Além disso, o pós-modernismo traz consigo também uma mudança de paradigma na gestão da informação, passando de um enfoque do produto — o documento de arquivo e respetivos instrumentos de pesquisa — para o processo — a descrição arquivística como uma atividade. Esta transição implica um repensar da abordagem de representação dos arquivos, estando progressivamente desacreditada uma posição autoritária, estática e unilateral da descrição dos fundos documentais, do geral para o particular, em favor da descrição ao nível do item de modo a permitir aos utilizadores construir as suas próprias pesquisas e contextos de interpretação dos recursos.

Relativamente à autenticidade, demonstra-se como esta constitui um dos fundamentos do código de ética dos arquivistas, carecendo da verificação da identidade e da demonstração da integridade. Outra característica-chave dos documentos de arquivo assenta na demonstração da sua fidedignidade, enquanto uma representação fiel das transações que documenta. Em conjunto, a garantia da autenticidade e a confiabilidade tem a finalidade de determinar se certo documento de arquivo é confiável. Vê-se ainda como estes procedimentos não encerram em si uma apreciação definitiva, sendo esta variável e subjetiva, dependendo das necessidades da comunidade que exerce o juízo de valor. Apresenta-se ainda a ideia do vínculo informacional como expansão do vínculo arquivístico.

Esta revisão e sistemática crítica da literatura permite dar resposta às questões de investigação lançadas previamente. Perguntou-se quais as vantagens e desvantagens da representação em grafo quanto à verificação da autenticidade dos documentos de arquivo. A resposta é esta: tal forma de representação permite enriquecer os contextos nos quais os recursos são interpretados, potenciando a descoberta e suavizando a imposição de uma narrativa pela instituição. Porém, uma desvantagem encontrada prende-se com a impreparação do mercado, que pode dificultar as implementações. Questionou-se também qual o paradigma de descrição arquivística subjacente a estas formas de representação e acesso à informação, chegando-se à conclusão de que parece ser o paradigma pós-moderno, caracterizado pela consciência da subjetividade e por uma recusa das narrativas impostas.

Outra pergunta indagava se seria possível construir um modelo teórico que identificasse os requisitos para avaliar a autenticidade e determinar a confiabilidade dos documentos de arquivo que possa ser seguido e aplicado no contexto da *Web Semântica*, e quais seriam os elementos de metadados que esse modelo deva incluir. Através do trabalho realizado, em particular no quinto e último capítulo, demonstrou-se que tal modelo não só é possível, como necessário de forma a reforçar a asserção da autenticidade e da confiabilidade dos documentos digitais. Os elementos desse modelo são expostos na Figura 18, uma vez que seria demasiado extenso enumerá-los aqui. Por fim, questionou-se ainda sobre os potenciais benefícios deste modelo de metadados, em termos de descoberta da informação e completude das descrições, quando comparado com os atualmente predominantes instrumentos de descrição arquivística. Uma pergunta deste carácter requer uma resposta mais aberta, uma vez que tais benefícios dependem das instituições adotantes. Porém, acredita-se que com a apresentação deste modelo se contribui para a valorização da autenticidade enquanto um elemento crucial a ter em consideração no processo de descrição em arquivos. Também se promove uma interpretação mais clara de como manifestar a autenticidade, com a particular vantagem de existir a garantia de que o modelo proposto tem em consideração as recomendações presentes na literatura técnica e científica dos últimos 10-12 anos. Assim, este modelo figura como um contributo completo, crítico e atualizado sobre a questão da autenticidade nas descrições arquivísticas.

Contudo, nem estas questões nem as soluções apresentadas são definitivas. Dado o carácter socialmente construído da informação e das representações, espera-se que estes conceitos, estas

correntes e estas práticas sofram alterações nos próximos anos. É já possível entrever algumas tecnologias que podem assumir uma posição preponderante no dia de amanhã, nomeadamente a inteligência artificial, o *blockchain* e as comunicações quânticas. Algumas destas estão já em curso, como *machine-learning* e *smart-contracts*, sendo de esperar que na próxima década se tornem *mainstream*. Seria fascinante desenvolver trabalhos que explorassem a forma como estas tecnologias e os hábitos sociais e culturais que elas acarretam influenciam a produção, o armazenamento e a comunicação da informação, que por sua vez se manifestam no tipo informação com que os arquivos têm de lidar.

Sobre os demais contributos desta investigação, entende-se que não devem ser os autores a pronunciar-se sobre eles. Tentar enumerar contributos seria impor uma narrativa de leitura deste documento, tal como os arquivistas tentaram impor as suas narrativas através da representação e descrição dos recursos. Os contributos terão de ser depreendidos pelos leitores, consoante a natureza do benefício que retirarem desta investigação, seja qual for, mesmo que o único contributo possa ser o de reconhecerem que, no seu caso, neste trabalho não há contributos.

## Bibliografia

---

- Abbott, T. (1994). *Internet World's on Internet 94: An International Guide to Electronic Journals, Newsletters, Texts, Discussion Lists, and Other Resources on the Internet*. London: Mecklermedia.
- Academia das Ciências de Lisboa. (2001). *Dicionário da língua portuguesa contemporânea da Academia das Ciências de Lisboa*. Lisboa: Verbo.
- Achinstein, P. (1964). Models, Analogies, and Theories. *Philosophy of Science*, 31(4), 328–350.
- Achinstein, P. (1965). Theoretical Models. *The British Journal for the Philosophy of Science*, 16(62), 102–120.
- Acker, A. (2017). When is a record?: a research framework for locating electronic records in infrastructure. In A. Gilliland, S. McKemmish, & A. Lau (Eds.), *Research in the Archival Multiverse* (pp. 286–323). Clayton: Monash University Publishing.
- Ackoff, R. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, 16(1), 3–9. <https://doi.org/citeulike-article-id:6930744>
- Ahmed, J. (2010). Documentary Research Method: New Dimensions. *Indus Journal of Management & Social Sciences*, 4(1), 1–14. Disponível em: <http://ideas.repec.org/s/iij/journl.html>
- Albarello, L., Digneffe, F., Hiernaux, J.-P., Maroy, C., Ruquoy, D., & Saint-Georges, P. de. (2005). *Práticas e Métodos de Investigação em Ciências Sociais* (2º). Lisboa: Gradiva.
- Alvesson, M., & Skoldberg, K. (2010). *Reflexive Methodology: New Vistas for Qualitative Research*. London: SAGE Publications.
- American Libraries. (1997). They Won! and Did It ALA's Way. *American Libraries*, 28(8), 70–77.
- American Library Association. (2020). Anglo-American Cataloging Rules.
- Anchor, R. (2013). 'More product, less process': method, madness or practice? *Archives & Records*, 34(2), 156–174. <https://doi.org/10.1080/23257962.2013.818937>
- Andersen, M. (1993). Studying across difference: Race, class, and gender in qualitative research. In J. Stanfield & R. Dennis (Eds.), *Race and ethnicity in research methods* (pp. 39–52). Newbury Park, CA: SAGE Publications.
- Anderson, G., & Arsenault, N. (1999). *Fundamentals of educational research*. London: Falmer Press Teachers Library.
- Andrade, M., Berti Júnior, D., Pereira, G., & Baptista, A. (2014). Esquemas de metadados utilizados por repositórios digitais científicos e de objetos educacionais : estudo preliminar. *Cadernos de Biblioteconomia, Arquivística e Documentação*, (2), 219–222.
- Ankersmit, F. (2001). *Historical representation*. Stanford University Press.
- Apostolou, C. (2009). *The role of metadata in managing knowledge*. Loughborough University.
- Apple, M. (1979). *Ideology and curriculum*. Boston, MA: Routledge & Kegan Paul.
- Ary, D., L., J., & Sorensen, C. (2010). *Introduction to research in education* (8th ed.). Belmont: Wadsworth.
- Aubert, M., Setiawan, P., Oktaviana, A. A., Brumm, A., Sulistyarto, P. H., Saptomo, E. W., ... Brand, H. E. A. (2018). Palaeolithic cave art in Borneo. *Nature*, 564(7735), 254–257. <https://doi.org/10.1038/s41586-018-0679-9>
- Austin, J. L. (1962). *How to do things with words*. Clarendon Press.
- Avram, H. (1975). *MARC: History and Implications*. Washington, D.C.: Library of Congress.
- Baca, M. (Ed.). (2016). *Introduction to metadata* (3º). Los Angeles: Getty Research Institute.
- Bagley, P. (1968). *Extension of programming language concepts*. Philadelphia: University City Science Center.
- Bailey, J. (2013). Disrespect des fonds: rethinking arrangement and description in born-digital archives. *Archive Journal*, (3). Disponível em: <http://www.archivejournal.net/issue/3/archives-remixed/disrespect-des-fonds-rethinking-arrangement-and-description-in-born-digital-archives>
- Bailey, K. (1982). *Methods of social research* (2nd ed.). New York: Free Press.

- Bailey, K. (1994). *Methods of social research*. New York: The Free Press.
- Banks, M. (2007). *Qualitative Research kit: Using visual data in qualitative research*. London: SAGE Publications. <https://doi.org/10.4135/9780857020260>
- Baños-Moreno, M., Valentin Ruiz, F., & Blázquez Martín de las Mulas, A. (2019). e-EMGDE, RIC, NEDA y las normas de codificación: más allá de un perfil de aplicación en archivos. In *Atas do IV Congresso ISKO Espanha-Portugal*. Barcelona.
- Bantin, P. C. (1998). Strategies for Managing Electronic Records: A New Archival Paradigm? An Affirmation of our Archival Traditions? *Archival Issues*, 23(1), 17–34.
- Baptista, A. M., & Gomes, M. V. (1995). Arte rupestre do Vale do Côa. 1. Canada do Inferno. Primeiras impressões. In *1º Congresso de Arqueologia Peninsular* (pp. 350–385). Sociedade Portuguesa de Antropologia e Etnologia.
- Barbedo, F., Ramalho, J. C., Corujo, L., Faria, L., Ferreira, M., & Castro, R. (2007). RODA: Repositório de Objetos Digitais Autênticos - Relatório Final.
- Bardin, L. (2009). *Análise de Conteúdo*. Lisboa: Edições 70.
- Barroso, H. (1999). *Forma e substância da expressão da Língua Portuguesa*. Almedina.
- Bartlett, N. (1992). Respect des Fonds: *Primary Sources & Original Works*, 1(1–2), 107–115. [https://doi.org/10.1300/J269V01N01\\_07](https://doi.org/10.1300/J269V01N01_07)
- Bates, M. (2011). Birger Hjørland's Manichean Misconstruction of Marcia Bates' Work. *Journal of the American Society for Information Science and Technology*, 62(10), 2038–2044. <https://doi.org/10.1002/asi>
- Beak, J. (2014). *A Child-Driven Metadata Schema: A Holistic Analysis of Children's Cognitive Processes During Book Selection*. University of Wisconsin-Milwaukee. Disponível em: [http://search.proquest.com/docview/1610812495?accountid=14643%5Cnhttp://mlbsfx.sibi.usp.br:3410/sfxlcl41?url\\_ver=Z39.88-2004&rft\\_val\\_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:ProQuest+Dissertations+%26+Theses+Global&ati](http://search.proquest.com/docview/1610812495?accountid=14643%5Cnhttp://mlbsfx.sibi.usp.br:3410/sfxlcl41?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:ProQuest+Dissertations+%26+Theses+Global&ati)
- Beall, J., & Kafadar, K. (2005). The proportion of NUC pre-56 titles represented in OCLC WorldCat. *College and Research Libraries*, 66(5), 431–435. <https://doi.org/10.5860/crl.66.5.431>
- Bearman, D. (1989). Archival methods. *Archives and Museum Informatics Technical Report*, 3(1). Disponível em: [http://www.archimuse.com/publishing/archival\\_methods/](http://www.archimuse.com/publishing/archival_methods/)
- Bearman, D. (1996). Item level control and electronic recordkeeping. *Archives and Museum Informatics*, 10(3), 195–245. <https://doi.org/10.1007/bf02802369>
- Bearman, D. (2006). Moments of risk: Identifying threats to electronic records. *Archivaria*, 62(1), 15–46. <https://doi.org/10.1016/j.jsat.2012.12.012.Criminal>
- Bell, G. (2013). Commercial and cultural sectors: potential for data collaboration? *JLIS.It*, 4(1), 291–304. <https://doi.org/10.4403/jlis.it-5487>
- Bennaceur, A., Tun, T., Yu, Y., & Nuseibeh, B. (2019). Requirements Engineering. In S. Cha, R. Taylor, & K. Kang (Eds.), *Handbook of Software Engineering* (pp. 51–92). Springer International Publishing. <https://doi.org/10.1007/978-3-030-00262-6>
- Berners-Lee, T. (1997). Web design issues. Disponível em: <https://www.w3.org/DesignIssues/UI.html>
- Berners-Lee, T. (2006). Linked Data. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *The Scientific American*, 5(284), 34–43. <https://doi.org/10.1038/scientificamerican0501-34>
- Beyene, W. (2017). Metadata and universal access in digital library environments. *Library Hi Tech*, 35(2), 210–221. <https://doi.org/10.1108/LHT-06-2016-0074>
- Beyene, W., & Godwin, T. (2018). Accessible search and the role of metadata. *Library Hi Tech*, 36(1), 2–17. <https://doi.org/10.1108/LHT-08-2017-0170>
- Beynon-Davies, P. (2009). *Business Information Systems*. Palgrave Macmillan.
- Bhatia, S., & Wright de Hernandez, A. D. (2019). Blockchain Is Already Here. What Does That Mean for Records Management and Archives? *Journal of Archival Organization*, 16(1), 1–10. <https://doi.org/10.1080/15332748.2019.1655614>
- Biddle, B., & Anderson, D. (1986). Theory, methods, knowledge and research on teaching. In M.



- Wittrock (Ed.), *Handbook of Research on teaching* (3<sup>o</sup>, pp. 230–252). New York: MacMillan.
- Bishop, G., Oldendick, R., Tuchfarber, A., & Bennett, S. (1980). Pseudo-Opinions on Public Affairs. *Public Opinion Quarterly*, 44, 198–209.
- Bisquerra, R. (1989). *Métodos de investigación educativa: guía práctica*. Barcelona: Ediciones CEAC.
- Blair, D. (1990). *Language and representation in information retrieval*. New York: Elsevier Science Publishers.
- Blake, V. (2002). Forging the Anglo-American Cataloging Alliance: Descriptive Cataloging, 1830–1908. *Cataloging & Classification Quarterly*, 35(1), 3–22. <https://doi.org/10.1300/J104v35n01>
- Bleicher, J. (1980). *Contemporary Hermeneutics: Hermeneutics as Method, Philosophy, and Critique*. London: Routledge.
- Blum, R. (1991). *Kallimachos: The Alexandrian Library and the Origins of Bibliography*. Madison: University of Wisconsin Press.
- Bogdan, R., & Biklen, S. (1994). *Investigação qualitativa em Educação: uma introdução à teoria e aos métodos*. Porto: Porto Editora.
- Boles, F. (1982). Disrespecting original order. *American Archivist*, 45(1), 26–32. <https://doi.org/10.17723/aarc.45.1.94g0502t2g81053g>
- Borges, M. M. (2006). *A Esfera: Comunicação Acadêmica e Novos Media*. Coimbra university. Disponível em: <http://hdl.handle.net/10316/8557>
- Boring, E. (1930). A new ambiguous figure. *The American Journal of Psychology*, 42(3), 444–445.
- Borrego Huerta, Á. (1999). La investigación cualitativa y sus aplicaciones en Biblioteconomía y Documentación. *Revista Española de Documentación Científica*, 22(2), 139–156. <https://doi.org/10.3989/redc.1999.v22.i2.335>
- Botão, A. (2011). *Recuperação da informação digital: a Norma Brasileira de Descrição Arquivística (NOBRADE) na descrição de material imagético*. Universidade Federal do Rio de Janeiro.
- Bountouri, L., Gratz, P., & Sanmartin, F. (2017). Digital preservation: how to be trustworthy. In *Conference of the Marie Skłodowska-Curie Initial Training Network for Digital Cultural Heritage (ITN-DCH)* (Vol. 10605, pp. 364–374). Olímje, Slovenia: Springer International Publishing. <https://doi.org/10.1007/978-3-319-75826-8>
- Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/qrj0902027>
- Bradburn, N., Sudman, S., Blair, E., Locander, W., Miles, C., Singer, E., & Stocking, C. (1979). *Improving Method and Questionnaire Design*. San Francisco: Jossey-Bass.
- Braut, N. (1972). *The Great Debate on Panizzi's Rules in 1847–1849: The Issues Discussed*. Los Angeles: School of Library Service, University of California.
- Bravo, P., & Eisman, L. (1998). *Investigación Educativa* (3<sup>o</sup>). Sevilla: Ediciones Alfar.
- Brewer, J., & Hunter, A. (2006). *Foundations of Multimethod Research: Synthesizing styles*. Thousand Oaks, CA: SAGE Publications.
- Brickley, D., & Miller, L. (2014). FOAF Vocabulary Specification 0.99.
- British Broadcasting Corporation. (2013, July 6). Transparency and secrets. *Moral Maze*.
- British Library. (2020). Diamond Sutra. Disponível em: <http://vll-minos.bl.uk/onlinegallery/sacredtexts/diamondsutra.html>
- Brooks, S. (2000). A History of the Card Catalog. In G. Eberhart (Ed.), *Whole library handbook 3* (pp. 311–314). American Library Association.
- Brothman, B. (1991). Orders of Value: Probing the Theoretical Terms of Archival Practice. *Archivaria*, 32((Summer 1991)), 78–100.
- Brothman, B. (1993). The Limits of Limits: Derridean Deconstruction and the Archival Institution. *Archivaria*, 36((Autumn 1993)), 205–220.
- Brothman, B. (1999). Declining Derrida: Integrity, tensegrity, and the preservation of archives from deconstruction. *Archivaria*, 48(Fall (1999)), 64–88.
- Brown, C. (2015). Archival Retrieval. In L. Duranti & P. Franks (Eds.), *Encyclopedia of Archival Science* (pp. 82–83). Lanham: Rowman & Littlefield.
- Brown, C. (Ed.). (2019). *Archival futures*. London: Facet Publishing.

- Brown, R. (1991a). Records Acquisition Strategy and its Theoretical Foundation: The Case for a Concept of Archival Hermeneutics. *Archivaria*, 33((Winter 1991-92)), 34–56.
- Brown, R. (1991b). The Value of “Narrativity” in the Appraisal of Historical Documents: Foundation for a Theory of Archival Hermeneutics. *Archivaria*, 32((Summer 1991)), 152–156.
- Brown, R. (1997). Death of a Renaissance Record-Keeper: The Murder of Tomasso da Tortona in Ferrara, 1385. *Archivaria*, 44((Fall 1997)), 1–43.
- Bryman, A. (1992). Quantitative and qualitative research: further reflections on their integration. In Julia Brannen (Ed.), *Mixing Methods: Qualitative and Quantitative Research* (pp. 57–80). Aldershot, England: Avebury.
- Buckland, M. (2012). What Kind of Science Can Information Science Be? *Journal of the American Society for Information Science and Technology*, 1(1), 1–7. <https://doi.org/10.1002/asi>
- Bunn, J. (2013). Developing descriptive standards: a renewed call to action. *Archives & Records*. Taylor & Francis. <https://doi.org/10.1080/23257962.2013.830066>
- Bureau Canadien des Archivistes. (1986). *Les normes de description en archivistique: une nécessité; rapport et recommandations du groupe de travail canadien sur les normes de description en archivistique*. Ottawa: Bureau Canadien des Archivistes.
- Bureau of Canadian Archivists. (2008). Rules for Archival Description. Ottawa, Canada: Bureau of Canadian Archivists. Disponível em: [http://www.cdncouncilarchives.ca/RAD/RADComplete\\_July2008.pdf](http://www.cdncouncilarchives.ca/RAD/RADComplete_July2008.pdf) (en);
- Caplan, P. (2000). International Metadata Initiatives: lessons in bibliographic control. In *Conference on Bibliographic Control in the New Millennium: Confronting the challenges of networked resources and the web*. Washington, D.C.
- Capurro, R., & Hjørland, B. (2003). The Concept of Information. *Annual Review of Information Science and Technology*, 37(chapter 8), 343–411. <https://doi.org/DOI:10.1002/aris.1440370109>
- Carmo, H., & Ferreira, M. M. (1998). *Metodologia da Investigação: guia para auto-aprendizagem*. Lisboa: Universidade Aberta.
- Carr, W., & Kemmis, S. (1988). *Teoría crítica de la enseñanza: La investigación-acción en la formación del profesorado*. Barcelona: Martínez Roca.
- Carta, G. (2017). Metadata and video games emulation: an effective bond to achieve authentic preservation? *Records Management Journal*, 27(2), 192–204. <https://doi.org/10.1108/RMJ-10-2016-0037> Permanent
- Casson, L. (2001). *Libraries in the ancient world*. New Haven: Yale University Press.
- Castells, M. (2002). *A era da informação: economia, sociedade e cultura*. Lisboa: Fundação Calouste Gulbenkian.
- Castells, M. (2010). *The Rise of the Network Society* (2º). Oxford: Wiley-Blackwell.
- CC:DA. (2000). Task Force on Metadata: Final report. Committee on Cataloging: Description and Access. Disponível em: <https://www.libraries.psu.edu/tas/jca/ccda/tf-meta6.html>
- Cencetti, G. (1939). Il fondamento teorico della dottrina archivistica. *Archivi II*, 6, 7–13.
- Cerf, V., & Kahn, R. (1974). A protocol for packet network intercommunication. *IEEE Transactions on Communications*, 22(5), 637–648. <https://doi.org/10.1145/1064413.1064423>
- Chan, L. M., & Zeng, M. L. (2006). Metadata Interoperability and Standardization - A Study of Methodology Part I. *D-Lib Magazine*, 12(6). <https://doi.org/10.1045/june2006-chan>
- Cheetham, A., Ayotte, D., Hung, J., Vass, J., Clark, C., Mitchell, J., & Treviranus, J. (2014). Accessible metadata generation. In C. Stephanidis & M. Antona (Eds.), *Universal Access in Human-Computer Interaction*. (pp. 101–110). Springer International Publishing. [https://doi.org/10.1007/978-3-319-07440-5\\_10](https://doi.org/10.1007/978-3-319-07440-5_10)
- Chen, Y.-N., Wen, C.-Y., Chen, H.-P., Lin, Y.-H., & Sum, H.-C. (2011). Metrics for metadata quality assurance and their implications for digital libraries. In C. Xing, F. Crestani, & A. Rauber (Eds.), *Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation. ICADL 2011. Lecture Notes in Computer Science, vol 7008* (pp. 138–147). Berlin: Springer. [https://doi.org/https://doi.org/10.1007/978-3-642-24826-9\\_19](https://doi.org/https://doi.org/10.1007/978-3-642-24826-9_19)
- Chu, H. (2010). *Information representation and retrieval in the Digital Age* (2nd ed.). Medford, NJ: American Society for Information Science and Technology.

- Clark, H., & Schober, M. (1982). Asking questions and influencing answers of attitude. In J. Tanur (Ed.), *Questions and Questions: Inquiries into the Cognitive Bases of Surveys* (pp. 14–48). New York: Russel Sage Foundation.
- Clifford, J., & Marcus, G. (Eds.). (1986). *Writing Culture: The Poetics and Politics of Ethnography*. Berkeley, CA: The University of California Press.
- Clobridge, A. (2010). *Building a Digital Repository Program with Limited Resources*. Oxford: Chandos Publishing.
- Connaway, L., & Powell, R. (2007). *Basic research methods for librarians* (5th ed.). Santa Barbara: Libraries Unlimited.
- Considine, J. (2011). Leibniz as lexicographer? In G. Hassler & G. Volkmann (Eds.), *History of Linguistics 2008: Selected Papers from the Eleventh International Conference on the History of the Language Sciences (ICHoLS XI), 28 August -2 September 2008, Potsdam*. John Benjamins Publishing.
- Contessa, G. (2007). Scientific Representation, Interpretation, and Surrogate Reasoning. *Philosophy of Science*, 74(1), 48–68. <https://doi.org/10.1086/519478>
- Conway, P. (n.d.). Modes of Seeing : Digitized. *American Archivist*, 73(2), 425–462.
- Conway, P. (1996). *Preservation in the Digital World*. Washington, D.C.: Commission on Preservation and Access.
- Cook, M. (1986). *The management of information from archives*. Brookfield, Vermont: Gower Publishing.
- Cook, M. (1999). *The management of information from archives* (2nd ed.). London: Routledge. <https://doi.org/10.4324/9781315238432>
- Cook, M., & Procter, M. (1989). *Manual of archival description* (2nd ed.). Aldershot, England: Gower Publishing.
- Cook, T. (1995). Keeping Our Electronic Memory: approaches for securing computer-generated records. *South African Archives Journal*, 37, 79–95.
- Cook, T. (1997). What is Past is Prologue: A History of Archival Ideas Since 1898, and the Future Paradigm Shift. *Archivaria*, 43((Spring 1997)), 17–63.
- Cook, T. (1999). Archival Appraisal and Collection: Issues, Challenges, New Approaches. Special Lecture Series to the University of Maryland and to NARA Staff. Disponível em: <http://www.mybestdocs.com/cookt-nara-990421-2.htm>
- Cook, T. (2001). Archival science and postmodernism: new formulations for old concepts. *Archival Science*, 1(March), 3–24. Disponível em: <http://www.polonistyka.uj.edu.pl/documents/41623/111f093d-a2af-4fc6-8f9a-e193d85712a5>
- Cook, T. (2007). Electronic records, paper minds: the revolution in Information management and archives in the post-custodial and post-modernist era. *Archives & Social Studies: A Journal of Interdisciplinary Research*, 1(0), 399–443. Disponível em: [http://archivo.cartagena.es/files/36-164-DOC\\_FICHERO1/06-cook\\_electronic.pdf](http://archivo.cartagena.es/files/36-164-DOC_FICHERO1/06-cook_electronic.pdf)
- Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3<sup>o</sup>). Thousand Oaks: SAGE Publications.
- Corujo, L. M. N. *Avaliação Arquivística de Informação Eletrônica: da emergência teórica ao consenso sobre um modelo*. 2020. University of Coimbra, 2020. Disponível em: <<https://estudogeral.sib.uc.pt/handle/10316/95276>>.
- Corujo, L., & Barbedo, F. (2012). MIP: Metainformação para Interoperabilidade. Disponível em: [http://arquivos.dglab.gov.pt/wp-content/uploads/sites/16/2013/10/MIP\\_v1-0c.pdf](http://arquivos.dglab.gov.pt/wp-content/uploads/sites/16/2013/10/MIP_v1-0c.pdf)
- Coutinho, C. (2019). *Metodologia de Investigação em Ciências Sociais e Humanas: Teoria e Prática* (2<sup>a</sup>). Coimbra: Almedina.
- Cox, R. (1997). Electronic Systems and Records Management in the Information Age: an introduction. *Bulletin of the American Society for Information Science and Technology*. <https://doi.org/10.1002/bult.59>
- Creswell, J. (2014). *Research design: Qualitative, Quantitative, and Mixed Methods Approaches* (4<sup>o</sup>). Thousand Oaks: SAGE Publications.
- Creswell, J., & Clark, V. (2013). *Pesquisa de métodos mistos* (2nd ed.). Porto Alegre: Penso.
- Creswell, J., & Clark, V. (2018). *Designing and conducting mixed methods research* (3<sup>o</sup>). Los

- Angeles: SAGE Publications.
- Cron, B. (2016). Metadata in Electronic Records Management. *Records Express*. Disponível em: <https://records-express.blogs.archives.gov/2016/11/21/metadata-in-electronic-records-management/>
- Crotty, M. (1998). *The foundations of social research: meaning and perspective in the research process*. London: SAGE.
- Culkin, J. M. (1967). A schoolman's guide to Marshall McLuhan. *Saturday Review*, 51–53, 70–72.
- Cumming, K. (2010). Ways of seeing: Contextualising the continuum. *Records Management Journal*, 20(1), 41–52. <https://doi.org/10.1108/09565691011036224>
- Cummins, R. (1996). *Representations, targets and attitudes*. MIT Press.
- Cutter, C. (1904). *Rules for a dictionary catalog* (4th ed.). Washington, D.C.: Government Printing Office. Disponível em: <https://archive.org/details/rulesforadictio06cuttgoog/page/n14/mode/2up>
- Cziko, G. (1989). Unpredictability and Indeterminism in Human Behavior: Arguments and Implications for Educational Research. *Educational Researcher*, 18(3), 17–25.
- Daniel, L., & Daniel, L. (2012). *Digital Forensics for Legal Professionals: Understanding Digital Evidence From the Warrant to the Courtroom*. Waltham, MA: Syngress.
- Day, M. (2001). Metadata in a nutshell. *Information Europe*, 6(2), 11. Disponível em: <http://www.ukoln.ac.uk/metadata/publications/nutshell/>
- Dempsey, L., & Weibel, S. L. (1996). The Warwick Metadata Workshop: a Framework for the Development of Resource Description. *D-Lib Magazine*.
- Dennis, M. A. (2019). Defense Advanced Research Projects Agency. In *Encyclopaedia Britannica*.
- Denton, W. (2007). FRBR and the history of cataloging. In A. Taylor (Ed.), *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools* (pp. 35–57).
- Denzin, N. (1989). *The research act: a theoretical introduction to sociological methods* (3<sup>o</sup>). London: Prentice Hall.
- Denzin, N. (2012). Triangulation 2.0. *Journal of Mixed Methods Research*, 6(2), 80–88. <https://doi.org/10.1177/1558689812437186>
- Denzin, N., & Lincoln, Y. (2005). Introduction: The Discipline and practice of Qualitative Research. In *The Sage Handbook of Qualitative Research* (pp. 1–32). Thousand Oaks, CA, US: SAGE.
- Dervos, D., & Coleman, A. (2006). A common sense approach to defining data, information and metadata. In *Knowledge organization for a global learning society: proceedings of the Ninth International ISKO Conference*. Berlin, Germany: Ergon.
- Desjardins, G. (1890). *Le Service des archives départementales, conférences faites aux élèves de l'École des chartes, les 10, 18, 25 et 30 juin 1890*. Paris: E. Bourloton.
- Devlin, K. (1999). *Info-Senso: Como Transformar a Informação em Conhecimento*. Lisboa: Livros do Brasil.
- Dicionário Priberam da Língua Portuguesa. (2020a). Meta-. Disponível em: <https://dicionario.priberam.org/meta->
- Dicionário Priberam da Língua Portuguesa. (2020b). Positivo.
- Digital Preservation Coalition. (2015). *Digital Preservation Handbook* (2nd ed.).
- Dodge, B. (1997). Places apart: Archives in dissolving space and time. *Archivaria*, 44(January), 118–131.
- DOMO. (2018). Data never sleeps. Disponível em: <https://www.domo.com/solution/data-never-sleeps-6>
- Dooley, J. (1998). *Encoded Archival Description: Context, Theory, and Case Studies*. Society of American Archivists.
- Douglas, J. (2016). Toward More Honest Description. *American Archivist*, 79(1), 26–55. <https://doi.org/10.17723/0360-9081.79.1.26>
- Dow, E. (2009). Encoded archival description as a halfway technology. *Journal of Archival Organization*, 7(3), 108–115. <https://doi.org/10.1080/15332740903117701>
- Dryden, J. (2009). The Open Archival Information System Reference Model. *Journal of Archival Organization*, 7(4), 214–217. <https://doi.org/10.1080/15332740903334116>
- Dryden, J. (2011). Measuring Trust: Standards for Trusted Digital Repositories. *Journal of Archival*

- Organization*, 9(2), 127–130. <https://doi.org/10.1080/15332748.2011.590744>
- Dublin Core Metadata Initiative. (2021a). DCMI Metadata Terms. Disponível em: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>
- Dublin Core Metadata Initiative. (2021b). Dublin Core. Disponível em: <https://www.dublincore.org/specifications/dublin-core/>
- Duchain, M. (1977). Le respect des fonds en Archivistique: principes théoriques et problèmes pratiques. *La Gazette Des Archives*, 2(97), 71–96.
- Duchain, M. (1983). Theoretical Principles and Practical Problems of Respect des fonds in Archival Science. *Archivaria*, 16(Summer), 64–82. Disponível em: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Theoretical+Principles+and+Practical+Problems+of+Respect+des+Fonds+in+Archival+Science#0>
- Ducheva, D., & Pennington, D. (2019). RDA in Europe: Implementations and perceptions. *Journal of Librarianship and Information Science*, 51(2), 387–402. <https://doi.org/10.1177/0961000617709060>
- Duff, W. (1995). Will metadata replace archival description: a commentary. *Archivaria*, (39), 33–38.
- Duff, W. (1998). Harnessing the Power of Warrant. *American Archivist*, 61(1), 88–105.
- Duff, W., & Harris, V. (2002). Stories and names: archival description as narrating records and constructing meanings. *Archival Science*, 2, 263–285.
- Duff, W., Yakel, E., & Tibbo, H. (2013). Archival reference knowledge. *American Archivist*, 76(1), 68–94.
- Duncan, C. (2009). Counterpoint: authenticity or bust. *Archivaria*, (68), 97–118.
- Duranti, L. (1989). Diplomatics: new uses for an old science (part I). *Archivaria*, 28(Summer), 7–27. Disponível em: <http://journals.sfu.ca/archivar/index.php/archivaria/article/viewFile/11567/12513>
- Duranti, L. (1993). The archival body of knowledge: archival theory, method, and practice, and graduate and continuing education. *Journal of Education for Library and Information Science*, 34(1), 8–24. <https://doi.org/10.2307/40323707>
- Duranti, L. (1995). Reliability and Authenticity: The Concepts and Their Implications. *Archivaria*, 39, 5–10.
- Duranti, L. (1996). Archival Science. In *Encyclopedia of Library and Information Sciences* (3<sup>o</sup>).
- Duranti, L. (1997). The Archival Bond. *Archives and Museum Informatics*, 11(3), 213–218. <https://doi.org/10.1023/A:1009025127463>
- Duranti, L. (2005). *The InterPARES Project: The Long-term Preservation of Authentic Electronic Records: the Findings of the InterPARES Project*. San Miniato: Archilab. Disponível em: <https://books.google.pt/books?id=FmBoAAAAMAAJ>
- Duranti, L. (2010). From Digital Diplomatics to Digital Records Forensics. *Archivaria*, 68, 39–66.
- Duranti, L. (2015). Archival bond. In L. Duranti & P. Franks (Eds.), *Encyclopedia of Archival Science* (pp. 28–29). Lanham: Rowman & Littlefield.
- Duranti, L. (2018). Whose truth? Records and archives as evidence in the era of post-truth and disinformation. In C. Brown (Ed.), *Archival Futures* (pp. 19–32). London: Facet Publishing.
- Duranti, L., & Blanchette, J. (2004). The authenticity of electronic records: the InterPARES approach. Disponível em: [http://www.interpares.org/display\\_file.cfm?doc=ip1-2\\_dissemination\\_cpr\\_duranti~blanchette\\_ist\\_2004.pdf](http://www.interpares.org/display_file.cfm?doc=ip1-2_dissemination_cpr_duranti~blanchette_ist_2004.pdf)
- Duranti, L., & Eastwood, T. (1995). Protecting electronic evidence: a progress report on research study and its methodology. *Archivi & Computer*, (3), 213–250.
- Duranti, L., Eastwood, T., & MacNeil, H. (2002). *Preservation of the Integrity of Electronic Records*. Dordrecht: Kluwer Academic Publishers.
- Duranti, L., & Franks, P. (Eds.). (2015). *Encyclopedia of archival science*. Lanham: Rowman & Littlefield.
- Duranti, L., Rogers, C., Mcleod, J., Shepherd, E., Foscarini, F., Shabou, B., ... Pearce-Moses, R. (2019). Trusting Records and Data in the Cloud: the creation, management, and preservation of trustworthy digital content. <https://doi.org/10.29085/9781783304042>
- Duval, E., Hodgins, W., Sutton, S., & Weibel, S. (2002). Metadata Principles. *D-Lib Magazine*, 8(4). Disponível em: <http://www.dlib.org/dlib/april02/weibel/04weibel.html>

- Eastwood, T. (1993). Reflections on the development of archives in Canada and Australia. In S. McKemmish & F. Upward (Eds.), *Archival Documents: Providing Accountability Through Recordkeeping*. Melbourne: Ancora Press.
- Eastwood, T. (1994). What is archival theory and why is it important? *Archivaria*, 37, 122–130. <https://doi.org/10.1590/198053142851>
- Eastwood, T. (2000). Putting the parts of the whole together: Systematic arrangement of archives. *Archivaria*, 50(1), 93–116.
- Eastwood, T. (2010). Archives, Democratic Accountability, and Truth. In C. Avery & M. Holmund (Eds.), *Better Off Forgetting?: Essays on Archives, Public Policy, and Collective Memory* (pp. 143–168). Toronto: University Toronto Press.
- ECLAP. (2011). Metadata descriptors Identification and Definition.
- EDItEUR. (2009). Onix standards: ONIX for Books, ONIX for Subscription Products, Licensing Terms and Rights Information.
- Eisenstein, E. (1980). *The Printing Press as an Agent of Change*. Cambridge: Cambridge University Press.
- Elings, M., & Waibel, G. (2007). Metadata for all: Descriptive standards and metadata sharing across libraries, archives and museums. *First Monday*, 12(3). <https://doi.org/10.5210/fm.v12i3.1628>
- Engvall, T. (2019). Situating trust challenges of online trade. *Records Management Journal*, 29(1/2), 272–286.
- Espanha. (2016). *Esquema de Metadatos para la Gestión del Documento Electrónico. Versión 2.0*. Disponível em: <http://www.mcu.es/ccbae/es/consulta/registro.cmd?id=220251>
- European Committee for Standardization. (2020). What is a standard? Disponível em: <https://www.cen.eu/work/endev/whatisen/pages/default.aspx>
- Europeana. (2020). About us. Disponível em: <https://pro.europeana.eu/about-us/mission>
- Eysenbach, G., Yihune, G., Lampe, K., Cross, P., & Brickley, D. (2000). MedCERTAIN: quality management, certification and rating of health information on the Net. *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, 230–234.
- Factor, M., Henis, E., Naor, D., Rabinovici-Cohen, S., Reshef, P., & Ronen, S. (2009). Authenticity and Provenance in Long Term Digital Preservation: Modeling and Implementation in Preservation Aware Storage. *Workshop on Theory and Practice of Provenance*. Disponível em: [https://www.usenix.org/legacy/event/tapp09/tech/full\\_papers/factor/factor.pdf](https://www.usenix.org/legacy/event/tapp09/tech/full_papers/factor/factor.pdf)
- Faria, M. I., & Pericão, M. da G. (2008). *Dicionário do Livro*. Coimbra: Edições Almedina.
- Ferreira, M., Saraiva, R., & Rodrigues, E. (2012). *Estado da arte em preservação digital*. Disponível em: <http://repositorium.sdum.uminho.pt/handle/1822/17049>
- Fielding, R. (2000). *Architectural styles and the design of network-based software architectures*. University of California.
- Finnegan, R. (1996). Using documents. In R. Sapsford & V. Jupp (Eds.), *Data Collection and Analysis* (pp. 138–151). London: SAGE.
- Flick, U. (2002). *An Introduction to Qualitative Research* (2º). Thousand Oaks, CA: SAGE Publications.
- Flick, U. (2007). *El diseño de Investigación Cualitativa*. Madrid: Ediciones Morata.
- Foster, S. (1993). *Ethnographic interviews in disability studies: The case of research with people who are deaf*. Paper presented at the annual meeting of the American Sociological Association, Miami, FL.
- Foucault, M. (1980). Questions on Geography. In *Power/Knowledge: Selected Interviews and Other Writings, 1972-1977* (pp. 173–182). New York: Pantheon.
- Frankfield, J. (2019). Smart Contracts. In *Investopedia*. Disponível em: <https://www.investopedia.com/terms/s/smart-contracts.asp>
- Freitas, Cristiana. (2012). Garantir a autenticidade e o acesso continuado à informação digital: os desafios da preservação digital em arquivos. *Actas Do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas*, (11). Disponível em: <https://www.bad.pt/publicacoes/index.php/congressosbad/article/viewFile/272/pdf>
- Freitas, Cristina. (2009). *A arquivística sob o signo da mudança: cenários arquivísticos (re)desenhados pelo documento eletrônico*. Universidade de Salamanca.

- Freitas, Cristina de. (2017). Revistas especializadas em temas arquivísticos e em acesso aberto: uma cartografia necessária. In *A Ciência Aberta: o Contributo da Ciência da Informação: atas do VIII Encontro Ibérico EDICIC* (pp. 211–226). Universidade de Coimbra. Centro de Estudos Interdisciplinares do Século XX - CEIS20.
- Freitas, Cristina de, & Silva, C. G. da. (2016). O novo e o atual na Arquivística internacional: a desmaterialização, a interoperabilidade, a organização e o uso da informação em evidência (2011-2016). In *12º Encontro de Arquivos Municipais*. Disponível em: [http://repositorio.ul.pt/bitstream/10451/30018/1/O\\_novo\\_e\\_o\\_atual\\_na\\_arquivistica\\_internacional.pdf](http://repositorio.ul.pt/bitstream/10451/30018/1/O_novo_e_o_atual_na_arquivistica_internacional.pdf)
- Frigg, R. (2002). Models and representation: why structures are not enough. *Measurement in Physics and Economics Project Discussion Paper Series*. Disponível em: <http://hdl.handle.net/10065/1042>
- Frigg, R., & Nguyen, J. (2020). Scientific representation. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Disponível em: <https://plato.stanford.edu/archives/spr2020/entries/scientific-representation/>
- Fritsch, A. (1664). *Tractatus de jure archivi et cancellariae*. Jena.
- Frogner, R. (2015). “Lord, save us from the Et cetera of the notary”: Archival appraisal, local custom, and colonial law. *Archivaria*, 79(Spring), 121–158.
- Fuller, B. (1982). *Critical path*. New York: St. Martin’s Press.
- Furner, J. (2020). Definitions of “Metadata”: A Brief Survey of International Standards. *Journal of the Association for Information Science and Technology*, 71(6), 33–42. <https://doi.org/10.1002/asi.24295>
- Gadamer, H.-G. (2013). *Truth and method*. London: Bloomsbury.
- García Roldán, J. (2003). *Cómo elaborar un proyecto de investigación*. Alicante: Publicaciones de la Universidad de Alicante.
- Garshol, L. M. (2004). Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all. *Journal of Information Science*, 30(4), 378–391. <https://doi.org/10.1177/0165551504045856>
- Gartner, R. (2015). An XML schema for enhancing the semantic interoperability of archival description. *Archival Science*, 15(3), 295–313. <https://doi.org/10.1007/s10502-014-9225-1>
- Gartner, R. (2016). *Metadata: shaping knowledge from antiquity to the Semantic Web*. Springer.
- Gilliland, A. (2016). Setting the stage. In M. Baca (Ed.), *Introduction to Metadata* (3rd ed, pp. 1–20). Los Angeles: Getty Research Institute.
- Gilliland, A. (2018). Evidence and exigency: Reconstructing and Reconciling Records for Life and Conflict. In K. Gracy (Ed.), *Emerging Trends in Archival Science* (pp. 1–26). Lanham, MD: Rowman & Littlefield. Disponível em: [https://books.google.pt/books?id=vIg6DwAAQBAJ&pg=PR11&lpg=PR11&dq=Evidence+and+exigency:+Reconstructing+and+Reconciling+Records+for+Life+and+Conflict&source=bl&ots=CjU\\_tYuy70&sig=ACfU3U3qUVHZFRths5RFC3O0K5tA\\_Pkc3A&hl=en&sa=X&ved=2ahUKÉwi2zMHHwO\\_pAhXYDGMBH](https://books.google.pt/books?id=vIg6DwAAQBAJ&pg=PR11&lpg=PR11&dq=Evidence+and+exigency:+Reconstructing+and+Reconciling+Records+for+Life+and+Conflict&source=bl&ots=CjU_tYuy70&sig=ACfU3U3qUVHZFRths5RFC3O0K5tA_Pkc3A&hl=en&sa=X&ved=2ahUKÉwi2zMHHwO_pAhXYDGMBH)
- Gilliland, A., Lindberg, L., McCargar, V., Langmead, A., & Evans, J. (2008). Part Six— Investigating the Roles and Requirements, Manifestations and Management of Metadata in the Creation of Reliable and Preservation of Authentic Digital Entities: Description Cross-domain Task Force Report. In L. Duranti & R. Preston (Eds.), *International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records*. Padova, Italy: Associazione Nazionale Archivistica Italiana.
- Giroux, H. (1983). *Theory and resistance in Education: a pedagogy for the opposition*. South Hadley, MA: Bergin & Garvey.
- Gladney, H. (2007). *Preserving digital information*. New York: Springer.
- Gladney, H. (2009). Long-term preservation of digital records: Trustworthy digital objects. *American Archivist*, 72(2), 401–435. <https://doi.org/10.17723/aarc.72.2.g513766100731832>
- Goddard, J., & Eccles, T. (2013). *Uncommon Sense, Common Nonsense: why Some Organisations Consistently Outperform Others*. London: Profile Books.
- Gómez, G., Flores, J., & Jimenez, E. (1996). *Tradición y enfoques en la investigación cualitativa. Metodología de la investigación cualitativa*. Málaga: Ediciones Aljibe. Disponível em:

- <http://www.albertomayol.cl/wp-content/uploads/2014/03/Rodriguez-Gil-y-Garcia-Metodologia-Investigacion-Cualitativa-Caps-1-y-2.pdf>
- Goodman, N. (1972). Seven Strictures on Similarity. In N. Goodman (Ed.), *Problems and Projects*. Indianapolis and New York: Bobs-Merril.
- Goodman, N. (1976). *Languages of Art* (2nd ed.). Indianapolis and Cambridge: Hackett.
- Goodson, I. (2015). *Narrativas em educação: a vida e a voz dos professores*. Porto: Porto Editora.
- Goody, J., & Watt, I. (1963). The Consequences of Literacy. *Comparative Studies in Society and History*, 5(3), 304–345. <https://doi.org/10.1017/S0010417500001730>
- Gorman, M. (2000). Seymour Lubetzky, Man of Principles. In T. H. Connell & R. L. Maxwell (Eds.), *The Future of Cataloging: Insights from the Lubetzky Symposium; April 18, 1988, University of California, Los Angeles* (pp. 12–21). Chicago: American Library Association.
- Gracy, K. (2015). Archival description and linked data: a preliminary study of opportunities and implementation challenges. *Archival Science*, 15(3), 239–294. <https://doi.org/10.1007/s10502-014-9216-2>
- Gracy, K., & Lambert, F. (2014). Who's ready to surf the next wave? A study of perceived challenges to implementing new and revised standards for archival description. *American Archivist*, 77(1), 96–132. <https://doi.org/10.17723/aarc.77.1.b241071w5r252612>
- Greene, J. (2007). *Mixed methods in social inquiry*. San Francisco: Jossey-Bass.
- Greene, M. (2009). The Power of Archives: Archivists' Values and Value in the Postmodern Age. *The American Archivist*, 72(1), 13–41. <https://doi.org/10.17723/aarc.72.1.k0322x0p38v44153>
- Greene, M., & Meissner, D. (2005). More product, less process: Revamping traditional archival processing. *American Archivist*, 68(2), 208–263. <https://doi.org/10.17723/aarc.68.2.c741823776k65863>
- Guba, E. (Ed.). (1990). *The paradigm dialog*. Thousand Oaks, CA: SAGE Publications.
- Guenther, R. (2003). MODS: The Metadata Object Description Schema. *Portal: Libraries and the Academy*, 3(1), 137–150. <https://doi.org/10.1353/pla.2003.0006>
- Habermas, J. (2008). *Teoría y praxis: Estudios de filosofía social* (5ª). Madrid: Editorial Tecnos.
- Hakim, C. (1993). Research analysis of administrative records. In M. Hammersely (Ed.), *Social research, philosophy, politics and practice*. London: Sage.
- Hammersley, M. (1992). Deconstructing the qualitative-quantitative debate. In J. Brannen (Ed.), *Mixing Methods: Qualitative and Quantitative Research* (pp. 39–55). Aldershot, England: Avebury.
- Hammersley, M. (1996). The relationship between qualitative and quantitative research: Paradigm loyalty versus methodological eclecticism. In J. Richardson (Ed.), *Handbook of Qualitative Research Methods for Psychology and the Social Sciences* (pp. 159–174). Leicester: BPS Books.
- Hammersley, M., & P. Atkinson. (1995). *Ethnography: Principles in Practice* (2º). London: Routledge.
- Hanson, N. (1958). *Patterns of discovery: an inquiry into the conceptual foundations of science*. Cambridge: Cambridge University Press.
- Harding, S. (1987). *Feminism and Methodology*. Bloomington: Indiana University Press.
- Harper, D. (2020). Facet. In *Online Etymology Dictionary*. Disponível em: [https://www.etymonline.com/word/facet#etymonline\\_v\\_1057](https://www.etymonline.com/word/facet#etymonline_v_1057)
- Harpring, P. (2019). *Categories for the Description of Works of Art (CDWA)*. (M. Baca & P. Harpring, Eds.). Los Angeles: The Getty. Disponível em: [https://www.getty.edu/research/publications/electronic\\_publications/cdwa/introduction.html](https://www.getty.edu/research/publications/electronic_publications/cdwa/introduction.html)
- Harris, V. (1997). Claiming less, delivering more: A critique of Positivist formulations on archives in South Africa. *Archivaria*, 44(January), 132–141.
- Harvey, F., Kuhn, W., Pundt, H., Bishr, Y., & Riedemann, C. (1999). Semantic interoperability: A central issue for sharing geographic information. *Annals of Regional Science*, 33(2), 213–232. <https://doi.org/10.1007/s001680050102>
- Houghton, B. (2011). What happened to the Great Library at Alexandria? In *Ancient History Encyclopedia*. Disponível em: <https://www.ancient.eu/article/207/what-happened-to-the-great-library-at-alexandria/>
- Haynes, D. (2018). *Metadata for information management and retrieval* (2º). London: Facet



- Publishing.
- Head, R. (2013). Documents, archives, and proof around 1700. *The Historical Journal*, 56(4), 909–930.
- Head, R. (2019). *Making Archives in Early Modern Europe: Proof, Information, and Political Record-Keeping, 1400–1700*. Cambridge: Cambridge University Press.
- Hedstrom, M. (1993). Descriptive practices for electronic records: deciding what is essential and imagining what is possible. *Archivaria*, 36, 53–63.
- Heidegger, M. (2004). *Being and Time*. Oxford: Blackwell.
- Henshilwood, C. S., d’Errico, F., van Niekerk, K. L., Dayet, L., Queffelec, A., & Pollarolo, L. (2018). An abstract drawing from the 73,000-year-old levels at Blombos Cave, South Africa. *Nature*, 562(7725), 115–118. <https://doi.org/10.1038/s41586-018-0514-3>
- Henttonen, P. (2009). A comparison of MoReq and SÄHKE metadata and functional requirements. *Records Management Journal*, 19(1), 26–36. <https://doi.org/10.1108/09565690910937227>
- Hjørland, B. (2007). Information: Objective or Subjective/Situational? *Journal of the American Society for Information Science and Technology*, 58(10), 1448–1456. <https://doi.org/10.1002/asi>
- Hjørland, B. (2009). The controversy over the concept of “information”: A rejoinder to professor bates. *Journal of the American Society for Information Science and Technology*, 60(3), 643. <https://doi.org/10.1002/asi.20972>
- Hlava, M. (2015). *The Taxobook: Principles and Practices of Taxonomy Construction. Part 2 of a 3-Part Series*. California: Morgan & Claypool Publishers.
- Hofman, D., Lemieux, V., Joo, A., & Batista, D. A. (2019). “The margin between the edge of the world and infinite possibility”: Blockchain, GDPR and information governance. *Records Management Journal*, 29(1–2), 240–257. <https://doi.org/10.1108/RMJ-12-2018-0045>
- Holiday, R. (2017). *Trust me, I’m Lying: Confessions of a Media Manipulator*. New York: Portfolio/Penguin.
- Holman, L., Darraj, E., Glaser, J., Hom, A., Mathieson, H., & Nettles, D. (2012). How users approach discovery tools. In M. Popp & D. Dallis (Eds.), *Planning and implementing resource discovery tools in academic libraries* (pp. 252–267). Hershey, PA: Information Science Reference.
- Holmes, O. (1964). Archival arrangement - Five different operations at five different levels. *American Archivist*, 27(1), 21–42. <https://doi.org/10.17723/aarc.27.1.1721857117617w15>
- Hooland, S. van, & Verborgh, R. (2014). *Linked data for libraries, archives and museums*. Croydon: Facet Publishing Production.
- Hopkins, J. (1992). The 1791 French Cataloging Code and the Origins of the Card Catalog. *Libraries and Culture*, 27(4), 378–404.
- Horsman, P. (1993). Taming the elephant: an orthodox approach to the principle of provenance. In S. Abukhanfusa & S. Sydbeck (Eds.), *The principle of provenance: report from the First Stockholm Conference on Archival Theory and the Principle of Provenance*. Stockholm: Swedish National Archives.
- Host, K., Jaksic, D., & Poscic, P. (2018). Overview and comparison of the selected relational data modelling tools. *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1592–1597. <https://doi.org/10.23919/MIPRO.2018.8400286>
- Hudgins, J., Agnew, G., & Brown, E. (1999). *Getting Mileage Out of Metadata: applications for the library*. Chicago: American Library Association.
- Hufford, J. R. (1991). The pragmatic basis of catalog codes: Has the user been ignored? *Cataloging and Classification Quarterly*, 14(1), 27–38. [https://doi.org/10.1300/J104v14n01\\_03](https://doi.org/10.1300/J104v14n01_03)
- IBM. (2006). *The toxic terabyte: how data-dumping threatens business efficiency*.
- International Council on Archives. (1992). *Statement of principles regarding archival description*.
- International Council on Archives. (1994). ISAD(G): General International Standard Archival Description. Ottawa. Disponível em: [http://www.hi-u-tokyo.ac.jp/personal/yokoyama/jugyo99/isad\(g\)e.html](http://www.hi-u-tokyo.ac.jp/personal/yokoyama/jugyo99/isad(g)e.html)
- International Council on Archives. (1996). Code of ethics.
- International Council on Archives. (1997). *Guide for managing electronic records from an archival*

- perspective*. Paris: ICA Studies. Disponível em: [http://www.ica.org/sites/default/files/ICA\\_Study\\_8\\_guide\\_eng.pdf](http://www.ica.org/sites/default/files/ICA_Study_8_guide_eng.pdf)
- International Council on Archives. (2000). *ISAD(G): General International Standard Archival Description* (2<sup>o</sup> ed). Ottawa: International Council on Archives.
- International Council on Archives. (2008). *Principles and Functional Requirements Guidelines and Functional Requirements for Records in Electronic Office Environments - Module 2: Guidelines and Functional Requirements for Electronic Records management SYSTEMS*. Disponível em: [www.ica.org](http://www.ica.org)
- International Council on Archives. (2012). *Principles of access to archives*. Disponível em: <http://www.ica.org/15369/toolkits-guides-manuals-and-guidelines/technical-guidance-on-managing-archives-with-restrictions.html>
- International Council on Archives. (2016). Records in contexts. A conceptual model for archival description. Consultation draft v0.1. <https://doi.org/10.1017/S0940739197000234>
- International Council on Archives. (2019). Records in Contexts: a Conceptual Model for Archival Description (Consultation Draft v0.2). Disponível em: [https://www.ica.org/sites/default/files/ric-cm-0.2\\_preview.pdf](https://www.ica.org/sites/default/files/ric-cm-0.2_preview.pdf)
- International Federation of Library Associations and Institutions. (1998). *Functional Requirements for Bibliographic Records. Final report*. Munich: K.G. Saur Verlag. <https://doi.org/10.11129/detail.9783955531478.8>
- InterPARES 1. (2002). Requirements for assessing and maintaining the authenticity of electronic records.
- InterPARES 2. (2008a). *International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records*. (L. Duranti & R. Preston, Eds.). Italy: Associazione Nazionale Archivistica Italiana. <https://doi.org/10.1108/rmj.2009.28119aae.003>
- InterPARES 2. (2008b). Metadata in the Chain of Preservation Model: Draft Metadata Specification Model.
- InterPARES 2. (2008c). The InterPARES 2 Project Glossary. In L. Duranti & R. Preston (Eds.), *International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records*. Padova, Italy: Associazione Nazionale Archivistica Italiana.
- ISO/IEC 15938. (2002). Information technology — Multimedia content description interface. Geneva: ISO.
- ISO/IEC TR 19583-1. (2019). Information technology — Concepts and usage of metadata — Part 1: Metadata concepts. Geneva: ISO. Disponível em: <https://www.iso.org/obp/ui#iso:std:iso-iec:tr:19583:-1:ed-1:v1:en>
- ISO. (2017). ISO 15836-1:2017: Information and documentation — The Dublin Core metadata element set — Part 1: Core elements.
- ISO 10303. (2020). Automation systems and integration: Product data representation and exchange. Geneva: ISO.
- Jackson, M. (1995). *Software requirements & specifications: a lexicon of practice, principles and prejudices*. New York: ACM Press/Addison-Wesley.
- James, T. (1986). *The first printed catalogue of the Bodleian Library, 1605: a facsimile*. Oxford: Clarendon Press.
- Janes, A. (2012). Of maps and meta-records: Eighty- five years of map cataloguing at the national archives of the United Kingdom. *Archivaria*, 74(Fall 2012), 119–165.
- Jansen, A. (2015). Object-oriented diplomatics: using archival diplomatics in software application development to support authenticity of digital records. *Records Management Journal*, 25(1), 45–55. <https://doi.org/10.1108/RMJ-04-2014-0022>
- Jenkinson, H. (1937). *A Manual of Archival Administration*. London: Percy, Lund, Humphries & Co.
- Jick, T. (1979). Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly*, 24(4), 602–611.
- Jimerson, R. (2015). Archives and memory. In L. Duranti & P. Franks (Eds.), *Encyclopedia of Archival Science* (pp. 99–101). Lanham: Rowman & Littlefield.

- Joe Rogan Experience. (2019). Joe Rogan Experience #1368 - Edward Snowden. Disponível em: <https://www.youtube.com/watch?v=efs3QRr8LWw&t=4175s>
- Johns Hopkins University. (2014). The Price of Privacy: Re-Evaluating the NSA. Disponível em: <https://www.youtube.com/watch?v=kV2HDM86XgI>
- Jones, M. (2018). From catalogues to contextual networks: reconfiguring collection documentation in museums. *Archives & Records*, 39(1), 4–20. <https://doi.org/10.1080/23257962.2017.1407750>
- Joudrey, D., & Taylor, A. (2018). *The organization of information* (4th ed.). Santa Barbara, CA: Libraries Unlimited.
- Kallberg, M. (2012). Archivists 2.0: Redefining the archivist's profession in the digital age. *Records Management Journal*, 22(2), 98–115. <https://doi.org/10.1108/09565691211268162>
- Kaplan, A. (1964). *The Conduct of Inquiry: Methodology for Behavioral Science*. San Francisco: Chandos Publishing.
- Kärberg, T. (2014). Digital preservation and knowledge in the public archives: for whom? *Archives and Records*, 35(2), 126–143. <https://doi.org/10.1080/23257962.2014.942606>
- Karvalics, L. (2007). Information society — what is it exactly? The meaning, history and conceptual framework of an expression.
- Kelton, K., Fleischmann, K., & Wallace, W. (2008). Trust in Digital Information. *Journal of the American Society for Information Science and Technology*, 59(3), 363–374.
- Ketelaar, E. (1999). Archivalisation and Archiving. *Archives and Manuscripts*, 27(1), 54–61.
- Ketelaar, E. (2001). Tacit narratives: the meanings of archives. *Archival Science*, 1, 131–141.
- Kirschenbaum, M. (2008). *Mechanisms: New Media and the Forensic Imagination*. MIT Press.
- Kleinrock, L. (1971). *Information Flow in Large Communication Nets*. Massachusetts Institute of Technology. Disponível em: <http://www.lk.cs.ucla.edu/LK/Bib/REPORT/PhD/proposal.html>
- Koch, T. (2006). Electronic thesis and dissertation services: semantic interoperability, subject access, multilinguality. In *E-Thesis Workshop*. Amsterdam. Disponível em: <http://www.ukoln.ac.uk/ukoln/staff/t.koch/publ/e-thesis-200601.html>
- Koltun, L. (1999). The promise and threat of digital options in an archival age. *Archivaria*, 47(Spring), 114–135.
- Kovačević, A., Ivanović, D., Milosavljević, B., Konjović, Z., & Surla, D. (2011). Automatic extraction of metadata from scientific publications for CRIS systems. *Program - Electronic Library and Information Systems*, 45(4), 376–396. <https://doi.org/10.1108/00330331111182094>
- Krajewski, M. (2011). *Paper Machines: About Cards & Catalogs, 1548-1929*. Cambridge, MA: MIT Press.
- Krippendorff, K. (2004). *Content analysis: an introduction to its methodology*. Los Angeles: SAGE Publications.
- Kuhn, T. (1970). *The Structure of Scientific Revolutions*. *Philosophical Review* (2<sup>a</sup>, Vol. II). Chicago: The University of Chicago Press. <https://doi.org/10.1119/1.1969660>
- Laplante, P. (2013). *Requirements engineering for software and systems*. Boston, MA: Auerbach Publications.
- Lather, P. (1992). Critical frames in educational research: Feminist and post structuralist perspectives. *Theory and Practice*, 31(1), 87–99.
- Latorre, A., Del Rincon, D., & Arnal, J. (1996). *Bases metodológicas de la investigación educativa*. Barcelona: Hurtado Ediciones.
- Lechêne, R. (2020). Printing. In *Encyclopaedia Britannica*.
- Lee, R. (2000). *Métodos não interreferentes em pesquisa social*. Lisboa: Gradiva.
- Lehnus, D. (1972). A comparison of Panizzi's 91 rules and the AACR of 1967. Urbana Champaign: University of Illinois Graduate School of Library Science.
- Leiner, B., Gerf, V., Clark, D., Kahn, R., Kleinrock, L., Lynch, D., ... Stephen Wolff. (1997). *Brief history of internet*. Internet Society. <https://doi.org/10.3233/978-1-60750-880-9-121>
- Lemieux, V. (2016). Trusting records: is Blockchain technology the answer? *Records Management Journal*, 26(2), 110–139. <https://doi.org/10.1108/RMJ-12-2015-0042>
- Li, C., & Sugimoto, S. (2017). Provenance Description of Metadata Vocabularies for the Long-term Maintenance of Metadata. *Journal of Data and Information Science*, 2(2), 41–55.

- <https://doi.org/10.1515/jdis-2017-0007>
- Library and Archives Canada. (2012). Discover Finding Aids! Disponível em: <https://thediscoverblog.com/2012/02/28/discover-finding-aids/>
- Library of Congress. (2009). Design Principles for Enhancements to MODS and MADS.
- Library of Congress. (2016). MODS: Uses and Features. Disponível em: <https://www.loc.gov/standards/mods/mods-overview.html>
- Library of Congress. (2017). *Card Catalog: Books, Cards, and Literary Treasures*. Chronicle Books.
- Library of Congress. (2019a). MARC21 format for bibliographic data. Appendix B: full level record examples. Disponível em: <https://www.loc.gov/marc/bibliographic/examples.html>
- Library of Congress. (2019b). MARC21 format for bibliographic data.
- Library of Congress. (2020a). MODS User Guidelines (ver. 3). Disponível em: <https://www.loc.gov/standards/mods/userguide/introduction.html>
- Library of Congress. (2020b). Outline of Elements and Attributes in MODS Version 3.7. Disponível em: <https://www.loc.gov/standards/mods/mods-outline-3-7.html>
- Library of Congress. (2020c). Union catalogs.
- Library of Congress. (2021). MODS Implementation Registry. Disponível em: <https://www.loc.gov/standards/mods/registry.php>
- Licklider, J. (1960). Man-Computer Symbiosis. In *IRE Transactions on Human Factors in Electronics* (Vol. HFE-1, pp. 4–11).
- Licklider, J. (1963). Memorandum For Members and Affiliates of the Intergalactic Computer Network. Disponível em: <https://www.kurzweilai.net/memorandum-for-members-and-affiliates-of-the-intergalactic-computer-network>
- Licklider, J., & Clark, W. (1962). On-line man-computer communication. In *AIEE-IRE '62 (Spring): Proceedings of the May 1-3, 1962, spring joint computer conference* (pp. 113–128). New York: Association for Computing Machinery.
- Licklider, J., & Taylor, R. (1968). The Computer as a Communication Device. *Science and Technology*, 21–41. <https://doi.org/10.1145/1036362.1036363>
- Lincoln, Y. (1990). The Making of a Constructivist: A Remembrance of Transformations Past. In E. Guba (Ed.), *The paradigm dialog* (pp. 67–87). Thousand Oaks, CA: SAGE Publications.
- Lincoln, Y., Lynham, S., & Guba, E. (2017). Paradigmatic controversies, contradictions, and emerging confluences revisited. In N. Denzin & Y. Lincoln (Eds.), *The Sage Handbook of Qualitative Research* (5°, pp. 97–128). Thousand Oaks, CA: SAGE Publications.
- Linden, A., & Fenn, J. (2003). Understanding Gartner's hype cycles. Gartner. <https://doi.org/10.1023/B:JMISC.0000013906.60034.e8>
- Linden, L. (2017). *O tratamento temático da informação em instrumentos normativos de descrição arquivística*. Universidade Federal de Santa Catarina. <https://doi.org/10.19132/1808-5245243.96-124>
- Lipinski, M., Yao, K., Breitingner, C., Beel, J., & Gipp, B. (2013). Evaluation of header metadata extraction approaches and tools for scientific PDF documents. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, 385–386. <https://doi.org/10.1145/2467696.2467753>
- Llanes-Padrón, D. (2011). *La representación normalizada de los documentos: estudio comparado de las normas de descripción archivística*. University of Salamanca.
- Llanes-Padrón, D. (2017). La representación de documentos y autoridades archivísticas: una mirada desde la perspectiva de la normalización. *Perspectivas Em Ciência Da Informação*, 22(2), 211–231. <https://doi.org/10.1590/1981-5344/2643>
- Llanes-Padrón, D. (2019). La descripción archivística y la curaduría digital: relaciones y perspectivas. In *Atas do IV Congresso ISKO Espanha-Portugal*. Barcelona. <https://doi.org/10.31229/osf.io/265vr>
- Llanes-Padrón, D., & Moro-Cabero, M. (2017). Records in contexts: un nuevo modelo para la representación de la información archivística en el entorno de la web semántica. *El Profesional de La Información*, 26(3), 525–533. <https://doi.org/10.3145/epi.2017.may.19>
- Llanes-Padrón, D., & Pastor-Sánchez, J. (2017). Records in context: el camino de los archivos hacia la interoperabilidad semántica. *Anuario ThinkEPI*, 11, 297–304. <https://doi.org/10.3145/thinkepi.2017.56>

- Lodolini, E. (1998). *Archivistica. Principi e Problemi* (8th ed.). Milano: Franco Angeli.
- Lopatin, L. (2010). Metadata practices in academic and non-academic libraries for digital projects: a survey. *Cataloging and Classification Quarterly*, 48(8), 716–742. <https://doi.org/10.1080/01639374.2010.509029>
- Losee, R. (2014). Combining high metainformation with high information content: The Information-Metainformation Utility hypothesis. *Knowledge Organization*, 41(2), 123–130. <https://doi.org/10.5771/0943-7444-2014-2-123>
- Lubas, R., Jackson, A., & Schneider, I. (2013). *The metadata manual: a practical handbook*. Oxford: Chandos Publishing.
- Lubetzky, S. (2001a). Cataloging rules and principles. In E. Svenonius & D. McGarry (Eds.), *Seymour Lubetzky: Writings on the Classical Art of Cataloging* (pp. 75–139). Englewood, CO: Libraries Unlimited.
- Lubetzky, S. (2001b). *Seymour Lubetzky: writings on the classical art of cataloging / compiled and edited by Elaine Svenonius and Dorothy McGarry*. Englewood, CO: Libraries Unlimited.
- Lubetzky, S. (2001c). The Fundamentals of Bibliographic Cataloging and AACR2. In E. Svenonius & D. McGarry (Eds.), *Seymour Lubetzky: Writings on the Classical Art of Cataloging* (pp. 367–378). Englewood, CO: Libraries Unlimited.
- Lubetzky, S. (2001d). The Vicissitudes of Ideology and Technology in Anglo-American Cataloging since Panizzi and a Prospective Reformation of the Catalog for the Next Century. In E. Svenonius & D. McGarry (Eds.), *Seymour Lubetzky: Writings on the Classical Art of Cataloging* (pp. 419–429). Englewood, CO: Libraries Unlimited.
- Lynch, C. (1998). Identifiers and Their Role in Networked and Information Applications. *Bulletin of the American Society for Information Science and Technology*, 24, 17–20. <https://doi.org/10.1002/bult.80>
- Lynch, C. (2004). *CNI White Paper on Networked Information Discovery and Retrieval*. Coalition for Networked Information.
- Macdonald, K., & Tipton, C. (1996). Using documents. In N. Gilbert (Ed.), *Researching Social Life* (pp. 187–200). London: Sage.
- Machado, L. M. O., Souza, R. R., & Graça Simões, M. da. (2019). Semantic Web or Web of Data? A diachronic study (1999 to 2017) of the publications of Tim Berners-Lee and the World Wide Web Consortium. *Journal of the Association for Information Science and Technology*, (June). <https://doi.org/10.1002/asi.24111>
- Macleod, R. (Ed.). (2004). *The Library of Alexandria: Centre of Learning in the Ancient World*. New York, UK: Bloomsbury Publishing PLC.
- Macneil, H. (2000). Providing grounds for trust: Developing conceptual requirements for the long-term preservation of authentic electronic records. *Archivaria*, 50(1), 52–78.
- MacNeil, H. (1998). *Trusting records: the evolution of legal, historical, and diplomatic methods of assessing the trustworthiness of records, from antiquity to the digital age*. The University of British Columbia.
- MacNeil, H. (2000). *Trusting Records: Legal, Historical, and Diplomatic Perspectives*. Dordrecht: Kluwer Academic Publishers.
- MacNeil, H. (2001). Trusting records in a postmodern world. *Archivaria*, 1(51), 36–47.
- MacNeil, H. (2005). Picking our text: archival description, authenticity, and the archivist as editor. *American Archivist*, 68(Fall/Winter), 264–278.
- MacNeil, H. (2008). Archivalterity: Rethinking Original Order. *Archivaria*, 66(Fall), 1–24.
- MacNeil, H. (2009). Trusting description: authenticity, accountability, and archival description standards. *Journal of Archival Organization*, 7(3), 89–107. <https://doi.org/10.1080/15332740903117693>
- MacNeil, H. (2011). Trust and professional identity: Narratives, counter-narratives and lingering ambiguities. *Archival Science*, 11(3–4), 175–192. <https://doi.org/10.1007/s10502-011-9150-5>
- MacNeil, H. (2012). What finding aids do: Archival description as rhetorical genre in traditional and web-based environments. *Archival Science*, 12(4), 485–500. <https://doi.org/10.1007/s10502-012-9175-4>
- MacNeil, H. (2017). Deciphering and Interpreting an Archival Fonds and Its parts. In A. Gilliland,

- S. McKemmish, & A. Lau (Eds.), *Research in the Archival Multiverse* (pp. 161–197). Clayton: Monash University Publishing.
- Mak, B. (2012). On the uses of authenticity. *Archivaria*, 73(Spring 2012), 1–17.
- Mak, B. (2015). Authenticity. In L. Duranti & P. Franks (Eds.), *Encyclopedia of Archival Science* (pp. 119–122). Lanham: Rowman & Littlefield.
- Mann, P. (1980). The publishing of scholarly monographs. *Journal of Documentation*, 36(1), 1–10. <https://doi.org/10.1108/eb026687>
- Massieu, J.-B. (2018). *Instruction pour procéder à la confection du catalogue de chacune des bibliothèques*. Paris: Hachette Livre - BNF.
- McCall, G. (1984). Systematic field observation. *Annual Review of Sociology*, 10, 263–282.
- McCarty, W. (2005). *Humanities Computing*. Humanities Computing. New York: Palgrave Macmillan. <https://doi.org/10.1057/9780230504219>
- McCrank, L. (2002). *Historical Information Science*. Medford, NJ: Information Today, Inc.
- Mckemmish, S. (2002). Placing Records Continuum Theory and Practice. *Archival Science*, 1, 333–359.
- Mckemmish, S., Glenda, A., Nigel, W., & Barbara, R. (1999). Describing Records in Context in the Continuum: the Australian Recordkeeping Metadata Schema. *Archivaria*, (48), 3–43. Disponível em: <http://journals.sfu.ca/archivar/index.php/archivaria/article/view/12715/13890>
- McLeod, J. (2014). Reinventing archival methods: Reconceptualising electronic records management as a wicked problem. *Archives & Manuscripts*, 42(2), 193–196. <https://doi.org/10.1080/01576895.2014.911687>
- McLeod, J., & Gormly, B. (2017). Using the cloud for records storage: issues of trust. *Archival Science*, 17(4), 349–370. <https://doi.org/10.1007/s10502-017-9280-5>
- McQueen, K. (2018). Ethical Issues of Knowledge Organization in Designing a Metadata Schema for the Leo Kottke Archives. *Knowledge Organization*, 42(5), 332–338. <https://doi.org/10.5771/0943-7444-2015-5-332>
- Meehan, J. (2009). Making the leap from parts to whole: Evidence and inference in archival arrangement and description. *American Archivist*, 72(1), 72–90. <https://doi.org/10.17723/aarc.72.1.kj672v4907m11x66>
- Meehan, J. (2010). Rethinking original order and personal records. *Archivaria*, 70, 27–44.
- Meho, L. I. (2019). Using Scopus's CiteScore for assessing the quality of computer science conferences. *Journal of Informetrics*, 13(1), 419–433. <https://doi.org/10.1016/j.joi.2019.02.006>
- Merriam, S. (1988). *Case study research in education: A qualitative approach*. San Francisco: Jossey-Bass.
- Mertens, D. (1998). *Research Methods in Education and Psychology*. Thousand Oaks, CA: SAGE Publications.
- Mertens, D. (2009). *Transformative research and evaluation*. New York: Guilford Press.
- Mertens, D. (2019). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative and mixed methods (5<sup>o</sup>)*. Thousand Oaks, CA: SAGE Publications.
- Mertens, D., & Tarsila, M. (2015). Mixed methods evaluation. In *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry* (pp. 426–446). Oxford: Oxford Press.
- Mertens, D., & Wilson, A. (2013). *Program evaluation theory and practice: A comprehensive guide*. New York: Guilford Press.
- Meshkin, C. C. (2010). Unchecked data: A tool for political corruption? *Engage*, 11(3), 44–51.
- Middleton, M. (1999). Metainformation incorporation in library digitisation projects. In T. Aparac, B, C, D, E, F, & G (Eds.), *Digital Libraries: interdisciplinary concepts, challenges and opportunities; proceedings of the Third International Conference on the Conceptions of Library and Information Science, Dubrovnik, Croatia, May 23-26, 1999* (pp. 334–338).
- Miksa, F. (Ed.). (1977). *Charles Amni Cutter: Library systematizer*. Littleton CO: Libraries Unlimited.
- Miller, E. (1988). *Prince of Librarians: The Life and Times of Antonio Panizzi of the British Museum*. The British Library Publishing.
- Miller, S. (2011). *Metadata for digital collections: a how-to-do-it manual*. New York: Neal-

- Schuman Publishers.
- Moen, W. (2001). Mapping the interoperability landscape for networked information retrieval. In *Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 50–51). New York: ACM Press. <https://doi.org/10.1145/379437.379447>
- Morin, E. (1983). *O paradigma perdido*. Lisboa: Publicações Europa-América.
- Moro-Cabero, M., Martín-Pozuelo, M., & Zazo, J. (2011). ISO 15489 and other standardized management systems: Analogies and synergies. *Records Management Journal*, 21(2), 104–121. <https://doi.org/10.1108/09565691111152044>
- Morrison, M. (2008). Models as representational structures. In L. Bovens, C. Hofer, & S. Hartmann (Eds.), *Nancy Cartwright's Philosophy of Science* (pp. 67–90). New York: Routledge.
- Morse, J. (2003). Principles of Mixed Methods and Multimethod Research Design. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of Mixed Methods In Social and Behavioral Research* (pp. 189–208). Thousand Oaks, CA: SAGE Publications.
- Mortensen, P. (1999). The place of theory in archival practice. *Archivaria*, 47((Spring 1999)), 1–26.
- Moss, M., & Thomas, D. (2017). Overlapping temporalities: the judge, the historian and the citizen. *Archives*, 52(134), 51–66. <https://doi.org/10.3828/archives.2017.4>
- Moyano Collado, J. (2013). La descripción archivística. De los instrumentos de descripción hacia la web semántica. *Anales de Documentacion*, 16(2). <https://doi.org/10.6018/analesdoc.16.2.171841>
- Mueller, S., Campello, B., & Dias, E. (1996). Disseminação da pesquisa em ciência da informação e biblioteconomia no Brasil. *Ciência Da Informação*, 25(3), 1–22.
- Müller, S., Feith, J., & Fruin, R. (2003). *Manual for the arrangement and description of archives / drawn up by direction of the Netherlands Association of Archivists by S. Muller, J.A. Feith and R. Fruin ; translation of the second edition by Arthur H. Leavitt ; with new introductions by Peter Ho*. Chicago: Society of American Archivists.
- Myler, J. (1658). *De Principibus & Statibus Imperii Rom. Germ. eorumvè praecipuis Juribus Succincta delineatio*. Tubingae: Cotta.
- NARA. (2015). Bulletin 2015-04: Metadata Guidance for the Transfer of Permanent Electronic Records. Disponível em: <https://www.archives.gov/records-mgmt/bulletins/2015/2015-04.html>
- Nart, D. de, Degl'Innocenti, D., Peressotti, M., & Tasso, C. (2017). Stratifying Semantic Data for Citation and Trust: An Introduction to RDFDF. In M. Agosti, M. Bertini, S. Ferilli, S. Marinai, & N. Orío (Eds.), *Digital Libraries and Multimedia Archives. IRCDL 2016. Communications in Computer and Information Science, vol 701* (pp. 104–111). Springer. [https://doi.org/10.1007/978-3-319-56300-8\\_10](https://doi.org/10.1007/978-3-319-56300-8_10)
- National Archives of Australia. (2015). Australian Government Recordkeeping Metadata Standard (AGRkMS). Disponível em: [http://www.americanbanker.com/issues/179\\_124/which-city-is-the-next-big-fintech-hub-new-york-stakes-its-claim-1068345-1.html%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/15003161%5Cnhttp://cid.oxfordjournals.org/lookup/doi/10.1093/cid/cir991%5Cnhttp://www.scielo](http://www.americanbanker.com/issues/179_124/which-city-is-the-next-big-fintech-hub-new-york-stakes-its-claim-1068345-1.html%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/15003161%5Cnhttp://cid.oxfordjournals.org/lookup/doi/10.1093/cid/cir991%5Cnhttp://www.scielo)
- Navarro, M. (1998). La normalización de la descripción y la recuperación de información en los archivos: vino viejo en odres nuevos. *Boletín de La ANABAD*, 68(1), 9–18.
- NBC News. (2011). Meet the Press transcript for Feb. 13, 2011. Disponível em: [http://www.nbcnews.com/id/41536793/ns/meet\\_the\\_press-transcripts/t/meet-press-transcript-feb/#.XtjbNJ70nH4](http://www.nbcnews.com/id/41536793/ns/meet_the_press-transcripts/t/meet-press-transcript-feb/#.XtjbNJ70nH4)
- Needham, J. (1994). *The Shorter Science and Civilisation in China, Volume 4*. Cambridge University Press.
- Nesmith, T. (1999). Still Fuzzy, But More Accurate: Some Thoughts on the “Ghosts” of Archival Theory. *Archivaria*, 47((Spring)), 136–150.
- Neuman, W. (2014). *Social research methods: Qualitative and quantitative approaches* (7<sup>o</sup>). Boston: Allyn and Bacon.
- Newman, I., & Ridenour, C. (1998). *Qualitative-quantitative research methodology: Exploring the interactive continuum*. Carbondale and Edwardsville: Southern Illinois University Press.

- NISO. (2004). *Understanding Metadata*. Bethesda: NISO Press.
- NISO. (2017). *Understanding metadata: what is metadata, and what is it for?* Bethesda: NISO Press. Disponível em: [http://www.niso.org/apps/group\\_public/download.php/17446/Understanding Metadata.pdf](http://www.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf)
- Niu, J. (2013). Recordkeeping metadata and archival description: A revisit. *Archives & Manuscripts*, 41(3), 203–215. <https://doi.org/10.1080/01576895.2013.829751>
- Niu, J. (2015a). Archival Intellectual Control in the Digital Age. *Journal of Archival Organization*, 12(3–4), 186–197. <https://doi.org/10.1080/15332748.2015.1154747>
- Niu, J. (2015b). Original order in the digital world. *Archives & Manuscripts*, 43(1), 61–72. <https://doi.org/10.1080/01576895.2014.958863>
- Niu, J. (2016). Linked data for archives. *Archivaria*, 82(Fall 2016), 83–110.
- Niu, X., Zhang, T., & Chen, H. (2014). Study of User Search Activities With Two Discovery Tools at an Academic Library. *International Journal of Human-Computer Interaction*, 30(5), 422–433. <https://doi.org/10.1080/10447318.2013.873281>
- Norris, D. (1939). *A history of cataloguing and cataloguing methods, 1100-1850: with an introductory survey of ancient times*. London: Grafton & Co.
- O’Toole, J. (1994). On the Idea of Uniqueness. *American Archivist*, 57(4), 632–658. <https://doi.org/10.1128/EC.00300-10>
- Obama, B. (2013). “Nobody is listening to your telephone calls,” Obama says. *CBS News*. Disponível em: <https://www.youtube.com/watch?v=hUD1ujK2BJc>
- Oldroyd, D. (1986). *The Arch of Knowledge: An Introductory Study of the History of the Philosophy and Methodology of Science*. New York: Methuen.
- Oliver, M. (1992). Changing the Social Relations of Research Production? *Disability, Handicap & Society*, 7(2), 101–114. <https://doi.org/10.1080/02674649266780141>
- Online Archive of California. (2020). *About OAC*. Disponível em: <http://www.oac.cdlib.org/about/>
- Oparina, O. D. (2012). The interdisciplinary aspects of the interaction of bibliography with social sciences and humanities. *Scientific and Technical Information Processing*, 39(1), 42–46. <https://doi.org/10.3103/S0147688212010133>
- Osborn, A. (1985). The Crisis in Cataloging. In M. Carpenter & E. Svenonius (Eds.), *Foundations of Cataloging: A Sourcebook* (pp. 92–103). Littleton, Colorado: Libraries Unlimited.
- Owen, K. (2015). Reliability. In L. Duranti & P. Franks (Eds.), *Encyclopedia of Archival Science* (pp. 361–363). Lanham: Rowman & Littlefield.
- Pacheco, A. (2017). *Informação digital: o vértice comum entre a Diplomática e a Ciência da Informação*. Braga: Edições Húmus.
- Pacheco, J. A. (1993). *O pensamento e a ação do professor em formação*. Universidade do Minho.
- Panizzi, A., & British Museum. (1841). *Rules for the compilation of the catalogue*. London: Printed by order of the Trustees.
- Park, S. (2017). Information is power. In *Digital Capital*. London: Palgrave Macmillan.
- Pattie, L. (1998). Henriette davidson avram, the great legacy. *Cataloging and Classification Quarterly*, 25(2–3), 67–81. [https://doi.org/10.1300/J104v25n02\\_05](https://doi.org/10.1300/J104v25n02_05)
- Pfeiffer, R. (Ed.). (1965). *Callimachus*. Oxonii: Typographeo Clarendoniano.
- Philip, I. (1983). *The Bodleian Library in the Seventeenth and Eighteenth Centuries*. Lyell Lectures.
- Pinheiro, O. (2003). *Sistema de Apoio à Decisão no Planeamento da Produção de Produtos Complexos: Identificação e Especificação de Requisitos*. University of Porto.
- Pinto, M. M. (2009). *PRESERVMAP: Um roteiro da preservação na era digital (1ª)*. Porto: Edições Afrontamento.
- Pohl, K. (2010). *Requirements engineering: fundamentals, principles, and techniques*. Heidelberg: Springer.
- Pomerantz, J. (2015). *Metadata*. Cambridge, MA: MIT Press.
- Presser, S., & Traugott, M. (1992). Little White Lies and Social Science Models: Correlated Response Errors in a Panel Study of Voting Author. *Public Opinion Quarterly*, 56(1), 77–86.
- Price, D. M., & Smith, J. J. (2011). The trust continuum in the information age: A Canadian perspective. *Archival Science*, 11(3–4), 253–276. <https://doi.org/10.1007/s10502-011-9148-z>
- Procter, M., & Cook, M. (2000). *Manual of archival description* (3rd ed.). London: Gower Publishing.



- Prom, C., & Frusciano, T. (Eds.). (2013). *Archival arrangement and description*. Society of American Archivists.
- Punch, K. (1998). *Introduction to Social Research: quantitative & qualitative approaches*. London: SAGE Publications.
- Quine, W. (1969). *Ontological Relativity and Other Essays*. New York: Columbia University press.
- Quivy, R., & Campenhoudt, L. (2013). *Manual de investigação em ciências sociais* (6<sup>a</sup>). Lisboa: Gradiva.
- Radnitzky, G. (1970). *Contemporary Schools of Metascience, Vols I-II*. Goteborg: Akademiforlaget.
- Ramesh, P., Vivekavardhan, J., & Bharathi, K. (2015). Metadata diversity, interoperability and resource discovery issues and challenges. *DESIDOC Journal of Library and Information Technology*, 35(3), 193–199. <https://doi.org/10.14429/djlit.35.3.8074>
- Ramsay, S. (2004). Databases. In *A Companion do Digital Humanities*. Blackwell.
- Ranganathan, S. R. (1951). *Philosophy of library classification*. New York: Hafner.
- Ranganathan, S. R. (1955). *Heading and Canons: Comparative Study of Five Catalogue Codes*. Madras: S. Viswanathan.
- Ranganathan, S. R. (1957). *Prolegomena to library classification*. London: Library Association.
- Rapley, T. (2007). *Qualitative research kit: Doing Conversation, Discourse and Document Analysis*. London: SAGE Publications. <https://doi.org/10.4135/9781849208901.n7>
- Rather, L., & Wiggins, B. (1989). “Mother Avram’s Remarkable Contribution”: Henriette D. Avram. *American Library Assciation*, 20(9), 855–857, 859–561.
- Reason, P. (1994). Three approaches to participative inquiry. In N. Denzin & Y. Lincoln (Eds.), *Handbook of Qualitative Research* (pp. 324–339). Thousand Oaks, CA: SAGE Publications.
- Reed, B. (1997). Electronic records management in Australia. *Records Management Journal*, 7(3), 191–204. <https://doi.org/10.1108/eb027111>
- Research Libraries Group. (2005). *Descriptive Metadata Guidelines for RLG Cutlrual Materials*. Mountain View, CA: RLG. Disponível em: [www.rlg.org](http://www.rlg.org)
- Riberette, P. (1970). *Les bibliothèques françaises pendant la révolution 1789-1795*. Paris: Bibliothèque nationale.
- Ridener, J. (2009). *From Polders to Postmodernism: A Concise History of Archival Theory*. Minnesota: Litwin Books.
- Riley, J. (2010). Seeing Standards: A Visualization of the Metadata Universe. <https://doi.org/10.1080/07317131.2011.598072>
- Roberts, J. (1987). Archival theory: Much ado about shelving. *American Archivist*, 50(1), 66–74. <https://doi.org/10.17723/aarc.50.1.1357257455776g52>
- Roberts, J. (1990). Archival Theory: Myth or Banality? *The American Archivist*, 53(1), 110–120.
- Roberts, L. (1967). Multiple computer networks and intercomputer communication. In *SOSP '67: Proceedings of the first ACM symposium on Operating System Principles* (pp. 3.1-3.6). New York: Association for Computing Machinery.
- Robertson, S., & Robertson, J. (2012). *Mastering the Requirements Process: Getting Requirements Right* (3<sup>o</sup>). New York: Addison-Wesley.
- Rogers, C. (2015). *Virtual Authenticity: Authenticity of Digital Records from Theory to Practice*. University of British Columbia. <https://doi.org/10.14288/1.0166169>
- Rogers, C., & Tennis, J. (2013). Authenticity as a Social Contract—We Are Our Records. In *iConference 2013 Proceedings* (pp. 801–803). <https://doi.org/10.9776/13393>
- Rolan, G. (2015). Towards archive 2.0: Issues in archival systems interoperability. *Archives & Manuscripts*, 43(1), 42–60. <https://doi.org/10.1080/01576895.2014.959535>
- Rolan, G. (2017). Towards interoperable recordkeeping systems: A meta-model for recordkeeping metadata. *Records Management Journal*, 27(2), 125–148. <https://doi.org/RMJ-09-2016-0027>
- Rolan, G., Humphries, G., Jeffrey, L., Samaras, E., Antsoupova, T., & Stuart, K. (2019). More human than human? Artificial intelligence in the archive. *Archives & Manuscripts*, 47(2), 179–203. <https://doi.org/10.1080/01576895.2018.1502088>
- Rosenberg, J. (1981). *Linguistic representation*. Boston, MA: D. Reidel Publishing Company.
- Ross, J. (2014). Art’s role in the origins of writing: the seal-carver, the Scribe, and the earliest lexical texts. In B. A. Brown & Marian H. Feldman (Eds.), *Critical Approaches to Ancient*

- Near Eastern Art* (pp. 295–318). Boston: De Gruyter.
- Rowat, T. (1993). The Record and Repository as a Cultural Form of Expression. *Archivaria*, 36(Autumn), 198–204.
- Samouelian, M. (2009). Embracing web 2.0: Archives and the newest generation of web applications. *American Archivist*, 72(1), 42–71. <https://doi.org/10.17723/aarc.72.1.k73112x7n0773111>
- Sanger, D., & Smale, A. (2013, December 16). U.S.-Germany Intelligence Partnership Falters Over Spying. *The New York Times*. Disponível em: <https://www.nytimes.com/2013/12/17/world/europe/us-germany-intelligence-partnership-falters-over-spying.html>
- Santos, Â. (2015). *Repositório Digital Seguro: um modelo de requisitos para um provedor de serviços de certificação*. Universidade do Porto. Disponível em: <http://hdl.handle.net/10216/79978>
- Schaeffer, B., Baranski, B., Foerster, T., & Brauner, J. (2012). A service-oriented framework for real-time and distributed geoprocessing. *Lecture Notes in Geoinformation and Cartography*, (199579), 3–20. [https://doi.org/10.1007/978-3-642-10595-1\\_1](https://doi.org/10.1007/978-3-642-10595-1_1)
- Schellenberg, T. (1988). *The Management of Archives*. Washington, D.C.: National Archives and Records Administration.
- Schellenberg, T. (2003). *Modern archives: principles & techniques*. Chicago: The Society of American Archivists.
- Schottlaender, B. (2003). Why metadata? Why me? Why now? *Cataloging and Classification Quarterly*, 36(3–4), 19–29. [https://doi.org/10.1300/J104v36n03\\_03](https://doi.org/10.1300/J104v36n03_03)
- Schwandt, T. (1994). Constructivist, interpretivist approaches to humanist inquirt. In N. Denzin & Y. Lincoln (Eds.), *Handbook of Qualitative Research* (pp. 221–259). Thousand Oaks, CA: SAGE Publications.
- Schwartz, J. (1995). “We Make Our Tools and Our Tools Make Us”: Lessons from Photographs for the Practice, Politics, and Poetics of Diplomats. *Archivaria*, 40(Fall), 40–74.
- Searle, J. (1986). Meaning, communication, and representation. In R. Grandy & R. Warner (Eds.), *Philosophical grounds of rationality* (pp. 209–226). Clarendon Press.
- Searle, J. (2002). *Consciousness and language*. Cambridge University Press.
- Serewicz, L. (2010). Do we need bigger buckets or better search engines?: the challenge of unlimited storage and semantic web search for records management. *Records Management Journal*, 20(2), 172–181. <https://doi.org/10.1108/09565691011064313>
- Shannon, C., & Weaver, W. (1963). *The mathematical theory of communication*. The University of Illinois Press. The University of Illinois Press. <https://doi.org/10.1145/584091.584093>
- Sharp, H. (1937). *Cataloguing: a textbook for use in libraries*. London: Grafton.
- Shaw, I. (1999). *Qualitative evaluation*. London: SAGE Publications.
- Sherry, J., & Brown, C. (2004). History of the internet. In *The Internet Encyclopedia, Volume 2 (G - O)* (pp. 114–123).
- Silva, A. M. da. (2009). Mediações e mediadores em Ciência da Informação. *Prisma.Com*, 9, 68–104. Disponível em: <http://revistas.ua.pt/index.php/prisma.com/article/view/700>
- Silva, A. M. da, Ribeiro, F., Ramos, J., & Real, M. L. (2009). *Arquivística: teoria e prática de uma ciência da informação* (3ª). Porto: Edições Afrontamento.
- Silva, C. G. da. (2019). *Investigação em Ciência da Informação*. Lisboa: Edições Colibri.
- Silva, C. G. da. (2021). Investigação Documental. In S. Gonçalves; J. Gonçalves & C. Marques (Eds.), *Manual de Investigação Qualitativa: Conceção, Análise e Aplicações*. Lisboa: Pactor.
- Silva, I. (2012). *A organização e a representação do conhecimento no domínio da arquivística*. Universidade Estadual Paulista.
- Simões, M. da G. (2010). *A representação de etnia e a sua evolução na CDU*. Universidad de Salamanca.
- Sinn, D. (2015). Collaborative Education between Classroom and Workplace for Archival Arrangement and Description: Aiming for Sustainable Professional Education. *The American Archivist*, 76(1), 237–262. <https://doi.org/10.17723/aarc.76.1.u86275j711w471m6>
- Slife, B., & Williams, R. (1995). *What’s behind the research? Discovering hidden assumptions in the behavioral sciences*. Thousand Oaks, CA, US: SAGE Publications.

- Smiraglia, R. (2005). Introducing metadata. In Richard Smiraglia (Ed.), *Metadata: a cataloger's primer* (pp. 1–15). Binghamton, NY: Haworth Information Press.
- Society of American Archivists. (2005). *Glossary of Archival and Records Terminology*. Disponível em: <http://www2.archivists.org/glossary/terms/f/fonds>
- Society of American Archivists. (2012). Society of American Archivists core values statement and code of ethics. Disponível em: <https://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics>
- Society of American Archivists. (2019a). Encoded Archival Description Tag Library Version EAD3 1.1.1. <https://doi.org/10.1353/mov.2003.0022>
- Society of American Archivists. (2019b). Fonds. Disponível em: <https://www2.archivists.org/glossary/terms/f/fonds>
- Society of American Archivists. (2020). *Describing Archives: A Content Standard - DACS 2019.0.3*. Disponível em: [https://files.archivists.org/pubs/DACS\\_2019.0.3\\_Version.pdf](https://files.archivists.org/pubs/DACS_2019.0.3_Version.pdf)
- Soeng, M. (2011). *The Diamond Sutra: Transforming the Way We Perceive the World*. Wisdom Publications.
- Sousa, A. B. (2005). *Investigação em educação*. Lisboa: Livros Horizonte.
- Stake, R. (1995). *The art of case study research*. Thousand Oaks, CA: SAGE Publications.
- Stamou, S., & Efthimiadis, E. (2010). Interpreting User Inactivity on Search Results. In *Advances in Information Retrieval. ECIR 2010. Lecture Notes in Computer Science, vol 5993*. Berlin: Springer. [https://doi.org/10.1007/978-3-642-12275-0\\_12](https://doi.org/10.1007/978-3-642-12275-0_12)
- Stewart, D. (1984). *Secondary research: information sources and methods*. Newbury Park, California: Sage.
- Strout, R. (1956). The Development of the Catalog and Cataloging Codes. *The Library Quarterly*, 26(4), 254–275.
- Suárez, M. (2003). Scientific representation: against similarity and isomorphism. *International Studies in the Philosophy of Science*, 17(3), 225–244. <https://doi.org/10.1080/0269859032000169442>
- Sudmanns, M., Tiede, D., Lang, S., & Baraldi, A. (2018). Semantic and syntactic interoperability in online processing of big Earth observation data. *International Journal of Digital Earth*, 11(1), 95–112. <https://doi.org/10.1080/17538947.2017.1332112>
- Suvak, W. (2015). Authentication. In L. Duranti & P. Franks (Eds.), *Encyclopedia of Archival Science* (pp. 116–119). Lanham: Rowman & Littlefield.
- Swoyer, C. (1991). Structural representation and surrogate reasoning. *Synthese*, 87(3), 449–508. <https://doi.org/10.1007/BF00499820>
- Talja, S., Tuominen, K., & Savolainen, R. (2005). “Isms” in information science: Constructivism, collectivism and constructionism. *Journal of Documentation*, 61(1), 79–101. <https://doi.org/10.1108/00220410510578023>
- Tanenbaum, A. (2003). *Computer Networks* (4<sup>o</sup>). Upper Saddle River, NJ: Prentice Hall PTR. Disponível em: <http://repositorio.unan.edu.ni/2986/1/5624.pdf>
- Tarraubella, J. (1998). Los archivos y sus usuarios. *Revista Lligall*, 12, 188–204.
- Tashakkori, A., & Creswell, J. (2007). The new era of mixed methods. *Journal of Mixed Methods Research*, 1(1), 3–7.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches. Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA, US: Sage Publications, Inc.
- Tashakkori, A., & Teddlie, C. (Eds.). (2003). *Handbook of mixed methods in social & behavioral research*. Thousand Oaks: SAGE Publications.
- Teddlie, C., & Tashakkori, A. (2009). *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Thousand Oaks, CA: SAGE Publications.
- Teddlie, C., & Tashakkori, A. (2015). Overview of Contemporary Issues in Mixed Methods Research. In A. Tashakkori & C. Teddlie (Eds.), *SAGE Handbook of Mixed Methods in Social & Behavioral Research* (pp. 1–42). Thousand Oaks, CA, US: Sage Publications, Inc. <https://doi.org/10.4135/9781506335193.n1>
- TEI. (2020). Text Encoding Initiative: history. Disponível em: <https://tei-c.org/about/history/>

- Tennant, R. (2001). Different Paths to Interoperability. *Library Journal*, 126(3), 118–119.
- Texier, C. (1839). *Description de l'Asie Mineure: faite par ordre du gouvernement français en 1833-1837... vol I*. Paris: Didot Frères. Disponível em: <https://digi.ub.uni-heidelberg.de/diglit/texier1839bd1/0054>
- The Consultative Committee for Space Data Systems. (2013). *Reference Architecture for Space Information Management Architecture for Space Information*.
- The National Archives of Finland. (2013). Finnish Conceptual Model for Archival Description.
- The National Archives of the UK. (2002). Functional Requirements for Electronic Records Management Systems - Reference document.
- Thibodeau, K. (2002). Overview of technological approaches to digital preservation and challenges in coming years. In *The state of digital preservation: an international perspective* (pp. 4–31). Washington, D.C.: Council on Library and Information Resources. <https://doi.org/10.1353/pla.2003.0087>
- TiECON East. (2014). Data is new oil. Disponível em: <http://www.tieconeast.org/2014/big-data-analytics>
- Timmermans, S., & Epstein, S. (2010). A World of Standards but not a Standard World: Toward a Sociology of Standards and Standardization. *Annual Review of Sociology*, 36(1), 69–89. <https://doi.org/10.1146/annurev.soc.012809.102629>
- Toffler, A. (1990). *Powershift: Knowledge, Wealth and Violence at the Edge of the 21st Century*. New York: Bantam Books.
- Tough, A. (2016). Thinking about and working with archives and records: a personal reflection on theory and practice. *Archives and Records*, 37(2), 225–238. <https://doi.org/10.1080/23257962.2016.1147343>
- Trace, C. (2015). Archival arrangement. In L. Duranti & P. Franks (Eds.), *Encyclopedia of Archival Science* (pp. 21–24). Lanham: Rowman & Littlefield.
- Treece, E., & Treece, J. (1982). *Elements of research in nursing*. St. Louis: The C.V. Mosby Company.
- TRIPLE. (2020). *D2.1 - Data Acquisition Plan*.
- Tuckman, B. (2012). *Manual de investigação em educação: metodologia para conceber e realizar o processo de investigação científica* (4th revise). Lisboa: Fundação Calouste Gulbenkian.
- UK Parliament. (2016). Investigatory Powers Act.
- Upward, F. (2000). Modelling the continuum as paradigm shift in recordkeeping and archiving processes, and beyond — a personal reflection. *Records Management Journal*, 10(3), 115–139. <https://doi.org/10.1108/EUM0000000007259>
- Usher, R. (1996). A critique of the neglected epistemological assumptions of educational research. In D. Scott & R. Usher (Eds.), *Understanding Educational Research*. London: Routledge.
- van Fraassen, B. (2008). *Scientific representation: paradoxes of perspective*. Oxford: Oxford University Press.
- Verborgh, R., Van Hooland, S., Cope, A. S., Chan, S., Mannens, E., & Van De Walle, R. (2015). The fallacy of the multi-API culture conceptual and practical benefits of representational state transfer. *Journal of Documentation*, 71(2), 233–252. <https://doi.org/10.1108/JD-07-2013-0098>
- Veyne, P. (1987). *Como se escreve a história*. Lisboa: Edições 70.
- Visual Resources Association. (1996). VRA Core 1.0 - Introduction. Disponível em: [https://www.loc.gov/standards/vracore/VRACore1\\_Introduction.pdf](https://www.loc.gov/standards/vracore/VRACore1_Introduction.pdf)
- Visual Resources Association. (1999). VRA Core 2.0 - Introduction. Disponível em: [https://www.loc.gov/standards/vracore/VRACore2\\_Introduction.pdf](https://www.loc.gov/standards/vracore/VRACore2_Introduction.pdf)
- Visual Resources Association. (2002). VRA Core 3.0. Disponível em: [https://www.loc.gov/standards/vracore/VRACore3\\_Introduction.pdf](https://www.loc.gov/standards/vracore/VRACore3_Introduction.pdf)
- Visual Resources Association. (2007a). VRA Core 4.0 - Element Outline. Disponível em: [https://www.loc.gov/standards/vracore/VRA\\_Core4\\_Outline.pdf](https://www.loc.gov/standards/vracore/VRA_Core4_Outline.pdf)
- Visual Resources Association. (2007b). VRA Core 4.0 Element Description.
- Visual Resources Association. (2014). VRA Core 4.0 Introduction. Disponível em: [http://www.loc.gov/standards/vracore/VRA\\_Core4\\_Intro.pdf](http://www.loc.gov/standards/vracore/VRA_Core4_Intro.pdf)
- Vrettas, G., & Sanderson, M. (2015). Conferences versus journals in Computer Science. *Journal of*

- the Association for Information Science and Technology*, 66(12), 2674–2684.  
<https://doi.org/10.1002/asi.23349>
- W3C. (2014a). RDF 1.1 Concepts and Abstract Syntax. Disponível em:  
<https://www.w3.org/TR/rdf11-concepts/>
- W3C. (2014b). RDF 1.1 Primer. Disponível em: <https://www.w3.org/TR/rdf11-primer/>
- Wartofsky, M. (1979). *Models: Representation and the Scientific Understanding* (Vol. 129). Dordrecht: D. Reidel Publishing Company.
- Webb, E., Campbell, D., Schwarz, R., & Sechrest, L. (1984). The use of archival sources in social research. In *Sociological research methods - an introduction*. London: MacMillan.
- Webb, Eugene, Campbell, D., Schwartz, R., & Schrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago: Rand McNally.
- Webb, Eugene, Campbell, D., Schwartz, R., & Schrest, L. (1981). *Nonreactive Measures in the Social Sciences*. Dallas: Houghton Mifflin.
- Weber, R. (1985). *Basic content analysis*. Beverly Hills: SAGE Publications.
- Weibel, S., Godby, J., Miller, E., & Daniel, R. (1995). OCLC/NCSA Metadata Workshop Report.
- Weinberger, D. (2009). Transparency is the new objectivity. Disponível em:  
<http://www.hyperorg.com/%0Ablogger/2009/07/19/transparency-is-the-new-objectivity/>
- Weinberger, D. (2010). The Problem with the Data-Information-Knowledge-Wisdom Hierarchy. *Harvard Business Review*. Disponível em: <https://hbr.org/2010/02/data-is-to-info-as-info-is-not>
- Wellisch, H. (1981). Ebla: the world 's oldest library. *The Journal of Library History*, 16(3), 488–500.
- Wersig, G., & Neveling, U. (1975). The phenomena of interest to Information Science. *Information Scientist*, n. 4, 9(4), 127–140. <https://doi.org/10.1017/CBO9781107415324.004>
- Westwell, C. (2012). Found in a Bricked-Up Chamber: The Reading Abbey Library Catalogue. Disponível em: <https://blogs.bl.uk/digitisedmanuscripts/2012/11/found-in-a-bricked-up-chamber-the-reading-abbey-library-catalogue.html>
- While, A. (1987). Records as a data source: the case for health visitor records. *Journal of Advanced Nursing*, 12, 757–763.
- Whitfield, R., & Farrer, A. (1990). *Caves of the thousand Buddhas: Chinese art from the silk route*. London: British Museum Publications.
- Wieggers, K., & Beatty, J. (2013). *Software Requirements*. Redmond: Microsoft Press. <https://doi.org/10.3362/9781780449357>
- Wiersma, W. (1995). *Research methods in Education: An Introduction*. Boston: Allyn and Bacon.
- Wilkinson, M., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3. <https://doi.org/10.1038/sdata.2016.18>
- Wilson, M. (2012). *Search user interface design. Synthesis Lectures on Information Concepts, Retrieval, and Services* (Vol. 3). San Rafael, CA: Morgan & Claypool Publishers.
- Winograd, T., & Flores, F. (1986). *Understanding computers and cognition*. Reading: Addison-Wesley.
- Witten, I., Bainbridge, D., & Nichols, D. (2010). *How to build a digital library*. Oxford: Elsevier, Morgan Kaufmann.
- Witty, F. (1958). The Pinakes of Callimachus. *Library Quarterly*, 28(2), 132–136. <https://doi.org/10.1086/618523>
- Woodley, M. (2016). Metadata Matters: Connection People and Information. In M. Baca (Ed.), *Introduction to Metadata* (3° ed, pp. 38–60). Los Angeles: Getty Research Institute.
- Woolley, L. (1950). *Ur of the Chaldees: a record of seven years of excavation*. Harmondsworth: Penguin Books.
- Wosh, P. (2011). *Waldo Gifford Leland and the Origins of the American Archival Profession*. Chicago: Society of American Archivists.
- Wright, K. (2014). Broadening the record and expanding the archives. *Archives & Manuscripts*, 42(2), 219–221. <https://doi.org/10.1080/01576895.2014.911693>
- Yakel, E. (2003). Archival representation. *Archival Science*, 3(1), 1–25. <https://doi.org/10.1007/BF02438926>

- Yardley, L. (Ed.). (2002). *Material Discourses of Health and Illness*. London and New York: Routledge.
- Yeo, G. (2007). Concepts of Record (1): evidence, information, and persistent representations. *American Archivist*, 70(2), 315–343. <https://doi.org/10.17723/aarc.70.2.u327764v1036756q>
- Yeo, G. (2008). Concepts of Record (2): prototypes and boundary objects. *American Archivist*, 71(1), 118–143. <https://doi.org/10.17723/aarc.71.1.p0675v40tr14q6w2>
- Yeo, G. (2011). Rising to the level of a record? Some thoughts on records and documents. *Records Management Journal*, 21(1), 8–27. <https://doi.org/10.1108/09565691111125071>
- Yeo, G. (2012). Bringing things together: Aggregate records in a digital age. *Archivaria*, 74(Fall 2012), 43–91.
- Yeo, G. (2013). Trust and context in cyberspace. *Archives & Records*. Taylor & Francis. <https://doi.org/10.1080/23257962.2013.825207>
- Yeo, G. (2018). *Records, information and data*. London: Facet Publishing.
- Yin, R. (2010). *Estudo de caso: planejamento e métodos* (4th ed.). Porto Alegre: Bookman.
- Youn, E. (2015). Adoption of ISAD(G) in practice: a close look at the standardization process of ISAD(G) in a manuscript archives of Korea. *Archives & Records*, 36(2), 128–145. <https://doi.org/10.1080/23257962.2015.1029892>
- Zappavigna, M. (2018). *Searchable talk : hashtags and social media metadiscourse*. London and New York: Bloomsbury Academic, an imprint of Bloomsbury Plc,.
- Zave, P. (1997). Classification of research efforts in requirements engineering. *ACM Computing Surveys*, 29(4), 315–321. <https://doi.org/10.1145/267580.267581>
- Zeichner, K. (1983). Alternative Paradigms of Teacher Education. *Journal of Teacher Education*, 34(3), 3–9. <https://doi.org/10.1177/002248718303400302>
- Zeng, M. L. (2017). Smart Data for Digital Humanities. *Journal of Data and Information Science*, 2(1), 1–12. <https://doi.org/10.1515/jdis-2017-0001>
- Zeng, M. L., & Chan, L. M. (2010). Semantic Interoperability. In M. Bates & M. Maack (Eds.), *Encyclopedia of Library and Information Sciences. Volume I* (3rd ed., pp. 4645–4662). New York: Dekker Encyclopedias, Taylor and Francis Group.
- Zeng, M. L., & Qin, J. (2016). *Metadata (2°)*. London: Facet Publishing.
- Zhang, J. (2012a). Archival Context, Digital Content, and the Ethics of Digital Archival Representation. *Knowledge Organization*, 39(5), 332–339. <https://doi.org/10.5771/0943-7444-2012-5-332>
- Zhang, J. (2012b). Archival representation in the digital age. *Journal of Archival Organization*, 10(1), 45–68. <https://doi.org/10.1080/15332748.2012.677671>
- Zhang, J. (2012c). Original order in digital archives. *Archivaria*, 74(Fall 2012), 167–193.
- Zhang, J., & Mauney, D. (2015). When Archival Description Meets Digital Object Metadata: A Typological Study of Digital Archival Representation. *American Archivist*, 76(1), 174–195. <https://doi.org/10.17723/aarc.76.1.121u85342062w155>
- Zins, C. (2007). Conceptual approaches for defining data, information and knowledge. *Journal of the American Society for Information Science and Technology*, 58(4), 479–493. <https://doi.org/10.1002/asi>

Tese realizada com o apoio de:

