



# TENDÊNCIAS ATUAIS E PERSPETIVAS FUTURAS EM ORGANIZAÇÃO DO CONHECIMENTO

ATAS DO III CONGRESSO ISKO ESPANHA-PORTUGAL  
XIII CONGRESSO ISKO ESPANHA

*Universidade de Coimbra, 23 e 24 de novembro de 2017*

Com a coordenação de

---

Maria da Graça Simões, Maria Manuel Borges

TÍTULO

Tendências Atuais e Perspetivas Futuras em Organização do Conhecimento: atas do III Congresso ISKO Espanha e Portugal - XIII Congresso ISKO Espanha

COORDENADORES

Maria da Graça Simões  
Maria Manuel Borges

EDIÇÃO

Universidade de Coimbra. Centro de Estudos Interdisciplinares do Século XX - CEIS20

ISBN

978-972-8627-75-1

ACESSO

<https://purl.org/sci/atas/isko2017>

COPYRIGHT

Este trabalho está licenciado com uma Licença Creative Commons - Atribuição 4.0 Internacional (<https://creativecommons.org/licenses/by/4.0/deed.pt>)

OBRA PUBLICADA COM O APOIO DE



FLUC FACULDADE DE LETRAS  
UNIVERSIDADE DE COIMBRA



CEIS 20  
CENTRO DE ESTUDOS  
INTERDISCIPLINARES  
DO SÉCULO XX  
UNIVERSIDADE DE COIMBRA

**FCT**  
Fundação para a Ciência e a Tecnologia  
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

PROJETO UID/HIS/00460/2013

# CLASSIFICAÇÃO DA INFORMAÇÃO ARQUIVÍSTICA SEGUNDO MÉTODOS FILOMEMÉTICOS: METADADOS COMO UNIDADES FENOTÍPICAS?

L. S. Ascensão de Macedo

*Universidade de Coimbra, 0000-0001-7251-7314, laureanomacedo@student.fl.uc.pt*

**RESUMO** Este artigo explora métodos filogenéticos aplicados à classificação da informação arquivística. Analisam-se paralelismos discursivos baseados na ‘metáfora biológica’ existente na literatura arquivística e na CI, bem como as aplicações deste método nas Humanidades Digitais, conhecida como filomemética. Para obter resultados que nos permitam analisar do ponto de vista evolucionista a estrutura de um sistema de informação arquivístico, analisamos uma amostra descrita numa matriz de distâncias/similaridade baseada em propriedades do Records in Context. Em resultado disto, efetua-se neste artigo uma análise em torno dos outputs das aplicações utilizadas no âmbito da análise filogenética. O contributo deste estudo exploratório visa contrapor os modelos top-down (base funcional) de representação da informação arquivística face a um modelo de representação científica e evolucionista dos conjuntos informacionais. Conclui-se que os metadados podem constituir-se como unidades fenotípicas por permitirem apresentar padrões de (dis)similaridade entre as unidades informacionais analisadas.

**PALAVRAS-CHAVE** *Classificação arquivística, metáfora biológica, filomemética, metadados, traços fenotípicos.*

**ABSTRACT** This paper explores phylogenetic methods applied to archival information and knowledge classification. We analyse discourses about the “biological metaphor” in the archival and information science literature, and their methodological applications in the Digital Humanities, known also as phylomemetics. We adopt an exploratory method using metadata as phenotypic traits and RiC properties to build distance/similarity matrices with dendritic visualizations based on set of taxa (typological records). This study analyses the outputs of phylogenetic tools to support our interpretation of the data. The importance of this study shows us alternative methods to functional analysis in archival classification theory to an evolutionary classification of archival information systems.

**KEYWORDS** *Archival classification, biological metaphor, phylomemetics, metadata, phenotypic traits.*

**COPYRIGHT** Este trabalho está licenciado com uma Licença Creative Commons - Atribuição 4.0 Internacional (<https://creativecommons.org/licenses/by/4.0/deed.pt>)

## INTRODUÇÃO

A classificação arquivística consiste numa atividade de organização sistemática e estruturada de conjuntos informacionais arquivísticos, com o objetivo de controlar e proporcionar meios para descrever, representar e recuperar informação (Simões & Freitas, 2013; Simões, Freitas, & Rodríguez-Bravo, 2016). A representação da informação arquivística materializa-se em instrumentos de pesquisa,

*u. g.*, roteiros, guias, inventários, catálogos, índices, *thesauri* e bases de dados relacionais (MacNeil, 2012), podendo conter *instrumenta* sintéticos como planos, grelhas, quadros ou tabelas de seleção, com objetivos distintos de utilização. No entanto, a classificação arquivística não se cinge apenas às necessidades de recuperação de informação, mas constitui a ‘coluna vertebral’ tanto para a tomada de decisão no âmbito da avaliação da informação como na definição de níveis de acesso, de proteção de dados, de confidencialidade, de preservação e conservação, inclusa a (re)engenharia dos processos de trabalho (Mokhtar & Yusof, 2015). Ainda assim, a maioria dos instrumentos de pesquisa adotou uma visualização hierarquizada e multinível das classes informacionais ou documentais (Gartner, 2016), de acordo com determinadas regras e convenções adotadas a nível internacional e nacional para a sua descrição (cf. DGRQ, 2011), com forte penetração em sistemas de informação eletrónicos (Niu, 2013). A representação da informação arquivística, segundo as convenções normativas supramencionadas, permanece configurada de forma hierarquizada, estática, unidimensional, idiossincrásica, onde o fundo/arquivo é descrito como um sistema fechado, qual “silo”, e tratado como resíduo da atividade das organizações. Os mais recentes estudos apontam para a representação multidimensional da informação arquivística segundo modelos conceituais e sua transposição para ontologias formais para suporte à descrição. Esta perspetiva aproxima-se dos pressupostos da ISO 21.127 (2014), como é o caso do *RiC: Records in Context* (Gueguen et al., 2013; ICA, 2016). Apesar de a literatura arquivística referir-se à ‘metáfora biológica’ como um traço distintivo em relação aos modelos de classificação em bibliotecas, esta perspetiva carece de uma análise à luz da Teoria Evolucionista.

Este artigo efetua um estudo exploratório aos métodos filomeméticos enquanto modelos de classificação científica aplicáveis à representação da informação arquivística. Analisar-se-ão os conceitos operatórios associados à ‘metáfora biológica’, com uma contextualização das aplicações de modelos filomeméticos aplicados nas Humanidades Digitais, seguido de uma exemplificação de aplicação em torno de unidades informacionais.

## CONTEXTUALIZAÇÃO

### A ‘METÁFORA BIOLÓGICA’ NA ARQUIVÍSTICA E O DISCURSO EVOLUCIONISTA EM CI

As comparações inspiradas em processos biológicos para explicar fenómenos na arquivística não são espúrias. Autores fundamentais da arquivística, como Muller, Feith & Fruin (1898), Jenkinson (1922), Casanova (1928) estabeleciam analogias em torno do processo de formação e natureza dos arquivos baseadas em ‘metáforas biológicas’ (Adell, 2010; Bak, 2012; Hurley, 1993). Na terminologia arquivística, existem vários conceitos definidos com base no símile biológico, materializado em termos como arquivo, fundo, série e, designadamente, organicidade (cf. Quadro 1).

Quadro 1. *Arquivo e Organicidade na terminologia arquivística (português).*

Entrada terminológica	Definição*	Fonte
<b>Arquivo</b>	<i>Conjunto <b>orgânico</b> de documentos, independentemente da sua data, forma e suporte material, produzidos ou recebidos por uma pessoa jurídica, singular ou colectiva, ou por um organismo público ou privado, no exercício da sua actividade e conservados a título de prova ou informação. É a mais ampla unidade arquivística. A cada proveniência corresponde um arquivo.</i>	(BNP, 2010, p. 379)
<b>Organicidade</b>	<i>Característica que decorre do facto de os arquivos reflectirem, enquanto <b>produto natural</b> da actividade de uma administração – no seu todo ou em cada uma das suas unidades, bem como nas relações entre elas – a vontade e o funcionamento dessa administração. Fundamenta os princípios da proveniência e do respeito pela ordem original.</i>	(BNP, 2010, p. 379)
	<i><b>Relação natural</b> entre documentos de um arquivo em decorrência das atividades da entidade produtora.</i>	(Brasil. Arquivo Nacional, 2005)

\* Sublinhados nossos.

Na terminologia inglesa, vários termos evidenciam esta perspectiva biológica, *u. g. records aggregation* (IP2, 2008), *archival bond* (Duranti, 2015), *archival sedimentation* (Mata Caravaca, 2015), *archival nature*, *accumulation*, *organic collection* (Pearce-Moses, 2005), incluindo equívocos, como expressões “archival darwinism” (Ham, 1993) ou “neo-Darwinian construction” (Cook, 2011).

A estrutura hierarquizada das classes arquivísticas – arquivo/fundo, secção, série, documento composto/simples – tem sido entendida como uma conexão entre unidades informacionais (Adell, 2010). O símile entre série e a *species* lineana também não é espúrio (Adell, 2010; Hurley, 1993; Janzen, 1994; Mai, 2011; Rieppel, 2010; Wilkins, 2009). Podemos encontrar em Jenkinson, em 1943, a formulação de uma estrutura inspirada na taxonomia lineana:

*In the case of Archives Species, Genera and Orders are represented by Classes, Groups and Categories of Records: which are, to an almost equal extent, natural, not artificial, divisions.*  
(Jenkinson, 1943/2003, p. 200)

O discurso evolucionista no âmbito da CI conta com importantes contributos que permitem dispor de uma base teórica e interdisciplinar no âmbito da filogenia cultural (Gnoli, 2006). Assiste-se a uma aplicação de modelos filomeméticos para explicar processos evolutivos em sistemas de informação enquanto abordagens emergentes, designadamente

- da curadoria digital (Thomer & Weber, 2014; Weber, 2015; Weber & Thomer, 2016);
- da computação forense (Costa et al., 2015; Lanterna & Barili, 2017; Oliveira et al., 2016);
- da estematologia/crítica textual (Andrews & Macé, 2012; Heikkilä, 2007), em alternativa aos métodos de *document clustering* (Huang, 2008).

## FILOGENÉTICA / FILOMEMÉTICA NAS HUMANIDADES E CIÊNCIAS SOCIAIS

Darwin defendeu que *all true classification is genealogical* (1859, p. 420), para reforçar a validade dos princípios da descendência com modificação e da hereditariedade. Tais princípios darwinianos constituem um dos pilares da filogenética sistemática (Lecointre, 2015).

No entanto, a filogenética não se tornou numa disciplina exclusiva da Biologia Evolutiva: as suas potencialidades transdisciplinares tiveram larga aplicação, por exemplo, na classificação de macrofamílias linguísticas (Atkinson, 2006; Fangerau, Geisler, Halling, & Martin, 2013), na crítica textual (Andrews & Macé, 2012; Heikkilä, 2007), na arqueologia evolutiva (García Rivero, 2013, 2016; O'Brien & Lyman, 2003), na teoria organizacional (McCarthy, 2005), nos processos evolutivos em tecnologias industriais (Lee, Jung, Yoon, Kim, & Jung, 2013) e na evolução cultural (Currie & Mace, 2011; Mesoudi, 2011, 2016). A transposição de métodos filogenéticos para o domínio das Humanidades e Ciências Sociais constitui uma abordagem emergente sobre a aplicação de métodos baseados na epistemologia evolucionista (Gontier, 2012) e na exploração de modelos algorítmicos meta-heurísticos (Eiben & Smith, 2012).

Os modelos de classificação da Sistemática Filogenética resultaram em três escolas filosóficas: a fenética, a cladística e a taxonomia filogenética (García Rivero, 2010). A diferença entre estas três escolas, exemplificadas sinteticamente *infra* (Quadro 2), consiste no facto de a fenética não se basear no princípio de descendência com modificação, ao contrário da cladística e da taxonomia evolutiva. A visualização das relações entre os *taxa* em dendrogramas, os quais constituem hipóteses de visualização (Morrison, 2014).

Quadro 2. Escolas filogenéticas pós-síntese

<i>Escola</i>	<i>Método de classificação</i>	<i>Modelo de visualização</i>	<i>Procedimentos</i>	<i>Autores fundamentais</i>	<i>Problemas</i>
<i>Fenética (numerical taxonomy)</i>	<i>Similaridade morfológica</i> <i>Unidade: OTU</i>	<i>Fenograma</i>	<i>Estatístico (análise fatorial, análise de conglomerados, análise coordenadas principais).</i>	(Sneath & Sokal, 1973; Sokal, 1986)	<i>- não explica a história evolutiva;</i> <i>- gera paralelismos, convergências evolutivas.</i>
<i>Cladística (sistemática filogenética)</i>	<i>abordagem “descendência com modificação”:</i> <i>carateres análogos, homologia, sinapomorfias, autapomorfias, simplesiomorfia.</i>	<i>Cladograma</i>	<i>Estatístico (máxima verossimilhança, de inferência bayesiana, método dos mínimos quadrados, parsimónia)</i>	(Hennig, 1975)	<i>- critério da parsimónia insuficiente;</i> <i>- desvalorização da anagénesse</i> <i>- ausência de critérios de ranking.</i>
<i>Taxonomia evolutiva</i>	<i>Similitude entre classes, grupos monofiléticos e parafiléticos, anagéneses.</i>	<i>Filograma</i>	<i>Estatístico (máxima verossimilhança, de inferência bayesiana, método dos mínimos quadrados, parcimónia)</i>	(Mayr, 1981; Mayr & Bock, 2002)	<i>- critério de semelhança e genealogia não é passível de ser representada simultaneamente.</i>

Fonte: adaptado a partir de García Rivero (2010).

Sendo os fenótipos o conjunto de propriedades morfológicas de um organismo (Mahner & Kary, 1997), em CI tem como possíveis equivalentes os metadados, ativos semânticos e propriedades significativas, uma vez que constituem elementos descritivos de artefactos informacionais (Gartner, 2016).

## MATERIAIS E MÉTODOS

Sendo o documento de arquivo a unidade mínima de um fundo na hierarquia de classes, a análise documental através dos métodos da diplomática contemporânea consiste na identificação de carateres externos e internos das tipologias documentais (Rogers, 2015, 2016).

Os esquemas de metadados – tanto de tipo descritivo, administrativo, técnico como de preservação,

além da variedade e diversidade de esquemas (cf. Riley, 2010) – podem ser analisados segundo processos evolutivos: *some metadata are re-used (inheritance)*, *other metadata are eliminated (extinction)* and *still others are updated or newly generated (mutation)* (Niu, 2013). De modo similar, os traços fenotípicos culturais são objeto de *recombination, copying error, and [...] can be the foundation for the production of new traits* (O'Brien, Lyman, Mesoudi, & VanPool, 2010).

Propomos a seguinte questão de investigação: os metadados possibilitam reconstituir um sistema de informação arquivístico segundo uma perspectiva evolucionista, constituindo-se como hipótese de classificação científica desse sistema informacional? Adota-se um método misto (qualitativo e quantitativo) segundo uma abordagem de descrição *bottom-up* (documento a documento). Para possibilitar um estudo filomemético, estabeleceu-se o seguinte fluxo de trabalho:

- (i) **Coleta de dados e processamento.** Seleccionámos aleatoriamente 9 documentos compostos (N= 41 peças documentais) de tipologia diversa classificados segundo MEF/ASIA (DGLAB, 2016) proveniente do sistema de informação do Governo Regional da Madeira (GRM), referente ao período de registo de 2014 a 2016.

Quadro 3. Classes MEF/ASIA

Classe MEF/ASIA	Nota de definição	Relação com outros PN
300.10.302 Eliminação de documentos e informação	Destruição controlada de documentação e informação administrativa, independentemente do suporte, decorridos os prazos legais. Inicia com a proposta e termina com a comunicação da eliminação à entidade competente. Inclui análise da documentação ou informação, verificação de prazos legais de conservação administrativa e de destino final, autorização de eliminação e elaboração de auto de eliminação.	Complementar#300.30.008#Registo de documentos e informação# Complementar#300.30.602#Avaliação de informação arquivística#
300.10.600 Depósito de bens	Aquisição da responsabilidade da gestão de bens de terceiros. Inicia com o pedido de depósito ou entrega de bens e termina com a devolução ou decisão sobre o destino dos bens. Inclui o registo dos bens a depositar, a sua receção e confirmação de depósito.	300.30.008#Registo de documentos e informação# Síntese(sintetizado)
300.30.008 Registo de documentos e informação	Inserção de dados relativos a documentos e informação, independentemente da sua natureza ou suporte. Inicia com a recolha de dados e termina com o seu registo. Inclui atualização de dados.	Complementar#300.30.602#Avaliação de informação arquivística# Complementar#300.10.302 Eliminação de documentos e informação#
<b>300.30.602 Avaliação de informação arquivística</b>	Atribuição de um valor informacional, de acordo com critérios de valorimetria aplicáveis, para efeitos de determinação do destino final da informação/documentação. Inicia com a identificação do objeto a avaliar e termina com a produção de instrumento de avaliação. Inclui estudo do contexto de produção, criação de estrutura de organização intelectual e justificação dos valores atribuídos.	Complementar#300.10.302 Eliminação de documentos e informação# Complementar#300.30.008 Registo de documentos e informação# Complementar#300.40.501 Conservação e proteção de informação#

Fonte: extraído de DGALB (DGLAB, 2016).

O tamanho da amostra é meramente experimental e constituímos uma matriz binária, acessível em Macedo (2017), onde se descreve cada documento simples pelas suas tipologias documentais identificadas no Quadro 4. Na matriz binária atribuíram-se valores numéricos: 0 significa que uma tipologia documental não possui determinada propriedade *RiC* e 1 possui essa propriedade, conforme Quadros 5.

Quadro 4. Relação de tipologias documentais

Abreviatura	Descrição	Tradição	Taxon	Obs.
auto_ceden	Auto de cedência a título devolutivo	original	I4	
auto_elim_d_pdg	Auto de eliminação, por via de portaria de gestão de documentos	cópia	C2, H2	
auto_elim_o_pdg	Auto de eliminação, por via de portaria de gestão de documentos	original	C1, H1	
auto_elim_rada_d	Auto de eliminação, por via de relatório de avaliação	duplicado	B2	
auto_elim_rada_o	Auto de eliminação, por via de relatório de avaliação	original	B1	
auto_de_entre	Auto de entrega	original	E1	
auto_incorp_d	Auto de incorporação	duplicado	F2	
auto_incorp_o	Auto de incorporação	original	F1	
guia-r_d_rada	Guia de remessa, por via de relatório de avaliação	duplicado	D2	
guia-r_o_rada	Guia de remessa, por via de relatório de avaliação	original	D1	
inf_int_o	Informação interna	original	A3, B3, C3, D3, G3, H3, I1	
idd_d	Instrumento de pesquisa	cópia	G2	
idd_o	Instrumento de pesquisa	original	G1	
ofic_c	Ofício	cópia	A5, B5, C5, D5, F5, G5, H5	
ofic_o	Ofício	original	A4, B4, C4, D4, F4, G4, H4, I2	
ofic_rem	Ofício	original	I3, F3, A6	Entidade externa
oficio_rem_o				
O_resp_oGA				
rada_d	Relatório de avaliação	duplicado	A2	
rada_o	Relatório de avaliação	original	A1	

Fonte: extraído de Macedo (Macedo, 2017).

Quadro 5. Extrato de matriz binária

Taxa	A1	A2	A3	A4	A5	A6	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	E1	F1	F2	F3	F4	F5	G1	G2	G3	G4	G5	H1	H2	H3	H4	H5	I1	I2	I3	I4			
RiC-R074	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0		
RiC-R075	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
RiC-R078	1	1	0	0	0	0	1	1	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	1	
RiC-R079	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
RiC-R080	1	1	0	0	0	0	1	1	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	
RiC-R082	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
RiC-R085	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RiC-R086	1	1	1	0	0	1	1	1	1	0	0	1	1	1	0	0	1	1	0	0	1	1	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	0	0	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Fonte: extraído de Macedo (Macedo, 2017), através de doi:10.7910/DVN/ERE8HL.

Quadro 6. Relação de entidades e de propriedades do RiC

Set of Properties	Properties	Set of Properties	Properties
Shared Properties of all Entities	RiC-P1 Global Persistent Identifier	Additional Property Specific to Person and Person Assumed Identity	RiC-P36 Gender
	RiC-P2 Local Identifier	Additional Property Specific to Delegate-Agent	RiC-P37 Technical Characteristics
	RiC-P3 Name	Additional Properties Specific to Corporate Body and Corporate Body Assumed Identity	RiC-P38 Services to the Public
	RiC-P4 General Note		RiC-P39 Contact Information
RiC-P5 Authenticity and Integrity Note	RiC-P40 Operating Hours		
Content Information	RiC-P6 Content Type	Properties of Position	RiC-P41 Facilities
	RiC-P7 Content Extent		RiC-P42 Type
	RiC-P8 Quality of Information		RiC-P43 Description
	RiC-P9 Scope and Content		RiC-P44 History
	RiC-P10 Encoding Format	Properties of Function	RiC-P45 Type
	RiC-P11 Language Information		RiC-P46 Description
	RiC-P12 Media Type		RiC-P47 History
	RiC-P13 Production Technique		RiC-P48 Type
Information about Carrier	RiC-P14 Medium	Property of Function (Abstract)	RiC-P49 Description
	RiC-P15 Physical or Logical Extent		RiC-P50 History
Information about Management and Use	RiC-P16 Physical Characteristics Note	Properties of Activity	RiC-P51 Description
	RiC-P17 Classification		RiC-P52 Type
	RiC-P18 Conditions of Access		RiC-P53 Description
	RiC-P19 Conditions of Use	RiC-P54 History	
	RiC-P20 History	Properties of Mandate	RiC-P55 Type
RiC-P21 Record State	RiC-P56 Description		
Properties of Record Set	RiC-P22 Authenticity and Integrity Note	Properties of Documentary Form	RiC-P57 History
	RiC-P23 Type		RiC-P58 Type
	RiC-P24 Accrual Note		RiC-P59 Description
	RiC-P25 Accrual Status	Properties of Date	RiC-P60 History
	RiC-P26 Arrangement		RiC-P61 Type
	RiC-P27 Classification		RiC-P62 Calendar
Properties Summarizing the Members of a Record Set	RiC-P28 History	Properties of Place	RiC-P63 Type
	RiC-P29 Content Extent		RiC-P64 Geographic Coordinates
	RiC-P30 Physical or Logical Extent		RiC-P65 Address
Properties Shared by All Agents	RiC-P31 Scope and Content	Properties of Concept/Thing	RiC-P66 Type
	RiC-P32 Type		RiC-P67 Description
	RiC-P33 Identity Type		
	RiC-P34 Language Information		
	RiC-P35 History		

Fonte: extraído de ICA.EGAD (ICA, 2016)

Refira-se que o esquema *RiC* possui 67 propriedades (Quadro 6) com 792 relações multientidades. As propriedades *RiC* foram reduzidas a 143 propriedades (marcadas em inglês pelo presente do indicativo).

De acordo com os Quadros 4 e 5 *supra* exarados, as referências aos documentos (*taxa*) são representadas na sequência de dois dígitos alfanuméricos em que a letra maiúscula representa o processo documental pertencente a um processo de negócio concreto e o número a tipologia documental (*u. g.*, *A1* decompõe-se em *A* equivale à classe 300.30.602 *Avaliação da Informação Arquivística*, e 1 corresponde a tipologia documental *relatório de avaliação*; *A4*: *A* = 300.30.602 *Avaliação da Informação Arquivística*; 4 = *ofício*).

Para aplicar relações de propriedades *RiC-R* com valor de 1, então *A1* consiste num *Record Component* (*RiC-R075*) que é parte de (*is part of*) de *Record* (*RiC-R14*), e inversamente (*RiC-R14-Record has part*

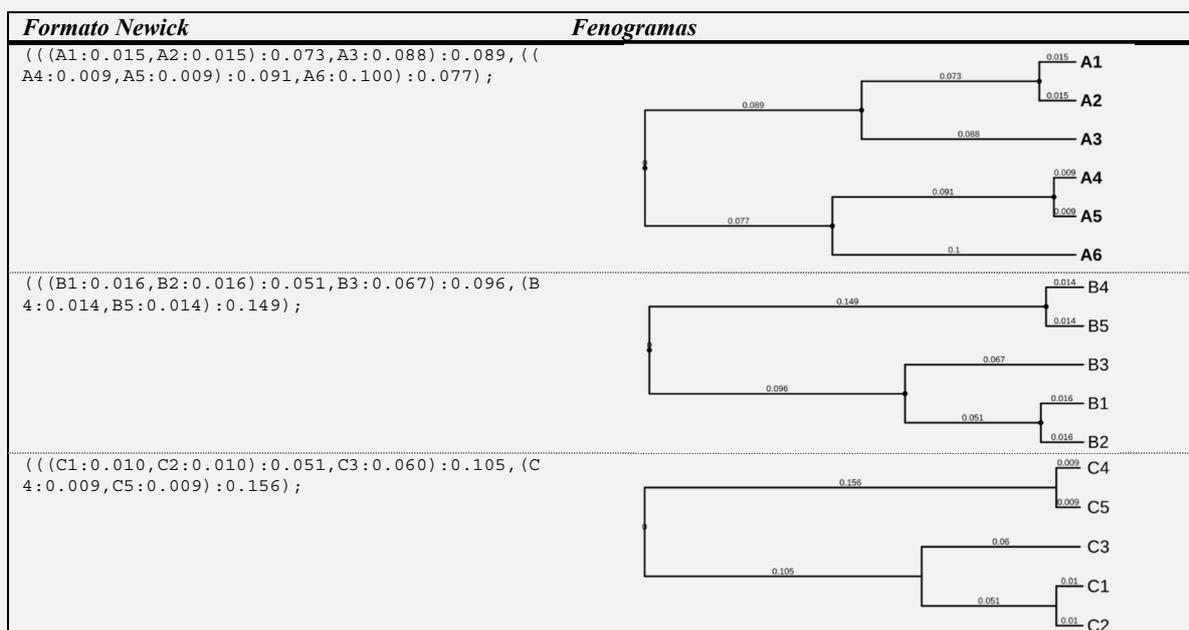
RiC-R075-Record Component). No caso de não ser explícita a presença de uma propriedade, recebe valor de 0. Obteríamos, por exemplo, uma sequência binária do tipo:

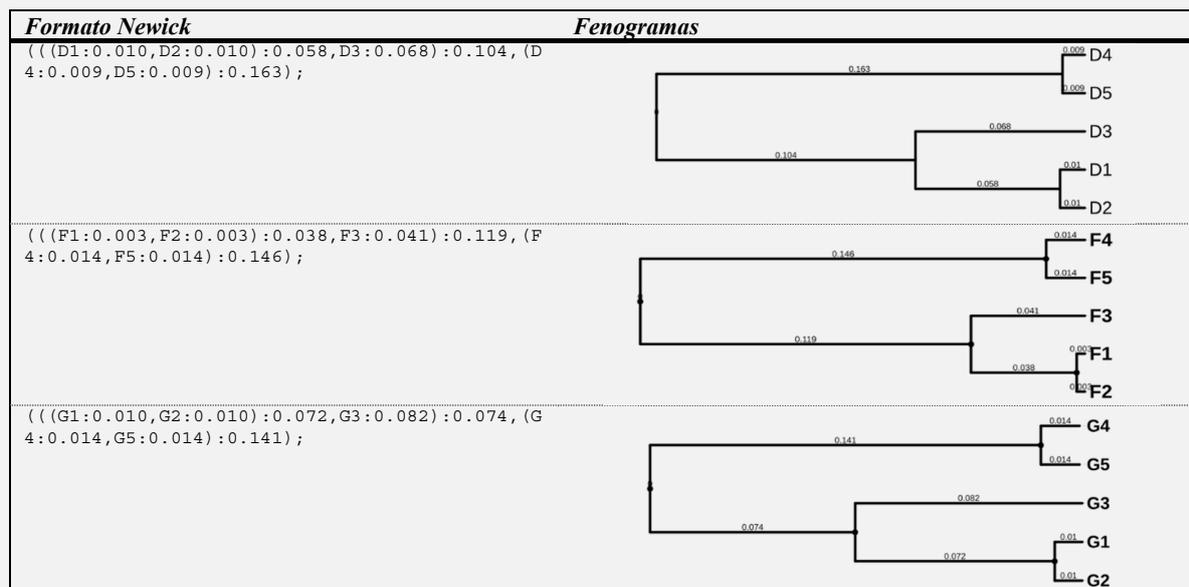
```
> A1
01101101111010111011011111110101011010011111101010100101010111110
1010000011010000110000000000110110111000011111100101101001111110
001111110000010000000000001000000001000000000000000000000000000100010
100010.
```

- (ii) **Análise e visualização de dados.** Seleccionámos as aplicações *web DendroUPGMA* (Garcia-Vallve, Palau, & Romeu, 1999) para processamento de dados binários e para extração de *outputs* e *iTOL* (Letunic & Bork, 2016), para visualização dendrífica. Convertemos automaticamente os *datasets* binários em matrizes de distâncias segundo o coeficiente de Sørensen–Dice e exportado para o formato Newick (Cardona, Rosselló, & Valiente, 2008), cujos *outputs* são acessíveis em Macedo (2017). Adotámos um método de agrupamento hierarquizado UPGMA (*Unweighted Pair Group Method with Arithmetic Mean*), cujo resultado consiste num dendrograma baseado na (dis)similaridade entre *taxa* (Pavlopoulos, Soldatos, Barbosa-Silva, & Schneider, 2010).

## RESULTADOS E DISCUSSÃO

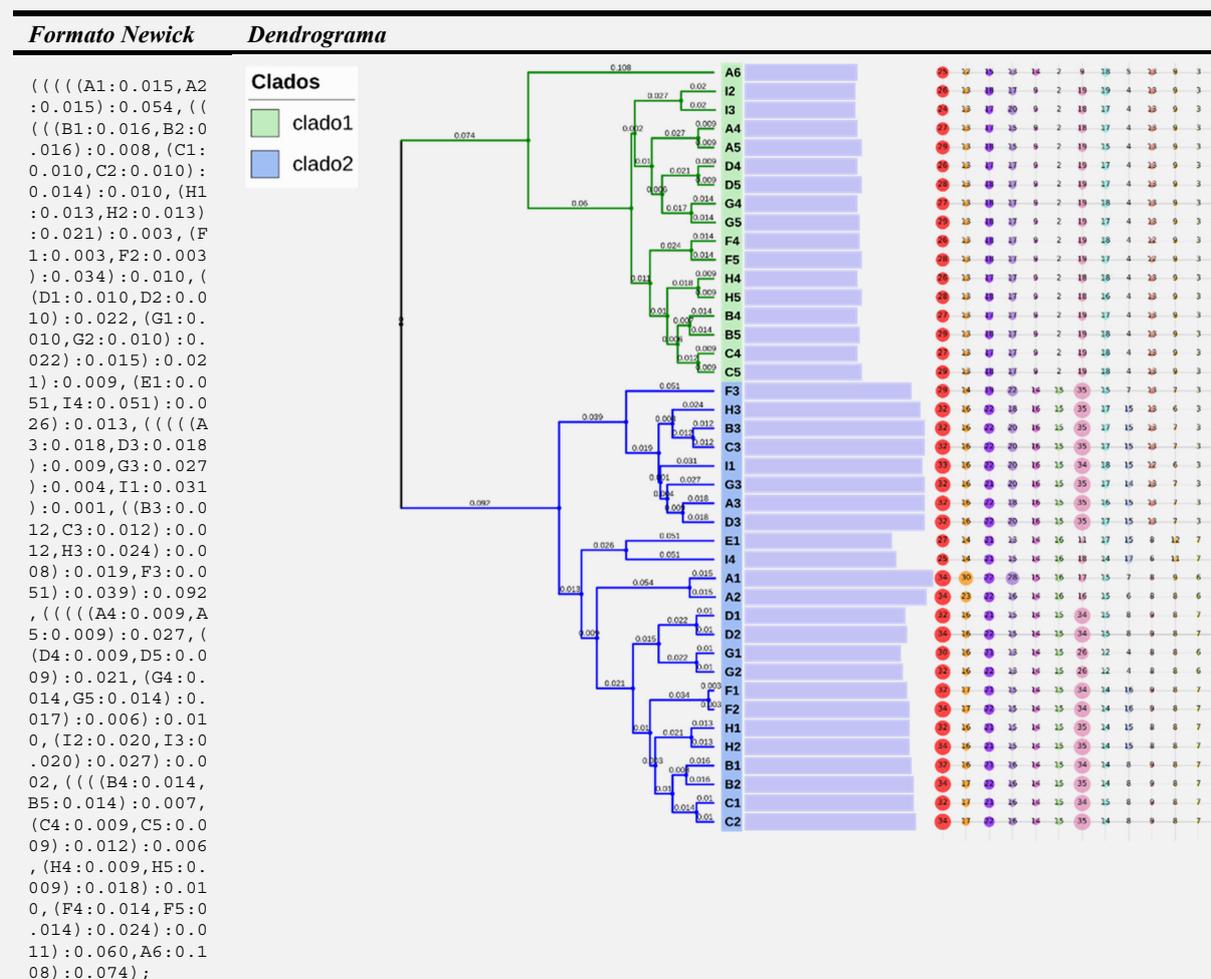
Analisada cada tipologia documental, obtivemos numa matriz de binária 201 variáveis (*RiC* propriedades-relações) em 41 *taxa* (documentos simples). Obtivemos três tipos de *outputs*, calculado automaticamente por *DendroUPGMA* através do coeficiente de Sørensen-Dice, que resultou (i) numa matriz de distâncias, (ii) numa matriz de similaridade e (iii) correspondente formato Newick e respetivo dendrograma (Figura 1). O algoritmo UPGMA constituiu um parâmetro pré-definido em *DendroUPGMA* para a constituição de *clusters* e visualização dendrífica.





**Figura 1.** Fenogramas de relações entre taxa (tipologias documentais), por processos documentais, processado em *DendroUPGMA* (Garcia-Vallve et al., 1999)

A partir da Figura 1, verifica-se que a metainformação do *RiC*, enquanto traços fenotípicos, permite uma diferenciação entre os *taxa*. As relações entre os *taxa* de cada dendrograma evidenciam similaridades, como A1-A2, B1-B2, C1-C2, D1-D2, F1-F2 e G1-G2 (cf. *supra* Quadro 4), distinguindo-se entre si ao nível das tipologias documentais e ao nível do suporte e da tradição documental (analógico/digital/híbrido). A mesma leitura pode ser feita para os pares A4-A5, B4-B5, C4-C5, D4-D5, F4-F5, G4-G5 (ofícios). No caso das informações internas (A3, B3, C3, D3 e F3), verifica-se que se encontram em ramos isolados mas vinculados a um *cluster*, respetivamente A1-A2, B1-B2, C1-C2, D1-D2, F1-F2 e G1-G2, numa relação, por exemplo, de tipo “*RiC-R18-Record*” *is member of* “*RiC-R13-Record Set*”, que determina a formação de agregações de documentos. Refira-se, contudo, que os dendrogramas apresentados constituem fenogramas, *i. e.*, baseado apenas na similaridade morfológica entre os *taxa*, portanto não indiciador de filogenia (Khanafiah & Situngkir, 2006; Morrison, 2014).



As barras do cladograma correspondem ao somatório das propriedades RiC por *taxon* e desdobrados em 13 esferas coloridas na seguinte sequência (esquerda/direita): *Record, Record Component, Record Set, Agent, Occupation, Position, Function, Activity, Mandate, Documentary Form, Date, Place, Concept/Thing*.

**Figura 2. Dendrograma de relações entre *taxa* (tipologias documentais), processado em *iTOL* (Letunic & Bork, 2016)**

Se agregarmos os *taxa* num único dendrograma, é possível observar a constituição de dois ramos, visível na Figura 2. Em primeiro lugar, o *clado 1* representa a comunicação interinstitucional por meio de ofícios, verificando-se A6 como a correspondência recebida (em ramo isolado) e os restantes como correspondência expedida. No *clado 2*, verifica-se a preparação da comunicação interna e da informação de suporte. A título exemplificativo, o caso do par A4-A5 (300.30.602) aparece vinculado aos *taxa* D4-D5 (*idem*) e G4-G5 (300.30.008), constituindo um *continuum* que nasce da avaliação de informação sobre conjuntos informacionais acumulados, em que o auto de entrega e guia de remessa e os instrumentos de pesquisa partilham propriedades similares entre si. Igualmente, no *clado 2*, verificam-se relações de similaridade entre pares D1-D2 (300.30.602) e G1-G2 (300.30.008), pelo facto de o relatório de avaliação partilhar propriedades comuns aos instrumentos de pesquisa. Os pares B1-B2, C1-C2, H1-H2, por seu turno, derivam de procedimentos similares de eliminação de informação arquivística (300.10.302), independentemente se resultante de um relatório de avaliação ou da execução de uma portaria de gestão de documentos.

Estes resultados constituem, contudo, uma aproximação preliminar aos métodos filomeméticos, contando com aspetos positivos e com limitações:

- (i) O recurso a uma matriz binária permite-nos, em princípio, descrever a informação a partir da identificação de propriedades observáveis em tipologias documentais de arquivo. Tendo como exemplo a Figura 1, verifica-se a presença de relações hierárquicas entre os *taxa* (documentos simples), que poderão indiciar uma hipótese de relação de ordem física entre tipologias documentais obtida a partir da matriz binária. As propriedades *RiC*, contudo, não significam que sejam suficientes para estabelecer relações de filogenia. Conforme Figura 2, os *clusters* representam apenas (dis)similaridades morfológicas entre os *taxa*, mais próximo dos métodos fenéticos do que cladísticos, por o algoritmo UPGMA proporcionar este tipo de resultados (Khanafiah & Situngkir, 2006). A título de exemplo, autos de eliminação (H1-H2, B1-B2, C1-C2) e os autos de incorporação (F1-F2) dissemelham-se em relação ao *cluster* H3, B3, C3, I1, G3, A3, D3 (informações internas) e não se encontram vinculados entre si conforme a ordem “natural” ou “original” atribuída pelo produtor.
- (ii) Os ramos constituídos pelo algoritmo UPGMA revelam aspetos pertinentes sobre as relações funcionais entre os *taxa*. Thomer e Weber utilizaram algoritmos distintos (máxima verossimilhança) para identificar “eventos de especiação” a partir de metadados com fins de preservação digital (Thomer & Weber, 2014; Weber & Thomer, 2016). O recurso a outros algoritmos depende da granularidade informativa dos metadados e dos objetivos de análise filogenética, que terá resultados distintos em função da questão de investigação e da natureza do objeto de estudo (Yu & Gen, 2010). A principal limitação deriva do facto de a maior parte das aplicações informáticas para o estudo filogenético está desenhada apenas para processar informação sobre entidades bióticas.

## CONCLUSÃO

Este artigo efetuou uma aproximação preliminar aos métodos filomeméticos aplicados à classificação da informação arquivística. Procuramos responder a uma questão atual que afeta à organização do conhecimento em sistemas de informação arquivísticos, na ausência de outros modelos disponíveis: de que modo podemos aplicar modelos de classificação científica, de base evolucionista, em sistemas de classificação arquivísticos? Propusemos uma análise contextual em torno da “metáfora biológica” baseada na literatura científica. A questão fundamental consiste em determinar que metainformação, perante a diversidade de esquemas, poderá ser significativa para uma classificação filomemética da informação arquivística baseada no princípio de hereditariedade (Niu, 2013). Recorremos, assim, a uma amostra de documentos de arquivo a partir da análise das tipologias documentais baseada numa matriz binária de propriedades *RiC*. Obtivemos várias hipóteses de vínculos de (dis)similaridade entre as unidades informacionais de nível de documento (*taxa*). As potencialidades deste modelo de prospeção de informação, independentemente da sua materialização, permitirão analisar conjuntos informacionais arquivísticos sob uma perspetiva evolucionista em torno de ecossistemas informacionais distintos.

Várias técnicas de classificação continuam a ser exploradas *ad haustum* no âmbito da aglomeração de dados por inferência computacional (Janssens, Zhang, Moor, & Glänzel, 2009; Jung, Bae, & Liu, 2008). No entanto, os modelos baseados em algoritmos evolucionistas estão mais próximos dos princípios arquivísticos, como proveniência, organicidade e ordem original. Importa, por conseguinte, indagar as

potencialidades e as limitações sobre os algoritmos meta-heurísticos aplicáveis à prospeção e reconstrução evolucionista da informação arquivística (Ventura & Luna, 2016; Yu & Gen, 2010).

Em última análise, procurámos explorar uma forma distinta das abordagens interpretivistas (base funcional) e encetámos uma metodologia filomemética. Revela-se cada vez mais necessário explorar na arquivística um discurso mais computacional no âmbito da organização do conhecimento, por permitir ambientes controlados para o processamento e interpretação da informação.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Adell, J. I. (2010). Clasificación de documentos y clasificación en biología. *Tábula*, 13, 105–124.
- Andrews, T. L., & Macé, C. Trees of Texts – Models and methods for an updated theory of medieval text stemmatology | Digital Humanities 2012, Digital Humanities (2012). Retrieved from <http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/trees-of-texts-models-and-methods-for-an-updated-theory-of-medieval-text-stemmatology.1.html>
- Atkinson, Q. D. (2006). *From species to languages : a phylogenetic approach to human prehistory*. ResearchSpace@Auckland.
- Bak, G. (2012). Continuous classification: Capturing dynamic relationships among information resources. *Archival Science*. <https://doi.org/10.1007/s10502-012-9171-8>
- Cardona, G., Rosselló, F., & Valiente, G. (2008). Extended Newick: it is time for a standard representation of phylogenetic networks. *BMC Bioinformatics*, 9(532). <https://doi.org/doi.org/10.1186/1471-2105-9-532>
- Casanova, E. (1928). *Archivistica*. Siena: Stab. arti grafiche Lazzeri.
- Cook, T. (2011). We Are What We Keep; We Keep What We Are’: Archival Appraisal Past, Present and Future. *Journal of the Society of Archivists*, 32(2), 173–189. <https://doi.org/10.1080/00379816.2011.619688>
- Costa, F. de O., Lameri, S., Bestagini, P., Dias, Z., Rocha, A., Tagliasacchi, M., & Tubaro, S. (2015). Phylogeny reconstruction for misaligned and compressed video sequences. In *Image Processing (ICIP), 2015 IEEE International Conference on* (pp. 301–305).
- Currie, T. E., & Mace, R. (2011). Mode and tempo in the evolution of socio-political organization: reconciling “Darwinian” and “Spencerian” evolutionary approaches in anthropology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567), 1108–1117. <https://doi.org/10.1098/rstb.2010.0318>
- Darwin, C. R. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray.
- DGARQ. (2011). *Orientações para a Descrição Arquivística 3.<sup>a</sup> Versão*. Lisboa: DGARQ.

DGLAB. (2016). Projeto ASIA – documento em consulta pública [dataset]. Lisboa: DGLAB. Retrieved from [http://arquivos.dglab.gov.pt/wp-content/uploads/sites/16/2016/03/ASIA\\_divulgacao\\_2016-03-11.xlsx](http://arquivos.dglab.gov.pt/wp-content/uploads/sites/16/2016/03/ASIA_divulgacao_2016-03-11.xlsx)

Duranti, L. (2015). Archival bond. In L. Duranti & P. Franks (Eds.), *Encyclopedia of Archival Science* (Luciana Du, pp. 28–29). Rowman & Littlefield.

Eiben, Á. E., & Smith, J. E. (2012). Evolutionary Algorithms. In F. Neri, C. Cotta, & P. Moscato (Eds.), *Handbook of Memetic Algorithms* (pp. 9–27). Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-23247-3\\_2](https://doi.org/10.1007/978-3-642-23247-3_2)

Fangerau, H., Geisler, H., Halling, T., & Martin, W. (Eds.). (2013). *Classification and Evolution in Biology, Linguistics and the History of Science Concepts – Methods – Visualization*. Stuttgart: Franz Steiner Verlag.

Garcia-Vallve, S., Palau, J., & Romeu, A. (1999). Horizontal gene transfer in glycosyl hydrolases inferred from codon usage in *Escherichia coli* and *Bacillus subtilis*. *Molecular Biology and Evolution*, 16(9), 1125–1134.

García Rivero, D. (2010). Introducción a la teoría de la clasificación ya las escuelas taxonómicas (Fenética, Cladística y Taxonomía evolutiva). In et al. Escacena (Ed.), *Clasificación y Arqueología: Enfoques y métodos taxonómicos a la luz de la evolución darwiniana* (pp. 61–94). Sevilla: Universidad de Sevilla.

García Rivero, D. (2013). *Arqueología y Evolución: a la búsqueda de filogenias culturales*. Sevilla: Universidad de Sevilla.

García Rivero, D. (2016). Darwinian Archaeology and Cultural Phylogenetics (pp. 43–72). [https://doi.org/10.1007/978-3-319-25928-4\\_3](https://doi.org/10.1007/978-3-319-25928-4_3)

Gartner, R. (2016). *Metadata Shaping Knowledge from Antiquity to the Semantic Web*. Heidelberg: Springer. <https://doi.org/10.1007/978-3-319-40893-4>

Gnoli, C. (2006). Phylogenetic classification. *Knowledge Organization*, 33(3), 138–152.

Gontier, N. (2012). Applied Evolutionary Epistemology: a new methodology to enhance interdisciplinary research between the life and human sciences”. *Kairos: Revista de Filosofia & Ciência*, 4, 7–49. Retrieved from [http://www.academia.edu/download/30865351/Applied\\_Evolutionary\\_Epistemology.pdf](http://www.academia.edu/download/30865351/Applied_Evolutionary_Epistemology.pdf)

Gueguen, G., Fonseca, V. M. M. da, Pitti, D. V., Grimouard, C. S., Marques, M., Fonseca, D., ... Sibille-De Grimouard, C. (2013). Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives’ Experts Group on Archival Description. *The American Archivist*, 76(2), 566–583.

Ham, F. G. (1993). *Selecting and appraising archives and manuscripts*. Chicago, IL: Society of American Archivists.

- Heikkilä, T. (2007). The Evolution of Texts: Confronting Stemmatological and Genetical Methods - Research database Tuhat - University of Helsinki. *Variants - Journal of the European Society for Textual Scholarship*.
- Huang, A. (2008). Similarity Measures for Text Document Clustering. In *NZCSRSC 2008, April 2008, Christchurch, New Zealand* (pp. 49–56).
- Hurley, C. (1993). What, If Anything, Is A Function? *Archives & Manuscripts*, 21(2), 208–220.
- ICA. (2016). *RiC-CM-0.1: Records in Contexts: a conceptual model for archival description*. Paris: ICA.
- IP2. (2008). The InterPARES 2 Project Dictionary. Retrieved from [http://www.interpares.org/ip2/ip2\\_terminology\\_db.cfm](http://www.interpares.org/ip2/ip2_terminology_db.cfm)
- ISO. (2014). *ISO 21127:2014 - Information and documentation -- A reference ontology for the interchange of cultural heritage information*. Geneva: International Organization for Standardization.
- Janssens, F., Zhang, L., Moor, B. De, & Glänzel, W. (2009). Hybrid clustering for validation and improvement of subject-classification schemes. *Information Processing and Management*. <https://doi.org/10.1016/j.ipm.2009.06.003>
- Janzen, L. M. (1994). *Series: History, Theory and Practice*. University of British Columbia. <https://doi.org/10.14288/1.0087494>
- Jenkinson, H. (1922). *A manual of archive administration including the problems of war archives and archive making*. Oxford: The Clarendon Press.
- Jenkinson, H. (2003). *Selected writings of Sir Hilary Jenkinson*. Chicago, IL: Society of American Archivists.
- Jung, J. Y., Bae, J., & Liu, L. (2008). Hierarchical business process clustering. In *Proceedings - 2008 IEEE International Conference on Services Computing, SCC 2008*. <https://doi.org/10.1109/SCC.2008.69>
- Khanafiah, D., & Situngkir, H. (2006). Visualizing the Phylomemetic Tree: Innovation as Evolutionary Process. *Journal of Social Complexity*, 22(22), 20–30.
- Lantern, D., & Barili, A. (2017). Forensic analysis of deduplicated file systems. *Digital Investigation*, 20, S99–S106. <https://doi.org/10.1016/j.diin.2017.01.008>
- Lecointre, G. (2015). Descent (Filiation). In T. Heams, P. Huneman, G. Lecointre, & M. Silberstein (Eds.), *Handbook of Evolutionary Thinking in the Sciences* (pp. 159–207). Dordrecht: Springer Netherlands. [https://doi.org/10.1007/978-94-017-9014-7\\_9](https://doi.org/10.1007/978-94-017-9014-7_9)
- Lee, J., Jung, E., Yoon, J., Kim, S., & Jung, H. (2013). Evolutionary Patterns of an Artifact : The Mobile Phone, 1469–1478.
- Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, 1. <https://doi.org/10.1093/nar/gkw290>

- Macedo, L. S. A. de. (2017). Matriz binária RiC: classificação filogenética da informação arquivística. Harvard Dataverse. <https://doi.org/10.7910/DVN/ERE8HL>
- 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>
- Mahner, M., & Kary, M. (1997). What Exactly Are Genomes, Genotypes and Phenotypes? And What About Phenomes? *Journal of Theoretical Biology*, 186(1), 55–63. <https://doi.org/10.1006/jtbi.1996.0335>
- Mai, J.-E. (2011). The modernity of classification. *Journal of Documentation*, 67(5), 710–730. <https://doi.org/10.1108/00220411111145061>
- Mata Caravaca, M. (2015). The concept of archival “sedimentation”: its meaning and use in the Italian context. *Archival Science*, 1–12. <https://doi.org/10.1007/s10502-015-9256-2>
- McCarthy, I. P. (2005). Toward a Phylogenetic Reconstruction of Organizational Life. *Journal of Bioeconomics*, 7(3), 271–307. <https://doi.org/10.1007/s10818-005-5245-5>
- Mesoudi, A. (2011). *Cultural Evolution. How Darwinian Theory can Explain Human Culture & Synthesize the Social Sciences*. University of Chicago Press (Vol. 53). Chicago and London: University of Chicago Press. <https://doi.org/10.1017/CBO9781107415324.004>
- Mesoudi, A. (2016). Cultural Evolution: A Review of Theory, Findings and Controversies. *Evolutionary Biology*, 43, 481–497. <https://doi.org/10.1007/s11692-015-9320-0>
- Mokhtar, U. A., & Yusof, Z. M. (2015). Classification: The understudied concept. *International Journal of Information Management*. <https://doi.org/10.1016/j.ijinfomgt.2014.12.002>
- Morrison, D. A. (2014). Is the tree of life the best metaphor, model, or heuristic for phylogenetics? *Systematic Biology*, 63(4), 628–38. <https://doi.org/10.1093/sysbio/syu026>
- Muller, S., Feith, J. A., & Fruin, R. (1898). *Handleiding voor het ordenen en beschrijven van archieven: ontworpen in opdracht van de Vereeniging van Archivarissen in Nederland*. Groningen: Erven B. van der Kamp.
- Niu, J. (2013). Recordkeeping metadata and archival description: a revisit. *Archives and Manuscripts*, 41(3), 203–215. <https://doi.org/10.1080/01576895.2013.829751>
- O’Brien, M., & Lyman, R. (2003). *Cladistics and archaeology*. Salt Lake City: University of Utah Press.
- O’Brien, M., Lyman, R., Mesoudi, A., & VanPool, T. (2010). Cultural traits as units of analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 3797–3806. <https://doi.org/10.1098/rstb.2010.0012>
- Oliveira, A. A. de, Ferrara, P., De Rosa, A., Piva, A., Barni, M., Goldenstein, S., ... Rocha, A. (2016). Multiple parenting phylogeny relationships in digital images. *IEEE Transactions on Information Forensics and Security*, 11(2), 328–343.

- Pavlopoulos, G. A., Soldatos, T. G., Barbosa-Silva, A., & Schneider, R. (2010). A reference guide for tree analysis and visualization. *BioData Mining*, 3(1), 1. <https://doi.org/10.1186/1756-0381-3-1>
- Pearce-Moses, R. (2005). *A glossary of archival and records terminology*. Chicago, IL: Society of American Archivists.
- Rieppel, O. (2010). The series, the network, and the tree: changing metaphors of order in nature. *Biology & Philosophy*, 25(4), 475–496. <https://doi.org/10.1007/s10539-010-9216-4>
- Riley, J. (2010). *Glossary of Metadata Standards*. Indiana University Libraries White Professional Development Award.
- Rogers, C. (2015). Diplomats of born digital documents – considering documentary form in a digital environment. *Records Management Journal*, 25(1), 6–20. <https://doi.org/http://dx.doi.org/10.1108/RMJ-03-2014-0021>
- Rogers, C. (2016). A Literature Review of Authenticity of Records in Digital Systems: From “Machine-Readable” to Records in the Cloud. *Acervo*, 29, 16–44. Retrieved from <http://revista.arquivonacional.gov.br/index.php/revistaacervo/article/view/715>
- Simões, M. da G., & Freitas, M. C. V. de. (2013). A Classificação em Arquivos e em Bibliotecas à Luz da Teoria da Classificação: pontos de convergência e de divergência. *Ponto de Acesso*, 7(1), 81–115.
- Simões, M. da G., Freitas, M. C. V. de, & Rodríguez-Bravo, B. (2016). Theory of Classification and Classification in Libraries and Archives: Convergences and Divergences. *Knowl. Org*, 43(7), 530–539.
- Thomer, A. K., & Weber, N. M. (2014). The phylogeny of a dataset. In *Proceedings of the ASIST Annual Meeting*. <https://doi.org/10.1002/meet.2014.14505101064>
- Ventura, S., & Luna, J. M. (2016). *Pattern Mining with Evolutionary Algorithms*. <https://doi.org/10.1007/978-3-319-33858-3>
- Weber, N. M. (2015). *A framework for analyzing the sustainability of peer produced science commons*. University of Illinois at Urbana-Champaign. Retrieved from <https://www.ideals.illinois.edu/bitstream/handle/2142/88062/WEBER-DISSERTATION-2015.pdf?sequence=1>
- Weber, N. M., & Thomer, A. K. (2016). Modeling Cultural Evolution with Metadata Collections: A Phylomemetic Approach. In *International Conference on Dublin Core and Metadata Applications DC-2016, Copenhagen, Denmark*. Copenhagen.
- Wilkins, J. S. (2009). *Species: A History of the Idea (Species and Systematics)*. Berkeley and Los Angeles, CA: University of California Press.
- Yu, X., & Gen, M. (2010). *Introduction to evolutionary algorithms*. London: Springer. <https://doi.org/10.1007/978-1-84996-129-5>