



Journal of the Text Encoding Initiative

Issue 8 | 2014

Selected Papers from the 2013 TEI Conference

TEI4LdoD: Textual Encoding and Social Editing in Web 2.0 Environments

António Rito Silva and Manuel Portela



Publisher
TEI Consortium

Electronic version

URL: <http://jtei.revues.org/1171>

ISSN: 2162-5603

Electronic reference

António Rito Silva and Manuel Portela, « TEI4LdoD: Textual Encoding and Social Editing in Web 2.0 Environments », *Journal of the Text Encoding Initiative* [Online], Issue 8 | December 2014 - December 2015, Online since 01 January 2015, connection on 29 November 2016. URL : <http://jtei.revues.org/1171> ; DOI : 10.4000/jtei.1171

The text is a facsimile of the print edition.

TEI Consortium 2015 (Creative Commons Attribution-NoDerivs 3.0 Unported License)

TEI4LdoD: Textual Encoding and Social Editing in Web 2.0 Environments

António Rito Silva and Manuel Portela

AUTHOR'S NOTE

We would like to thank Timothy Thompson for his contributions to the TEI template for *LdoD*.

This work was supported by Portuguese national funds through FCT – Fundação para a Ciência e a Tecnologia, under projects PTDC/CLE-LLI/118713/2010 and UID/CEC/50021/2013.

1. Introduction

- 1 Fernando Pessoa's *Book of Disquiet* (*Livro do Desassossego*, abbreviated *LdoD*) is an unfinished book project. Pessoa wrote more than five hundred texts meant for this work between 1913 and 1935, the year of his death. The first edition of *LdoD* was not published until 1982, and another three major versions have been published since then (1990, 1998, 2010), most of which have been further revised. As it exists today, *LdoD* may be characterized as (1) a set of autograph (manuscript and typescript) fragments, (2) mostly unpublished at the time of Pessoa's death, which have been (3) transcribed, selected, and organized into four different editions, implying (4) various

interpretations of what constitutes this book according to the extant print editions. Editions show four major types of variation: in readings of particular passages, in selection of fragments, in their ordering, and also in heteronym attribution.¹

- 2 The goal of the *LdoD Archive*² is twofold: on the one hand, we want to provide a “standard” archive where experts can study and compare *LdoD*’s authorial witnesses to their different editions, and each one of these to each other; on the other hand, we want to design a virtual archive that allows both experts and non-experts to experiment with the production of different editions of *LdoD*, and also with the writing of their own fragments based on *LdoD*’s original fragments. Therefore, this latter goal, which is built on top of the genetic and critical archival goal, extends a scholarly understanding of *LdoD* as both authorial project and editorial construct to a new perspective of *LdoD* as an exploratory environment for individual and/or community editing and writing based on authorial witnesses and editorial versions.
- 3 The rationale for allowing experts and non-experts to experiment with re-editing and rewriting this work originates in social editing theories and in the increasing appeal of the *Book of Disquiet* for a general and global audience.³ We want to explore the collaborative dimension of the Web as a reading and writing space in the context of a digital archive in ways that enhance its scholarly, pedagogical, ludic, and expressive uses. Through this double strategy, we also expect to build a tool for investigating the relation between writing processes and material and conceptual notions of the book. By allowing users to perform various roles within the *LdoD Archive*, we suggest that one-way flow of text from the expert edition to the novice-remixed digital edition can be one way of handling questions of authority concerning the work’s textual form. We argue that it is possible to create a textual environment where professional and general readers can interact, and that such an environment extends theories about the social nature of textual production to current social media technologies.
- 4 The novelty of this project is built on the possibility of dynamically interacting with the repository. Therefore, we intend to provide to end users, either TEI experts or non-experts, a web interface through which they can interact with the *LdoD* repository to create their own interpretations of Pessoa’s work. To achieve the project’s goal we designed and implemented a collaborative archive where experts and non-experts can interact, while preserving the relative autonomy of their interventions. The archive enables the coexistence of three different dimensions of *LdoD*: genetic,

social, and virtual. This coexistence is supported by a Web 2.0 environment that integrates an object-oriented database with a file repository for TEI-encoded files. Web 2.0 interactions are done through the object-oriented database while experts continue to use their TEI XML editor of choice—which stores TEI files in the file repository—to interact with the environment.

- 5 In this article we focus on two aspects: (1) how to encode *LdoD*, and (2) how to support a dynamic environment. For (1) we will show how the existing witnesses, both authorial and editorial, can be encoded in TEI markup and how the proposed encoding structure will support dynamic interactions with the book, such as, for instance, the creation of further editions.⁴ Regarding (2), we will propose a software architecture that supports the traditional process of TEI encoding (Vanhoutte 2006; TEI Consortium 2012; Barney 2012; Earhart 2012) with the possibility of dynamically extending the interpretations of *LdoD* on top of a TEI representation.
- 6 In section 2 we briefly refer to related work. Section 3 presents the aims and rationale of the project. Section 4 states the problem, and in section 5 we present our solution strategy. Sections 6 and 7 describe our approach as far as, respectively, the TEI encoding and the software architecture are concerned.

2. Related Work

- 7 Our proposal explores current approaches to editing in electronic environments and attempts to integrate them with TEI conceptual and processing models.
- 8 In a recent article on the theory of digital editions, Peter Robinson shows how printed editions have traditionally focused on the relationship of text to work, while digital editions have been more focused on the relationship of text to document. He suggests that “a scholarly edition must, so far as it can, illuminate both aspects of the text, both text-as-work and text-as-document” (Robinson 2013, 123), and he calls attention to the presence of reading in any production of text through acts of transcription. The *LdoD Archive* embodies a similar understanding of the nature of textual semiosis as a process involving self and object in a continuous and co-dependent process of meaning production through acts of reading. Editing Pessoa’s centrifugal and reticular body of unpublished work—a work that is constantly relocating its centre in a network of ever-expanding writing projects—is an especially acute experience of the productive function of reading in activating the material and interpretative fields that allow us to

move back and forth from document to text to book to work. To the extent that each text of each edition is contextualizable in an archive of authorial and editorial witnesses, it is the very process of construction of text from document and edition from text that the genetic and social dimensions of the *LdoD Archive* place in evidence. Our conceptual model reframes established digital archive practices by using the dynamic and social affordances of electronic space itself for simulating literary processes.

- 9 The object representation of transcriptions is related to the work on data structure for representing multi-version objects (Schmidt and Colomb 2009) because we also parse the TEI files into a customized data structure, in our case a persistent object model that is manipulated in memory to avoid round trips to the databases. We emphasize the need identified by Schlitz and Bodine (2009) to have a clear separation between content and presentation in order to simplify and empower presentation tools. With regard to a Web 2.0 for digital humanities, we are indebted to proposals on cooperative annotations by Tummarello, Morbidoni, and Pierazzo (2005) and the advantages and vision of Web 2.0 and collaboration discussed in Benel and Lejeune (2009), Fraistat and Jones (2009), and Siemens et al. (2010). More recent research work raises the need to have several views of the encoding (Brüning, Henzel, and Pravida 2013). In our approach, different views are also helpful for interoperability (interchange) and to simplify the implementation of user interfaces. The work of Wittern (2013) stresses the need to allow dynamic edition of texts and management of versions, while Muñoz, Viglianti, and Fraistat (2013) highlight the challenges of participatory and distributed encoding in the context of the recently launched Shelley-Godwin Archive.
- 10 One of the differences related to other existing TEI projects is that in the *LdoD* project, TEI is envisioned more as an interchange mechanism (Bauman 2011), through the export functionalities, than as the standard for a fully interoperable service-oriented architecture. As a consequence, we can experiment with technologies that are not usually used in TEI environments. To do this we had to implement TEI import and export functionalities to allow the manipulation of the TEI-encoded information in an object-oriented database. Additionally, the export functionality allows backup in a safe repository in a human-readable format. On the other hand, there is information in the

object-oriented database that is specific to a Web 2.0 environment, like the information required to support access control policies, and which is not explicitly encoded using TEI because it depends on the dynamic aspects of the Web 2.0 tools.

3. *LdoD* Project

- 11 The main goal of the project is to create a dynamic literary archive: *LdoD Archive—Collaborative Digital Archive of the Book of Disquiet*. This archive will aggregate digital facsimiles of all documentary materials of the *Livro do Desassossego* and loosely topographic textual transcriptions of these materials. Our topographic transcription represents four types of spatial marks in the autographs: line breaks (using the <lb> element), spacing between paragraphs (using the <space> element), dividing rulers (using the <pb> element), and revision sites (using the @place attribute on <add>). The main goal of the topographic transcription is to facilitate the side-by-side reading of the facsimile and its transcription. The archive will also include textual representations of the four major Portuguese editions published between 1982 and 2013, respectively by Jacinto do Prado Coelho ([1982]), Teresa Sobral Cunha ([1990–91] 2008), Richard Zenith ([1998] 2012), and Jerónimo Pizarro ([2010] 2013). At this level the archive combines a genetic edition with a social text edition of *LdoD*, showing this work both as a network of multiple authorial intentions and as a conjectural construction of its successive editors. Readers will engage with the genetic edition when they see Pessoa’s acts of revision in each fragment as well as his different partial plans for the *Book of Disquiet*. Future annotations will also mark intertextual relations of fragments to Pessoa’s readings, including references to passages, authors, and marginalia in his own personal library. Readers will engage with the social text edition in two ways: first, by seeing actual historical instances of the *Book of Disquiet* as it has been edited in the four versions selected for inclusion in the archive, and in our own XML-encoded topographic transcription; and second, by creating new virtual instances of the *Book of Disquiet* understood as a particular selection and organization of fragments. Tools for textual analysis, collation, and annotation will represent the dynamics of the acts of writing and editing in the production and reproduction of *LdoD*.
- 12 Although the extension and detail of the critical apparatus varies considerably in Coelho, Cunha, Zenith, and Pizarro, all of those editions make explicit the interpretative criteria used for selecting and ordering the pieces of text. Our digital representation of the dynamics of textual and

bibliographic variation depends on both XML encoding of variation sites (deletions, additions, substitutions, etc.) and metatextual information concerning authorial and editorial witnesses (such as date, order, or heteronym). While TEI markup may be considered as a particular kind of critical apparatus on its own, it is through visualization tools that users will be able to critically engage with the dynamics of variation in authorial and editorial witnesses. They will be able to examine not only the differences between autograph sources or between manuscripts and print editions, but also the differences among the various print editions.

- 13 Besides using TEI XML encoding and programming to recreate the history of the authorial and editorial dynamics, the *LdoD Archive* also explores the simulative potential of the digital medium as a space for virtualizing the *Book of Disquiet* in ways that will enable users to experiment with the processes of editing and writing in relation to this work. Expert and non-expert users will be able to collaborate by making their own editions and by adding further textual fragments to the archive. Interventions can take two forms: selecting, ordering, and annotating fragments as part of a user-defined virtual edition; or selecting fragments or parts of fragments as textual seeds for creating variations and extensions based on *LdoD* as part of a user-defined virtual writing process. This interactive feature of the archive is further enhanced by search and navigation functions that will allow a strong integration between the initial closed set of scholarly materials and the open virtual editing and writing additions.
- 14 The participatory affordance of the digital medium has two major aspects: an environment for collaboration and social interaction, on the one hand, and the possibility of marking material changes at the level of code, on the other. Material changes can be marked up in the XML encoding, but also as new data and metadata generated by the users' interaction with the archive, which are stored in the database. We believe that these features of networked computational media can be used to redesign the digital archive as a dynamic environment for different kinds of practice, not limited to research and teaching. A scholarly remediation of *LdoD* according to tested principles of electronic philology would be part of a larger interactive environment where reading and writing practices around the *Book of Disquiet* could be socialized within the digital medium itself.⁵ Aggregation of genetic and editorial witnesses according to criteria defined by readers could also

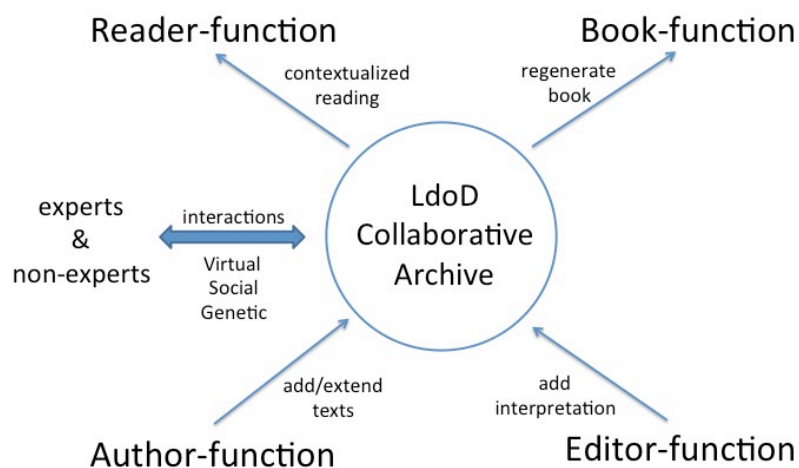
occur within a virtual space that allows users to make critical annotations and write variations based on the fragments. Thus knowledge of textual form and textual transmission would be complemented by experiments with writing processes and bibliographic structures.

4. Problem: The LdoD Archive Model

- 15 In this section we offer a synthesis of the four functions and three dimensions of our model for a virtual *LdoD*, as described in Portela and Silva (2014). This description has been diagrammed in figure 1. Our virtual model for *LdoD* establishes a framework of interactions distributed across four functions: reader-function, editor-function, book-function, and author-function. These functions contain a model of the ecosystem of literary inscriptions. With regard to facsimile surrogates and textual transcriptions, seen here as digital objects available for interaction with expert and non-expert users, we may say that the archive contains three related dimensions: a genetic dimension that allows users to create a narrative of authorial composition; a social dimension that allows users to create a narrative of scholarly editing and textual reception; and a virtual dimension that enables users to explore the possibilities of both writing and editing while interacting with the *Book of Disquiet*.
- 16 We should clarify here that “social dimension” is used in this article to refer to the socialization of texts embodied in particular textual and bibliographic forms in the historical archive—a notion derived from social editing theories (McGann 2006). The “virtual dimension” refers to the collaborative additions and interactive transformations of the archive in a web environment. Thus the “virtual dimension” should be understood as a particular expression of the “social dimension” in this constructed electronic reading, editing, and writing space, which can be freely used by either Pessoa experts or non-experts. The final relevance of each virtual edition will result from its popularity—how often it is accessed and used in the construction of other virtual editions. “Social editing” in the title of this article subsumes both processes of textual socialization, that is, our digital representation of the historical socialization of Pessoa’s *LdoD* in a set of expert editions (“social dimension”), and our design of a digital platform for furthering textual collaborations (“virtual dimension”). The social/virtual distinction is pertinent from both

a design and a theoretical perspective, since we need to distinguish the static and dynamic layers of the archive. On the other hand, we also want to emphasize the social nature of editing processes through a series of built-in affordances in our digital archive.

Figure 1: A functional model of the *LdoD* Archive



- 17 The reader-function supports a contextualized reading of the *LdoD* fragments. It enables readers to visualize and compare fragments and variations according to several authorial and editorial witnesses. The book-function enables the construction and reconstruction of *LdoD* based on textual and metatextual information in the archive. The output of these flexible remakings of the *LdoD* can be either a given form that already exists in the work's archive, such as the Coelho 1982 or the Pizarro 2010 edition, or a virtual reorganization of the book. The editor-function, in its turn, is meant to provide interpretations of *LdoD* as a book project based on a pre-existing (but variable and unstable) corpus of documents. This function allows users to produce textual sequences and textual aggregations according to user-defined criteria. It also enables users to add annotations and tags to particular passages. Finally, the author-function enables the extension of *LdoD* with new texts based on original fragments from *LdoD*. Thus, a reader of *LdoD* is able to write a new

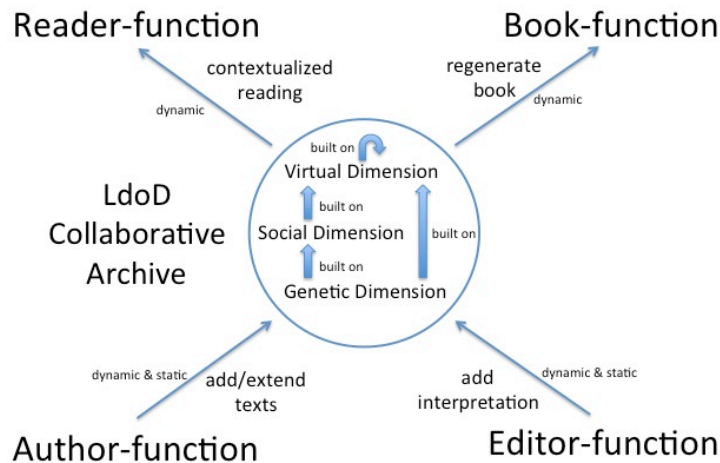
text derived from fragments of *LdoD*, becoming an author in the context of a virtual edition. These newly authored fragments can result both from textual remixes of the archive's content, and from textual additions of user-created content.

- 18 We can conclude that what is needed is (1) a scholarly archive where experts can study and compare *LdoD*'s authorial witnesses with their critical editions; and (2) a virtual archive that allows experts and non-experts to experiment with the production of different editions of *LdoD*, by means of (2a) editing (aggregating, sequencing, annotating, tagging) and (2b) writing (extensions and variations based on Pessoa's texts).
- 19 Given the above set of goals, the *LdoD Archive* has to accommodate scholarly standards and requirements concerning digital archives, for instance the use of TEI as a specification to encode literary texts, and the virtual communities and social software features to support the social edition of *LdoD* by both other experts and non-experts. This aspect implies additional constraints on how to coordinate the interactions within the archive. Therefore, there is the need for a dynamic archive where users can edit their own versions of *LdoD*, and write extensions of the original fragments, while the archive's initial set of experts' interpretations and analyses of *LdoD* are kept "unchanged" and clearly separated from the socialized virtual editions and writings (Silva and Portela 2013). This constraint will actually shape the formulation of the solution strategy.

5. Solution Strategy: A Collaborative Archive Framework

- 20 To address the separation of the experts' interpretations from the socialized virtual editions and writings we establish the following principles:
- Definition of Expert and Virtual (Social) Editions,⁶ to explicitly separate the experts' interpretations, represented in expert editions, from the socializations of the book, represented as virtual editions;
 - Coexistence and Separation of Editions, to enable users to use the experts' editions to build their virtual editions without interfering with expert interpretations. This is achieved by imposing a set of restrictions:
 - Virtual editions are built on top of other expert and virtual editions, and thus use their interpretations;

- Expert editions do not refer to virtual editions, in order to keep these experts' interpretations separate from the virtual editions' extensions;
 - By default only the expert editions are presented, so as to preserve an "official" experts' archive, which means that users have to explicitly access the virtual editions. This explicit access makes them aware of the existence, and separation, of experts' and socialized virtual interpretations.
- 21 These principles are illustrated in [figure 2](#). In the center we show how the archive repository is built using three layers (the three aforementioned dimensions), and their recursive construction. The original set of expert editions (that is, our own topographic transcription and the editions by Coelho, Cunha, Zenith, and Pizarro) instantiate the genetic and social dimensions while virtual editions are built on them and on other virtual editions. As we can see, although the virtual level of the archive is separate from its scholarly level, the virtual dimension is built on the genetic and on the social dimensions. For instance, when reading takes place, it is necessary to show what kind of edition is being read: the project's topographic transcription of an authorial document (genetic view), one of the four expert editorial versions that are pre-encoded in the archive (social view), or one of an open set of virtual editions (virtual view). Each textual unit of any virtual edition, in its turn, will contain the stemma of its sources, including authorial witnesses, critical editions, and other virtual editions.

Figure 2: Dynamic and static aspects in the *LdoD Archive*.

- 22 Additionally, figure 2 also enriches the representation of the four functions in figure 1 by describing to what extent they support static and dynamic manipulations of the archive repository. In contrast with most digital literary archives, which are built using XSLT technology to transform TEI representations into HTML for visualization, and where only the reader- and book-functions are dynamic, in the *LdoD Archive* we allow users to create their own editorial interpretations of the *LdoD*, and extend *LdoD* written fragments using a web interface. Therefore, users can dynamically change the archive repository through its virtual dimension. However, we still continue to support the traditional scholarly work on the genetic and social dimensions where the project editors do static offline TEI encoding of the *LdoD* using their XML editor of choice.

6. TEI Encoding

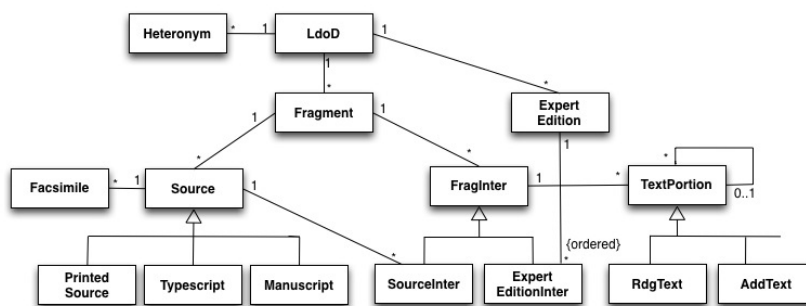
- 23 Because of the dynamic requirements of *LdoD* project, and the strategy to create a collaborative archive as described above, the TEI encoding needs to address the three dimensions. Actually, while the description of the TEI encoding of the genetic and social dimensions is driven by how TEI can be used to express the authorial witnesses and their expert interpretations through the

concept of edition, the encoding of the virtual dimension focuses on how the TEI encoding chosen for the genetic and social dimensions can be used to support the creation of a non-predefined number of virtual editions.

- 24 In Portela and Silva (2014) we describe a UML (Unified Modeling Language) model for a virtual *LdoD* that supports the project requirements. In this section we show how this model is encoded in TEI.

6.1 Genetic and Social Dimensions

Figure 3: Model of the genetic and social interpretations of *LdoD*.



- 25 Figure 3 shows a UML model of the genetic and social interpretations of *LdoD*. The root concept, which has a single instance, is *LdoD*, and it contains a set of *Fragment* instances and several interpretations (*FragInter* instances) for each fragment. Interpretation is the name we give in the model to refer to a textual instance of a fragment according to one of the sources, authorial or editorial, thus stressing transcription as an interpretive act. These interpretations are encoded in TEI using the `<app>` element with an `<rdg>` element for each different transcription. There are two types of interpretations, authorial and editorial. Authorial interpretations, instances of *SourceInter*, refer to their *Source* instance, which can be a *PrintedSource*, *Typescript*, or *Manuscript*. On the other hand, editorial interpretations, *ExpertEditionInter* instances, are contained, and ordered, in the context of expert editions (*ExpertEdition* instances). Currently, the project is encoding four *ExpertEdition* instances. Finally, an interpretation contains several *TextPortions* that represent the use of TEI elements in the fragments' transcriptions. *TextPortions* abstract elements which constitute the transcription with the objective of capturing their variations.

- 26 To encode this model in TEI we use the `<teiCorpus>` element to contain all the fragments and aggregate in its header the entities that are common, and can be reused, in each of the fragments. The encoding below shows a part of the TEI Corpus header where the different editions are encoded inside a `<sourceDesc>` element as a list of `<bibl>` elements. The TEI encoding of Pessoa's heteronyms is done within `<profileDesc>` and `<particDesc>` elements as a list of `<person>` elements with attribute `@type` set to the "heteronyms" value. In the example, the four experts' editions and the two heteronyms assigned by experts to the *LdoD* are declared.

```

<teiHeader type="corpus">
  <fileDesc>
    <titleStmt>...</titleStmt>
    <sourceDesc>
      <listBibl xml:id="ED.CRIT">
        <bibl xml:id="ED.CRIT.C">...</bibl>
        <bibl xml:id="ED.CRIT.SC">...</bibl>
        <bibl xml:id="ED.CRIT.Z">...</bibl>
        <bibl xml:id="ED.CRIT.P">...</bibl>
      </listBibl>
    </sourceDesc>
  </fileDesc>
  <profileDesc>
    <particDesc>
      <person xml:id="FP">
        <persName type="orthonym">...</persName>
      </person>
      <listPerson type="heteronyms" xml:id="HT">
        <person xml:id="HT.BS">...</person>
        <person xml:id="HT.VG">...</person>
      </listPerson>
    </particDesc>
  </profileDesc>
</teiHeader>

```

- 27 For each fragment we use a `<TEI>` element inside the global *LdoD* `<teiCorpus>` element. This `<TEI>` element contains a `<teiHeader type="text">` element where the different interpretations are declared; several `<facsimile>` elements to declare the fragment's facsimiles; and a `<text>` element

which contains the interpretations' transcriptions, represented in [figure 3](#) by `TextPortion` instances. The encoding example below presents part of the structure of the header for a fragment selected for the four expert editions that has two manuscripts and one printed source.

```

<teiHeader type="text">
  <fileDesc>
    <titleStmt>...</titleStmt>
    <publicationStmt>...</publicationStmt>
    <sourceDesc>
      <listBibl xml:id="Fr1.SRC">
        <listBibl xml:id="Fr1.SRC.MS">
          <msDesc xml:id="Fr1.SRC.MS.Fr1a">...</msDesc>
          <msDesc xml:id="Fr1.SRC.MS.Fr1b">...</msDesc>
        </listBibl>
        <listBibl xml:id="Fr1.SRC.ED">
          <bibl xml:id="Fr1.SRC.ED.Fr1c">...</bibl>
        </listBibl>
      </listBibl>
      <listWit xml:id="Fr1.WIT">
        <listWit xml:id="Fr1.WIT.MS">
          <witness xml:id="Fr1.WIT.MS.Fr1a.274">
            <ref target="#Fr1.SRC.MS.Fr1a"/>
          </witness>
          <witness xml:id="Fr1.WIT.MS.Fr1b.273">
            <ref target="#Fr1.SRC.MS.Fr1b"/>
          </witness>
        </listWit>
        <listWit xml:id="Fr1.WIT.ED">
          <listWit xml:id="Fr1.WIT.ED.ORIG">
            <witness xml:id="Fr1.WIT.ED.ORIG.Fr1c.405">
              <ref target="#Fr1.SRC.ED.Fr1c"/>
            </witness>
          </listWit>
          <!-- Bibliographic info about the edition -->
            </witness>
          </listWit>
          <listWit xml:id="Fr1.WIT.ED.CRIT">
            <witness xml:id="Fr1.WIT.ED.CRIT.C">
              <ref target="#ED.CRIT.C"/>
            </witness>
          </listWit>
          <!--Info about the edition interpretation -->
            </witness>
          </listWit>
        </listWit>
      </listWit>
    </sourceDesc>
  </fileDesc>
</teiHeader>

```



```

    <witness xml:id="Fr1.WIT.ED.CRIT.SC">
      <ref target="#ED.CRIT.SC"/>

<!--Info about the edition interpretation -->
    </witness>
    <witness xml:id="Fr1.WIT.ED.CRIT.Z">
      <ref target="#ED.CRIT.Z"/>

<!--Info about the edition interpretation -->
    </witness>
    <witness xml:id="Fr1.WIT.ED.CRIT.P">
      <ref target="#ED.CRIT.P"/>

<!--Info about the edition interpretation -->
    </witness>
  </listWit>
</listWit>
</listWit>
</sourceDesc>
</fileDesc>
</teiHeader>

```

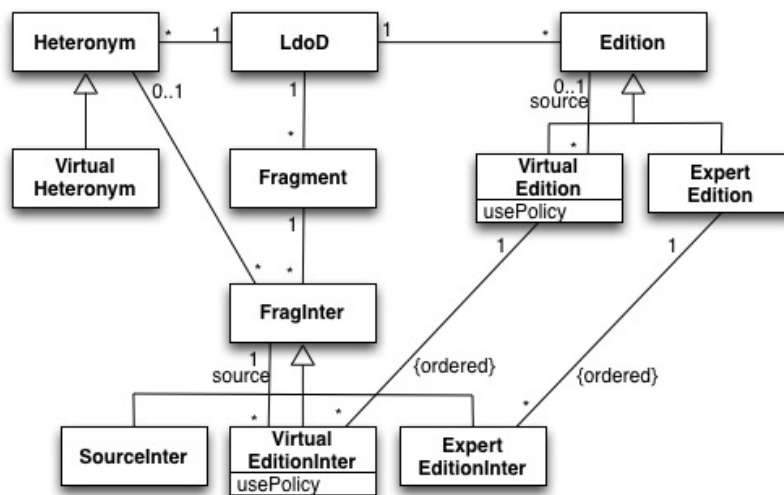
- 28 The Source concept is implemented within the `<sourceDesc>` element and the distinction between Manuscript, Typescript, and PrintedSource is indicated by convention through the structure of the `@xml:id` attribute value. Note that Manuscript and Typescript are implemented within the `<msDesc>` element, while PrintedSource is used in the `<bibl>` element. On the other hand, the FragInter concept is implemented by the `<witness>` element and the distinction between editorial (ExpertEditionInter) and authorial (SourceInter) witnesses is also indicated by convention through the structure of the `@xml:id` attribute value. A `<ref>` element is used to associate the witness with its source, for SourceInter witnesses, or with its edition, for ExpertEditionInter witnesses. The editions are declared in the corpus header. The editorial contextual information of the fragment (metatextual information) is encoded within the `<witness>` element. Finally, the TextPortion instances depicted in [figure 3](#) are represented within the `<text>` element and refer to their respective FragInter through the `@wit` attribute of `<rdg>` elements, which contain the witness identifier declared in the fragment header.⁷ As mentioned

above, an apparatus is used to distinguish the different readings of the fragments. The <rdg> element is useful in the context of the editions to represent variations in their readings of the source documents.

- 29 This approach allows us to associate interpretation metadata in the context of each witness. Users will be able to compare digital facsimile representations of authorial documents (and topographic transcriptions of those documents) to editorial transcriptions. The latter can also be compared against each other in order to highlight their interpretations of the source.
- 30 When a TEI-encoded file for a *LdoD* fragment is uploaded to the system, it is parsed, and if it does not contain any errors, a new Fragment instance is created associated with a new instance of FragInter for each different transcription of the text. In addition to the verification of the TEI syntax, the parser does a semantic verification of the encoded fragment. For instance, it verifies the existence of the entities referenced by @xml:id attributes. This supports the encoding work because it allows the early detection of errors.

6.2 Virtual Dimension

Figure 4: Model of the virtual interpretations of *LdoD*.



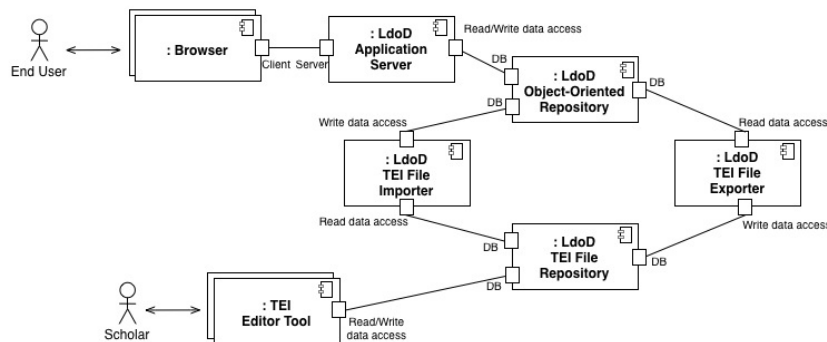
- 31 [Figure 4](#) describes how the model for the genetic and social interpretations is enriched to support virtual interpretations. Three new concepts are introduced: `VirtualEdition`, which represents virtual editions; `VirtualEditionInter`, which represents the interpretations of fragments in the context of virtual editions; and `VirtualHeteronym`, which represents the new heteronyms that end users may create.
- 32 These three new concepts implement the principle of the separation of expert and virtual editions. To support the coexistence of expert and virtual editions, we define an optional association between a `VirtualEdition` and an `Edition` and a mandatory association between a `VirtualEditionInter` and a `FragInter`. The former association means that all the interpretations of the virtual edition use the interpretations of the source edition. Therefore, this association is optional at the edition level because we allow the creation of a virtual edition that uses interpretations of fragments from different source editions. The association between a `VirtualEditionInter` and a `FragInter` is mandatory because virtual interpretations of fragments should be based on an existing interpretation—either another virtual interpretation or an authorial or editorial interpretation. Ultimately, the transcriptions in a virtual interpretation are identical to either an authorial or an editorial transcription. Note that a virtual interpretation comprises one transcription, either authorial or editorial, and a set of tags and annotations on it.
- 33 The semantics of this relationship is defined by the `@usePolicy` attribute that can take two values: "import" and "inherit". When the "import" value is used, the virtual interpretation of the fragment is built on top of the source interpretation and can change it. Any further change done in the source interpretation does not impact on the virtual interpretation. The source interpretation is copied, at virtual interpretation creation time, to the virtual interpretation context and can be freely changed by the user. When the "inherit" policy is used, the source interpretation is extended in the virtual interpretation and cannot be changed. For example, if the source interpretation is changed by the source community, then the changes are propagated to its virtual interpretations. These policies only apply to the tags and annotations associated with the virtual edition, because, currently, the archive does not allow any changes to the authorial and editorial transcriptions.

- 34 The TEI encoding of the genetic and social interpretations is extended to support the virtual interpretations. Therefore, the virtual editions are declared in the corpus header as a new `<listBibl>` element containing a `<bibl>` element for each new virtual edition. However, the distinction between the set of critical editions and the set of virtual editions is done by convention through the `@xml:id` attribute value. Virtual heteronyms are similarly declared in the corpus header. As regards virtual interpretations, they are encoded in the fragment header using `<witness>` elements and a convention based on the `@xml:id` attribute value to distinguish them from the other editorial interpretations, namely the experts' interpretations. Additionally, the editorial contextual information (metatextual information) of the virtual fragment contains, among other information, a reference to the source interpretation and the fragment order in the virtual edition.
- 35 Following the proposed TEI encoding we are able to implement the solution principles for the *LdoD* project identified in [section 4](#). However, some of the distinctions associated with the separation principle, or the references between virtual interpretations and their source interpretations, are set by convention and require automatic tools to be aware of them. On the other hand, TEI does not support the encoding of some information that is required in a Web 2.0 application, namely access control information. For instance, we cannot express in TEI whether a virtual edition is private or public, and which specific users may access the virtual edition. We intend, as one of the final results of the project, to define a TEI customization that accommodates all the identified aspects that cannot be expressed using the TEI core.
- 36 There are some other open issues. The main problem is related to the dynamic evolution of the archive in terms of Web 2.0 requirements: how can TEI code be changed as a result of users' interactions with the archive? Note that the traditional approach to encoding in TEI is done statically, through tools like oXygen. However, we want to support the evolution of *LdoD* as a continuously reeditable and rewritable book. This means that it is necessary to enable the dynamic addition of new virtual editions and heteronyms in the Corpus and the addition of new fragments that extend the original ones. Additionally, users can define their own interpretation of any of *LdoD*'s fragments, for instance by using tags, which results in the generation of new editions of the book through the execution of a categorization algorithm. This open issue is addressed by the software architecture we propose in the [next section](#).

7. Architecture

- 37 Most digital scholarly archives are static. By static we mean that the construction of the archive is separated from its use. The former is done using TEI and XML editors, and the latter is supported by XSLT transformations. This software architectural approach is not feasible if we want to provide Web 2.0 functionality to the archive. Since we do not want to disregard the existing practice of encoding in TEI by expert users, the architecture needs to support the traditional encoding in TEI by the experts while enabling dynamic user interactions with the platform, as highlighted in the static and dynamic aspects of figure 2.
- 38 Figure 5 presents a component-and-connector view of the *LdoD Archive* software architecture using the repository and the client-server architectural styles. Components are depicted using UML component instances; connectors are represented by linking UML component ports. The port roles are labelled according to the architectural style that the connector implements: that is, the roles between the `:Browser` and `:LdoD Application Server` components denote the client-server architectural style.

Figure 5: *LdoD Archive* Software Architecture.



- 39 Traditional TEI encoding, where scholars use an XML editor of choice statically, is represented in figure 5 by the interactions between components `:TEI Editor Tool` and `:LdoD TEI File Repository`. However, instead of supporting the presentation through XSLT transformation of files in the repository, we provide an importer component (`:LdoD TEI File Importer`) that loads the TEI-encoded files into an object-oriented database (`:LdoD Object-Oriented Repository`). This database contains the object model described in Portela and Silva (2014). For the support

of Web 2.0 interactions, the `:LdoD Application Server` component provides a web client-server interface that users, experts and non-experts alike, use to interact with the object-oriented repository.⁸ Therefore, the changes occur on the object-oriented representation of the TEI-encoded files. Finally, in order to preserve the TEI interchange qualities of the archive, the `:LdoD TEI File Export` component allows the regeneration of TEI-encoded files from the object-oriented repository. Note that this component allows the selection of which parts of the repository to generate and the strategy of the encoding. For instance, it is possible to choose only a subset of the interpretations, and decide between different methods for linking critical apparatus to the text. Since the dynamic write interactions can only occur in the context of a virtual edition, it is possible to export the original TEI-encoded files as they were imported, except for their formatting, which is not stored when files are uploaded and parsed.

- 40 When comparing the software architecture in [figure 5](#) with [figure 2](#), we see that the static aspects of the author and editor functions occur through the `:TEI Editor Tool` component and all the other dynamic aspects occur through the `:Browser` component.
- 41 The key point of the proposed architecture is the use of an object domain model to represent the *LdoD* archive. Using this approach we, at first, transform *LdoD* encoded in TEI to the object model, and allow the visualization and editing of this object model through a web user interface. Additionally, TEI files can be regenerated from the object model. This approach has several advantages: (1) the archives' experts continue using editor tools like oXygen to do their work; (2) users (experts and non-experts) can create their virtual editions and fragment extensions through the web user interface; (3) the object model preserves a semantically consistent *LdoD* archive by checking the consistency of users' operations; (4) interoperability (interchange) can be supported by exporting the regenerated TEI files; (5) it is possible to customize the generation of TEI files.
- 42 This architecture is implemented by a full-fledged prototype that the *LdoD* encoders use on a daily basis to see the result of their static encoding. Additionally, we are starting to plan experiments to study how users interact with the repository using the dynamic Web 2.0 features. The prototype is implemented in the JAVA programming language, using the [Spring MVC framework](#)⁹ for server-side interaction, [Bootstrap](#)¹⁰ and [jQuery](#)¹¹ for client-side interactions, and the [Fénix Framework](#)¹² to support a transactional and persistent domain model.

8. Conclusion

- 43 The specific correlation of static and dynamic goals in the *LdoD* Digital Archive means that our emphasis falls on open changes that feed back into the archive. The TEI encoding and software design implications of this project make us address both the conceptual aspects of TEI schemas for modeling texts and documents, and the processing problems posed by user-oriented virtualization of Pessoa's writing and bibliographic imagination.
- 44 In this article we have presented how to encode *LdoD* editions and fragments using TEI,¹³ and also the software architecture of a Web 2.0 environment through which the encoded fragments are fed. These two approaches create an environment where experts and non-experts can collaborate in four roles: reader, book, editor, and author. The environment separates the different contributions while allowing certain levels of information-sharing that enhance collaboration and the continuous enrichment of the repository.
- 45 In this phase we have already implemented a full-fledged prototype that is used on a daily basis by the TEI encoders. Currently we are extending the architecture and prototype with more affordances for the editor and author functions. Meanwhile we intend to start experiments with end users to assess the dynamic aspects of the environment and study how to foster the creation of communities around a virtual *LdoD*.

BIBLIOGRAPHY

- Barney, Brett. 2012. "Digital Editing with the TEI Yesterday, Today, and Tomorrow." *Textual Cultures* 7(1): 29–41.
- Bauman, Syd. 2011. "Interchange vs. Interoperability." In *Proceedings of Balisage: The Markup Conference 2011*. Balisage Series on Markup Technologies, vol. 7. doi:10.4242/BalisageVol7.Bauman01.
- Benel, Aurelien, and Christophe Lejeune. 2009. "Humanities 2.0: Documents, Interpretation and Intersubjectivity in the Digital Age." *International Journal of Web Based Communities* 5(4): 562–76. doi:10.1504/ijwbc.2009.028090.
- Brüning, Gerrit, Katrin Henzel, and Dietmar Pravida. 2013. "Multiple Encoding in Genetic Editions: The Case of 'Faust'." *Journal of the Text Encoding Initiative* 4 (March). <http://jtei.revues.org/697>. doi:10.4000/jtei.697.
- Coelho, Jacinto do Prado, ed. 1982. Fernando Pessoa, *Livro do Desassossego*. 2 vols. Lisboa: Ática.

- Cunha, Teresa Sobral, ed. 1990–91. Fernando Pessoa, *Livro do Desassossego*. 2 vols. Lisboa: Relógio d'Água.
- . 2008. Fernando Pessoa, *Livro do Desassossego*. Lisboa: Relógio d'Água.
- . 2013. Fernando Pessoa, *Livro do Desassossego*. Lisboa: Relógio d'Água.
- Earhart, Amy E. 2012. "The Digital Edition and the Digital Humanities." *Textual Cultures* 7(1): 18–28.
- Fraistat, Neil, and Steven E. Jones. 2009. "Editing Environments: The Architecture of Electronic Texts." *Literary and Linguistic Computing* 24(1): 9–18. doi:10.1093/llc/fqn032.
- McGann, Jerome. 2006. "From Text to Work: Digital Tools and the Emergence of the Social Text." *Text* 16: 49–62.
- Muñoz, Trevor, Raffaele Viglianti, and Neil Fraistat. 2013. "Texts and Documents: New Challenges for TEI Interchange and the Possibilities for Participatory Archives." *The Linked TEI: Text Encoding in the Web. Book of Abstracts. Abstracts of the TEI Conference and Members Meeting 2013*, edited by Fabio Ciotti and Arianna Ciula, 91–96. Rome: DIGILAB Sapienza University and TEI Consortium. <http://digilab2.let.uniroma1.it/teiconf2013/wp-content/uploads/2013/09/book-abstracts.pdf>.
- Pizarro, Jerónimo, ed. 2010. Fernando Pessoa, *Livro do Desasocego*. Edited by Jerónimo Pizarro. 2 vols. Lisboa: Imprensa Nacional-Casa da Moeda.
- . 2013. Fernando Pessoa, *Livro do Desassossego*. Lisboa: Tinta-da-China.
- Portela, Manuel, and António Rito Silva. 2014. "A Model for a Virtual LdoD." *Digital Scholarship in the Humanities* March 2014. doi:10.1093/llc/fqu004.
- Robinson, Peter. 2013. "Towards a Theory of Digital Editions." *Variants* 10: 105–31.
- Schlitz, Stephanie A., and Garrick S. Bodine. 2009. "The TEIViewer: Facilitating the Transition from XML to Web Display." *Literary and Linguistic Computing* 24(3): 339–46. doi:10.1093/llc/fqp022.
- Schmidt, Desmond, and Robert Colomb. 2009. "A Data Structure for Representing Multi-version Texts Online." *International Journal of Human-Computer Studies* 67(6): 497–514. doi:10.1016/j.ijhcs.2009.02.001.
- Siemens, Ray, Mike Elkins, Alastair McColl, Karin Armstrong, James Dixon, Angelsea Saby, Brett D. Hirsch, Cara Leitch, et al. 2010. "Underpinnings of the Social Edition? A Narrative, 2004–9, for the Renaissance English Knowledgebase (REKn) and Professional Reading Environment (PReE) Projects." *Online Humanities Scholarship: The Shape of Things to Come. Proceedings of the Mellon Foundation Online Humanities Conference at the University of Virginia, March 26–28, 2010*, edited by Jerome McGann, Andrew M. Stauffer, Dana Wheelers, and Michael Pickard, 401–60. Houston, TX: Rice University Press.
- Silva, António Rito, and Manuel Portela. 2013. "Social Edition 4 *The Book of Disquiet*: The Disquiet of Experts with Common Users." In *ECSCW 2013 Adjunct Proceedings: The 13th European Conference on Computer-Supported Cooperative Work*, edited by Matthias Korn, Tommaso Colombino, and Myriam Lewkowicz, 45–50. Aarhus, Denmark: Department of Computer Science, Aarhus University. http://cs.au.dk/~mkorn/ECSCW_2013_Adjunct_Proceedings-web.pdf.

- TEI Consortium. 2012. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 2.5.0. Last updated October 25. N.p.: TEI Consortium. <http://www.tei-c.org/Vault/P5/2.2.0/doc/tei-p5-doc/en/html/>.
- Tummarello, Giovanni, Christian Morbidoni, and Elena Pierazzo. 2005. "Toward Textual Encoding Based on RDF." In *From Author to Reader: Challenges for the Digital Content Chain: Proceedings of the 9th ICC International Conference on Electronic Publishing*, 57–63. Leuven-Heverlee, Belgium: Peeters Publishing Leuven. <http://elpub.scix.net/data/works/att/206elpub2005.content.pdf>.
- Vanhoutte, Edward. 2006. "Prose Fiction and Modern Manuscripts: Limitations and Possibilities of Text- Encoding for Electronic Editions." In *Electronic Textual Editing*, edited by Lou Burnard, Katherine O'Brien O'Keefe, and John Unsworth, 161–80. New York: Modern Language Association of America.
- Wittern, Christian. 2013. "Beyond TEI: Returning the Text to the Reader." *Journal of the Text Encoding Initiative* 4 (March). <http://jte.revues.org/691>. doi:10.4000/jtei.691.
- Zenith, Richard, ed. [1998] 2012. Fernando Pessoa, *Livro do Desassossego*. Lisboa: Assírio & Alvim.

NOTES

- 1 The concept "heteronym" was developed by Fernando Pessoa himself – heteronyms are fictional authors who have a specific writing style and a unique psychology. Many of Pessoa's works have been written by a heteronym. In the case of the *Book of Disquiet*, the work was assigned by Pessoa to heteronym Vicente Guedes during the first stage of writing, and to Bernardo Soares during the later stage. One editor has assigned this work to both Vicente Guedes and Bernardo Soares; two editors have assigned it to Bernardo Soares; and one editor has assigned it to Fernando Pessoa.
- 2 "No Problem Has a Solution: A Digital Archive of the *Book of Disquiet*," a research project of the Centre for Portuguese Literature at the University of Coimbra, Portugal, is funded by FCT (Foundation for Science and Technology). Principal investigator: Manuel Portela. Reference: PTDC/CLE-LLI/118713/2010. Co-funded by FEDER (European Regional Development Fund), through Axis 1 of the Operational Competitiveness Program (POFC) of the National Strategic Framework (QREN). COMPETE: FCOMP-01-0124-FEDER-019715.
- 3 During the last three decades, the *Livro do Desassossego* has been translated several times, and new translations continue to appear in Spanish, French, English, Italian, German, Dutch, Swedish, Polish, and several other languages. *LdoD* is now regarded by many critics as one of the major achievements of Fernando Pessoa and one of the major works of European modernism. In Portugal, two new revised editions (Sobral Cunha 2013; Pizarro 2013) were published in 2013, and

excerpts from *LdoD* were recently selected for inclusion in the secondary school syllabus. Pessoa's heteronyms, such as Álvaro de Campos, Alberto Caeiro, Ricardo Reis, and Bernardo Soares, are among the most quoted authors in Brazilian and Portuguese blogs. A feature film based on this unfinished book, *Filme do Desassossego*, directed by João Botelho, was released to critical acclaim in 2010.

4 Our XML encoding of what we refer to as “authorial” and “editorial” witnesses is in itself a new edition of *LdoD*. The distinction that we make in this article between “authorial” and “editorial” serves the rhetorical purpose of showing two types of relation between our XML encoding and its source texts: in the case of “authorial” witnesses we are referring to our topographic transcription of autograph documents, which will also be accessible as digital image facsimiles; in the case of “editorial” witnesses we are referring to our textual transcription of *LdoD* texts as they have been transcribed and organized in four major editions published between 1982 and 2013. It can be argued that our XML topographic transcription of the “authorial” witnesses is of course a fifth editorial witness of the *LdoD*, since it cannot be considered an authorial source. It must always be a particular reading and representation of that source, even if we intend to give readers a less mediated text by placing it in the context of the document facsimiles. It should, however, be noted that, at this level, we are not proposing a specific selection or organization for the fragments other than the semi-arbitrary shelf marks of the National Library. In fact, the fragments included for representation in the archive equal the total sum of the fragments included in those four editions. Our topographic transcription could be described as a new implicit edition by the research team focused on representing revision layers and describing material details of the source documents. An explicit edition by the research team, with its own selection and organization rationale, will take place only at the virtual level of the archive.

5 A second stage of development of this project (2016–2018) will add a corpus of reception texts (reviews and essays) that will be fully integrated with the *Book of Disquiet* fragments. This new stage will also develop software tools—including text editors, textual generators and other e-lit tools—for enabling writing interactions with the archive's textual database. Six applications for mobile media, called “Machines of Disquiet,” are currently being tested.

6 Although we refer to expert and virtual editions, the latter may also be constructed by Pessoa experts. “Expert editions” refers to the four printed editions published prior to the existence of the *LdoD Archive*, and also to the project’s topographic transcription.

7 Although we are not, at this stage of the project, creating a critical apparatus for variants for these major four editions, the possibility remains open for such a representation in the future, if resources permit. Except for Prado Coelho, whose 1982 edition was not reedited (he died soon after publication), the other three editors have never reissued *LdoD* under the exact same form. Every new publication becomes an occasion for revision. There are now more than fifteen editions that may be considered variations of those three sets. Our copyright agreement with the *LdoD* editors is that we encode in the archive the most recent version at the time of work.

8 The interactions with the repository are done on top of the object model described in figures 3 and 4. It is possible to generate TEI-encoded files from the repository. On the other hand, the system allows tagging and annotation of transcriptions but does not allow them to be changed.

9 Spring MVC framework: <http://spring.io/>.

10 Bootstrap: <http://getbootstrap.com/>.

11 JQuery: <http://jquery.com/>.

12 Fénix Framework: <http://fenix-framework.github.io/>.

13 A complete example of an encoded fragment is included in appendix 1.

ABSTRACT

In this article we describe how textual encoding is used in our current project of constructing a digital archive of Fernando Pessoa’s *Livro do Desassossego [LdoD]*. Our model for virtualizing the authorial and editorial forms of this unfinished work aims to create a dynamic archive that is open to user interaction and collaboration. A brief introduction to the theoretical rationale of the archive is followed by a description of our technical solution for TEI XML encoding that is responsive to dynamic changes over time. With our software architecture proposal for processing TEI markup, the Collaborative Digital Archive of the Book of Disquiet will be able to instantiate the cooperative and social editing functionalities of Web 2.0 environments.

INDEX

Keywords: Fernando Pessoa, Book of Disquiet, digital archive, Web 2.0, collaborative archive

AUTHORS

ANTÓNIO RITO SILVA

António Rito Silva is associate professor in the Department of Computer Science and Engineering, University of Lisbon, and researcher at the [INESC-ID](#). He has worked in the field of collaborative systems and social software, particularly in the domain of business processes tools that blend the roles of producer and consumer. In this project he intends to apply some of these techniques for connecting the roles of producer and reader of the literary work.

MANUEL PORTELA

Manuel Portela is assistant professor with Habilitation in the Department of Languages, Literatures and Cultures, University of Coimbra, Portugal, where he directs the Doctoral Program in [Advanced Studies in the Materialities of Literature](#). He is also a researcher at the [Centre for Portuguese Literature](#) at the University of Coimbra, and the principal investigator of the project [No Problem Has a Solution: A Digital Archive of the Book of Disquiet](#) (2012–15). His latest book is *Scripting Reading Motions: The Codex and the Computer as Self-Reflexive Machines* (MIT Press, 2013).