

A Web Application for Exploring Primary Sources: The DanteSources Case Study

Valentina Bartalesi

ISTI-CNR, via Moruzzi 1, 56124 Pisa, Italy. E-mail: valentina.bartalesi@isti.cnr.it

Carlo Meghini

ISTI-CNR, via Moruzzi 1, 56124 Pisa, Italy. E-mail: carlo.meghini@isti.cnr.it

Daniele Metilli

ISTI-CNR, via Moruzzi 1, 56124 Pisa, Italy. E-mail: daniele.metilli@isti.cnr.it

Mirko Tavoni

Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa, Piazza Torricelli 2, 56126, Pisa, Italy. E-mail: mirko.tavoni@gmail.com

Paola Andriani

Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa, Piazza Torricelli 2, 56126, Pisa, Italy. E-mail: paola.andriani@gmail.com

16

17

18

19 **Abstract**

20 We present the methodological and technical process we adopted to develop DanteSources, a Web application that allows free
21 access to the knowledge about Dante Alighieri's primary sources, i.e. the works of other authors that Dante cites in his texts. Up
22 to now, this knowledge has been collected in many paper books, making it difficult for the scholars to retrieve it and to produce a
23 complete overview of these data. Using Semantic Web technologies, we developed an ontology expressed in the RDF/S
24 vocabulary providing the terms to represent this knowledge in a machine-readable form. A semi-automatic tool helps the scholars
25 to populate the ontology with the data included in authoritative paper commentaries to Dante's works. Then, the tool
26 automatically saves the resulting RDF graph in a triple store. On top of this graph, we developed DanteSources, a Web
27 application that allows users to extract and display the information stored in the knowledge base in the form of charts and tables.
28 Finally, we report the results of a survey to collect suggestions from end-users on their interactions with DanteSources. The
29 methodology and the tools we developed are easily reusable, e.g. to represent the knowledge about primary sources of other
30 authors of the Italian and the international literature.

31

32 **Introduction**

33 One important kind of knowledge studied by scholars in the Humanities field is represented by the primary sources of literary
34 texts, which are crucial in order to reconstruct the history of the creation of a text. Primary sources are the literary works of other
35 authors that an author refers to in her/his texts (Hellqvist, 2010; Dalton and Charnigo, 2004; Malkmus, 2010). In particular, the
36 primary sources we considered are literary texts, we did not take into account different primary sources like monumental
37 inscriptions. This study is a crucial aspect of this research field since it allows reconstructing the cultural background of an
38 author and its evolution in time. Usually, this knowledge is reported in commentaries, written by scholars in natural language,
39 which the readers use to study and decode several aspects of the literary texts. These commentaries are called secondary sources,
40 i.e. documents that relate or discuss information originally presented in a primary source. A secondary source contrasts with a
41 primary source, which is an original source of the information being discussed. Secondary sources involve generalisation,
42 analysis, and interpretation of the original information. Unfortunately, commentaries are usually in natural language form and the
43 knowledge contained in them is not machine-readable (Hendler, 2012), thereby limiting the possibility for scholars to obtain
44 automatic inferences in order to derive new knowledge. For instance, mapping the distribution of the primary sources cited in a
45 work would allow scholars to obtain a precise understanding of which sources influenced the author's work at a particular
46 moment in time. In addition, some of the main research questions still unsolved that are at the basis of our research are: (i) is it
47 possible to retrieve from a text the textual fragments that refer to different primary sources? (ii) what kinds of references are
48 present in a text (e.g. explicit references, strict references)? (iii) is it possible to retrieve from a text the primary sources organised
49 by thematic area (e.g. Aristotelianism, Astronomy, Physics)?

50 In order to overcome these limitations in the study of literary texts, we developed an ontology for representing the knowledge
51 about primary sources and a Web application that allows visualising this knowledge in a user-friendly format, such as charts and
52 tables.

53 Our study is part of the "Towards a Digital Dante Encyclopedia" project (2013–2016), an Italian National Research project
54 supporting scholars in formally expressing and accessing the knowledge about primary sources cited in the works of Dante
55 Alighieri — the major Italian poet of the late Middle Ages — and more generally in literary texts.

56 In this paper, we describe the methodology and the technical development process we followed to create DanteSources
57 (<http://dantesources.org>), a Web application that allows free access to the information about Dante Alighieri's primary sources.
58 We illustrate the entire process from the creation of the ontology for representing the knowledge on the primary sources to the
59 development of the Web application.

60 One of the main goals of our work was to create an ontology providing a formal representation of the knowledge on primary
61 sources of literary texts. To this aim, we adopted the Semantic Web technologies, in particular the RDF/S (Resource Description

62 Framework Schema) vocabulary, which provides the terms to represent the knowledge in a machine-readable form. In the Digital
63 Humanities literature, there are many ontologies focusing on different aspects of textual information. Each of these ontologies
64 represents a set of possible interpretations of the source text(s). Our developed ontology (Bartalesi and Meghini, 2015) is the first
65 attempt to formally represent this knowledge, since up to now vocabularies to express this information have not been developed
66 and, to the best of our knowledge, at the moment there are no other research projects focussing on giving a logical representation
67 of this knowledge. Furthermore, our ontology extends well beyond the specific author's primary sources we are considering in
68 our project.

69 On top of our ontology, we developed the Web application DanteSources that allows visualising the data in form of charts,
70 tables, and exporting them in CSV format. In our project, these visualisations provide useful knowledge to the scholars who are
71 working on the creation of a complete encyclopedia of Dante's works. Furthermore, to the best of our knowledge, this is the first
72 time that the information about Dante's primary sources is available in machine-readable format. Indeed, in the commentaries the
73 knowledge about primary sources is expressed in natural language, thereby preventing automatic inferences of new information
74 that may be useful for the studies of the scholars, e.g. the inferences can concern the total amount of references to a certain work.
75 In addition to scholars and Dante's experts, our application could be useful for high school professors who want to analyse the
76 Dante's works with their students, discovering their primary sources in a quick and easy way. A benefit could derive also for
77 university students in the Humanities and Digital Humanities who write their thesis on the primary sources used by Dante. For
78 example, starting from the results of the DanteSources queries they could compare the primary sources used by Dante before and
79 after his exile¹.

80 The methodology and the tools we developed within the project are freely available on request for research aims, by e-mail to
81 the authors. The tools are open-source and released under the GPLv3 license². Indeed, they are easily reusable in order to
82 represent the knowledge about primary sources of other authors of the Italian and international literature, in addition to Dante
83 Alighieri. Our methodology could also be applied to scientific literature in order to formally represent knowledge about its
84 primary sources.

85 The paper is structured as follows: in order to understand the functionalities of the DanteSources Web application, Section
86 Background reports the definitions of the main notions used in the Humanities field. The Related Works Section presents several
87 applications developed about Dante Alighieri's works and a list of ontologies to represent textual knowledge. The Methodology
88 Section describes the methodology we followed to develop our Web application. In the Ontology Section we present the details
89 of the semantic model and its implementation to represent the knowledge about primary sources. In the Population Tool Section,
90 we describe the software to populate the developed ontology. Then, the DanteSources Web Application Section reports the

91 description of the functionalities of our application and how the knowledge is presented to the final user. In the Evaluation
92 Section, we report data about a user test we conducted in order to evaluate the usability of DanteSources. The Conclusions
93 Section reports our final remarks.

94 **Background**

95 The knowledge about the primary sources of literary works is usually reported in commentaries. A commentary is a text, usually
96 reported below the text of the literary work, which contains a series of comments, or explanations about the work. Authoritative
97 commentaries are those written by notable scholars and the scientific validity of them is recognised by the scientific reference
98 community. These commentaries are composed of notes. A note is a textual fragment that refers to a particular expression in the
99 original work. A note is composed of two parts: a piece of text of the original work and a detailed comment of this piece of text
100 developed by a scholar, where also information about primary sources are reported. An example is shown in Figure 1.
101 Commentaries and their notes, neither have a standard structure, nor follow a standard codification, thus each scholar can codify
102 her/his commentary using her/his preferred codification model. However, these codifications are usually very similar between
103 them, thus it is possible to develop software that, using different rules, automatically extracts the two components of the
104 commentary, i.e. the reported text of the literary work and the related comment. An example of a commentary with a different
105 codification from Figure 1 is shown in Figure 2.

106 Up to now, in our research we have taken into account six authoritative commentaries related to the following Dante Alighieri's
107 works: *Convivio* (Alighieri, 2014a), a philosophical essay written between 1304 and 1307; *Vita Nova* (Alighieri, 2011b) and *Vita*
108 *Nuova* (Alighieri, 1980), are two different versions of commentaries written by two different scholars on the same Dante's
109 work, published in 1295, an expression of the medieval genre of courtly love. *De vulgari eloquentia* (Alighieri, 2011a), a Latin
110 essay written between 1302 and 1305 discussing the relationship between Latin and vernacular; *Monarchia* (Alighieri, 2014b), a
111 Latin treatise on secular and religious power; and *Rime* (Alighieri, 2011c), lyric poems that Dante wrote throughout his life,
112 based on the poet's varied existential and stylistic experiences.

113

114 Insert Figure 1 here

115

116 Insert Figure 2 here

117

118 **Related Works**

119 *Applications of Dante's Works*

120 Several Web applications that allow investigating different aspects of Dante's works are currently available. Unfortunately, none
121 of them focuses on the primary sources nor uses a Semantic Web approach to represent this information. In the following, we
122 report an overview of the most important projects about Dante Alighieri's work.

123 The Dartmouth Dante Project (DDP)³ aims at publishing the full searchable text of the *Divine Comedy* and several of its
124 commentaries in digital format (Hollander, 1989). On top of the DDP, the Web application Dante Lab⁴ was developed. This
125 application allows the concurrent visualisation of the original text of the *Divine Comedy*, some translations i.e. English, French
126 and German, and more than 75 commentaries, each of them fully searchable. The Princeton Dante Project⁵ includes the full
127 searchable text of the *Divine Comedy*, Dante's minor works, several commentaries and multimedia resources (Hollander, 2013).

128 Differently from our approach, in the previous projects, the results of the queries are not exportable from the Web application
129 and it is not possible to automatically extract and aggregate the primary sources for each text. It is only possible to retrieve this
130 information by reading the content of the commentaries on the computer screen.

131 The World of Dante⁶ is a multimedia research tool that contains the whole text of the *Divine Comedy* with a basic semantic
132 annotation that identifies people, places, deities and structures, and connections to digital objects such as images and music.
133 However, the semantic knowledge is not represented through a formal ontology and is not available in a machine-readable format
134 (Hendler, 2012).

135 Digital Dante⁷ offers the full text of all Dante's works with commentaries, illustrations and recorded readings but not in a
136 machine-readable form. Digital Dante provides a functionality, namely Intertextual Dante, which links the *Divine Comedy*'s text
137 to the corresponding Ovid's works. This functionality is limited to a single author, Ovid, and its static implementation is based on
138 a manual annotation of the texts.

139 Dante Online⁸ is a Web site that allows visualising and searching the full text of all Dante's works. It also includes a biography
140 with an interactive timeline, and a database listing hundreds of manuscripts of the *Divine Comedy* with bibliographic information.
141 This site does not contain the text of commentaries to Dante's works nor information on the primary sources.

142 DanteSearch⁹ allows users to perform morphological and syntactic queries on the full text of the Dante's works. This application
143 focusses on the morpho-syntactic aspects of the Dante's texts but no semantic annotation is performed on them.

144 Finally, DaMA¹⁰ is a digital archive containing the full text of Dante's works and several medieval primary sources in XML
145 (eXtensible Markup Language)-TEI (Text Encoding Initiative) format. DaMA allows visualising the text of the medieval primary
146 sources used by Dante but, differently from DanteSources, it does not provide links between each fragment of a Dante's work
147 and the corresponding primary sources.

148 *Ontologies*

149 In the Semantic Data Modeling field, several ontologies have been developed in order to represent textual knowledge. We have
150 investigated the scientific literature and the existing standards in the Digital Libraries field (Altman, 2006), both official and *de*
151 *facto*. Up to now, a formal ontology for representing knowledge about the primary sources of literary texts is not present in the
152 scientific literature. As explained in Bartalesi and Meghini (2015), we did not find a single ontology able to represent all the
153 classes and properties we identified when analysing the commentaries on Dante's works. In order to maximise the
154 interoperability of our ontology, we reused classes and properties from the following ontologies¹¹, adding our own classes and
155 properties only if necessary.

- 156 • **Functional Requirements for Bibliographic Records Object Oriented (FRBRoo)**. It is essentially the Functional
157 Requirements for Bibliographic Records (FRBR) ontology (Riva, Doerr and Zumer, 2008) expressed in an object-
158 oriented form that harmonises the FRBR and the CIDOC Conceptual Reference Model (CRM) (Doerr, 2002) to better
159 share library and museum information.
- 160 • **Dublin Core**. The Dublin Core Metadata Element Set¹² is a vocabulary of fifteen properties for usage in resource
161 description. These elements of the Dublin Core are part of a larger set of metadata vocabularies and technical
162 specifications developed by the Dublin Core Metadata Initiative.
- 163 • **Simple Knowledge Organization System (SKOS)**. It is a model for sharing and linking knowledge organisation
164 systems such as thesauri, classification schemes, subject heading systems, and taxonomies within the framework of the
165 Semantic Web (Matthews, Wilson, and Brickley, 2005).
- 166 • **Friend of a Friend (FOAF)**¹³. It is a machine-readable ontology linking people and information using the Web. In
167 particular, FOAF describes persons, their activities and their relations to other people and objects.
- 168 • **Document Components Ontology (DoCO)**. This ontology (Constantin, Peroni, Pettifer, Shotton, and Vitali, 2016)
169 allows describing document components in RDF. Such components are both structural (e.g. block, chapter, heading,
170 paragraph, section) and rhetorical (e.g. Abstract, Introduction, Results, Discussion, Conclusions, Acknowledgements,
171 Bibliography).
- 172 • **FRBR-aligned Bibliographic Ontology (FaBiO)**. It is an ontology (Peroni and Shotton, 2012) to record and publish
173 bibliographic records created by scholars on the Semantic Web. FaBiO entities are primarily textual publications such
174 as books, magazines, newspapers and journals, and items of their content such as poems, conference papers and
175 editorials.
- 176 • **Citation Typing Ontology (CiTO)**. It is an ontology (Shotton and Peroni, 2011) for the characterisation of
177 bibliographic citations, both factually and rhetorically and for their publication on the Web.

178 • **The Open Annotation Core Data Model.** The Open Annotation Core Data Model (Sanderson, Ciccarese, and Van de
179 Sompel, 2013) specifies an approach to associate annotations with resources, using a methodology that conforms to the
180 Architecture of the World Wide Web and the Linked Data initiative.

181 In the last decades, ontologies and Semantic Web technologies have been applied in the Humanities field in order to improve
182 the interoperability of the data, e.g. through analysis of iconographic representation in manuscripts, data aggregation for digital
183 history, semantic annotation of literary texts. For instance, in González-Blanco, Seláf, Del Rio Riande, Martínez Cantón, and
184 Martos Pérez (2014) a semantic model to connect repertoires of poetic writings is described. Another example is Biblissima
185 (Gehrke, Frunzeanu, Charbonnier, and Muffat, 2015), an online digital library, which provides easy and coordinated access to a
186 huge and complex mass of documentation on manuscripts and early printed books from the 8th to 18th Centuries. In order to
187 handle the heterogeneity of the database formats and the variety of the data (manuscript cataloguing databases, textual editions,
188 iconographic databases), in the Biblissima project the CIDOC Conceptual Reference Model and FRBRoo are chosen as reference
189 ontologies to facilitate the internal mapping to a single common model and to allow the partners of the project to expose their
190 data in RDF compliant to a globally established standard.

191 In the WarSampo project (Hyvönen et al., 2016), the Semantic Web and Linked Data technologies are used for modeling,
192 harmonising, and aggregating data about Second World War. This approach makes it possible, for both historians and laymen, to
193 study history in a contextualised way where linked datasets enrich each other.

194 Wagner and Caesar (2016) describe a digital collection of more than a hundred sources from Iberian theologians and jurists of the
195 16th and 17th Century. To formally represent data about the primary sources and the authors, the FOAF, BIO¹⁴, Relationship¹⁵,
196 and SPAR¹⁶ ontologies were used. Niccolucci and D'Andrea (2006) show an ontology for the 3D visualisation of cultural
197 heritage; Lana et al. (2014) present an ontology for annotating geographical places in texts. Currently, to the best of our
198 knowledge, a research project for the formal representation of the primary sources of literary texts using the Semantic Web
199 technologies does not exist yet. Taking into account the previous projects about the works of Dante Alighieri and the developed
200 ontologies for representing aspects of literary texts, we developed DanteSources, where we applied the technologies of the
201 Semantic Web to literary texts in order to represent the knowledge on Dante's primary sources.

202 **Methodology**

203 The main phases of the development process of DanteSources were: (i) analysing the representation requirements about citations
204 to primary sources, in collaboration with a Dante's expert; (ii) creating an ontology responding to the requirements, providing a
205 formal representation of the terms required to express knowledge about the primary sources, using the Semantic Web language;
206 (iii) developing a semi-automatic tool for populating the ontology and storing the resulting data in a knowledge base; (iv)

207 developing a Web application running on top of the knowledge base that makes inferences based on the contents of the
208 knowledge base and visualises these inferences using charts and tables. Figure 3 summarises the main phases of our approach.

209

210 Insert Figure 3 here

211

212 First of all, semi-structured interviews were performed with four Dante's experts who collaborated with us during the whole
213 project, in order to clarify the requirements of the project, in particular, the meaning of the data and the functionalities needed for
214 the citations to primary sources. We asked to the experts to identify the relevant knowledge we had to extract from the
215 commentaries and to suggest to us how this knowledge had to be further elaborated.

216 According to the scholars, our ontology has to represent the following pieces of knowledge:

- 217 ▪ the textual fragment of Dante's works which the note refers to (e.g. "As the Philosopher says in the beginning of the first
218 Philosophy");
- 219 ▪ the textual fragment of the note citing a primary source (e.g. "These are the words that open the Aristotle's *Metaphysics* I 1,
220 980a 21");
- 221 ▪ the structure of Dante's work containing the fragment (e.g. book, chapter, and paragraph, poem, *cantica*, etc.);
- 222 ▪ the name of the author of the primary source (e.g. Aristotle);
- 223 ▪ the title of the primary source (e.g. *Metaphysics*);
- 224 ▪ the thematic area of the primary source (e.g. Aristotelianism, Bible, Medical Science, Scholastic Philosophy).
- 225 ▪ the entire text of the commentary, thus making available to the scholars the context of the reference;
- 226 ▪ the type of reference, which can be:
 - 227 ▪ *explicit*, if the reference is explicitly made by Dante, as in "As the Philosopher says at the beginning of the First
228 Philosophy", where the Philosopher is Aristotle and the First Philosophy is *Metaphysics*;
 - 229 ▪ *strict*, if the reference is indicated by a scholar, and refers to a specific work as in "SI MANUCA: it is the bread of the
230 angels, the *manna* as called in the Old Testament (Ps. 77, 25 *Panem angelorum manducavit homo*)";
 - 231 ▪ *generic*, if the reference is indicated by a scholar, and refers to a concept (e.g. "Medieval comments to Aristotle's
232 works").

233 Once the semantics was understood, we created an ontology to represent it, with the additional requirement of maximising
234 reusability. To this end, we investigated several existing vocabularies to identify useful classes and properties to represent this
235 knowledge. In particular, we analysed the ontologies addressing issues related to the textual domain, reported in the Related
236 Works Section. We chose the classes and properties from those ontologies that we considered the most appropriate. When we
237 could not find appropriate classes and properties in the existing ontologies to represent the necessary kind of knowledge, we

238 created our own. Following the Linked Data recommendations, we used HTTP URIs (Uniform Resource Identifier) for denoting
239 resources and we provided a description for every resource. Then, we validated our ontology by showing to the experts the
240 semantic network resulting from encoding of the citations of *Convivio* (Alighieri, 2014a). Once the ontology was validated, we
241 invited the domain experts participating in the research project to extract the relevant information about primary sources from
242 authoritative commentaries about other Dante's works, and to encode such information in our ontology. In order to facilitate the
243 experts in this task, we developed a semi-automatic tool allowing them to easily export the extracted knowledge in XML form.
244 The resulting XML files were automatically exported in RDF format. The resulting RDF graph was saved into a Virtuoso triple
245 store.

246 In parallel with the ontology population phase, we developed a Web application to explore the semantic network of Dante's
247 works and their primary sources. The requirements to develop the Web application we identified by interviewing the Dante's
248 experts are the following:

- 249 ▪ Extracting either the primary sources or the cited authors or the thematic areas for each Dante's work
- 250 ▪ Extracting all the Dante's works citing a specific primary source, author and thematic area
- 251 ▪ Extracting the types of citation (explicit, strict, generic) from each Dante's work and primary source

252 The Web application provides search functionalities corresponding to each requirement and visualises the results through
253 charts and tables. In the final stage of the project, a group of external users were asked to fill in and submit a survey for the
254 evaluation of the Web application.

255

256 **The Ontology**

257 We developed an ontology for the semantic representation of the structure of the works of Dante and of the knowledge related to
258 their primary sources. An ontology is a set of classes and properties that describe a knowledge domain, along with axioms that
259 capture the formal meaning of these classes and properties. Our ontology is expressed in RDF Schema, a vocabulary for data
260 modeling recommended by the W3C (Manola, Miller and McBride, 2004). The schema is available on the DanteSources Web
261 site¹⁷. The classes and properties defined in RDF Schema are then used in an RDF graph, which is a collection of RDF statements
262 representing some slice of reality. An RDF statement is a triple in the form of subject-predicate-object (World Wide Web
263 Consortium, 2014).

264 In order to obtain an ontology for the semantic representation of the pieces of knowledge defined as requirements (see
265 Methodology Section), we analysed several existing ontologies as reported in the Related Works Section, and we chose the
266 classes and the properties that we considered the most appropriate for our purposes, respecting the constraints established in the
267 source ontologies, i.e. we preserved the definitions of the classes and the domains and ranges of the properties. Furthermore, we

268 added our own classes and properties for representing pieces of knowledge that the existing ontologies had not defined. Almost
269 90% of the classes and properties of our ontology come from existing ontologies.

270 In the following, the classes and properties we re-used as well as the new ones are reported. A more detailed description of the
271 ontology is reported in Bartalesi and Meghini (2015).

272 As a notational convention, we use prefixed qnames (e.g. foaf:Person) to denote terms re-used from other vocabularies,
273 whereas we use local names (e.g. hasCommentary) to denote terms of our own ontology.

274 From the FRBROo ontology we reused two classes to represent the related aspects of our textual knowledge: efrbroo:Work,
275 corresponding to a specific work, cited by the commentaries, (e.g., *De consolatione philosophiae*) without reference to a specific
276 edition; efrbroo:ExpressionFragment, corresponding to (i) a fragment of a Dante's work a commentary refers to and also (ii) to a
277 fragment of a commentary that refers to a specific primary source.

278 In our model a work has an author (e.g. Albertus Magnus) and a thematic area (e.g. Rhetoric, Scholasticism, Medical science,
279 etc.). We represented these pieces of knowledge using the classes foaf:Person and skos:Concept, respectively. To link the work,
280 author and thematic area with their corresponding names in natural language, we used the property dcterms:alternative.

281 To link the frbroo:Work class with its author and thematic area, we used two Dublin Core properties: dc:creator, and dc:subject,
282 respectively.

283 To describe the structure of the resource which the fragment is extracted from, we reused three classes of FaBiO and DoCo
284 ontologies:

- 285 1. doco:Paragraph, a self-contained unit of discourse that deals with a particular point or idea;
- 286 2. fabio:Poem, an artistic work written with an intensity or beauty of language more characteristic of poetry than of prose;
- 287 3. doco:Line, a line in poetry, i.e. a unit of language into which a poem is divided;
- 288 4. doco:Chapter, a principle division of the body matter of a large document;
- 289 5. fabio:Book, a non-serial document that is complete in one volume or a designated finite number of volumes.

290 The Open Annotation Model is the standard adopted by the W3C (World Wide Web Consortium), for this reason we decided to
291 be compliant with this ontology and to guarantee the interoperability with it. The Open Annotation Model defines an annotation
292 as composed of "a single Body, which is the comment or other descriptive resource, and a single Target that the Body is
293 somehow about".

294 In order to link the annotation with its components, we used the following classes and properties:

- 295 1. oa:Annotation class represents a commentary;
- 296 2. oa:TextualBody, which identifies the textual content of the Annotation;
- 297 3. oa:hasBody, which is the property that links the Annotation and its textual body;

298 4. oa:hasTarget, which is the property that links the Annotation and its related resource (in our case, it represents the primary
299 source);

300 5. frbroo:ExpressionFragment, i.e. a fragment of an annotation that refers to a specific primary source.

301 We also defined a new property hasCommentary to directly link the fragment of Dante's text to the body of the note that
302 describes it and identifies a related primary source.

303 Since in the content of the same commentary there can be several references to different primary sources, we added the new
304 property hasCitingFragment to relate the commentary with its own citing fragments. We defined hasCitingFragment as a
305 subproperty of frbroo:hasFragment.

306 The fragment of a commentary refers to a primary source through three subproperties of the cito:cites property we defined: (i)
307 citesAsExplicitCitation, (ii) citesAsStrictCitation, (iii) citesAsGeneralCitation.

308 Figure 4 shows an example of the representation of a note in our ontology. The individual danteExpressionFragment represents
309 the fragment of Dante's text "Si come dice lo Filosofo nel principio de la Prima Filosofia, tutti li uomini naturalmente desiderano
310 di sapere" (As the Philosopher says in the beginning of the First Philosophy, all men by nature desire to know). Dante's fragment
311 has a note, the individual noteBody, which is connected to it through the property oa:hasNote. The note contains another
312 fragment of text, noteExpressionFragment, which states "sono le parole con cui si apre la Metafisica di Aristotele" (these are the
313 words that open Aristotle's *Metaphysics*). This fragment is connected to the primary source through the property
314 citesAsExplicitCitation. The cited work is *Metaphysics*, which is Aristotle's First Philosophy. *Metaphysics* has dc:creator
315 Aristotle and dc:subject Aristotelianism.

316

317 Insert Figure 4 here

318

319 Using the semi-automatic tool described in the Population Tool Section, three Dante's experts populated the ontology with the
320 knowledge extracted from the authoritative commentaries listed in Section Background, i.e. *Convivio*, *Vita Nova*, *Vita Nuova*, *De*
321 *vulgari eloquentia*, *Monarchia*, *Rime*. The obtained RDF graph was eventually stored in a Virtuoso triple store (Erling and
322 Mikhailov, 2009).

323 The information stored in our knowledge base may be incomplete and can provide an incoherent representation of Dante's
324 works. For instance, two commentaries written by two different scholars who identify two different primary sources for the same
325 fragment of Dante's text may raise incoherency. Indeed, on the one hand our ontology is able to represent such conflicting
326 annotations, because it does not impose a unicity constraint on the primary sources related to one piece of text. On the other,

327 incompleteness may arise since in our knowledge base we stored only some commentaries to Dante's works. However, the
328 ontology allows the addition of new commentaries at any time in order to increase knowledge completeness.

329 A further problem in order to formally represent knowledge is the fact that the commentary of a literary text can be open to
330 different interpretations that can produce inconsistencies in the ontology. To avoid these types of inconsistencies, we are
331 considering extending the ontology adding to the commentary the information about who has interpreted it if s/he is a different
332 person from the author of the commentary itself.

333 The main advantages of using an ontology instead of a traditional database model are extensibility (researchers can add classes
334 and properties in order to refine the ontology) and interoperability (researcher can re-use the data by integrating our ontology
335 with theirs, and this is much easier with Linked Data than with traditional data). Furthermore, it is possible to automatically
336 check the consistency of the knowledge base with respect to the ontological model using a semantic reasoner (Mishra and Kumar,
337 2011). Moreover, the semantic reasoners allow to automatically infer new implicit knowledge. Furthermore, the ontology can be
338 linked to other existing ontologies to extend the represented domain. Finally, any user can download and use the ontology
339 schema freely, using the paradigm of Linked Data.

340 **The Population Tool**

341 In order to facilitate the population process of the ontology, we implemented a semi-automatic tool. The population of the
342 ontology was carried out by three of the four Dante's experts who participated in our project (under the supervision of the fourth
343 scholar, who coordinated the Dante's expert group). The six commentaries (the full list is reported in The Ontology Section) we
344 started from were in PDF and MS Word format. The total size of the commentaries is about 1,600,000 words. Through a Java
345 software we developed, we pre-processed the texts of the commentaries in order to automatically extract the following
346 information: (i) the fragment of Dante's text which the note refers to, (ii) the entire text of the note, (iii) the title of the Dante's
347 work along with the number of the book, chapter, paragraph which the fragment belongs to. Since in the Humanities a standard
348 format to write a commentary does not exist and each scholar writes her/his commentary using different formats, in our software
349 for each commentary we used different regular expressions in order to identify specific textual fragments and pieces of
350 knowledge. Once extracted, these data were stored in an XML file. This automatic phase can be estimated in 6 person-days (7
351 hours per day) for two Information Science researchers, including the development of different regular expressions for each
352 commentary.

353 After this first phase of pre-processing, we developed a tool with a Web interface in order to facilitate the insertion of the
354 remaining data by the experts. The tool is developed in HTML5 and JavaScript using the jQuery¹⁸ and jQuery UI¹⁹ libraries. It
355 takes as input the XML file, where the data extracted from the commentaries are stored, and automatically shows this data in the

356 corresponding fields of the interface. It also facilitates the user's work by storing the inserted data in an IndexedDB²⁰ database
357 and by suggesting it to the user through autocomplete fields.

358 After analysing the notes to Dante's works reported in the commentaries, the experts inserted through the interface the
359 following remaining pieces of knowledge: (i) the fragment of the note referring to a primary source, (ii) the title of the primary
360 source, (iii) the name of the author, (iv) the thematic area which the primary source belongs to and (v) the type of reference. For
361 primary sources, authors and thematic areas, the experts also inserted the corresponding URIs. For primary sources and authors,
362 the URIs were imported from the DBpedia²¹ knowledge base. For thematic areas, they were imported from the Nuovo
363 Soggettario²² thesaurus. For instance, the URI for Aristotle's Poetics is [http://dbpedia.org/page/Poetics_\(Aristotle\)](http://dbpedia.org/page/Poetics_(Aristotle)), and the URI
364 for Aristotelianism is <http://purl.org/bnfc/tid/19144>. When the URIs were not available on DBpedia or in the Nuovo Soggettario,
365 custom URIs were automatically assigned by the population tool.

366 Figure 5 shows the interface of the tool.

367

368 Insert Figure 5 here

369

370

371 As the data were gradually added, they were saved as lists in a JSON (JavaScript Object Notation) file, which was used to
372 implement the auto-completion menus of the interface. In particular, the auto-completion menus were implemented for: the
373 primary sources' titles, the names of authors and of thematic areas, and the corresponding URIs. In the JSON file a mapping
374 between primary sources, authors, thematic areas and URIs was also defined. Therefore, when an expert chooses a title of a
375 primary source, the interface automatically completes the related fields (i.e. author, thematic area, URIs). This functionality
376 allowed the experts to reduce (i) the time for populating the ontology and (ii) the possibility to make mistakes by inserting the
377 data manually. The insertion in the knowledge base of the data extracted from a portion of about 15,000 words of the first
378 commentary analysed, i.e. *Convivio*, was done manually by a single expert, since the tool was under development. We compared
379 the number of errors identified in the manually-inserted data with the number of errors in the data inserted using the population
380 tool for another portion of *Convivio* of the same length. We noticed that the number of errors was about 20% greater in the
381 manually-inserted data than in the other. Furthermore, the manual insertion required about 30% more time than the insertion
382 using the tool.

383 At the beginning of our research, before the development of the Java software that pre-process the texts of the commentaries
384 and of the Web interface used by the experts, we estimated that inserting all data manually (about 15,000 words for *Convivio*),

385 the work of an expert is about 8 person-days (7 hours per day). Using the Java SW and the Web interface the time was reduced to
386 2,5 person-days (7 hours per day).

387 We estimated that, using the tool, the manual work of each expert for inserting all data was about 100 person-days (7 hours per
388 day). In addition, we can compute a further revision of the quality of the inserted data that we can estimate in about 15 person-
389 days (7 hours per day).

390 The data added by the experts are saved in the same XML file that the interface takes as input, enriching it. The XML Schema
391 follows the structure of our ontology. Figure 6 shows a piece of an XML file produced by the tool.

392

393 Insert Figure 6 here

394

395 The XML generated by the tool is loaded into a Java triplifier that converts it to intermediate Java objects. The knowledge
396 stored in the Java objects is finally converted to RDF using the free and open source Java framework Apache Jena²³, which
397 provides an API to write RDF graphs. The resulting graph is automatically stored into a Virtuoso triple store. The RDF file stored
398 in Virtuoso has a size of 29,2 MB and contains 187,328 triples at present.

399 **The DanteSources Web Application**

400 The DanteSources Web application²⁴ allows extracting and displaying the information stored in the knowledge base to support
401 scholars in discovering and exploring Dante's primary sources.

402 In the last year DanteSources had 4,956 unique visitors, for a total of 30,501 visited pages. The majority of the users are from
403 Italy (66%), 7% of the users are from the United States, 3% from Germany, 3% from France, 2% from the U.K. The remaining
404 19% of the users are from different countries in the world. 85% of the users access DanteSources from a desktop computer and
405 15% from a mobile device.

406 *Software Development*

407 DanteSources is developed in Java and JavaServer Pages (JSP). JavaBeans and the Apache Jena library are used to perform
408 SPARQL (Prud'Hommeaux and Seaborne, 2008) queries in order to extract information from the underlying RDF knowledge
409 base stored in Virtuoso. It currently uses Virtuoso version 7.2 and Jena version 3.1.

410 The Web interface is written in HTML5 and CSS3, using JavaScript and Ajax functions. In particular, we used the jQuery
411 JavaScript library and Ajax functions to implement the search menus, which are described in detail in The User Interface
412 subsection. jQuery was also used to implement the tables that show the content of the textual fragments (an example of the table
413 visualising textual fragments is reported in Figure 5).

414 DanteSources shows the knowledge about the primary sources cited by Dante in the form of tables and column bar charts. To
415 implement the charts, we used the Highcharts²⁵ JavaScript library. Highcharts allows exporting the charts in the following
416 formats: PDF, PNG, JPEG, SVG. Furthermore, we implemented an additional JavaScript function allowing users to
417 automatically export and download the data in CSV (Comma Separated Value) format²⁶. This feature is important since it allows
418 scholars to apply further data analyses on raw data, in addition to the ones already provided by the application's queries.
419 DanteSources is provided with a responsive layout that allows an easy reading and navigation with a minimum of resizing,
420 panning, and scrolling across a wide range of devices, from desktop computers to mobile phones (Bartalesi, Meghini, Metilli, and
421 Andriani, 2016).

422 *The User Interface*

423 Currently, DanteSources provides eight different predefined SPARQL queries to extract data. They can be distinguished into
424 three different groups. This distinction is based on the results produced by the queries. Three queries are in the first group, which
425 return knowledge about the distribution of the works, the authors, and the thematic areas mentioned by Dante. Three queries are
426 in the second group, which return the distribution of a particular primary source, a cited author, and a thematic area. Two queries
427 are in the last group, which return the distribution of the three types of reference to primary sources.

428 The three queries belonging to the first group produce column bar charts regarding the distribution of the works, the authors
429 and the thematic areas mentioned by Dante. Through the search form shown in Figure 7, users can choose either one work of
430 Dante or all his works and, in addition, a specific subpart of the work (e.g. a book).

431 Figure 8 shows the chart of the distribution of the ten most-cited primary sources in *Monarchia*. This visualisation allows the
432 scholar to immediately get an idea of the main sources that have influenced Dante in writing a specific work.

433

434 Insert Figure 7 here

435

436 Clicking on the name of each primary source, author and thematic area visualised on the charts, it is possible to receive more
437 information. In particular, clicking on the title of a primary source, the tool shows a table reporting the following details: (i) the
438 book, (ii) the chapter, (iii) the paragraph or poem and (iv) the fragment of the Dante's work in which the primary source is cited,
439 (v) the type of reference, (vi) the reference to a fragment of the primary source cited in the commentary, (vii) the thematic area
440 and (viii) the author of the primary source.

441

442 Insert Figure 8 here

443

444 Insert Figure 9 here

445

446 Clicking on a table row, it is possible to visualise the fragment of Dante's text and the fragment of the corresponding primary
447 source. An example of this table is shown in Figure 9. We are importing in DanteSources the text of the primary sources included
448 in DaMA in XML/TEI format. Up to now, the application includes the text of five primary sources but we are working to add the
449 complete list. Figure 10 shows the chart of the distribution of the ten most-cited authors in *Convivio*.

450 Similarly, as shown in Figure 11, clicking on the name of a cited author, the tool shows a table reporting all the works of that
451 author cited in the Dante's text chosen by the user. Furthermore, for each primary source the tool reports the book, the chapter
452 and the paragraph (or poem) of the Dante's text in which the author is cited. Through this table, the scholar can directly access
453 the paragraph in a Dante's work in which the poet refers to a specific primary source.

454

455 Insert Figure 10 here

456

457

458 Insert Figure 11 here

459

460 Figure 12 shows the chart of the ten thematic areas with the highest distribution cited in *De vulgari eloquentia*.

461

462 Insert Figure 12 here

463

464 Clicking on one name of the thematic areas, the user obtains the following information: the primary sources included in that area,
465 their authors and the book, chapter and paragraph (or poem) of the Dante's work where the thematic area is cited. Figure 13
466 shows the first rows of the table regarding the thematic area Scholasticism in *De vulgari eloquentia*. Using this search, a scholar
467 can immediately access all the primary sources belonging to a specific thematic area.

468

469

470 Insert Figure 13 here

471

472 The three queries of the second group allow visualising the distribution of a particular primary source, a cited author or a
473 thematic area respectively, in the form of charts. In order to improve the usability of the search, the application provides two
474 different search forms for these queries: (i) an autocomplete menu where the user can type the title of the cited work, the author,
475 or the thematic area; (ii) an alphabetically ordered list where the user can select the title of the primary source, the name of the
476 author or the thematic area.

477 The distribution data are available for both an entire Dante's work, e.g. *Monarchia*, and its subparts, e.g. books, chapters. By
478 clicking on one bar in the chart, the application allows to visualise details about the subparts of one work. Figure 14 shows the
479 distribution of Aristotle's *Nicomachean Ethics* in Dante's works and in some of their subparts.

480

481 Insert Figure 14 here

482

483 The two queries of the last group allow visualising the distribution of the three types of reference to primary sources: explicit,
484 strict and generic, focusing either on Dante's works or on a single primary source.

485 Figure 15 shows the distribution of the three types of reference in the works of Dante. By clicking on the bars of the chart, it is
486 possible to visualise the data regarding the distribution of the types of reference also for the subparts of Dante's works, i.e. books
487 and chapters.

488

489 Insert Figure 15 here

490

491 Figure 16 reports the distribution of the three types of reference for the works of Seneca cited by Dante in his texts. Having the
492 type of reference for each author is a useful indication for the scholar to understand the influence of that author on Dante's work
493 (e.g. several explicit references to a particular author indicate that that author had a significant influence for the poet).

494

495 Insert Figure 16 here

496

497 Finally, DanteSources makes available a SPARQL endpoint so that any user can make her/his own query to the knowledge
498 base.

499 Analysing the logs of the Apache Web server, we discovered that the most performed query is the one by primary source
500 (37%), followed by the one showing all cited authors in a specific Dante's work (20%). The third is the one visualising the
501 distribution of a specific cited author in all Dante's works (16%).

502 **Usability Evaluation**

503 We conducted a survey on the interaction with DanteSources via PCs and mobile devices to collect suggestions and comments on
504 possible usability issues observed by the users. For a detailed description of the questions and the answers of our usability test,
505 and for an aggregate analysis of the answers see Bartalesi et al., 2016²⁷.

506 The survey included 28 questions about the main aspects regarding the usability of the Web application. We collected
507 suggestions and comments from 26 users. 25% of the users were students, 20% were researchers, 12% were university
508 professors, 4% were school teachers and 37% of the users are generic users. Likert 5-scale values from 1 (very negative) to 5
509 (very positive) were used to express the opinion.

510 Positive responses on the general interaction have appeared from the survey: 85% of the users expressed a distinctly positive
511 opinion (score 4 and 5). About 80% reported a satisfactory opinion on the navigation (score 4 and 5) and on the readability of
512 charts and tables (score 4 and 5) as well. 100% of the users who accessed DanteSources from a mobile device (5 users) declared
513 that the interaction was easy and clear (score 4 and 5). On the other hand, the main problems identified concern: (i) the presence
514 of a large blank space between the search forms and the results, (ii) the hidden position of the explanation of the charts'
515 functionalities, (iii) the lack of text visualisation for the cited fragments and for the entire texts of the primary sources and
516 Dante's works. The users also suggested to enrich the knowledge base with other Dante's works, especially with the *Divine*
517 *Comedy*. We are working to modify and enrich DanteSources taking into account the difficulties highlighted by the users and the
518 given suggestions.

519 The scholars who took part in the evaluation confirmed that, to the best of their knowledge, this is the first time that the
520 information about Dante's primary sources is available in digital format. This availability improves and makes the searches by
521 Dante's experts more efficient. Indeed, in the commentaries in paper format, the knowledge on the primary sources is expressed
522 in natural language. This situation limits scholars in their advances insofar as it prevents automatic inferences of new information
523 that may be useful for their studies, e.g. regarding the primary sources, the inferences can concern the total amount of references
524 to a certain work. Eventually, the digitisation of the knowledge about primary sources allows scholars to have a comprehensive
525 overview of the data that can be easily exported to make further analyses.

526 **Conclusions and Future Works**

527 In this paper we have presented the methodology and the development process we adopted to create DanteSources, a Web
528 application that allows free access to the knowledge on Dante Alighieri's primary sources. In order to represent this knowledge,

529 we developed an ontology, expressed in RDF/S, on top of which DanteSources was created. The main advantages of using an
530 ontology instead of a traditional database are the following: (i) it allows researchers to add classes and properties thereby refining
531 the ontology, (ii) the ontology can be linked to other ontologies to extend the represented domain, (iii) any user can download
532 and use the knowledge base freely, using the paradigm of Linked Data. Furthermore, this is the first time that the knowledge
533 about Dante's primary sources is coded in digital format. This availability allows improving and making more efficient the
534 scholars' searches. Eventually, the digitisation of the knowledge about primary sources and its visualisation in form of charts and
535 tables allow the scholars to have a complete overview of the data. This simplifies the study on the evolution of Dante's cultural
536 background in time, also related to the diverse stages of his biography.

537 Summarising, we have developed a methodology and a suite of tools to represent the knowledge about primary sources cited in
538 literary texts. We have created a semantic model and an ontology for representing such knowledge, a semi-automatic tool to
539 populate the ontology, and a Web application that allows exploring the knowledge base. Our ontology, methodology and tools
540 are not created *ad hoc* for being exclusively applied to Dante's works, but they are easily reusable in order to represent the
541 knowledge about the primary sources of other authors of the Italian and international literature. Furthermore, they could also be
542 applied to other fields where it is important to represent primary sources, for instance scientific literature.

543 As future works, we envision: (i) further developing to the ontology in order to make it more adaptable to different kinds of
544 textual works; (ii) refining the population tool in order to make it more efficient, e.g. by making it easier for the user to import
545 URIs from other repositories such as Wikidata²⁸ or VIAF²⁹; (iii) building new visualisations to present the extracted data to the
546 user from different points of view; (iv) updating the software architecture to adopt newer technologies, such as JavaServer Faces
547 (JSF). We also believe that our method could lead to the creation of entirely new applications, such as a new generation of
548 digital critical editions built on top of structured knowledge.

549

550

551 **Notes**

552 ¹ Up to now, two theses have been written using DanteSources:

553 Mussati, V. Le fonti della Vita Nuova. Trattamento informatico e formazione della "biblioteca di Dante". University of Pisa,

554 2014.

555 Gibert, Z. Jenseitsvisionen im islamischen Orient und christlichen Okzident: Jenseitreisen im Islam und Dantes Göttliche

556 Komödie im Vergleich. Georg-August-Universität Göttingen, 2015

557 ² <https://www.gnu.org/licenses/gpl-3.0.en.html>

558 ³ <http://dante.dartmouth.edu>

559 ⁴ <http://dantelab.dartmouth.edu>

560 ⁵ <http://etcweb.princeton.edu/dante/>

561 ⁶ <http://www.worldofdante.org>

562 ⁷ <http://digitaldante.columbia.edu>

563 ⁸ <http://danteonline.it>

564 ⁹ <http://www.perunaenciclopediadantescadigitale.eu:8080/dantesearch/>

565 ¹⁰ <http://perunaenciclopediadantescadigitale.eu/istidama/>

566 ¹¹ For the full list of the analysed ontologies see Bartalesi and Meghini (2015)

567 ¹² <http://dublincore.org/documents/dces/>

568 ¹³ <http://www.foaf-project.org>

569 ¹⁴ <http://vocab.org/bio/>

570 ¹⁵ <http://vocab.org/relationship/>

571 ¹⁶ <http://www.sparontologies.net>

572 ¹⁷ <http://perunaenciclopediadantescadigitale.eu/dantesources/schema/dante.rdf>

573 ¹⁸ <https://jquery.com>

574 ¹⁹ <https://jqueryui.com>

575 ²⁰ <https://www.w3.org/TR/IndexedDB-2/>

576 ²¹ <http://dbpedia.org>

577 ²² <http://thes.bncf.firenze.sbn.it>

578 ²³ <https://jena.apache.org/index.html>

579 ²⁴ DanteSources was the winner of the Digital Humanities Awards 2015 in the category “Best DH Tool or Suite of Tools”,
580 <http://dhawards.org/dhawards2015/>
581 ²⁵ <http://www.highcharts.com>
582 ²⁶ <http://www.w3.org/2013/05/lcsv-charter.html>
583 ²⁷ http://link.springer.com/chapter/10.1007/978-3-319-39513-5_18
584 ²⁸ <https://wikidata.org>
585 ²⁹ <https://viaf.org>

587 **Captions**

- 588 Fig. 1 Example of commentary to Dante's *Convivio*. The highlighted note states that the fragment "si manuca" refers to Psalm 77,
589 the Gospel of John, and the *Enarrationes in Psalmos*
- 590 Fig. 2 Example of commentary to Dante's *Monarchia*. The format of the commentary is different from that of *Convivio*
- 591 Fig. 3 The main phases of our approach: from the textual knowledge to the web application
- 592 Fig. 4 An example of the representation of a note in our ontology
- 593 Fig. 5 The interface of the population tool
- 594 Fig. 6 A piece of XML file produced by the population tool
- 595 Fig. 7 The search form for the first group of queries
- 596 Fig. 8 The ten primary sources with the highest distribution in *Monarchia*
- 597 Fig. 9 The first rows of the table regarding Virgil's Aeneid cited in *Monarchia*
- 598 Fig. 10 The ten cited authors with the highest distribution in *Convivio*
- 599 Fig. 11 The first rows of the table regarding Saint Augustine cited in *Convivio*
- 600 Fig. 12 The ten thematic areas with the highest distribution cited in *De vulgari eloquentia*
- 601 Fig. 13 The first rows of the table regarding the thematic area Scholasticism in *De vulgari eloquentia*
- 602 Fig. 14 The distribution of *Nicomachean Ethics* in Dante's works (a), in the three books of *Monarchia* (b) and in the chapters of
603 the first book of *Monarchia* (c)
- 604 Fig. 15 The distribution of the three types of reference in Dante's works
- 605 Fig. 16 The distribution of the three types of reference for the Seneca's works

607 **References**

- 608 Alighieri, D. (1980). In D. De Robertis (eds) *Vita nuova*, Milano, Napoli, Ricciardi.
- 609 Alighieri, D. (2011a). De vulgari eloquentia. In C. Giunta, G. Gorni, and M. Tavoni (eds), *Opere*. Vol. I: Rime, Vita Nova, De
610 vulgari eloquentia. Mondadori, Milano.
- 611 Alighieri, D. (2011b). Vita Nova. In C. Giunta, G. Gorni, and M. Tavoni (eds), *Opere*. Vol. I: Rime, Vita Nova, De vulgari
612 eloquentia. Mondadori, Milano.
- 613 Alighieri, D. (2011c). Rime. In C. Giunta, G. Gorni, and M. Tavoni (eds), *Opere*. Vol. I: Rime, Vita Nova, De vulgari eloquentia.
614 Mondadori, Milano.
- 615 Alighieri, D. (2104a). Convivio. In G. Fioravanti, C. Giunta, D. Quaglioni, C. Villa, and G. Albanese (eds), *Opere*. Vol. II:
616 Convivio, Monarchia, Epistole, Egloghe. Mondadori, Milano.
- 617 Alighieri, D. (2014b). Monarchia. In G. Fioravanti, C. Giunta, D. Quaglioni, C. Villa, and G. Albanese (eds), *Opere*. Vol. II:
618 Convivio, Monarchia, Epistole, Egloghe. Mondadori, Milano.
- 619 Altman, M. (2006). An overview of digital libraries: issues and trends, Academic Research Microsoft, Cambridge, 1-10.
- 620 Bartalesi, V., and Meghini, C. (2015) Using an ontology for representing the knowledge on literary texts: The Dante Alighieri
621 case study, *Semantic Web Journal*, (Preprint), 1-10.
- 622 Bartalesi, V., Meghini, C., Metilli, D., and Andriani, P. (2016). Usability evaluation of the digital library DanteSources,
623 *Proceedings of International Conference on Human-Computer Interaction* (pp. 191-203), July 2016. LNCS, Springer
624 International Publishing.
- 625 Ciccarese, P., Ocana, M., Castro, L. J. G., Das, S., and Clark, T. (2011). An open annotation ontology for science on web 3.0,
626 *Journal of biomedical semantics*, 2(2), 1.
- 627 Constantin, A., Peroni, S., Pettifer, S., Shotton, D., and Vitali, F. (2016). The document components ontology (DoCO), *Semantic
628 Web*, 7(2), 167-181.
- 629 Dalton, M. S., and Charnigo, L. (2004). Historians and their information sources, *College & Research Libraries*, 65(5), 400-425.
- 630 Doerr, M. (2002). The CIDOC CRM: An ontological approach to semantic interoperability of metadata, *AI Magazine*, Special
631 Issue, 24(3), 75-92.
- 632 Doerr, M., Gradmann, S., Henniecke, S., Isaac, A., Meghini, C., and van de Sompel, H. (2010). The Europeana Data Model
633 (EDM), *Proceedings of World Library and Information Congress: 76th IFLA general conference and assembly* (pp. 10-15),
634 August 2010.
- 635 Erling, O., and Mikhailov, I. (2009). RDF support in the Virtuoso DBMS, In *Networked Knowledge-Networked Media* (pp. 7-
636 24). Springer Berlin Heidelberg.

637 Gehrke, S., Frunzeanu, E., Charbonnier, P., & Muffat, M. (2015). *Biblissima's Prototype on Medieval Manuscript Illuminations*
638 *and their Context*, Proceedings of the SW4SHD@ESWC (pp. 43-48).

639 González-Blanco, E., Seláf, L., Del Rio Riande, M.G., Martínez Cantón, C.I., Martos Pérez, M.D. (2014). *Building a metrical*
640 *ontology as a model to link digital poetic repertoires*, Proceedings of the Digital Humanities Conference.
641 <http://dharchive.org/paper/DH2014/Paper-674.xml> (accessed 12 July 2017)

642 Hedges, M., Jordanous, A., Faith Lawrence, K., Tupman, C. (2012). *Exploring manuscripts: sharing ancient wisdoms across the*
643 *semantic Web*, Proceedings of WIMS 2012.

644 Hellqvist, B. (2010). *Referencing in the humanities and its implications for citation analysis*, Journal of the American Society for
645 *Information Science and Technology*, 61(2), 310-318.

646 Hendler, J. (2012). *A primer on machine readability for online documents and data*.
647 <https://www.data.gov/developers/blog/primer-machine-readability-online-documents-and-data> (accessed 12 July 2017)

648 Hyvönen, E., Heino, E., Leskinen, P., Ikkala, E., Koho, M., Tamper, M., ... & Mäkelä, E. *Publishing Second World War History*
649 *as Linked Data Events on the Semantic Web*, Proceedings of the Digital Humanities Conference 2016 (pp. 571-573), July 2016.

650 Hollander, R. (1989). *The Dartmouth Dante Project*, *Quaderni d'italianistica*, 10(1-2), 287-298.

651 Hollander, R. (2013). *The Princeton Dante Project*, *Humanist Studies & the Digital Age*, 3(1), 53-59.

652 Lana, M., Ciotti, F., Magro, D., Peroni, S., Tomasi, F., and Vitali, F. (2014). *Annotating texts with ontologies, from geography to*
653 *persons and events*, Poster at Digital Humanities Conference 2014.

654 Malkmus, D. (2010). "Old stuff" for new teaching methods: outreach to history faculty teaching with primary sources, *Libraries*
655 *and the Academy*, 10(4), 413-435.

656 Manola, F., Miller, E., and McBride, B. (2004). *RDF primer*, W3C recommendation, 10(1-107), 6.

657 Miles, A., Matthews, B., Wilson, M., and Brickley, D. (2005). *SKOS core: simple knowledge organisation for the web*,
658 *Proceedings of International Conference on Dublin Core and Metadata Applications* (pp. 3-10), September 2005.

659 Mishra, R. B., and Kumar, S. (2011). *Semantic web reasoners and languages*. *Artificial Intelligence Review*, 35(4), 339-368.

660 Niccolucci, F., and D'Andrea, A. (2006). *An ontology for 3D cultural objects*, Proceedings of the 7th International conference on
661 *Virtual Reality, Archaeology and Intelligent Cultural Heritage* (pp. 203-210), October 2006. Eurographics Association.

662 Parker, D. (2001). *The World of Dante: a hypermedia archive for the study of the Inferno*, *Literary and Linguistic Computing*,
663 16(3), 287-297.

664 Peroni, S., and Shotton, D. (2012). *FaBiO and CiTO: ontologies for describing bibliographic resources and citations*, *Web*
665 *Semantics: Science, Services and Agents on the World Wide Web*, 17, 33-43.

666 Prud'Hommeaux, E., and Seaborne, A. (2008). *SPARQL query language for RDF*, W3C recommendation, 15.

667 Riva, P., Doerr, M., and Zumer, M. (2008). FRBRoo: enabling a common view of information from memory institutions,
668 Proceedings of World Library and Information Congress: 74th IFLA General Conference and Council, August 2008.

669 Sanderson, R., Ciccarese, P., and Van de Sompel, H. (2013). Designing the W3C open annotation data model, Proceedings of the
670 5th Annual ACM Web Science Conference (pp. 366-375), May 2013. ACM.

671 Shotton, D. (2010). CiTO, the citation typing ontology, *Journal of Biomedical Semantics*, 1(1), 1.

672 Van Peteghem, J. (2015). Digital readers of allusive texts: ovidian intertextuality in the 'Commedia' and the digital concordance
673 on 'Intertextual Dante', *Humanist Studies & the Digital Age*, 4(1), 39-59.

674 Wagner, A., and Caesar, I. (2016). The School of Salamanca on the Semantic Web, Proceedings of Digital Humanities
675 Conference 2016 (pp. 704-705), July 2016.

676 World Wide Web Consortium. (2014). RDF 1.1 concepts and abstract syntax.