Umanistica Digitale - ISSN 2532-8816 - n.1, 2017
V. Bartalesi et al. - DanteSources: a Digital Library for Studying Dante Alighieri's Primary Sources
DOI: https://doi.org/10.6092/issn.2532-8816/7250

DanteSources: a Digital Library for Studying Dante Alighieri's Primary Sources

Valentina Bartalesi[1], Carlo Meghini[1], Daniele Metilli[1], Paola Andriani[2], e Mirko Tavoni[2]

[1] Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo", Consiglio Nazionale delle Ricerche (CNR)
[2] Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa
valentina.bartalesi@isti.cnr.it, carlo.meghini@isti.cnr.it, daniele.metilli@isti.cnr.it, paola.andriani@gmail.com, mirko.tavoni@gmail.com

Abstract. We present DanteSources, a focused Digital Library that aims at building innovative services supporting scholars in creating, evolving and consulting a digital encyclopaedia of Dante Alighieri's works. DanteSources is endowed with Web services that allow extracting and visualizing data about Dante Alighieri's primary sources, i.e. other authors' works which Dante referred to in his own works. The heart of the digital library is a knowledge base built on an ontology that was developed in the RDF/S language, representing Dante's works and the knowledge about them, extracted from some authoritative commentaries. Finally, on top of this graph, we developed DanteSources, which allows users to extract and display the knowledge stored in the knowledge base in the form of charts and tables. The methodology and the tool we developed are easily reusable to represent the knowledge about primary sources of other authors of the Italian or the international literature.

Abstract. In questo articolo presentiamo DanteSources, una Digital Library il cui scopo è fornire servizi innovativi di supporto alla creazione, evoluzione e consultazione dell'enciclopedia digitale delle fonti di Dante Alighieri. DanteSources è dotato di servizi che consentono l'estrazione e la visualizzazione delle fonti primarie di Dante, i.e. opere di altri autori a cui Dante ha fatto riferimento nelle proprie opere. Il cuore del progetto è una base di conoscenza costruita attorno ad uno Schema RDF rappresentante le opere di Dante e le relative informazioni estratte da autorevoli commentari. A partire dal grafo è stato sviluppata l'applicazione DanteSources, che permette agli utenti di estrarre e visualizzare i contenuti sotto forma di grafici e tabelle. La metodologia e il tool realizzato sono di facile riutilizzo per la rappresentazione di conoscenza relativa a fonti primarie, siano esse italiane o internazionali.

Keywords: Dante Alighieri, Digital Libraries, RDF/S, Semantic Web

1 Introduction

One kind of knowledge that the scholars consider essential for of the study of old literary works is the analysis of their primary sources, i.e. the works of other authors that an author makes reference to in his/her texts. Usually, this knowledge is reported in modern paper commentaries, which the readers use to understand several aspects of the literary works. Having this knowledge available in a machine-readable form instead of in natural language could allow using computers to make inferences that could facilitate and improve the study to scholars, e.g. automatic calculation of the distribution of the primary sources and of the authors cited in an old text.

Our research is part of the *Towards a Digital Dante Encyclopedia* project, an Italian National Research project supporting scholars in formally expressing the knowledge about primary sources presented in Dante Alighieri's works and more generally in literary texts. In order to represent this information we (i) created an ontology providing a formal representation of the terms required for expressing knowledge about the primary sources, and (ii) develop a semantic Digital Library (DL) based on the ontology that allows scholars to make inferences on the collected data and visualize them in a friendly and easy way.

In the Digital Humanities field, there are ontologies focusing on different aspects of textual information but, up to now, an ontology for representing knowledge about primary sources of literary texts has not been developed yet. For this reason, we created an ontology in the RDF/S language [13] to represent Dante's works and the knowledge about their primary sources. To develop the ontology, we reused existing vocabularies where possible, in order to maximize its interoperability, adding our own classes and relationships for the representation of the categories of knowledge that were not addressed by the existing ontologies. For a detailed description of the ontology we refer to [7].

Then, we populated the ontology with the knowledge extracted from some authoritative commentaries, obtaining an RDF knowledge base. On top of this knowledge base, we developed DanteSources,[1] a focused Digital Library that allows extracting and visualizing information on Dante Alighieri's primary sources in form of charts and tables and in CSV format. In this paper we describe the DL DanteSources and the methodology we used to develop it.

The paper is structured as follows: Section 2 reports a review of related works; in Section 3 the functionalities of DanteSources are described in detail; in Section 4 conclusions are reported.

2 Related Works

In the Semantic Data Modeling field, several ontologies have been developed in order to represent textual knowledge. Up to now, a formal ontology for representing knowledge about the primary sources of literary texts has not appeared yet. So we developed a new ontology taking into account the existing ones, reusing their classes and properties in order to improve its interoperability. In particular, we reused, among others, some classes and prop-

[1] http://dantesources.org

erties from authoritative ontologies like FRBRoo [15], SKOS [14], Dublin Core,[2] DoCo [16], and the CIDOC CRM [8].

In the Digital Humanities field, several projects about Dante Alighieri and his works have been published online. Most of them offer digital editions of Dante's works with commentaries, textual search, and multimedia resources. For example, the *Dartmouth Dante Project* (DDP)[3] was launched in 1985 with the aim of publishing the full searchable text of the *Divine Comedy* and several of its commentaries in digital format [10]. On top of the DDP in 2013 the web application Dante Lab[4] was developed. This application allows the concurrent visualization of the original text of the Comedy, some translations and more than 75 commentaries. The *Princeton Dante Project*[5] includes the full searchable text of the *Divine Comedy*, Dante's minor works, several commentaries and multimedia resources [11]. *Digital Dante*[6] offers the full text of all Dante's works with commentaries, illustrations and recorded readings. It also features a subproject called *Intertextual Dante*,[7] the first digital attempt at connecting passages of Dante's works with the corresponding fragments of cited primary sources. However, the project is limited to the works of the Roman poet Ovid [17]. Recently, two Italian research projects have been published: *DanteSearch*[8] and *Dante Medieval Archive* (DaMA).[9] *DanteSearch* is a complete lemmatization, grammatical and syntactic annotation of Dante's works allowing users to perform morphological and syntactic queries on the full text of the author's works. On the other hand, DaMA is a digital archive containing the full text of Dante's works, commentaries and several primary sources in XML-TEI[10] format.

With the advent of the Semantic Web, several projects were started with a focus on the application of the new semantic technologies (RDF, OWL)[11] to the Humanities. For instance, [9] describes a semantic model to connect repertoires of poetic writings. [18] presents an ontology for the 3D visualization of cultural heritage; [12] describe an ontology for annotating geographical places in texts. In this context, we developed DanteSources DL, which uses the technologies of the Semantic Web to represent the knowledge included in the works of Dante Alighieri, focusing on primary sources.

3 The Digital Library DanteSources

First of all, in order to develop the DL DanteSources, we analyzed the representation requirements about citations of primary sources, in collaboration with a Dante's expert. Then, on the basis of this analysis, we created an ontology responding to the requirements, providing a formal representation of the terms required for expressing knowledge about the

[2] http://dublincore.org/documents/dcmi-type vocabulary/index.shtml
[3] `http://dante.dartmouth.edu`
[4] `http://dantelab.dartmouth.edu`
[5] `http://etcweb.princeton.edu/dante/`
[6] `http://digitaldante.columbia.edu`
[7] http://digitaldante.columbia.edu/digital-projects/intertexual-dante/
[8] http://www.perunaenciclopediadantescadigitale.eu:8080/dantesearch/
[9] http://perunaenciclopediadantescadigitale.eu/istidama/
[10] http://www.tei-c.org/index.xml
[11] https://www.w3.org/TR/owl2-overview/

primary sources, using the RDF/S Semantic Web language. After this phase, we developed a semi-automatic tool for populating the ontology and storing the resulting data in a knowledge base. Finally, we developed DanteSources, a web application running on top of the knowledge base that allows making inferences on the contents of the knowledge base and visualizes them in a friendly and easy way, using charts and tables.

The authoritative commentaries used to populate the ontology are the following: *Vita Nova* [5], *Vita Nuova* [6], *De vulgari eloquentia* [2], *Convivio* [1], *Monarchia* [3], and Rime [4]. DanteSources is developed in Java using JavaScript and Ajax functions. The DL extracts knowledge by running SPARQL queries on the RDF knowledge base. DanteSources shows the knowledge about the primary sources cited by Dante in the form of tables and column bar charts. In particular, we used the Highcharts[12] JavaScript library to implement these charts. Highcharts allows exporting the charts in various well-known formats: PDF, PNG, JPEG, SVG. Furthermore, we implemented an additional JavaScript function allowing users to automatically export and download all the data in CSV format.[13] This feature was particularly appreciated by the scholars since it allows obtaining and managing raw data in a simple and standard format, in order to apply further data analyses in addition to the ones already provided by the DL.

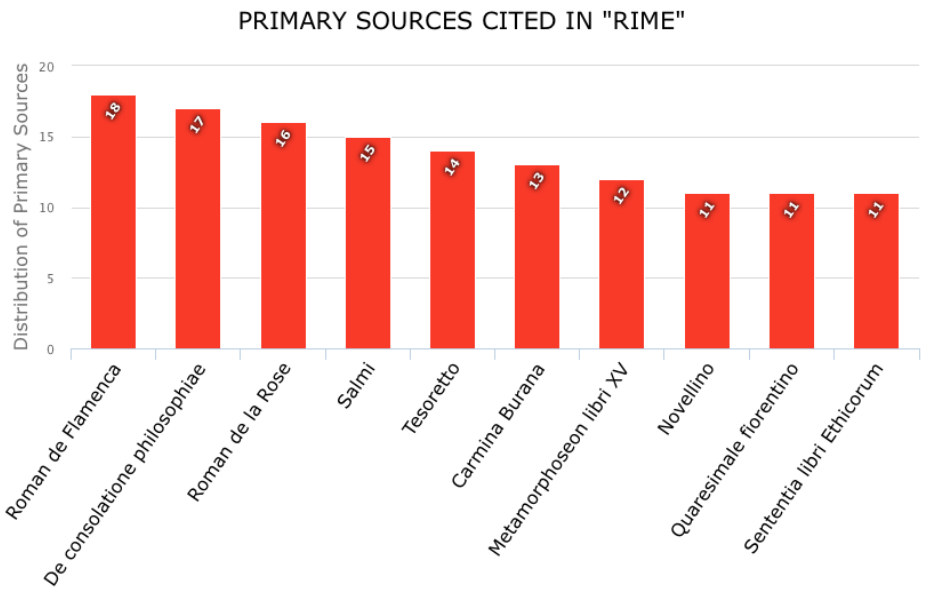

Fig. 1. Distribution of the ten most-cited primary sources in Rime

Currently, DanteSources makes available eight different predefined SPARQL queries to extract data. They can be distinguished into three different groups. The first group includes

[12] http://www.highcharts.com/
[13] http://www.w3.org/2013/05/lcsv-charter.html

three queries. In order to make these queries, a search form allows users to choose either one Dante's work or all his works and, in addition, a specific subpart of the work (e.g. a book). The queries produce column bar charts regarding the distribution of the works, the authors and the thematic areas cited by Dante. For example, Fig. 1 shows the chart of the distribution of the ten most-cited primary sources in Dante's work *Rime*.

Further information about primary sources, authors and thematic areas can be visualized by clicking on their names on the charts. In particular, for each primary source the DL shows a table reporting information about: (i) the book, (ii) the chapter, (ii) the paragraph or verse and (iv) the fragment of the Dante's work in which the primary source is cited, (v) the type of reference, (vii) the reference to a fragment of the primary source cited in the commentary, (viii) the thematic area and (ix) the author of the primary source. An example of this table, related to Brunetto Latini's work *Tesoretto*, is shown in Fig. 2.

Dante's work	Book	Chapter *or* Poem	Paragraph *or* Verse	Fragment of Dante's work	Reference type	Fragment of the primary source	Thematic area of the primary source	Author of the primary source
Rime	1	24	6	MI RISPOSE ... GRECO	Strict reference	Brunetto Latini, Tesoretto 2582 «in greco salisti» 'insuperbisti';	Poesia didascalica italiana	Brunetto Latini
Rime	1	27	54-57	VILLAN DILETTO	Strict reference	esoretto 1452-6 «e molto m'è rubello / chi dispende in bordello / e va perdendo 'l giorno / in femine d'intorno»	Poesia didascalica italiana	Brunetto Latini
Rime	1	27	1-133	POSCIA CH'AMOR ... TUTTI CONTRA	Strict reference	Tesoretto – dove non a caso si am- maestra non «mastro Burnetto» ma «un bel cavalero» (v. 1367)	Poesia didascalica italiana	Brunetto Latini
Rime	1	27	22	CAPERE ... STANNO	Strict reference	Brunet- to Latini (Tesoretto 1720-3 «chi non dura fatica / sì che possa valere, / non si creda capere / tra gli uomini valenti»	Poesia didascalica italiana	Brunetto Latini

Fig. 2. Table listing the first four citations of Brunetto Latini's Tesoretto in Rime

Similarly, when clicking on the name of a cited author, a table reporting all the works of that author cited in the Dante's text chosen by the user is shown. Furthermore, for each primary source the DL reports the book, the chapter (or poem) and the paragraph (or verse) of the Dante's text in which the author is cited. Fig. 3 shows an example of table related to cited author Brunetto Latini.

The same happened for the thematic areas, indeed clicking on one name the user obtains the following information: the primary sources included in that area, their authors and the

Dante's Work	Book	Chapter / Poem	Paragraph / Verse	Primary Source
De vulgari eloquentia	1	6	3	Trésor
De vulgari eloquentia	1	8	3	Trésor
De vulgari eloquentia	1	8	3	Trésor
De vulgari eloquentia	1	8	6	Trésor
De vulgari eloquentia	1	8	6	Trésor
De vulgari eloquentia	1	10	2	Trésor
De vulgari eloquentia	1	10	2	Trésor
De vulgari eloquentia	1	10	5	Trésor
De vulgari eloquentia	1	10	6	Trésor
De vulgari eloquentia	1	13	1	S'eo son distretto inamoratamente

Fig. 3. Table listing the first ten citations of Brunetto Latini in De vulgari eloquentia

book, chapter (or poem) and paragraph (or verse) of the Dante's work where the thematic area is cited. Fig. 4 shows an example of table related to thematic area Aristotelianism.

Dante's Work	Book	Chapter / Poem	Paragraph / Verse	Author	Primary Source
Convivio	1	1	1	Aristotele	Metaphysica
Convivio	1	1	1	Aristotele	Metaphysica
Convivio	1	1	9	Aristotele	Ethica Nicomachea
Convivio	1	3	5	Aristotele	Meteorologica
Convivio	1	4	3	Aristotele	Rhetorica
Convivio	1	4	6	Aristotele	Rhetorica
Convivio	1	5	11	Aristotele	Ethica Nicomachea
Convivio	1	5	11	Aristotele	Ethica Nicomachea
Convivio	1	5	14	Aristotele	Metaphysica
Convivio	1	6	5	Aristotele	Ethica Nicomachea

Fig. 4. Table listing the first ten citations of thematic area Aristotelianism in the first book of Monarchia

The three queries of the second group allow visualizing several charts that report the distribution of a particular primary source, a cited author or a thematic area respectively.

The data regarding the distributions are not only available for an entire Dante's work, like *Convivio*, but also for its subparts like books, chapters or poem. Indeed, by clicking on

one bar in the chart representing the distribution of the information onto a Dante's work, it is possible to visualize information about its subparts. For example, it is possible to visualize the distribution of the selected knowledge in all three books of Dante's *Monarchia* and also in each chapter of the first book of *Monarchia*. Fig. 5 shows the distribution of a primary source on three different levels: all Dante's works (a), one particular Dante's work (b), a subpart of a Dante's work (c).

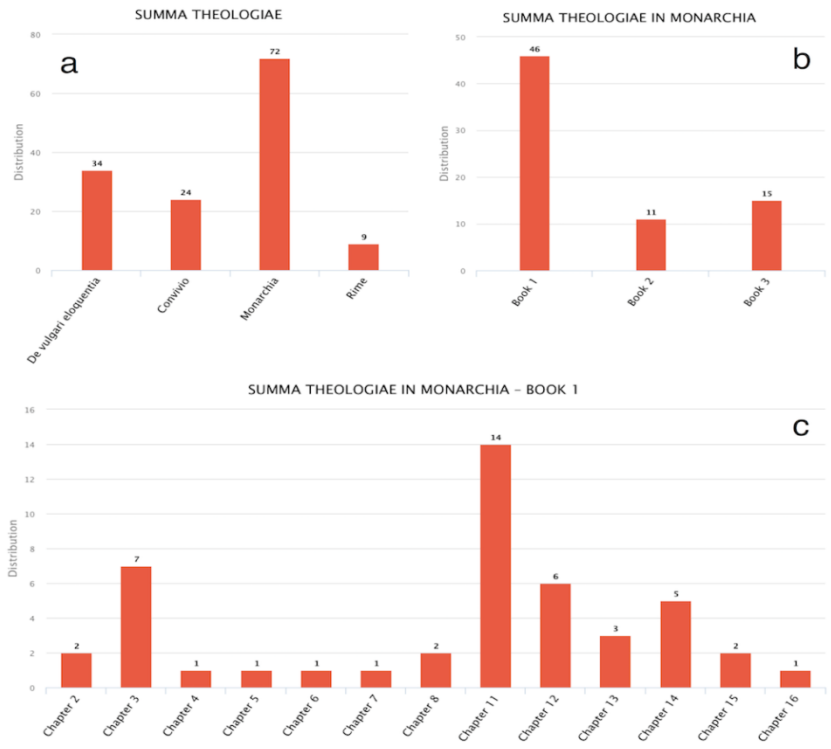

Fig. 5. The distribution of Summa Theologiae in all Dante's works (a), in the three books of Monarchia (b) and in the chapters of the first book of Monarchia (c).

The last group includes two queries that visualize the distribution of the three types of reference to primary sources: explicit, strict and generic, focusing on either on Dante's works or on all primary sources written by a specific author. Indeed, with the support of Dante's experts, we defined three types of reference:

1. *explicit*, if the reference is explicitly made by Dante, e.g. "As the Philosopher says at the beginning of the First Philosophy", where the Philosopher is Aristotle and the First Philosophy is *Metaphysics*;

2. *strict*, if the reference is indicated by a scholar and refers to a specific work, e.g. "SI MANUCA: it is the bread of the angels, the *manna* as called in the Old Testament (Ps. 77, 25 *Panem angelorum manducavit homo*)";

3. *generic*, if the reference is indicated by a scholar, and refers to a concept (e.g. Medieval comments to Aristotle's works).

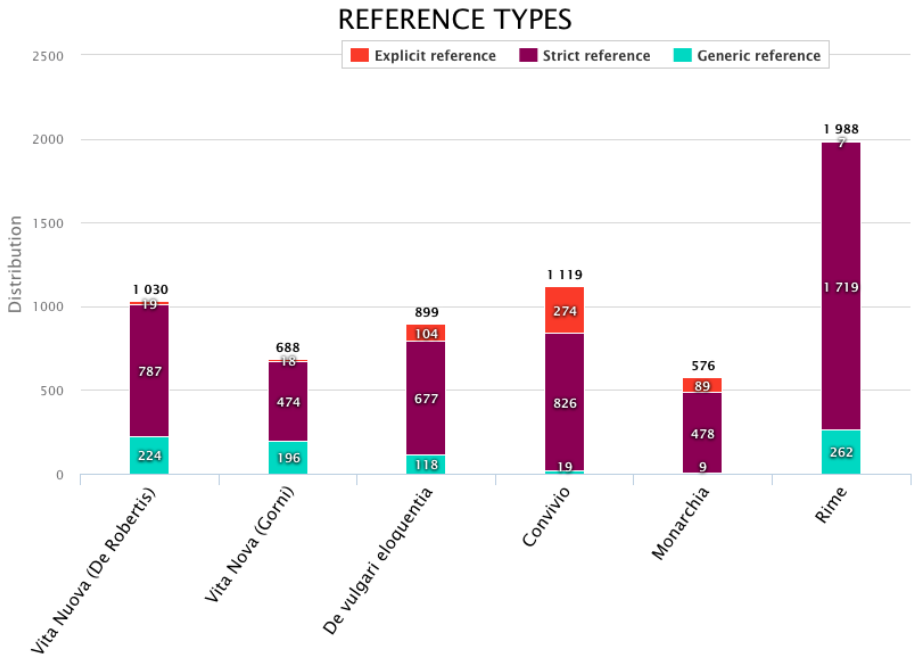

Fig. 6. The distribution of the three types of reference in Dante's works

Fig. 6 shows the distribution of the three types of reference in a specific primary source. By clicking on the bars of the chart, the data regarding the distribution of the types of reference are available both for an entire Dante's work and for its subparts, i.e. books and chapters. Fig. 7 shows the distribution of the three types of reference in all primary sources written by a specific author, Albertus Magnus.

4 Conclusions

In this paper we have presented DanteSources, a Digital Library that allows free access to the knowledge on Dante Alighieri's primary sources. In order to represent this knowledge, we developed an ontology, expressed in RDF/S, on top of which DanteSources was created. The main advantages of using an ontology instead of a traditional database are the following: (i) it allows researchers to add classes and relationships thereby refining the ontology, (ii)

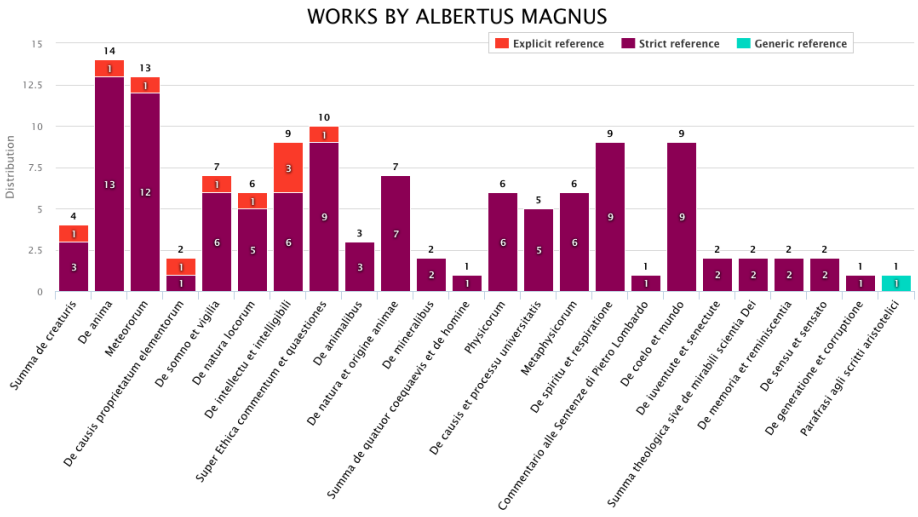

Fig. 7. Distribution of the three types of reference in all primary sources authored by Albertus Magnus

the ontology can be linked to other ontologies to extend the represented domain, (iii) any user can download and use the knowledge base freely, using the paradigm of Linked Data.

Furthermore, this is the first time that the knowledge about Dante's primary sources is coded in digital format. This availability allows improving and making the scholars' searches more efficient. Eventually, the digitization of the knowledge about primary sources and its visualization in form of charts and tables allow the scholars to have a complete overview of the data. This simplifies the study by the scholars on the evolution in time of Dante's cultural background, also in relation with the diverse stages of his biography.

The ontology, the tool to populate it and the developed DL are easily and freely reusable in order to represent the knowledge about primary sources of other authors of the Italian and international literature.

DanteSources was the winner of the Digital Humanities Awards 2015[14] in the category *Best DH Tool or Suite of Tools*.[15]

References

1. Alighieri, Dante. "Convivio." In *Opere*, a cura di Gianfranco Fioravanti, Vol. 2. Milano: Mondadori, 2014.
2. Alighieri, Dante. "De Vulgari Eloquentia." In *Opere*, a cura di Mirko Tavoni, Vol. 1. Milano: Mondadori, 2011.
3. Alighieri, Dante. "Monarchia." In *Opere*, a cura di Diego Quaglioni, Vol. 2. Milano: Mondadori, 2014.
4. Alighieri, Dante. "Rime." In *Opere*, a cura di Claudio Giunta, Vol. 1. Milano: Mondadori, 2011.

[14] http://dhawards.org/dhawards2015/
[15] http://dhawards.org/dhawards2015/results/

5. Alighieri, Dante. “Vita nova.” In *Opere*, a cura di Guglielmo Gorni, Vol. 1. Milano: Mondadori, 2011.
6. Alighieri, Dante. *Vita nuova*. a cura di Domenico De Robertis. Milano-Napoli: Ricciardi, 1980.
7. Bartalesi, Valentina e Carlo Meghini. “Using an ontology for representing the knowledge on literary texts: the Dante Alighieri case study.” *Semantic Web* 8, num. 3 (2017): 385-94.
8. Doerr, Martin. “The CIDOC CRM - An ontological approach to semantic interoperability of metadata,” 2002.
9. González-Blanco, Elena, Levente Seláf, Gimena Del Rio Riande, Clara Martínez Cantón e María Dolores Martos Pérez. “Building a metrical ontology as a model to link digital poetic repertoires.” Lausanne, 2014.
10. Hollander, Robert. “The Dartmouth Dante project.” *Quaderni d’italianistica* 10, num. 1-2 (1989): 287-98.
11. Hollander, Robert. “The Princeton Dante project.” *Humanist studies & the digital age* 3, num. 1 (2013): 53-59.
12. Lana, Maurizio, Fabio Ciotti, Diego Magro, Silvio Peroni, Francesca Tomasi e Fabio Vitali. “Annotating Texts with Ontologies, from Geography to Persons and Events,” 494-96. Lausanne, 2014.
13. Manola, Frank, Eric Miller e Brian McBride. “RDF Primer.” W3C, 2004. https://www.w3.org/TR/2004/REC-rdf-primer-20040210/.
14. Miles, Alistair, Brian Matthews, Michael Wilson e Dan Brickley. “SKOS Core: simple knowledge organisation for the Web.” *International conference on Dublin Core and metadata applications*, 2005, 3-10.
15. Riva, Pat, Martin Doerr, and Maja Žumer. “FRBRoo: enabling a common view of information from memory institutions,” 2008.
16. Shotton, David. “CiTO, the citation typing ontology.” *Journal of Biomedical Semantics* 1, num. 1 (June 22, 2010): S6.
17. Van Peteghem, Julie. “Digital readers of allusive texts: Ovidian intertextuality in the Commedia and the digital concordance on intertextual Dante.” *Humanist studies & the digital age* 4, num. 1 (2015): 39-59.
18. Vitale, Valeria. *An ontology for 3D visualization in cultural heritage*. MP4. Digital classicist. London, 2013. `http://www.digitalclassicist.org/wip/wip2013-02vv.html`.