Towards a Semantic Network of Dante's Works and Their Contextual Knowledge

Valentina Bartalesi and Carlo Meghini
ISTI – CNR, Via G. Moruzzi, 1 - 56124, Pisa

Paola Andriani and Mirko Tavoni
Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa, Piazza Torricelli 2 - 56126, Pisa

Abstract

We present the achievements of the 'Towards a Digital Dante Encyclopaedia' project, a 3-year Italian National Research Project that aims at building a digital library endowed with services supporting scholars in creating, evolving, and consulting a digital encyclopaedia of Dante Alighieri and his works. Our main goal is to represent and visualize the knowledge of the primary sources that Dante refers to in his works. Currently, this information is scattered in many books, making it difficult to systematically overview the culture of Dante in order to obtain a well-founded perception of how this culture was gradually constructed over time. We describe a semantic representation of the involved texts and knowledge based on the Semantic Web languages. The representation is freely available as an XML document, and it can be easily acquired, e.g. by other researchers in order to extend the representation with additional texts and knowledge. We also present a web application that allows users to extract and display these texts and knowledge in the form of charts and tables. In particular, the application supports the visualization of the data related to the primary sources according to different parameters (e.g. in chronological order, by author, by work). Furthermore, it allows exploring the dynamics of the multi-faceted culture of Dante in relation to the diverse and often conflicting stages of his biography.

Correspondence:
Valentina Bartalesi,
ISTI-CNR, via G. Moruzzi,
1 – 56124 Pisa.
E-mail: valentina.bartalesi
@isti.cnr.it

1 Introduction

The 'Towards a Digital Dante Encyclopaedia' project is an Italian National Research Project, started in 2012, that aims at building a digital library endowed with services supporting scholars in creating, evolving, and consulting a digital encyclopaedia of Dante Alighieri and of his works.

In order to achieve this goal, we have first developed a semantic representation of the texts of Dante and of the knowledge related to Dante's primary sources. A semantic representation is a set of classes and relationships that describes a knowledge domain. Our representation is expressed in the RDF language (Manola *et al.*, 2004). In RDF, every piece of knowledge is represented as a triple (subject–predicate–object), and a set of triples forms an RDF graph, generally called semantic network, whose aim is to highlight the formal linguistic nature of the representation. To the best of our

Digital Scholarship in the Humanities, Vol. 30, Supplement 1, 2015.

doi:10.1093/llc/fqv044 Advance Access published on 1 October 2015

knowledge, this is the first attempt to such semantic representation in the context of Dante studies.

Then, we created a user-friendly web application that displays the semantic representation to the user in various ways, and offers various services addressing tasks carried out by the scholars building the encyclopaedia. Namely: (1) the visualization of references to primary sources (i.e. other authors' works that Dante refers to in his own works), (2) the types of references (e.g. explicit citation), and (3) the distribution of references both in time and in the works of Dante. The overall goal is to shed light into Dante's cultural context and into the development of Dante's reference library over time.

The article is organized as follows: in the next section, we review related works, and in Section 3, we briefly describe the methodology that we followed. Section 4 illustrates the semantic representation that we developed for Dante's works and the related knowledge, while Section 5 presents the web services built on top of this representation to support the work of the scholars. Finally, in Section 6, we draw our conclusions.

2 Related Works

Several web applications are currently available that allow the investigation of different aspects of Dante's culture. Unfortunately, none of them focuses on the semantics of the texts and in particular on the Dante's primary sources.

The Princeton Dante Project[1] is a web application that facilitates a textual search of Dante's work and in several commentaries. DaMA[2] is a digital archive collecting main classic, late-ancient, and medieval sources, of Dante's works and commentaries about Dante's theoretical and literary thought. The World of Dante[3] is a multi-media research tool intended to facilitate the study of the 'Divine Comedy' through a wide range of offerings. These include an encoded Italian text allowing structured searches and analyses, an English translation, interactive maps, diagrams, music, a database, a timeline, and a gallery of illustrations. The Dartmouth Dante Project[4] is a searchable full-text database containing more than 70 commentaries on Dante's Divine Comedy. DanteSearch[5] is a web application allowing users to do morphological and syntactic queries on the lemmatized and grammatically annotated Dante's works. All these applications focus on text search and syntactic analysis, while our application focuses on the semantic analysis.

We did not find specific ontologies that are able to represent 'all' the knowledge that we came across with while analysing the commentaries to literature texts. Thus, we selected among the existing ontologies some classes and properties that can be used to express at least some aspects of annotations. We analysed the following ontologies (Bartalesi *et al.*, 2013): CIDOC-CRM (Doerr, 2003), FRBR (Madison, 2000; Tillett, 2004), FRBRoo (Riva *et al.*, 2008), Dublin Core (Baker, 1998), SKOS (Alistair *et al.*, 2004), FOAF (Brickley and Miller, 2012), DoCO (Shotton and Peroni, 2011), FaBiO (Peroni and Shotton, 2012), CiTO (Shotton, 2010), and The Open Annotation Core Data Model (Sanderson *et al.*, 2013).

3 Methodology

We divided our work into several phases. In the first one, we created a simple ontology for representing the knowledge embedded in scholarly commentaries to Dante's Convivio (Fioravanti, 2014), the philosophical treatise that we choose as an initial case study. In the second phase, we generalized the ontology to represent the knowledge embedded in the scholarly commentaries to other Dante's works. So far, in this phase, we have represented the knowledge embedded in Vita Nova (Gorni, 2011), De Vulgari Eloquentia (Tavoni, 2011), and Monarchia (Quaglioni, 2014). In the third phase, in parallel with the second, three domain experts manually extracted the information related to the primary sources (cited works, authors, thematic areas, etc.) from the commentaries. Then, using a Java program, the experts inserted the texts of Dante's works along with the selected information extracted by their commentaries into the Virtuoso triple store (Erling and Mikhailov, 2007), that forms the backbone of our digital library. In the fourth phase, we developed several web services and corresponding

web applications allowing scholars to browse the semantic network of Dante's works, of the primary sources, and of the references linking the works to the primary sources. The web applications represent the primary sources' data in an intuitive way through tables and charts. Their main aim is to highlight data distribution in Dante's works over time.

4 The Ontology

In our digital library, we need to represent:

- passages of Dante's text (e.g. 'Sì come dice lo Filosofo nel principio della Prima Filosofia' ['*As the Philosopher says in the beginning of the first Philosophy*']) to which a note refers to;
- passages of the note citing a primary source (e.g. 'sono le parole con cui si apre la Metafisica di Aristotele (I 1, 980a 21)') ['*These are the words that open the Aristotle's Metaphysics (I 1, 980a 21)*'];
- the corresponding book, chapter, and paragraph of Dante's text (e.g. 1.2.5);
- the author of the work referenced in the commentary (e.g. Aristotle);
- the title of the work referenced in the commentary (e.g. Metaphysics);
- the thematic area of the work referenced in the commentary (e.g. Aristotelianism, Bible, Medical Science, Scholastic Philosophy, Stilnovismo).

In order to re-use available and standard resources, experts in Dante selected thematic areas from Nuovo Soggettario[6], a subject index edited by the National Central Library of Florence, in compliance with the International Federation of Library Associations and Institutions recommendations[7]. A RDF version is freely available on-line[8].

Fig. 1 shows an annotated representation of several elements of the above list, (book, chapter, and paragraph classes are omitted for readability). In order to obtain an ontology for the semantic representation of the above information, we investigated several existing ontologies (see Section 2), and we chose the classes and the properties that we considered the most appropriate for our purposes. Furthermore, we added our own classes and properties for the representation of the knowledge that was not addressed by the existing ontologies. Then, we transformed the initial commentary into an RDF graph structured according to the ontology (Bartalesi *et al.*, 2013). Fig. 2 shows the classes of our ontology.

It is important to note that the works of Dante, as well as most of their primary sources, already exist in digital format. However, to the best of our knowledge, there is no semantic representation that integrates this information into a unique body of knowledge, expressed through a formal ontology. The built knowledge base is not expected to give a coherent view of Dante's works. The knowledge in our RDF graph may be incoherent and incomplete. Indeed, incoherency may arise from two commentaries written by two different scholars that identify two different primary sources for the same piece of Dante's text. Our ontology allows representing such conflicting annotations, because it does not impose a unicity constraint on the primary sources related to one piece of text. On the other hand, incompleteness may arise from the fact that we considered only some commentaries to Dante's works while allowing the addition of new commentaries at any moment in order to increase knowledge completeness.

5 The Web Application

In order to extract and display the knowledge stored in our digital library, we developed a web application[9]. The application is meant to support scholars in writing a complete encyclopaedia about Dante's works. It is able to produce column bar charts in order to show the data about primary sources cited by Dante. We used the Highcharts[10] JavaScript library to implement these charts. This library 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 extracted by the application in CSV format[11]. This feature is particularly important since it allows scholars to obtain and manage raw data, in order to apply further data analyses.

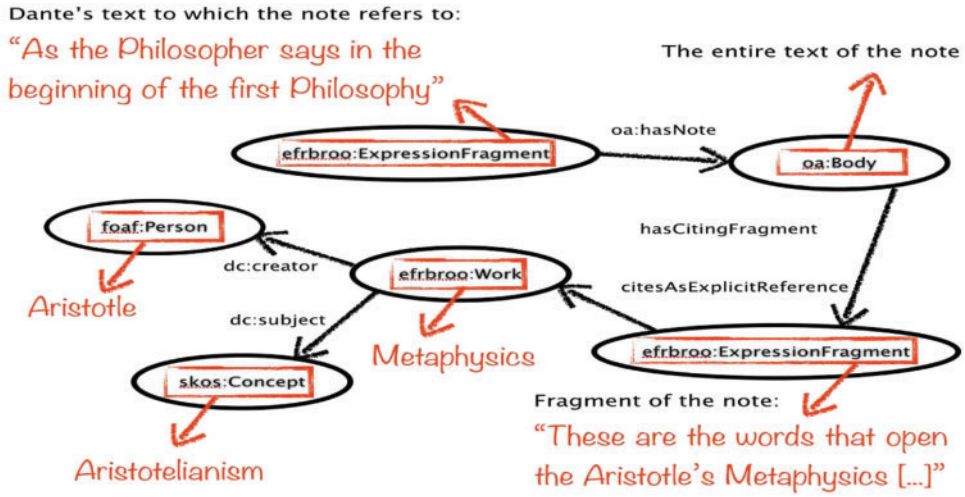

Fig. 1 A schematic representation of a part of our semantic model, with examples of instances of the classes

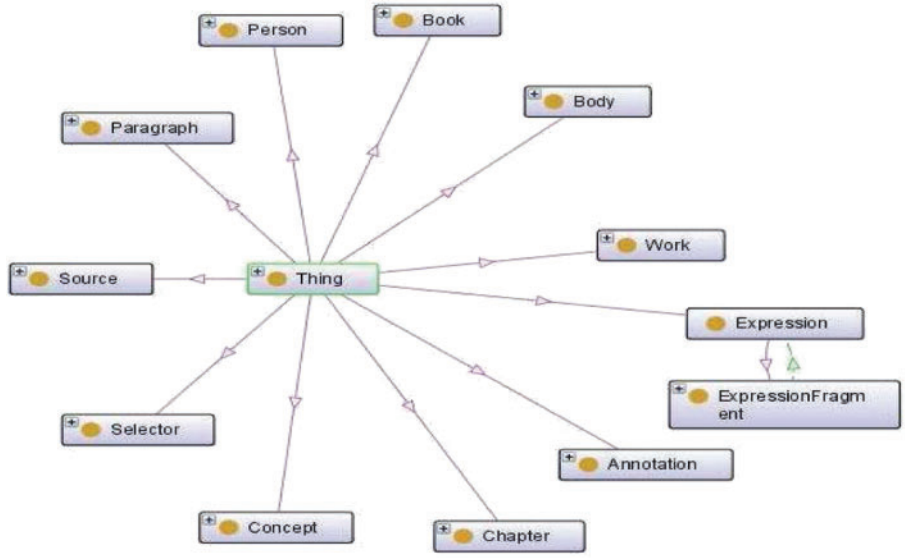

Fig. 2 The classes of our ontology

Up to now, six different predefined queries are available to extract data for a representation using column bar charts. They can be distinguished into two different groups. The first group includes three queries. For these queries, the user can choose one Dante's works (or even all his works) and, in addition, a specific subpart of the work (e.g. a book). Fig. 3 shows the search form.

The three queries of the first group are the following:

(1) **Cited works**. The query returns data regarding the distribution of the works cited by Dante. If the user clicks on the name of a cited work on the *x*-axis, a table appears reporting the following information: (1) where

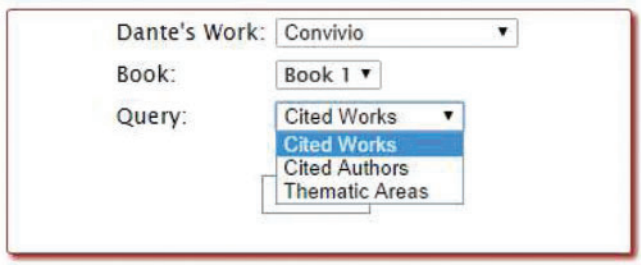

Fig. 3 The search form for the first group of queries

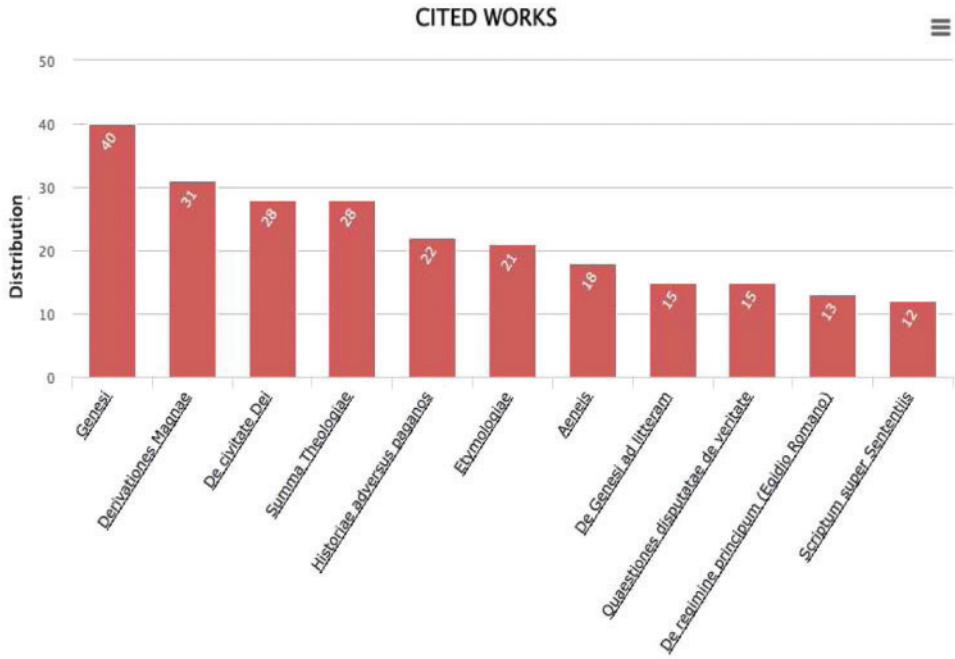

Fig. 4 Data regarding the distribution of some works cited in the first book of De Vulgari Eloquentia

the work is cited (Dante's work, book, chapter, paragraph, or poem); (2) the kind of citation (explicit, strict, or generic reference); (3) the fragment of the Dante's text the note refers to; (4) the precise fragment of the cited works (if present in the note); (5) the thematic area of the cited work; (6) the author of the cited work. An example of a cited works graph is reported in Fig. 4.

(2) **Cited authors.** The query returns data regarding the distribution of the authors of the primary sources cited by Dante. Clicking on the name of the cited author on the *x*-axis, a table appears reporting: (1) the list of the author's works cited by Dante; and (2) where the author is cited in Dante's work, including information about the specific book, chapter, paragraph, or poem.

(3) **Cited thematic areas.** The query returns data regarding the distribution of the thematic areas of the works cited by Dante. If the user clicks on the name of the cited thematic area, a table appears reporting: (1) the cited work belonging to the selected thematic area;

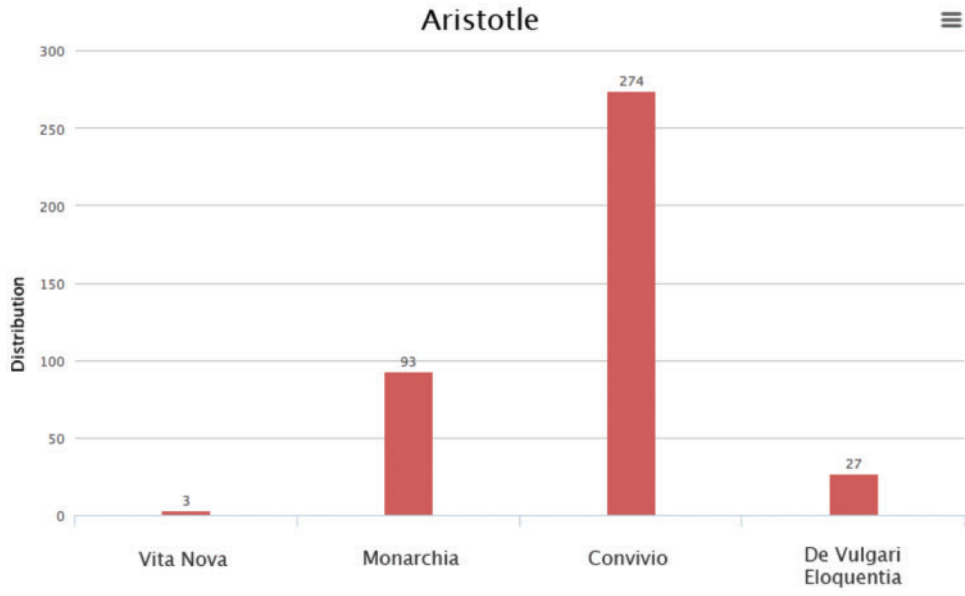

Fig. 5 The distribution of Aristotle into Dante's works

(2) the author of the cited work; (3) the Dante's works, books, chapters, paragraphs, or poems in which the thematic area is cited.

The second group contains three other queries. In order to improve the usability of the web application, we implemented two different search forms:

(1) An autocomplete menu where the user can type the title of the cited work, the author, or the thematic area;
(2) An alphabetically ordered list in which the user can select the title of the cited work, the author, or the thematic area.

The last three queries return several column bar charts representing the following information:

(1) **A specific cited work.** The charts represent the distribution of a chosen cited work progressively in: Dante's works, books, chapters, paragraphs, or poems.
(2) **A specific cited author.** The charts report the distribution of a chosen cited author progressively in: Dante's works, their books, chapters, paragraphs, or poems. Fig. 5 reports the graph regarding the distribution of the cited author Aristotle in Dante's works.
(3) **A specific cited thematic area.** The charts show the distribution of a chosen cited thematic area progressively in: Dante's works, books, chapters, and paragraphs or poems.

The user can shift from high granularity charts to finer ones by clicking on the column bars.

6 Conclusions

We presented an RDF representation of the knowledge included in Dante Alighieri's works, focusing in particular on the primary sources. Then, a web application has been developed in order to visualize the knowledge stored in our RDF graph, generally called a semantic network.

Until now, there was not a single ontology which allowed describing the knowledge about primary sources included in literary works. The main advantages to have such ontology available are that: (1) using the Linked Data approach, researchers can obtain and re-use our representation; (2) our ontology can be linked with other ontologies to extend the represented

domain; and (3) any user can download and use our representation to build more applications on it.

The creation of the semantic network is a very time-consuming and knowledge-intensive process. It requires researching the most appropriate ontologies for representing all aspects, and in several cases, it requires developing a new ontology to fill the gaps of the existing ones. The fact of gathering the current information on the primary sources used by Dante in his works, and the fact of having this information available in digital format, can improve and make more efficient the research of the scholars. Having such information dispersed on paper books, indeed, makes impossible a systematic overview of the culture of Dante and a well-ordered perception of how it was gradually set up in time. We developed a web application that allows visualizing the data about primary resources, according to different parameters (in chronological order, or by type of source, or by author, by work, etc.). This application allows scholars to automatically explore the dynamics of the multi-faceted culture of Dante in relation to the diverse and often conflicting stages of his biography and to study the evolution in time of Dante's cultural background.

References

Alighieri, D. (2011). De Vulgari Eloquentia. In Giunta, C., Gorni, G. and Tavoni, M. (eds), *Opere. Vol. I: Rime, Vita Nova, De vulgari eloquentia*. Milano: Mondadori.

Alighieri, D. (2011). Vita Nova. In Giunta, C., Gorni, G. and Tavoni, M. (eds), *Opere. Vol. I: Rime, Vita Nova, De vulgari eloquentia*. Milano: Mondadori.

Alighieri, D. (2014). Convivio. In Fioravanti, G., **AA.VV.** (eds), *Opere. Vol. II: Convivio, Monarchia, Epistole, Egloghe*. Milano: Mondadori.

Alighieri, D. (2014). Monarchia. In Quaglioni, D. (ed.), *Opere. Vol. II: Convivio, Monarchia, Epistole, Egloghe*. Milano: Mondadori.

Baker, T. (1998). *Languages for Dublin Core*. D-Lib Magazine. http://dlib.org/dlib/december98/12baker.html (accessed on 23 September 2015).

Bartalesi, V., Meghini, C., Locuratolo, E. and Versienti, L. (2013). A preliminary study on the semantic representation of the notes to Dante Alighieri's Convivio. *Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: metadata, vocabularies and techniques in the Digital Humanities (DH-CASE '13)*. Florence, Italy, September 2013. New York: ACM Press.

Brickley, D. and Miller, L. (2012). *FOAF vocabulary specification 0.98*. Namespace Document, August 9 (2010). http://xmlns.com/foaf/spec/ (accessed on 23 September 2015).

Doerr, M. (2003). The CIDOC CRM—an ontological approach to semantic interoperability of metadata. *AI Magazine*, 24(3): 75–92.

Erling, O. and Mikhailov, I. (2007). RDF Support in the Virtuoso DBMS. In Networked Knowledge-Networked Media, 7–24. Springer Berlin Heidelberg.

Madison, O. M. A. (2000). The IFLA functional requirements for bibliographic records: International standards for universal bibliographic control. *Library Resources &Technical Services*, 44(3): 153–159.

Manola F., Miller, E. and McBride, B. (2004). RDF Primer. W3C Recommendation 10 February 2004. http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/ (accessed 30 September 2014).

Miles, A., Matthews, B., Wilson, M. and Brickley, D. (2005). SKOS Core: Simple knowledge organisation for the Web, *Proceedings of the 2005 international conference on Dublin Core and metadata applications: vocabularies in practice (DCMI '05)*. Dublin Core Metadata Initiative. Madrid, Spain.

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

Riva, P., Doerr, M. and Žumer, M. (2008). FRBRoo: enabling a common view of information from memory institutions. *International Cataloguing and Bibliographic Control*, 38: 30–4

Sanderson, R., Ciccarese, P. and Van de Sompel, H. (2013). Designing the W3C open annotation data model. *Proceedings of the 5th Annual ACM Web Science Con-ference*. New York: ACM Press.

Shotton, D. (2010). CiTO, the citation typing ontology. *Biomedical Semantics*, 1(S-1): S6.

Shotton, D. and Peroni, S. (2011). *DoCO, the Document Components Ontology*. http://purl.org/spar/doco (accessed 30 September 2014).

Tillett, B. (2004). What is FRBR? A conceptual model for the bibliographic universe. *Library of Congress Cataloging Distribution Service*, 25(5): 1–8.

Notes

1 http://etcweb.princeton.edu/dante/index.html
2 http://perunaenciclopediadantescadigitale.eu/istidama/index.php?id=6&L=1
3 http://www.worldofdante.org/
4 http://dante.dartmouth.edu/
5 http://www.perunaenciclopediadantescadigitale.eu:8080/dantesearch/
6 http://thes.bncf.firenze.sbn.it/index_eng.html
7 http://www.ifla.org/
8 http://thes.bncf.firenze.sbn.it/thes-dati_eng.htm
9 http://dante1.isti.cnr.it:8080/perunaenciclopediadantescadigitale/
10 http://www.highcharts.com/
11 http://www.w3.org/2013/05/lcsv-charter.html