

# From the Clavius Correspondence to Linked Data: the CoW-LD Project

Angelica Lo Duca<sup>1</sup>, Andrea Marchetti<sup>2</sup>

<sup>1</sup>Institute of informatics and Telematics - CNR, Italy - angelica.loduca@iit.cnr.it

<sup>2</sup>Institute of informatics and Telematics - CNR, Italy - andrea.marchetti@iit.cnr.it

## ABSTRACT

Over the last years, a great effort has been made in the field of Digital Humanities to digitize documents and collections in different formats. As a result a great amount of data has been made available thus providing new opportunities to build applications allowing the navigation of collections. In this paper, we focus on the Clavius Correspondence, a set of 266 handwritten letters, sent to Christopher Clavius, a known mathematician of the Sixteenth Century, from other important scientists of the same period. In order to bring all these precious manuscripts on the Web, the Clavius on the Web (CoW) project was built. CoW enriched all these letters with lexical and semantic annotations. In addition, the project implemented a Web platform where all this material was made available and annotations within letters could be done. Among the many activities of CoW, one focused on building a Linked Dataset about all contextual resources associated to manuscripts, people and places involved in those letters. This paper describes the CoW Linked Dataset (CoW-LD), built around these manuscripts. CoW-LD was built partially manually and partially through automatic programs, which enriched the dataset with external resources, such as those contained in DBpedia and GeoNames. As a result, CoW-LD contains also letters annotations and events associated to the authors of the letters, as well as aliases to external sources.

## KEYWORDS

Linked Data, Digital Humanities, Europeana Data Model, Christopher Clavius.

## 1. INTRODUCTION

One of the preliminary activities of Digital Humanities consists in digitizing documents and manuscripts and organizing them in collections. After that, the next activity consists in making them accessible to common users through Web portals or online catalogs [1]. Many projects and initiatives have been proposed following this line, such as the Levy Dead Sea Scrolls Digital Library<sup>1</sup>, which shows the famous Qumran scrolls in a Web portal. Another recent initiative is the Clavius on the Web project (CoW) [2], which aims at making accessible on the Web the Clavius Correspondence, owned by the Historical Archives of the Pontifical Gregorian University (APUG)<sup>2</sup> in Rome. Christopher Clavius (1537-1612) was a jesuit mathematician and astronomer and one of the most important characters in the scientific scene of the late Sixteenth Century. These manuscripts consist of two volumes of correspondence (266 letters) and seven volumes of works, some of these ones printed and some not still published. The importance of the correspondence becomes clear just looking at the authors of the letters: Galileo Galilei, Tycho Brahe, Joseph Scaliger, Guido Ubaldo Dal Monte and many others. Within CoW many modules were implemented to visualize, annotate and extract knowledge from manuscripts.

In this paper, we describe the Clavius on the Web Linked Dataset (CoW-LD)<sup>3</sup> which represents all the contextual resources associated to the Clavius Correspondence as Linked Data. CoW-LD contains many links to external sources, i.e. DBpedia [3], GeoNames<sup>4</sup>, Wikipedia<sup>5</sup>, VIAF<sup>6</sup>, Treccani<sup>7</sup> and Google Books<sup>8</sup>. The dataset is released under the CC BY-SA 4.0 license<sup>9</sup> and is published as a Linked Data node, which can be queried through a SPARQL endpoint<sup>10</sup>.

---

<sup>1</sup> <<http://www.deadseascrolls.org.il/>> (last access 2017/09/29).

<sup>2</sup> <[https://www.unigre.it/Archivio/archivio\\_en.php](https://www.unigre.it/Archivio/archivio_en.php)> (last access 2017/09/29).

<sup>3</sup> <<http://claviusontheweb.it/lod>> (last access 2017/09/29).

<sup>4</sup> <<http://www.geonames.org/>> (last access 2017/09/29).

<sup>5</sup> <<https://www.wikipedia.org/>> (last access 2017/09/29).

<sup>6</sup> <<https://viaf.org/>> (last access 2017/09/29).

<sup>7</sup> <<http://www.treccani.it/>> (last access 2017/09/29).

<sup>8</sup> <<https://books.google.com/>> (last access 2017/09/29).

<sup>9</sup> <<https://creativecommons.org/licenses/by-sa/4.0/>> (last access 2017/09/29).

<sup>10</sup> <<http://claviusontheweb.it/dataset/snorql/>> (last access 2017/09/29).

## 2. RELATED WORK

Many important initiatives have been rising to publish collections of documents and manuscripts as Linked Data. DM2E [4] is a project which provides links to manuscripts and their contextual resources from fifteen Cultural Heritage institutions across Europe. Other important initiatives are the Europeana Linked Open Data [5], the Burckhardtsource project [6], BookSampo dataset [7] for fiction literature, the Amsterdam Museum [8] and the Rijksmuseum Collection [9].

## 3. THE COW-LD DATASET

From manuscripts owned by APUG, the following resources were extracted: *letters* (all handwritten letters), *agents/persons* (all the authors of letters), *places* (all the places where the letters were written), *events* (all events related to the lives of the characters as well as the process of creating letters), *annotations* (all the metadata marked in the letters, such as geometric entities), *concepts* (other entities). Entities in CoW-LD contain both internal links and external links. An internal connection establishes a relationship between two nodes within the same dataset. External links, instead, establish relationships between nodes in the dataset and other nodes belonging to external sources, such as DBpedia. In CoW-LD almost all the external links have been modeled through the *owl:same* property. CoW-LD contains external links to the following Web sources: DBpedia, VIAF and Treccani, for the people; DBpedia and GeoNames, for places; DBpedia for events. The dataset contains also links to GoogleBooks, with regard to the works of the authors of the letters. While external links were built manually, external links were built through automatic scripts and then filtered manually.

One of the best practices of Linked Data asserts ontology reuse. For this reason, the COW-LD dataset exploits different well-known ontologies to represent entities and properties. Firstly, as main ontology, it uses the Europeana Data Model ontology [10], the official ontology developed as part of Europeana<sup>11</sup>. EDM uses standard ontologies for the representation of entities (such as FOAF [11] and Dublin Core<sup>12</sup>), and enriches them with classes and specific properties related to the domain of Cultural Heritage. Secondly, CoW-LD uses some properties and classes defined in the Simple Knowledge Organization System RDF Schema (SKOS) [12] and Web Ontology Language (OWL) [13] ontologies. Finally, the CoW-LD dataset exploits the Web Annotation Vocabulary<sup>13</sup> to represent annotated concepts. Figure 1 shows the employed classes and the main relationships. Some concrete examples of resources can be found on the COW-LD website<sup>14</sup>.

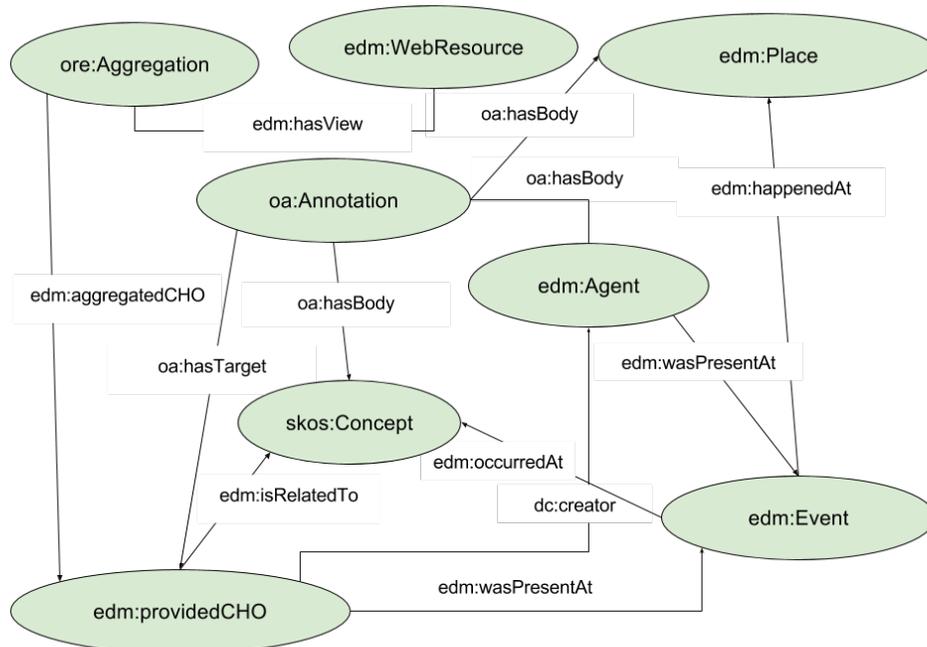


Figure 1: Classes and their main relationships within CoW-LD.

<sup>11</sup> <<http://www.europeana.eu/>> (last access 2017/09/29).

<sup>12</sup> <<http://dublincore.org/>> (last access 2017/09/29).

<sup>13</sup> <<https://www.w3.org/TR/annotation-vocab/>> (last access 2017/09/29).

<sup>14</sup> <<http://claviusontheweb.it/lod/resource.html>> (last access 2017/12/26).

#### 4. CONCLUSIONS AND FUTURE WORK

In this paper we have illustrated CoW-LD, the Linked Data version of all the Clavius Correspondence. At the moment, the dataset contains 13219 triples and 1583 resource. The next step consists in enriching the dataset by extracting all the information related to letters sent to Christopher Clavius to the other scientists. This process has already started and consists in building a table with all the information contained in the Clavius Correspondence book, written by Ugo Baldini [14]. As further step, all the Clavius Correspondence could be represented through some specific ontologies related to the domain of letters. Such ontologies should be firstly identified and then used.

#### 5. ACKNOWLEDGMENTS

The CoW-LD dataset is part of the Clavius on the Web project, which was funded by Registro .it.

#### 6. BIBLIOGRAPHY

- [1] BORGMAN, Christine L. *What are digital libraries? Competing visions*. Information Processing & Management, 1999, 35.3: pp. 227-243.
- [2] PEDRETTI, Irene et. al. *The Clavius on the Web Project: Digitization, Annotation and Visualization of Early Modern Manuscripts*. In Proceedings of the Third AIUCD Annual Conference on Humanities and Their Methods in the Digital Ecosystem (AIUCD '14), 2014, ACM, New York, NY, USA.
- [3] AUER, Soren. et al. *Dbpedia: A nucleus for a web of open data*. The Semantic Web, 2007, 722-735.
- [4] BAIERER, Konstantin, et al. *DM2E: A Linked Data Source of Digitised Manuscripts for the Digital Humanities*. Semantic Web, 2017, Preprint: 1-13.
- [5] ISAAC, Antoine; HASLHOFER, Bernhard. *Europeana linked open data–data. europeana. eu*. Semantic Web, 2013, 4.3: 291-297.
- [6] DI DONATO, Francesca, et. al. *Biblioteche digitali semantiche. il progetto burckhardtsource. org*. Bibliotime, 2013, 16.1.
- [7] MÄKELÄ, Eetu, et. al.. *Fiction literature as linked open data - The BookSampo dataset*. Semantic Web, 2013, 4.3: 299-306.
- [8] DE BOER, Victor, et al. *Amsterdam museum linked open data*. Semantic Web, 2013, 4.3: 237-243.
- [9] DIJKSHOORN, Chris, et al. *The Rijksmuseum Collection as Linked Data*. Semantic Web, 2014, Preprint: 1-10.
- [10] ISAAC, Antoine. et al. *Europeana data model primer*. 2013. Available at: [https://pro.europeana.eu/files/Europeana\\_Professional/Share\\_your\\_data/Technical\\_requirements/EDM\\_Documentation/EDM\\_Primer\\_130714.pdf](https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf) (last access 2017/09/29).
- [11] BRICKLEY, Dan et. al. *FOAF vocabulary specification 0.99*. 2007. Available at: <http://xmlns.com/foaf/spec/> (last access 2017/09/29).
- [12] MILES, Alistair, et. al.. *SKOS simple knowledge organization system reference*. 2009. Available at: <https://www.w3.org/TR/skos-reference/> (last access 2017/09/29).
- [13] MCGUINNESS, Deborah L., et al. *OWL web ontology language overview*. W3C recommendation, 2004, 10.10: 2004. <https://www.w3.org/TR/owl-features/> (last access 2017/09/29).
- [14] CLAVIUS, Christoph. *CORRISPONDENZA*. Edizione critica a cura di. Ugo Baldini e Pier Daniele Napolitani. VOLUME II. 1570 - 1592. Parte I: lettere e testi.