The IMAGO Project: Towards a Knowledge Base of Medieval and Renaissance Geographical Works

Valentina Bartalesi[1][0000−0001−9024−0822] and Nicolò Pratelli[1][0000−0003−0364−922X]

ISTI-CNR, Via G. Moruzzi 1, 56124 Pisa, Italy
{valentina.bartalesi,nicolo.pratelli}@isti.cnr.it

Abstract. The image of the world created by the Medieval and Renaissance culture was crucial to the development of Western thought in European history. To the best of our knowledge Medieval and Renaissance geographical works have not been studied using digital methods. The three years (2020-2023) Italian National research project IMAGO - Index Medii Aevi Geographiae Operum - aims at providing a systematic overview of this literature using Semantic Web technologies. As the first step to develop tools to support scholars in creating, evolving and consulting a knowledge base (KB) of the geographical works, we created an OWL 2 DL ontology. Following the re-use logic and to maximize the interoperability, we developed the ontology as an extension of two reference ontologies, that is the CIDOC CRM vocabulary and its extension FRBRoo, including its in-progress reformulation, LRMoo. In this paper, we present the project, the ontology and the tool to populate it that we developed. Furthermore, we present a preliminary study to map the works collected in the IMAGO KB and the manuscripts stored in the KB of the Mapping Manuscript Migrations project.

Keywords: Semantic Web · Medieval geographical Works · Digital Humanities.

1 Introduction

The image of the world created by the Medieval and Renaissance culture was crucial to the development of Western thought in European history. During the Middle Ages, geographical descriptions were mostly used to collect the human knowledge into encyclopedic works or to provide universal chronicles [11]. Specific descriptions of lands, cities, places, monuments and buildings were also supplied as a guide to the pilgrims travelling to the Holy Land, Rome and Santiago de Compostela [10]. By the end of the Middle Ages and the beginnings of Renaissance Humanism, a more clear image of the world was defined thanks to the discovery of ancient geographical models (especially the works of Ptolemy and Strabo): detailed information from the past helped to produce more accurate geographical descriptions and maps. Furthermore, the genre of geographical

description had a further and decisive turning point during the period of the exploration travels and discoveries: the description and representation of the New World, along with the reassessment of the physical space, gave the basis of modern geography [4].

Until now, Medieval and Renaissance geographical works have not been studied using digital methods. The three years (2020-2023) Italian National research project IMAGO - Index Medii Aevi Geographiae Operum - aims at providing a systematic overview of this literature using the Semantic Web technologies to make available this knowledge as Linked Open Data (LOD) [2] and to develop automatic search and visualisation services on the collected data. In particular, the project aims to produce and make available to the users a complete survey of Medieval and Humanistic geographical works, providing: (i) a classification of authors, genres and contents; (ii) a list of the manuscript tradition and printed editions for each work; (iii) a list of critical editions of some more representative works; (iv) a Medieval Latin toponymy index.

As the first step in order to develop tools to support scholars in creating, evolving and consulting a knowledge base (KB) of the geographical works, we created an OWL 2 DL ontology [9] that formally represents this knowledge. Following the re-use logic and in order to maximize the interoperability, we developed our ontology as an extension of two reference ontologies, that is the CIDOC CRM [5] vocabulary and its extension FRBRoo [6], including its in-progress reformulation, LRMoo [12].

The final aim of the project is the creation of a Web application allowing scholars to freely access and visualise the data collected in the IMAGO knowledge base. The idea is to improve the studies of Medieval and Renaissance Humanism geography by providing scholars a better insight into this field from many perspectives, such as the Medieval Latin toponymy and the identification of historical places. The Web application will host a special section of Medieval and Renaissance cartography as well, in order to provide a digital collection of the most interesting maps and drawings.

2   The IMAGO Ontology

As the first step to develop tools to support scholars in creating, evolving and consulting a KB of the Medieval and Renaissance geographical works, we created an ontology that formally represents this knowledge. The IMAGO ontology is derived from a strict collaboration between ISTI-CNR and the scholars from the University of Pisa and the University of Salento - expert in Latin and Italian Literature and Linguistics - who are involved in the project. The methodology we followed to develop the ontology is well known and it is the one usually adopted to create formal vocabularies in the Semantic Web research field. The main novelty introduced by our research is the use of the Semantic Web technologies to formally represent the scientific domain of the geographical Latin works written during the Middle Ages and the Renaissance. Despite in other research projects semantic technologies have been used to represent ancient manuscript

corpora [7, 1, 3, 8], no scientific research that applies a Semantic Web approach has been conducted in this specific research field. Furthermore, the information about the geographical Latin works written during the Middle Ages and the Renaissance is dispersed on paper books, and this makes a systematic overview of the geographic literature impossible, preventing a well-ordered perception of how it was gradually set up in time. The IMAGO project aims at making this information available in digital form to both scholars and general users. We defined a conceptualisation of the domain of knowledge and then we formalised this conceptualisation using classes and properties from two existing ontologies we chose as reference vocabularies, that is the CIDOC CRM and its extension FRBRoo, including its in-progress reformulation LRMoo. We adopted a lot of terms from these ontologies to maximize the interoperability of our representation. Finally, we added our own classes and properties to represent the terms that we did not find in the reference vocabularies. The resulting ontology is expressed in OWL 2 DL language. Our conceptual idea is that the domain of the geographical work can be represented using some main categories. The first ones are the author and title of a work. For each work, the literary genre is specified along with the toponyms that represent the places that are described or reported into the work. Furthermore, for each work, several metadata about the related manuscripts and printed editions are added. In particular, for each manuscript the following knowledge is reported : the name of the author and the title of the work in the forms that appear in the manuscript; the library in which the manuscript is collected; the location of the library; the signature and the folios of the manuscript; the incipit and explicit of the dedication/proem, if they exist; the incipit and explicit of the text; the date of the creation of the manuscript; the secondary sources.

On the other hand, for each printed edition the following knowledge is reported: the author, the title, and curator's name of the edition; the place and the date of publication; the publisher; the format of the edition; the number of pages; the information about the images reported in the edition; some general notes that the scholars intend to add as comment to the edition; the name of the author of the introduction, the text of the introduction, the text of the dedications; information about whether the edition is a first edition or a reprint; primary and secondary sources of the edition; the ecdotic typology. In Table 1 we reported the classes we used to represent our main concepts and in Table 2 the properties we used to express the semantic relationship among concepts are listed. As a notational convention, the CIDOC CRM uses the letters "E" and "P" to indicate classes and properties respectively, whereas FRBRoo (and its revisions LRMoo) uses the letters "F" and "R" to indicate classes and properties, respectively. Note that we intended dates as time intervals and that we preferred the classes of FRBRoo and LRMoo instead of the corresponding classes of CRM when the concepts to capture and represent underlay the semantics of bibliographic information. Furthermore, we used F2 Expression instead of F1 Work for representing a work since for *work* we intend a particular edition of that work.

Table 1. Classes used to represent our main concepts

Concept	Class
Author	subclass of E39 Actor
Work	equivalent to F2 Expression
Work creation	equivalent to F28 Expression Creation
Genre	subclass of E55 Type
Toponym	subclass of E41 Appellation
Manuscript	subclass of F5 Item
Printed Edition	subclass of F3 Manifestation
Library	subclass of F11 Corporate Body
Place	equivalent to E53 Place
Geographical Coordinate	equivalent to E94 Space Primitive
Signature	equivalent to E42 Identifier
Folios	subclass to E19 Physical Object
Date	equivalent to E52 Time-Span
Curator/Publisher	subclass of E39 Actor

Table 2. Properties used to represent relation among the main concepts

Relation (R) between concepts	Property
R(Work creation event,Author)	equivalent to P14 is carried out by
R(Work creation event,Work)	equivalent to R17 created
R(Manuscript,Title)	equivalent to P102 has title
R(Printed Edition,Title)	equivalent to P102 has title
R(Manuscript,Library)	equivalent to P50 has current keeper
R(Place,Geographical coordinates)	equivalent to P168 place is defined by
R(Manuscript,Signature)	equivalent to P1 is identified by
R(Manuscript,Folios)	equivalent to P46 is composed of
R(Manuscript,Date)	equivalent to P4 has time span
R(Printed edition,Date)	equivalent to P4 has time span
R(Printed edition,Curator)	subproperty of P14 carried out by
R(Printed edition,Publisher)	subproperty of P14 carried out by
R(Printed edition,Format)	equivalent to R69 specifies physical form
R(Printed edition,Page)	equivalent to P106 is composed of

To improve the level of interoperability of the ontology, we used specific terminological resources, when possibile. For the individuals of the class Genre, we used the Soggettario Nazionale[1], a standard thesaurus created and maintained by the National Central Library of Florence. To represent the instances of the following classes we used Wikidata [14] as reference KB: (i) Toponym, which represents the places that are described or reported into the work, (ii) Library, in which the manuscript is stored, (iii) Place, which represents the location of a library. For the ecdotic typology of the printed edition, we did not find a suitable terminology, thus we created a short controlled vocabulary to satisfy our representational aims.

---

[1] https://thes.bncf.firenze.sbn.it/

To populate the ontology, we developed a semi-automatic Web tool to allow scholars to insert knowledge through a user-friendly interface. The tool was created to reduce the time to insert knowledge and to avoid the insertion of mistakes thanks to the use of predefined lists of works, authors, libraries, places, literary genres. The geographical coordinates of the places are also automatically assigned. The labels and the corresponding IRIs[2] contained in these predefined lists are extracted from the Wikidata knowledge base [14] and the MIRABILE database[3]. Figure 1 shows the main interface of the tool.

At the current stage of the project, our KB includes 250 works, 206 authors and 614 libraries and the scholars have started to insert detailed knowledge about manuscripts and printed editions of these works. The KB also includes seven different literary genres, four types of editions, six ecdotic typologies.

Fig. 1. The interface of the tool used by the scholars to insert the knowledge

3   Adding Knowledge about Manuscript Migration to the IMAGO KB

Mapping Manuscript Migrations (MMM) [3] is a project developed with funding from the Trans-Atlantic Platform under its Digging into Data Challenge

---

(2017-2019). By using Linked Open Data principles and Web Semantic technologies, MMM unite records from three datasets: the Schoenberg Database of Manuscripts[4] at the University of Pennsylvania, the Bibale[5] database at the Institut de recherche et d'histoire des textes, and the Medieval Manuscripts Catalogue[6] at the University of Oxford. Within the MMM, a data model was developed to serve the aims of the project, but it is general enough to be used by anyone who would want to represent the knowledge about the manuscript provenance data. It incorporates concepts from several existing ontologies, including Erlangen CIDOC-CRM for events, FRBRoo for bibliographic information, and the Getty Thesaurus of Geographic Names for physical locations. The data model includes also its own classes and properties that serve both unique instances in the source datasets and manuscript studies in general. The knowledge stored in the MMM KB is interesting for the IMAGO project. In particular, the information on how the manuscripts have traveled across time and space from their place of production to their current locations could significantly enrich the IMAGO KB. Since both MMM and IMAGO use the same reference vocabularies, the level of interoperability between the two ontologies is high. We have conducted a preliminary study to map our works and the manuscripts stored in the MMM KB. Querying the MMM KB, we measured that about 20% of the works collected in the IMAGO KB is also present in the MMM KB. We plan to integrate the knowledge related to these shared manuscripts in order to give more complete information to the users of the IMAGO Web application.

4   Conclusion and Future Work

In this paper, we have presented the research developed within the Italian National Research Project IMAGO - Index Medii Aevi Geographiae Operum (2020-2023). IMAGO aims at creating a KB of the Medieval and Renaissance geographical works which report the description and representation of the world in the VI-XV centuries. The knowledge included in the KB is formally represented following the Linked Open Data paradigm and using the languages of the Semantic Web (OWL 2 DL). Indeed, to the best of our knowledge, until now no scientific research has applied digital methods in a systematic way in this field of studies. We have presented the ontology we have developed to formally represent the knowledge about these geographical works. The IMAGO ontology has been implemented as an extension of two standard vocabularies: CIDOC CRM and FRBRoo (and its ongoing extension LRMoo). On the basis of the ontology, we have developed a tool that is used by the scholars who are inserting data in our KB. We have also presented a preliminary study to map the works collected in the IMAGO KB and the manuscripts stored in the KB of the Mapping Manuscript Migrations project. First of all, as future work we have planned to evaluate the ontology. In particular, we plan to conduct two different types

---

[4] https://sdbm.library.upenn.edu/

[5] https://bibale.irht.cnrs.fr/

[6] https://medieval.bodleian.ox.ac.uk/

of evaluation: an automatic evaluation and an evaluation involving users. For the first type of evaluation, we plan to use the automatic OntoQA system [13] that allows us to evaluate both the model and the KB. For the second type of evaluation, we plan to propose a specific questionnaire to the scholars who are currently populating the ontology. After the analysis of the evaluation results, if necessary, we will review and extent our ontology. The long term aim of the project is to develop a Web application that allows retrieving and consulting the data collected in the IMAGO KB in a user-friendly way (e.g. tables, maps, CSV files) for scholars and general users.

References

1. Barzaghi, S., Palmirani, M., Peroni, S.: Development of an ontology for modelling medieval manuscripts: The case of progetto irnerio. Umanistica Digitale **9**, 117–140 (2020)
2. Bauer, F., Kaltenböck, M.: Linked open data: The essentials. Edition mono/monochrom, Vienna **710** (2011)
3. Burrows, T., Emery, D., Fraas, M., Hyvönen, E., Ikkala, E., Koho, M., Lewis, D., Morrison, A., Page, K., Ransom, L., Thomson, E., Tuominen, J., Velios, A., Wijsman, H.: Mapping manuscript migrations: Digging into data for researching the history and provenance of medieval and renaissance manuscripts: White paper (August 2020), https://diggingintodata.org/file/1281/download?token=x59u8fFQ
4. Defilippis, D.: Da flavio biondo a leandro alberti. corografia e antiquaria tra quattro e cinquecento, atti del convegno di studi (foggia, 2 febbraio 2006), a cura di d. defilippis (2009)
5. Doerr, M.: The cidoc conceptual reference module: An ontological approach to semantic interoperability of metadata. AI Mag. **24**(3), 75–92 (Sep 2003)
6. Doerr, M., Bekiari, C., LeBoeuf, P.: Frbroo, a conceptual model for performing arts. In: 2008 Annual Conference of CIDOC. pp. 06–18. CIDOC – ICOM International Committee for Documentation (2008)
7. Gehrke, S., Frunzeanu, E., Charbonnier, P., Muffat, M.: Biblissima's prototype on medieval manuscript illuminations and their context. In: SW4SHD@ ESWC. pp. 43–48 (2015)
8. Jordanous, A., Lawrence, K.F., Hedges, M., Tupman, C.: Exploring manuscripts: Sharing ancient wisdoms across the semantic web. In: Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics. pp. 1–12 (2012)
9. Krötzsch, M.: Owl 2 profiles: An introduction to lightweight ontology languages. In: Reasoning Web International Summer School. pp. 112–183. Springer (2012)
10. Menestò, E.: Relazioni di viaggi e di ambasciatori. Cavallo, G pp. 535–599 (1994)
11. Potthast, A.: Repertorium fontium historiae medii aevi (1962)
12. Riva, P., Žumer, M.: Frbroo, the ifla library reference model, and now lrmoo: a circle of development. In: IFLA WLIC 2018 Conference, Transform Libraries, Transform Societies (2017)
13. Tartir, S., Arpinar, I.B., Moore, M., Sheth, A.P., Aleman-Meza, B.: Ontoqa: Metric-based ontology quality analysis (2005)
14. Vrandečić, D.: Wikidata: A new platform for collaborative data collection. In: Proceedings of the 21st international conference on world wide web. pp. 1063–1064 (2012)