OXFORD UNIVERSITY PRESS | Digital Scholarship in the Humanities

Towards a Knowledge Base of Geographical Latin Works of Medieval and European Humanism: The IMAGO Ontology

Journal:	*Digital Scholarship in the Humanities*
Manuscript ID	DSH-2020-0180
Manuscript Type:	Full Paper
Date Submitted by the Author:	02-Dec-2020
Complete List of Authors:	Bartalesi, Valentina; CNR, ISTI Metilli, Daniele; CNR, ISTI; University of Pisa, Department of Computer Science Pratelli, Nicolò; CNR, ISTI Pontari, Paolo; University of Pisa Department of Philology Literature and Linguistics
Keywords:	Latin works, Ontology, Semantic Web, Medieval geography, Latin geographic literature, Linked Open Data

Towards a Knowledge Base of Geographical Latin Works of Medieval and European Humanism: the IMAGO Ontology.

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

Daniele Metilli
ISTI-CNR, via Moruzzi 1, 56124 Pisa, Italy & Dipartimento di Informatica, Università di Pisa, Largo Bruno Pontecorvo 3, 56127 Pisa, Italy.
E-mail: daniele.metilli@isti.cnr.it

Nicolò Pratelli
ISTI-CNR, via Moruzzi 1, 56124 Pisa, Italy.
E-mail: nicolo.pratelli@isti.cnr.it

Paolo Pontari
Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa
Via S. Maria 36, 56126 Pisa, Italy.
E-mail: paolo.pontari@unipi.it

Abstract

In this paper we present the first achievement of the The IMAGO – Index Medii Aevi Geographiae Operum – Italian National Research Project (2020-2023), that is the ontology we have created in order to formally represent the knowledge about the geographical works written in Medieval and Renaissance Humanism (VI-XV centuries). The IMAGO ontology is derived from a strict collaboration between ISTI-CNR and the scholars who are involved in the project, who have supported ISTI-CNR in defining a conceptualisation of the domain of knowledge. Following the re-use logic, we have selected as reference ontologies the CIDOC CRM (Doerr, 2003) vocabulary and its extension FRBRoo (Doerr et al., 2008), including its in-progress reformulation, LRMoo (Riva and Žumer, 2017). This research is included in a wider project context whose final aim is the creation of a knowledge base of Latin geographic literature of the Middle Ages and Renaissance Humanism in which the data are formally represented following the Linked Open Data paradigm and using the Semantic Web languages. At the end of the project, this knowledge base will be accessed through a Web application that allows retrieving and consulting the collected data in a user-friendly way for scholars and general users, e.g. tables, maps, CSV files.

1. Introduction

IMAGO – Index Medii Aevi Geographiae Operum – is a three-year (2020-2023) Italian National Research Project (PRIN) that aims at realising new tools, based on Semantic Web technologies, to support scholars in the study of geographical works written in Medieval and Renaissance Humanism.

The tools will allow scholars to create and access a collection of Latin works which define the knowledge, description and representation of the world in the VI-XV centuries. In particular, the project aims at creating a knowledge base in which the data are formally represented following the Linked Open Data paradigm and using the languages of the Semantic Web. This knowledge base will be accessed through a Web application that allows retrieving and consulting the collected data in a user-friendly way for scholars and general users, e.g. tables, maps, CSV files.

The image of the world that the Medieval and Renaissance culture created throughout ten centuries is crucial to understand the level of geographical knowledge and the development of western thought in European history. During the Middle Ages, geographical descriptions were mostly functional to collect the human knowledge into encyclopedic works or to provide universal chronicles with an essential overview. Specific descriptions of lands, cities, places, monuments and buildings were also supplied as a guide to the pilgrims traveling to the Holy Land, Rome and Santiago de Compostela. By the end of the Middle Ages and the beginning of Renaissance Humanism, a more and more clear image of the World was defined thanks to the discovery of ancient geographical models (especially Greek works by Ptolemy and Strabo). After this period, the genre of geographical description had a further and decisive turning point, due to the exploration travels and discoveries: the description and representation of the New World, together with the reassessment of the physical space, brought about an epochal revolution. To the best of our knowledge, until now in this field of studies, no scientific research that has applied digital methods in a systematic way was conducted and an overall study which highlights the importance of this literature from a historical-literary point of view is needed.
As the first step in order to develop tools to support scholars in creating, evolving and consulting a knowledge base of the geographical works written in Medieval and Renaissance Humanism, we created an ontology that formally represents this knowledge. The IMAGO ontology is derived from a strict collaboration between ISTI-CNR and the scholars who are involved in the project who defined together a conceptualisation of the domain of knowledge. Following the re-use logic, we selected as reference ontologies the CIDOC CRM (Doerr, 2003) vocabulary and its extension FRBRoo (Doerr et al., 2008), including its in-progress reformulation, LRMoo (Riva and Žumer, 2017).

The paper is organised as follows: in Section 2 we report the state of the art of the studies of Latin geographic literature of the Middle Ages and Renaissance Humanism and the digital projects, archives and ontologies that are useful to represent knowledge in this field. Section 3 describes the methodological approach we have followed to develop an ontology for the IMAGO project. Section 4 introduces in an informal way the knowledge about the geographical works we are interested in representing. In Section 5 we formally express the conceptualisation and we describe the IMAGO ontology. In Section 6 our conclusions are reported.

2. State of the Art

Latin geographic literature of the Middle Ages and Renaissance Humanism (VI-XV centuries) has never been the subject of an overall and systematic scientific examination using digital methods so far. To start a research in this field, the recovery of the Medieval and Renaissance Humanistic geographical texts is necessary in order to make a full and complete screening of this literature from both historical-critical and philological-ecdotic point of views.
A fundamental framework to categorise Medieval travel literature is in (Menestò, 1994), in which the author defines the specific features of this literature across the Medieval centuries, making a clear classification of the literary genres (i.e. *itineraria*, *descriptiones*, narrations of the crusades, ambassadors and missionaries reports, imaginary journeys and *mirabilia*). About the rebirth of

geographical science in Renaissance Humanism, the most extensive and comprehensive survey is reported in (Defilippis, 2001). However, historical-geographical overviews are also reported in (Bouloux, 1999) and in (Defilippis, 2009). A significant starting point for the study of Latin travel and geographical texts in the Middle Ages and Renaissance Humanism and of their manuscript tradition can be found in critical editions and critical studies of specific works such as the ones reported in (de Rubrouck, 2011; Chiesa, 2020; Stocchi, 1963; Pontari, 2016). Furthermore, studies and critical editions realised by SISMEL[1] (Società Internazionale per lo Studio del Medioevo Latino/Italian Society for the study of Latin Middle Ages) constitute an authoritative corpus to study Medieval geography. Another important source of medieval geographical texts is the "Repertorium fontium Historiae Medii Aevi" by August Potthast (Potthast, 1962), which contains many information about works belonging to the geographical genre, their tradition and critical bibliography.

To perform a complete and systematic analysis of geographical works, authoritative digital archives that are particularly interesting in our research are ALIM[2] – Archivio della Latinità Italiana del Medioevo (Archive of the Italian Latinity of the Middle Ages), MIRABILE[3] – Archivio digitale della cultura medievale (Digital Archive for Medieval Culture), and ENSU – Edizione Nazionale dei testi della Storiografia Umanistica (National Edition of Texts of the Humanism Storeography).

With regard to specific Web resources devoted to the field of geography, there are several on-line dictionaries, especially useful for detecting and normalizing toponyms. Among these, the most relevant in our project are the Getty Thesaurus of Geographical Names[4], Histograph[5] and Trismegistos[6]. However, these geographic dictionaries are limited resources that do not allow data interconnection. A step forward in sharing and reusing data is represented by two collaborative projects such as Pelagios[7] and Pleiades[8].

During the last years, some specific vocabularies have been developed to represent geographic knowledge. For example, the GeoNames Ontology[9] allows representing the features of geographic places using the Web Ontology Language (OWL) (McGuinness and Van Harmelen, 2004). The GeoNames knowledge base has collected over 11 million geographic places represented using the terms defined in the GeoNames ontology, and each place is denoted by an IRI, following the Linked Data paradigm. The GO! ontology (Lana and Tambassi, 2011) is another vocabulary, developed within the Geolat project, that allows access to the geographical knowledge contained in the classical Latin texts included in the digilibLT[10] digital library. GO! describes the geographical entities with their boundaries, the mereological and topological relationships, the coordinates, their spatial representation and their literary, historical and cultural features. A further geographical vocabulary is the Geographical Entity Ontology[11]. This ontology was developed to represent geopolitical entities (such as sovereign states and their administrative subdivisions) as well as various geographical regions (including but not limited to the specific ones over which the governments have jurisdiction). The Geographical Entity Ontology is implemented in OWL and based on the Basic Formal Ontology

[1] https://www.sismelfirenze.it/
[2] http://en.alim.unisi.it
[3] http://www.mirabileweb.it
[4] http://www.getty.edu/research/tools/vocabularies/tgn/index.html
[5] http://histograph.io
[6] https://www.trismegistos.org
[7] http://pelagios.org
[8] https://pleiades.stoa.org
[9] http://geonames.org
[10] http://digiliblt.lett.unipmn.it
[11] https://www.ebi.ac.uk/ols/ontologies/geo

(BFO)[12]. Finally, the Wikidata project[13] has defined a very large set of terms for representing geographic knowledge, including more than 29,000 classes and more than 700 properties expressing geographical relations[14].

3. Methodology

As the first step of the project, the scholars have started working on a census of the Medieval and Renaissance Humanism geographical Latin texts. They are using as reference study the "Repertorium fontium Historiae Medii Aevi" by August Potthast (Potthast, 1962). However, the work of the census will not be limited to the collection of data from the repertory by Potthast, but it will be extended to other bibliographic tools and catalogues, such as the "Iter italicum" by Paul Oskar Kristeller (Kristeller, 1963). A strong contribution comes from the MIRABILE and ENSU databases, especially with regards to the methods for classifying authors, texts and genres, manuscripts, editions and historical-critical bibliography. At the same time, the scholars plan to create a Medieval Latin toponymy index. This index will be the first step towards the realization of an exhaustive catalogue that will collect specific lemmas related to Medieval Latin toponyms, providing a reference point, not available until now, for detecting recurring place names into the texts of the Middle Ages and Renaissance Humanism.

The IMAGO project started on 20th January 2020, thus the census of the Medieval and Renaissance Humanism geographical Latin texts is still an ongoing activity. At the same time, we have begun working on the creation of the IMAGO ontology, which aims at formally representing the knowledge collected by the scholars. The methodology we adopted to develop this ontology can be summarised in the following steps: (i) definition of a conceptualisation of the domain of knowledge; (ii) formalisation of the conceptualisation using standard ontologies as reference vocabularies; (iii) development of the IMAGO ontology starting from the reference vocabularies; (iv) population of the ontology; (v) evaluation and refinement of the ontology.

At the current stage of the project, we have gone through the first three steps. The conceptualisation of the research domain has been defined and it is reported in Section 4. In order to express the conceptualisation in a formal way, we analysed some existing ontologies that are relevant to our work, with a preference for standards, for interoperability reasons. We adopted as reference vocabularies the CIDOC CRM and FRBRoo (and its ongoing reformulation LRMoo). Finally, we developed the IMAGO ontology as an extension of these two vocabularies, as described in Section 5.

4. Conceptualisation

This Section introduces in an informal way the knowledge about the geographical works that we are interested in representing. On the basis of the studies and the methodological approach reported in Sections 2 and 3 respectively, the idea is that the domain of the geographical work can be represented using some main categories. The first ones are the author (this is the author's name in Italian) and title (in Latin) of the works that were analysed. For each work, the literary genre has to be specified along with the toponyms that represent the locations that are described or reported into the work.

[12] https://github.com/bfo-ontology/BFO/wiki

[13] https://www.wikidata.org

[14] The classes are subclasses of "Q27096213 geographic entity" and can be retrieved through the following query: https://w.wiki/hsF. A list of geography-related properties is available here: https://www.wikidata.org/wiki/Wikidata:List_of_properties/geography

Furthermore, for each work, several metadata about the related manuscripts and print editions are reported.

Each manuscript and print edition related to a work is described using several pieces of information. 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 stored; 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, if they exist; the date of the creation of the manuscript; the secondary sources.

On the other hand, for each print 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. Figure 1 shows the categories described above in tabular format and for each category a value was reported from the work we chose as case study, that is "Descriptio insulae Cretae" by Cristoforo Buondelmonti (Florence, 1380/1390 – 1430).

Insert Figure 1 here

5. The IMAGO Ontology

In order to formally express the conceptualisation, we took into account some existing ontologies as references vocabularies. Of course, existing ontologies have been extended with notions that are suited to describe the domain we are interested in. However, it was paramount to minimise the number of such extensions, in order to reduce the idiosyncrasies in our research.

Two top ontologies were analysed in order to understand whether these are rich enough to capture the concepts described in Section 4. The first ontology is the CIDOC CRM (CRM for short), a high-level ontology that allows integrating the information contained in data of the cultural heritage domain along with their correlation with knowledge stored in libraries and archives (Doerr, 2003). The CRM achieves this by providing definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation and of general interest for the querying and exploration of such data. Since December 2006, the CRM has been recognized as an official ISO standard. This status was renewed in 2014 and can be found at ISO 21127:2014[15]. As such, it offers a stronger guarantee under many aspects: it is widely known, it is regularly revised, and it is universally accessible. The Special Interest Group[16] of the CRM continuously works for expanding the domain of applicability of the ontology, and a number of extensions have been already devised[17]. The CRM has been successfully applied to the representation of knowledge in several fields, including narrative representation (Bartalesi et al., 2017; Meghini et al., forthcoming), biography modeling (Tuominen et al., 2018), Craft Heritage (Zabulis et al., 2020) and archaeology (Niccolucci, 2017).

[15] https://www.iso.org/standard/57832.html
[16] http://www.cidoc-crm.org/sig-members
[17] http://www.cidoc-crm.org/collaborations

The second ontology we took into account is FRBRoo (Doerr et al., 2008), including its in-progress reformulation, LRMoo[18]. FRBRoo is a formal ontology intended to capture and represent the underlying semantics of bibliographic information and to facilitate the integration, mediation, and interchange of bibliographic and museum information. The FRBR model was originally designed as an entity-relationship model by a study group appointed by the International Federation of Library Associations and Institutions (IFLA) during the period 1991–1997, and was published in 1998. At the same time, the CRM was being developed independently from 1996 by the ICOM-CIDOC (International Council for Museums – International Committee on Documentation) Documentation Standards Working Group. FRBRoo is based on the idea that both the library and museum communities might benefit from harmonising FRBR with the CRM. A first version of FRBRoo was expressed in 2000 and was expanded in the following years. The latest major version of FRBRoo was published in October 2017, and a new version called LRMoo is currently in draft status. FRBRoo provides fundamental notions for text modelling that are important for our aims.
Analysing these two ontologies, we verified that they contain terms for representing all the categories and their characteristics described in the conceptualisation. In the following subsections, a detailed mapping is reported.

5.1 Representing Authors, Works, Literary Genres, Manuscripts and Print Editions

As a notational convention, the CIDOC CRM uses the letters "E" and "P" to indicate classes and properties respectively. On the other hand, FRBRoo (and its revisions LRMoo) uses the letters "F" and "R" to indicate classes and properties, respectively.

The two main categories of the conceptualisation are Author and Work. In the IMAGO ontology, this knowledge is represented using the classes E39 Actor and F2 Expression. As reported in Figure 2, the class F2 Expression is related to the class E39 Actor through the class F28 Expression Creation. F28 Expression Creation is linked to F2 Expression by the property R17 created and to the class E39 Actor by the property P14 is carried out by.

Insert Figure 2 here

The class E39 Actor is linked with the class E41 Appellation (the author's name in Italian) through the property P1 is identified by. In general, to link the Appellation IRI with the corresponding literal (a string), we use the CRM property P190 has symbolic content.

The literary genre of the F2 Expression is represented using the Genre class that we defined as a subclass of E55 Type. F2 Expression is linked to the class Genre by the property "has genre" we defined as a subproperty of P2 has type. The individuals of the class Genre are: geographic work and travel literature.

As shown in Figure 3, the toponyms are represented using the class Toponym we defined as a subclass of E41 Appellation. The class F2 Expression representing a work is linked to the class E35 Place by the property P67 refers to. The Place is linked to the class Toponym by the property "is identified by toponym" that we defined as a subproperty of P1 is identified by. For each place, the corresponding geographical coordinates are reported in order to show this knowledge on a map in a

[18] http://www.cidoc-crm.org/frbroo/ModelVersion/lrmoo-f.k.a.-frbroo-v.0.6

later stage of the project. To represent this knowledge, the class Place is linked to the class E94 Space Primitive, representing the geographical coordinates, by the property P168 place is defined by. The Expression is linked to the Toponym contained in it by the property P106 is composed of.

Insert Figure 3 here

The IMAGO ontology has to represent two types of resources: Manuscript and Print Edition.
The manuscript is represented using the class Manuscript that we defined as a subclass of F5 Item, and the print edition using the class Print Edition that we defined as a subclass of F3 Manifestation.
In the following subsections 5.2 and 5.3, we report the classes and properties we used to represent the knowledge about manuscripts and print editions.

5.2 Representing Knowledge about Manuscripts

The manuscript of a work is represented through the class Manuscript that we defined as a subclass of F5 Item. The class Manuscript is linked to the corresponding F3 Manifestation through the property R7i is materialized in. The Manifestation R4 embodies F2 Expression. Figure 4 shows the representation of the manuscript.

Insert Figure 4 here

As reported in the conceptualisation, we are interested in representing the following knowledge about a manuscript:

- *The name of the author as it is reported in the manuscript.* To represent this knowledge we linked the class F28 Expression Creation with the class Manuscript through the direct property R18 created. F28 is related to the name of the author as it is reported in the manuscript using the class E41 Appellation. To link each appellation to the manuscript in which it appears, we use the property P106i forms part of.
- *The title of the work as it is reported in the manuscript* (this is the title of the whole manuscript). The Manuscript class is linked to the class E35 Title using the property P102 has title.
- *The library in which the manuscript is stored.* It is represented with the class F11 Corporate Body and it is related to the class Manuscript using the property P50 has current keeper.
- *The location of the library***.** It is represented with the class E53 Place and it is linked to the class F11 Corporate Body through the property P74 has current or former residence. The class E53 Place is linked to the class E94 Space Primitive, representing the geographical coordinates, by the property P168 place is defined by.
- *Signature*. It is represented using the class E42 Identifier and it is related to the class Manuscript through the property P1 is identified by. To link the signature IRI with the corresponding string, we use the CRM property P190 has symbolic content.
- *Folios*. The class Manuscript is P46 is composed of E19 Physical Object that is P1 is identified by E41 Appellation.
- *The Incipit dedication/proem*. It is represented with the class E90 Symbolic Object. Each instance of the Symbolic Object class is linked to the corresponding string using the CRM

property P190 has symbolic content. The Symbolic Object class is related with the corresponding manuscript using the property "is incipit dedication of" that we defined as a subproperty of P106 is composed of.

- *The Explicit dedication/proem*. It is represented with the class E90 Symbolic Object. Each instance of the Symbolic Object class is linked to the corresponding string using the CRM property P190 has symbolic content. The Symbolic Object class is related with the corresponding manuscript using the property "is explicit dedication of" that we defined as a subproperty of P106 is composed of.
- *The Incipit of the text*. It is represented with the class E90 Symbolic Object. Each instance of the Symbolic Object class is linked to the corresponding string using the CRM property P190 has symbolic content. The Symbolic Object class is related with the corresponding manuscript using the property "is text incipit of" that we defined as a subproperty of P106 is composed of.
- *The Explicit of the text*. It is represented with the class E90 Symbolic Object. Each instance of the Symbolic Object class is linked to the corresponding string using the CRM property P190 has symbolic content. The Symbolic Object class is related with the corresponding manuscript using the property "is text explicit of"' that we defined as a subproperty of P106 is composed of.
- *Date*. The date is represented with the class E52 Time Span and it is related to the class F30 Manifestation Creation using the property P4 has time span. F30 Manifestation Creation is linked to the Manuscript using the property R24 created.
- *Secondary sources*. To represent the secondary sources we used the class Secondary Sources that we defined as a subclass of F3 Manifestation. We linked the class Secondary Sources with the class Manuscript using the class P129 is about.

5.3 Representing Knowledge about Print Editions

As reported in the conceptualisation (Section 4), we are interested in representing the following knowledge about a print edition:

- *Author*. To represent this knowledge we linked the class F28 Expression Creation with the class F2 Expression using the property R17 created. Then, we linked the class Print Edition with the class F2 Expression using the property R4 embodies. Finally, we linked the F28 Expression Creation with the E39 Actor and then Actor with the E41 Appellation. To link each Appellation to the Print Edition in which it appears, we use the property P106i forms part of. Figure 5 shows this formal representation.

Insert Figure 5 here

- *Title*. The class Print Edition is linked to the class E35 Title using the property P102 has title. To link the title IRI with the corresponding string we use the CRM property P190 has symbolic content.
- *Curator*. To represent this knowledge we have introduced the class F30 Manifestation Creation that is linked to the class Printed Edition using the property R24i was created through. F30 is linked to the class Curator we defined as a subclass of the class E39 Actor using the property "has curator" that is a subproperty of P14 carried out by. We linked the Curator class to the E41 Appellation.

- *Place*. We use the class E35 Place to represent this knowledge. The class F30 Manifestation Creation (linked to the class Print Edition) is linked to the class E35 Place using the property P7 took place at.
- *Date*. We use the class E52 Time Span to represent a date. The class F30 Manifestation Creation is linked to the E52 Time Span using the class P4 has time span.
- *Editor/Publisher*. F30 Manifestation Creation is linked to the class Publisher we defined as a subclass of the class E39 using the property "has publisher" that we defined as a subproperty of P14 carried out by. We linked the Publisher class to the E41 Appellation.
- *Format*. To represent this knowledge we used a subclass of the E55 Type, we called Format. The class Print Edition is linked to the class Format using the property R69 specifies physical form. We linked the class Format to E41 Appellation.
- *Pages*. To represent the pages of a print edition, we adopted the class E90 Symbolic Object. The print edition is linked to the class E90 using the property P106 is composed of. The class E90 is P1 identified by E41 Appellation.
- *Information about figures*. The class Print Edition is linked to a literal (string) that reports information about the figures present in the edition using the property "has figure note" that we defined as subproperty of P3 has note.
- *Notes*. The class Print Edition is linked to a literal (string) using the property P3 has note.
- *Author of the introduction, dedications, introductions*. The class Print Edition is linked to a literal (string) using the property "has introduction note" that we defined as a subproperty of P3 has note.
- *First edition/reprint*. This knowledge is represented using the class Edition that we defined as a subclass of the class E55 Type. The individuals of this class are: first edition, reprint (associated with the year of publication) facsimile and anastatica. The class Edition is linked to the related print edition through the property P2 has type.
- *Primary Sources*. To represent the primary sources we used the class Primary Source that we defined as a subclass of F3 Manifestation. We linked the class Primary Source with the class Print Edition using the property P67 refers to.
- *Ecdotic typology*. This knowledge is represented using the class Typology that we defined as a subclass of the class E55 Type. The class Typology is linked to the related print edition through the property P2 has type.
- *Secondary Sources*. To represent the secondary sources we used the class Secondary Source that we defined as a subclass of F3 Manifestation. We linked the class Secondary Source with the class Print Edition using the property P129 is about.

5.4 Representing Resources using IRIs

Following the Linked Open Data paradigm, each resource that we will create in the knowledge base (KB) will be identified by an IRI that allows accessing a description of the resource. For identifying authors and works, we decided to use, where possible, IRIs from two existing KBs: the Wikidata knowledge base and the MIRABILE database[19]. We selected Wikidata because it is one of the largest general-purpose KBs and contains thousands of descriptions of geographic entities, and MIRABILE because it is a specialized KB that describes many of the works, authors, and manuscripts that we aim to represent. The scholars provided us with a list of works and authors they intend to investigate during the project, and we mapped the entries of this list to the corresponding IRIs that we found in Wikipedia and MIRABILE.

The mapping has been accomplished using a semi-automatic tool that we developed. The tool queries the relevant knowledge bases and retrieves a set of matching IRIs. These IRIs are then checked by a

[19] http://www.mirabileweb.it

human, who approves the result or, in case of multiple results, selects the correct one. If the tool finds an existing connection between the knowledge bases (e.g. Wikidata links to MIRABILE), this connection is automatically imported in our KB.
For representing manuscripts, we are currently evaluating whether to also use MIRABILE IRIs.
For identifying geographic places and libraries, where possible, we use IRIs extracted from Wikidata.
For the other resources will be collected in our KB (e.g print edition, format, editor, signature, folios etc.), we automatically assign IRIs in the form https://imagoarchive.it/resource/ID.

6. Conclusion

In this paper we have presented an ontology developed within the Italian National Research Project IMAGO (Index Medii Aevi Geographiae Operum). The ontology aims to formally represent the knowledge about geographical Latin works, including manuscripts and print editions, which report the description and representation of the world in VI-XV centuries. Generally speaking, IMAGO aims at creating a knowledge base in which the data about these works are formally represented following the Linked Open Data paradigm and using the languages of the Semantic Web. 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. In the paper, we have reported the methodological approach that we have followed to develop the ontology. First we have defined a conceptualisation of our domain of interest, and then we have formally expressed it using two standard ontologies as reference vocabularies: the CIDOC CRM and FRBRoo (and its ongoing revision LRMoo). A detailed mapping between the concepts of the conceptualisation and the classes of these ontologies are also reported.
Now, we are working to develop a Web-based tool allowing scholars to populate the ontology with the data they are retrieving and collecting about geographical Latin works. We plan to deliver this tool at the beginning of the second year of the project, i.e. February 2021. 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.

References

Bartalesi, V., Meghini, C. and Metilli, D. (2017). A conceptualisation of narratives and its expression in the CRM. International Journal of Metadata, Semantics and Ontologies, 12(1), pp. 35-46.
Chiesa, P. (2002). Introduction, in The Travels of Friar Odoric., transl. by Sir Henry Yule, Grand Rapids (Michigan), Cambridge, William B. Eerdmans Publishing Company, pp. 1-62.
de Rubrouck, G. (2011). Viaggio in Mongolia: (Itinerarium). A cura di Paolo Chiesa. Fondazione Lorenzo Valla, Arnoldo Mondadori Editore.
Defilippis, D. (2001). La rinascita della corografia tra scienza ed erudizione.
Defilippis, D. (2009). Da Flavio Biondo a Leandro Alberti. Corografia e antiquaria tra Quattro e
Doerr, M. (2003). The CIDOC conceptual reference module: An ontological approach to semantic interoperability of metadata. AI magazine, 24(3), 75.
Doerr, M., Bekiari, C., & LeBoeuf, P. (2008). FRBRoo, a conceptual model for performing arts. In 2008 Annual Conference of CIDOC, Athens (pp. 15-18).

Meghini, C., Bartalesi, V. & Metilli, D. (forthcoming). Representing Narratives in Digital Libraries: The Narrative Ontology. Semantic Web Journal, Special Issue on Semantic Web for Cultural Heritage. Accepted for publication.

Bouloux, N. (1999). Culture et savoirs géographiques dans l'Italie du XIVe siècle (Doctoral dissertation, Paris 1).

Kristeller, P. O. (Ed.) (1963). Iter Italicum. Vol. 6: Italy III and Alia itinera IV. Brill.

Lana, M., & Tambassi, T. (2011). GO! – An ontology for the geographical knowledge contained in classical Latin texts.

McGuinness, D. L., & Van Harmelen, F. (2004). OWL web ontology language overview. W3C recommendation, 10(10), 2004.

Menestò, E. (1994). Relazioni di viaggi e di ambasciatori. Cavallo, G, 535-599.

Niccolucci, F. (2017). Documenting archaeological science with CIDOC CRM. International Journal on Digital Libraries, 18(3), pp. 223-231.

Pontari, P. (2016). «Nedum mille qui effluxerunt annorum gesta sciamus». L'Italia di Biondo e l'invenzione del Medioevo.

Potthast, A. (1962). Repertorium fontium historiae medii aevi.

Riva, P., & Žumer, M. (2017). FRBRoo, the IFLA Library Reference Model, and now LRMoo: a circle of development.

Stocchi, M. P. (1963). Tradizione medievale e gusto umanistico nel" De montibus" del Boccaccio (Vol. 39). CEDAM.

Tuominen, J.A., Hyvönen, E.A. and Leskinen, P. (2018). Bio CRM: A data model for representing biographical data for prosopographical research. In Proceedings of the Second Conference on Biographical Data in a Digital World 2017 (BD2017). CEUR Workshop Proceedings.

Zabulis, X., Meghini, C., Partarakis, N., Beisswenger, C., Dubois, A., Fasoula, M., et al. (2020). Representation and Preservation of Heritage Crafts. Sustainability, 12(4), 1461.

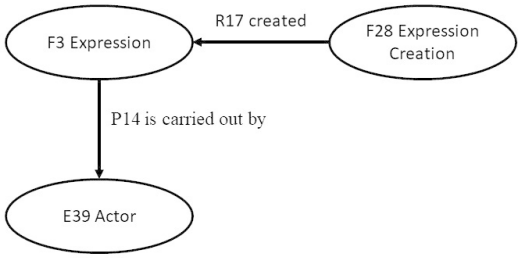
WORK	
AUTHOR	Buondelmonti Cristoforo (Firenze, 1380/1390 – 1430)
WORK	*Descriptio insulae Cretae*

MANUSCRIPT	
AUTHOR	Christofori Bondelmontis
WORK	Descriptio insulae Cretae
PLACE	Vatican City
LIBRARY	Vatican Apostolic Library
SEGNATURE	Rossiano 703
FOLIOS	ff. 1r-50v
INCIPIT OF THE DEDICATION/PROEM	-
EXPLICIT OF THE DEDICATION/PROEM	-
INCIPIT OF THE TEXT	LOREM IPSUM...
EXPLICIT OF THE TEXT	... LOREM IPSUM
DATE	1417/1422
SECONDARY SOURCES	Potthast, p. 1967, p. 606; DBI, XV, p. 199

PRINT EDITION	
AUTHOR	Cristoforo
WORK	Buondelmonti
CURATOR	E. Legrand
PLACE	Paris
DATE	1897
PUBLISHER	E. Leroux
FORMAT	-
PAGES	I-XL, 1-258
IMAGES	44 images of geographic maps out of the text
NOTES	-
AUTHOR OF INTRODUCTION – INTRODUCTION-DEDICATIONS	-
FIRST EDITION/REPRINT	First edition
PRIMARY SOURCES	Manuscrit du Serail
ECDOTIC TYPOLOGY	Critical edition with commentary with French translation
SECONDARY SOURCES	Potthast, p. 1967, p. 606; DBI, XV, p. 199

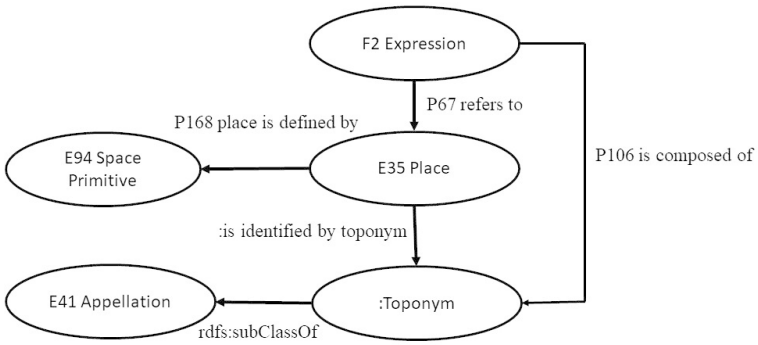
The main categories of the conceptualisation with the corresponding values extracted from the "Descriptio insulae Cretae" by Cristoforo Buondelmonti (Florence, 1380/1390 – 1430).

121x197mm (300 x 300 DPI)



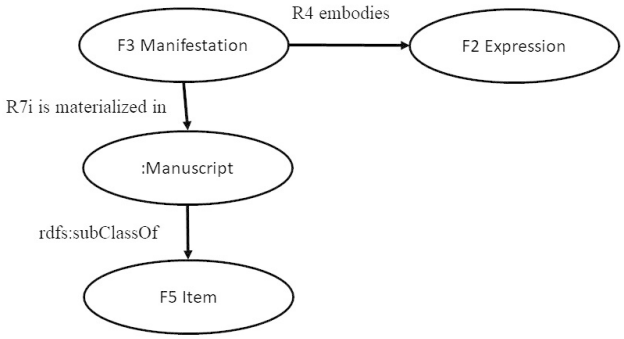
The formal representation of the author of a work.

108x60mm (300 x 300 DPI)



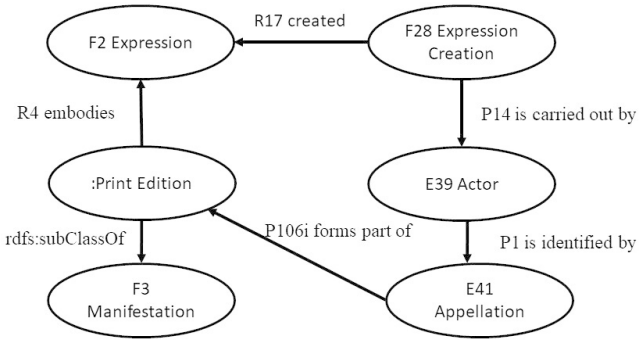
The formal representation of the toponyms to which a work refers to.

108x60mm (300 x 300 DPI)



A graphical view of the classes and properties used to represent the manuscript.

108x60mm (300 x 300 DPI)



A graphical view of the classes and properties used to represent the printed edition.

108x60mm (300 x 300 DPI)