Noname manuscript No.
(will be inserted by the editor)

Linking Different Scientific Digital Libraries in the Digital Humanities Domain: The IMAGO Case Study

Velentina Bartalesi · Nicolò Pratelli · Emanuele Lenzi

Received: date / Accepted: date

Abstract In the last years, several Scientific Digital Libraries (DLs) in Digital Humanities (DH) field have been developed following the Open Science principles. These DLs aim at sharing the research outcomes, in several cases as FAIR data, and at creating linked information spaces. In several cases, to reach these aims the Semantic Web technologies and Linked Data have been used. This paper presents how the current scientific DLs in the DH field can provide the creation of linked information spaces and navigational services that allow users to navigate them, using Semantic Web technologies to formally represent, search and browsing knowledge. To support the argument, we present our experience in developing a scientific DL supporting scholars in creating, evolving and consulting a knowledge base related to Medieval and Renaissance geographical works within the three years (2020-2023) Italian National research project IMAGO - Index Medii Aevi Geographiae Operum. In the presented case study, a linked information space was created to allow users to discover and navigate knowledge across multiple repositories, thanks to the extensive use of ontologies. In particular, the linked information spaces created within the IMAGO project make use of five different datasets, i.e. Wikidata, the MIRABILE digital archive, the Nuovo Soggettario thesaurus, Mapping Manuscript Migration knowledge base, and the Pleiades gazetteer. The linking among different datasets allows to considerably enrich the knowledge collected in the IMAGO KB.

Keywords Linking information spaces · Semantic Web · Ontology · Linked Open Data · IMAGO project

Valentina Bartalesi
Institute of Information Science and Technologies "Alessandro Faedo", CNR, via G. Moruzzi 1, 56124 Pisa (Italy)
E-mail: valentina.bartalesi@isti.cnr.it

Nicolò Pratelli
Institute of Information Science and Technologies "Alessandro Faedo", CRN, via G. Moruzzi 1, 56124 Pisa (Italy)
E-mail: nicolo.pratelli@isti.cnr.it

Emanuele Lenzi
Institute of Information Science and Technologies "Alessandro Faedo", CNR, via G. Moruzzi 1, 56124 Pisa (Italy)
E-mail: emanuele.lenzi@isti.cnr.it

1 Introduction

In the last years, in Digital Humanities (DH) domain, several scientific Digital Libraries (DLs) have been created following the Open Science principles [7]. This approach allows sharing the knowledge collected in the DLs as FAIR (Findable Accessible Interoperable and Reusable) data [23] and creating linked information spaces. In several cases, to reach these aims the Semantic Web (SW) technologies [22] and Linked Data (LD) [8] have been used.
Initially, these technologies have supported semantic interoperability and data aggregation, creating data models for harmonizing data (e.g. RDF, SKOS) and vocabularies able to provide names and define the classes and relations to formally represent different domains of reality (e.g. CIDOC CRM, FOAF, FRBRoo, LRMoo, Europeana Data Model). Several DLs were developed based on harmonizing data. These DLs made available facilities for information retrieval such as semantic searches that highlight the relations among entities and semantic recommendation systems [33] for exploring and browsing data in meaningful ways. Furthermore, several ontologies have been developed to represent particular aspects of domains of reality that needed specific classes and relations. These ontologies were usually developed as extensions of standard vocabularies.

On the basis of these ontologies, DLs were developed that allow users to access and visualise the collected knowledge not only through a SPARQL end-point [53] but also using predefined searches that show the retrieved knowledge in user-friendly formats like maps, tables, graphs, etc. At the same time, these DLs publish this knowledge as Linked Open Data [6]. This paper presents how the current scientific DLs developed in the DH field can provide the creation of linked information spaces and navigational services using Semantic Web technologies to formally represent, search and browse knowledge. To support this argument, we present our experience within the three years (2020-2023) Italian National research project IMAGO - Index Medii Aevi Geographiae Operum, where we are developing a scientific DL supporting scholars in creating, evolving and consulting a knowledge base (KB) of Medieval and Renaissance geographical works and toponyms. In particular, we created a linked information space using five different datasets, i.e. Wikidata, the MIRABILE digital archive, the Nuovo Soggettario thesaurus, Mapping Manuscript Migration knowledge base, and the Pleiades gazetteer, which has allowed to significantly enrich the knowledge collected in the IMAGO KB.

The article is structured as follows: to understand the context and the last developments of the DLs using SW technologies in the DH field, Section 2 reports an overview of these systems in different DH domains. Section 3 presents the IMAGO project and the ontology that was developed to formally represent the knowledge related to a collection of Medieval and Renaissance geographical works. In Section 4 a detailed description of how different scientific DLs are linked to populate the IMAGO ontology was reported. Section 5 describes a semi-automatic Web tool we developed to easy populate the ontology, and then to create the IMAGO KB as an OWL graph. Section 6 reports our final remarks and future work.

2 DLs and Semantic Web in the DH domain: An Overview

The advent of the Semantic Web and the Linked Open Data paradigm has fostered the development of DH projects that use these technologies to formally represent the collected data. These technologies support the production of findable, accessible, interoperable, and reusable (FAIR)[1] data [50]. In recent years, several projects have focused on different aspects of DH, successfully using the Semantic Web technologies. Initially, the Semantic Web technologies supported semantic interoperability and data aggregation, creating models for harmonising data. Examples of well known data models are RDF [44], SKOS[2], Dublin Core[3]. These data models were used to represent relations between terms of several thesauri that are useful sources for Linked Open Data (LOD) approaches, e.g. Nuovo Soggettario Nazionale[4] and iDAI.thesauri[5] that both use the semantic relations defined in SKOS. At the same time, several ontologies have been developed to represent knowledge formally. Well known examples of these are CIDOC CRM [17], FRBR [19], Europeana Data Model [20], DOLCE upper level ontology [24]. Taking these semantic models as reference vocabularies, many other ontologies in the DH field were developed to represent more specific knowledge domains, e.g. the Sampo model [29] or DanteSources ontology [4]. We have to notice that other ontologies, especially those created during the years shortly after the Semantic Web advent, were developed as *had hoc* project vocabularies with a very low level of interoperability. Based on the above ontology models, several DLs were developed that allow users to access and visualise the data collected in the KBs through SPARQL end-points and user-friendly interfaces. Some of these DLs, which are more interesting for the aims of this paper, are reported and described in the following, organised by the research field.

A very explored domain in DH is the Cultural Heritage field, in which several DLs were developed that use their own or external ontology models for representing data. For example, Storyspace [51] is an ontology-based system that allows describing stories based on events that span museum objects. The focus of the system is the creation of curatorial narratives from an exhibition. Each digital object has a linked creation event in its associated heritage object story. The system is based on an *ad hoc* ontology to formally represent heritage content by capturing the underlying stories of both the individual heritage object and the curated exhibits as a whole. The ontology and software tool were used by two Irish heritage institutions, the Irish Museum of Modern Art (IMMA) and the National Gallery of Ireland (NGI) to create a DL collecting the stories of two exhibitions, one on the Dutch painter Gabriel Metsu and one on Irish Modernism. The stories are created using the digital objects collected by the two heritage institutions. Another example is Bletchley Park Text [40], an application that helps users to explore museums' collections. Visitors express their interests in some spe-

[1] https://www.go-fair.org/fair-principles/

[2] https://www.w3.org/2004/02/skos/

[3] https://dublincore.org/

[4] https://thes.bncf.firenze.sbn.it/index_eng.html

[5] http://thesauri.dainst.org/de.html

cific topics using SMS messages containing keywords. The semantic description of the resources is used to organise a collection into a personalised website based on the keywords chosen by the user. The system relies on a *ad hoc* ontology of story developed within the Story Fountain project [41]. The stories represented in the system are exploited to create relations among the entities in the online collections, allowing the user to query the system for a semantic path among entities. The system was tested for the Bletchley Park Museum, based in Milton Keynes, UK. Bletchley Park was a wartime code-breaking centre that developed the Colossus computer. In the early 1990s, it became a heritage centre focusing on the history of computing and code-breaking. The museum DL describes the history of Bletchley Park, its buildings, important characters, work and inventions. Furthermore, Bletchley Park DL has an extensive archive which includes hundreds of historical accounts and interviews with people connected to the museum. Another CH project is the Labyrinth 3D system [15] that integrates the semantic annotation of cultural objects with the interaction style of 3D games. The system immerses users into a virtual reality, where a user can explore the collection of a DL using paths representing the semantic relations over cultural objects. The exploration of the DL is mediated through a set of cultural archetypes, or narrative structures, which are modelled according to the Archetype Ontology [14]. This ontology was designed with the goal of interoperability with some standard ontologies, such as Functional Requirements for Bibliographic Records (FRBR) and Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE). Another successful example of the use of the Semantic Web technologies in the CH domain is CultureSampo [31], a DL and a publication channel for Finnish cultural heritage. CultureSampo was the first project that applied these technologies to the Finnish cultural heritage. The contents of CultureSampo come from 22 museums, archives, and libraries. The DL contains digital objects of 67 different content types: artefacts, paintings, drawings, sculptures, abstract art, novels, comics, web pages, three types of folklore, five types of folk music, photos, aerial photos, persons, organisations, biographies, historical events, skills, videos, buildings, archaeological sites, and others. The metadata is represented using 204 different properties and 253 additional properties used in several ontologies. The aggregated knowledge base contains 134,000 cultural objects and 285,000 additional cultural resources (persons, places etc.), described by 2.7 million triples. CultureSampo uses a CRM-based ontological model that allows linking events with digital objects. Following the CultureSampo experience, BiographySampo [30] aims to develop a system to extract narratives from biographical dictionaries, represent them in a formal way using the CIDOC CRM model and publish them on the Web as Linked Open Data (LOD). The system has been used to build a DL containing more than 13,000 biographies of historical people extracted from the Finnish Literature Society[6]. The aims of the project are to 1) transform textual biographies into LOD by using knowledge extraction techniques; 2) enrich the data by linking it to internal and external data sources and by reasoning; 3) publish the data as a Linked Data service and a SPARQL endpoint on the web; and 4) create a DL on top of the service, enriched with data-analytic tools and visualisations for distant reading of Big Data, i.e., for Digital Humanities research. Using a CRM-based ontological model, these two "Sampo" projects facilitate the semantic interoperability and reasoning about the events of the world, digital objects and stories. These advantages are exploited in the implementations of a dozen of online cultural heritage "Sampo portals" in the period 2002—2021[7]. A more recent and interesting example of a DL based on the Sampo model is the one developed within the Mapping Manuscript Migrations (MMM) project [35]. MMM harmonized and published heterogeneous premodern manuscript metadata as LOD. A semantic portal and a LOD service were created to integrate distinct manuscript datasets into a shared platform for search and discovery. In particular, thanks to the use of the Sampo model, MMM (i) supports collaborative data creation and publishing, (ii) uses a shared ontology infrastructure; (iii) provides multiple perspectives to the same data through a set of different visualizations; (iv) supports knowledge discovery in addition to data exploration, using a reasoning engine. The MMM services are available online[8] and are based on metadata of over 220,000 manuscripts from the Schoenberg Database of Manuscripts[9] of the Schoenberg Institute for Manuscript Studies (University of Pennsylvania), the Medieval Manuscripts in Oxford Libraries[10], and Bibale[11] of Institut de recherche et d'histoire des textes in Paris. Finally, we have to notice that the *Sampo* model is based on standards and best practices of W3C for Linked Data publishing and supports FAIR principles [29]. Thus all the *Sampo*-based projects are compliant with these paradigms.

[6] https://www.finlit.fi/en
[7] https://seco.cs.aalto.fi/applications/sampo/
[8] https://mappingmanuscriptmigrations.org/en/
[9] https://sdbm.library.upenn.edu/
[10] https://medieval.bodleian.ox.ac.uk/
[11] https://bibale.irht.cnrs.fr/

Regarding the DH projects in the historical domain, two well-known examples of the semantic approaches are Histropedia and Chronas.
Histropedia[12] is a system that allows users to create or view timelines on topics of their choice by importing statements from Wikidata. Links to related Wikipedia articles and Wikimedia Commons images are automatically added, resulting in rich spatio-temporal visualisations. The scope of the project includes research, education, and tourism applications. On the other hand, Chronas[13] is an open-source and semantic-based chronological and cartographical application with a special focus on visualising maps of the world across human history. The corresponding Wikipedia page is visualised for each piece of knowledge reported in the application (e.g. Byzantine Empire, Battle of Thessalonica, Persians). Furthermore, it is possible to browse the history by clicking the Wikidata entities linked to the corresponding Wikipedia pages.

Semantic Web technologies have recently been used to explore a corpus of data in the history of the science field, that is, the Henri Poincaré correspondence [10]. The Poincaré correspondence is a corpus of letters sent and received by the mathematician. The edition of this correspondence is a long-term project that began during the 1990s. Since 1999, the Henri Poincaré website[14] has been devoted to publishing a DL of this correspondence with digitized letters. The corpus consists of around 2,100 letters, 1,126 sent by Henri Poincaré and 956 received by him. Original letters come from 63 different archives and libraries in 14 countries. The letters collected on the DL are semantically annotated using the Dublin Core extended metadata model, and an RDF/S KB of the metadata was created. Dublin Core is a basic semantic model, and the researchers involved in the project stated that an epistemological investigation to enrich the model and link it to other standard ontologies in digital humanities, such as the CIDOC CRM, is required [10]. In particular, for each letter, the following information is reported: a physical description (e.g. type of letter - telegram, autograph letter, minutes, etc. –, number of pages, location of the letter within the archive) and the content of the letters (e.g. people or publication quoted, mathematical theories or formulae, philosophical concepts).

The LOD approach is also used for exploring datasets in the ancient geography field. The Pelagios - Enable Linked Ancient Geodata In Open Systems [45] is a collective of projects connected by a shared vision of a world in which the geography of the past is as interconnected, interactive and interesting as the present. The Pelagios project created a Linked Open Data source to interlink online resources that refer to places in the historical past using the SKOS data model to represent the similarity between terms (i.e. skos:closeMatch).
Pleiades[15] provides historical geographic information about the ancient world in digital form. At present, Pleiades has extensive coverage for the Greek and Roman world and is expanding into Ancient Near Eastern, Byzantine, Celtic, and Early Medieval geography. To represent the data, Pleiades uses terms from different data models, that is Dublin Core, SKOS, FOAF[16], Citation Ontology[17]. The Pleiades dataset may be downloaded in RDF format.

Semantic web technologies were also applied to archaeological collections. Collections are groupings of archaeological resources related to some archaeological activity (such as excavations) or to some research activity in the context of which the members of the collection have been created [39]. These collections often include textual documents or images, artefacts and other findings. In addition, a considerable number of published papers and reports is available that contain the results of an analysis activity of the found artefacts. In this domain, we can cite DECHO, a framework for the acquisition, ontological representation and visualisation of knowledge about archaeological objects [2]. In particular, DECHO was designed to help archaeology researchers and educators to reconstruct and analyse the historical context of cultural heritage objects. DECHO provides an environment able to address different types of data like images, pictures, X-rays, drawings, several types of metadata such as from chemical tests, and 3D models of digitised objects. The information core of DECHO is a semantic model that provides a higher-level view of the database contents used in the project, integrates data stored in other databases, and supports object retrieval, browsing, and knowledge navigation. The DECHO semantic model is derived from the CIDOC CRM by selecting only the classes and properties useful for the project. Other two projects that deserve to be mentioned in this research domain are ARIADNE (Advanced Research Infrastructure for Archaeological Dataset Networking in Europe) and ARIADNE+. Within these two European projects, a model for representing archaeological collections was developed. The ARIADNE Catalogue Data Model (ACDM for short) [32]

[12] http://histropedia.com
[13] http://chronas.org.
[14] http://henripoincare.fr/
[15] https://pleiades.stoa.org/home
[16] http://xmlns.com/foaf/0.1/
[17] http://purl.org/spar/cito/

was specifically developed to produce a detailed, formal, and unambiguous representation of the archaeological information of legacy archives. The ACDM was built around the DCAT vocabulary [37], which was expanded by adding classes and properties needed for best describing the ARIADNE assets. DCAT[18] is an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web. DCAT enables a publisher to describe datasets and data services in a catalogue using a standard model that facilitates the consumption and aggregation of metadata from multiple catalogues. ACDM was enriched with properties for specifying the access policy and the original identifier of the resources and specialised classes, e.g. ACDM:DataResource, whose instances represent the various types of data containers owned by the ARIADNE partners and lent to the project for integration, ACDM:LanguageResource, having as instances vocabularies, metadata schemas, gazetteers, and mappings, ACDM:Services, whose instances represent the services owned by the ARIADNE partners. The interoperability of the semantic models developed in DECHO and ARIADNE/ARIADNE+ is guaranteed since both are CRM-based ontologies.

In the literary domain, an interesting example of the use of the Semantic Web technologies and LOD paradigm is the Digitised Manuscripts to Europeana (DM2E) project. Within DM2E, a Linked Data source [3] of digitised manuscripts was developed. The Linked Data source provides metadata and links for direct access to digitised content from various cultural heritage institutions across Europe. The data model used in the project is a specialisation of the Europeana Data Model (EDM) [20] and satisfies specific requirements from the domain of manuscripts and old prints, as well as from developers who want to create applications on top of the data. EDM itself is very generic to represent resources provided by museums, libraries, archives and galleries all over Europe. It is based on top-level data models like Dublin Core and SKOS. The DM2E model adds mostly subclasses and properties for the domain of manuscripts to existing terms of the EDM. Main additions were made for person roles, e.g. dm2e:composer, for classes that are used to indicate the cultural heritage object type like dm2e:Manuscript, for properties to describe specifics of manuscripts or other document types of the model, e.g. dm2e:incipit for the opening words of a manuscript or dm2e:receivedOn for a date on which a letter was received. In this domain, Semantic Web technologies are also used to formally represent and explore Dante Alighieri's primary sources. DanteSources[19] [4] was the first project to provide a semantic representation of the knowledge related to the primary sources of Dante Alighieri's so-called minor works. To do so, DanteSources employed the standard technologies of the Semantic Web, such as RDF and SPARQL. The data collected in the DanteSources KB is compliant with the FAIR principles. The Hypermedia Dante Network is a current Italian National Research project that plans to extend the experience of DanteSources, focusing on the semantic representation of the primary sources of Divine Comedy. Following the Linked Data paradigm, HDN aims at reusing contents and knowledge, making them standardised and accessible. This is why the commentaries used by HDN to identify Dante Alighieri's primary sources are taken from the corpus collected by the Dartmouth Dante Project[20], which includes commentaries from the 14th to the 20th century written in Latin, Italian and English. Like DanteSources, the knowledge collected in the HDN KB is compliant with the FAIR principles. Another example of the use of the Semantic Web technologies in the literary domain is the semantic digital scholarly edition of Paolo Bufalini's notebook[21]. This is a project that aims to identify, analyse and enhance the intratextual and extratextual relations which characterise the texts of the notebook. The results are presented as a knowledge graph where nodes represent the entities of the edition, and the arcs represent different types of relations between such entities. The project uses existing ontologies for the representation of the collected knowledge (in particular, SPAR Ontologies[22], Dublin Core and Open Annotation[23]). Furthermore, it makes the data accessible through a SPARQL endpoint equipped with a graphical interface.

Finally, we have to mention PARTHENOS[24], a European project and a research infrastructure whose objective is to strengthen the cohesion of research across a number of related domains associated with the humanities. This broad sector includes linguistic studies, cultural heritage, history and archaeology and existing research structures such as ARIADNE (archaeology), CLARIN (languages) and DARIAH (arts and humanities). Within the PARTHENOS project, the Parthenos Entities (PE) model [11] was developed. It is an ontological model and RDF schema to encode data of

[18] https://www.w3.org/TR/vocab-dcat-3/

[19] https://dantesources.dantenetwork.it

[20] https://dante.dartmouth.edu/

[21] https://projects.dharc.unibo.it/bufalini-notebook/

[22] https://sparontologies.github.io/

[23] http://www.openannotation.org/spec/core/

[24] https://www.parthenos-project.eu/

use in supporting the activities and aims of research infrastructures to pool and connect services, software, and datasets and to enable users of such services to reach the actors and understand the knowledge generation processes which generated the offered datasets. PE is modelled as an extension of CIDOC CRM, and CRMdig [21], an extension of the CRM which models provenance information in digitization processes.

To sum up, we have to notice that most of the projects reported above extend standard ISO or *the facto* vocabularies to represent their data. We can mention all the *Sampo* projects but also the projects in the archaeological and literary domains. Some other projects, e.g. the projects developed in the historical domain, are completely based on already created ontological models (e.g. Wikidata). On a higher conceptual level, ontological models were also developed in the European research infrastructures' projects for pooling and connecting services, software and datasets. A reported example is the Parthenos Entities model. Other projects used simple data models to represent their data, e.g. the projects for representing ancient geography data, which use Dublin Core and SKOS. Finally, few systems, which are developed within the oldest projects, created their own vocabularies to represent data, significantly limiting their interoperability, i.e. Storyspace, Bletchley Park Text.

The main lesson we learned from the previously analysed projects is the importance of reusing the available standard technologies, semantic models and KBs. Indeed, following the PARTHENOS approach, we created the IMAGO ontology as an extension of the ISO standard CIDOC CRM vocabulary. This maximises its interoperability and allows linking the IMAGO KB with other CRM-based KBs (e.g. MMM or all *Sampo* projects). Furthermore, like ACDM was enriched with properties for specifying the original identifier of the resources and specialised classes, IMAGO ontology defined its own classes and properties to describe the specific knowledge on the manuscripts and toponyms collected in the project. Considering the Chronas application, IMAGO aims to implement an open-source and semantic-based cartographical functionality to visualise the Medieval and Renaissance Latin toponyms collected in the project on a map. Furthermore, following the Historipedia example, the corresponding Wikidata entity is visualised for each toponym. The Wikidata entity is linked to the corresponding Wikipedia and Pleiades pages. Following the requirements collected in the *Sampo*-based, DanteSources and HDN projects, we designed and implemented IMAGO to get its KB compliant with the FAIR principles and LOD paradigm.

3 The IMAGO Project and Ontology

IMAGO - Index Medii Aevi Geographiae Operum is a three years (2020-2023) Italian National research project that aims at providing a systematic overview of the Medieval and Humanistic geographical literature using the Semantic Web technologies. In particular, the project aims to make available this knowledge as Linked Open Data (LOD) and develop automatic search and visualisation services on the collected data. Indeed, the image of the world created by Medieval and Renaissance culture was crucial to the development of Western thought in European history. During the Middle Ages, geographical descriptions of lands, cities, places, monuments and buildings were used as guides to the pilgrims travelling to the Holy Land, Rome and Santiago de Compostela. Furthermore, geographical descriptions were often reported in encyclopedic works or universal chronicles. The genre of geographical description had a further and decisive turning point during the period of explorations, 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. Despite the importance of this literature, until now, Medieval and Renaissance geographical works have not been studied using digital methods.

3.1 Methodology

As the first step of the project, the scholars have started creating a collection of Medieval and Renaissance geographical Latin works. They are using the Repertorium fontium Historiae Medii Aevi [43] as reference study, along with other bibliographic tools and catalogues, such as the 'Iter italicum' by Kristeller [36], MIRABILE - Archivio digitale della cultura medievale (Digital archive of the medieval culture)[25], and ALIM - Archivio della Latinità Italiana del Medioevo (Archive of the Italian Latinity of the Middle Ages)[26]. At the same time, the scholars are creating a Medieval Latin toponymy index, starting from the toponyms reported in the Latin works of the Italian poet Dante Alighieri. This index will be the first step towards the realisation of an exhaustive catalogue that will collect specific lemmas related to Medieval Latin toponyms, providing a reference point not available until now for detecting

[25] http://www.mirabileweb.it/
[26] http://en.alim.unisi.it/

recurring place names in the works of the Middle Ages and Renaissance Humanism.

As the second step, semi-structured interviews were performed with four scholars who are involved in the project in order to define a conceptualisation of our knowledge domain. We asked the scholars to identify the relevant knowledge to represent and to suggest how this knowledge should be further elaborated. On the basis of the collected requirements, we created an ontology providing a formal representation of the knowledge about Medieval and Renaissance geographical Latin literature. At the same time, we developed a semi-automatic tool for populating the ontology and storing the resulting data in a KB as an OWL graph. The final aim of the project is developing a Web application running on top of the KB that (i) makes inferences based on the contents of the knowledge base and (ii) visualises these inferences in user-friendly formats like charts, tables or maps.

Following the Linked Open Data paradigm, each resource of the KB is identified by an Internationalized Resource Identifier (IRI) that allows accessing a description of the resource itself. Following the re-use logic, we used IRIs already defined in existing KBs that we adopted as reference sources.

3.2 Conceptualisation

The conceptual idea is that the domain of the Latin geographical works can be represented using some main categories. These categories are derived from authoritative studies on the Medieval and Renaissance Latin geographic manuscripts and print editions carried out by the scholars involved in the IMAGO project and other experts. In particular, the authoritative studies from which we derived our conceptualisation are [42], [43], [38], [36], [13], [16], [9]. The first categories we identified 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 in the work. Furthermore, for each work, metadata about the related manuscripts and printed editions are added. For example, among others, for each manuscript, the library in which the manuscript is collected, the location of the library, the signature and the folios of the manuscript are reported. For each printed edition, the main metadata are: the curator's name, the place and the date of the publication, the publisher, the format of the edition, and the number of pages.

Table 1 shows the main categories described above in tabular format and for each category, a value was reported from the work we chose as case study, i.e. "De temporibus suis" by Leonardus Brunus (Arezzo (Italy), 1370/1344).

Table 1 Example of the main categories of the conceptualisation with the corresponding values extracted form the work "De temporibus suis" by Leonardus Brunus

Category	value
Author	Leonardus Brunus, n. 1370, m. 1444
Work	De temporibus suis
Toponym	Florence
Genre	Historiography
Manuscript	
Library	Vatican Apostolic Library
Library Location	Vatican city
Signature	Pal. lat. 1598
Folios	III+146+I'
Printed Edition	
Curator	C. Di Pierro
Place of publication	Bologna
Date of publication	1926
Publisher	N. Zanichelli
Format of edition	*empty*
Page	407-458

3.3 The IMAGO ontology

To formally express the conceptualisation, we developed the IMAGO ontology [5]. The methodology we followed to develop the ontology is well known and usually adopted to create formal vocabularies in the Semantic Web research field [27], that is: (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. 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. Indeed, no scientific research that has applied digital methods, and in particular the Semantic Web approach, in a systematic way has been conducted in this specific field of research. Currently, such information is dispersed on paper books, and this makes a systematic overview of the geographic Latin 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 scholars, students and general users. The IMAGO ontology is created as an extension of two standard vocabularies: the CIDOC CRM and FRBRoo (and its ongoing version LRMoo). The CIDOC CRM [17] (CRM for short) is a

high-level ontology that allows integrating the information contained in data of the Cultural Heritage domain along with the correlation with knowledge stored in libraries and archives. 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 for the querying and exploration of such data. Since December 2006, CRM has been recognised as an official ISO standard. This status was renewed in 2014 and can be found in ISO 21127:2014. Properties and classes defined in CRM provide an efficient model for an event-based representation of the information. An event-based approach has several advantages compared to traditional, non-event-based approaches. First of all, events provide a meaningful way of describing links between things and the actions of human beings. Secondly, events that form the history of an item and new events can be added at any time. These advantages are exploited in the IMAGO ontology, for example, to represent the link between the event of a manuscript creation and the manuscript itself or to add to the same conceptual work, both the events of the creation of the related manuscript and of the creation of the related print edition.

The second ontology we took into account is FRBRoo [18], including its in-progress reformulation, LRMoo. 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. FRBRoo is based on the idea that both the library and museum communities might benefit from harmonising FRBR with the CRM. FRBRoo provides fundamental concepts for text modelling that are important for our aims. These two 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 to reduce the idiosyncrasies in our research.

Table 2 reports the mapping between the main categories identified in the conceptualisation and the classes of the IMAGO ontology, which are equivalent or subclasses of the CRM and FRBRoo classes. Table 3 shows the properties used to represent the relations between the concepts.

Table 2 Classes used to represent our main concepts

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

Table 3 Properties used to represent the relations between the main concepts reported in Table 2

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(Work,Genre)	subproperty of P2 has type
R(Work,Toponym)	subproperty of P1 is identified by
R(Manuscript,Library)	equivalent to P50 has current keeper
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

4 Linking different scientific DLs to Populate the IMAGO Ontology

Following the Semantic Web approach, all the resources collected in the IMAGO KB are identified by IRIs. These IRIs are retrieved from different sources freely available on the Web. Indeed, several DLs are linked, and different pieces of knowledge were extracted and reused to enrich the IMAGO KB. This approach, which extensively uses ontologies, allows discovering knowledge across multiple repositories and allows users to navigate the linked scientific information space. In particular, we linked the following datasets:

- Wikidata knowledge base
- MIRABILE digital archive
- Nuovo Soggettario thesaurus
- Mapping Manuscript Migration knowledge base
- Pleiades gazeteer

Figure 1 shows a graphical representation of the linked datasets along with the entity types extracted by each of them. Basically, we used SPARQL queries to detect and retrieve the entities' IRIs from the reference datasets. However, more detailed descriptions of the methods and software solutions we used are reported in the following subsections.

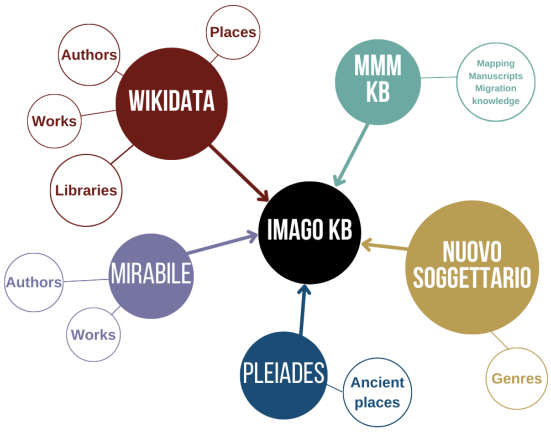

Fig. 1 Graphical representation of the IMAGO KB and the linked datasets, along with the extracted entity types

When IRIs are not available in the previous datasets, we automatically created and assigned custom IRIs to the entities.
In Table 3, for each entity type of the IMAGO ontology the percentage of retrieved IRSs is reported, along with the dataset(s) in which the IRIs are collected.

Table 4 % of linked IRIs for each dataset

% Entity Type	%	Dataset
Author	96%	Wikidata and Mirabile
Work	98%	Wikidata and Mirabile
Place	95%	Wikidata
Genre	$\simeq$ 95%	Nuovo Soggettario
Library	41%	Wikidata
Ancient place	$\simeq$ 50%	Pleiades
Manuscript Migration Knowledge	$\simeq$ 20%	Mapping Manuscript Migration

4.1 Wikidata and MIRABILE

For identifying authors and works, we decided to use, where possible, IRIs from two existing sources: the Wikidata KB and the MIRABILE database. We selected Wikidata because it is one of the largest general-purpose KBs and contains thousands of descriptions of geographic entities [49]. MIRABILE is a specialized digital archive 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 software that we developed. The tool queries the two sources and retrieves a set of matching IRIs. These IRIs are then checked by a human, who approves the result or, in case of multiple results, selects the correct one. If the software finds an existing connection between the two sources (e.g. Wikidata links to MIRABILE), this connection is automatically imported into the IMAGO KB. To identify the libraries in which the works are collected, we decided to use the IRIs of Wikidata. For each library, we retrieved the name in English and the place in which it is located. For each place, we also retrieved the name in English, the coordinates, and the State to which the place belongs to. Using IRIs from Wikidata and MIRABILE allows having a description for each resource of our KB linking other digital libraries as the LOD paradigm states. We were able to automatically retrieve 98% of the IRIs of the works that the scholars provided us in MIRABILE and Wikidata. For what concerns the authors, we found 96% of the IRIs from the same sources. Regarding the libraries, we retrieved about 41% of the IRIs in Wikidata.

4.2 Nuovo Soggettario

Scholars defined a set of literary genres composed of 40 entries. The IRIs identifying the literary genres (e.g. letter writing, Latin historiography, Latin dramaturgy) are imported from the Nuovo Soggettario[27], a standard thesaurus created and maintained by the National Central Library of Florence. Nuovo Soggettario has been developed in compliance with the International Federation of Library Associations and Institutions (IFLA)[28] recommendations and other international standards in the field of subject indexing. This thesaurus was created for general and specialized Italian libraries, especially those participating in the National Library Service (SBN)[29], and for museums, multimedia libraries, archives and documentation centres. The Nuovo Soggettario system has been used by the Italian National Bibliography since 2007 and is continuously enriched. We were able to map about 95% of the genres defined by the scholars to the entries of the Nuovo Soggettario.

4.3 Mapping Manuscript Migration

Mapping Manuscript Migrations (MMM) [12] 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 unites records from three differ-

[27] https://thes.bncf.firenze.sbn.it/index_eng.html
[28] https://www.ifla.org/
[29] https://opac.sbn.it/opacsbn/opac/iccu/informazioni_en.jsp

ent datasets: the Schoenberg Database of Manuscripts[30] at the University of Pennsylvania, the Bibale database[31] at the Institut de recherche et d'histoire des textes, and the Medieval Manuscripts Catalogue[32] at the University of Oxford. Within the MMM, an ontology was developed to satisfy the aims of the project, but it is general enough to be used by anyone who wants to represent the knowledge about the manuscript provenance data. It incorporates concepts from several existing vocabularies, including Erlangen CIDOC CRM [25] for events, FRBRoo for bibliographic information, and the Getty Thesaurus of Geographic Names for physical locations. The MMM ontology also includes 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 can enrich and extend the IMAGO KB, including the knowledge about how the manuscripts have travelled across time and space from their places of production to their current locations.
Since both MMM and IMAGO use the same reference vocabularies, the level of interoperability between the two ontologies is high, and it makes simple to merge the data from these two different KBs. Thus, we conducted a first exploratory phase of MMM KB. Indeed, to understand if some ancient manuscript in our collection is also included in the MMM KB, we performed some SPARQL queries on the MMM SPARQL endpoint. For this exploratory phase, we searched for the title and the author of the manuscript. Querying the MMM KB, we measured that about 20% of the works collected in the IMAGO KB are also present in the MMM KB. This preliminary study aims to integrate the knowledge included in MMM in our KB to enrich it. Currently, we are working on this integration.

4.4 Pleiades

One of the main aims of the IMAGO project is the creation of an index of Latin toponyms. To create this index, we have started with the toponyms used by the Italian poet Dante Alighieri in his Latin works. The scholars provided a list of these toponyms manually created by consulting DanteSearch[33], a web application that allows querying Dante's works that were previously lemmatized and grammatically annotated. The list is composed of 93 entries. They include names of cities, mountains, woods, rivers, regions, and continents.

[30] https://sdbm.library.upenn.edu/
[31] https://bibale.irht.cnrs.fr/
[32] http://www.earlymedievalmonasticism.org/Catalogues-of-Latin-Manuscripts.html
[33] https://dantesearch.dantenetwork.it/

First of all, we mapped these toponyms with the corresponding entities included in Wikidata. We retrieve 95% of the Wikidata IRIs of the toponyms. Then, we mapped these Wikidata entities with those included in the Pleiades gazetteer. Pleiades is an open-access digital gazetteer for ancient places. It provides stable Uniform Resource Identifiers (URIs) for tens of thousands (and growing) of geographic entities. Built on the Classical Atlas Project (1988-2000), which produced the Barrington Atlas of the Greek and Roman World [46], Pleiades is co-organized by the Institute for the Study of the Ancient World (NYU)[34] and the Ancient World Mapping Center (UNC Chapel Hill)[35].
Mapping the Wikidata IRIs with the ones included in the Pleiades provides each resource with a description of the current place (as Wikidata reports) and the corresponding ancient place. We were able to automatically map about 50% of the entities, that is 50% of the toponyms cited by Dante Alighieri in his Latin works are included in the Pleiades gazetteer. We performed this mapping automatically, calculating the string similarity between the names of the places as reported in our KB and Pleiades.

5 Supporting the Scholars to Populate the IMAGO Knowledge Base

To allow an easy ontology population and the creation of the IMAGO KB, we decided to develop a semi-automatic Web tool. The tool supports the scholars in reducing the time to populate the ontology and minimizing the cognitive and technical burden in selecting and identifying the resources. In particular, the tool provides the following features to support the data entry activity:

- The identification of works, authors and libraries from the specific lists of pre-defined options that the scholars defined and sent us.
- The automatic association of the IRIs to the resources collected through the tool (e.g. the IRIs associated with: the authors, the titles of the manuscripts, the literary genres, the libraries, the places and the corresponding geographic coordinates).
- The direct search of the geographical places that are interesting for the scholars in Wikidata and the automatic saving of the corresponding geographic coordinates.

Figure 2 shows the main annotation fields (author, work, place/s (i.e toponym/s) and literary genre/s) of a work by Leonardo Bruni[36], along with the annotation

[34] https://isaw.nyu.edu/
[35] https://awmc.unc.edu/wordpress/
[36] https://en.wikipedia.org/wiki/Leonardo_Bruni

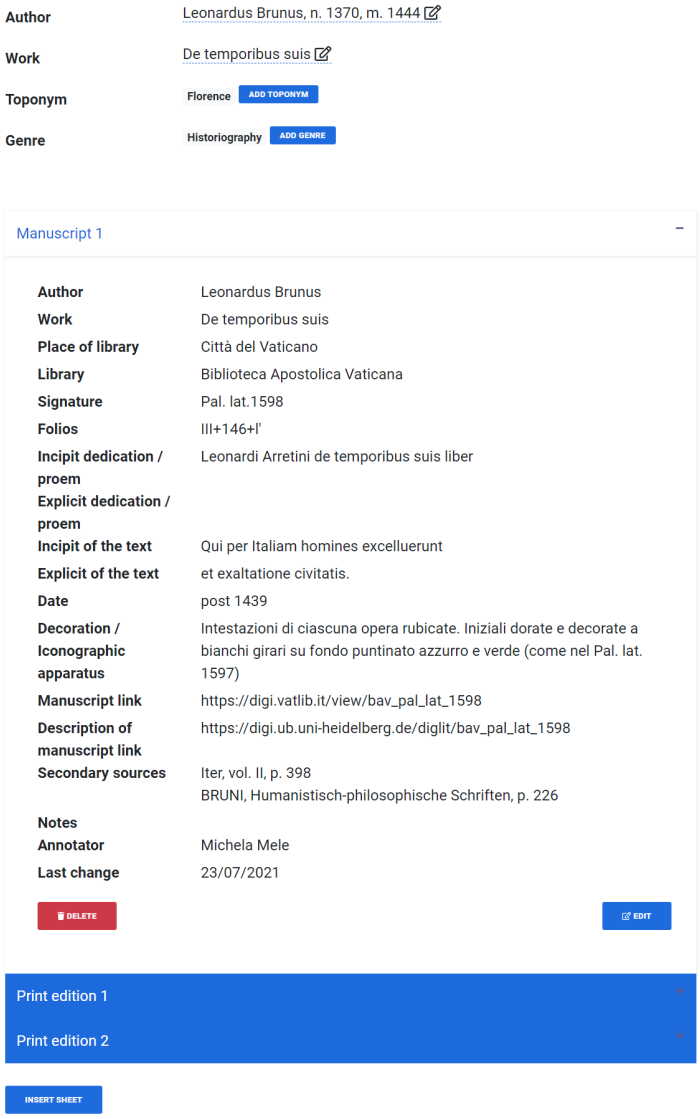

Fig. 2 The start page of the IMAGO Web tool

summary of a related manuscript. Figure 3 reports the annotation form of a printed edition for the same work. Each tool field reported in both figures corresponds to a class of the IMAGO ontology.

The tool automatically assigns the IRIs to works, authors, places, genres, and libraries. For the works and authors, the IRIs are imported from MIRABILE or Wikidata. For the places and libraries, the IRIs are imported from Wikidata. For the toponyms, the corresponding IRIs are extracted from Wikidata and the Pleiades gazeteer. The set of predefined genres is imported from the Nuovo Soggettario.
Note that Wikidata KB changes over time, just as all the other KBs we took as reference datasets. However, the tool can accommodate these changes. Indeed, following the Linked Data paradigm, the tool collects the IRIs that uniquely identify the resources. To access their descriptions, our users go directly to the corresponding pages of the Wikidata site or the pages of the other datasets. Thus, we can automatically synchronize our KB with the Wikidata and other dataset changes or updates. However, to avoid mistakes in the IMAGO KB, the descriptions associated with the IRIs are checked and assessed by the scholars involved in the project.
When IRIs are unavailable in Wikidata, MIRABILE or the Nuovo Soggettario, custom IRIs are automatically assigned by the tool.
This automatic assignment of the IRIs to the entities and the possibility of choosing an entity in a predefined list allow the scholars to reduce (1) the time for populating the ontology and (2) the possibility of making

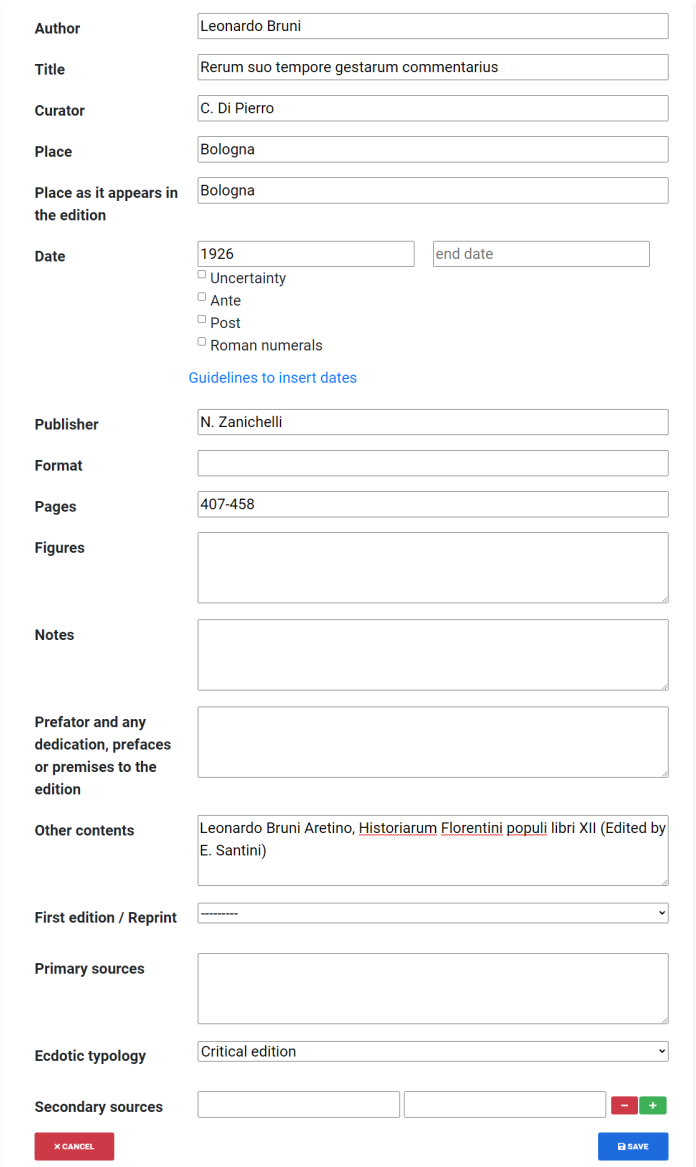

Fig. 3 An example of the printed edition annotation form of the IMAGO Web tool

mistakes while inserting the data manually. The knowledge that is inserted by scholars through the tool is later automatically converted into an OWL graph according to our ontology model.

We are currently conducting a user test to assess and estimate the time saved and the cognitive and technical burden reduced by using the tool. Hopefully, the results will be reported in a subsequent paper.

At the current stage of the project, our KB includes 250 works, 206 authors, 614 libraries, and 313 places. The population of our ontology is far from the conclusion, and at the moment, it is not possible to calculate the total amount of references of works and toponyms to the libraries and places we have already collected in our KB. We have a reference for each library and place, but we do not know how many libraries are related to the same place or how many times a place is cited in the works.

5.1 A First Evaluation of the Ontological Model

We performed a first evaluation of the IMAGO ontology. In particular, we conducted two different types of assessment: an automatic evaluation and an evaluation involving users. At this stage of the project, the ontology population has just started, thus we only evaluated the ontological model, but we plan to also evaluate the IMAGO KB as future work.

For the automatic evaluation of the ontological model, we used the automatic OntoQA system [47]. OntoQA

is a feature-based approach for evaluating ontologies that does not require data training. OntoQA evaluates the ontologies using a predefined set of metrics. It provides a user-friendly application system that automatically computes metrics out of an input ontology. A deep description of the used metrics and the results of the IMAGO ontological model evaluation are reported in [5]. The two main results of this assessment show that the IMAGO ontology has a value of Relationship Richness (RR) equal to 68.75%. RR is a metric that evaluates the diversity of relations in the ontology, equal to 68.75%. This value denotes that the ontology is significantly richer than a simple taxonomy. The Inheritance Richness (IR) has a value of 1.66. IR indicates how the knowledge is grouped into different categories and subcategories in the ontology. This value indicates the vertical nature of the ontology and defines it as a domain-specific ontology. Notably, the IR values of general-purpose ontologies like SWETO [1], or TAP [26] have 4 and 5.36 IR values, respectively.

Regarding the evaluation of the model involving users, the IMAGO ontology is currently being populated by four scholars, two University professors and two PhD students, all experts in Italian and Latin literature and linguistics, with a special focus on the Medieval and Renaissance geographical literature. The tool was used for three weeks by the scholars, and after that, we organised a meeting to gather and discuss their comments about the ontological model. Since this was the first evaluation of the ontology, we chose the think-aloud method [48] to allow scholars to freely express problems, suggestions and requests. The scholars have appreciated the ability of the model to represent in a satisfactory way all concepts and relations reported in the conceptualisation. However, managing the model through the tool, they have realised that some pieces of knowledge are still missing. In particular, all the scholars agreed that the following additional information has to be represented by the model to provide a complete description of the manuscript and the printed edition: (i) the manuscript has to be linked to a Web page containing a description or a digitalised reproduction of the manuscript; (ii) the manuscript has to be linked to a text field (a string) that reports the comments of the scholars involved in the project about the manuscript. In this field, the scholars report notes regarding the discoveries or insights resulting from studying and analysing the original manuscript. The scholars use this field to share their notes with the other scholars involved in the project. At the moment, the notes collected are for scholars' internal use only; (iii) the manuscript has to be linked to a string that reports the information about its iconographic apparatus; (iv) the print edition has to be linked to the name of the publication place as it is reported in it, e.g. *Argentoratum*, the ancient name of Strasbourg.

The requests of the scholars were easily integrated into our ontological model. The ability of the ontology to be expanded in a simple way denotes the strength of the model, thanks to its semantic interoperability. The technical implementation of the new requirements in the ontology is reported in [5].

6 Conclusion and Future Work

This paper discusses how the current scientific Digital Libraries in the Digital Humanities field can create linked information spaces and navigational services, making use of the Semantic Web technologies to formally represent, search and browse knowledge. We have reported a quite extensive overview of the DLs that are developed using the Semantic Web technologies in different fields (e.g. archaeology, literary studies, cultural heritage) of the Digital Humanities domain. We presented as a case study the IMAGO - Index Medii Aevi Geographiae Operum, a three years (2020-2023) Italian National research project that aims at providing a systematic overview of the Medieval and Renaissance geographical literature. Currently, such information 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 KB is based on an ontology expressed in OWL 2 language. To maximize the interoperability of the ontology, it was developed as an extension of two standard ontologies: CIDOC CRM and FRBRoo (and its extension LRMoo). In the presented case study, making use of the Semantic Web technologies, linked information spaces were created that allow users to discover and navigate knowledge across multiple repositories. In particular, IMAGO links five different datasets, i.e. Wikidata, the MIRABILE digital archive, the Nuovo Soggettario thesaurus, the Mapping Manuscript Migration knowledge base, and the Pleiades gazetteer. The linking among different datasets allows to considerably enrich the knowledge collected in the IMAGO KB. To populate the ontology, we developed a web tool that allows automatically assigning IRIs to the resources collected in the KB, starting from the datasets we choose as reference sources. We have also reported the results of a first assessment of our ontological model. We conducted an automatic evaluation using the automatic OntoQA system [47] and an evaluation involving users. As future work, we have planned to carry out a complete assessment that includes not only the model but also the KB of the project. 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). In particular, with the support of the scholars involved in the project, we plan to define predefined SPARQL queries to extract data that can be interesting for scholars, students but also for general users. At the same time, the application will make available a SPARQL endpoint so that any user can make her/his own queries to the knowledge base.

References

1. Aleman-Meza, B., Halaschek, C., Sheth, A., Arpinar, I.B., Sannapareddy, G. (2004). SWETO: Large-Scale Semantic Web Test-bed. In: Proceedings of the 16th International Conference on Software Engineering and Knowledge Engineering (SEKE 2004): Workshop on Ontology in Action, Banff, Canada, June 21-24, pp. 490–493
2. Aliaga, D.G., Bertino, E., Valtolina, S. DECHO: a framework for the digital exploration of cultural heritage objects. J. Comput. Cultural Herit (JOCCH) 3(3), 12 (2011)
3. Baierer, K., Dröge, E., Eckert, K., Goldfarb, D., Iwanowa, J., Morbidoni, C., Ritze, D. DM2E: A Linked Data Source of Digitised Manuscripts for the Digital Humanities. Semantic Web 733 – 745 (2017), DOI: 10.3233/SW-160234
4. Bartalesi, V., Meghini, C., Metilli, D., Tavoni, M. and Andriani, P. A web application for exploring primary sources: The DanteSources case study. Digital Scholarship in the Humanities, 33(4), 705-723 (2018), DOI: https://doi.org/10.1093/llc/fqy002
5. Bartalesi, V., Metilli, D., Pratelli, N., and Pontari, P. (2021). Towards a knowledge base of medieval and renaissance geographical Latin works: The IMAGO ontology. Digital Scholarship in the Humanities. Oxford University Press, DOI: https://doi.org/10.1093/llc/fqab060
6. Bauer, F., and Kaltenböck, M. (2011). Linked open data: The essentials. Edition mono/monochrom, Vienna, 710
7. Bezjak, S., Clyburne-Sherin, A., Conzett, P., Fernandes, P. L., Görögh, E., Helbig, K., ... and Tennant, J. (2018). The open science training handbook.Genève, Switzerland: Zenodo.
8. Bizer, C., Heath, T., Idehen, K., and Berners-Lee, T. (2008, April). Linked data on the web (LDOW2008). In: Proceedings of the 17th international conference on World Wide Web (pp. 1265-1266).
9. Bouloux, N. (1999). Culture et savoirs géographiques dans l'Italie du XIVe siècle (Doctoral dissertation, Paris 1).
10. Bruneau, O., Lasolle, N., Lieber J., Nauer, E., Pavlova, S., Rollet, L. Applying and Developing Semantic Web Technologies for Exploiting a Corpus in History of Science: The Case Study of the Henri Poincaré Correspondence. Semantic Web, 12(2), 359-378. IOS Press (2021), DOI: 10.3233/SW-200400
11. Bruseker, G., and Meghini, C. (2018). Make it Happen–Carrying out Research and Analysing Data (PARTHENOS Webinar)
12. Burrows, T., Emery, D., Fraas, M., Hyvönen, E., Ikkala, E., Koho, M., ... and Wijsman, H. Mapping manuscript migrations knowledge graph: Data for tracing the history and provenance of medieval and renaissance manuscripts. Journal of open humanities data, 6(1), 3 (2020)
13. 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.
14. Damiano, R., and Lieto, A. (2013). Ontological representations of narratives: a case study on stories and actions. In 2013 Workshop on Computational Models of Narrative (pp. 76-93). Schloss Dagstuhl Leibniz-Zentrum für Informatik.
15. Damiano, R., Lieto, A., Lombardo, V.: Ontology-based visualisation of cultural heritage. In: Proceedings of the Eighth International Conference on Complex, Intelligent and Software Intensive Systems (CISIS), 558–563, IEEE (2014)
16. Defilippis, D. (2009). Da Flavio Biondo a Leandro Alberti. Corografia e antiquaria tra Quattro e Cinquecento, Atti del Convegno di Studi (Foggia, 2 febbraio 2006)
17. Doerr, M. (2003). The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. AI magazine, 24(3), 75-75
18. Doerr, M., Bekiari, C., LeBoeuf, P., and nationale de France, B. (2008, September). FRBRoo, a conceptual model for performing arts. In 2008 Annual Conference of CIDOC, Athens (pp. 15-18)
19. Doerr, M., LeBoeuf, P. (2007, February). Modelling intellectual processes: the FRBR-CRM harmonization. In International DELOS Conference (pp. 114-123). Springer, Berlin, Heidelberg
20. Doerr, M., Gradmann, S., Hennicke, S., Isaac, A., Meghini, C., and Van de Sompel, H. (2010, August). The europeana data model (edm). In World Library and Information Congress: 76th IFLA general conference and assembly (Vol. 10, p. 15)
21. Doerr, M., and Theodoridou, M. (2011). CRMdig: A Generic Digital Provenance Model for Scientific Observation. TaPP, 11, 20-21
22. Domingue, J., Fensel, D., and Hendler, J. A. (Eds.). (2011). Handbook of semantic web technologies. Springer Science & Business Media
23. Dunning, A., De Smaele, M., and Böhmer, J. (1970). Are the fair data principles fair?. International Journal of digital curation, 12(2), 177-195
24. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., and Schneider, L. (2002, October). Sweetening ontologies with DOLCE. In International Conference on Knowledge Engineering and Knowledge Management (pp. 166-181). Springer, Berlin, Heidelberg
25. Goerz, G., Oischinger, M., and Schiemann, B. (2008, September). An implementation of the CIDOC conceptual reference model (4.2. 4) in OWL-DL. In: Proceedings of the 2008 Annual Conference of CIDOC-The Digital Curation of Cultural Heritage
26. Guha, R., and McCool, R. (2003). TAP: A semantic web test-bed. Journal of Web Semantics, 1(1), 81-87.
27. Guarino, N., Oberle, D., and Staab, S. (2009). What is an ontology?. In Handbook on ontologies (pp. 1-17). Springer, Berlin, Heidelberg.
28. Hall, M., Agirre, E., Aletras, N., Bergheim, R., Chandrinos, K., Clough, P., ... and Stevenson, M. PATHS–Exploring Digital Cultural Heritage Spaces. In: International Conference on Theory and Practice of Digital Libraries, 500-503, Springer, Berlin, Heidelberg (2012)
29. Hyvönen, E., Ikkala, E., Koho, M., Tuominen, J., Burrows, T., Ransom, L., and Wijsman, H. (2021, October). Mapping Manuscript Migrations on the Semantic Web: A Semantic Portal and Linked Open Data Service for Premodern Manuscript Research. In International Semantic Web Conference (pp. 615-630). Springer, Cham

30. Hyvönen, E., Leskinen, P., Tamper, M., Rantala, H., Ikkala, E., Tuominen, J., and Keravuori, K. BiographySampo–publishing and enriching biographies on the semantic web for digital humanities research. In: European Semantic Web Conference, 574-589, Springer, Cham (2019)
31. Hyvönen, E., Takala, J., Alm, O., Ruotsalo, T., Mäkelä, E. Semantic Kalevala-accessing cultural contents through semantically annotated stories. In: Proceedings of the Cultural Heritageon the Semantic Web Workshop at the 6th International Semantic Web Conference (ISWC 2007), Busan, Korea (2007)
32. Hollander H. S., Meghini C., Scopigno R., Richards J., Wright H., Geser G. ... and Felicetti A. ARIADNE: A Research Infrastructure for Archaeology. ACM Journal on Computing and Cultural Heritage, 10(3) (2017) DOI: https://doi.org/10.1145/3064527
33. Jannach, D., Zanker, M., Felfernig, A., Friedrich, G., Recommender systems: an introduction. Cambridge University Press, Cambridge (2010)
34. Kilfeather, E., McAuley, J., Corns, A., and McHugh, O. (2003, September). An ontological application for archaeological narratives. In proceeedig of 14th International Workshop on Database and Expert Systems Applications, 2003, pp. 110-114). IEEE
35. Koho M., Burrows T., Hyvönen, E., Ikkala, E., Page, K., Ransom, L., ... and Wijsman, H. Harmonizing and publishing heterogeneous premodern manuscript metadata as Linked Open Data. Journal of the Association for Information Science and Technology (JASIST) (2021), DOI: https://doi.org/10.1002/asi.24499
36. Kristeller, P. O. (ed.) (1963). Iter Italicum. Vol. 6: Italy IIIand Alia itinera IV. Leiden, Netherlands: Brill.
37. Maali, F., and Erickson , J. World Wide Web Consortium. Data catalog vocabulary (DCAT). (2014). Available at https://www.w3.org/TR/vocab-dcat/. Accessed 23 September 2021
38. Menestò, E. (1994). Relazioni di viaggi e di ambasciatori in Lo spazio letterario del Medioevo. Edited by G. Cavallo et alii, pp. 535-600. Roma-Salerno (1993)
39. Miller, T. M. (2012). Specify for Archaeology: A Proposed Data Model for Archaeological Collection Database Management (Doctoral dissertation, University of Kansas)
40. Mulholland, P., Collins, T. Using digital narratives to support the collaborative learning and exploration of cultural heritage. In: Proceedings of the 13th International Workshop on Database and Expert Systems Applications, 527–531. IEEE (2002)
41. Mulholland, P., Collins, T., and Zdrahal, Z. (2004). Story fountain: intelligent support for story research and exploration. In: Proceedings of the 9th international conference on Intelligent user interfaces (pp. 62-69).
42. Pontari, P. (2016). "Nedum mille qui effluxerunt annorum gesta sciamus". L'Italia di Biondo e l'invenzione del Medioevo. A NEW SENSE OF THE PAST - THE SCHOLARSHIP OF BIONDO FLAVIO (1392-1463). Edited by Angelo MAZZOCCO and Marc LAUREYS. Leuven University Press (2016)
43. Potthast, A. Repertorium fontium historiae medii aevi. Istituto storico italiano per il medio evo, Rome (1962)
44. Schreiber, A. T., and Raimond, Y. (2014). RDF 1.1 Primer.
45. Simon, R., Isaksen, L., Barker, E. T., and de Soto Canamares, P. The Pleiades gazetteer and the Pelagios project. Bloomington: Indiana University Press (2016)
46. Talbert, R. J. (Ed.). (2000). Barrington Atlas of the Greek and Roman World: Map-by-map Directory (Vol. 1). Princeton University Press.
47. Tartir, S., Arpinar, I. B., Moore, M., Sheth, A. P., Aleman-Meza, B. (2005). In: Proceedings of the IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources, pages 45–53, 2005.
48. Van Someren, M., Barnard, Y. F., and Sandberg, J. (1994). The think aloud method: a practical approach to modelling cognitive. London: AcademicPress.
49. Vrandečić, D., and Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. Communications of the ACM, 57(10), 78-85
50. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... and Mons, B. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1), 1-9 (2016)
51. Wolff, A., Mulholland, P., Collins, T. Storyspace: a story-driven approach for creating museum narratives. In: Proceedings of the 23rd ACM Conference on Hypertext and Social Media, 89–98. ACM (2012)
52. World Wide Web Consortium. (2012). OWL 2 web ontology language document overview.
53. World Wide Web Consortium. (2013). SPARQL 1.1 overview.