Umanistica Digitale - ISSN:2532-8816 - n.9, 2020
S. Barzaghi. M. Palmirani, S. Peroni – Development of an ontology for modelling medieval manuscripts: the case of Progetto IRNERIO
DOI: http://doi.org/10.6092/issn.2532-8816/11187

Development of an ontology for modelling medieval manuscripts: the case of Progetto IRNERIO

[1]Sebastian Barzaghi, [2]Monica Palmirani, [3]Silvio Peroni

[1,2,3]University of Bologna, Italy
[1]sebastian.barzaghi2@unibo.it
[2]monica.palmirani@unibo.it
[3]silvio.peroni@unibo.it

Abstract

The catalogue of digitized medieval texts managed by Progetto IRNERIO contains a rich amount of legal, cultural and historical data that is neither easy to access nor linked to relevant external information. This paper introduces an ontology, called Medieval Manuscripts Ontology (MeMO), to model these texts so as to allow researchers to represent, identify, analyse and retrieve the information within and related to them. MeMO has been designed with a solid methodology in order to be compliant with the requirements of the domain experts (the preservation of the historical narrative, the representation of the context and the description of the collection). Also, it is flexible and richly documented so as to be easily reusable and extensible for further expansions. Thus MeMO allows one to model the resources managed by Progetto IRNERIO with precise semantics so as to avoid information loss and to support a better representation of the conceptual complexities that characterize the collection.

Il catalogo digitale di manoscritti medievali gestito dal Progetto IRNERIO è ricco di materiale di interesse legale, culturale e storico difficilmente accessibile e separato rispetto a contenuti esterni ad esso rilevanti. Questo articolo introduce un'ontologia chiamata Medieval Manuscripts Ontology (MeMO) e sviluppata per modellare questi manoscritti in modo da permettere agli studiosi di rappresentare, identificare, analizzare ed estrarre le informazioni in essi contenute. MeMO è stata sviluppata seguendo una metodologia robusta e agile, in modo da rispettare i requisiti richiesti dall'esperto di dominio (la preservazione della narrazione storica, la rappresentazione del contesto e la descrizione della collezione dei testi). MeMO, inoltre, è flessibile e riccamente documentata in modo da essere facilmente riusabile e estendibile in possibili scenari futuri. Quindi MeMO permette di modellare le risorse del Progetto IRNERIO con termini precisi per evitare una perdita di informazioni e supportare una miglior rappresentazione delle complessità concettuali che caratterizzano la collezione.

Introduction

The Royal College of Spain in Bologna (http://www.realecollegiospagna.it/) hosts the prestigious collection of medieval texts on Roman and Canon Law that has been carried out, since the XII century, by jurists and law scholars hailing from every part of the continent and operating in Bologna, Italy. This work includes important information of legal, cultural and historical nature, which dates to the period of time between the X-XI centuries to the beginning of the XVI century.

Due to the varied fields of study involved in its history, the collection is extremely rich and heterogeneous in terms of nature, provenance and expected use. The subjects range greatly, from legal documentation to philosophical and theological dissertations. Next to illuminated manuscripts there are texts supplemented by marginalia, bare volumes for scholastic use, letters, and other types of ancient writings made on parchment or paper. Many codices also collect a notable quantity of different manuscripts, whose text is summarized and enriched by intricate apparatuses of glosses. These volumes, manuscripts, annotations, illustrations and their descriptions attract jurists, historians, theologians, codicologists and other types of scholars due to their content, their features and the historical context in which they thrive. For example, the real value of a page in a codex often lies within the glosses, the comments made in later times by different scholars, that are clearly separated from the original text. Text, glosses, metadata and other commentaries, layered on the top of each other, form a scientific and historical narrative which has been weaved by multiple different agents through the centuries.

A work of digitization of the collection, led by the publishing house CLUEB (https://clueb.it/) and consisting in the complete scanning of all of the codices from beginning to end, allowed these documents to be available for remote consultation and documentation, thus increasing public accessibility to the collection. Such digital reproductions (more than 138.000) have been published on a digital catalogue managed by *Progetto IRNERIO* (http://irnerio.cirsfid.unibo.it/), a project maintained by the Interdepartmental Centre for Research in History, Philosophy and Sociology of Law and in Computer Science and Law (CIRSFID, http://www.cirsfid.unibo.it/) at the University of Bologna.

The infrastructure of the digital catalogue is based on the TEI P5 ENRICH schema (http://projects.oucs.ox.ac.uk/ENRICH/), which provides a complete, integrated framework to encode, catalogue and describe the manuscripts and their digital images, as well as their descriptive material, text and metadata. However, the way such information is currently rendered lacks the depth of representation that is required to model effectively the peculiarities of this collection. For example, direct answers to queries pointed to the data itself cannot be provided in a clear and straightforward way. Despite the abundance of material and the availability of structured data that accompanies the manuscripts, there are only a few access points to it in the catalogue such as the author's name, the title of the manuscript or the time period related to the codex. In addition, the current structure of the records does not provide links to relevant information found in external resources (e.g. bibliographic citations to critical editions). This presents certain problems for the user who is not familiar with the content of a specific codex, a manuscript, the description or the identity of a certain author responsible for

the realization of a manuscript or a gloss. Moreover, the descriptors employed by TEI ENRICH, albeit useful, are not enough. TEI-based analytic descriptions of the physical and intellectual contents of this kind of resources, as well as digital images obtained from a process of digital scanning, function as digital surrogates of these artefacts to further support scholarly research. Nonetheless, these materials are often assembled together with little regard for existing complementary resources, leaving it to the end-user to make and sustain the connections across collections, that remain fundamentally siloed, with no way to establish permanent semantic connections of their contents [4].

Another fundamental issue in using TEI is the segmentation it operates on the descriptive and scientific narrative of the text. The structure and the content of historical manuscripts – such as those that are part of the catalogue of Progetto IRNERIO – are held together by a scholarly description that is usually tied to multiple factors (e.g. historical context, legal discourse, additional annotation layered through time, etc.) and prone to the overlapping problem [22]. Manuscripts do not exist as isolated entities: they are part of a larger story that builds on their collections, layering of annotations, authors' names variations, and so on [35]. An alternative data model that is capable of handling and expressing such level of complexity is thus needed.

As highlighted in past works, e.g. [1], problems related to data integration, knowledge formalization, information retrieval and mapping can be addressed and solved by Semantic Web technologies. A Web of data would allow humanities researchers to apply these technologies to retrieve adequate answers to multiple different kinds of research topics [2] and to use them as a safeguard to guarantee interoperability, usefulness, openness, dissemination, communication, sharing and integration of the data and metadata of collections on the Web [23].

One of the main aspects that characterises the Semantic Web is the notion of Web ontologies, a way to encode a data model to share knowledge on the Web, which comprises a set of concepts, their definitions and a series of semantic interrelationships between them [37].[1]

There are important benefits from using ontologies. Auer and Herre [3] argue that ontologies capture the semantics of the knowledge in a format that is designed to be easy to maintain and efficient to process by reasoning algorithms. The organization of knowledge and the knowledge itself about the modelled objects are expressed in a clear, meaningful and documented way, for both human and software agents, and the information becomes inferable, reusable and accessible to scientific communities, researchers, companies and general public. Finally, using an ontology entices new means of scholar inquiry, such as to operate searches on implicit information based on automatic reasoning or to link a resource to multiple other resources existing on the Web. Overall, ontological data modelling can be used as a method for organizing discrete facts into a coherent information system where semantic information is structured, managed and made available to a larger portion of target users.

This article presents the *Medieval Manuscripts Ontology* (*MeMO*). MeMO is an OWL 2 DL

1 See [15] for an overview on the history of the term 'ontology' and its usage in both philosophy and computer science.

ontology that has been developed for modelling in a formal way the body of knowledge related to the collection of medieval texts of the Royal College of Spain in Bologna, Italy, which has been digitised within Progetto IRNERIO. The goal was to provide a data model that enables:

- the *preservation of the flow of historical narrative* and the layering of information provided by different agents through the course of time;
- the *representation of the context* of the manuscripts, i.e. the correlations between the material in the collection and the other entities that contribute to shaping its informational context;
- the *description of the structure* of the material, from both a physical and a conceptual point of view, in its particular components and features.

In order to address the needs listed above, the development of MeMO was based on a thorough analysis of the manuscripts catalogue and on existing, well-founded models such as the Functional Requirements for Bibliographic Records (FRBR) [18], the Semantic Publishing and Referencing Ontologies (SPAR) [32] and Ontology Design Patterns (ODP) [12].

The rest of the article is structured as follows. Related works provides a review of some of the most important works about ontology modelling in the domain of manuscript studies. Methodology tracks the workflow of the process through which MeMO has been designed and developed. The Medieval Manuscripts Ontology provides a high-level description of MeMO, while Example of use shows how it can be used within the body of knowledge included in the digital catalogue of the Progetto IRNERIO. Current status describes the current status of the implementation of MeMO, highlighting the main scenarios that have been handled. Conclusions concludes the paper and proposes some further developments.

Related works

In the last two decades, several projects have been carried out in order to develop an ontological model of the world of discourse related to manuscript texts. Being a formalized interpretation of reality, an ontological model may be developed through a few different approaches, each with its own advantages and disadvantages. With regard to the Digital Humanities, a fundamental difference has emerged between two different modelling approaches: a *document-centric* approach and an *event-centric* approach. To date, most of these projects have relied on global, high-level models, such as the CIDOC Conceptual Reference Model (CIDOC-CRM) [8]. Other models, based on a document-centric approach to describe manuscript information, have been reused as well, although to a lesser extent.

The document-centric approach has been employed by Dröge et al. [10] in Digitised Manuscripts to Europeana (DM2E), a project born with the objective to represent metadata in the domain of handwritten manuscripts. DM2E specializes the Europeana Data Model (EDM) [19], a generic representation of the semantics in the cultural heritage domain, for the domain of manuscripts. The main objective of the project is to create a model tailored on the data

providers' needs and that would enable rich semantic mappings of the provided data through the reuse of other ontologies like the DCMI Metadata Terms (DCTerms), the Bibliographic Ontology (BIBO), the FRBR-aligned Bibliographic Ontology (FaBiO) [30] and the Open Archives Initiative Object Reuse and Exchange vocabulary (OAI-ORE).

The majority of the existing ontological schemes for historical manuscript representation, however, embrace an event-centric approach, which focuses on developing fundamental ontological models to provide a more complete semantic representation of the object [17], which is understood as an entity whose ontological status heavily depends on its contextual information. Given their vast descriptive breadth, event-centric ontologies have been also employed to resolve data integration problem, an issue of gaining interoperability among heterogeneous schemata, formats and metadata which has emerged as an effect of unresolved semantic problems and proliferation of different mappings.

Vieira and Ciula [38] developed an ontology (FRH3) to model the information contained in the Fine Rolls of Henry III. The Fine Rolls are historical documents that record monetary transfers done to the king of England by municipal and religious individuals or corporate bodies in exchange of concessions and favours of social, political and economic nature. One of the main goals of the project is to produce richer indices and search mechanisms to help researchers in the retrieval and interpretation of the source material. The approach they use is to extract data from TEI-XML files, each representing a roll which was marked up as to include structural information, temporal information and semantic content, and to plot them into an ontological model constructed by reusing CIDOC-CRM and other ontologies such as DCTerms, WGS84 Geo Positioning (GEO), Simple Knowledge Organization System (SKOS) and the Time ontology (TIME) [16].

Similar work has been pursued by others as well. The objective of the *Sharing Ancient WisdomS* project (SAWS) [20] has been to create a model to capture the knowledge contained inside TEI-annotated manuscripts related to Greek and Arabic ancient wisdom literature. The main objective of SAWS is to represent, identify and analyse this flow of knowledge across these texts, as well as the evolution patterns of these sayings within their cultural contexts, and to facilitate the process of sharing research work and publishing digital editions of the material. The relationships within SAWS manuscripts have been encapsulated within an ontological model which formally defines the vocabulary being used to express the RDF information. In developing the model, the FRBR[2] object-oriented model (FRBRoo) [36] has been reused as the base ontology to express relationships among the texts, the excerpts of the texts and analogous bodies of material. Other ontologies reused by SAWS are CRM Digital (CRMdig) [9], BIBO, SKOS and DCTerms.

Biblissima [14] is one of the first attempts to use Semantic Web technologies for modelling

2 The Functional Requirements for Bibliographic Records standard (FRBR) [18] is a model for describing bibliographic resources proposed by IFLA (https://www.ifla.org/publications/functional-requirements-for-bibliographic-records). FRBRoo is an ontology based on both CIDOC-CRM and FRBR, and expressed in RDF Schema (RDFS).

descriptions of manuscripts. It focuses on the management of the information related to a huge and complex mass of documentation on manuscripts and early printed books, dated from the 8th to 18th centuries. The majority of Biblissima databases contain descriptive and structural metadata for medieval manuscripts, but the project also includes digital editions in TEI-XML format. In order to handle the heterogeneity of formats and data, it uses a mixture of CIDOC-CRM and FRBRoo as a common framework to facilitate internal mapping and allow other people to expose their data in RDF, in compliance to a globally established standard. The ontology also exploits a thesaurus based on technical terms and descriptors that are commonly used for indexing medieval resources. In addition, the project's data on people, corporate bodies, places and titles have been aligned with existing authority lists, such as the Virtual International Authority File (VIAF) and GeoNames (GN). By adopting open standards for both the ontology and the thesaurus, the data might also be aggregated and used in other projects.

Zhitomirsky-Geffet and Prebor [39] provide a review of recent techniques and present a comparison against earlier methods, such as DM2E and FRH3. In their paper they propose to design an ontological model to represent the narrative of historical handwritten Hebrew manuscripts, in order to enable a systematic research of the knowledge embedded into them. The underlying approach they took consisted in treating manuscripts as 'living entities', with a life cycle based on events, and in developing a data model to describe such life cycle accordingly. In order to explicitly provide an adequate semantic data representation of the manuscript and its biography, they use an event-centric ontological model which was built as an extension to existing ontologies such as CIDOC-CRM, FRBRoo, DM2E, SKOS, BIBO, the Citation Typing Ontology (CiTO) [30], DCTerms, the Friend Of A Friend vocabulary (FOAF) [6], the Biographical Information vocabulary (BIO), the EAC-CPF Descriptions Ontology for Linked Archival Data (EAC-CPF), the VIVO Core Ontology (VIVO), the Linking Open Descriptions of Events ontology (LODE) and the SEM Ontology (SEM).

In order to capture the specific semantics needed to represent the resources in the digital catalogue of Progetto IRNERIO without compromising on either precision or practicability (or both), great care must be taken with the development of the model from at least two points of view. On the one hand, the model should be able to deal with all the three complexities of representing manuscripts mentioned in the introduction – i.e. preservation of the flow of historical narrative, representation of the context of the manuscripts, and description of the structure of the material. On the other hand, it should prove to be adequate for *pragmatic* uses within an existing digital catalogue implemented by means of specific technologies.

From this perspective, ontologies that are heavily based on either CIDOC-CRM or other models emanated from it (e.g. FRBRoo), such as FRH3, SAWS, Biblissima, and Zhitomirsky-Geffet and Prebor's model, would not be a good fit for the project. CIDOC-CRM, albeit being an ontology specifically developed for cultural objects, provides only primitives with high-level semantics that inevitably abstract them from any concrete implementation. According to Nussbaumer and Haslhofer [28], this is due to the fact that it has been developed as a global ontology with the objective to provide a generalized framework according to which it is possible to operate decentralized data integration between different metadata schemas, thus

resulting, more often than not, into similar entities and relations being mapped into different ontological chains or, vice versa, different entities and relations being mapped to identical chains. The double bond between high-level semantics and lack of implementation guidelines implies the need to add more information from controlled vocabularies to disambiguate a model that was already tortuous to begin with. This causes CIDOC-CRM to be over-engineered, too difficult to comprehend and to be used successfully with respect to this case study, which instead is related to a precise project with its very specific needs and users – which may not necessarily be practical of the nomenclature and structure imposed by CIDOC-CRM.

By contrast, the Functional Requirements for Bibliographic Records standard (FRBR) [18], a well-known and robust model proposed by the International Federation of Library Association (IFLA) for representing bibliographic resources and metadata, would be a good basis to build on due to its flexibility in representing complex and layered objects. In addition, some of the SPAR Ontologies such as FaBiO and CiTO, by explicitly focusing on documents and their description in FRBR terms, can further expand the possibilities offered by FRBR through the definition of additional bibliographic entities and the relations between them. Finally, certain conceptual issues that might entail significant cognitive effort in their modelling (such as recording changes in values through time) can be easily dealt with by plugging Ontology Design Patterns (ODP)[3] into other models. For example, Time Interval (TI) [13] is an ODP which enables an intuitive description of periods of time. TI, in turn, is reused by other convenient ODPs such as Time-indexed Value in Context (TVC) [34], useful to describe situations in which entities have values during a certain time interval and within a particular context, and Literal Reification [13], which allows modelling certain literals as individuals of a class so that one may use them as proper subjects or objects of RDF statements within an ontology.

Methodology

This section explains in detail the methodology we used to develop MeMO. In Preliminary metadata analysis, the inquiry carried out on the catalogue of Progetto IRNERIO is described to detail how and which metadata have been drawn from it. In Ontology design and development, the reused ontologies and some other additional tools are presented to provide a complete overview on the ontology development process.

Preliminary metadata analysis

The objective of the analysis was to determine an initial set of primitives (classes, attributes and properties), originally encoded into the catalogue records as metadata, which could then be used as a basis to develop the ontology. To this end, the work on the catalogue focused on identifying different types of fields in the catalogue records and their decomposition into

3 ODPs are small, documented and reusable ontologies that can be used as modelling components in ontology design and engineering [12].

atomic ontological units.

The single codex record provides a description of all the available information and metadata from the catalogue about the codex (such as materials, dimensions, etc.), a list of the manuscripts it contains, a bibliography and a list of unnumbered elements such as flyleaves, plates and covers, also belonging to the codex. A codex can be incomplete, with a number of missing folios from it. Neither title nor authorship are provided, so each codex is organized according to a specific identifier made up by either a combination of three numbers, ranging from 000 to 286, or a letter, ranging from A to P in alphabetical order (with the exception of the letters D, E, M, N and O).

The single manuscript record provides access to the sequence of images of the digitized folios, a set of metadata (e.g. *title*, *author*, *editions*, etc.) and to a list of related manuscripts contained in the same codex. In the collection there are more than 800 registered authors, and often, beside their *name*, each author has one *nickname* or even multiple variations of name. More often than not, there is also a misalignment between the identity of an author of a manuscript and the identity of the glossators who commented on such manuscript.

The metadata have been categorized in terms of their complexity with respect to the information they describe, while using the metadata schema of the catalogue as a guiding tool and a yardstick for evaluating the consistency of certain metadata.

Identifier, *materials* and *title* can be converted into ontological elements with relative ease. other metadata cover different types of data (e.g. *description*, *century*). The *description* field contains a broad array of miscellaneous information which can be gathered under several distinct metadata, such as *style of script*, *decorations*, *alternative identifiers*, and so on. The *century* metadata describes temporal information as well as spatial indications of some sort (i.e. a *place*), thus recording two very different things. Other metadata present inconsistent or unclear information (e.g. *author*, *dimension*, *columns*). There are many inconsistencies between the identity of the author of the manuscript and that of the glossators. The *dimensions* metadata describe the size of the codex and (possibly) the size of the folios, but the attribution of size to a codex is problematic in terms of conceptualization, since it technically refers to the size of the binding of that codex[4]. In addition, many metadata express meaningful relationships between the data contained in the catalogue that need further explanation (e.g. *foliation*, *works*, *bibliography*, *edition*, *nickname*, *name*, *citation*, *incipit*, *explicit*, *notes*, and so on).

Ontology design and development

This subsection presents the methodology used for developing the ontology, along with a list of reused ontologies and a series of supporting applications used during the development process.

SAMOD. The Simplified Agile Methodology for Ontology Development (SAMOD) [29] is an agile methodology for developing ontologies that is partially inspired to the Test-Driven Development process in software engineering and other existing agile ontology development

4 This point is actually addressed in the TEI Guidelines. The <dimensions> element has attribute @type, which can have "binding" or "leaf" as a value (among others), but not "codex".

methodologies, such as eXtreme Design (XD) [5]. As shown in Figure 1, SAMOD consists in an iterative process made up by three main phases: 1) the development of a *modelet* formalizing a scenario that belongs to the domain of discourse, and the creation of a *test case* comprising the modelet and a series of additional resources such as glossaries, diagrams and query examples; 2) the merging of the modelet to the current model, developed in the previous iteration (if any); 3) the refactoring of the new current model resulting from the previous step. Before moving on to the next step, each test case must pass a testing phase made up by a model test, a data test and a query test. Each step of the methodology ends with the formal implementation of the ontology in its current state, called 'milestone', accompanied by all the previous test cases, updated accordingly.

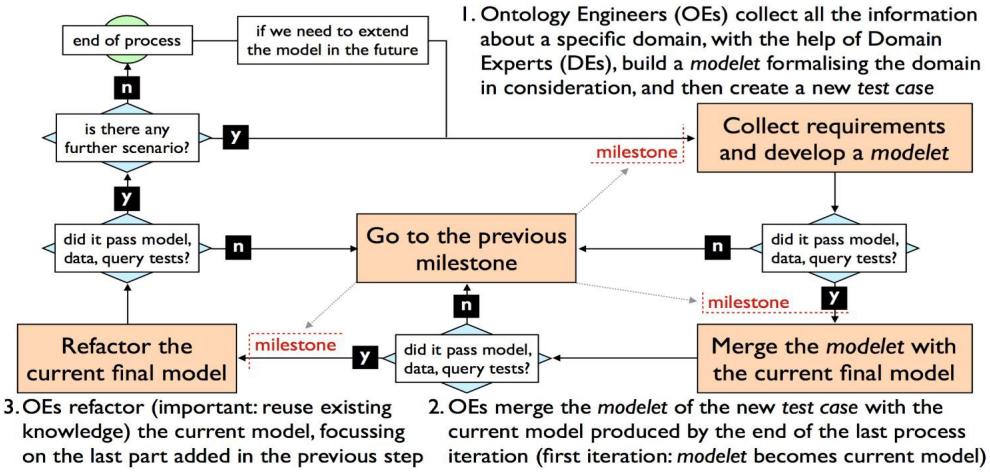

Figure 1: A summary of the three steps of SAMOD.

With respect to the project scope and the domain expert's needs, SAMOD has proven to be extremely beneficial as a methodology for ontology development, since it allows one single person to build up – within a reasonable period of time – a well-developed, documented, reusable ontology by using exemplars of data and testing phases. In this way, the ontology turns out to be both compliant with the domain expert's specific demands and easily embeddable with other models so as to be extendable to a more precise description of the domain through further iterations.

Reused ontologies. In the third step of every iteration, SAMOD suggests importing other models into the ontology whenever possible, in order to maximize the ontology reusability in other contexts. FaBiO – being an ontology focused on entities that are textual in nature, and/or referred to by bibliographic references – has been reused as the basis on which MeMO was built. It defines its own set of entities (that are subclasses of the original FRBR entities) by reusing a well-known RDF vocabulary (FRBRcore) that incorporates the basic concepts and

relations of the FRBR model. CiTO was used to model the numerous references existing between glosses, manuscripts, codices, metadata, and other relevant entities that are part of the collection (e.g. a gloss citing another gloss in a given manuscript). Literal Reification was used to record how certain elements, normally modelled as literal strings, might change over time (e.g. the name of an author), while TVC was used as a basis to model certain metadata related to a manuscript and the way they might change over time on the basis of their textual context and their role within it (metadata referring to the incipit of the manuscript, for example). TI was used in conjunction with Literal Reification and TVC to express the concept of time interval. Finally, DCTerms and FOAF were used for modelling date times, names, titles, and so on.

Supporting tools. A series of supporting applications have been used during the development process:

- the Live OWL Documentation Environment (LODE) [33] is an open source service that automatically extracts classes, object properties, data properties, named individuals, annotation properties, general axioms and namespace declarations from an OWL ontology, and renders them as ordered lists, together with their textual definitions, in a human-readable HTML page designed for browsing and navigation;
- the Graphical Framework for OWL Ontologies (Graffoo) [11] is an open source tool that can be used to present the classes, object properties, data properties, individuals, general axioms, namespace declarations and restrictions within OWL ontologies as user-friendly diagrams;
- Protégé [27] is an open-source ontology editor developed at Stanford University which provides a graphic user interface, deductive classifiers and OWL 2 DL reasoning engines (e.g. HermiT and Pellet) to validate the consistency of an ontology and to infer new knowledge from it;
- Apache Jena Fuseki is a SPARQL 1.1 server with a web interface, backed by the Apache Jena TBD RDF triple store, which provides the SPARQL 1.1 protocols for query and update as well as the SPARQL Graph Store protocol.

During the development of MeMO, both LODE and Graffoo proved to be stable tools with a good level of usability. LODE has been used to produce the HTML documentation of the ontology by extracting the labels, comments and provenance information of the ontology elements. Instead, Graffoo has been used to create the diagrams of the various modelets, the refactored models and the full ontology. Protégé, arguably the most widely used open source software for building and maintaining ontologies [24], has been used multiple times in each iteration for testing the consistency of the modelet and the dataset, the merged model, and the refactored model. Although it is not the best solution in terms of query execution time [25], Fuseki can be quite useful for easy, quick testing [26]. In particular, it has been used in each iteration of SAMOD as a query engine for testing the correctness of the formal competency questions and for addressing the particular requirements they expressed.

The Medieval Manuscripts Ontology

The Medieval Manuscripts Ontology (MeMO) is an OWL 2 DL ontology that aims at providing a framework for the formal description of the digital catalogue of medieval texts managed by Progetto IRNERIO. MeMO has been made available with a Creative Commons Attribution License 4.0.

The Graffoo diagram in Figure 2 visualizes the current version of MeMO. The prefixes and relative base URIs of the models which MeMO reuses are listed in Table 1.

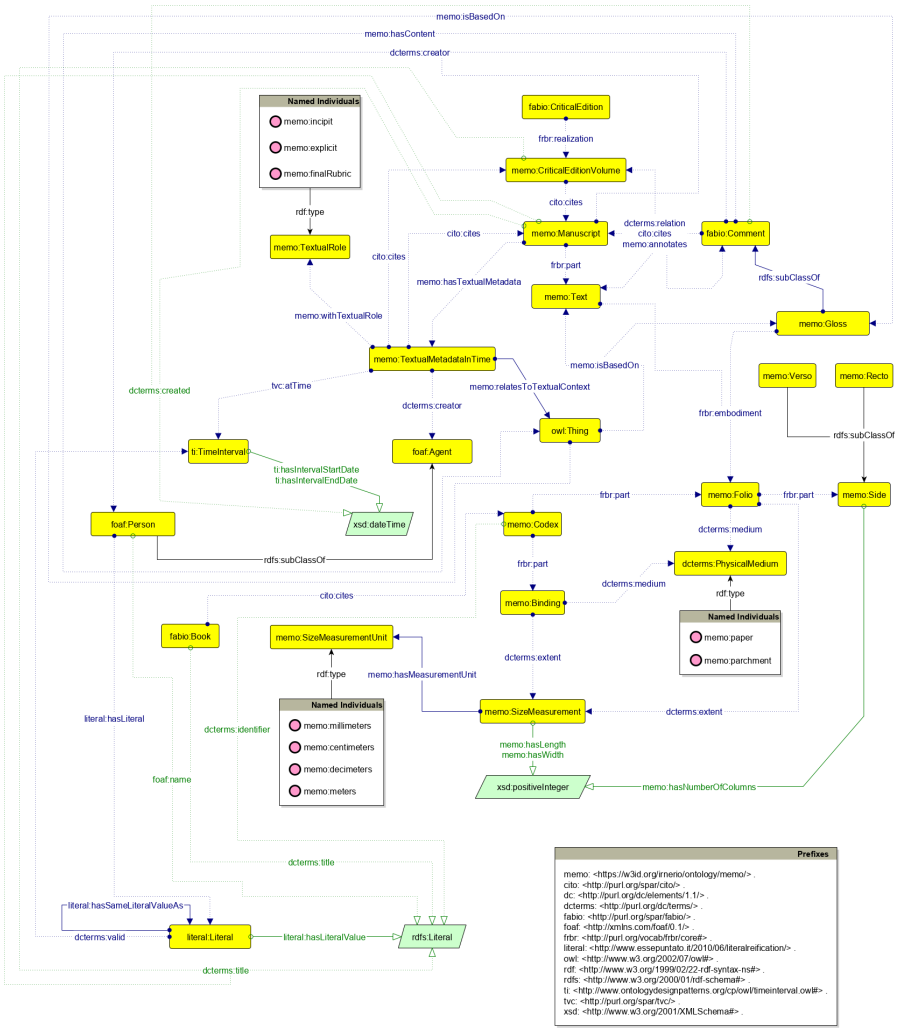

Figure 2: The Graffoo diagram of MeMO.

The resources described by the catalogue of Progetto IRNERIO are artefacts characterized by multiple levels of complexity: non-immediate compositionality, references to both internal and external resources, variability of authorship identity, historical layered commentaries, and so on. In order to adequately represent such complexity, MeMO has been designed around FRBR, which describes bibliographic resources from four different conceptual points of view that are interliked with each other and are defined by the categories of *Work*, *Expression*, *Manifestation*, and *Item*.[5]

FRBR allows one to have a holistic perspective about the resource, on multiple levels of conceptualization, by breaking down the semantic and conceptual ambiguities related to objects into different but related and layered concepts, and by allowing the description of an artefact and its relations with other entities to be more expressive, precise and dynamic. Nonetheless, FRBR has some limitations. Even though the definitions of its concepts are quite straightforward, FRBR is not easily understood by the common user, who finds the terms *Work*, *Expression*, *Manifestation* and *Item* quite ambiguous. According to FRBR, any object (such as a manuscript, for example) has to be described by taking into consideration all the four levels at the same time, in order to have a complete view over it. This multi-layered conceptualization is difficult for an average user to comprehend, because it is much more intuitive to expect the concept of that object to exist at a single FRBR level. In order to avoid this issue without giving up the expressiveness of FRBR, a good solution is to place that object at the level that is deemed more appropriate in relation to the scenario taken into consideration.

Prefix	**Base URI**
memo	https://w3id.org/irnerio/ontology/memo/
cito	http://purl.org/spar/cito/
dc	http://purl.org/dc/elements/1.1/
dcterms	http://purl.org/dc/terms/
fabio	http://purl.org/spar/fabio/
foaf	http://xmlns.com/foaf/0.1/

5 A FRBR Work is the high-level description of the essence of a particular resource, which does not depend on any concrete representation. It is realized through one or more Expressions. A FRBR Expression is the form taken by a Work when it is realized in terms of content. It is the realization of one and only one Work and is embodied in one or more Manifestations. A FRBR Manifestation is a particular embodiment in the physical world of an Expression, according to a specific format. It embodies one or more Expressions and is exemplified by one or more Items. A FRBR Item is the single, tangible, and located exemplar of a certain Manifestation. It exemplifies one and only one Manifestation.

Prefix	Base URI
frbr	http://purl.org/vocab/frbr/core#
literal	http://www.essepuntato.it/2010/06/literalreification/
owl	http://www.w3.org/2002/07/owl#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
ti	http://www.ontologydesignpatterns.org/cp/owl/timeinterval.owl#
tvc	http://purl.org/spar/tvc/
xsd	http://www.w3.org/2001/XMLSchema#

Table 1: The prefixes and the corresponding base URIs of the models reused in MeMO.

For example, this approach has been used systematically for developing FaBiO, which is the main reason why we decided to use it as a foundation for MeMO. In order to embrace this approach and be consistent with the scenarios that are present in the case study, MeMO reuses FaBiO interpretations of FRBR entities as super classes of its entities. In particular: *manuscript*, *text* and *gloss* are conceptualized as subclasses of `fabio:Expression` (since they carry a content and are not inherently related to a precise format), while *codex* and *folio* are conceptualized as subclasses of `fabio:Manifestation`, since they carry a format that functions as a container for both text and glosses, and thus for manuscripts as well.

A *manuscript* (represented by the class `memo:Manuscript`) is a handwritten composition made up by one *text* (the primary text of a manuscript, represented by the class `memo:Text`) and zero or more *glosses* (marginal annotations, represented by the class `memo:Gloss`). A *codex* (represented by the class `memo:Codex`) is a set of one or more folios, in which one or more manuscripts are embodied. The MeMO understanding of the term 'codex' refers to a physical object that should not be confused with '*the* Codex' (i.e. the *Codex Justinianeus*), nor with the concept of manuscript (classically understood as a physical document written by hand by someone). Instead, a codex is a composite entity, without a clear authorship and title, that embodies manuscripts. A *folio* (a leaf or sheet of paper or another material, represented by the class `memo:Folio`) is a physical object that is usually – but not necessarily – part of a codex. It is made up of two sides, one called *recto* (the front side of a folio) and the other called *verso* (the back side of a folio). It carries a specific format and layout. According to how MeMO has

been designed, text and glosses of a manuscript attain a physical realization in the moment they are embodied in the physical format carried by a folio. In this way, it is possible to represent many different scenarios that are effectively present in the collection of Progetto IRNERIO (e.g. multiples manuscript in the same folio, non-sequential foliation, bilocation of the same manuscript in two codices, etc.) without incurring in logical inconsistencies.

Example of use

As aforementioned in the introduction, MeMO has been developed first and foremost with the objective of addressing the requirements of the domain experts – namely jurists, historians, theologians, codicologists, and other scholars – interested in the collection. In this subsection we illustrate one of the possible applications of MeMO within the digital catalogue of the Progetto IRNERIO.

In particular, glosses and citations play a fundamental role in the description and study of the resources of Progetto IRNERIO.[6] Explanatory glosses of various kinds occur frequently in manuscripts, especially in those with biblical or legal content, as sophisticated tools for studying, recovering, restoring and producing new knowledge on the basis of existing work, like the medieval glossators from the juridical school of Bologna did with the mass of Justinian legislation [21]. Glossators from Bologna managed to construct an increasingly large apparatus of glosses which layered up on each other in the form of observations, references, links and comments. Through some of these glosses, they effectively created fully-fledged authoritative standards of reference for avoiding contradictions between statutes and for determining the constitutionality of certain legal stipulations. Historians of law might be interested in the glosses and other forms of commentary that take place in the margins of the page, because the explanatory comments and annotations that medieval manuscripts hoarded over the centuries can present the scholars with a living record of use, study and reference of that resource. A medievalist's research might be centered around the reconstruction of the network of references between a manuscript text and a series of glosses. A modern historian might focus on the way glosses and other resources (such as textual metadata) comment on other glosses, thus spreading the volume of commentaries beyond the collection itself. A contemporary historian might be interested in retracing how a certain manuscript has been cited by critical editions over the course of time.

Overall, MeMO allows to easily represent such articulated data and to compute the level of annotation of each gloss, changing dynamically the perspective on a certain source and inferring new information. In order to demonstrate this, an example of use of MeMO, related to glosses and citations, is discussed below. The example is based on test data, encoded in

6 In our context, a citation is defined as 'a conceptual directional link from a citing entity to a cited entity, for the purpose of acknowledging or ascribing credit for the contribution made by the author(s) of the cited entity. The citing and cited entities may be scholarly publications, online documents, blog posts, datasets, or any other authored entities capable of giving or receiving citations.' [31].

Turtle, a RDF serialization that allows one to write RDF statements in a form that is intuitive and compact, with the possibility of concatenating prefixes with base URIs in order to abbreviate them into prefixed names. All the entities that refer to the test data have a specific base URI (https://w3id.org/irnerio/data/memo/) abbreviated through the prefix `ex`. These Turtle sources are publicly available on the GitHub repository https://github.com/irnerio-opendata/memo, in the `data` directory.[7] In particular, the following scenario is expressed in Turtle as shown in Text 1.

A manuscript (`ex:manuscript_1`) is made up by a text (`ex:text_1`) and six glosses (`ex:gloss_a`, `ex:gloss_b`, `ex:gloss_c`, `ex:gloss_d`, `ex:gloss_e`, `ex:gloss_f`). These entities are related to each other through a series of relationships, which build up a complex record of interpretations, commentaries and references. In fact, glosses made by different people in different time periods layer on the top of each other, either citing or annotating or referring to the text, or another gloss, or a whole manuscript. These relationships are indicated by a series of properties.[8] The property `dcterms:relation` represents the most general way in which an entity can address another entity. Its subproperty `memo:annotates` indicates that some entity provides a critical or explanatory observation to another entity. The property `cito:cites`, defined in CiTO and aligned as a subclass of `dcterms:relation`, expresses the fact that a citing entity references a cited entity. In this way, it becomes possible to infer the annotation level of a gloss and how it varies in relation to other entities.

A scholar might be interested in the glosses that are part of the manuscript and cite its text, or in the second-level glosses that are part of the manuscript and cite or annotate its text, or that annotate or refer to another gloss, such as `ex:gloss_a`.

7 For the prefixes used to abbreviate the base URIs of the ontologies, please refer to Table 1.

8 When developing MeMO, domain and range constraints of these properties have not been defined so as to allow an easy integration with other models and augment the model extensibility in addition to what has already been defined, thus allowing the further addition of other types of properties in the future.

```
ex:manuscript_1 a memo:Manuscript ;
       frbr:part ex:text_1 , ex:gloss_a , ex:gloss_b , ex:gloss_c ,
       ex:gloss_d     , ex:gloss_e , ex:gloss_f .
ex:gloss_a a memo:Gloss ;
       cito:cites ex:text_1 .
ex:gloss_b a memo:Gloss ;
       dcterms:relation ex:text_1 .
ex:gloss_c a memo:Gloss ;
       memo:annotates ex:gloss_a .
ex:gloss_d a memo:Gloss ;
       dcterms:relation ex:gloss_a .
ex:gloss_e a memo:Gloss ;
       memo:annotates ex:gloss_b , ex:gloss_c .
ex:gloss_f a memo:Gloss ;
       cito:cites ex:text_1 ;
dcterms:relation ex:gloss_d ;
       memo:annotates ex:gloss_e .
ex:text_1 a memo:Text .
```

Text 1: Turtle code representing the relations between a manuscript, its text and a set of glosses that are part of it.

Due to the way MeMO has been modelled, it becomes quite easy to navigate this network in a way that allows the desired glosses to be returned according to a level of annotation that shifts dynamically with the scholar's perspective on the source. The level of annotation can be computed according to different types of relationship between a `memo:Gloss` and another entity (e.g. `memo:Manuscript`, `memo:Text`, `memo:Gloss`) by counting the property edges of one or more types of relationship. This is addressed in SPARQL queries through the use of *property paths*. Property paths are a way of expressing chains of properties (forward and backward) without the need to bind all the individual resources along the way, which is especially important if a variable number of edges are to be allowed. The path language for SPARQL property paths is described in Section 9.1 'Property Path Syntax' of the SPARQL 1.1 Query Language Recommendation.

Text 2 shows an example of a simple SPARQL query that exploits property paths to navigate the network of references between glosses. The query returns all the second-level glosses that are part of `ex:manuscript_1` and annotate or cite `ex:text_1`.

```
SELECT ?gloss
WHERE {
        ?gloss a memo:Gloss ;
        frbr:partOf ex:manuscript_1 ;
        (memo:annotates|cito:cites) / (memo:annotates|cito:cites) ex:text_1
}
```

Text 2: A SPARQL query which returns all the second level glosses that are part of a specific manuscript and annotate or cite its text.

The two occurrences of `(memo:annotates|cito:cites)` represent two elements of the property path (each counting as one level of annotation) that are sequenced one after the other, from left to right, through the use of a slash (`/`), which in the property path syntax serves as a sequencing operator. Each path element consists of a combination of two object properties `memo:annotates` and `cito:cites` enclosed together in brackets and separated from each other by a vertical bar (`|`), which in the property path syntax indicates an alternative path of one property or the other. The sequencing of grouped alternative properties allows one to test the existence of the desired path between any gloss and the text `ex:text_1` by trying all the alternative paths. As a result, all the glosses which exist at the other end of such paths are returned (i.e. `ex:gloss_c`).

Current status

The development of MeMO has been carried out in six iterations of SAMOD, each resulting in a model that is responsible for the description of a specific aspect of the domain taken into consideration.

First iteration: *Glosses*. The model which was developed by the end of this iteration enables the description of the gloss apparatus of a manuscript, made of glosses and their relationships with other entities such as manuscripts, manuscript texts, and other glosses. In particular, a `memo:Manuscript` is a handwritten composition that is made up by exactly one `memo:Text` and zero or more instances of `memo:Gloss`. The class `memo:Text` represents the primary textual content that is part of a manuscript. Both `memo:Text` and `memo:Manuscript` are subclasses of `fabio:Expression`. The class `memo:Gloss` is an annotation that comments on the manuscript text and accompanies it (as in the margin of the page or between the lines of the text). `memo:Gloss` has been modelled as a subclass of `fabio:Comment`. The object property `frbr:part` has been used to describe the part-whole relationship existing between `memo:Manuscript` and its parts. The object property `memo:hasContent` serves as a pointer to anything that represents the content of another entity. The property `dcterms:relation` represents the most general way in which an entity can address another entity. One of its subproperties, `memo:annotates`, indicates that some entity makes a critical or explanatory observation for another entity, while the other (`cito:cites`) represents the fact of an entity citing another entity. It is defined in CiTO and, in the process of being imported to MeMO, it

has been aligned as a subproperty of `dcterms:relation`.

Second iteration: ***Textual metadata***. The model developed by the end of this iteration, which is based on TVC, allows the description of textual metadata associated to a manuscript (e.g. metadata referring to the incipit, explicit, or final rubric of a manuscript). In particular, `memo:TextualMetadataInTime` is a subclass of `tvc:ValueInTime` and is characterized by a series of properties. It is related to a manuscript through the object property `memo:hasTextualMetadata`, a sub property of `tvc:hasValue`. `memo:withTextualRole`, a subproperty of `tvc:withValue`, relates a textual metadata in time with its respective textual role, modelled as a class (`memo:TextualRole`) that can assume the following controlled values: `memo:incipit`, `memo:explicit`, `memo:finalRubric`. The object property `tvc:atTime` specifies the particular time interval that has been associated with the textual metadata by linking it to the `ti:TimeInterval` class, which has two data properties that set its start and end dates (`ti:hasIntervalStartDate` and `ti:hasIntervalEndDate`, respectively). The object property `dcterms:creator` is used to link `memo:TextualMetadataInTime` with the agent (`foaf:Agent`) that created it. The object property `memo:relatesToTextualContext` relates a `memo:TextualMetadataInTime` to a textual context that is, in turn, related to the classes `memo:Text` and `memo:Gloss` through the object property `memo:isBasedOn`.

Third iteration: ***Citations***. The model developed by the end of this iteration enables the description of the structure of citations that exists between manuscripts, glosses, textual metadata, editions and other similar resources. A codex is a set of at least one folio and is modelled through the class `memo:Codex`, while folios belong to the class `memo:Folio`. Both are subclasses of `fabio:Manifestation`. The object property `frbr:part` links the classes `memo:Codex` and `memo:Folio` with each other. A codex and a manuscript are related with each other through the idea of the manuscript text and glosses being distributed through the folios that are part of that codex. This situation is modelled in MeMO by using the object property `frbr:embodiment` to relate the classes `memo:Text` and `memo:Gloss` with `memo:Folio`. The model reuses FaBiO for describing certain entities that are external to the catalogue, such as `fabio:Book` and `fabio:CriticalEdition`. Since it deals with citations, it also leans heavily on the model that has been developed in the first iteration, by reusing CiTO for representing the citation network existing between the various entities (with the object property `cito:cites`) along with other relationships, such as `dcterms:relation` and `memo:annotates`. In addition, critical editions have been conceptually separated from their realizations in two specific and distinct layers, characterized by two classes: `fabio:CriticalEdition` and `memo:CriticalEditionVolume`. The `fabio:CriticalEdition` class is a subclass of `fabio:Work` and describes the edition essence, independently from the revisions that can characterize it in time. The `memo:CriticalEditionVolume` class is a subclass of `fabio:Expression` and is used for pointing to a specific realization of a critical edition. The data property `dcterms:title` has been used to model the titles of entities such as `fabio:Book` and `memo:Manuscript`.

Fourth iteration: ***Names***. The model developed by the end of this iteration allows one to

describe of the variations of people's names in time. It is based on the Literal Reification pattern in combination with OWL 2 punning and defines a `literal:Literal` individual that also belongs to the property `foaf:name` which has been meta-modelled as a class. Each literal individual is then assigned a time interval via the `dcterms:valid` property, whose range has been appropriately adapted to accommodate the class `ti:TimeInterval`, so as to represent the period in which the name is valid. The properties `dcterms:creator` and `dcterms:created` have been used respectively to relate the resources to the person who created them (represented with `foaf:Person`) and the dates in which they have been created by that person (represented with `xsd:dateTime`).

Fifth iteration: ***Foliation***. The model developed by the end of this iteration enables the representation of the arrangement of the folios which make up a codex and contain its manuscripts. In particular, the main structure of the codex, with each part related to the other through the `frbr:part` property, is modelled. One or more instances of `memo:Folio` are part of `memo:Codex`, as already anticipated in the third iteration. The classes `memo:Recto` and `memo:Verso`, both subclasses of the class `memo:Side`, are part of `memo:Folio`.

Sixth iteration: ***Codex metadata***. The model developed by the end of this iteration allows describing some features of a codex (e.g. size, materials, number of columns, etc.). The classes `memo:Folio` and `memo:Binding` (which represents the binding of the codex) have been associated with the class `dcterms:PhysicalMedium`, which represents the materials of a folio or a binding, via the property `dcterms:medium`. The named individuals `memo:paper` and `memo:parchment` are the possible controlled values that `dcterms:PhysicalMedium` can take. `memo:Folio` and `memo:Binding` are also related to the class `memo:SizeMeasurement`, subclass of `dcterms:SizeOrDuration`, via the property `dcterms:extent`. In order to express a size in terms of length, width and with millimeters as unit of measurement, `memo:SizeMeasurement` has two data properties called `memo:hasLength` and `memo:hasWidth` and an object property `memo:hasSizeMeasurementUnit` that allows it to be associated with the class `memo:SizeMeasurementUnit`, a class used to express the concept of measurement unit that can assume a series of values, according to the unit used, as aptly named individuals. When it comes to Progetto IRNERIO, all measures are implicitly expressed in millimeters, so the named individual `memo:millimeters` is included in the model. For completeness, other plausible measures have been included in the model through the named individuals `memo:centimeters`, `memo:decimeters` and `memo:meters`. In addition, the information related to the number of columns, expressed via the data property `memo:hasNumberOfColumns`, is not related directly to a codex; since many types of variations are possible within the same codex (between ranges of folios, between single folios or even within the same folio), it has been put in relation with the class `memo:Side` instead. The identifier or identifiers of a codex are expressed via the data property `dcterms:identifier`.

The MeMO GitHub repository contains all the source files of the elements which form the documentation of the ontology. The `development` directory contains a folder for each iteration, thus constituting a full test case with a motivating scenario, a list of informal

Competency Questions, a glossary of terms, a Graffoo diagram of the model in `.png` format (along with its `.graphml` file), a list of formal Competency Questions written in SPARQL, a modelet and a dataset (both written in the Turtle RDF serialization).

The `data` directory contains a set of refactored datasets, one for each iteration, written in the Turtle RDF serialization. The `diagrams` directory contains a set of Graffoo diagrams, each representing the refactored model of its respective iteration. The `docs` directory contains all the ontology files and its versions in time. The `sparql` directory contains a set of refactored formal Competency Questions.

Conclusions

This article introduced MeMO, an ontology for modelling the collection of medieval texts and their catalogue information published and managed by Progetto IRNERIO. MeMO was designed on the basis of a metadata analysis conducted on the catalogue and was developed by using SAMOD, a data-centric and pattern-based methodology for ontology development. This approach allowed to create a fully documented, extendible and dynamic ontological model that allows to faithfully represent the material in the digital catalogue without excessive conceptual clutter and with an eye towards potential expansions to cover more information related to the manuscript studies domain. The methodology proved to be efficient and effective for the task. The metadata analysis allowed to have a general overview on the material and provided the researcher with a solid basis on which the ontological model has been built by using SAMOD. The complexities of the domain and of the collection proved that MeMO is sufficient to meet the domain expert's requirements. Still, further work is needed to semantically refine and expand the model. For example, codex metadata such as *Century*, *Description*, and those related to the style of script, the state of condition, the decoration, and the ruling, need to be integrated into MeMO in order to take into consideration additional useful information. In addition, a web service could be implemented to allow users to query and visualize the data modelled by the ontology.

References

[1] Aljalbout, Sahar, and Gilles Falquet. 2018. 'A Semantic Model for Historical Manuscripts', 12. https://arxiv.org/abs/1802.00295.

[2] Antezana, Erick, Martin Kuiper, and Vladimir Mironov. 2009. 'Biological Knowledge Management: The Emerging Role of the Semantic Web Technologies'. *Briefings in Bioinformatics* 10 (4): 392–407. https://doi.org/10.1093/bib/bbp024.

[3] Auer, Sören, and Heinrich Herre. 2007. 'RapidOWL – An Agile Knowledge Engineering Methodology'. In *Perspectives of Systems Informatics*, edited by Irina Virbitskaite and Andrei Voronkov, 424–30. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-70881-0_36.

[4] Barbera, Michele, Michele Nucci, Daniel Hahn, and Christian Morbidoni. 2008. 'A Semantic Web Powered Distributed Digital Library System'. In *Sustainability in the Age of Web 2.0 - Proceedings of the 12th International Conference on Electronic Publishing*, 130–39. https://elpub.architexturez.net/doc/oai-elpub-id-130-elpub2008.

[5] Blomqvist, Eva, Valentina Presutti, Enrico Daga, and Aldo Gangemi. 2010. 'Experimenting with EXtreme Design'. In *Knowledge Engineering and Management by the Masses*, edited by Philipp Cimiano and H. Sofia Pinto, 120–34. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-16438-5_9.

[6] Brickley, Dan, and Libby Miller. 2014. 'FOAF vocabulary specification 0.91'. Namespace Document 14 January 2014 – Paddington Edition. http://xmlns.com/foaf/spec/ (last visited 15 June 2020).

[7] Brickley, Dan, V. Guha Ramanathan, and Brian McBride. 2014. 'RDF Schema 1.1'. 2014. https://www.w3.org/TR/rdf-schema/.

[8] Doerr, Martin. 2003. 'The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata'. *AI Magazine* 24 (3): 75–75. https://doi.org/10.1609/aimag.v24i3.1720.

[9] Doerr, Martin, and Maria Theodoridou. 2011. 'CRMdig: A Generic Digital Provenance Model for Scientific Observation'. In *Proceedings of the 3rd Workshop on the Theory and Practice of Provenance*, edited by Peter Buneman and Juliana Freire. https://static.usenix.org/events/tapp11/tech/final_files/Doerr.pdf.

[10] Dröge, Evelyn, Julia Iwanowa, and Steffen Hennicke. 2014. 'A Specialization of the Europeana Data Model for the Representation of Manuscripts: The DM2E Model.' *Libraries in the Digital Age (LIDA) Proceedings*. http://ozk.unizd.hr/proceedings/index.php/lida/article/download/117/120.

[11] Falco, Riccardo, Aldo Gangemi, Silvio Peroni, David Shotton, and Fabio Vitali. 2014. 'Modelling OWL Ontologies with Graffoo'. In *The Semantic Web: ESWC 2014 Satellite Events*, edited by Valentina Presutti, Eva Blomqvist, Raphael Troncy, Harald Sack, Ioannis Papadakis, and Anna Tordai, 320–25. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-11955-7_42.

[12] Gangemi, Aldo, and Valentina Presutti. 2009. 'Ontology Design Patterns'. In *Handbook on Ontologies*, edited by Steffen Staab and Rudi Studer, 221–43. International Handbooks on Information Systems. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-92673-3_10.

[13] Gangemi, Aldo, Silvio Peroni, and Fabio Vitali. 2010. 'Literal reification'. In *Proceedings of the 2nd International Workshop on Ontology Patterns - WOP2010*,

edited by Eva Blomqvist, Vinay K. Chaudhri, Oscar Corcho, Valentina Presutti, and Kurt Sandkuhl, 65–6. http://ceur-ws.org/Vol-671/pat04.pdf.

[14] Gehrke, Stefanie, Eduard Frunzeanu, Pauline Charbonnier, and Marie Muffat. 2015. 'Biblissima's Prototype on Medieval Manuscript Illuminations and Their Context'. In *Proceedings of the First International Workshop Semantic Web for Scientific Heritage at the 12th ESWC 2015 Conference*, edited by Arnaud Zucker, Isabelle Draelants, Catherine Faron Zucker, and Alexandre Monnin, 43–8. http://ceur-ws.org/Vol-1364/paper5.pdf.

[15] Guarino, Nicola, Daniel Oberle, and Steffen Staab. 2009. 'What Is an Ontology?' In *Handbook on Ontologies*, edited by Steffen Staab and Rudi Studer, 1–17. International Handbooks on Information Systems. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-92673-3_0.

[16] Hobbs, Jerry R., and Feng Pan. 2004. 'An Ontology of Time for the Semantic Web'. *ACM Transactions on Asian Language Information Processing* 3 (1): 66–85. https://doi.org/10.1145/1017068.1017073.

[17] Hyvönen, Eero. 2020. 'Using the Semantic Web in Digital Humanities: Shift from Data Publishing to Data-Analysis and Serendipitous Knowledge Discovery'. *Semantic Web* 11 (1): 187–93. https://doi.org/10.3233/SW-190386.

[18] IFLA Study Group. 2008. 'Functional Requirements for Bibliographic Records'. Final Report 19. IFLA Series on Bibliographic Control. Munich: International Federation of Library Associations and Institutions. http://www.ifla.org/VII/s13/frbr (last visited 15 June 2020).

[19] Isaac, Antoine. 2013. 'Europeana Data Model Primer.' https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf (last visited 15 June 2020).

[20] Jordanous, Anna, K. Faith Lawrence, Mark Hedges, and Charlotte Tupman. 2012. 'Exploring Manuscripts: Sharing Ancient Wisdoms Across the Semantic Web'. In *Proceedings of the 2Nd International Conference on Web Intelligence, Mining and Semantics*, 44:1–44:12. WIMS '12. New York, NY, USA: ACM. https://doi.org/10.1145/2254129.2254184.

[21] Loschiavo, Luca. 2011. 'La riscoperta dell'Authenticum e la prima esegesi dei glossatori'. In *Novellae Constitutiones. L'ultima legislazione di Giustiniano tra oriente e occidente da Triboniano a Savigny*, edited by Luca Loschiavo, Giovanna Mancini, and Cristina Vano, 111–39. http://hdl.handle.net/11575/3258.

[22] Mah, Carole, Julia Flanders, and John Lavagnino. 1997. 'Some Problems of TEI Markup and Early Printed Books'. *Computers and the Humanities* 31 (1): 31–46. https://doi.org/10.1023/A:1000464519769.

[23] Meroño-Peñuela, Albert. 2013. 'Semantic Web for the Humanities'. In *The*

Semantic Web: Semantics and Big Data, edited by Philipp Cimiano, Oscar Corcho, Valentina Presutti, Laura Hollink, and Sebastian Rudolph, 645–49. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-38288-8_44.

[24] Musen, Mark A. 2015. 'The Protégé Project: A Look Back and a Look Forward'. *AI Matters* 1 (4): 4–12. https://doi.org/10.1145/2757001.2757003.

[25] Nikolić, Nikola, Goran Savić, Milan Segedinac, Stevan Gostojić, and Zora Konjović. 2015. 'RDF Stores Performance Test on Servers with Average Specification'. In *ICIST 2015 Proceedings*, 67–72. Society for Information Systems and Computer Networks. http://www.eventiotic.com/eventiotic/library/paper/94.

[26] Nouwt, B. 2017. 'Tight Integration of Web APIs with Semantic Web'. CEUR-WS. http://resolver.tudelft.nl/uuid:47a68a19-8f33-45be-8163-191ae72f74de.

[27] Noy, N.F., M. Sintek, S. Decker, M. Crubezy, R.W. Fergerson, and M.A. Musen. 2001. 'Creating Semantic Web Contents with Protege-2000'. *IEEE Intelligent Systems* 16 (2): 60–71. https://doi.org/10.1109/5254.920601.

[28] Nussbaumer, Philipp, and Bernhard Haslhofer. 2007. 'Putting the CIDOC CRM into Practice - Experiences and Challenges'. Technical Report. University of Vienna. September 2007. https://eprints.cs.univie.ac.at/404/.

[29] Peroni, Silvio. 2016. 'A Simplified Agile Methodology for Ontology Development'. In *OWL: Experiences and Directions – Reasoner Evaluation - 13th International Workshop, OWLED 2016, and 5th International Workshop, ORE 2016, Bologna, Italy, November 20, 2016, Revised Selected Papers*, edited by Mauro Dragoni, Maria Poveda-Villalón, and Ernesto Jimenez-Ruiz, 55–69. Lecture Notes in Computer Science. Cham , Switzerland: Springer. https://doi.org/10.1007/978-3-319-54627-8_5.

[30] Peroni, Silvio, and David Shotton. 2012. 'FaBiO and CiTO: Ontologies for Describing Bibliographic Resources and Citations'. *Journal of Web Semantics* 17 (December): 33–43. https://doi.org/10.1016/j.websem.2012.08.001.

[31] Peroni, Silvio, and David Shotton. 2018. 'Open Citation: Definition'. Figshare. https://doi.org/10.6084/m9.figshare.6683855.v1.

[32] Peroni, Silvio, and David Shotton. 2018. 'The SPAR Ontologies'. In *The Semantic Web – ISWC 2018*, edited by Denny Vrandečić, Kalina Bontcheva, Mari Carmen Suárez-Figueroa, Valentina Presutti, Irene Celino, Marta Sabou, Lucie-Aimée Kaffee, and Elena Simperl, 119–36. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-00668-6_8.

[33] Peroni, Silvio, David Shotton, and Fabio Vitali. 2012. 'The Live OWL Documentation Environment: A Tool for the Automatic Generation of Ontology

Documentation'. In *Knowledge Engineering and Knowledge Management - 18th International Conference, EKAW 2012, Galway City, Ireland, October 8-12, 2012. Proceedings*, edited by Annette ten Teije, Johanna Völker, Siegfried Handschuh, Heiner Stuckenschmidt, Mathieu d'Acquin, Andriy Nikolov, Nathalie Aussenac-Gilles, and Nathalie Hernandez, 398–412. Lecture Notes in Computer Science. Berlin, Germany: Springer. https://doi.org/10.1007/978-3-642-33876-2_35.

[34] Peroni, Silvio, David Shotton, and Fabio Vitali. 2012. 'Scholarly Publishing and Linked Data: Describing Roles, Statuses, Temporal and Contextual Extents'. In *Proceedings of the 8th International Conference on Semantic Systems*, 9–16. I-SEMANTICS '12. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2362499.2362502.

[35] Resmini, Andrea. 2010. 'Information Architecture Modeling for Historical and Juridical Manuscript Collections.' PhD diss., Università di Bologna. http://amsdottorato.unibo.it/2941/1/andrea_resmini_tesi.pdf.

[36] Riva, Pat, Martin Doerr, and Maja Žumer. 2008. 'FRBRoo: Enabling a Common View of Information from Memory Institutions'. Conference paper. 2008. https://repozitorij.uni-lj.si/IzpisGradiva.php?id=35461.

[37] Uschold, Mike, and Michael Gruninger. 1996. 'Ontologies: Principles, Methods and Applications'. *The Knowledge Engineering Review* 11 (2): 93–136. https://doi.org/10.1017/S0269888900007797.

[38] Vieira, Jose Miguel, and Arianna Ciula. 2007. 'Implementing an RDF/OWL Ontology on Henry the III Fine Rolls'. In *Proceedings of the OWLED 2007 Workshop on OWL: Experiences and Directions*, edited by Christine Golbreich, Aditya Kalyanpur, and Bijan Parsia. http://ceur-ws.org/Vol-258/paper06.pdf.

[39] Zhitomirsky-Geffet, Maayan, and Gila Prebor. 2016. 'Toward an Ontopedia for Historical Hebrew Manuscripts'. *Frontiers in Digital Humanities* 3. https://doi.org/10.3389/fdigh.2016.00003.