Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing

Schriften des
Instituts für Dokumentologie und Editorik

herausgegeben von:

Band 15

Schriften des Instituts für Dokumentologie und Editorik — Band 15

Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing

edited by

Elena Spadini, Francesca Tomasi, Georg Vogeler

2021

BoD, Norderstedt

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über **http://dnb.d-nb.de/** abrufbar.

Digitale Parallelfassung der gedruckten Publikation zur Archivierung im Kölner Universitäts-Publikations-Server (KUPS). Stand 5. Dezember 2021.

Herstellung und Verlag: Books on Demand GmbH, Norderstedt
ISBN: 978-3-7543-4369-2
Einbandgestaltung: Stefan Dumont nach Vorarbeiten von Johanna Puhl und Katharina Weber
Satz: LuaTEX, Bernhard Assmann

Contents

Projects and Editions

Appendices

Infrastructures and Technologies

Formal Models

Formal Semantics for Scholarly Editions

Hans Cools (†), Roberta Padlina

Abstract

In the NIE-INE project based at the University of Basel, Switzerland, a national IT-infrastructure is being developed to support scholarly editions. In a first phase, it focuses on data representation and publication, and in a second phase on online editing. The paper deals with the Semantic Web technological (SWT) part, based on W3C standards, used to express edition project data and related knowledge in a formal way. A section on the project objectives discusses the benefits of formalizing the editions, ensuring the FAIR-principles. The hurdles of this process are also described. The state of the art deals with the components of the infrastructure. The development has a complex group of dependencies, of which the core technology provider (graph database and API) has the most impact. A first step, and major part of the SWT, is the development of formal vocabularies (ontologies) to express the editions' semantics. Also, external generic ontologies (e.g., CIDOC-CRM) are used, together with basic modeling patterns. The important features of the modelling process are described, for example, that expressing data in machine-interpretable standard languages requires explicitness. The results describe the differentiation and integration of ontologies, which are highly reusable for future projects. Different ontological graphics are used for different purposes.

1 Introduction

The aim of this paper is to present the implementation and development of Semantic Web technology (SWT) within the ongoing project *National Infrastructure for Editions* (NIE-INE 2019a).

The overall goal of NIE-INE, funded by swissuniversities[1] in 2016, is to create a Swiss infrastructure that includes an environment for the online publishing and editing of scholarly editions produced by projects in the Humanities. The chosen core technologies are the SWT standards developed by W3C (W3C 2001). The work comprises creating vocabularies or ontologies based on these standards, to convert

[1] https://www.swissuniversities.ch/en/themen/digitalisierung/p-5-wissenschaftliche-information/projekte/nie-ine.

Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing. Ed. by Elena Spadini, Francesca Tomasi and Georg Vogeler. Schriften des Instituts für Dokumentologie und Editorik 15. Norderstedt: Books on Demand, 2021. 97–124.

existing or new scientific data models and research data into formal machine interpretable expressions for long-term preservation, as well as further processing with machine reasoning.

Semantic Web technologies are particularly suitable for representing highly complex entities (like scholarly editions) and their relationships. Since questions of meaning and interpretation are central in the Humanities, it is essential to provide scholars with tools to express their research data and results in a formal, explicit, and self-descriptive way – and this is exactly where SWT proves its efficacy. The implementation of this technology initially corresponds to the creation of ontologies which, in the Semantic Web context, are formal data models that specify the concepts (classes), and the relations between concepts (properties), belonging to a domain knowledge. The W3C gives two concise definitions of an ontology in SWT: "a conceptualization of a domain to enable knowledge sharing" (W3C 2009) and "a representation of terms and their interrelationships" (W3C 2004b).

The most innovative aspect of the Semantic Web, in comparison to traditional formats and standards (e.g., TEI[2]/XML[3]), is that in an ontology, not only entities and their relations (linked data) are formally expressed, but also the tacit domain knowledge that forms the foundation of research data (for example, editorial principles and decisions, domain assumptions and perspectives, methodologies and workflows, and so on). This knowledge is crucial for an overall understanding, but is often implicitly hidden in the data. Texts and their editions, for example, are embedded in complex webs of discourse and narratives setting multidimensional relations (Robinson 2013, 107; Gabler 2010, 44). There is a well-known tension between two editorial perspectives (text-as-work and text-as-document), but, even if these perspectives set different interpretation contexts, they are indissolubly connected, and a scholarly edition must account for both (Robinson 2013, 123). SWT is precisely designed to cope with a wide range of information variance, and to allow multiple (even opposite) viewpoints to coexist within a single system. The conceptual description of the document-text-work triad, as well as concepts like authority, intention, agency and meaning, can hugely vary from one edition to another, but, from the moment that these fundamentally interpretative conceptualizations are explicitly formalized, the Semantic Web allows them to exist in a common semantic space wherein they are also explicitly related, and not locked in self-contained models or semantic silos.

Of course, to introduce variance, you first need commonality, and this is why knowledge formalization requires a minimal commitment and consensus on the part of domain experts. Achieving this is not without challenges, but the effort required

---

[2] The Text Encoding Initiative is a consortium that maintains a standard for the representation of text in digital form, http://www.tei-c.org/index.xml.

[3] The eXtensible Markup Language is another standard maintained by the W3C, https://www.w3.org/TR/2008/REC-xml-20081126/.

pays off, as it dramatically increases the expressiveness and the reusability of data and domain knowledge. It also eases interdisciplinary exchange, representing a clear benefit for other existing or future editions and humanities projects. Indeed, the lack of consistent authority files is one of the major challenges that Linked Open Data is facing at the moment. Therefore, the adoption of SWT will facilitate the establishment of formal linked data encompassing the FAIR-principles[4]. We believe that developing and implementing SWT represents a service for research in the Humanities, which, given the ever-increasing amount of digital data that scholars are producing, will prove more and more crucial.

2 Objectives

The main objective of the NIE-INE project in using Semantic Web technologies is the formal expression of scholarly editions, enabling semantic interoperability across editing projects, and, in the long run, creating a semantic space for an interdisciplinary formal knowledge domain for the Humanities. This interoperability is meant for both humans and machines, and, hopefully, will contribute to the long-term preservation of scientific data. To the best of our knowledge, the development of a variety of domain-specific ontologies for scientific editing in Humanities is quite a new niche, in which no comparable work has been done. The aim of NIE-INE is to create such a new digital environment, via the synthesis of existing technologies, and by creating new components i.e., formal ontologies, queries and rules, as well as a back- and front-end environment.

NIE-INE adheres to the FAIR-principles for scientific data management: scientific data should be **f**indable, **a**ccessible, **i**nteroperable, and **r**eusable. This requires appropriate data storage, editing and publication of scientific data.

The advantages of SWT translate to the FAIR-principles as follows:

- *Accessibility* is a feature (not equal to openness) of the Internet and Web technology, thus also of SWT.
- On the Internet and Web, resource locations are *findable* by their identifiers (URL[5]). On the Semantic Web all resources are identified (IRI[6]). Having an identifier ensures the ability to talk about the same thing, and having ontological elements as semantic identifiers enables mutual understanding.
- *Reusability* is enabled by the use of explicit, self-descriptive semantics to express information, with minimal ambiguity, in formal ontologies, providing data transparency and quality assurance. The terminological and conceptual consensus

[4] The FAIR-principles are: Findable, Accessible, Interoperable, and Reusable (Wilkinson et al. 2016).
[5] Uniform Resource Locator.
[6] Internationalized Resource Identifier. A URL is a type of IRI.

process for including domain knowledge in the ontologies also contributes to this principle.
- Due to the built-in logic (Berners-Lee 1998) of the formal standard languages, the ontology and data expressions are machine interpretable, enabling semi-automated *semantic interoperability*, going beyond the mere linking of hardware and software. The abstract expressions are natural language independent, enabling global information exchange. Thus, it is possible to link data from disparate sources and knowledge domains, adding domain knowledge and, therefore, facilitating reuse and multidisciplinarity.

The machine-interpretable semantics of the formal ontologies, data, and rules hugely increases the information expressiveness, and also allows rule-based machine reasoning. This kind of Artificial Intelligence further improves data quality through consistency and validity constraints checks. It is used to enhance data expressiveness by means of many kinds of calculations, for example, interval calculus for temporal reasoning, calendar conversion, data comparison and analysis.

The formal data models in SWT are much more flexible, and independent from the technological implementation, than the models in relational databases or XML; new data models can therefore be implemented in a more sustainable way. In comparison to SQL[7] databases, where column names and identifiers are local, IRIs make links and values explicit and globally identifiable. Another main advantage of the Semantic Web model is that it is non-hierarchical. That is why it can represent complex entities and their relationships in a better way than is possible in a hierarchical, tree-based data model like XML. Thus, SWT can overcome problems resulting from XML implementations (e.g., developing a native stand-off solution), which can also have repercussions for the usage of the TEI guidelines. TEI is considered the standard for the representation of textual resources, but, being an XML-based language, TEI is more a syntactic and serialization language than a formal semantic model. TEI's usage in the community is largely influenced by circumstances and pragmatic factors, since the TEI guidelines allow for many ways of expressing the same textual feature, while the same TEI element can be subject to different interpretations depending on the context. Moreover, in XML and TEI the text has to be organized according to a single "ordered hierarchy of content objects" (OHCO), which makes it very difficult to represent concurrent hierarchies and overlapping features of textuality (DeRose et al. 1990; Renear et al. 1996). This is why some proposals to formalize the semantics of the TEI framework have already been made (Ciotti 2018).

The implementation of SWT, including the migration of existing data to the RDF format, isn't effortless, but entails some hurdles and challenges, of which it is important to be aware:

---

[7] Structured Query Language, used for managing data in relational databases.

- SWT is often perceived as replacing existing database systems (mainly SQL and XML). However, being one of the most modern data technologies, it is, rather, offering a complementary environment to the classical databases and data processing tools.
- SWT is not yet perceived as a standard in the Humanities. On one hand, being situated in the back-end as a middleware, it is not directly visible for non-IT specialists; on the other hand, being declarative and quite different from other, imperative programming, it is perceived as an extra challenge: notably the "open world assumption" often makes people feel uncomfortable.
- SWT needs the implementation of a complex environment with different elements covering it end-to-end, from input database, through formal semantic data model, to user interface. It is not merely another data model, but also includes reflections on semiotics, semantics, linguistics (in relation to different natural languages), logic, and IT, comprising standard formal languages and their ontologies, a query language, and rule-based machine reasoning. According, a learning curve has to be taken into account.
- Developing an ontology for a new given domain also requires an initial overhead building up the foundation of a semantic space, needing discussions and consensus with domain experts.

3 State of the Art

In the NIE-INE project, different technological domains are brought together: SWT, an overall Web framework called *inseri* (former NIE-OS)[8], as well as the coordination and development of digital editing tools[9]. For development using SWT, the W3C formal standard languages are used:

- the Resource Description Framework (RDF) (W3C 2004a; W3C 2014a), RDF Schema (RDFS) (W3C 2014b), and Web Ontology Language (OWL) (W3C 2004b; W3C 2012), are used to express formal ontologies and data;
- the RDF Query Language SPARQL (W3C 2013), for information retrieval from an RDF graph database;
- Notation3 (N3) (W3C 2011), to express formal inference rules for machine reasoning.

These standards are based on set theory, model or interpretation theory, and first order logic.

---

[8] The source code is published in open access as a Git repository, https://github.com/nie-ine/inseri. A test instance is available at http://test-nieos.nie-ine.ch/.

[9] The presentation of the state of the art of *inseri* and editing tools is beyond the scope of this paper.

The machine reasoner used is Euler Yet another proof Engine (EYE) (Verborg & De Roo 2015).

The NIE-ontologies are serialized in Turtle (W3C 2014c), a subset of N3, which is much more human-readable than the initial RDF-XML standard format.

Of course, the NIE-INE project did not start creating formal ontologies from scratch. Concerning modelling OWL-ontologies in general, we refer to Allemang and Hendler (2011), and in Humanities we refer to Eide and Ore (2018).

Standard ontologies such as Friend of a Friend (FOAF) (Brickley & Miller 2014), Dublin Core Metadata Initiative (DCMI 1995), and Simple Knowledge Organization System (SKOS, Miles & Bechhofer 2009) are used, together with some of the basic ontologies of the SWEET series (ESIP 2019).

Concerning domain-oriented terminology for the Humanities, NIE-INE ontologies are based on:

- International Committee for Documentation – Conceptual Reference Model (CIDOC-CRM) (CIDOC 2006);
- Functional Requirements for Bibliographic Records object-oriented extension to the CIDOC-CRM (FRBRoo) (Dunsire 2014; IFLA 1998).

4 Methodologies

4.1 Dependencies

In the NIE-INE project there are a series of dependencies that directly influence the implementation of SWT, especially the modelling of ontologies:

- The project makes use of the W3C standards, the aforementioned standard and existing domain ontologies, basic modelling patterns, and best practices.
- NIE-INE supports overall eleven projects, eight of which are currently using SWT. Moreover, several new projects have already indicated their willingness to collaborate with us. The input dependency is represented by the original data model and data, mostly in XML (5 projects, which primarily, but not exclusively, use TEI), SQL (2 projects), or mixed (1 project).
- Last but not least, there is a dependency on the tacit knowledge of the domain specialists working on the editing projects, which is required for the implementation of domain knowledge beyond the restricted database models. To uncover this knowledge, a commitment to modelling on the part of domain specialists is needed, in order to reach consensus on semantics beyond their own specific research objectives.

All these dependencies represent a major challenge and, initially, a substantial overhead, but the return on investment (ROI) is big, and will be even bigger the more

project database models are formalized using SWT. This formalization will also contribute to the more general cross-project semantics, removing the need to model the same concepts for new projects multiple times. Reaching a wide consensus (ideally, on an international scale) on domain terminology contributes to the ROI and semantic interoperability.

4.2 Preliminary Analysis

Knowledge of the W3C standards and the aforementioned standard and basic ontologies, together with knowledge of input data model formats like XML, especially TEI, and SQL, was present from the start of the NIE-INE project. Existing domain ontologies, like CIDOC and FRBRoo, and the applied framework[10] had to be analysed. For every edition project that NIE-INE supports, an in-depth investigation of the data model is also provided. This implies a deep understanding of digital scholarly editing, requiring models to be produced as the result of an iterative epistemological hermeneutic process (Pierazzo 2015).

4.3 Basic Modelling Patterns

By conceiving or adopting basic modelling patterns that will be abundantly used in ontologies, modelling begins before an ontology is actually declared. Such patterns comprise one or more basic concepts and their sets of relations (properties). Examples of such patterns are *event* and *role*[11]. *Event* and its consecutively derived concepts *process*, *action*, and *procedure* are considered as four-dimensional entities described in time (start, end) and place (geographic names and coordinates), having inputs and outputs (process), with agents (person, organization, and even software) having roles (functions) in an action. *Procedure* is a particularly useful pattern. It captures both the perspectives of a time-space entity and the result thereof, making it possible to clearly distinguish between, for example, an edition as a procedure (editing), and an edition as a resulting product; it is therefore well-suited for describing scientific editing as a variety of procedures, involving agents with different roles, and having a variety of in- and outputs.

Moreover, in the Humanities, it is essential that data are positioned in time and space. Therefore, indicators for dates and places are primordial. Even approximate indicators can be used (which is especially useful for data concerning ancient times).

A more domain-oriented modelling pattern concerns the concept of *reference*, which

[10] https://www.knora.org/

[11] For more information on basic modelling patterns, http://e-editiones.ch/ontology-modeling#basic-modeling-patterns.

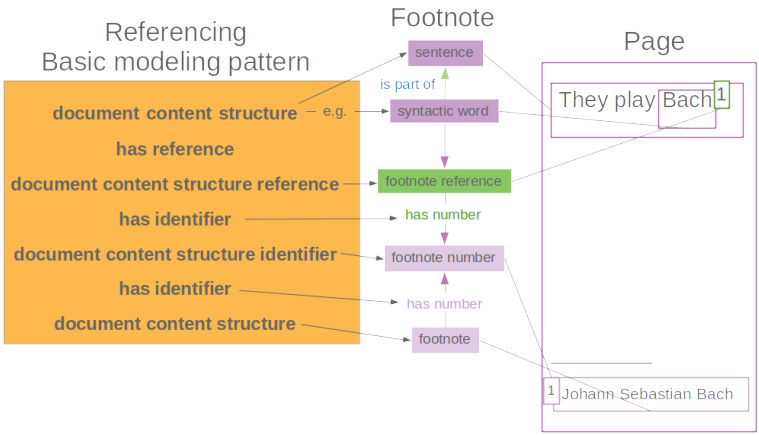

Figure 1. Basic modelling pattern for the concept *reference*.

is especially important for the Humanities. Figure 1 shows the concepts involved using the *footnote* entity as an example.

4.4 Development

Ontologies, SPARQL queries and N3 rules are created with a text editor, and ontologies are checked on syntax and logic consistency in the open source editor Protégé[12].

The whole development evolves in a very iterative way, requiring the connection of different roles and expertise. It is very important to have regular discussions between NIE-INE modelers (4) and edition projects contacts (5), as well as with domain specialists (12), as it is impossible to capture the required project semantics all at once.

4.5 Identification and Structure

The basic expression unit structure in SWT is the triple, consisting of a *subject*, a *property* (predicate), and an *object*, each being an IRI, or, in case of an object, also a plain or typed literal value. A set of triples makes a graph: an ontology and RDF data are graphs; the result of a SPARQL query is also a graph. Figure 2 shows an example of such an identification and serialization of a sentence in natural language, encoded in XML, and converted into RDF triples.

[12] Protégé is an ontology editor developed by the Stanford Center for Biomedical Informatics Research, https://protege.stanford.edu/.

Natural Language expression as literal:

```
''The person has the name Petrus Lombardus.''
```

In XML:

```
<person>
  <name>Petrus Lombardus</name>
</person>
```

RDF triples in Turtle:

```
example:personX a human:person . # 'a' is syntactic sugar for rdf:type
example:personX human:hasNameLiteral ''Petrus Lombardus'' .
```

Note:
`example:`, `rdf:` and `human:` are replacments of the respective namespace part of a full IRI, e.g. `<http://www.e-editiones.ch/ontology/human\#>`, making Turtle more redable.

Figure 2. Sentence in natural language and XML converted in RDF triples.

4.6 Modelling

Explicit Statements

Since the W3C OWL ontologies contain the built-in logic of the RDF model theory, they are machine interpretable. When basing an own ontology on these, its elements have to be declared in an explicit way, with minimal hidden assumptions, to keep them machine-interpretable. Being explicit also represents a good way to reveal flaws in the original or formal data model.

It is important to point out that this process of introducing standards and being explicit does not reduce the expressiveness of the data. On the contrary, RDF permits the co-existence of distinctive expressions of domain knowledge, as long as they are made explicit.

For human development and usage, ontologies and ontological elements obtain clear multilingual labels and a description. We think it is important to keep the latter concise, because the longer it is, the more scope there is for interpretation, rather than clarifying meaning.

Reification and Abstraction

CIDOC-CRM and FRBRoo come with extra levels of abstraction and reification. Examples of reification are the class `cidoc:E41_Appellation`, and the subclass

`cidoc:E44_Place_Appellation`. An instance of the latter has an IRI – that is, it is a thing, not a string – which, in turn, has a name as a literal string value, e.g., "Lausanne". Similarly, all instances of the class `cidoc:E33_Linguistic_Object` have an IRI, which in turn are linked to literal expressions. Examples of abstraction levels are the basic classes `frbroo:F1_Work`, `frbroo:F2_Expression`, `frbroo:F4_Manifestation_Singleton`, and `frbroo:Item`.

Building further on these FRBRoo classes, in the NIE-INE ontologies there are e.g., four different classes dealing with *page*:

1. `information-carrier:Page`: a physical surface of a leaf, e.g., in a manuscript or a book;
2. `document:Page`: a content structure with information, e.g., text lines and graphs, as part of a document expression, on an information carrier page;
3. `text-structure:Page`: text structure as part of a document page and a text expression; and
4. `text-editing:Page`: text page of a scientific edition.

Middle-Out

Adhering to the principles *as simple as possible, as complex as needed,* and *the least ontologic commitment* (i.e., providing the necessary and sufficient semantics), the modelling faces a challenge in finding the right balance between ground elements and details. Deep grounding is provided by very basic or upper ontologies, e.g., CIDOC-CRM. On the other end of the spectrum, there is the rather ad hoc project-specific modelling in a stand-alone way, which is not useful for semantic interoperability. Middle-out modelling begins with some required points of depth and detail, in a way that makes it possible to easily extend ontologies.

Multitude of Namespaces

In creating a semantic space for broad applicability, ontology size, and consequently the number of ontologies developed, also matters. Having an *all-in-one* approach is rather difficult if the intended research-space is broad, implying multiple domains. A scientific project always needs general concepts such as *person*, *document*, *text*, *image*; these concepts are typically reused from more general ontologies. A project ontology can exist, but should be based on these general ontologies, and only contain project-specific elements, for example, the concept of *Dreissiger* (a set of about 30 verses) used in the *Parzival*-project.

Partitioning knowledge over ontologies upfront is also a challenge, but doing so is important in order to avoid the need for later division of an ontology. Moreover, new

knowledge is constantly produced, possibly leading to deprecation of old knowledge and, thus, requiring the splitting of ontologies.

Property Oriented

Entities obtain meaning based on the way, and in the extent to which, they are related to other entities via properties. Databases often contain implicit, condensed, or shortcut semantics. In order to be explicit, it is important that these semantics are unravelled, and consequently that more properties are added: a simple example is the conversion of a name to a family name and a given name. Together with the previous modelling feature (multitude of namespaces), this leads to a *network* or a web of OWL ontologies and RDF data graphs.

Consensus

Last, and certainly not least, iterative discussions with the project domain specialists must lead to a minimal consensus about domain concepts. As previously stated, consensus is also essential for enabling semantic interoperability. This makes modelling a very *collaborative* and multidisciplinary activity.

4.7 Database to RDF Mapping

Once the semantics of a project are covered in the ontologies, the original data model can be mapped to RDF using the classes and properties defined in the ontologies. This is done in a spreadsheet, using the following methodology.

For XML, parent nodes are mapped to RDF classes, and child nodes and attributes are mapped to properties. Content and attribute values become property values of instances.

For SQL, table names are mapped to RDF classes, and column headers are mapped to properties. Keys are used to instantiate a row as a member of a class, with a series of properties mapping to the field values.

After the mapping is complete, a script is written (e.g., in XSLT or Python) to convert project data in XML or SQL to RDF, in order to import the data into an RDF database or triple store. During conversion, the data elements obtain an IRI, and are checked for consistency with the built-in logic of the applied W3C standards part.

Once in the triple store, RDF data can be checked using large SPARQL queries, covering the whole semantics of an edition. From this point forward, RDF data can be retrieved with specific SPARQL queries, converted to JSON and then either consumed by JavaScript-based web frontends, or further processed. Repeated representation or functionality can be supported by existing SPARQL query libraries, which can retrieve the necessary parts of RDF data.

The following is a summary of the tasks in the workflow of formal semantic modelling in NIE-INE:

1. Ontology development as servicexs and products
   1.1. Initiating new project
   1.2. General semantics emerged, but independent from individual edition project
   1.3. Modelling iteratively, multiple projects simultaneously
   1.4. Change management (e.g. dependency on application)
   1.5. Graphics
   1.6. Information
2. Mapping original data model to RDF using ontologies
   2.1. Test ontology XML-schema generation (application)
3. Import scripts for bulk data import
4. Test triple store data with SPARQL
5. Create SPARQL query sets for apps

Future:

6. Notation3 rules for machine reasoning
   6.1. Second setp (semantic conversion) in the 2-step formalization
7. 2-step formalization

The most demanding tasks are 1.2, 1.3, 2, 3, 6 and 7.

4.8 Future

Machine Reasoning

In the near future, we intend to apply machine reasoning on the formal data using the built-in logic of the formalisms. Besides the ontologies, these formalisms will comprise N3-rules for different purposes: to provide automated data quality assurance through consistency checks (e.g. cardinality of instance existence); to infer new data from existing data using the built-in domain knowledge of N3-rules, thus enhancing data expressiveness in a multidisciplinary environment; to perform temporal reasoning (interval calculus). Temporal reasoning can be applied, for example, in case of an event (basic modelling pattern) that is lacking start or end date, e.g., when a birthdate of a life (as an event) is missing. By defining a maximum life span, a certainty period can be established, calculating the earliest birthdate. Similarly, in a case where two different birthdates are mentioned for the same person, a certainty interval for the birth event, comprising both dates, can be calculated, together with the certainty interval of the derived life event, with the earliest birthdate as the start. Concerning time indicators NIE adopts the EDTF level 0 (Library of Congress 2019). The other

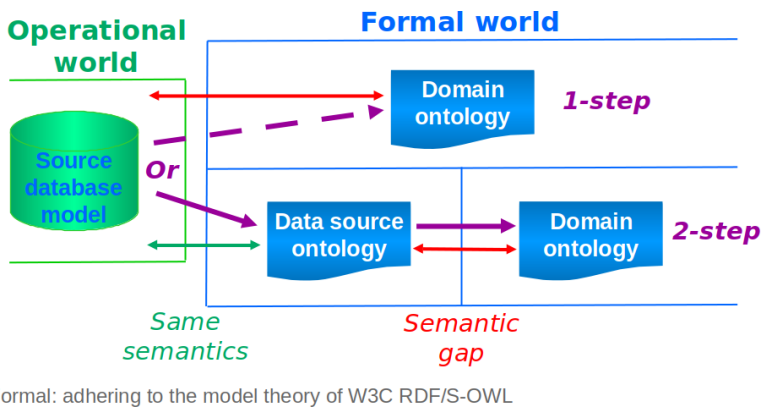

Figure 3. 1-step vs. 2-step formalization from source database model to formal domain ontology.

levels are already partially supported with N3-rules, and can be further implemented. RDF result sets of the machine reasoning process are then added to the triple store. As already mentioned, the reasoner being used is EYE[13].

2-Step Formalization

To formalize the data in a flexible and convenient way, a so-called 2-step formalization is planned for the near future, decoupling the database semantics from the formal domain knowledge. Until now, RDF data representing formal domain ontologies are created in one step, and have a direct link to the original source data. In a 2-step system, the first step corresponds to a 1-1 conversion of the original data to RDF, without interpretation, or making explicit all of the semantics (see Figure 3). The data can then be stored in this format in the RDF database. In a second step, the RDF data are converted to explicit, enriched semantics using domain ontologies and N3-rules. In this way, it is possible to enable semantic interoperability in a more decoupled way, that is less dependent on data source and application specificities.

Another example of temporal reasoning can be seen at the second step: the simple literal time indicators are converted to typed literal data values, i.e., *year*, *month*, and *day*, go from being three fields in a relational data table, to being represented as an interval of one year with a start and end `xsd:dateTime` (e.g., from *1886*, *9*, and *4* to start: `1886-09-04T00:00:00.0Z^^xsd:dateTime` and end: `1886-09-04T23:59:59.999997854Z^^xsd:dateTime`).

[13] EYE is an open source reasoning engine, https://github.com/josd/eye.

Number of → in ↓	Ontology	Rdfs:Class	owl:ObjectProperty owl:DatatypeProperty
Generic ontologies	37	671	497
Project ontologies	8	135	44
Total	45	806	541

Table 1. NIE ontologies in numbers.

Collaboration

It is also our intent to strengthen the collaboration with other projects using SWT, in particular with *histHub*[14], the Scholastic Commentaries and Text Archive (SCTA[15]), SARI (Swiss Art Research Infrastructure), and the Zentrum für Informationsmodellierung at the University of Graz[16].

5 Results

All *ontologies* created in the NIE-INE project are open source and published on GitHub as Turtle files (NIE-INE 2019b). There are two series: generic and project ontologies. The former is, *grosso modo,* further divided into four levels: 1) general domain, 2) general Humanities, 3) specific Humanities, and 4) external terminology and code systems ontologies (see Figure 4). They differ quite a lot in size, granularity and specificity. This division is somewhat arbitrary, meaning that it isn't formalized, but it is convenient to illustrate the articulation of ontologies and their interrelations. Table 1 shows the status of the modelling at the end of September 2019. The most populated ontologies are *human*, *information carrier*, *document*, *text*, *text-expression*, *text-structure*, *scholarly-editing*, *publishing*, *literature*, and *linguistic-morphology*. The *time*-ontology contains properties mainly for N3-rule declarations.

Ontologies can easily be added to or extended. In the development phase until end 2019, we still foresee possible splits of faster growing ontologies, and the transfer or replacement of elements.

Ontologies and their elements can be analysed from different perspectives and entry points. In NIE-INE, four kinds of graphics are used to illustrate the ontologies, to show classes and properties and instances in different ways.

[14] The aim of histHub is to establish and operate a research platform for the historical sciences, https://histhub.ch/.

[15] SCTA's aims is to connect and freely distribute the intellectual history of the scholastic tradition, https://scta.info/.

[16] https://informationsmodellierung.uni-graz.at/.

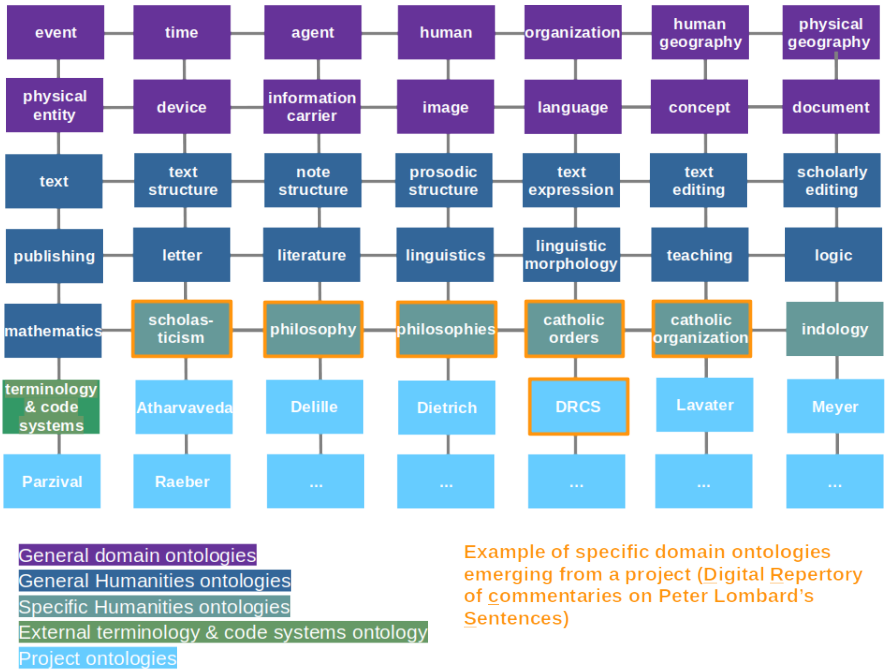

Figure 4. A simplified representation of the NIE-INE web of ontologies.

A first, manually created, graphic type (Figure 5) shows different core domain ontological elements in a simplified way, offering a broad overview of the dependencies between the ontologies, while enabling a focus on certain semantics, (e.g., on text and critical editing,) while also providing more general semantics.

A second type of graphic (Figure 6) centers one entity, e.g., *event*, relating it to elements from different domain ontologies, and adhering (in a reduced way) to the RDF structure, e.g., indicating prefixes and full property names. Classes are represented by circles; properties by rectangles. Classes from external ontologies are colored orange in the upper part. The subclass property is represented by a dotted arrow. This graphic is created with Grafo[17].

A third type of graphic (Figure 7) focuses on (a part of) an ontology, also representing properties as nodes. It is created with the open source SPARQL-visualizer[18],

[17] Available at https://gra.fo/.
[18] Available at https://github.com/MadsHolten/sparql-visualizer.

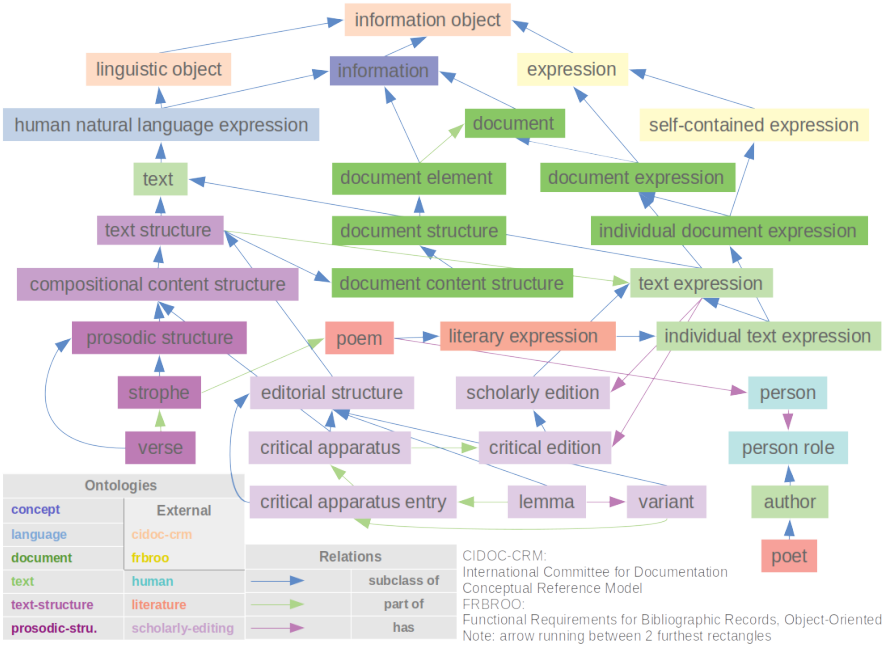

Figure 5. Graphic representing core classes and properties from different domain ontologies.

filtering some ontological elements out (labels, comments, and blank nodes) to enhance readability. It is particularly suitable for discussions on terminology with domain specialists.

A fourth type (Figure 8), created with Protégé, depicts a subsumption tree of a merged group of ontologies. It is very useful to get a quick and broad hierarchical overview during the modelling stage. In the example, two trees are shown, building on CIDOC-CRM and FRBRoo. The left tree contains document expression, text expression, and subclasses. The right tree contains document structure, text structure, and subclasses (note: not all document structure subclasses are shown).

For further graphic examples we refer to the NIE-INE GitHub site. We will now discuss in more detail the different levels of ontologies.

5.1 General Domain Ontologies

Although all ontological classes are instantiable, the more general ones will often function as *glue* to search in an RDF-database with SPARQL queries, and to enhance

Figure 6. Graphic representing classes and properties from different ontologies concerning the concept *event*.

Figure 7. The concept-ontology.

machine reasoning, e.g., for subsumption (subclassing). For example, if a scholar wants to search for all the languages in which a text is translated, the property *is translated into* can be used without specifying the language. In other words, all instances of all subclasses of the class *human natural language* are retrieved. In another case, one would like to find all information carriers (e.g., manuscripts and prints) bearing creations from of a certain author, across more than one project: one project refers to prints existing in an archive and having a signature, and another project uses manuscripts preserved in a library with a manuscript identifier; in this instance, both the manuscripts and prints can be found with a super-property *is on carrier*.

As a general domain ontology, the *concept*-ontology (see Figure 8) describes, among other things, concepts created by a person as abstract ideas, e.g., symbolic and propositional, basing on CIDOC-CRM and FRBRoo. It contains entities such as *information*, *identifier*, and *thought-method*, and their relations to each other, and to other entities,

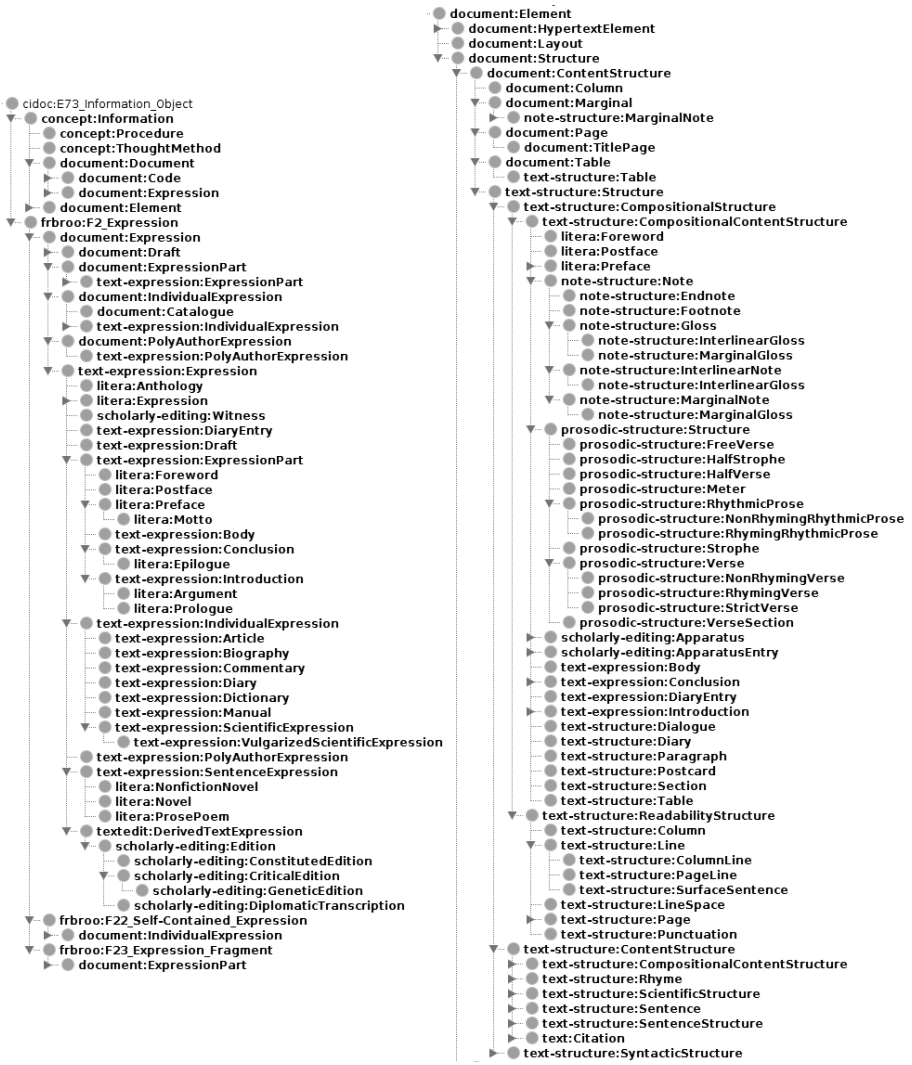

Figure 8. Two subsumption trees representing classes from different ontologies concerning the concepts document expression and text expression on the left and document structure and text structure on the right.

e.g., persons. The *document*-ontology (see also Figures 5 and 8) describes documents and document structures, e.g., tables, identifiers, and references, such as footnotes. It also contains the abstract document expression as based on FRBRoo[19], and the different relations between document structures.

5.2 General Humanities Ontologies

The *general Humanities ontologies* comprise the core concepts of scholarly editions, and the major part of entities common to the individual models of single editions. The following is a description of five core vocabularies used for scholarly editing in Humanities (see Figures 5 and 8, above).

- Text: this ontology describes text as a human natural language expression, serialized in writeable form. It contains all kinds of text forms (e.g., written, typewritten, transcribed, and printed) and the roles of persons processing text (e.g., editor, annotator, and citer). It serves as the basis for all text-related ontologies: text-expression, text-structure, text-editing, and literature ontology.
- Text expression: the eponymous core concept bases (via the document-ontology) on FRBRoo, that abstracts text from its carrier (see Figure 8). The ontology defines further related roles (e.g., author and commentator) and general expression types (e.g., draft and commentary). It provides the basis for the literature- and scholarly-editing-ontology.
- Text structure: this ontology describes all kinds of textual structures (see Figure 8), e.g., syntactic, compositional, content, scientific, readability. They form an upper layer to enable more flexible and extensive search, as well as machine reasoning. More specific entities are word, sentence, paragraph, section, line, and text column. Extensions of this ontology are the prosodic-structure-ontology and the note-structure-ontology, containing entities such as verse and strophe, and marginal note and gloss, respectively. An important relation between structures is *part of*, which, by its transitive nature, enables searching and machine-reasoning in a way that does not require explicitly stating all possible relations between structures in the data, since they can be inferred via transitivity.
- Text editing and scholarly editing both of these terms describe the necessary semantics for editing, the latter extending the former with specific scholarly entities, e.g., diplomatic transcription, critical edition, different types of apparatus, lemma, variant, editorial comment, witness, siglum and so on. Also, related roles are declared, e.g., editor, glossator, and critical text editor. An extensive set of properties relates these entities to one other, as well as to text, text structure and expression elements (see Figures 5 and 8).

[19] FRBRoo provides the concept of expression, abstracted from its carrier.

- Publishing: this ontology describes classes and properties related to publishing, that is, to publication and its subclasses: printed and web publication, serial-like newspaper, periodical, magazine, etc. There is a substantial set of properties relating expressions and other entities to elements in this ontology.
- Literature: this ontology describes literary genres such as narrative, different kinds of poetry, and further different types of literary expressions (e.g., poem, hymn, novel) and their subclasses. It also contains related roles, e.g., poet and novelist. Different types of literary structures are declared, e.g., foreword, preface, prologue, and epilogue, and related properties (see Figures 5 and 8).

5.3 Specific Humanities Ontologies

This series comprises more specific entities, as used in different specialized domains in the Humanities, e.g., about scholasticism in medieval philosophy. Some ontologies (e.g., *indology*, *catholic organization* and *philosophy*) are for the moment only providing the more general concepts as needed for the current projects, but they can, and will, be extended. *Catholic orders* and *philosophies* describe subclasses of classes in aforementioned ontologies, which will also be extended.

Although the scope of these ontologies is narrower, i.e., more project-oriented, the entities can be reused in another context, if applicable, meaning that they do not need to be restricted to a specific project.

External terminology and code systems ontology This ontology contains formal descriptions of terminology and code systems, and their datatypes, as links between such systems' data, OWL-ontologies and RDF-data. Examples of terminology and code systems in the Humanities are the ISO[20] standard OAIS[21], and the GND[22] for the DACH countries. A generic system, for example, is using hexadecimal color coding to describe e.g., text color. Other datatypes are declared in respective domain ontologies. Examples are `calendar:julianDate` in the calendar-ontology, to type a Julian date literal, `languages:iso639-2` in the languages-ontology, to type ISO language standard codes, and `text:characterSize` in the text-ontology, to type a numeral representing character size.

Project ontologies These ontologies contain entities that are only used in their respective projects, but which are still usable outside those projects, if applicable. For example, the Parzival-ontology contains the concept of *Dreissiger*, (being a set of about 30 verses,) which would be reusable in another project about the verse novel *Parzival*.

---

[20] International Organization for Standardization. Cf. https://www.iso.org.
[21] Open Archival Information System. Cf. http://www.oais.info/.
[22] Gemeinsame Normdatei. Cf. https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd_node.html.

All other generic ontologies are, to different degrees, re-used in the project ontologies by creating subclasses or subproperties from definitions in the generic ontologies.

Generally, another project about a same author could reuse some ontological elements in order to be linkable with one of the projects supported by NIE-INE. It then would be possible to query the two different SPARQL endpoints simultaneously, which is essential for research, since the triple stores can contain complementary information on the same subject. This is actually the case for the DRCS project (dealing with the commentaries on the *Sentences* of Peter Lombard), which is linked to another project at the University of Baltimore, U.S. This case demonstrates the added value of semantic interoperability between disparate databases, facilitated by the use of the same external ontologies CIDOC-CRM and FRBRoo.

Atharvaveda-ontology This ontology represents the formal description of specific concepts in the critical edition of the Paippalāda Recension of the Indian Atharvaveda (UZH 2020), an anonymous collection of verse songs about everyday life (c. 1200-1000 BC). It contains subclasses and subproperties of elements in the text, text expression, text editing, prosodic structure, literature, and Indology ontologies, thus, modelling a sub-set of the Indian Veda literature.

Delille-ontology This ontology represents the formal description of specific concepts in the scientific edition of the third canto of Jacques Delille's (1738–1813) verse poem *L'homme des champs* (*The Rural Philosopher*) (Marchal 2020). One of the main goals of the project is the reconstitution of the reception of Delille's poetry. The main concepts concern poetic expression and its structures, the works (e.g., anthologies, dictionaries, and articles) and their authors (and other actors) that cite verses of the canto, and the scientific commentaries on the citers and their works.

Dietrich-ontology This ontology represents the formal description of specific concepts in the critical edition about father Joseph Dietrich's monastery diary (1670-1704) (WSU 2020). The concepts are diary-related. The majority of the semantics are expressed using more generic ontologies.

DRCS-ontology This ontology represents the formal description of specific concepts in the scientific study Digital Repertory of Commentaries on Peter Lombard's *Sentences* (DRCS, Peter Lombard c.1096–1160) (Zahnd 2020). The project collected over 1700 commentaries, from which further data has been extracted, such as the identification of the authors, their names and life dates, their membership of religious orders, their philosophical thinking and tradition, their roles in the editing processes (e.g., editor, abbreviator, corrector), along with all graspable information about the text itself, like genre, location and time of creation, as well as bibliographic information. The specific Humanities ontologies containing

those concepts are shown in Figure 4. The DRCS-ontology itself mostly contains the different types of commentaries and the Stegmueller-related concepts. Interlinking this collected data through our ontologies provides highly dynamic possibilities for evaluating and detecting as-yet-unknown patterns, interrelations and networks in medieval philosophy and intellectual history. The reception of texts and writers, thought patterns, writing traditions or thought methods can be detected and traced down through time.

Kuno Raeber-ontology This ontology represents the formal description of specific concepts in the online publication of the lyric work of the Swiss poet Kuno Raeber (1922–1992) (Morgenthaler 2020). As the first model of a project-ontology in the learning curve of NIE, it contains more than the average number of multiparent subclasses. On the other hand, there is also the need for single instance classes leading to a more extensive ontology, than for other projects. The concepts mainly concern the different expression formats (written, typed, etc.), their carriers, and their convolutes, strongly representing the FRBRoo ontology.

Lavater-ontology This ontology represents the formal description of specific concepts in the critical edition of correspondence of Johann Caspar Lavater (1741–1801) (UZH Deutsches Seminar 2020). All concepts in this ontology are letter-related, as subclasses of more generic classes, enabling the internal structure to be kept close to other sources, as in the example e-manuscripta.ch.

Meyer-ontology This ontology represents the formal description of specific concepts in the critical edition of the correspondence of Conrad Ferdinand Meyer (1825–1898) (Lukas 2020). As in the Lavater-ontology, above, in this ontology all the concepts are letter-related, as subclasses of more generic classes; this occurs on the level of expressions and information carriers, as defined by FRBRoo.

Parzival-ontology This ontology represents the formal description of specific concepts in the critical and digital edition of the verse novel *Parzival* by Wolfram von Eschenbach (c. 1160-c. 1220) (Stolz 2020). One part of the concepts describes a series of transcriptions (or parts) of the verse novel, and corresponds therefore with the levels of expressions and information carriers in FRBRoo; the other part of the concepts describes the critical edition with different apparatus. This can be achieved by making extensive use of the scholarly editing ontology.

Wölfflin-ontology This ontology represents the formal description of specific concepts in the scientific study Heinrich Wölfflins gesammelte Werke (1864–1945) (UZH Kunsthistorisches Institut 2020). Due to the current absence of a consolidated source data model, the ontology is not yet published. However, it should be noted that most of the required semantics is already covered in more generic ontologies.

The Kritische Robert Walser-Ausgabe (1878–1956) (UNIBAS Deutsches Seminar 2020) and the Anton Webern Gesamtausgabe (1883–1945) (UNIBAS Musikwissenschaftlichen Seminar 2020) govern their own ontologies, which they intend to link to our ontologies at a later stage.

5.4 Expected Results

We will further develop application-independent SWT, enabling semantic interoperability and reusability adhering to the FAIR-principles. The ontology library will grow, and the generic ontologies will stabilize, but still be extendable. We will increasingly model to allow for broader usage in the Humanities, beyond scholarly editing, and in consensus with aforementioned parties. The ontologies will be declared with enhanced expressiveness, considering the whole of RDF/S and OWL ontological elements. We will also start implementing machine-reasoning in different ways, to enrich research data with inferred domain knowledge; temporal reasoning will be an important part of this. We will have more ontological graphics, and extend the documentation, particularly regarding our modelling methodology, workflow, and best practices.

6 Conclusion

Being able to connect different research projects using formal semantics is more a consequence of the implementation of the SWT standards and modelling, than the primary intention of research projects. Of the eleven editions we support, only DRCS initially expressed the need for semantic interoperability (in this case with another US project with the same subject).

Even if researchers tend to focus on the specificities of their topics and objects, they have an important part of the basic semantics of research projects in common. It, therefore, makes sense to invest a little more effort to obtain broadly reusable models, which capture these common semantics. Such models, which come about as the result of consensus (without loss of expressivity), and which are based on standards of W3C and the domain of Humanities (e.g. CIDOC-CRM and derivatives), are necessary in order to avoid project-specific semantic silos, in which the same concepts are modelled over and over again for different projects, in a multiplication of effort and cost. Only in this way is the need for tedious reverse engineering minimized.

In our experience, researchers on an individual project quickly become aware of the advantages of *RDF-izing* their data, from the quality assurance on different levels (due to the explicitness of the formal data), to performing machine-reasoning on that data to answer research questions. Domain specialists don't have to dive into SWT themselves to understand its functionality and potential. We are aware that the

following is a bold statement, but we are convinced that with RDF-data, one can infer new data in a way that is impossible with other data models, because of the lack of the built-in logic of the formal languages RDF/S, OWL and Notation3 (N3).

Although the notion of interoperability is included in the FAIR-principles, the understanding of semantic (machine) interoperability is growing slower. It is not enough to use the SWT standards and adopt *upper* ontologies. Consensus about domain semantics at different levels of specificity is indispensable. The willingness of domain specialists to discuss terminology in order to obtain a critical mass of consensus models is growing due to the aforementioned direct RDF advantages. Although domain specialists do have discussions among themselves to reach a semantic consensus, this consensus remains on the level of human understanding and is not explicitly formalized in the data models to be machine-interpretable and reusable. This is why there is a real need for SWT experts who are able to liaise between domain specialists and IT staff, in order to bring the semantic consensus to a new level. Also, the creation of a collaborative online environment to create a library of ontologies has been very helpful, e.g., the NIE-INE GitHub repository (NIE-INE 2019b).

The project terminology discussions concerning the formalization of domain knowledge, facilitated by a SWT expert who operates across various projects and domains, complies with the tendencies of *multidisciplinaritization* and Linked Open Data (LOD) promoted by politics and policies. Of course, the intellectual property rights aspects have to be dealt with appropriately.

Because the formal data are independent from natural language, foreign collaboration is also facilitated, internationalizing domain semantics.

Different projects that deal with the same topics are likely to cover different aspects that are complementary, and very worthwhile to connect on the formal semantic level, increasing the chance of new research findings. Here, the next step in the technological implementation comes into play: machine-reasoning based on N3-rules consuming RDF-data and OWL-ontologies, for semantic conversion and for answering research questions.

Bibliography

Allemang, Dean, and James Hendler, *Semantic Web for the Working Ontologist* (Oxford: Elsevier LTD, 2011)

Berners-Lee, Tim, 'The Semantic Web as a Language of Logic', 1998 <https://www.w3.org/DesignIssues/Logic.html>

Brickley, Dan, and Libby Miller, 'FOAF Vocabulary Specification 0.99', *FOAF Vocabulary Specification 0.99*, 2014 <http://xmlns.com/foaf/spec/>

CIDOC, 'CIDOC Documentation Standards Working Group, and CIDOC CRM SIG', 2006 <http://www.cidoc-crm.org/>

Ciotti, Fabio, 'A Formal Ontology for the Text Encoding Initiative', *Umanistica Digitale*, 3 (2018) <https://umanisticadigitale.unibo.it/article/view/8174>

Ciotti, Fabio, and Francesca Tomasi, 'Formal Ontologies, Linked Data, and TEI Semantics', *Journal of the Text Encoding Initiative*, 9 (2016) <https://doi.org/10.4000/jtei.1480>

DCMI, *Dublin Core Metadata Initiative Schemas*, 1995 <https://web.archive.org/web/20190930150058/https://www.dublincore.org/schemas/>

DeRose, Steven J., David G. Durand, Elli Mylonas, and Allen H. Renear, 'What Is Text, Really?', *Journal of Computing in Higher Education*, 1.2 (1990), 3–26

Dunsire, Gordon, 'Functional Requirements for Bibliographic Records Object-Oriented Extension to the the CIDOC Conceptual Reference Model', 2014 <http://metadataregistry.org/schema/show/id/94.html>

Eide, Øyvind, and Christian-Emil Ore, 'Ontologies and Data Modeling', in *The Shape of Data in Digital Humanities* (Routledge, 2018)

ESIP, 'Official Repository for Semantic Web for Earth and Environmental Terminology (SWEET) Ontologies', 2019 <https://github.com/ESIPFed/sweet>

Gabler, Hans Walter, 'Theorizing the Digital Scholarly Edition', *Literature Compass*, 7.2 (2010), 43–56 <https://doi.org/10.1111/j.1741-4113.2009.00675.x>

IFLA Study Group, *Functional Requirements for Bibliographic Records*, IFLA Series on Bibliographic Control 19 (Munich, 1998) <https://web.archive.org/save/https://www.ifla.org/publications/functional-requirements-for-bibliographic-records>

Library of Congress, *Extended Date/Time Format (EDTF) Specification* (Library of Congress, 2019) <https://www.loc.gov/standards/datetime/>

Lukas, Wolfgang, 'C. F. Meyers Briefwechsel-Kritische Ausgabe', *C. F. Meyers Briefwechsel-Kritische Ausgabe*, 2020 <http://www.cfmeyer.ch/>

Marchal, Hugues, 'Reconstruire Delille', *Reconstruire Delille*, 2020 <https://delille.philhist.unibas.ch/>

Miles, Alistair, and Sean Bechhofer, 'SKOS Simple Knowledge Organization System Namespace Document', 2009 <https://www.w3.org/2009/08/skos-reference/skos.html>

Morgenthaler, Walter, 'Kuno Raeber Lyrik', *Kuno Raeber Lyrik*, 2020 <https://www.kunoraeber.ch/lyrik/>

NIE-INE, 'Nationalen Infrastruktur Für Editionen - Infrastructure Nationale Pour Les Éditions', *Nationalen Infrastruktur Für Editionen*, 2019a <http://e-editiones.ch/about>

———, 'NIE-INE Ontologies', 2019b <https://github.com/nie-ine/Ontologies>

Pierazzo, Elena, *Digital Scholarly Editing: Theories, Models and Methods* (Farnham, Surrey: Ashgate, 2015)

Renear, Allan H., David G. Durand, and Elli Mylonas, 'Refining Our Notion of What Text Really Is: The Problem of Overlapping Hierarchies', in *Research in Humanities Computing*, ed. by S. Hockey and N. Ide (presented at the ALLC/ACH Conference, Christ Church, Oxford, April 1992, Oxford: Oxford University Press, 1996) <http://cds.library.brown.edu/resources/stg/monographs/ohco.html>

Robinson, Peter, 'Towards a Theory of Digital Editions', *The Journal of the European Society for Textual Scholarship*, Variants, 10 (2013) <https://doi.org/10.1163/9789401209021_009>

Stolz, Michael, 'Parzival-Projekt', *Parzival-Projekt*, 2020 <http://www.parzival.unibe.ch/home.html>

UNIBAS Deutsches Seminar, 'Kritische Robert Walser-Ausgabe', *Kritische Robert Walser-Ausgabe*, 2020 <https://kritische-walser-ausgabe.ch/>

UNIBAS Musikwissenschaftlichen Seminar, 'Anton Webern Gesamtausgabe', *Anton Webern Gesamtausgabe*, 2020 <https://anton-webern.ch/index.php?id=17>

UZH, 'Online Edition of the Paippalāda Recension of the Atharvaveda', *Online Edition of the Paippalāda Recension of the Atharvaveda*, 2020 <https://www.atharvavedapaippalada.uzh.ch/en.html>

UZH Deutsches Seminar, 'Lavater-Edition', *Edition Johann Caspar Lavater*, 2020 <https://www.lavater.uzh.ch/de.html>

UZH Kunsthistorisches Institut, 'Heinrich Wölfflin - Gesammelte Werke', *Heinrich Wölfflin - Gesammelte Werke*, 2020 <https://www.woelfflin.uzh.ch/de.html>

Verborgh, Ruben, and Jos De Roo, 'Drawing Conclusions from Linked Data on the Web: The EYE Reasoner', *IEEE Software*, 32.3 (2015), 23–27 <https://doi.org/10.1109/MS.2015.63>

W3C, 'Notation3 (N3): A Readable RDF Syntax', *Notation3 (N3): A Readable RDF Syntax*, 2011 <https://www.w3.org/TeamSubmission/n3/>

———, 'OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax (Second Edition)', *OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax (Second Edition)*, 2012 <https://www.w3.org/TR/owl2-syntax/>

———, 'OWL Web Ontology Language Overview (W3C Recommendation 10 February 2004)', *OWL Web Ontology Language*, 2004b <https://www.w3.org/TR/2004/REC-owl-features-20040210/>

———, 'Product Modelling Using Semantic Web Technologies', *Product Modelling Using Semantic Web Technologies*, 2009 <https://www.w3.org/2005/Incubator/w3pm/XGR-w3pm-20091008/>

———, 'RDF 1.1 Primer', *RDF 1.1 Primer*, 2014a <https://www.w3.org/TR/rdf11-primer/>

———, 'RDF 1.1 Turtle RDF Triple Language', *RDF 1.1 Turtle*, 2014c <https://www.w3.org/TR/turtle/>

———, 'RDF Primer (W3C Recommendation 10 February 2004)', *RDF Primer*, 2004a <https://www.w3.org/TR/rdf-primer/>

———, 'RDF Schema 1.1', *RDF Schema 1.1*, 2014b <https://www.w3.org/TR/rdf-schema/>

———, 'SPARQL 1.1 Overview', *SPARQL 1.1 Overview*, 2013 <https://www.w3.org/TR/sparql11-overview/>

———, 'W3C Semantic Web', *Semantic Web*, 2001 <https://www.w3.org/standards/semanticweb/>

Wilkinson, Mark D., Michel Dumontier, and I. Aalbersberg, 'The FAIR Guiding Principles for Scientific Data Management and Stewardship', *Scientific Data*, 3 (2016) <https://doi.org/10.1038/sdata.2016.18>

WSU, 'Das Kloster-Tagebuch Des Einsiedler Paters Joseph Dietrich, 1670–1704', *Das Kloster-Tagebuch Des Einsiedler Paters Joseph Dietrich, 1670–1704*, 2020 <http://www.dietrich-edition.unibe.ch/>

Zahnd, Ueli, 'A Digital Repertory of Commentaries on Peter Lombard's Sentences', *A Digital*

Repertory of Commentaries on Peter Lombard's Sentences, 2020 <https://drcs.zahnd.be/index.php>

Projects and Editions

Appendices

Biographical Notes

Thomas Ahrend (University of Basel, Switzerland – thomas.ahrend@unibas.ch) studied Musicology, Philosophy and Literary Studies in Frankfurt a. M. and Berlin. He received his MA 1996, and his PhD 2005 at Technische Universität Berlin with a dissertation on the instrumental music of Hanns Eisler. 1997–2010 member of the editorial staff of the Hanns Eisler Gesamtausgabe in Berlin. Since September 2010, member of the editorial staff of the Anton Webern Gesamtausgabe at Musikwissenschaftliches Seminar at University of Basel.

Peter Boot (Huygens ING, The Netherlands – peter.boot@huygens.knaw.nl) studied mathematics and Dutch language and literature; he wrote his PhD thesis about annotation in scholarly digital editions and its implications for humanities scholarship. He oversaw the creation of the digital edition of the letters of Vincent van Gogh. He is employed as a senior researcher at the Huygens Institute for the History of the Netherlands where he works, among other things, as a consultant in several edition projects.

Manuel Burghardt (University of Leipzig, Germany – burghardt@informatik.uni-leipzig.de) is head of the Computational Humanities Group at Leipzig University. He is interested in the use of digital tools and computational techniques to explore new modes of doing research in the humanities. His most recent areas of research are Sentiment Analysis in the Humanities, Drametrics, Computational Intertextuality, Computational Analysis of Movies and Series and Music Information Retrieval.

Toby Burrows (University of Oxford, United Kingdom – toby.burrows@oerc.ox.ac.uk) is a Senior Researcher in the Oxford e-Research Centre at the University of Oxford, and a Senior Honorary Research Fellow in the School of Humanities at the University of Western Australia.

Hugh Cayless (Duke University, USA - hugh.cayless@duke.edu) is Senior Digital Humanities Developer at the Duke Collaboratory for Classics Computing. Hugh has over a decade of software engineering expertise in both academic and industrial settings. He also holds a Ph.D. in Classics and a Master's in Information Science. He is one of the founders of the EpiDoc collaborative and currently serves on the Technical Council of the Text Encoding Initiative.

Hans Cools (University of Basel, Switzerland – 1961-2021) had a master degree in medicine and a specialization in orthopaedic surgery and traumatology (Universities of Ghent and Antwerp, Belgium, 1997), a bachelor's degree in physical

therapy, and a standalone degree in informatics (1999). Through various research and project management positions, in both companies and academic institutions, he gained expertise in different aspects of the Semantic Web technologies, focusing particularly on formal data modeling and machine reasoning. Those positions were in internationally collaborative research projects in a biomedical setting, mainly of the 5-7th EU Framework Program. Foremost in these projects were semantic interoperability and reusability of data. Since 2016, he worked in the humanities, as knowledge engineer, ontologist, and Semantic Web technology expert, at the University of Basel, as part of the NIE-INE project, which highlights scholarly editing. He (co-)published several articles, and gave workshops on the implementation of Semantic Web technologies in biomedicine and the humanities. He passed away in April 2021.

Francesca Giovannetti (University of Bologna, Italy – francesc.giovannett6@unibo.it) is a second-year PhD student in Digital Humanities at the Department of Classical Philology and Italian Studies, University of Bologna. She received an MA in Digital Humanities from King's College London and a second cycle degree in Digital Humanities and Digital Knowledge from the University of Bologna. She is interested in combining digital scholarly editing with semantic web technologies and in the use of digital technologies in education.

Matthew Holford (University of Oxford, United Kingdom – matthew.holford@bodleian.ox.ac.uk) is Tolkien Curator of Medieval Manuscripts at the Bodleian Library, University of Oxford.

Marijn Koolen (Royal Netherlands Academy of Arts and Sciences - Humanities Cluster, The Netherlands – marijn.koolen@gmail.coml) studied artificial intelligence and wrote his PhD thesis on using hyperlinks in information retrieval algorithms. He has worked on scholarly annotation for digital humanities research and on annotation-related information behaviour and information systems. He works as a researcher and developer at the Humanities Cluster of the Royal Netherlands Academy of Arts and Sciences, where he leads a project on developing annotation support within the *CLARIAH research infrastructure* project.

David Lewis (University of Oxford, United Kingdom – david.lewis@oerc.ox.ac.uk) is a Research Associate in the Oxford e-Research Centre at the University of Oxford.

Andrew Morrison (University of Oxford, United Kingdom – andrew.morrison@bodleian.ox.ac.uk) is a Software Engineer in the Bodleian Digital Library Systems and Services, Bodleian Library, University of Oxford.

Stefan Münnich (University of Basel, Switzerland – stefan.muennich@unibas.ch) studied musicology and communication science at the Technische Universität Berlin, MA 2011 with a thesis on cantional setting in Heinrich Schütz's Becker-Psalter. 2012 research assistant, 2013–2015 research associate of the Felix Mendelssohn Bartholdy. Sämtliche Briefe edition at University of Leipzig (co-editor of vols. 9 & 12). Since October 2015 research associate of the Anton Webern Gesamtausgabe, Basel; received his Doctorate degree in 2020 at the department of musicology at the University of Basel with a dissertation about music notation and its codes.

Iian Neill (Digital Academy of the Academy of Sciences and Literature, University of Mainz - Iian.Neill@adwmainz.de) is a visiting researcher at the Digital Academy of the Academy of Sciences and Literature Department at the University of Mainz, Germany. He is the creator of Codex, a text annotation environment which uses standoff property annotation to generate entities in a graph meta-model. Codex is currently being used to produce a digital edition of the epistles of Hildegard von Bingen at the Digital Academy in Mainz.

Roberta Padlina (University of Basel, Switzerland – roberta.padlina@unibas.ch) studied medieval philosophy at the University of Fribourg, Switzerland, obtaining a doctoral degree in June 2020. She has twelve years of professional experience in the field of Digital Humanities, thanks to which she has been able to work closely with different actors involved in the online publication of open access research. Roberta has worked for several years for e-codices –Virtual Library of Manuscripts in Switzerland and currently coordinates the National Infrastructure for Editions (NIE-INE) project. Roberta's main focus is on the opportunities and challenges that the digital shift poses for traditional education and research institutions, including developing semantic web strategies for scholarly publications and cultural goods.

Kevin Page (University of Oxford, United Kingdom – kevin.page@oerc.ox.ac.uk) is a Senior Researcher in the Oxford e-Research Centre and Associate Member of Faculty in the Department of Engineering in the University of Oxford.

Miller C. Prosser (University of Chicago, USA – m-prosser@uchicago.edu) earned his Ph.D. in Northwest Semitic Philology from the University of Chicago. His academic interests include the social and economic structure of Late Bronze Age Ras Shamra-Ugarit and the use of computational methods for philological and archaeological research. Miller is the Associate Director of the Digital Studies MA program at the University of Chicago where he teaches courses on Data Management and Data Publication for the Humanities. He also works as a

researcher at the OCHRE Data Service of the Oriental Institute of the University of Chicago where he consults with and supports research projects using the Online Cultural and Historical Research Environment (OCHRE). He has also worked as a tablet photographer for the Mission de Ras Shamra (Ugarit) and the Persepolis Fortification Archive Project, employing advanced digital photographic methods such as reflectance transformation imaging, photogrammetry, and high-resolution digital scanning.

Matteo Romanello (Université de Lausanne, Switzerland - matteo.romanello@unil.ch) is Ambizione SNF Lecturer at the University of Lausanne, where he conducts a project on the commentary tradition of Sophocles' Ajax. Matteo is a Classicist and a Digital Humanities specialist with expertise in various areas of the Humanities, including archaeology and history. After obtaining his PhD from King's College London, he worked as a research scientist at EPFL's DHLAB on the Linked Books and Impresso projects, before moving to his current position. He was also teaching fellow at the University of Rostock, researcher at the German Archaeological Institute, and visiting research scholar at Tufts University.

Sandra Schloen (University of Chicago, USA – sschloen@uchicago.edu) is the Manager of the OCHRE Data Service at the Oriental Institute of the University of Chicago, and is the co-designer and developer of the Online Cultural and Historical Research Environment (OCHRE). Trained in computer science and mathematics (B.Sc. University of Toronto; M.Ed. Harvard University), Sandra has spent over 30 years working with technology as a systems analyst, technical trainer, and software developer. A long association with colleagues in the academic community has enabled her to develop a specialty in solving problems in the Digital Humanities where challenges of data capture, data representation and data management abound. Specifically, she has served extensively as a database manager for several archaeological projects in Israel and Turkey, and supports a wide range of research projects at the Oriental Institute and at other universities.

Desmond Schmidt (University of Bologna - desmond.allan.schmidt@gmail.com) has a background in classical Greek philology, information security and eResearch. He has worked on several scholarly edition projects, including the Vienna Wittgenstein Edition (1990–2001), Digital Variants (2004–2008), the Australian Electronic Scholarly Editions project (2012–2013), the Charles Harpur Critical Archive (2014-) and a pilot edition of Gianfrano Leopardi's Idilli (2018-). He currently works on developing practical web-based tools for making, visualising and publishing digital scholarly editions.

Colin Sippl (University of Regensburg, Germany – colin.sippl@ur.de) is currently a project employee at the University Library of Regensburg. Since 2017, he has been working on extending the open access services of the Electronic Journals Library (EZB). More recently, he has started developing and setting up a digital repository for literature, artefacts and experiments relating to the early life sciences based on the Invenio framework. He specialised in textual data mining and the development of media services in the institutional domain.

Elena Spadini (University of Lausanne - elena.spadini@unil.ch) is a postdoctoral researcher at the University of Lausanne. She holds a Ph.D. in Romance Philology from the University of Rome Sapienza (2016) and a M.A. in Digital Humanities from the École nationale des chartes (2014). She was a Marie Curie fellow in the IT Network DiXiT and co-directed the related volume Advances in Digital Scholarly Editing (Sidestone Press, 2017). She published in international journals and taught specialized courses in various European countries in the field of Digital Philology.

Francesca Tomasi (University of Bologna - francesca.tomasi@unibo.it) is associate professor in Archival Science, Bibliography and Librarianship at the University of Bologna (Italy). Her research is mostly devoted to digital cultural heritage, with a special attention to documentary digital edition, and a focus on knowledge organization methods in archives and libraries. She is member of different scientific committees of both associations and journals. In particular, she is President of the Library of the School of Humanities in the University of Bologna (BDU - Biblioteca di Discipline Umanistiche), Director of the international second cycle degree in Digital Humanities and Digital Knowledge (DHDK), President of the Italian Association of Digital Humanities (AIUCD – Associazione per l'Informatica Umanistica e la Cultura Digitale), and co-head of the Digital Humanities Advanced Research Center (/DH.ARC). She wrote about 100 papers and 4 monographs related to DH topics. She is editor and scientific director of several digital scholarly environments.

Athanasios Velios (University of the Arts London, United Kingdom – a.velios@arts.ac.uk) is Reader in Documentation at the University of the Arts London.

Georg Vogeler (University of Graz - georg.vogeler@uni-graz.at) is professor for Digital Humanities at the University of Graz and scientific director of the Austrian Center for Digital Humanities and Cultural Heritage at the Austrian Academy of Sciences. He is a trained historian (Historical Auxiliary Sciences). He spent several years in Italy (Lecce, Venice). In 2011, he became member of faculty at the Centre for Information Modelling at Graz University, where he was nominated

full professor for Digital Humanities in 2016 and head of department in 2019. His research interests lie in late medieval and early modern administrative records, diplomatics (digital and non digital), digital scholarly editing and the history of Frederic II of Hohenstaufen (1194–1250). He was and is part in several national and international research projects related to his research interests.

Christian Wolff (University of Regensburg, Germany – christian.wolff@ur.de) has been Professor of Media Informatics at the Institute for Information and Media, Language and Culture at the University of Regensburg since 2003. He holds a PhD in information science and is a habilitated computer scientist. His research interests include: human-computer interaction, multimedia and web-based information systems, (multimedia) software engineering and information retrieval (in particular information literacy and social media).