Patterns in modeling and querying a knowledge graph for literary history

Maria Hinzmann (Trier University), Matthias Bremm (Trier University), Tinghui Duan (Trier University), Anne Klee (Trier University), Johanna Konstanciak (Trier University), Julia Röttgermann (Trier University), Moritz Steffes (Trier University), Christof Schöch (Trier University), Joëlle Weis (Trier University)

0 Abstract

This paper investigates the role of patterns for knowledge production in the field of data-based literary historiography. The starting point and key object is the knowledge base on the French Enlightenment novel that has been created in the project *Mining and Modeling Text: Interdisciplinary applications*, *informational development*, *legal perspectives (MiMoText)*. This knowledge base offers structured knowledge in the field of literary history — modeled as a graph and queryable via SPARQL as query language. After the introduction, the contribution first proposes a theoretical framework that contextualizes the approach of "atomizing literary history" as a specific way of operationalization. Then, we illustrate the added value of the project-specific knowledge graph both through rather simple query patterns and through showcasing the importance of federated queries that are actually substantial regarding the potential of the Linked Open Data paradigm. Finally, we turn to infrastructure issues in the light of interoperability and reusability.

1 Introduction

This paper investigates the role of patterns for knowledge production in the field of data-based literary historiography. The starting point and key object is the knowledge base on the French enlightenment novel that has been created in the project *Mining and Modeling Text: Interdisciplinary applications*, *informational development*, *legal perspectives*

(MiMoText) at Trier University, Germany.[1] This knowledge base offers structured knowledge in the field of literary history — modeled as a graph and queryable via SPARQL as query language. The paper reflects on the meaning of patterns at different levels and in different dimensions that are relevant to the design, construction and use of our knowledge graph. The added value of our discussion of patterns lies in linking the different contexts in which this notion matters in the fields relevant to the MiMoText project, namely Digital Humanities and specifically Computational Literary Studies as well as Linked Open Data, and the modeling efforts associated with it.

1.1 Aims

We pursue the following goals within the three main sections of our contribution:

(1.) **Theoretical framework:** In Section 2, we propose a conceptual space in which patterns are situated between elements and models, and contextualize this conception within our approach of atomizing literary history. We delve into this concept in relation to the significance of operationalization within the Digital Humanities. We use the term atomization to describe an approach that breaks down the knowledge domain of literary history into elements as the smallest units which can be understood as particles of knowledge and can be viewed at three modeling levels (conceptual, formal / logical, physical / infrastructural). The patterns that arise from elements lie on the next higher level of abstraction and can be observed accordingly on the three levels. They represent an important formation at the intersection both between the three modeling levels and between the construction and the use of the knowledge network. This epistemological conceptualization is intended to contribute to further thinking about both patterns in the Digital Humanities and patterns in the context of an “atomization” of literary history as well as their interconnection. At the same time, there is a transferability insofar as we understand “atomization” as an operationalization approach that can be useful for the construction of knowledge resources such as a Linked Open Data graph database, independently of individual research questions.

(2.) **Constructing and querying a knowledge graph:** Our second objective is to demonstrate in Section 3 the significance of patterns in both creating and querying our knowledge graph on 18th-century French literary history. Additionally, we show how

[1] For more details, see the project website: https://mimotext.uni-trier.de.

distinct patterns built from various elements manifest themselves explicitly within the graph and the queries. These patterns function as a kind of stencil that can be laid over data (graph patterns via SPARQL), whereby the data stored in our triplestore becomes "re-sortable" through "recombining the atoms". The recombination of the atoms through the queries opens up a wide variety of possibilities for exploration. This section is central as it clarifies the added value of the knowledge graph through specific queries and thus illustrates the nature and usefulness of the rather abstract approach of an atomization of literary history.

(3.) **Infrastructural implications:** Finally, we address in Section 4 the pivotal role of technical infrastructures and related data models in realizing the Linked Open Data vision and achieving interoperability and reusability. To actually explore patterns across project boundaries using federated queries and to attain a new quality of research, the continuation and deepening of the discussion on the influence of infrastructures and the concept of federation within the Linked Open Data paradigm is needed.

An overarching aim of this article is to show that patterns are located at different levels and can have different functions, and that they are also discussed at these levels. This is not a question of providing one definition, but rather of opening up a broad horizon based on the structure and use of the project-specific knowledge database, and thus also of showing lines of connection between the different levels. The conceptualization proposal presented in Section 2.1.2 pursues less of a definitional claim but rather aims to stimulate reflection on connections between the construction and the exploratory as well as interpretive use of resources in the digital humanities. Section 4, on the other hand, illustrates that thinking about "patterns" has an important function in realizing the Linked Open Data vision.

1.2 Project context

The overarching goal of the project *Mining and Modeling Text* is to develop algorithmic methods for building a knowledge network constructed from different sources of information: metadata from bibliographic resources, textual features from primary literary texts and statements from scholarly publications. The Linked Open Data paradigm is fundamental for the modeling approach as well as for the infrastructure, because it allows us not only to aggregate the data obtained through information extraction from different sources but also to link them to additional resources of the wider Semantic Web. Quantitative methods (mining) and knowledge graph design (modeling) are thus directly related. The resulting data is freely available in the MiMoTextBase knowledge graph. An accompanying tutorial[2] introduces users to SPARQL (Hinzmann et al. 2022a). It aims to support and encourage users from literary studies and other disciplines in the humanities in their research based on the graph and other Linked Open Data resources. Additionally, we hope that it also enables the development of innovative approaches to our data and at the same time shows the potential of a knowledge resource that can be queried in this way.

Since the three different sources of information in our project have already been described in more detail elsewhere (Schöch 2021; Schöch et al. 2022; Röttgermann et al. 2022; Hinzmann et al. 2022b), we keep the following sketch short. Our knowledge graph is constructed from three source types:

(1.) Bibliographic metadata: The primary source of bibliographic information is the *Bibliographie du genre romanesque français* (Martin, Mylne & Frautschi 1977). This bibliography is special because it not only covers the universe of production of novels — or more precisely fictional prose — published between 1751 and 1800 but also contains keywords providing rich metadata (on narrative perspective, plot location, characters, themes/plot, style/tone) for many of the entries. Although this bibliography is a particular case, bibliographies or library catalogs in general can be thought of as relevant sources of similar information.

(2.) Scholarly literature: Various types of statements can be found in the scholarly literature (overviews of literary history and articles or chapters on more specific topics). From scholarly literature, information can be extracted that is rarely contained explicitly in the other two source types, e.g., relationships and influence between authors and works.

(3.) French novels 1751-1800: The results of various quantitative analysis methods applied to primary works serve as the basis for new statements as well. We have applied topic modeling (Schöch et al. 2022; Röttgermann et al. 2022), named entity recognition of places (Hinzmann et al. 2022b) as well as character and text matching. Further relevant mining methods include sentiment analysis and stylometry.

It is important to us not only to establish comparability between the different sources of information (because only this enables us to integrate the data in a knowledge

[2] See https://docs.mimotext.uni-trier.de.

base) but also to make our triples linkable to further, external data, following the Linked Open Data paradigm. In two pilot projects, we focused on statement types for which we could extract data from all three source types. These were thematic statements and spatial statements. In addition, we further “semantified” the keywords of the bibliographic data originally extracted by Lüschow (2020) and, for example, imported triples on the narrative form of the novels into our knowledge network. In the most recent project phase, we focused on further statements concerning authors and novels. Following the Linked Open Data paradigm, we link, for example, the narrative locations of the novels with the corresponding Wikidata identifiers, which makes it possible to use information stored in Wikidata about these spaces like geographical coordinates.

1.3 Related projects and standards

In the following, we reflect on important modeling decisions in the construction of the knowledge graph with regard to standards, relevant data models and related projects and discuss the relevance of the Linked Open Data paradigm.

1.3.1 Atomizing literary history within the Linked Open Data paradigm

Our approach of atomization, meaning the breaking down of information in its most basic statements, is closely intertwined with the decision to model the knowledge network in the Linked Open Data paradigm (Schöch 2021; Schöch et al. 2022). The Linked Open Data paradigm and the structure of the graph imply a very elementary, reduced basic structure, namely the simple triple structure. This structure, however, due to the scale of the graph — the MiMoTextBase, currently includes 331,671 triples involving 772 different authors, 1,774 different works and 375 thematic concepts[3] — allows new ways of research within the domain of literary history.

This goes along with the fact that our graph does not focus on canonized literary works but that we are able to aggregate statements of 1750 novels (on the basis of the bibliography, Martin, Mylne & Frautschi 1977) and further statements regarding a subset of about 200 novels for which we have established reliable full texts (with statements obtained based on Named Entity Recognition and Topic Modeling as well as further methods), in addition to statements extracted from the scholarly literature (often about canonical

[3] See the results of the following query 1: https://purl.org/mmt/patterns/query1.

works). In contrast to Wikidata, where information on a smaller range of individual works is available, the MiMoTextBase has an enormously high coverage and density of statements.

Not only in cultural and memory institutions but also in Digital Humanities projects, an increasing uptake of the Linked Open Data paradigm is currently visible. For example, the *International Journal of Humanities and Arts Computing* recently published a special issue on “Linked Open Data in Digital Humanities” (Alves 2022). In recent years, we have seen an increase in the number of projects using Wikidata. From a survey by Zhao (2022), it becomes clear that there are certain main application areas for Wikidata in humanities projects that use Linked Open Data, notably annotation, data enrichment, metadata curation and disambiguation. So far, as Zhao shows, most projects consume rather than produce Linked Open Data, for example using existing identifiers from Wikidata or other authority files to uniquely identify and disambiguate entities in their own data. However, a certain number of projects producing new Linked Open Data do exist, such as Factgrid (Simons 2022) or WeChangEd (Thornton et al. 2022) and of course MiMoText.

In general, Linked Open Data is already more integrated in the fields of history (Zhou et al. 2020; Bartalesi, Pratelli & Lenzi 2022; Hyvönen, Leskinen & Tuominen 2023), art history, and linguistics (Passarotti et al. 2020) than in literary history. Some examples do exist, however, like statements on a Serbian subcollection of the ELTeC (European Literary Text Collection) that have been integrated into Wikidata (Ikonić Nešić, Stanković & Rujević 2021), the POSTDATA project modeling European poetry (Bermúdez-Sabel et al. 2022) or the multilingual drama corpus DraCor (Fischer et al. 2019) that links the available works to Wikidata identifiers. The potential of Linked Open Data is also currently being explored for confessional-historical aspects in German Baroque poetics (Haider et al. 2022) and in the context of storytelling (Pasqual & Tomasi 2023).

There are several projects that link their data to Wikidata and also design an ontology in this framework but without claiming to build a systematic ontology of a specific domain in the humanities. Our goal cannot be to model the domain completely but at least to take the first steps and make suggestions that can be further discussed in the Digital Humanities community (especially Computational Literary Studies). A tendency towards more exchange in ontology development is emerging and was recently stimulated, for example, by a workshop organized within the GOLEM project (Pianzola et al. 2023).[4] Before

[4] See https://golemlab.eu/news/ontology-workshop/.

this, there have been individual attempts but with relatively clearly defined goals of the respective ontology, such as the connection of an image and a text database on narratives of the Middle Ages (Nicka et al. 2020) or the ontology in the context of a digital library on Dante Alighieri's works (Bartalesi & Meghini 2016).

1.3.2 Data modeling standards in the humanities

In terms of frameworks or standards for modeling humanities data, the CIDOC Conceptual Reference Model (CRM) (Bekiari et al. 2022; Liu, Hindmarch & Hess 2023) has been a kind of (continually evolving) quasi-standard for some time now — a status that is sometimes too little questioned and does not sufficiently consider the origin and scope of the respective model as well as the implications of its use. Originally, CIDOC-CRM has its roots the cultural heritage field and it has already been addressed that, for example, related to scholarly editions, it does not meet all requirements even in combination with *Functional Requirements for Bibliographic Records* (Spadini & Tomasi 2021). CIDOC-CRM allows "the description of humanities data to a high level of accuracy" (Kräutli, Chen & Valleriani 2021: 208) — but it is an accuracy that is not always necessary or useful and a precision that imposes strong constraints in the modeling of classes and properties, so that its reuse is limited. Often overlooked are the implications in terms of CIDOC-CRM strongly emphasizing the modeling of actors, roles and events, which may be useful in some contexts (e.g. modeling works of classical music and their associated revisions, adaptations and performances, Achichi et al. 2018) but has hardly any relevance in the context of MiMoText, insofar as MiMoText focuses on modeling statements primarily about the content and features of literary works.

In the absence of alternatives, the quasi-standard from the cultural heritage field is often transferred to the Digital Humanities: projects work through the complexity (and the constraints and hierarchies) of the CIDOC-CRM model although it is sometimes questionable whether a high granularity and hierarchization supports or hinders the exploration and research of patterns. In this sense, a differentiation of multiple conceptual levels between real places, particular conceptual representations of places and their representations in fictional worlds was avoided in MiMoText. Such intermediate levels tend to lead to duplications and redundancies, which on the one hand allow for a more precise representation but on the other hand make it difficult to use these data for pattern search.

In the design of the MiMoTextBase, the focus was not on precision but on pragmatism as well as simplicity and usability of the graph. In the development of the associated ontology, the transferability and reusability of the modules to other domains was of primary concern.

In the field of modeling bibliographic data, various standards have emerged and evolved, including METS, MODS, FRBR, FRBRoo etc. In MiMoText, the central bibliographic data (Martin, Mylne & Frautschi 1977) was modeled using the SPAR ontologies which are a kind of integrating umbrella ontology in the field of Semantic Publishing and Referencing Ontologies (Lüschow 2020). Various levels in this RDF representation had an overly detailed granularity in the context of the MiMoTextBase as a literary-historical knowledge graph (editions etc.). Instead, precise modeling of the keywords of the bibliography, allowing integration of this information with the corresponding information annotated in the scholarly literature as well as that obtained by textmining the novels, was realized only within MiMoText.

In the area of text data, XML following the guidelines of the Text Encoding Initiative (TEI) is the quasi-standard and a system in its own right, which has also already been discussed in its relationship to Linked Open Data as well as ontologies (e.g. Eide 2014). However, it differs essentially in that it represents semi-structured data and a stronger subdivision of elements within a view of text as an Ordered Hierarchy of Content Objects (OHCO). Ciotti and Tomasi (2016) consider that “the formalisms offered by the Semantic Web paradigm are mature enough to build a workable semantic extension of the TEI” and discuss various “semantic layers” in TEI but also challenges. The interfaces between TEI and Linked Open Data have been discussed and also realized in an integration of a large part of data from the *European Literary Text Collection* (ELTeC) into Wikidata (Ikonić Nešić, Stanković & Rujević 2021) and in the context of a study on Linguistic Linked Open Data (Stanković et al. 2023). Although a corpus of about 200 full texts modeled in TEI could be built in a subproject of MiMoText, the overarching focus is not on the edition of novels but on the modeling of data extracted from heterogeneous sources. To reduce complexity and to make the MiMoTextBase user-friendly, the decision was made to model the “literary works” as central items exclusively on the work level but not on the level of the individual editions. Thus, within the framework of the WEMI model (work — expression — manifestation — item) associated with the FRBR standard, the items are located at the level with the highest degree of abstraction (Coyle 2022).

Further related projects

The modeling of statements about statements (reification) is important in MiMoText as will be discussed below and is also discussed in other projects. The relevance of such "meta-statements" manifests itself in the particular emphasis on the provenance of or evidence for specific pieces of information. Including provenance information in the data model also, crucially, allows for a simultaneous representation of ambivalent or even contradictory statements within the same knowledge base (Baillie et al. 2021). Other important modeling dimensions for data in the humanities are the uncertainty and doubt that can arise, for example, from multiple perspectives on the provenance of data (Kuczera, Wübbena & Kollatz 2019; Massari et al. 2023).

Moreover, there are precursors of our approach in databases, in the construction of which atomizing decisions have always had to be made, also concretely for our domain of 18th century literary history (Burrows et al. 2021) . In general, the added value of Linked Open Data becomes obvious from these examples in an *ex negativo* way: if all the data generated to date on the domain of the 18th century were freely available and could be queried across project boundaries, a new quality of linking knowledge and the resulting new insights could be gained. The *Banque de données d'histoire littéraire*, for example, is an early project in the field of datafied literary history which was created in 1985 by a team of researchers from the Université de la Sorbonne-Nouvelle (Paris III). The database included metadata on authors, works, publishers, institutions, translations, and libraries, and was launched with the early idea of a computerized vision of literary history (cf. Bernard 1999). Unfortunately, it is no longer available online. The *French Book Trade in Enlightenment Europe* (FBTEE) project is a Digital Humanities project led by Simon Burrows (2021). It uses database technology to map the French book trade 1769–1794 charting best-selling texts and authors over time. The follow-up project *Mapping Print*, *Charting Enlightenment* aims to reinterpret eighteenth-century culture through historical bibliometrics.[5] Finally, the ongoing European Research Council (ERC) project *Measuring Enlightenment: Disseminating Ideas*, *Authors and Texts in Europe (MEDIATE*) aims to study the transnational circulation of books during the Enlightenment (1665–1820) (Montoya & Chartier 2017).

[5] See http://fbtee.uws.edu.au/mpce/.

2 Patterns in the atomization of literary history

2.1 General context and theoretical reflection on the atomization approach

With this in mind, the following aims to illustrate how an atomization of elements of literary history and their recombination can lead to new insights. We assume that changes in the system — in our case, literary history — can be seen at the level of the smallest units and that these can play a central role in the study of the domain. This also implies that we do not turn only to canonical works, but rather, in the context of a data-based literary history, we link a large amount of data on a wide variety of novels for our domain, thus providing a flexible resource that should enable innovative research. The enormously wide-ranging possibilities for combinations that result from the plurality of elements are reflected in the multitude of query options that we will exemplify in the third section.

2.1.1 Operationalization in the Digital Humanities

A fundamental aspect of research in the Digital Humanities is to look for patterns in larger datasets than humanities disciplines can usually handle. A relatively extensive and substantial part of the work is to build the data sets and resources (e.g. corpora) that make such a search for patterns possible in the first place. Data modeling as well as operationalization play an important role in generating the data sets and resources on the one hand and analyzing and interpreting them on the other hand.

Some aspects of such an approach of "atomization" can be related on a more general level to computational thinking, insofar as it is about breaking down a complex problem into simple, individual steps. By bringing together information that was previously only available in separate sources and in some cases not even digitally, in a way that is readable by both machines and humans, numerous new possibilities emerge. This step of decomposing into smaller observable units is central to our notion of atomization as a way of operationalizing.

Like data modeling (Flanders & Jannidis 2018; McCarty 2005), and closely related to it, operationalization can be seen as a "core activity" (Pichler & Reiter 2022) of the Digital Humanities. In the context of "the challenge of 'bridging the gap' from theoretical concepts [...] to results derived from data" (Pichler & Reiter 2022), operationalization plays an important role at the intersection of informatic methods and traditional humanities. Its theoretical and epistemological reflection, however, is an ongoing process. The fragmentation of fields of knowledge into small units as an important step of operationalization has also played a role in other approaches within the Digital Humanities. We would like to relate our notion of elements to factoids in order to sharpen our approach.

The concept of “factoids” has been important in the field of prosopography for a long time (Bradley 2005) and is still being further developed (Hadden, Schlögl & Vogeler 2022). A factoid is defined as “a kind of prosopographical assertion that centers on statements made by [a] historical source” or in other words “is a spot in a source that says something about a person or persons” (Bradley 2017). There are parallels between factoids and elements, especially because a domain of knowledge is decomposed into smaller units. However, we do not see the elements we address with our approach as “structured interpretation” but rather as parts of statement types that represent a kind of stencil with which statements across different source types can not only be extracted or annotated but also newly generated with certain methods. The text snippets referred to in generating statements can range (also depending on the heterogeneity of the source types) from short strings via longer passages to whole texts.

The notion of “factoid” for the outlined particles is rather misleading, in our opinion, insofar as it connotes a facticity that precisely does not correspond to our understanding of humanities data, for which we rather emphasize and model “perspectivity”. Knowledge in the humanities can rather be described as perspectivized or “situated knowledge” (Haraway 1988), not as factual knowledge. Johanna Drucker’s distinction between data and capta reflects “the situated, partial and constitutive character of knowledge production” and the fact that “knowledge is constructed, taken, not simply given as a natural representation of pre-existing fact” (Drucker 2011; cf. the proposal of “situated data” by Lavin 2021). (The ways in which we implement this notion of perspectivized knowledge will be discussed below.)

2.1.2 Patterns within an epistemological space of atomization

The following proposal to illuminate patterns as a kind of bridging concept between elements and models also aims to shed new light on the practice of operationalizing as well as on the relevance of patterns in the Digital Humanities. The visualization (see Fig. 1) serves to reflect on the epistemological status of patterns within research designs and in the construction of data sets and resources.

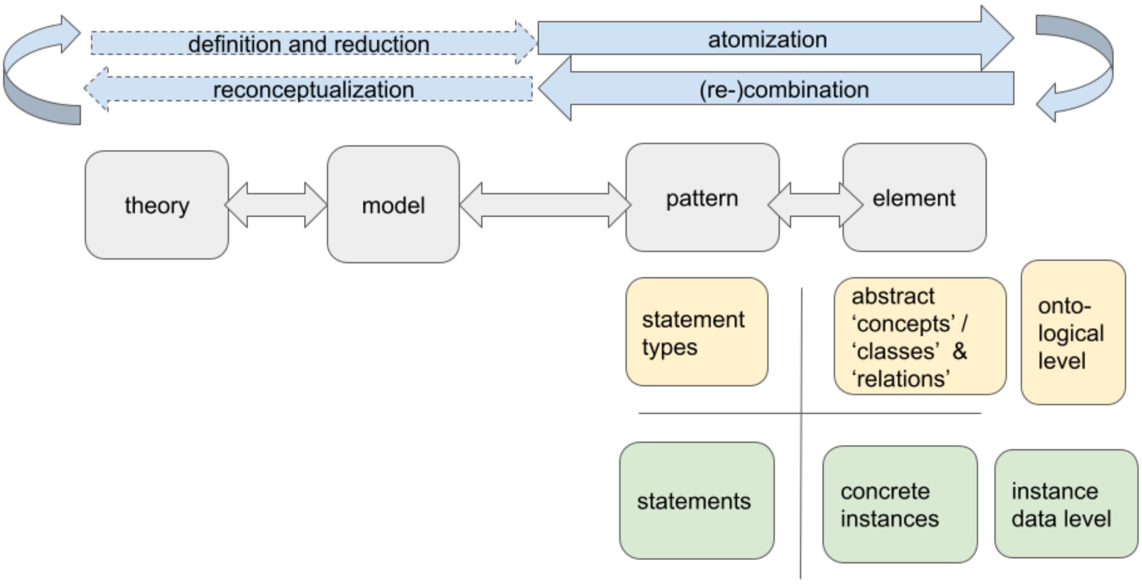


Figure 1: Epistemological contextualization of atomization

In this conceptual space, patterns stand in a kind of intermediate position between elements as the smallest units of information relevant to literary history, on the one hand, and models, on the other. On a scale that connects different epistemological functions and activities, a kind of continuum can be assumed from the individual elements or identifiable elements via patterns and models to theories. Models and patterns lie adjacent to each other in this conceptual space and yet have clearly distinguishable functions: Patterns mediate between the elements and models, whereas models are located between patterns and theories. The four different areas (element, pattern, model, theory) are not connected in a kind of process chain in one direction or the other; rather, their relationship is defined in an iterative or recursive process.

The notion of patterns is omnipresent in the humanities but neither systematically reflected nor consolidated in the sense of an established definition. In the conception of Bod (2018), patterns are at the empirical surface, while the underlying principles belong to another level. This can be related to our understanding of patterns, although the underlying principles should be differentiated, in our understanding, into models and theories. In his *Literary Pamphlet 15*, "Patterns and interpretation," Moretti defines patterns as "the shadows of forms over data," emphasizing the responsibility of the Humanist to establish causation (Moretti 2017: 5). In Moretti's case, as in Bod's, a spatial component also plays a role in the conception of the term which also implies the importance of the relationality of elements for the formation of patterns, which he describes as a “relationship of elements” (Moretti 2017: 5).

The plurality of uses of the pattern concept in the Digital Humanities to date, with no prospect of a consolidated definition, certainly has several causes that can be differentiated. Our aim is not to propose a definition but to reflect upon two differentiations that could serve to make it more precise in different contexts.

(1.) Patterns can play an important role in both operationalization and interpretation: (a.) in the course of operationalization and generation of data sets, insofar as they are a level on which modeling decisions become visible and at the same time also represent characteristics of the data; (b.) on the level of interpretation insofar as interpretation can be seen as a process of detecting patterns in the data sets, which means to connect them with hypotheses, theories, etc. regarding this data. Questions that revolve around this differentiation between interpretation and operationalization are, for example, the following: Are the patterns what is to be interpreted (interpretandum) or what only becomes visible in the interpretation or even the interpretation itself (or a part of it)? To what extent or under what conditions can these three be separated at all? And how does interpreting interrelate with explaining and understanding?

Making the epistemological conceptualization of patterns and its neighbors more explicit regarding the heuristic processes of operationalization and interpretation seems to be important, especially because patterns can thus assume an important function for reflecting on the connection between “becoming visible” and “making visible”. Existing metaphors that are often used in the context of patterns assume either an active (“patterns are discovered” or “revealed”) or not very active (“patterns emerge”) cognitive process. In the light of an iterative process, this is epistemologically undecidable and must remain the subject of reflection.

(2.) In addition, we would like to take a more detailed look at the distinction between different levels of data modeling here with regard to the approach of atomization and the notion of patterns. A distinction can be made between the “conceptual” level, the “logical” (or “formal”) level and a “physical” level (Jannidis 2017). According to this, patterns can be conceived abstractly, can be formalized on a logical level (e.g. in a specific modeling standard), or be implemented or stored on the concrete physical (or technical, infrastructural) level. We will return to this distinction in the following sections and illustrate it by using data modeling in the project.

2.2 Atomizing a domain and constructing a knowledge graph in MiMoText

The approach of atomizing a domain is crucial to the construction of the knowledge network, insofar as the information relevant to the domain of literary history is modeled in the form of a large number of simple triples, that is, statements made up of subject, predicate and object. In this respect, there is an interdependency between decomposing into very simple elements and allowing complexity through the recombination of these elements in queries of the knowledge network.

Whereas in general, the starting point of the operationalization lies in "one or more (theoretical) concepts which are traced back to phenomena on the text's surface via potentially several intermediate steps" (Pichler & Reiter 2022: 7) , the atomization as a specific way of operationalization starts from the segmentation of a domain into elements. Our aim is not the "development of a measurement for a given concept" (Pichler & Reiter 2022: 4) but the representation of a specific domain through the accumulation of multiple elements. In the context of MiMoText, the goal is to be flexible and open with regard to the research questions that users can work on with the queryable data. Due to the central objective to build a versatile knowledge resource, the operationalization process is different from projects that have a more specific focus in this respect.

2.2.1 Atomizing literary history as a knowledge domain

With regard to the history of science, our atomization approach could be considered in the context of skepticism towards grand narratives that has grown since the "linguistic turn" and is often associated with François Lyotard (see e.g. Browning 2000). For the domain of literary history, there is already work that points in this direction, both in analog approaches (e.g. Hollier 1994) and within computational literary studies (e.g. Paige 2020). The atomization of literary history, as exemplified by these works, entails a process of dissecting the narrative or literary works in general into discrete elements, themes or categories. The aim is to resist an all-encompassing narrative and, rather than presenting a unified view of literary history, explore the multifaceted and interconnected nature of (literary) history through diverse lenses and categories. This allows for a more granular understanding of the literary landscape, facilitating the exploration of intricate connections and patterns that may easily be overlooked in a traditional linear narrative.

Beyond modeling, there are additional challenges in building an ontology, insofar as there is no consensus on the central types of statements in literary history, nor on the goals of literary historiography (Borkowski & Heine 2013). It is noteworthy that the potential of the concept of models and modeling has recently received increased attention in literary studies (Erdbeer, Kläger & Stierstorfer 2018; Flanders & Jannidis 2018; Jannidis 2017; Matuschek & Kerschbaumer 2019). However, this has not yet led to a consolidation process from which an understanding or structuring of the domain could be derived. Rather, a pluralism of methods is prevalent in the humanities in general and in literary studies in particular. In this context, we understand literary history as a domain of knowledge in contrast to literary studies as a discipline.

The concept of knowledge can operate within a specific discipline or in an inter- or transdisciplinary setting. From MiMoText's standpoint, literary history is not confined solely to a subfield of literary studies. Instead, as shown in Figure 2, it intersects with various disciplines, contributing to knowledge through diverse methods and perspectives in their engagement with literary materials (primary, bibliographical, or secondary). The particularity of the discipline of literary studies can be found in the fact that there are different fields, which are either more methodological (narratology) or more substantive (postcolonial studies). There are no sharp contours of the domain — it is precisely these inter- and transdisciplinary tendencies that characterize the domain. We are not focused on any of these subdomains or methodological frameworks. Rather, our goal is to relate knowledge particles of varying granularity within the Linked Open Data paradigm, enabling valuable insights across different research interests.

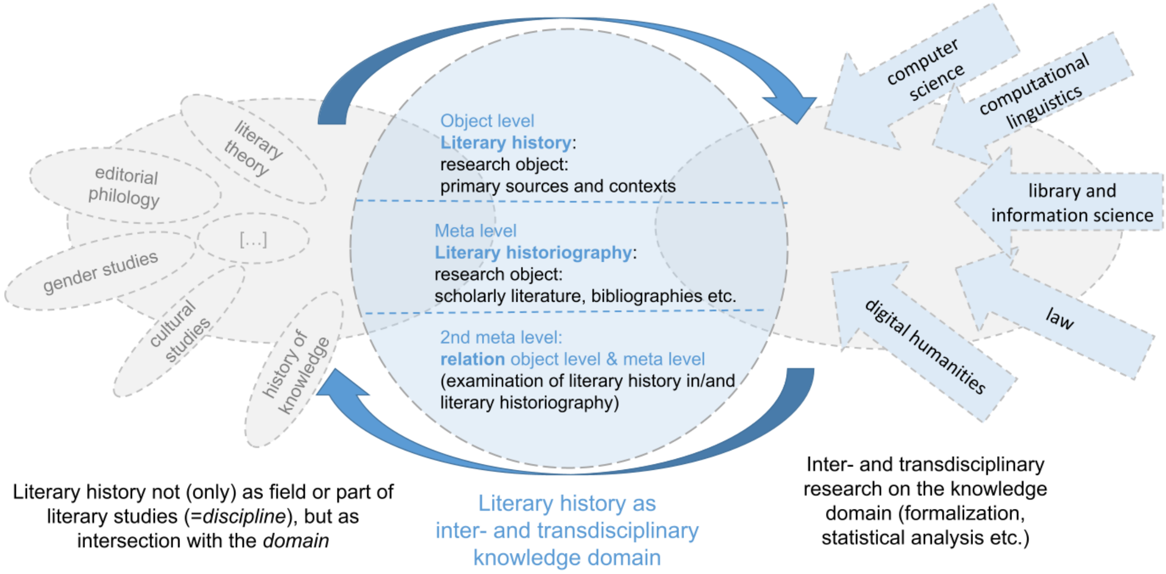

computer science
computational linguistics
library and information science
law
digital humanities
literary theory
editorial philology
[...]
gender studies
cultural studies
history of knowledge
Object level
Literary history:
research object:
primary sources and contexts
Meta level
Literary historiography:
research object:
scholarly literature, bibliographies etc.
2nd meta level:
relation object level & meta level
(examination of literary history in/and literary historiography)
Literary history not (only) as field or part of literary studies (=discipline), but as intersection with the domain
Literary history as inter- and transdisciplinary knowledge domain
Inter- and transdisciplinary research on the knowledge domain (formalization, statistical analysis etc.)

Figure 2: Literary history as a knowledge domain

2.2.2 Structuring a domain

In the process of structuring a domain, the already discussed steps of data modeling can be distinguished.

(1.) On a conceptual level, the relevant elements can be identified via competency questions, which represent a central starting point in ontology development.

(2.) On a logical or formal level, the concrete representation of the elements and their connections in a specific standard or with a related technology must then be established. In our case, this is the modeling in RDF triples within the Linked Open Data paradigm. Related to this, we will discuss certain modeling standards that provide a reference for the statement types of certain data independently of the concrete technology.

(3.) On the physical or infrastructural level, we have opted for a representation of the data in a project-specific Wikibase. This presupposes or entails a data model, whereby it becomes clear that all levels have overlaps or dependencies. We will discuss these in the fourth section of our article.

Conceptual level: Defining the central elements and considering important dimensions of the domain

An ontology is built, generally speaking, to define and share a common understanding of the information structure of a domain. This concerns domain-specific knowledge but is also related to the specific context of application of the ontology. In our case, this is as open as possible, i.e., we would like to design the ontology as a basis for meaningful query scenarios of our own data but also data pertaining to literary history in other projects. We do not want to prescribe research questions or directions but aim to provide a freely available knowledge base that is as open as possible to different methods, interests and research contexts.

As the definition of the elements that form the knowledge network determines what can be queried later, constructing and querying are directly related. The development of ontologies, defined as an "explicit specification of a conceptualization" (Gruber 1995), necessitates the formulation of so-called "competency questions" that help define the domain's scope (Noy & McGuinness 2001). Concerning our domain of literary history in general and the French Enlightenment novel in particular, these competency questions are, for example:

- What are the most common themes in novels of a given period?
- To what extent can we identify connections between narrative locations and themes in the French Enlightenment novel?
- Do developments in the book market and trade routes reflect changes in the most frequent places of publication?
- How are certain authors evaluated in the scholarly literature and do these evaluations change over time?
- Can specific changes (in themes, narrative locations, or tonalities of the novels) be detected since the French Revolution or since other striking events?

Basic elements — usually referred to as concepts and relations between them — play a role in answering these questions. An essential stage in developing the ontology involves determining pertinent concepts (respectively "classes" as types of "concepts") and their relations (respectively "properties").

Knowledge in the humanities consists less of facts and more of perspectives, as we have contextualized above (see 2.2.1). It is this particularity of humanities data that makes the modeling of statements about statements particularly important, so that information about the source, reliability, or status of specific statements can be stored along with each statement.

Logical level: Representing statement types and concrete statements

In this regard, the ontological level must be taken into account. Within the context of the Semantic Web, the Resource Description Framework (RDF) has emerged as a standard method for expressing simple statements in the form of "subject-predicate-object" triples (Hitzler, Krötzsch & Rudolph 2009; Dengel 2012). These triples consist of two nodes (concepts) linked by an edge (relation). The subject and predicate are consistently represented using a uniform resource identifier (URI), while the object position can accommodate either a resource (URI) or a value. The triple structure enables statements to serve as subjects in additional triples, leading to the creation of statements about statements, a process known as "reification" (Hernández, Hogan & Krötzsch 2015: 33).

The atomization within the ontology development involves identifying the smallest units making up our knowledge base. These are defined and conceptualized in a modular ontology, in which each module covers a specific part of our domain (like themes or places) and cross-domain issues (like a kind of reference module which defines how a reference of a statement is stored as a meta-statement). Insofar as ontologies are "shared conceptualizations" (Gruber 1995) and reuse plays an important role in the development of ontologies, we consider the thirteen modules of the ontology merely as a beginning and rather a proposal for a modular, extensible ontology.[6] As mentioned earlier, our domain is the French novel of the second half of the 18th century. Some modules are relatively specific to the domain of literary history, such as the one regarding narrative form. Other modules, such as the referencing module, function as transdisciplinary or cross-domain modules. Therefore, we consider our approach of atomization and the way to construct a Linked Open Data knowledge network to be transferable to other domains and disciplines.

Beyond the ontological level, interdependencies between ontological and instance data level are equally significant. One must take into account that the idealized notion of ontology development as a representation of a domain is only conditionally true. Ultimately, the ontology also and primarily serves to structure the data in such a way that they form a queryable knowledge graph. Those elements that have been obtained through the various information extraction processes (*mining*) as well as manual annotations are included into the knowledge network following the definition of statement types (*modeling*).

[6] See: https://github.com/MiMoText/ontology.

An essential dimension in the operationalization is the intertwining of mining and modeling: step by step, different methods were used to extract information from the three different source types and the resulting data was stored as RDF triples. The conceptual structure — defining what triples are possible — is represented in thirteen modules of an ontology. The central added value of the graph lies in the combination of the various methods and the data obtained from them. Even if well-established methods of text mining (such as topic modeling, named entity recognition, etc.) are used, the innovation lies in the combination of these methods for extracting information from heterogeneous sources. These are modeled in such a way that data extracted from one source type can be compared to data extracted from another source type, e.g. thematic statements generated via Topic Modeling in comparison to thematic keywords extracted from the bibliography (Martin, Mylne & Frautschi 1977). By linking the elements generated in this way, a knowledge network on the French novel of the 18th century is incrementally growing, with a density of information that has not existed before.

As evident from the aforementioned competency questions, the primary class in our knowledge domain are literary works. Instances of this class occupy the subject position of various statements that can be formulated as RDF triples, as illustrated in Figure 3.

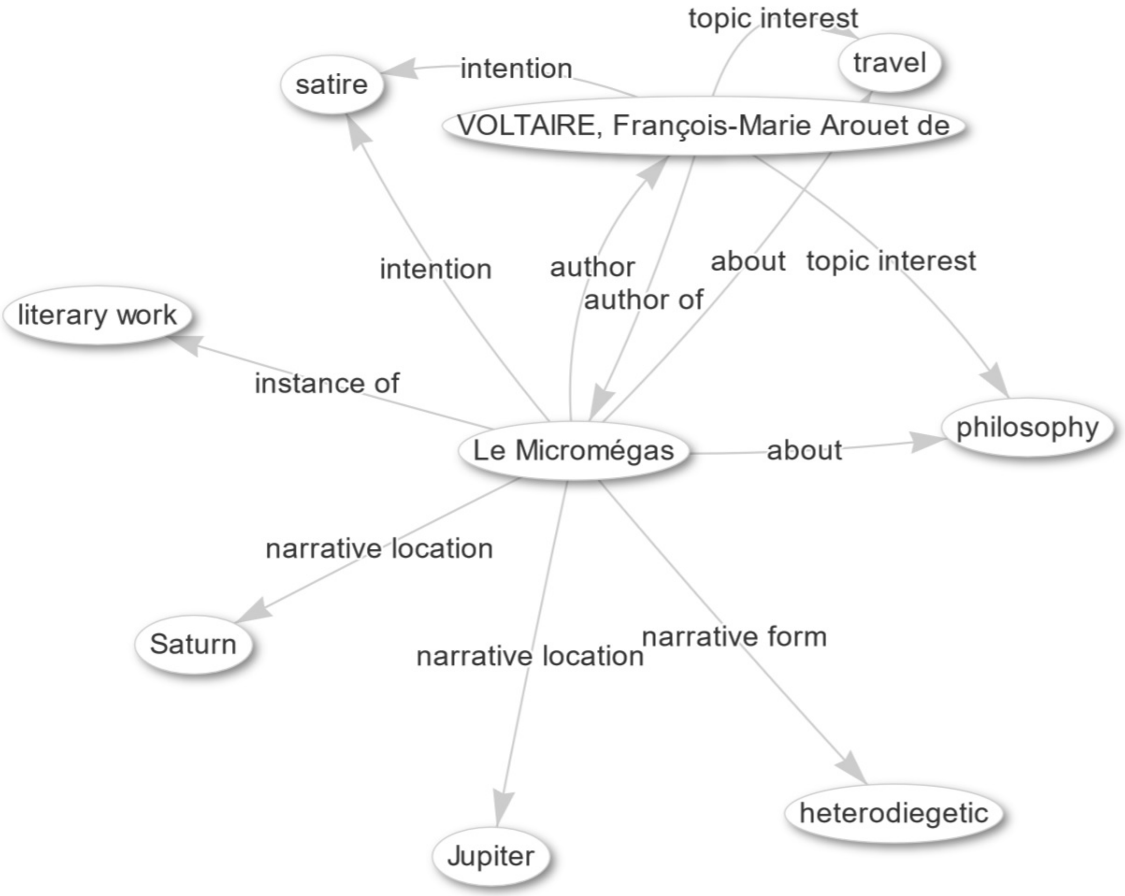


Figure 3: Authors (here: Voltaire) and works (here: *Le Micromégas*) in subject position. Query 0: https://purl.org/mmt/patterns/query0.

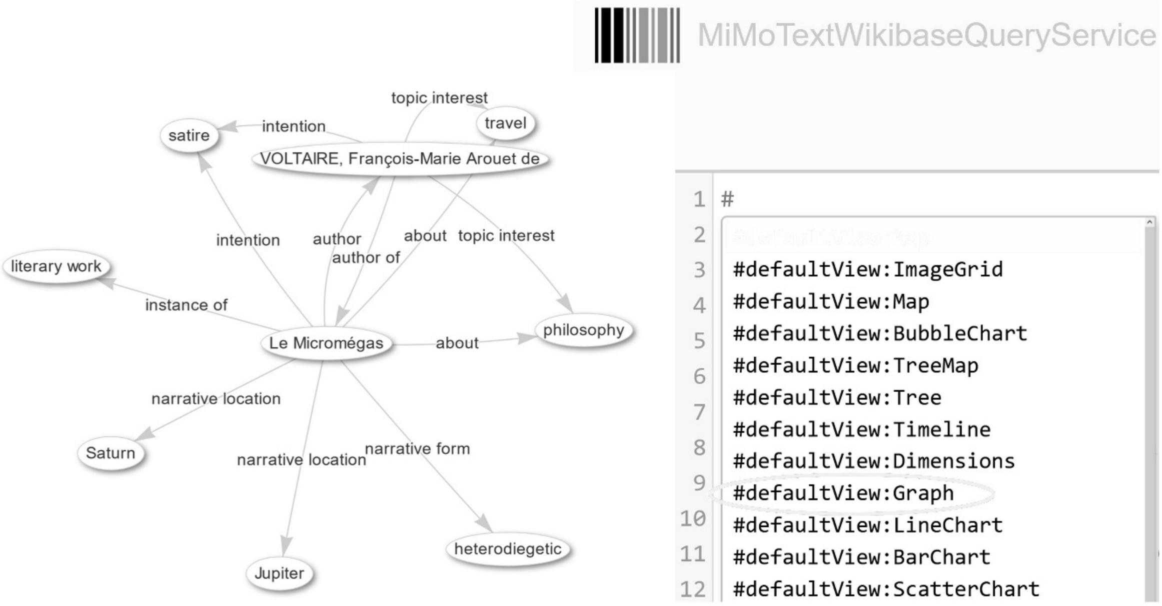
3 SPARQL: a query language for exploring and retrieving patterns

Patterns do not only play a decisive role in ontology design and the modeling of a knowledge graph but also — in the other direction, so to speak — in querying: What we have fed into the knowledge network as RDF triples can be queried via the RDF query language SPARQL (SPARQL Protocol and RDF Query Language). One refers to graph patterns of queries as so-called query patterns. The simplest patterns are triple patterns. More complex graph patterns can be formed by combining several smaller patterns. In turn, several basic graph patterns — defined as sets of triple patterns — can be combined to group graph patterns.[7] In the context of Semantic Web technologies, patterns can emerge from linking, querying and modeling data, with "graph patterns" being particularly vital for information retrieval from SPARQL-enabled knowledge bases. Unbehauen et al. (2013)

[7] See W3C SPARQL Working Group (2013) and Harris and Seaborn (2013) for the terminology and standard and Arenas et al. (2010) as well as DuCharme (2013) for further explanations.

describe the graph pattern of a query as the fundamental concept of SPARQL, defining the part of the RDF graph utilized in generating the query result.

At this point, showcasing the concrete possibilities that may arise via query patterns for literary scholars interacting with our MiMoText knowledge base seems more important than developing a typology of these patterns. Therefore, in the following we will (1.) illustrate some basic triple patterns and possible combinations, (2.) present further analysis and exploration possibilities and (3.) exemplify the potential of federated queries (in combination with further query patterns). In all three subsections, it becomes clear that numerous visualization options provided by the Wikibase framework, more precisely via the Wikidata Query Service[8], allow exploration and analysis of data at different levels of granularity (Fig. 4).



```
#
#defaultView:Map
#defaultView:ImageGrid
#defaultView:Map
#defaultView:BubbleChart
#defaultView:TreeMap
#defaultView:Tree
#defaultView:Timeline
#defaultView:Dimensions
#defaultView:Graph
#defaultView:LineChart
#defaultView:BarChart
#defaultView:ScatterChart
```

Figure 4: Overview over the default views of the Wikidata Query Service, here: “Graph”.

3.1 Basic triple patterns and their combinations

A simple triple pattern would be the narrative location of a literary work (?item mmdt:P32 ?narrLoc), where we could additionally display the labels of both the work and the narrative

[8] See https://wikitech.wikimedia.org/wiki/Wikidata_Query_Service.The name “Wikidata Query Service” is confusing, as this service is provided for all Wikibase instances and should therefore be called “Wikibase Query Service”. The usage has been inconsistent so far, we follow the current official name “Wikidata Query Service”.

location.[9] In our current dataset, such a query results in about 1824 triples about narrative locations of novels published between 1751 and 1800. Based on the same pattern, we can query a subset of all the novels, for example, only the novels set in imaginary places. The spatial concept for imaginary place has the identifier Q3371, which leads to the following triple pattern: ?item mmdt:P32 md:Q3371.[10] Similarly, we can query the themes of literary works and find a subset of texts that have, for example, *miracle* as a theme.[11] In the same way, more and more triple patterns can be combined, e.g. with a query on novels published in Paris that have *philosophy* as a theme and were first published between 1780 and 1790.[12]

3.2 Further analysis and exploration

Since it is possible to combine as many triple patterns (and graph patterns as sets of them) as needed, we can write increasingly complex queries. With query 6, we can get an overview of the novels within the MiMoTextBase.[13] However, using the Wikidata Query Service interface offers many additional functions that are useful for analysis and exploration.

With simple count queries, an overview of the most common topics of e.g. satirical novels, can be obtained using the #defaultView:BubbleChart (see Fig. 5).[14] With a more complex query that integrates a group pattern, the development of a particular topic/theme (e.g. *nature*) can be examined in the light of historical variation.[15]

[9] See query 2: https://purl.org/mmt/patterns/query2. All queries are documented under https://mimotext.github.io/MiMoTextBase_Tutorial/queries_patterns as well as via PURL.org (see Appendix).

[10] See query 3: https://purl.org/mmt/patterns/query3.

[11] See query 4: https://purl.org/mmt/patterns/query4.

[12] See query 5: https://purl.org/mmt/patterns/query5.

[13] See query 6: https://purl.org/mmt/patterns/query6.

[14] See query 7: https://purl.org/mmt/patterns/query7.

[15] See query 8: https://purl.org/mmt/patterns/query8.

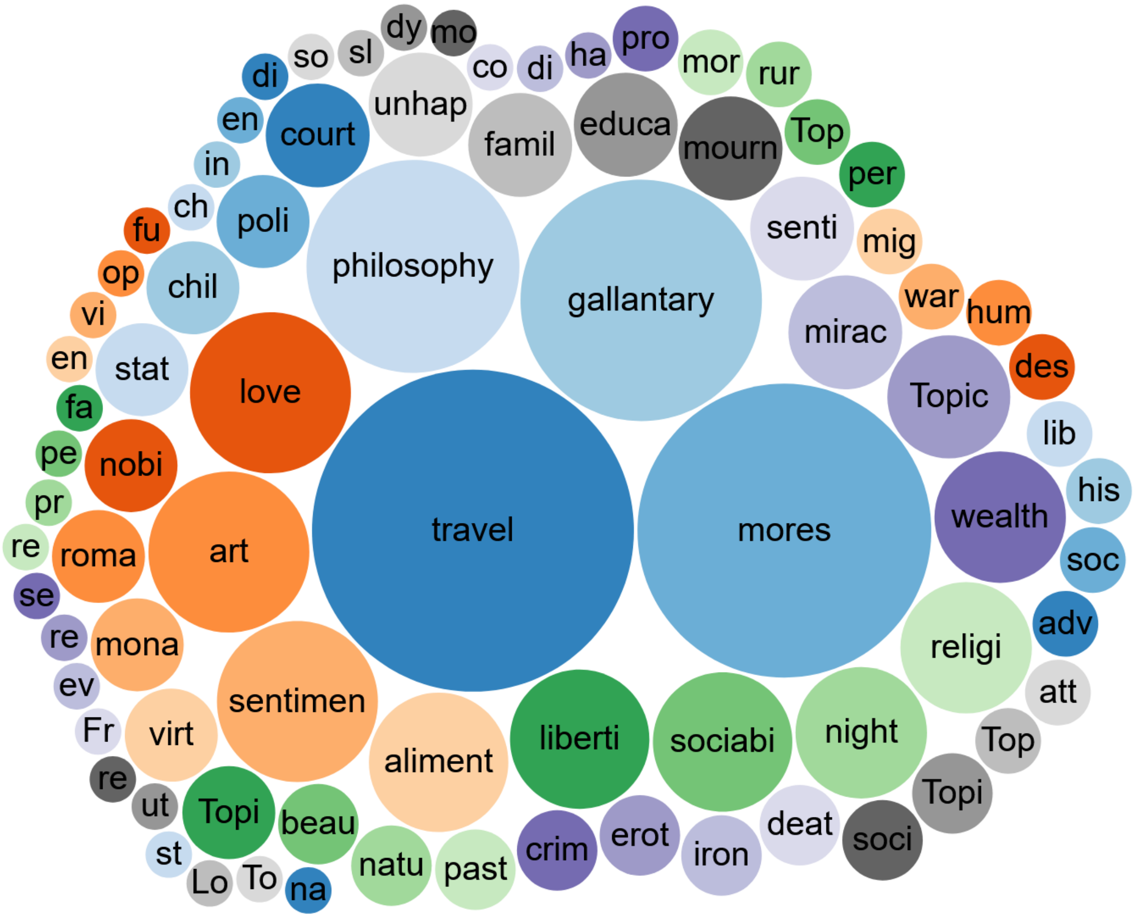


Figure 5: Overview of topics of satirical novels 1751–1800. Query 7: https://purl.org/mmt/patterns/query7.

A query pattern specific to the Wikibase / Wikidata data model could be called "referencing pattern".[16] From the precise referencing of the source for each individual triple, interesting possibilities for comparison arise. For example, a subset of triples can be filtered by source type, e.g. only the thematic statements that are documented by topic modeling.[17] These can then be examined more closely in comparison to the (e.g. thematic) statements that the bibliographers have annotated. In addition, the referencing pattern can be used to focus on only those triples that are supported by both sources.[18] There is likely to be only a relatively small number of such triples so that, depending on the researcher's interest, they could be a starting point for further analysis (Fig. 6).

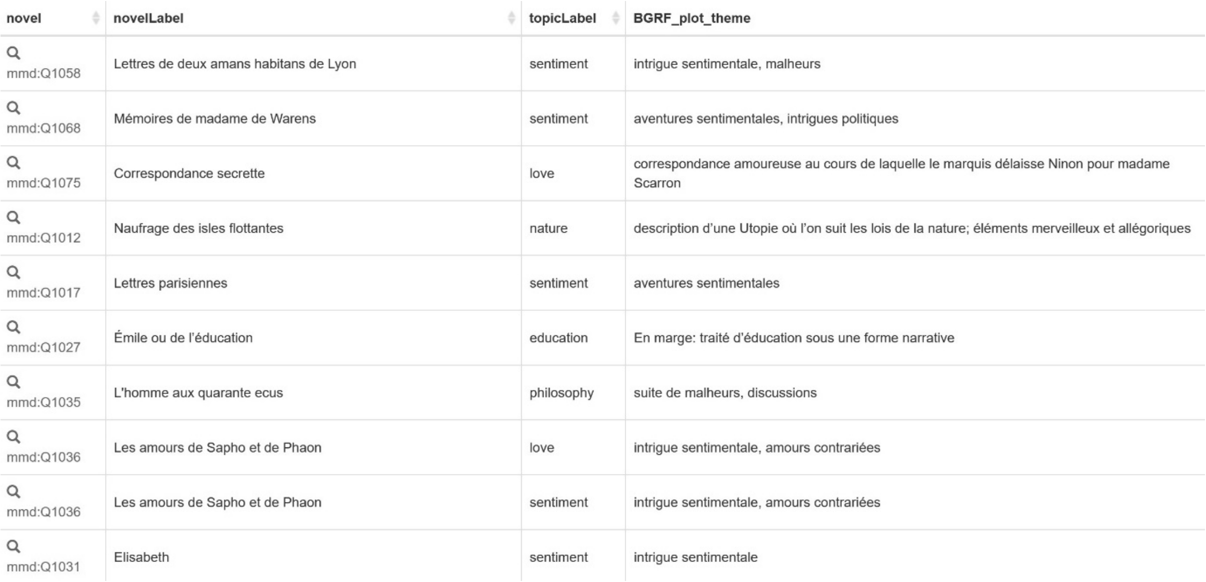
[16] In our referencing module, we model the different statement types used to specify references both at the claim level (first level triples) and at the reification level (meta triples), reusing Wikidata properties. See: https://github.com/MiMoText/ontology/tree/main/module7_referencing.

[17] See query 9: https://purl.org/mmt/patterns/query9.

[18] See query 10: https://purl.org/mmt/patterns/query10.

As we used different sources in the graph, we created several controlled vocabularies to standardize the information in order to provide comparable and queryable items.[19] As of January 2024, we use 1136 items that form part of five different vocabularies for which we have a total of 876 “exact” or “close” matches to Wikidata entries.[20] That enables researchers to obtain overviews on various topics in a comparable way, as in query 12, which gives an outline of the development of preferred combinations of narrative form and dominant intention (a category derived from the *Bibliographie du genre romanesque français*) over time.[21]

Being a multilingual knowledge graph, the labels on the MiMoTextBase can be queried directly in three languages, English, German and French.[22] However, due to the mapping of the concepts to Wikidata items, it is also possible to query labels in other languages, e.g. labels in all languages entered for Wikidata-item *Voltaire*.[23] For this purpose we can use federated queries, which will be explained in the next section.



novel	novelLabel	topicLabel	BGRF_plot_theme
mmd:Q1058	Lettres de deux amans habitans de Lyon	sentiment	intrigue sentimentale, malheurs
mmd:Q1068	Mémoires de madame de Warens	sentiment	aventures sentimentales, intrigues politiques
mmd:Q1075	Correspondance secrette	love	correspondance amoureuse au cours de laquelle le marquis délaisse Ninon pour madame Scarron
mmd:Q1012	Naufrage des isles flottantes	nature	description d’une Utopie où l’on suit les lois de la nature; éléments merveilleux et allégoriques
mmd:Q1017	Lettres parisiennes	sentiment	aventures sentimentales
mmd:Q1027	Émile ou de l’éducation	education	En marge: traité d’éducation sous une forme narrative
mmd:Q1035	L’homme aux quarante ecus	philosophy	suite de malheurs, discussions
mmd:Q1036	Les amours de Sapho et de Phaon	love	intrigue sentimentale, amours contrariées
mmd:Q1036	Les amours de Sapho et de Phaon	sentiment	intrigue sentimentale, amours contrariées
mmd:Q1031	Elisabeth	sentiment	intrigue sentimentale

Figure 6: Thematic statements referenced by topic modeling and the bibliography. Query 10: https://purl.org/mmt/patterns/query10.

[19] For more information, see https://github.com/MiMoText/vocabularies.

[20] See query 11 https://purl.org/mmt/patterns/query11.

[21] See query 12: https://purl.org/mmt/patterns/query12.

[22] The queries 13 (https://purl.org/mmt/patterns/query13) and 14 (https://purl.org/mmt/patterns/query14) will get the same result.

[23] See query 15: https://purl.org/mmt/patterns/query15.

3.3 Federated queries

Considering that every knowledge graph is using its own formalism and its own patterns — even though there is a common ambition to reuse existing ontologies if possible — how can knowledge graphs interact with each other? How is querying and reasoning across several knowledge graphs possible?

In the case of the MiMoText knowledge graph and the Wikidata knowledge graph, the crucial point is that the link between both knowledge graphs was made explicit on the level of the individual items that are in the object or subject position concerning certain categories (themes, locations, authors, works). This means that statements were added in our graph that link these concepts to the Wikidata graph via the property "exact match" (P13).

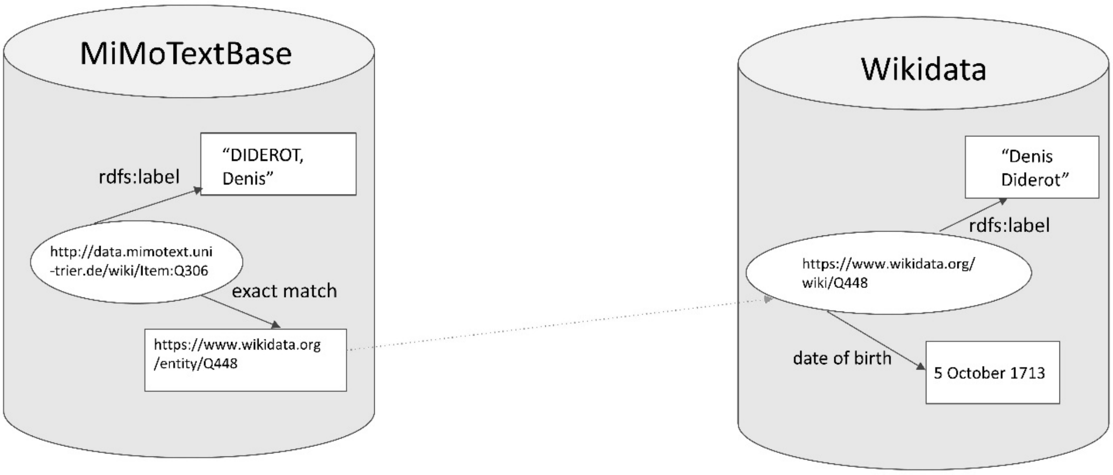

MiMoTextBase
rdfs:label
"DIDEROT, Denis"
http://data.mimotext.uni-trier.de/wiki/Item:Q306
exact match
https://www.wikidata.org/entity/Q448
Wikidata
"Denis Diderot"
rdfs:label
https://www.wikidata.org/wiki/Q448
date of birth
5 October 1713

Figure 7: Federation between two knowledge graphs: MiMoTextBase and Wikidata (inspired by Abel (2019: 5).

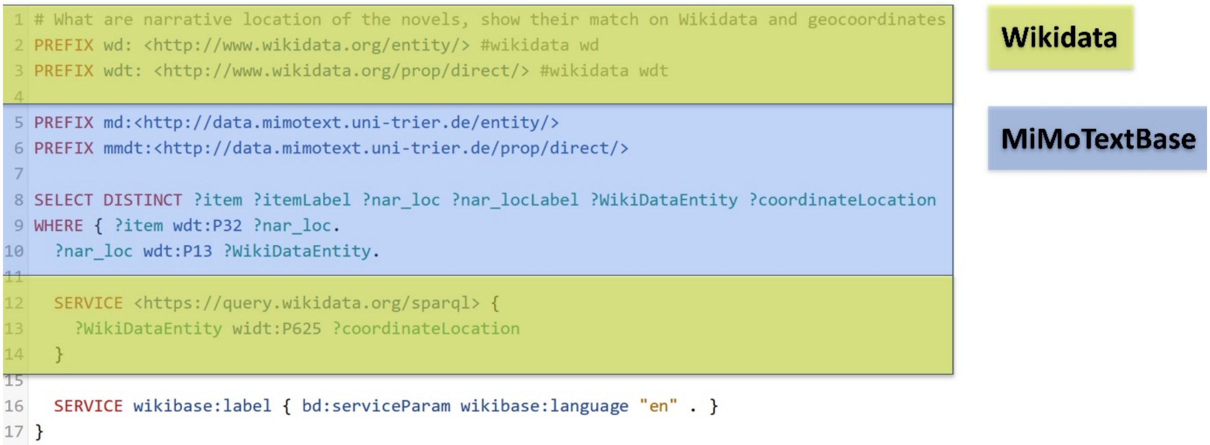
To exemplify this: The knowledge graph does have to carry the information that item Q448[24] of the Wikidata graph is equivalent to Q306[25] in the MiMoText graph, meaning that the item of the French writer, philosopher and encyclopedist Denis Diderot (Q448) on Wikidata corresponds to the item of Denis Diderot in the MiMoText-Graph (Q306). As shown in Figure

[24] See: https://www.wikidata.org/wiki/Q448.

[25] See: https://data.mimotext.uni-trier.de/wiki/Item:Q306.

7, this allows us to retrieve further information from Wikidata, for example the “date of birth” property (P569).[26]

Adding these crucial RDF-triples in either one of the two graphs (aligning the identifiers on the item-level) enables to query both graphs and their multitude of triples in so-called federated queries in SPARQL, even though their ontologies might differ in other respects. Federated queries in general allow the potential of Linked Open Data to be realized by querying data across different databases, which requires a federation infrastructure driven by the RDF and SPARQL standards (see Fig. 8).[27]



```
# What are narrative location of the novels, show their match on Wikidata and geocoordinates
PREFIX wd: <http://www.wikidata.org/entity/> #wikidata wd
PREFIX wdt: <http://www.wikidata.org/prop/direct/> #wikidata wdt

PREFIX md:<http://data.mimotext.uni-trier.de/entity/>
PREFIX mmdt:<http://data.mimotext.uni-trier.de/prop/direct/>

SELECT DISTINCT ?item ?itemLabel ?nar_loc ?nar_locLabel ?WikiDataEntity ?coordinateLocation
WHERE { ?item wdt:P32 ?nar_loc.
  ?nar_loc wdt:P13 ?WikiDataEntity.

  SERVICE <https://query.wikidata.org/sparql> {
    ?WikiDataEntity widt:P625 ?coordinateLocation
  }

  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" . }
}
```

Figure 8: Example of a federated SPARQL query syntax using Wikidata (green) and MiMoTextBase (blue).

An example for the use of federated queries is enriching our spatial concepts with geographical data from Wikidata, which is based on the matching of spatial concepts established in our spatial vocabulary.[28] These matches are part of our Wikibase as a triple pattern according to the scheme: “[spatial concept item] -> exact match -> wikidata URL”. Such a triple links the two Wikibase instances: our domain-specific MiMoText graph and the large Wikidata graph. One of the advantages of this approach is that information does not have to be stored redundantly. Instead, we can use the values of Wikidata property “coordinate location”[29] for a query on narrative locations of French novels from the second

[26] See: https://www.wikidata.org/wiki/Property:P569.

[27] See Görlitz and Staab (2011) and Prud’hommeaux and Buil-Aranda (2013) as well as https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/Federated_queries and https://www.mediawiki.org/wiki/Wikidata_Query_Service/User_Manual#Federation for specific aspects in the ‘wikiverse’.

[28] See https://github.com/MiMoText/vocabularies/blob/main/spatial_vocabulary.tsv.

[29] See: https://www.wikidata.org/wiki/Property:P625.

half of the 18th century contained in our MiMoTextBase. Using the map view provided by the Wikidata Query Service for each Wikibase instance, we can get an interactive overview with individual nodes for narrative locations, enabling users to click on them and access additional information for further exploration (see Fig. 9).[30]

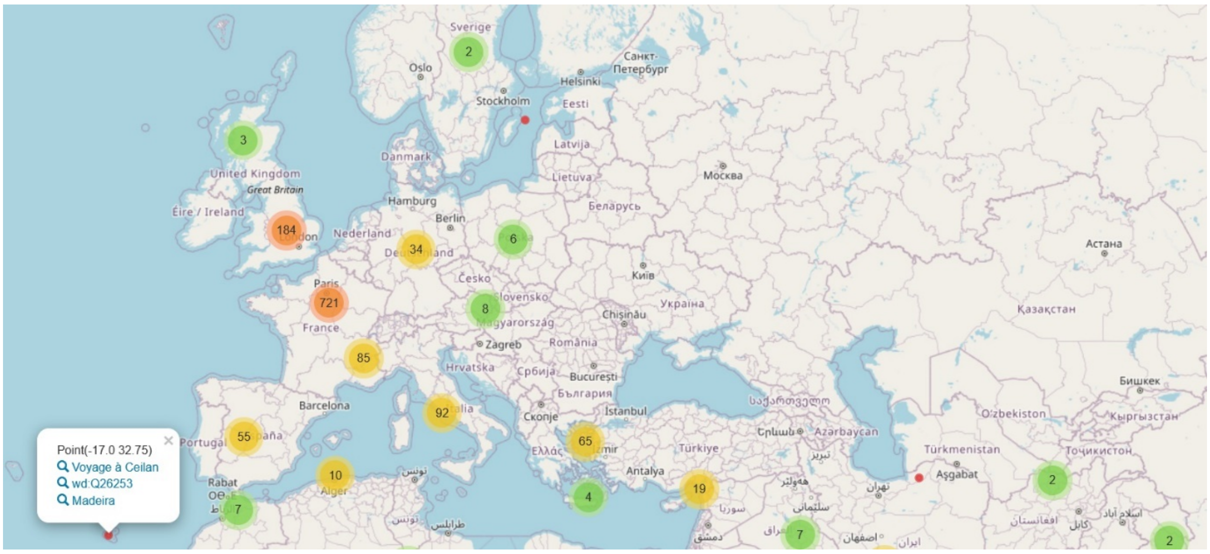


Figure 9: Overview of narrative locations in about 1700 French novels, 1750–1800. Query 16: https://purl.org/mmt/patterns/query16.

As already mentioned, we have matched not only the spatial concepts but also the author entities with Wikidata items. Via federated queries, we can retrieve useful information about alias labels or multilingual labels and use it further (e.g. for the controlled vocabulary labels in additional languages). For each Wikidata item, there are alternative labels stored in the infobox under "also known as". These are formalized in the Simple Knowledge Organization System (SKOS) standard and can be queried via the property skos:altLabel.[31] By running this query on the author item labeled as *Voltaire* we get the information that "François-Marie Arouet" and "Francois Marie Arouet de Voltaire" are designated as "alternative labels".[32] This information could be used for further analysis, such as an improvement of named entity recognition tasks.

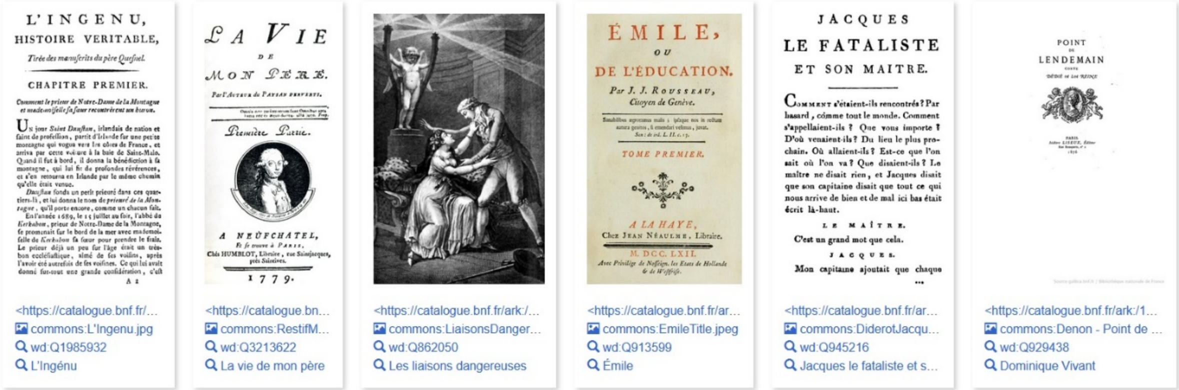
Another fruitful property of Wikidata is the collection of identifiers pointing at other knowledge bases, for example at the *Bibliothèque nationale de France* (BnF). We can

[30] See query 16: https://purl.org/mmt/patterns/query16. For more visualization options, see https://mimotext.github.io/MiMoTextBase_Tutorial/visualizations.html.

[31] See query 17: https://purl.org/mmt/patterns/query17.

[32] See: https://www.wikidata.org/wiki/Q9068.

retrieve information on the *BnF* via the Wikidata identifier by using the property "Bibliothèque nationale de France ID"[33] and "formatter URL"[34]. Query 18 exemplifies this by retrieving authority data on eighteenth-century French novels in Wikidata and by transforming the identifier (?bnfid) to a URL (?bnfurl) with the help of regular expressions (see Fig. 10).[35]



<https://catalogue.bnf.fr/...
commons:L'Ingenu.jpg
wd:Q1985932
L'Ingénu
<https://catalogue.bn...
commons:RestifM...
wd:Q3213622
La vie de mon père
<https://catalogue.bnf.fr/ark:/...
commons:LiaisonsDanger...
wd:Q862050
Les liaisons dangereuses
<https://catalogue.bnf.fr/ar...
commons:EmileTitle.jpeg
wd:Q913599
Émile
<https://catalogue.bnf.fr/ar...
commons:DiderotJacqu...
wd:Q945216
Jacques le fataliste et s...
<https://catalogue.bnf.fr/ark:/1...
commons:Denon - Point de ...
wd:Q929438
Dominique Vivant

Figure 10: MiMoText novels with identifiers on other knowledge graphs, here: Bibliothèque nationale de France (BnF) identifiers. Query 18: https://purl.org/mmt/patterns/query18.

The further potential of federated queries can be illustrated by using the property "influenced by". Researchers might be interested in influence networks. In Wikidata, relevant data is available in statements with the property "influenced by".[36] We can integrate them into a federated query. If we do not only ask for direct influences but allow for sequences of this graph pattern, then possible intermediaries connecting the subnetworks become visible. For this purpose, we can use a "property path" which queries "influenced by" multiple times in sequence (the syntax for this is wdt:P737/wdt:P737 for length two or wdt:P737/wdt:P737* for unlimited length of the path, Fig. 11).[37]

One can see in Figure 11 that work items like *Candide* (Q1022) by Voltaire or *Jacques le Fataliste* (Q1088) by Denis Diderot have a lot of connecting relations and are placed in central positions.[38]

[33] See: https://www.wikidata.org/wiki/Property:P268.
[34] See: https://www.wikidata.org/wiki/Property:P1630.
[35] See query 18: https://purl.org/mmt/patterns/query18.
[36] See https://www.wikidata.org/wiki/Property:P737.
[37] See query 19: https://purl.org/mmt/patterns/query19, as well as e.g. https://www.w3.org/TR/sparql11-property-paths/ regarding property paths.
[38] See: https://data.mimotext.uni-trier.de/wiki/Item:Q1022 and https://data.mimotext.uni-trier.de/wiki/Item:Q1088.

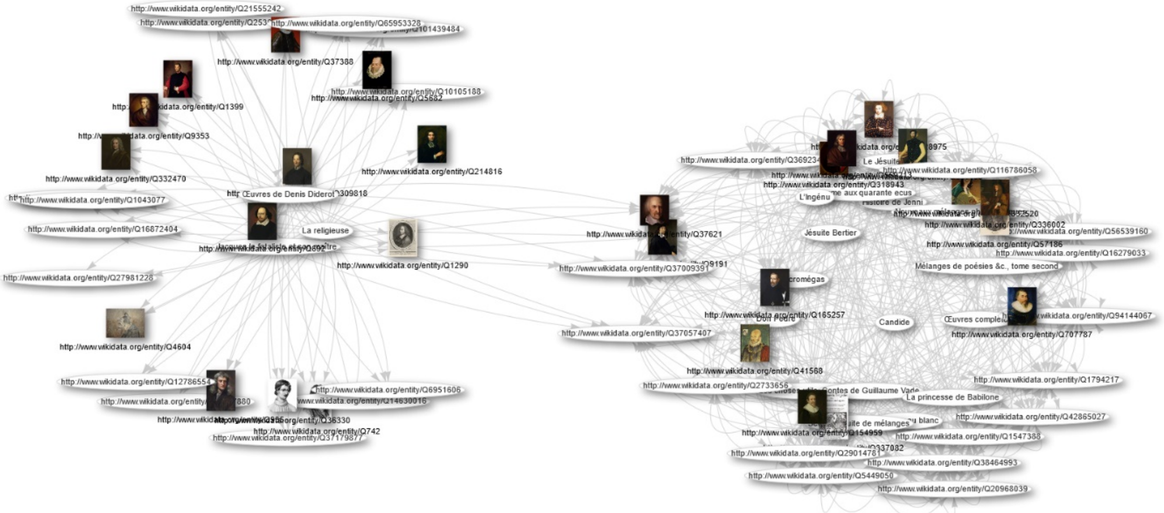

http://www.wikidata.org/entity/Q21555242
http://www.wikidata.org/entity/Q65953328
http://www.wikidata.org/entity/Q37388
http://www.wikidata.org/entity/Q10105188
http://www.wikidata.org/entity/Q5682
kidata.org/entity/Q1399
data.org/entity/Q9353
http://www.wikidata.org/entity/Q214816
http://www.wikidata.org/entity/Q332470
Œuvres de Denis Diderot
http://www.wikidata.org/entity/Q1043077
http://www.wikidata.org/entity/Q16872404
La religieuse
http://www.wikidata.org/entity/Q1290
http://www.wikidata.org/entity/Q27981228
http://www.wikidata.org/entity/Q4604
http://www.wikidata.org/entity/Q12786554
http://www.wikidata.org/entity/Q6951606
www.wikidata.org/entity/Q14630016
org/entity/Q36330
http://www.wikidata.org/entity/Q742
http://www.wikidata.org/entity/Q37179877
http://www.wikidata.org/entity/Q36923
Le Jésuite
http://www.wikidata.org/entity/Q116786058
http://www.wikidata.org/entity/Q318943
L'Ingénu
Histoire de Jenni
http://www.wikidata.org/entity/Q37621
http://www.wikidata.org/entity/Q336002
Jésuite Bertier
http://www.wikidata.org/entity/Q56539160
http://www.wikidata.org/entity/Q57186
http://www.wikidata.org/entity/Q16279033
http://www.wikidata.org/entity/Q37009391
Mélanges de poésies &c., tome second
http://www.wikidata.org/entity/Q165257
Candide
wikidata.org/entity/Q94144067
http://www.wikidata.org/entity/Q37057407
http://www.wikidata.org/entity/Q707787
http://www.wikidata.org/entity/Q41568
http://www.wikidata.org/entity/Q1794217
http://www.wikidata.org/entity/Q2733656
Contes de Guillaume Vade
La princesse de Babilone
http://www.wikidata.org/entity/Q42865027
http://www.wikidata.org/entity/Q154959
http://www.wikidata.org/entity/Q1547388
http://www.wikidata.org/entity/Q29014781
http://www.wikidata.org/entity/Q38464993
http://www.wikidata.org/entity/Q5449050
http://www.wikidata.org/entity/Q20968039

Figure 11: Influence networks of authors via federated query based on Wikidata matches. Query 19: https://purl.org/mmt/patterns/query19.

Property paths are also a very useful way to explore how data is modeled in detail or which subclasses exist. For example, the path P31/P279* can be used on Wikidata to get an overview of instances/subclasses* of novels. However, the concrete example also shows that for the interpretation of the data, it is important to be able to understand the patterns. In a corresponding count query, *penny dreadful* surprisingly ranks second.[39] The paths of the graph partly lead off the beaten track and, above all, show that it is important to take into account the social dimensions that (can) cause biases in the data.

In part, this is due to over- or underrepresentations in Wikidata created by particularly active or invisible communities. This must be taken into account as a facet of analyses based on this data. Solutions for dealing with non-uniformly modeled data can only be found if one knows about the non-uniformities. For example, the Alternative Graph Pattern can be used to merge different query patterns if, for example, two different properties were used that actually conceptually represent one property.[40] This highlights the importance of transparency in both modeling and querying patterns, shows how strongly

[39] See for the query on Wikidata https://w.wiki/7DJd and https://www.wikidata.org/wiki/Q3374808 for the item labeled “penny dreadful”.

[40] See Prud’hommeaux and Seaborne (2008) and for example the alternative “wdt:P577|P571” which are often used in the same way (P577=publication date; P571=inception).

the two are interrelated, and reminds us of the importance of considering the social aspects of data production.

4 Infrastructures and patterns

Federated queries across multiple databases, as illustrated in the previous section, can best be enabled, and more generally a landscape of disconnected data silos only be avoided, if the issue of technical infrastructure (software and fundamental data model) and its implications are carefully considered. Therefore, we would like to discuss in the following (1.) our decisions in the context of Open Science and of the understanding of MiMoTextBase as part of the Wikibase ecosystem and, (2.) ontology design patterns as a way to overcome infrastructural boundaries and (3.) modeling and patterns in graph databases.

4.1 Infrastructure in MiMoText and the MiMoTextBase as part of the Wikibase ecosystem

The publication of FAIR data and the use of open source tools according to Open Science principles guide our project as a whole (Röttgermann & Schöch 2020). In particular, we decided to rely on a Wikibase instance as our project infrastructure with an adapted PyWikibot for automated imports.[41] Wikibase is an open source software developed by the Wikimedia Foundation. The largest instance of the Wikibase software is the Wikidata Knowledge Graph (Diefenbach, De Wilde & Alipio 2021: 1). Using Wikibase rather than alternative software solutions has implications for the data model and thus for questions of reusability.

We chose Wikibase as an infrastructure for the project, not only because it is a free and open software but also because it is especially suited for multilingual data. Wikibase can be customized to meet specific data management and ontology design needs, making it a flexible and adaptable tool. The integrated Wikidata Query Service (WDQS) provides a SPARQL endpoint and comes along with several built-in visualization options, which facilitate plotting data patterns in various ways, as illustrated in the previous section. The Wikibase framework can be seen as a way to overcome “many obstacles to a persistent, transparent, and reusable resource” (Eells et al. 2021: 11). It is noteworthy that our MiMoText knowledge graph as an RDF graph database can be queried independently of

[41] See the repository for our customized WikibaseBot: https://github.com/MiMoText/wikibase-bot. In addition, we also used the tool “QuickStatements” written by Magnus Manske, see: https://www.wikidata.org/wiki/Help:QuickStatements.

Wikibase or the Wikidata Query Service.[42] We have released an RDF dump on our project website that can be downloaded and imported into other SPARQL query services such as Virtuoso SPARQL and GraphDB.[43]

A common criticism is that in Wikiverse (i.e. in Wikidata or in project-specific Wikibase instances), there is no ontological distinction between classes and instance data. Also, both levels are (to the regret of stricter ontologists) not systematically separated within the Wikibase/Wikidata model (Q-Items stand for classes as well as instance data) and there is a considerable number of different property types which have a different logic than property types in the Web Ontology Language (OWL). As a consequence, Wikidata statements cannot easily be represented in OWL, which is the common W3C standard. Nevertheless, they can be represented as an ontology. In our eyes, both perspectives are true: Wikidata is of great importance as a "linking hub" (Neubert 2017) for linking humanities data across project boundaries. The simple alignment of single entities across knowledge bases, which already enables federated queries, is an important step and already extremely useful, even without exhausting all the possibilities envisioned in the Semantic Web (of reasoning and inferencing etc.). At the same time, the criticism of a lack of systematic ontology and formal semantics is justified (Sack 2022). Much remains to be done in terms of alignment efforts, as Eells et al. (2021: 11) sum up, but such alignments can be reused. We see MiMoTextBase as part of the growing Wikibase ecosystem. The more projects in this ecosystem share the same infrastructure, the denser and more significant the whole graph becomes. With the Wikibase infrastructure, we are embedded in a larger framework and share the data model with Wikidata as the largest public Wikibase instance. Federated queries are of course possible across different types of infrastructures but are clearly made easier by a common basic data model (e.g. modeling of reifications).

4.2 Ontology Design Patterns as an interoperability bridge

To enable or strengthen interoperability and reusability within the Semantic Web, it is important that certain elements of an ontology are grouped as modules and that there are bridges between different infrastructures. The Wikibase data model is characterized by considerable complexity resulting from the fact that there is a multi-layered reification

[42] A comparison between a classical RDF infrastructure and Wikibase can be found in Diefenbach et al. (2021: 14).

[43] See https://vos.openlinksw.com/owiki/wiki/VOS/VOSSPARQL and https://www.ontotext.com/products/graphdb, respectively.

system. This reification system supports statements about statements or meta-triples that store, for example, the probability, ranking, time-frame and/or source of a statement. In addition, the representation of multilingualism and the management of community participation (with differentiated, group-based user rights, etc.) both play an important role in this infrastructure. In this respect, the data model has requirements and possibilities but also a structure that are different from those of an ontology modeled in the W3C standard OWL. The first and second point are key arguments that led us to prefer Wikibase over an OWL-based technical infrastructure.[44]

Ontology Design Patterns (ODP) play a central role in this context because they provide “semantic interoperability [...] without restricting heterogeneity” (Janowicz et al. 2016). In the early days of the Semantic Web, the promise that one could do federated queries across several query endpoints turned out to be difficult to realize due to a lack of shared vocabularies and alignments (cf. Janowicz et al. 2016). ODPs were a direct response to this problem. They served as "reusable solutions to frequently occurring ontology design problems" (Shimizu, Hammar & Hitzler 2022b: 10).[45]

There are different requirements for and types of ODPs, among them the so-called Content ODPs (knowledge patterns), which are “typically modeled for frequently recurring aspects of more complex ontologies and thus act as building blocks” that are “not limited to domain-specific cases” (Janowicz et al. 2016: 2).

A certain complexity also arises when one tries to harmonize an ontology oriented to the Wikibase model with the semantically more expressive Web Ontology Language (OWL). In their article *Aligning Patterns to the Wikibase Model*, Eells et al. (2021) propose a small library of patterns that provide a link between a traditional ontology design pattern and the underlying Wikibase data model. The authors demonstrate that such an alignment is possible but do not hide the fact that it was “not as straightforward as [...] expected” (Eells et al. 2021: 2). At the same time, the proposed “library of ontology design patterns that have been specifically engineered to explicitly represent how Wikibase models data ‘under-

[44] In practice, an ontology modeled for example in the ontology editor Protégé cannot be easily imported (i.e. not without the still relatively large effort associated with ‘alignments’) into a Wikibase. The very different data types of the properties play an important role here, i.e. the different approaches to model the position of the predicate in an RDF triple (between subject and object).

[45] Shimizu et al. (2022b: 10) aim to “reimagine ontology design patterns and their use”. Shimizu et al. (2022a) try to bridge the gap via “axioms”.

the-hood'" (Eells et al. 2021: 2) can potentially serve as a useful kind of interoperability bridge.

An example of such a pattern is the case "Quantity as Qualifier". Figure 12 shows the quantity pattern in Wikibase. The abbreviated version on the left (a) shows how a wd:Entity is linked to a xsd:decimal value via a property. The expanded version (b) shows how under the hood lie quasi reifications that support quality assurance measures like rankings of contradictory statements (wikibase:Statement, wikibase:BestRank or wikibase:NormalRank) or another hashed node as qualifier.

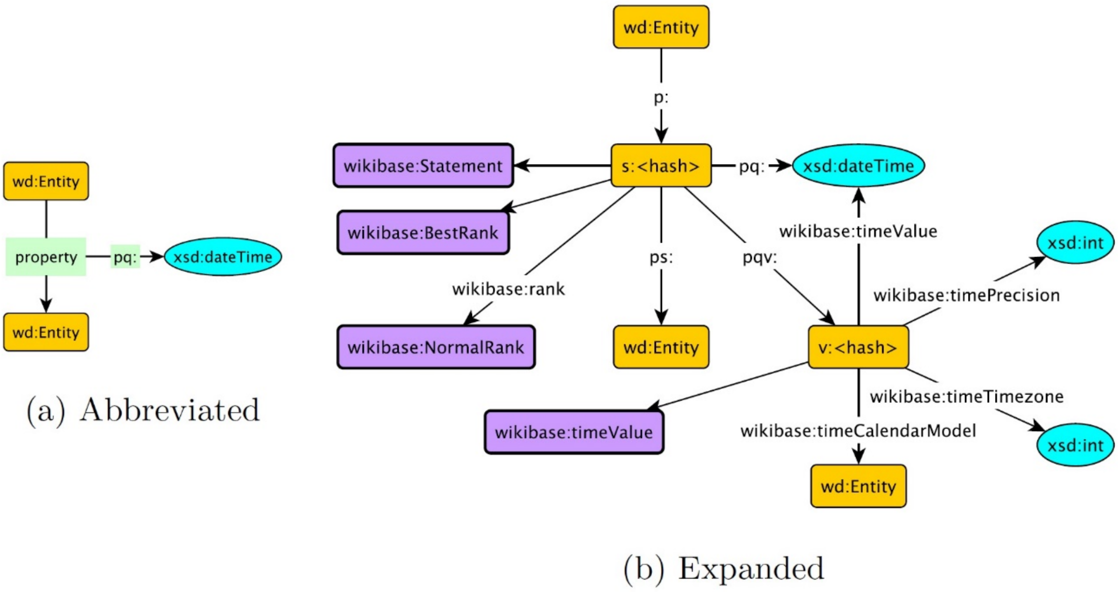


Figure 12: Aligned Pattern for "Quantity as Qualifier" (Eells et al. 2021, Fig. 8).

4.3 Modeling and patterns in graph databases

The potentials of the Linked Open Data universe — openness, flexibility, extensibility and possibly also community participation — have encountered various hurdles which have, so far, stopped the paradigm from experiencing a full breakthrough (Hooland & Verborgh 2014: 110). An example of such challenges are the difficulties around the "alignment" of authority data encountered in the project "DNB goes Wikibase".[46]

Infrastructures and data models are closely related, as we have already illustrated with the Ontology Design Patterns. In the following, we would like to increase awareness of the fact that comparable bridges are relevant for more interoperability and reusability on

[46] DNB is the abbreviation for Deutsche Nationalbibliothek (German National Library). In a joint effort, the bibliography's authority file records are made available within Wikibase structures. The process is described in Fischer (2022: 283–290).

the Semantic Web. These are not exclusively at the ontology level but are also related to the nature of the graph database. Concerning the realization of graph models, a multitude of options are currently available (Donkers, Yang & Baken 2020: 25) with labeled property graphs, RDF and RDF* being of particular relevance to the field of Digital Humanities.

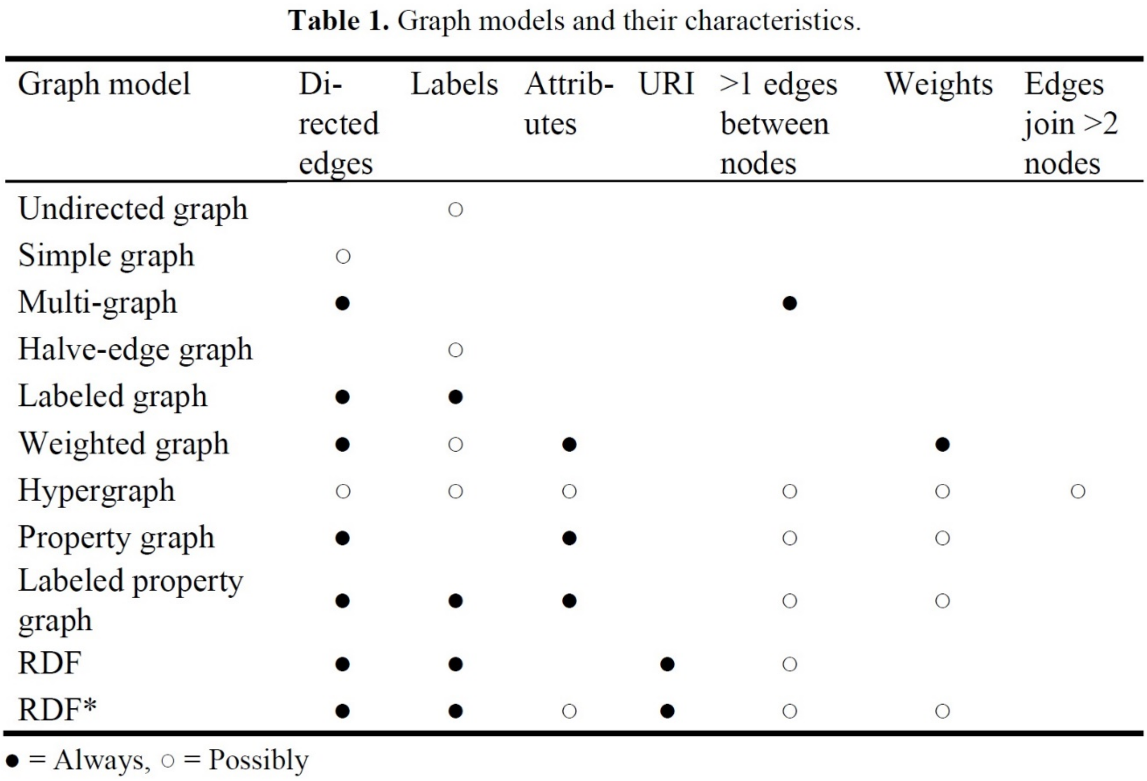


Table 1. Graph models and their characteristics.

Graph model	Directed edges	Labels	Attributes	URI	>1 edges between nodes	Weights	Edges join >2 nodes
Undirected graph		○					
Simple graph	○						
Multi-graph	●				●		
Halve-edge graph		○					
Labeled graph	●	●					
Weighted graph	●	○	●			●	
Hypergraph	○	○	○		○	○	○
Property graph	●		●		○	○	
Labeled property graph	●	●	●		○	○	
RDF	●	●		●	○		
RDF*	●	●	○	●	○	○	

● = Always, ○ = Possibly

Figure 13: Graph models according to Donkers et al. (2020: 25)

While labeled property graphs and RDF share the features of directed edges and labels, they differ in using attributes and weights. Different graph model patterns entail different query patterns (e.g. Cypher vs. SPARQL), so that the realization of a complete transformation of the HTML-based web into a Semantic Web as envisioned by Tim Berners-Lee (2006) seems utopian, given the current state of infrastructural patterns. Such non-matching patterns (graph models, query syntax) hinder the realization of overarching queries between various project-specific knowledge bases.

For humanities data, particular challenges arise in modeling more complex structures as meta-statements (perspectivity, reliability and sources/referencing, etc.). In various projects implementing graph databases, the relevance of meta-statements (e.g. source citations) is emphasized (Alassi 2023; Ammann, Alassi & Rosenthaler 2023; Baillie et al. 2021). However, each new project appears to develop solutions for this problem that may already exist in comparable projects.

It is difficult to assess to what extent certain infrastructures or data models will dominate in the future (e.g., RDF* or the Wikibase ecosystem). However, due to the heterogeneity of humanities projects with respect to the disciplines involved and the corresponding kinds of data, there will be no “one size fits all” solution, which makes the discussion about appropriate “bridging” strategies all the more important. The concept of patterns concerned with decomposing domains or models into elements and assembling units that connect several elements, as well as managing the transfer between different models, could play an even more important role here.

5 Conclusion

The paper illustrates that patterns play an important role in the construction and use of a literary history knowledge graph in particular and the Linked Open Data vision in general. This concerns the atomization of a domain, the linking of atomized elements into a queryable knowledge network and ways to strengthen interoperability and reuse through specific patterns.

A data-rich literary history and access to data via a knowledge graph have the potential to bring to the surface and make explorable a variety of aspects of the domain that have so far remained hidden due to various reasons, such as canonization processes, highly specialized vocabularies and heterogeneous sources. Infrastructure and associated substandards play a significant role in linking triples on the Semantic Web, and so far this has been a considerable constraint on the realization of the Linked Open Data paradigm.

As of 2023, the rapid developments in the field of Large Language Models (LLM) such as GPT4 have implications for the conception of a Semantic Web or the future viability of this vision. It remains to be seen whether LLMs will soon make knowledge graphs obsolete, whether they could facilitate interaction with knowledge graphs, or whether knowledge graphs, when integrated into LLMs, could improve their quality and be an important contribution to the goal of enabling algorithmic “reasoning”.

Contributions

Maria Hinzmann: writing – original draft, methodology, conceptualization, project administration.

Matthias Bremm: software.

Tinghui Duan: writing – review & editing.

Anne Klee: writing – review & editing, data curation.

Johanna Konstanciak: writing – review & editing, data curation.

Julia Röttgermann: writing – original draft, data curation.

Moritz Steffes: software.

Christof Schöch: funding acquisition, methodology, writing – review & editing.

Joëlle Weis: writing – review & editing.

Acknowledgements

We thank the anonymous reviewers from Trier Center for Digital Humanities for their careful review of our manuscript and their many insightful comments and valuable suggestions.

We would also like to thank all members of the project network at Trier University, in particular Folke Gernert, Ulrich Port, Benjamin Raue, Achim Rettinger and Ralf Schenkel, for their essential contributions to the project.

Further thanks go to the research fund "Forschungsinitiative Rheinland-Pfalz", which funded the research in Mining and Modeling Text in the period 2019–2023.

Appendix

You can find an overview of all SPARQL queries of this contribution on the following page: https://mimotext.github.io/MiMoTextBase_Tutorial/queries_patterns.

We additionally use the PURL.org service, an initiative of the Internet Archive, to guarantee long-term availability.

- Overview: https://purl.org/mmt/patterns
- Query 0: https://purl.org/mmt/patterns/query0
- Query 1: https://purl.org/mmt/patterns/query1
- Query 2: https://purl.org/mmt/patterns/query2
- Query 3: https://purl.org/mmt/patterns/query3
- Query 4: https://purl.org/mmt/patterns/query4
- Query 5: https://purl.org/mmt/patterns/query5
- Query 6: https://purl.org/mmt/patterns/query6
- Query 7: https://purl.org/mmt/patterns/query7
- Query 8: https://purl.org/mmt/patterns/query8
- Query 9: https://purl.org/mmt/patterns/query9
- Query 10: https://purl.org/mmt/patterns/query10
- Query 11: https://purl.org/mmt/patterns/query11
- Query 12: https://purl.org/mmt/patterns/query12
- Query 13: https://purl.org/mmt/patterns/query13
- Query 14: https://purl.org/mmt/patterns/query14
- Query 15: https://purl.org/mmt/patterns/query15
- Query 16: https://purl.org/mmt/patterns/query16
- Query 17: https://purl.org/mmt/patterns/query17
- Query 18: https://purl.org/mmt/patterns/query18
- Query 19: https://purl.org/mmt/patterns/query19

References

Abel, Antoine. 2019. *Faster SPARQL Federated Queries*. Université Rennes1. https://inria.hal.science/hal-02269417. (26 January, 2024).

Achichi, Manel, Pasquale Lisena, Konstantin Todorov, Raphaël Troncy & Jean Delahousse. 2018. DOREMUS: A Graph of Linked Musical Works. In Denny Vrandečić, Kalina Bontcheva, Mari Carmen Suárez-Figueroa, Valentina Presutti, Irene Celino, Marta Sabou, Lucie-Aimée Kaffee & Elena Simperl (eds.), *The Semantic Web – ISWC 2018* (Lecture Notes in Computer Science), vol. 11137, 3–19. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-00668-6_1.

Alassi, Sepideh. 2023. From unstructured texts to RDF-star-based open research data queryable by references. Zenodo. https://doi.org/10.5281/ZENODO.8107643.

Alves, Daniel (ed.). 2022. *IJHAC: A Journal of Digital Humanities. Special Issue: Linked Open Data in the Arts and the Humanities. 16(1)*. https://www.euppublishing.com/doi/10.3366/ijhac.2022.0271. (3 January, 2024).

Ammann, Nora Olivia, Sepideh Alassi & Lukas Rosenthaler. 2023. Jacob Bernoulli's Reisbüchlein an RDF-star-based Edition. In Walter Scholger, Georg Vogeler, Toma Tasovac, Anne Baillot & Patrick Helling (eds.), *Digital Humanities 2023: Book of Abstracts*. Graz: Zenodo. https://zenodo.org/record/8108020. (3 January, 2024).

Arenas, Marcelo, Claudio Gutierrez & Jorge Pérez. 2010. On the Semantics of SPARQL. In Roberto de Virgilio, Fausto Giunchiglia & Letizia Tanca (eds.), *Semantic Web Information Management: A Model-Based Perspective*, 281–307. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-04329-1_13.

Baillie, James, Tara Andrews, Maxim Romanov, Daniel Knox & Maria Vargha. 2021. Modelling Historical Information with Structured Assertion Records. *Digital History Berlin (Blog)*. https://dhistory.hypotheses.org/518. (5 January, 2024).

Bartalesi, Valentina & Carlo Meghini. 2016. Using an ontology for representing the knowledge on literary texts: The Dante Alighieri case study. (Ed.) Eero Hyvönen. *Semantic Web* 8(3). 385–394. https://doi.org/10.3233/SW-150198.

Bartalesi, Valentina, Nicolò Pratelli & Emanuele Lenzi. 2022. Linking different scientific digital libraries in Digital Humanities: the IMAGO case study. *International Journal on Digital Libraries* 23(4). 303–317. https://doi.org/10.1007/s00799-022-00331-4.

Bekiari, Chryssoula, George Bruseker, Eri Canning, Martin Doerr, Philippe Michon, Christian-Emil Ore, Stephen Stead & Velios Athanasios. 2022. *Definition of the CIDOC Conceptual Reference Model*. CIDOC CRM Special Interest Group. Version 7.1.2. https://www.cidoc-crm.org/sites/default/files/cidoc_crm_v7.1.2.pdf. (26 January, 2024).

Bermúdez-Sabel, Helena, María Luisa Díez Platas, Salvador Ros & Elena González-Blanco. 2022. Towards a common model for European poetry: Challenges and solutions. *Digital Scholarship in the Humanities* 37(4). 921–933. https://doi.org/10.1093/llc/fqab106.

Bernard, Michel. 1999. *Introduction aux études littéraires assistées par ordinateur*. Paris: Presses Universitaires de France.

Berners-Lee, Tim. 2006. Linked Data — Design Issues. https://www.w3.org/DesignIssues/LinkedData.html. (6 April, 2022).

Bod, Rens. 2018. Modelling in the Humanities: Linking Patterns to Principles. *Historical Social Research*. GESIS - Leibniz Institute for the Social Sciences 31. 78–95. https://doi.org/10.12759/HSR.SUPPL.31.2018.78-95.

Borkowski, Jan & Philipp David Heine. 2013. Ziele der Literaturgeschichtsschreibung. *Journal of Literary Theory* 7(1–2). https://doi.org/10.1515/jlt-2013-0002.

Bradley, John. 2005. Texts into Databases: The Evolving Field of New-style Prosopography. *Literary and Linguistic Computing* 20(Suppl 1). 3–24. https://doi.org/10.1093/llc/fqi022.

Bradley, John. 2017. Factoids: A site that introduces Factoid Prosopograph. http://factoid-dighum.kcl.ac.uk. (26 January, 2024).

Browning, Gary K. 2000. *Lyotard and the end of grand narratives* (Political Philosophy Now). Cardiff: University of Wales Press.

Burrows, Simon. 2021. *Enlightenment Bestsellers*. London, New York, Oxford, New Delhi, Sydney: Bloomsbury Academic.

Burrows, Simon, Michael Falk, Rachel Hendery & Katherine McDonough. 2021. Stationers, Papetiers and the Supply Networks of a Swiss Publisher: The Sociéte Typographique de Neuchâtel and the Paper Trade 1769–1789. In Daniel Bellingradt & Anna Reynolds (eds.), *The Paper Trade in Early Modern Europe*, 266–301. Leiden: BRILL. https://doi.org/10.1163/9789004424005_013.

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

Coyle, Karen. 2022. Works, Expressions, Manifestations, Items: An Ontology. *The Code4Lib Journal* 53(2). https://journal.code4lib.org/articles/16491. (4 January, 2024).

Dengel, Andreas (ed.). 2012. *Semantische Technologien: Grundlagen. Konzepte. Anwendungen*. Heidelberg: Spektrum. https://doi.org/10.1007/978-3-8274-2664-2.

Diefenbach, Dennis, Max De Wilde & Samantha Alipio. 2021. Wikibase as an Infrastructure for Knowledge Graphs: The EU Knowledge Graph. In Andreas Hotho, Eva Blomqvist, Stefan Dietze, Achille Fokoue, Ying Ding, Payam Barnaghi, Armin Haller, Mauro Dragoni & Harith Alani (eds.), *The Semantic Web – ISWC 2021* (Lecture Notes in Computer Science), vol. 12922, 631–647. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-88361-4_37.

Donkers, Alex, Dujuan Yang & Nico Baken. 2020. Linked data for smart homes: comparing RDF and labeled property graphs. *Proceedings of the 8th Linked Data in Architecture and Construction Workshop — LDAC2020*. https://ceur-ws.org/Vol-2636/02paper.pdf.

Drucker, Johanna. 2011. Humanities Approaches to Graphical Display. *Digital Humanities Quarterly* 5(1). https://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html. (2 February, 2024).

DuCharme, Bob. 2013. *Learning SPARQL*. Sebastopol, United States: O'Reilly Media.

Eells, Andrew, Cogan Shimizu, Lu Zhou, Pascal Hitzler, Seila Gonzales & Dean Rehberger. 2021. Aligning Patterns to the Wikibase Model. https://raw.githubusercontent.com/odpa/WOP2021/main/paper2.pdf. (26 January, 2024).

Eide, Øyvind. 2014. Ontologies, Data Modeling, and TEI. *Journal of the Text Encoding Initiative* (8). https://doi.org/10.4000/jtei.1191.

Erdbeer, Robert Matthias, Florian Kläger & Klaus Stierstorfer (eds.). 2018. *Literarische Form: Theorien, Dynamiken, Kulturen: Beiträge zur literarischen Modellforschung = Literary form: theories, dynamics, cultures: perspectives on literary modelling* (Beiträge zur neueren Literaturgeschichte). Vol. 371. Heidelberg: Universitätsverlag Winter.

Fischer, Barbara. 2022. Towards an open and collaborative Authority Control. *JLIS.it* 13(1). 283–290. https://doi.org/10.4403/jlis.it-12767.

Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hechtl, Christopher Kittel, Carsten Milling & Peer Trilcke. 2019. Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama. *Digital Humanities 2019: "Complexities" (DH2019)*, *Utrecht*, *9–12 July 2019*. Zenodo. https://doi.org/10.5281/ZENODO.4284002.

Flanders, Julia & Fotis Jannidis (eds.). 2018. *The Shape of Data in the Digital Humanities: Modeling Texts and Text-based Resources* (Digital Research in the Arts and Humanities). 1st edn. Abingdon, Oxon; New York, NY: Routledge. https://doi.org/10.4324/9781315552941.

Görlitz, Olaf & Steffen Staab. 2011. Federated Data Management and Query Optimization for Linked Open Data. In Athena Vakali & Lakhmi C. Jain (eds.), *New Directions in Web Data Management 1* (Studies in Computational Intelligence), vol. 331, 109–137. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-17551-0_5.

Gruber, Thomas R. 1995. Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies* 43(5–6). 907–928. https://doi.org/10.1006/ijhc.1995.1081.

Hadden, Richard, Matthias Schlögl & Georg Vogeler. 2022. Towards a prosopographical ecosystem: modelling, design, and implementation issues. In Yifan Wang, Tomohiro Murase, Kiyonori Nagasaki, Yoshihiro Sato & Shintaro Seki (eds.), *Digital Humanities 2022 : Conference Abstracts*, *The University of Tokyo*, *Japan 25–29 July 2022*, 472–473. Tokyo. https://dh2022.dhii.asia/dh2022bookofabsts.pdf. (2 February, 2024).

Haider, Thomas Nikolaus, Stephanie Schennach, Julius Thelen & Jörg Wesche. 2022. Barockpoetik als Wikibase: Eine Datenbank zu konfessionsgeschichtlichen Aspekten in deutschen Barockpoetiken. Zenodo. https://doi.org/10.5281/ZENODO.7715343.

Haraway, Donna. 1988. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies* 14(3). 575–599. https://doi.org/10.2307/3178066.

Harris, Steve & Andy Seaborne. 2013. SPARQL 1.1 Query Language — 5 Graph Patterns. *W3C Recommendation*. https://www.w3.org/TR/2013/REC-sparql11-query-20130321/#GraphPattern. (26 January, 2024).

Hernández, Daniel, Aidan Hogan & Markus Krötzsch. 2015. Reifying RDF: What works well with wikidata? *Proceedings of the 11th International Workshop on Scalable Semantic Web Knowledge Base Systems (CEUR Workshop Proceedings)* 1457. 32–47.

Hinzmann, Maria, Anne Klee, Johanna Konstanciak, Julia Röttgermann, Christof Schöch & Moritz Steffes. 2022a. MiMoTextBase Tutorial. https://mimotext.github.io/MiMoTextBase_Tutorial/. (1 August, 2022).

Hinzmann, Maria, Julia Röttgermann, Anne Klee, Moritz Steffes & Christof Schöch. 2022b. The French Enlightenment Novel as a Graph? Potentials and Challenges in the Construction of a Knowledge Network (extended abstract). Zenodo. https://doi.org/10.5281/ZENODO.5840089.

Hitzler, Pascal, Markus Krötzsch & Sebastian Rudolph. 2009. *Foundations of Semantic Web Technologies*. New York: Chapman and Hall/CRC. https://doi.org/10.1201/9781420090512.

Hollier, Denis (ed.). 1994. *A new history of French literature*. 1st edn. Cambridge, MA: Harvard University Press.

Hooland, Seth van & Ruben Verborgh. 2014. *Linked Data for Libraries, Archives and Museums: How to clean, link and publish your metadata*. London: Facet Publishing.

Hyvönen, Eero, Petri Leskinen & Jouni Tuominen. 2023. LetterSampo — Historical Letters on the Semantic Web: A Framework and Its Application to Publishing and Using Epistolary Data. *Journal on Computing and Cultural Heritage* 16(1). 1–23. https://doi.org/10.1145/3569372.

Ikonić Nešić, Milica, Ranka Stanković & Biljana Rujević. 2021. Serbian ELTeC Sub-Collection in Wikidata. *Infotheca* 21(2). 60–86. https://doi.org/10.18485/infotheca.2021.21.2.4.

Jannidis, Fotis. 2017. Datenmodellierung. In Fotis Jannidis, Hubertus Kohle & Malte Rehbein (eds.), *Digital Humanities: Eine Einführung*, 99–108. Stuttgart: Metzler.

Janowicz, Krzysztof, Aldo Gangemi, Pascal Hitzler, Adila Krisnadhi & Valentina Presutti. 2016. Introduction: Ontology Design Patterns in a Nutshell. In Pascal Hitzler, Aldo Gangemi, Krzysztof Janowicz, Adila Krisnadhi & Valentina Presutti (eds.), *Ontology engineering with ontology design patterns: Foundations and applications* (Studies on the Semantic Web), vol. 25, 3–21. Berlin, Amsterdam: Akademische Verlagsgesellschaft ; IOS Press.

Kräutli, Florian, Esther Chen & Matteo Valleriani. 2021. Linked data strategies for conserving digital research outputs. The shelf life of digital humanities. In Koraljka Golub & Ying-Hsang Liu (eds.), *Information and Knowledge Organisation in Digital Humanities: Global Perspectives*, 206–224. London: Routledge. https://doi.org/10.4324/9781003131816.

Kuczera, Andreas, Thorsten Wübbena & Thomas Kollatz. 2019. Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten: Ausgewählte Beiträge der Tagung 19.-20.01.2018 an der Akademie der Wissenschaften und der Literatur, Mainz. Wolfenbüttel,: Herzog August Bibliothek. https://doi.org/10.17175/SB004.

Lavin, Matthew. 2021. Why Digital Humanists Should Emphasize Situated Data over Capta. *Digital Humanities Quarterly* 15(2). http://www.digitalhumanities.org/dhq/vol/15/2/000556/000556.html. (2 February, 2024).

Liu, Fangchao, John Hindmarch & Mona Hess. 2023. A Review of the Cultural Heritage Linked Open Data Ontologies and Models. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLVIII-M-2–2023. 943–950. https://doi.org/10.5194/isprs-archives-XLVIII-M-2-2023-943-2023.

Lüschow, Andreas. 2020. Automatische Extraktion und semantische Modellierung der Einträge einer Bibliographie französischsprachiger Romane. In Christof Schöch (ed.), *Spielräume: Digital Humanities zwischen Modellierung und Interpretation. Konferenzabstracts*, 80–84. Paderborn: Dhd-Verband. https://doi.org/10.5281/ZENODO.3666690.

Martin, Angus, Vivienne Mylne & Richard L. Frautschi. 1977. *Bibliographie du genre romanesque français, 1751-1800*. London: Mansell.

Massari, Arcangelo, Silvio Peroni, Francesca Tomasi & Ivan Heibi. 2023. Representing provenance and track changes of cultural heritage metadata in RDF: a survey of existing approaches. Zenodo. https://doi.org/10.5281/ZENODO.8108101.

Matuschek, Stefan & Sandra Kerschbaumer (eds.). 2019. *Romantik erkennen — Modelle finden*. Boston: BRILL.

McCarty, Willard. 2005. Modelling. In *Humanities computing*, 20–72. Basingstoke, New York:

Palgrave Macmillan. https://doi.org/10.1002/9780470999875.

Montoya, Alicia C. & Roger Chartier. 2017. Middlebrow, Religion, and the European Enlightenment. A New Bibliometric Project, MEDIATE (1665-1820). *French History and Civilization* 7. 66–79.

Moretti, Franco. 2017. Patterns and Interpretation. *Literary Lab Pamphlets* 15. https://litlab.stanford.edu/LiteraryLabPamphlet15.pdf. (2 February, 2024).

Neubert, Joachim. 2017. Wikidata as a Linking Hub for Knowledge Organization Systems? Integrating an Authority Mapping into Wikidata and Learning Lessons for KOS Mappings. *Proceedings of the 17th European NKOS workshop*. https://ceur-ws.org/Vol-1937/paper2.pdf.

Nicka, Isabella, Peter Hinkelmanns, Miriam Landkammer, Manuel Schwembacher & Katharina Zeppezauer-Wachauer. 2020. Erzählerische Spielräume. Medienübergreifende Erforschung von Narrativen im Mittelalter mit ONAMA. In Christof Schöch (ed.), *Spielräume: Digital Humanities zwischen Modellierung und Interpretation. Konferenzabstracts*, 131–135. Paderborn: Dhd-Verband. https://doi.org/10.5281/ZENODO.3666690.

Noy, Natalya F. & Deborah L. McGuinness. 2001. Ontology Development 101: A Guide to Creating Your First Ontology. https://protege.stanford.edu/publications/ontology_development/ontology101.pdf. (2 February, 2024).

Paige, Nicholas D. 2020. *Technologies of the Novel: Quantitative Data and the Evolution of Literary Systems*. New York: Cambridge University Press. https://doi.org/10.1017/9781108890861.

Pasqual, Valentina & Francesca Tomasi. 2023. Data narratives with Linked Open Data, the case of mythLOD storytelling. (Ed.) Walter Scholger, Georg Vogeler, Toma Tasovac, Anne Baillot & Patrick Helling. *Digital Humanities 2023*. Zenodo. https://doi.org/10.5281/ZENODO.8107673.

Passarotti, Marco Carlo, Francesco Mambrini, Greta Franzini, Flavio Massimiliano Cecchini, Eleonora Litta, Giovanni Moretti, Paolo Ruffolo & Rachele Sprugnoli. 2020. Interlinking through Lemmas. The Lexical Collection of the LiLa Knowledge Base of Linguistic Resources for Latin. Zenodo. https://doi.org/10.5281/ZENODO.4017229.

Pianzola, Federico, Xiaoyan Yang, Noa Visser, Michiel van der Ree & Andreas van Cranenburgh. 2023. Constructing the GOLEM: Graphs and Ontologies for Literary Evolution Models. Zenodo. https://doi.org/10.5281/ZENODO.8206543.

Pichler, Axel & Nils Reiter. 2022. From Concepts to Texts and Back: Operationalization as a Core Activity of Digital Humanities. *Journal of Cultural Analytics* 7(4). https://doi.org/10.22148/001c.57195.

Prud'hommeaux, Eric & Carlos Buil-Aranda. 2013. SPARQL 1.1 Federated Query. *W3C Recommendation*. https://www.w3.org/TR/sparql11-federated-query/. (1 August, 2022).

Prud'hommeaux, Eric & Andy Seaborne. 2008. SPARQL Query Language for RDF - 7 Matching Alternatives. *W3C Recommendation*. https://www.w3.org/TR/rdf-sparql-query/. (2 February, 2024).

Röttgermann, Julia, Maria Hinzmann, Henning Gebhard, Anne Klee, Johanna Konstanciak, Christof Schöch & Moritz Steffes. 2022. Mining and Modeling Spaces and Places for Literary History as Linked Open Data. In Ikki Ohmukai & Taizo Yamada (eds.), *DH 2022 — Conference Abstracts*. Tokyo: DH2022 Local Organizing Committee.

https://zenodo.org/record/6948236. (15 January, 2024).

Röttgermann, Julia & Christof Schöch. 2020. FAIRe Daten in den Literaturwissenschaften? Das Beispiel „Mining and Modeling Text“ und der französische Roman des 18. Jahrhunderts. *Romanistik-Blog. Blog des Fachinformationsdienstes*. https://blog.fid-romanistik.de/2020/11/05/faire-daten-in-den-literaturwissenschaften/. (26 January, 2024).

Sack, Harald. 2022. nfdi4Culture: Knowledge Graphs (and Wikibase) for Research Data Management. *NFDI InfraTalks | online*. Zenodo. https://doi.org/10.5281/ZENODO.6372897.

Schöch, Christof. 2021. Open Access für die Maschinen. In Maria Effinger & Hubertus Kohle (eds.), *Die Zukunft des kunsthistorischen Publizierens*, 79–94. arthistoricum.net. https://doi.org/10.11588/ARTHISTORICUM.663.

Schöch, Christof, Maria Hinzmann, Julia Röttgermann, Anne Klee & Katharina Dietz. 2022. Smart Modelling for Literary History. *IJHAC: International Journal of Humanities and Arts Computing [Special issue on Linked Open Data]* 16(1). 78–93. https://doi.org/10.3366/ijhac.2022.0278.

Shimizu, Cogan, Andrew Eells, Seila Gonzalez, Lu Zhou, Pascal Hitzler, Alicia Sheill, Catherine Foley & Dean Rehberger. 2022a. Ontology Design Facilitating Wikibase Integration -- and a Worked Example for Historical Data. arXiv. https://doi.org/10.48550/ARXIV.2205.14032.

Shimizu, Cogan, Karl Hammar & Pascal Hitzler. 2022b. Modular Ontology Modeling. *Semantic Web — Interoperability, Usability, Applicability*. https://doi.org/10.3233/SW-222886.

Simons, Olaf. 2022. FactGrid. Forschungszentrum Gotha der Universität Erfurt. http://doi.org/10.17616/R31NJMQR. (26 January, 2024).

Spadini, Elena & Francesca Tomasi. 2021. Introduction. In Elena Spadini, Francesca Tomasi & Georg Vogeler (eds.), *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing* (Schriften Des Instituts Für Dokumentologie Und Editorik), vol. 15, 1–6. Norderstedt: Books on Demand.

Stanković, Rana, Christian Chiarcos, Miloš Utvić & Olivera Kitanović. 2023. Towards ELTeC-LLOD: European Literary Text Collection Linguistic Linked Open Data. In Sara Carvalho, Anas Fahad Khan, Ana Ostroški Anić, Blerina Spahiu, Jorge Gracia, John P. McCrae, Dagmar Gromann, Barbara Heinisch & Ana Salgado (eds.), *Proceedings of the 4th Conference on Language, Data and Knowledge, Vienna*, 180–191. Portugal: NOVA CLUNL. https://aclanthology.org/2023.ldk-1.16. (25 January, 2024).

Thornton, Katherine, Kenneth Seals-Nutt, Marianne van Remoortel, Julie M. Birkholz & Pieterjan de Potter. 2022. Linking Women Editors of Periodicals to the Wikidata Knowledge Graph. *Semantic Web Journal [Special Issue Cultural Heritage 2021]*. https://doi.org/10.3233/SW-222845.

Unbehauen, Jörg, Claus Stadler & Sören Auer. 2013. Accessing Relational Data on the Web with SparqlMap. In Hideaki Takeda, Yuzhong Qu, Riichiro Mizoguchi & Yoshinobu Kitamura (eds.), *Semantic Technology* (Lecture Notes in Computer Science), vol. 7774, 65–80. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-37996-3_5.

W3C SPARQL Working group. 2013. W3C SPARQL Working Group SPARQL 1.1 Overview W3C - Recommendation 21 March 2013. http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/. (26 January, 2024).

Zhao, Fudie. 2022. A systematic review of Wikidata in Digital Humanities projects. *Digital Scholarship in the Humanities*. https://doi.org/10.1093/llc/fqac083.

Zhou, Lu, Cogan Shimizu, Pascal Hitzler, Alicia M. Sheill, Seila Gonzalez Estrecha, Catherine Foley, Duncan Tarr & Dean Rehberger. 2020. The Enslaved Dataset: A Real-world Complex Ontology Alignment Benchmark using Wikibase. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* 3197–3204. https://doi.org/10.1145/3340531.3412768.