Encoding and Querying Historic Map Content

Simon Scheider, Jim Jones, Alber Sánchez and Carsten Keßler

Abstract Libraries have large collections of map documents with rich spatio-temporal information encoded in the visual representation of the map. Currently, historic map content is covered by the provided metadata only to a very limited degree, and thus is not available in a machine-readable form. A formal representation would support querying for and reasoning over detailed semantic contents of maps, instead of only map documents. From a historian's perspective, this would support search for map resources which contain information that answers very specific questions, such as *maps that show the cities of Prussia in 1830*, without manually searching through maps. A particular challenge lies in the wealth and ambiguity of map content for queries. In this chapter, we propose an approach to describe map contents more explicitly. We suggest ways to formally encode historic map content in an approximate *intensional* manner which still allows useful queries. We discuss tools for georeferencing and enriching historic map descriptions by external sources, such as DBpedia. We demonstrate the use of this approach by content queries on map examples.

S. Scheider · J. Jones · A. Sánchez
Institute for Geoinformatics, University of Münster, Münster, Germany
e-mail: simon.scheider@uni-muenster.de

J. Jones
e-mail: jim.jones@uni-muenster.de

A. Sánchez
e-mail: a.sanchez@uni-muenster.de

C. Keßler (✉)
CARSI, Department of Geography, Hunter College,
City University of New York, New York, USA
e-mail: carsten.kessler@hunter.cuny.edu

J. Huerta et al. (eds.), *Connecting a Digital Europe Through Location and Place*,
Lecture Notes in Geoinformation and Cartography, DOI: 10.1007/978-3-319-03611-3_15,

1 Introduction

Historic maps provide rich knowledge resources that graphically encode information about the state of a fraction of the real world at a certain point in time. As such, libraries and archives with very large collections such as the Library of Congress with its 5.5 million maps[1] offer an invaluable data source for historians and other researchers. However, libraries and historians cannot make full use of the encoded knowledge to date, as it is currently not possible to automatically retrieve maps that contain the answers to specific questions such as *what were the cities of Prussia in 1830*? or *was Posen part of Prussia in 1802*? While the typical metadata for a historical map contain the title, author, year of production, year represented, and a number of standardized keywords, it often remains unclear whether or not a map is able to help answering such a specific question. If a historian wants to find an answer, it depends on her background knowledge and a fair bit of luck in picking the right keywords for the query to find maps that potentially contain the answer. If those maps have not been digitized yet, it may require searching through a large number of actual paper maps. This is even the case for the fairly small number of maps in the University and State Library (about 2,000) or the Institute for Comparative Urban History's library[2] (about 20,000) at the University of Münster, Germany, that we have dealt with in this work.

A particular problem concerns the way how such detailed map contents should be encoded in order to be machine-readable and in order to allow such detailed queries. Furthermore, the *wealth* as well as the *ambiguity* of the content, even of a single map, poses a challenge for libraries. Extending the metadata for a map to cover all potential keywords about content is clearly not feasible, especially for very large collections.

To overcome this problem, we propose an approach for encoding and querying historic map content based on Semantic Web technology. It makes use of the fact that all map content is *semantically describable in an approximate manner*, *spatially* arranged on the map, and *temporally* referenced through metadata. In a nutshell, we discuss how to make visually encoded map information semantically explicit and queryable to different degrees. We demonstrate how to draw on external knowledge resources, such as DBpedia, for *named explicit content*, and discuss how *implicit contents* can be described in an *intensional* manner which is still useful for search. We base our discussion on a number of competency questions and query scenarios with three map examples in Sect. 2. Section 3 gives an overview of relevant related work. In Sect. 4, we discuss formal ways of encoding map contents, corresponding vocabularies, and review existing software tools for semantic enrichment. Section 5 applies this to the map examples and Sect. 6 demonstrates queries which correspond to the competency questions.

---

[1] http://www.loc.gov/rr/geogmap/guides.html

[2] http://www.uni-muenster.de/Staedtegeschichte/portal/datenbanken

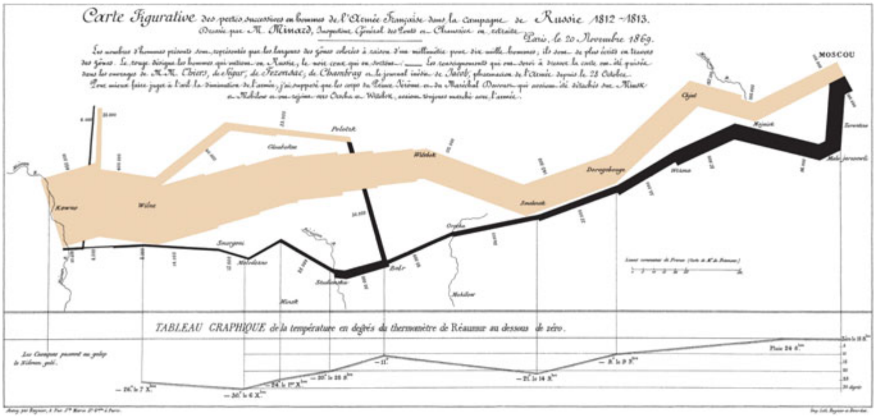


Fig. 1 Charles Minards map from 1861 about Napoleon's 1812 march on Russia

2 Motivating Examples

Following the methodology proposed by Gangemi and Presutti (2009), we start with discussing three map examples together with some *competency questions* that may be asked about them. These questions make some of the semantic content of the maps explicit, and they inform ontology engineers how such contents might be represented. We are aware that our selection does not in any way cover many relevant types of historic maps and possible questions that one might have. Yet, we believe that it already shows considerable variety and depth in content, and therefore serves as a good starting point for research. We focus on three challenging historic maps whose contents are not straightforward to describe.

The first example we discuss is the famous map of Napoleon's 1812 march on Russia by Charles Joseph Minard (see Fig. 1). It depicts the losses of Napoleon's army during his Russian campaign, showing the advance and retreat paths with cartographic signs depicting the number of people of the campaign, the visited places, as well as the temperature. Note that the map is not accurate in terms of the locations and the shape of the paths. The details of content and alternative ways of visualizing it are discussed in Kraak (2003). We focus here on some of the informative questions that might be asked about this map:

1. Where did Napoleon's 1812 campaign to Russia happen?
2. How many people did Napoleon's army have when soldiers arrived in Smolensk during his 1812 campaign?
3. What were the lowest temperatures during Napoleon's campaign?
4. Which places did Napoleon's army come across during the 1812 campaign?

Our second example is a political map which depicts the administrative parts of Prussia in the 17th century (see Fig. 2). It shows the different sub-territories and the Prussian dukes involved in obtaining them in the course of this century.

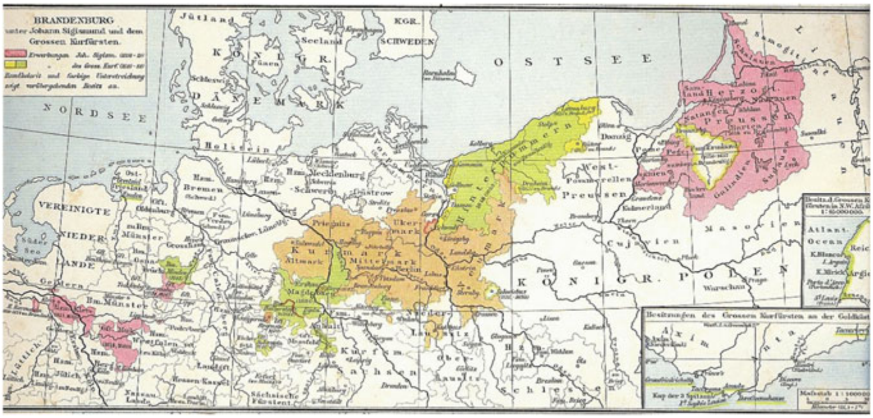


Fig. 2 Prussia (Brandenburg) in the 17th century. *Source* (public domain) G. Droysens Historischer Handatlas, 1886

1. Where was Prussia in 1688?
2. Which territories were part of Prussia in 1688?
3. Which Prussian territories were acquired by Friedrich-Wilhelm of Brandenburg, the great elector?

Our third example is a topographic map of Hildesheim from 1839 (see Fig. 3). This map is very rich in detail and in thematic content, including roads, landscape features, landcover, as well as the built environment. However only few of these depicted kinds of things are actually named entities.

1. Where was Hildesheim in 1839?
2. What were the types of landcover around Hildesheim in 1839?

The problem we address in this chapter is how an explicit semantic representation of the content should look like in order to answer these questions in terms of queries. Based on our examples, we explore a number of related research challenges: Which language and which expressivity is needed for encoding? In particular, how can we encode *complex spatio-temporal map contents*, such as the story behind Minard's map? How can we describe contents involving well known *named entities*, such as the Prussian territories and their associations with kings? How can we encode contents involving *nameless entities*, such as the landscape around Hildesheim or city blocks? And, since encoding each and every detail of a map's content is clearly not feasible: how can map content be encoded in an *intensional form*, i.e., such that only aspects of the content are made explicit?

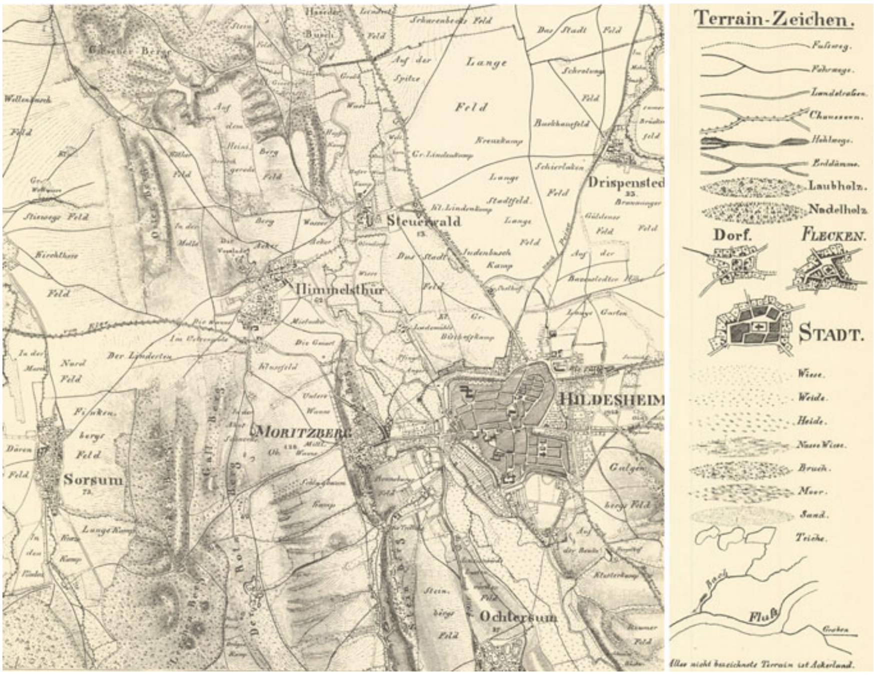


Fig. 3 Excerpt of a map of Hildesheim of the “Gaußsche Landesaufnahme” from 1839. *Source* Historische Kommission für Niedersachsen, Hannover 1963

3 Related Work

Making the content of historic maps and libraries available to the public is an ongoing topic of research as well as development. Library records are commonly encoded in the MARC21 standard or similar, however, more flexible and precise ways of representing document contents are needed. One focus is on combining historical databases with geographical information systems (GIS) (see Grossner (2010) for an overview) and on novel approaches of visual map representation Kraak (2003). A recent line of developments is concerned with *historical gazetteers*, such as CHGIS[3] and Simon et al. (2012). However, only few authors have focused on using Semantic Web technology for map descriptions so far. In the context of digital libraries, there is recent work on tools for annotating historic maps (Simon et al. 2011; Haslhofer et al. 2013), extracting content (Arteaga 2013) and on semantic vocabularies and ontologies (Gkadolou and Stefanakis 2013; Gkadolou et al. 2013; Grossner 2010). Gkadolou and Stefanakis (2013) propose an extension of CIDOC-CRM[4] for

---

[3] http://www.fas.harvard.edu/~chgis/

[4] The International Committee for Museum Documentation’s conceptual reference model for cultural heritage documentation, see http://cidoc.ics.forth.gr.

annotating map documents. Their historic maps ontology[5] covers document descriptions as well as content classes, however it is less suitable for content encoding as proposed in this chapter. Grossner (2010) suggests a general ontology to represent historic knowledge which is event based, providing relevant insights for this chapter. However, our focus is not on content vocabularies but rather on content encoding methods and queries. Simon et al. (2011) have suggested map annotation methods useful for content descriptions, but they do not focus on encoding complex content. In another chapter (Carral et al. 2013), we have proposed an ontology design pattern for map scaling in order to make maps of different scale discoverable in the Web. Hyvönen et al. (2011) discuss the problem of temporal identity of regions in historic data sets. There is also more general work on publishing cultural heritage data as Linked Open Data (Ruotsalo et al. 2013) and on using Linked Data in geographic contexts (Hart and Dolbear 2013).

4 Describing Historic Map Contents

In this section, we discuss semantic approaches to describing map contents. In particular, we address the problems of (a) finding a syntax for encoding these contents; and (b) describing contents in an approximate manner.

4.1 Formally Encoding Map Contents

What exactly is the content of a map? This question, at least in this form, might be too big to yield an answer, given the contested nature of maps as complex signs (MacEachren 2004). It may be as difficult to answer as the question about the semantic content of texts. However, from a pragmatic viewpoint and without being overly reductionist, one could say that the content of a map as a document is the *set of assertions* which can be extracted by looking at it (see Fig. 4).

Such *map interpretation* has a long tradition in cartography and geography, and it consists in drawing explicit conclusions from looking at a map. It is well known that conclusions can be different depending on who is looking at a map. Furthermore, an interpreter may have difficulties in actually writing down *all* contained assertions in an exhaustive manner. Still, it makes sense to think about map content in terms of a set of assertions which *can* be made by some interpreter. For example, the assertions extractable from the Minard map may include one statement saying that upon Napoleon's arrival in Moscow, the temperature was 0 degree on the Reaumur scale. And the map content of the Prussia map may contain assertions saying that Hinterpommern is a territory, that was part of Prussia 1806, and another one saying that it was acquired by elector Friedrich-Wilhelm. All of these assertions are visually

---

[5] http://gaia.gge.unb.ca/eg/HistoricalMap.owl

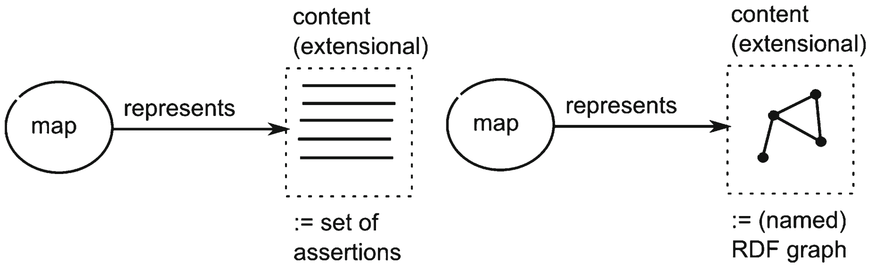


Fig. 4 Map contents as sets of assertions. A useful form of encoding such assertions are RDF graphs. Encoding sets of such assertions can be done in terms of "named graphs"

encoded in the corresponding map and can thus be extracted by human beings without drawing on further knowledge sources. In the following paragraphs, we discuss formal encoding schemes for such content.

Encoding map content as a named graph In order to formally encode these assertions and for making them machine-readable, however, we need a formal language and a vocabulary which allows us to talk about *individual represented phenomena*. That is, we need names for represented individual entities, like "Prussia" and "Hinterpommern", and we need also logical constants for classes and relations, such as "territory" or "state" and "is a", "part of" or "temperature was". Furthermore, we need a way of linking maps with their content.

One simple form of encoding such assertions are *RDF*[6] *Triples*, i.e., logical assertions with *subject*, *predicate*, and *object*. They form edges of a labeled graph in which subject and object are nodes and predicates denote types of edges. RDF assertions written down in this manner[7] look like this:

$$dbp{:}Hinterpommern \quad rdf{:}type \quad phen{:}Territory. \tag{1}$$

$$dbp{:}Hinterpommern \quad phen{:}partOfObject \quad dbp{:}Prussia. \tag{2}$$

where *dbp:*, *rdf:* and *phen:* specify namespaces for shared vocabularies, such as DBpedia[8] and RDF.[9] In the following, we will use the abbreviation *a* for *rdf:type*. RDF triples of this form do not contain any variables, only constant names. Names can either be Web addresses (URIs) or strings (called *literals*). The resulting assertions then form a *Linked Data graph* of explicit map contents which can be published on the Web (Bizer et al. 2009).

How should this content description be linked to a map? Maps can easily be encoded as single node which denotes some document. We then use a predicate

---

[6] http://www.w3.org/RDF/

[7] Throughout the chapter, we use the Turtle syntax to write down triples; see http://www.w3.org/TeamSubmission/turtle/.

[8] http://dbpedia.org

[9] http://www.w3.org/1999/02/22-rdf-syntax-ns

`maps:represents` as a semiotic short hand for the fact that some part of the map image, namely a certain *map sign*, represents something. Note that this "something" needs to be a graph in our case: *Each map represents one graph which encodes its content*, see Fig. 4. This requires to link maps to whole graphs, not to single nodes, which is only possible using a *named graph*,[10] i.e., a graph of assertions which has been assigned a URI for identification. The latter can be linked via `maps:represents` to a map and then serves to answer corresponding map queries. However, while triple stores such as OWLIM[11] support named graphs, they actually break the language barriers of RDF. They require, in effect, quadruples instead of triples.

Encoding map content by direct links If one wants to stay in RDF (or in some other Semantic Web logic) and if one is willing to sacrifice content relations for content nodes, it is possible to link contents without a named graph:

$$:map \quad maps{:}represents \quad dbp{:}Prussia. \tag{3}$$

It is furthermore possible and indeed makes sense to use both of these approaches for encoding map contents. In the following, we assume that all nodes of a content graph are also directly linked to the corresponding map.

The problem of encoding content implicitly The decision which names, classes and relations are needed to encode contents is not easy to take. Furthermore, it is almost unfeasible to write down map content assertions in an exhaustive manner. The reason is not only that it is too much of an effort to write them down, but also that for many assertions visually encoded in a map, we lack corresponding unambiguous names and constants. A look at the Hildesheim map (see Fig. 3) makes this clear: Most assertions depicted in this map talk about things that are not explicitly named, and about relations for which we may not know any adequate established constant names. For example, there are landscape objects like hills without names, and the question is open which geometrical relations we should choose.

Intensional description of map contents by graph patterns This makes it very hard to turn map content into a machine-readable form. Using formal logic, however, there is still a way of describing map contents in an approximate manner. If for some reason, we are unable to list content assertions, we may still write down some logical description of these assertions useful for search. For example, it is not required to explicitly encode all names and content assertions in terms of ground sentences. We can instead use variables that range over potential entities in order to avoid talking explicitly about some content. This allows us to encode map content in an *intensional* instead of an *extensional form*.

A useful form of such an intensional map content description is a *constraint network* or a *graph pattern*. This is a graph with variables. Assume each variable is implicitly existentially quantified (see Fig. 5). Such a graph is like a successful query that would deliver results if fired against an exhaustive set of RDF statements which

---

[10] http://en.wikipedia.org/wiki/Named_graph

[11] http://www.ontotext.com/owlim

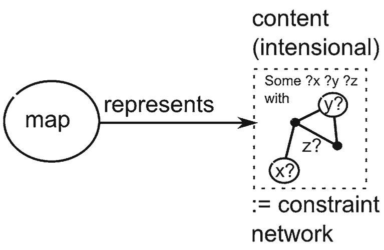


Fig. 5 Intensional description of map content using a constraint network. *Black nodes* are constants and question marks indicate variables

constitutes a map's content (or that of a large map collection). And just like such a query, it says something about the map's content. Graph patterns are therefore also used as bodies in the SPARQL query language.[12]

For example, without any knowledge of the historic names, one could encode the unnamed content of the Prussian map simply as the following query, where question marks indicate variables:

$$\begin{aligned} ?x\ {:}partOfObject\ ?y.\\ ?y\ a\ {:}State.\\ ?x\ {:}wasAcquiredBy\ ?z.\\ ?w\ {:}isSettingForPerson\ ?z.\\ ?w\ {:}isSettingForRole\ {:}King.\\ ?w\ {:}rulesOver\ ?y.\\ ?w\ {:}isSettingatTime\ ?q. \end{aligned} \tag{4}$$

The query says that some person *?z*, such as Friedrich the Great, who took the role of king of some state *?y*, such as Prussia, during time interval *?q*, such as 1740–1786, acquired some *?x*, such as Silesia, which was part of Prussia.[13]

A useful way of encoding such content graph patterns is to write them down as *graphs with blank nodes* instead of variables. Blank nodes, denoted by `_:id`, where *id* can be any string, can be interpreted as existentially quantified variables, because they can be interpreted into any other node satisfying a query. In our implementation, we encoded intensional map contents in terms of *named graphs with blank nodes*, as this allows to use the query mechanism of a triple store.

A simple intensional map content graph For further illustration, we discuss an example which corresponds to a very simple but useful constraint network. Because of

---

[12] http://www.w3.org/TR/rdf-sparql-query/

[13] In order to stay historically correct, one would need to say that Silesia was part of Prussia only after its conquest in 1763. This would require to introduce time-indexed *partOf* relations. Similarly, *wasAcquiredby* reflects some event. However, as a matter of fact, such kind of information is actually not contained in the map, and thus should not be represented by the content graph. Moreover, representing such time-indexed relationships presents a challenge of its own (Trame et al. 2013) which goes beyond the scope of this chapter.

their simplicity, graphs of this form can be encoded also in terms of constructs supported by the Web Ontology Language (OWL)[14] instead of named graphs with blank nodes. OWL encodes a fragment of Description Logic (DL) (Krötzsch et al. 2012), a subset of First-order Logic (FOL) which allows quantifying variables to a limited extent.

The simplest but still useful case is when we ask whether an instance of a certain class is contained in a map. People who catalog contents may only want to say that “some phenomenon of a class” is depicted in a map. For example, one may want to say that a map contains some administrative units, some roads, rivers and cities, without telling exactly which units, roads and cities are present. This directly corresponds to the following pattern:

$$?x\ a\ {:}River. \tag{5}$$

Without making use of a named graph, this content description can also be directly linked to a map by some Description Logic (DL) existential:

$$:map\ a\ \exists{:}represents.{:}River. \tag{6}$$

In this expression, the class $\exists represents.River$ stands for an OWL class restriction (where square brackets stand for a blank node), which is encoded in RDF as:

$$\begin{aligned}[a\ owl{:}Restriction;\\ owl{:}onProperty\ {:}represents;\\ owl{:}someValuesFrom\ {:}River].\end{aligned} \tag{7}$$

The expression literally means “the class of maps which represent something which is a river”. A future task is to identify similar intensional content patterns which can be encoded directly into some DL fragment and which can then be used for reasoning.

4.2 Vocabularies for Historic Map Content

Map descriptions need to make use of content vocabularies as well as vocabularies for describing documents. As these are two fundamentally different matters, they need to be distinguished. We first discuss vocabularies for geographic phenomena, and then propose vocabularies for describing semantic, spatial and temporal reference of maps as documents.

Geographic phenomena A geographic phenomenon (denoted by the class *GeoPhenomenon*) is any phenomenon on a geographic scale (Montello 1993) which can be

---

[14] http://www.w3.org/TR/owl-ref/

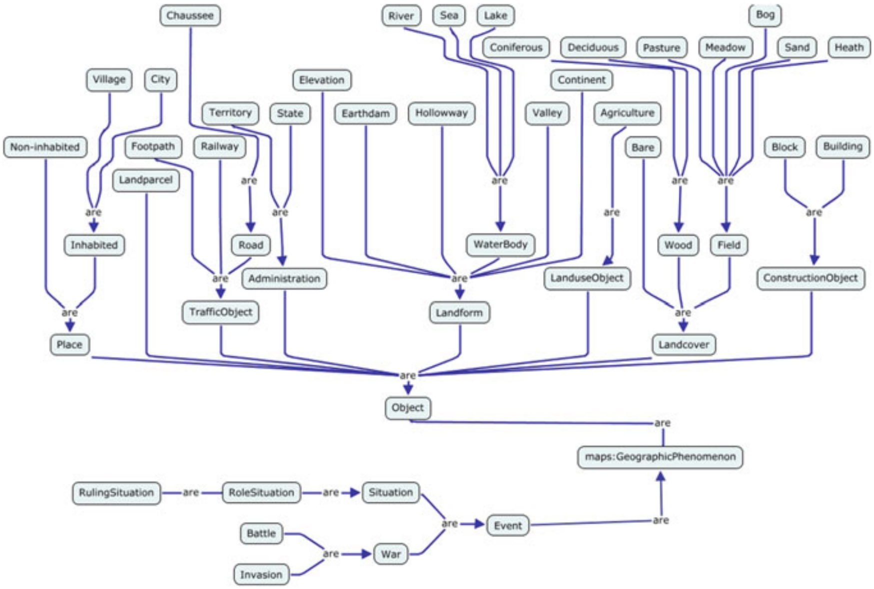


Fig. 6 Geographic phenomenon classes relevant for describing the content of historic maps

mapped. It roughly corresponds to what the Open Geospatial Consortium (OGC) calls a “geographic feature” (Kottmann and Reed 2009). For the domain of historic maps, the specific cartographic techniques used constrain the range of types of phenomena which are depicted in such maps. Grossner (2010) argues for an event-centric approach and suggests a corresponding spatial history ontology. A semi-structured overview of phenomenon classes is often available in the form *map legends*. We chose concepts needed based on our examples (see Fig. 6) and put them into the *historicmapsphen* ontology.[15] We thereby reused concepts from other geographic ontologies.[16] However, the question of ontological coverage is beyond the scope of this paper and we expect that the phenomenon ontology needs to be extended for each particular collection of maps under consideration.

Different classes of geographic phenomena correspond to conventional *cartographic perspectives* on geographic space, and thus to conventionalized domains of experience. These include geographic objects such as *landforms* (based on the shape and texture or the ground surface), *landcover* (based on the form of vegetation on the ground surface), *landuse* (based on cultivation of the ground surface), *construction* (based on the presence of buildings), *traffic* (based on the presence of

---

[15] Available at http://geographicknowledge.de/vocab/historicmapsphen [.rdf/.jpg], denoted by prefix *phen*.

[16] In fact, our classes cover the classes of geographic kinds suggested by Smith and Mark (2001), which was based on an empirical study. We furthermore imported the ontology for Linking Open Descriptions of Events (LODE) http://linkedevents.org/ontology/.

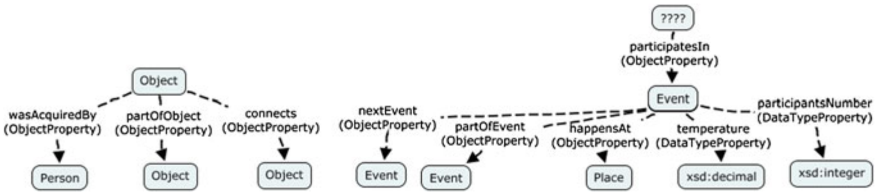


Fig. 7 Relations between phenomena

traffic infrastructure), *place* (based on the presence of named locales, inhabited or not), *administration* (based on the presence of institutionalized entities), as well as *landparcel* objects (objects of ownership).

These objects can *connect*[17] to other objects (e.g. a road connects to some village) and they can be *part of* other objects (such as administrative units) (see Fig. 7).

Furthermore, *events* and their *participants*, the latter including *human* as well as *institutional agents*, take on a central role in historical information. Events *happenAt* places, they can be *part of* other events, and they can be a *member of an event sequence*. Furthermore, all phenomena may have observable properties. A particularly relevant kind of historic event is a *role situation* in which some person takes on some role, e.g., of being a king. The complex relation between a person, his role as a king, his country, and the time of reign can be encoded using Gangemi's *time indexed person role* (Gangemi and Presutti 2009).[18] We included a simplified version of this pattern in our ontology which follows the example given in Sect. 4.1.

We assume that an individual phenomenon may be instantiated by more than one phenomenon class, depending on perspective. For example, a landuse instance such as an agricultural area may at the same time be considered a landcover area (bare), a landform (hill), as well as an object of ownership (landparcel). This corresponds to the common habit of intermixing geographic categories in a map legend, giving rise, e.g., to forest as a type of landuse. We therefore do not assume that the phenomena classes are mutually exclusive.

Space and time of map contents A spatial geometry, denoted by the class *Geometry*, denotes a region, i.e., a subset of point locations in some spatial reference system. We use the *GeoSPARQL ontology*[19] (Battle and Kolas 2012) in order to encode geometries together with their reference system. A geometry has one or more RDF literals which encode its region. We express the geometry as a WKT literal[20] using the GeoSPARQL predicate *asWKT*, which maps from geometries to WKT literals:

---

[17] With this predicate, we express that phenomena are visually connected in the map image, without making any further implications.

[18] http://ontologydesignpatterns.org/wiki/Submissions:Time_indexed_person_role

[19] Available at http://www.opengis.net/ont/geosparql/1.0, prefix *geo*.

[20] A serialization of geometry based on OGC's *simple feature* standard.

$$geo{:}asWKT\ rdfs{:}domain\ geo{:}Geometry. \tag{8}$$

$$geo{:}asWKT\ rdfs{:}range\ geo{:}WKT(Literal). \tag{9}$$

Geographic phenomena have locations. In order to express that a phenomenon is located at a geospatial geometry, we use a sub predicate *where* of the GeoSPARQL predicate *hasGeometry*, which maps from geographic phenomena to their geometric representation:

$$phen{:}where\ rdfs{:}domain\ phen{:}GeoPhenomenon. \tag{10}$$

$$phen{:}where\ rdfs{:}range\ geo{:}Geometry. \tag{11}$$

Maps *always* (per definitionem) represent the location of geographic phenomena. This can be added as a corresponding DL axiom (where we omitted namespaces):

$$GeoPhenomenon \sqcap \exists represents^{-}.Map \sqsubseteq \exists where.Geometry \tag{12}$$

This axiom says that if some geographic phenomenon is represented by a map, then the location of that phenomenon is also represented. Such an axiom can be expressed in OWL and automatically adds blank nodes as geometries of encoded geographic phenomena.[21]

For historic maps, the temporal coverage is as important as the spatial coverage. We denote temporal entities based on *OWL time*,[22] which encodes intervals simply as entities with start and end literals encoded as *xsd:dateTime*.

Maps as documents In the *maps ontology*,[23] we describe maps as image documents which represent windows of space, time as well as spatio-temporally referenced content. For this purpose, we subclass the class *Map* of the *Bibo* ontology,[24] which denotes image documents, add properties such as scale, and reuse *Dublin Core Terms*[25] predicates, such as title, author, medium, size (see Fig. 8). We also linked to *HistoricalMap* in http://gaia.gge.unb.ca/eg/HistoricalMap.owl. Maps represent assertions about geographic phenomena (encoded in terms of named graphs with blank nodes), the phenomena themselves (encoded by direct links), as well as regions in spatial and temporal reference systems. For this purpose, we use the semiotic predicate `maps:represents` and add corresponding sub-properties.

For example, maps directly represent the *spatial area of their map window*, a particular geometry which covers the spatial extent of the map, and which is denoted by a sub-predicate *mapsArea*:

---

[21] Alternatively, one may add corresponding blank nodes by some SPARQL construct.

[22] Available at http://www.w3.org/2006/time; prefix *time*.

[23] Available at http://geographicknowledge.de/vocab/maps [.rdf/.jpg], prefix *maps*.

[24] http://purl.org/ontology/bibo/

[25] http://purl.org/dc/terms/

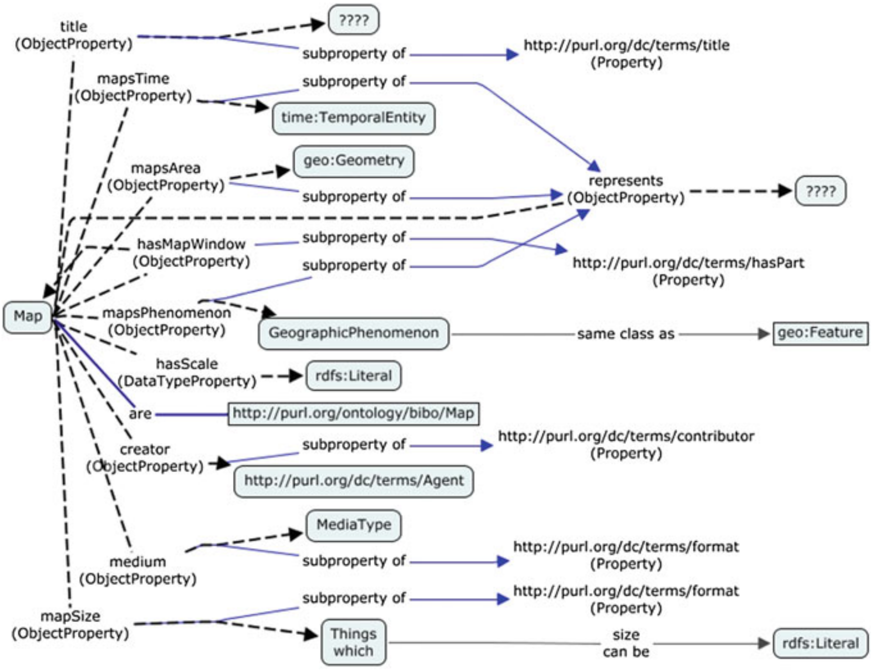


Fig. 8 Ontology of maps as documents

$$maps{:}mapsArea\ rdfs{:}subPropertyOf\ maps{:}represents. \tag{13}$$

$$maps{:}mapsArea\ rdfs{:}range\ geo{:}Geometry. \tag{14}$$

Correspondingly, maps have a *temporal window*, which corresponds to the time which is represented in the map:

$$maps{:}mapsTime\ rdfs{:}subPropertyOf\ maps{:}represents. \tag{15}$$

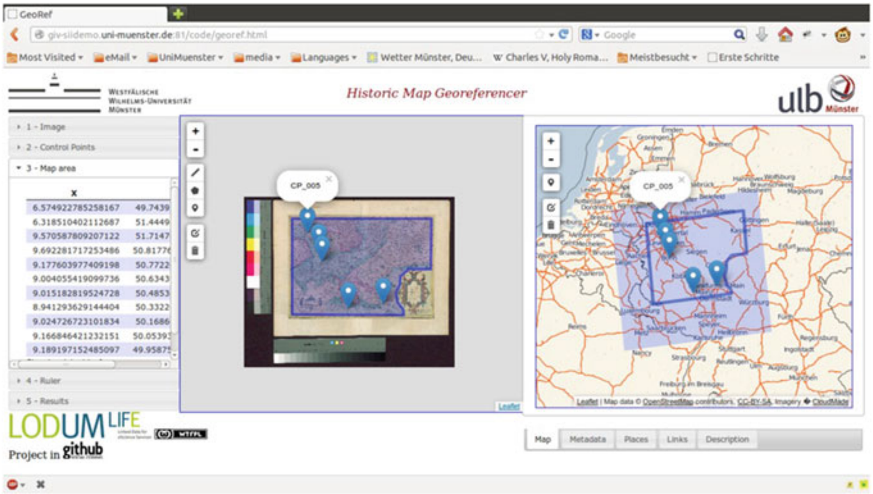
$$maps{:}mapsTime\ rdfs{:}range\ time{:}TemporalEntity. \tag{16}$$

4.3 Georeferencing Historic Maps and Finding Contents

The traditional way to find a historic map, for example in a library, relies on tagging maps with key words and expecting them to match user queries. In contrast to this, our approach consists in spatio-temporal and semantic content descriptions. The combination of semantic descriptions and georeferencing enables the use of explicit historical background knowledge published in the Web, such as DBpedia.[26] We

---

[26] http://dbpedia.org



x	
6.574922785258167	49.7439
6.318510402112687	51.4449
9.570587809207122	51.7147
9.692281717253486	50.8177
9.177603977409198	50.7722
9.004055419099736	50.6343
9.015182819524728	50.4853
8.941293629144404	50.3322
9.024726723101834	50.1686
9.166846421232151	50.0539
9.189197152485097	49.9587

Fig. 9 A georeferencing tool which allows combined spatial, temporal and semantic descriptions of historic maps

have developed a *georeferencing tool* which allows to combine complex content descriptions with georeferencing.[27] The map area and time are determined by the users through *georeferencing* a map image. This information allows the tool to search for entries located inside the map area or space-time window in DBpedia. Since DBpedia does not support GeoSPARQL yet, the spatial component is formulated using the bounding box of the map area. The result is a list of resources potentially related to the map (i.e. modern locations and historical events overlapping the historic map), which is presented to the user, who chooses those resources to be included in the map's description. Furthermore, it allows to encode intensional content in a simple way, as nameless typed entities (see Sect. 4.1). The application is still under development, but it may be turned into a more comprehensive map annotation tool, which allows to extend and reuse content ontologies and to encode arbitrary content in an intensional form (Fig. 9).

A remaining question is how to let users extract content in an automated fashion. One possibility is to use an image vectorization approach. This was employed in the *Map polygon and feature extractor* tool (Arteaga 2013).[28]

5 Encoding the Example Maps

In this section, we present the content encodings of the three example maps using the solutions discussed in Sect. 4. The following vocabularies are used:

```
@prefix maps:<http://www.geographicknowledge.de/vocab/maps#> .
```

---

[27] https://github.com/lodum/georef

[28] https://github.com/NYPL/map-vectorizer

```
@prefix phen:<http://www.geographicknowledge.de/vocab/historicmapsphen#> .
@prefix dbp:<http://dbpedia.org/resource/> .
@prefix dbp-de:<http://de.dbpedia.org/resource/> .
@prefix xsd:<http://www.w3.org/2001/XMLSchema#> .
@prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs:<http://www.w3.org/2000/01/rdf-schema#> .
@prefix time:<http://www.w3.org/2006/time#> .
@prefix sf:<http://www.opengis.net/ont/sf#> .
@prefix geo:<http://www.opengis.net/ont/geosparql#> .
```

We first present descriptions of the map as a document and then discuss the named graph which represents its contents.

5.1 Encoding the Map of Hildesheim

Map document The map of Hildesheim, denoted by the URI +`:4354_Hildesheim`+, is georeferenced and described as a map document in the following listing:

```
:4354_Hildesheim a maps:Map;
    maps:digitalImageVersion :4354_Hildesheim.jpeg;
    maps:represents :hildesheim;
    maps:hasScale "1:28526.1"^^xsd:string;
    maps:mapSize"62.4 * 55.5 cm"^^xsd:string;
    maps:medium maps:Paper;
    maps:mapsTime"1840"^^xsd:gYear;
    maps:mapsArea _:4354_Hildesheim_geom.
    _:4354_Hildesheim_geom a geo:Geometry
    geo:asWKT"<http://www.opengis.net/def/crs/EPSG/0/4326>
POLYGON((9.874690102339652 52.25156096729222, 9.874324681594004
52.126487663211606, 10.07547489355107 52.1268449901813,
10.073392224324136 52.252405987705664, 9.874690102339652
52.25156096729222))"^^sf:wktLiteral.
```

Content graph (*:hildesheim*) Our encoding of the Hildesheim map content makes heavy use of blank nodes for nameless and approximate contents. In this way, it becomes possible to say that villages and buildings and city blocks are shown in the map, the latter being part of the city of Hildesheim, that Hildesheim is connected to some road as well as the river “Innerste”, and that the map also contains wood, pasture, elevations and non-inhabited named places, such as “Steinbergsfeld”.

```
_:someroad a phen:Road ;
    phen:connects _:somevillage ;
    phen:connects dbp:Hildesheim .
_:someroad2 a phen:Footpath .
dbp:Hildesheim a phen:City .
_:somevillage a phen:Village .
_:someblock a phen:Block ;
    phen:partOfObject dbp:Hildesheim .
_:somebuilding a phen:Building .
_:somehill  a phen:Elevation .
```

```
 _:someforest a phen:Wood .
 _:somepasture a phen:Pasture .
 _:someplace a phen:Non-inhabited .
 dbp:Innerste a phen:River ;
     phen:connects dbp:Hildesheim .
...
```

5.2 Encoding Minard's Map

Map document The following triples describe the map and link it to the named graph `:french_invasion+`. We omit further document descriptions as they are equivalent to our first example.

```
:minard_map a maps:Map;
    maps:represents :french_invasion ;
    maps:mapsTime "1812"^^xsd:gYear .
...
```

Content graph (*:french_invasion*) In this map, a logical sequence of nameless stationary war events is visually encoded in terms of a spatio-temporal flow band. Correspondingly, these events are encoded as blank nodes. The map itself tells us nothing about the type of event (e.g. whether it is a battle or just a campaign arrival). The map is also vague about the path taken by the army in between these stationary events. The only information that we can get from the map is the logical event sequence and some of the event properties, such as the temperature during an event, the (remaining) number of people in the campaign, as well as the geographic place (and the time) of the event. The following listing shows an excerpt of this event sequence:

```
_:invasionOfRussia a phen:Invasion .
_:EVENT_13 a phen:War ;
    phen:partOfEvent _:invasionOfRussia ;
    phen:happensAt dbp:Smolensk ;
    phen:participantsNumber "37000"^^xsd:decimal ;
    phen:temperature"-21"^^dbp-de:Raumur-Skala ;
    phen:nextEvent _:EVENT_14 .
_:EVENT_14 a phen:War .
    phen:partOfEvent _:invasionOfRussia ;
    phen:happensAt dbp:Orscha ;
    phen:participantsNumber"24000"^^xsd:decimal ;
    phen:nextEvent _:EVENT_15 .
...
```

5.3 Encoding the Map of Prussia

Map document.

```
:Prussia_map a maps:Map;
    maps:represents :prussia ;
    maps:mapsTime "1688"^^xsd:gYear .
...
```

Content graph (*:prussia*) The following listing encodes the information that Johann Sigismund, a duke of Prussia from 1618–1619, acquired the "Grafschaft Mark", which is a subterritory of Prussia. Note that the situation of Johann Sigismund being the duke of Prussia is not given any name (and thus represented by a blank node, i.e., a variable), and similarly, the time interval of his ruling.

```
dbp:Grafschaft_Mark phen:wasAcquiredBy dbp:Johann_Sigismund ;
    phen:partOfObject dbp:Prussia .
_:EVENT_DUCHY_JS_PREUSSEN a phen:Event ;
    phen:isSettingForPerson dbp:Johann_Sigismund ;
    phen:isSettingForRole dbp:Duke ;
    phen:rulesOver dbp:Prussia ;
    phen:isSettingAtTime _:TIME_INTERVAL_JS_PREUSSEN .
_:TIME_INTERVAL_JS_PREUSSEN a time:TemporalInterval ;
    time:hasBeginning _:INSTANT_BEGINNING_JS ;
    time:hasEnd _:INSTANT_END_JS .
_:INSTANT_BEGINNING_JS a time:Instant ;
    time:inXSDDateTime "1618"^^xsd:gYear .
_:INSTANT_END_JS a time:Instant ;
    time:inXSDDateTime"1619"^^xsd:gYear .
...
```

This is continued for all subterritories shown in the map.

6 Querying Historic Map Contents

If historians search in map collections, they are primarily interested in *finding maps which help answer their question*. While this is currently done through searching for maps which have certain properties, acting as a proxy, the former is a much more general problem. We translate this question into the following *map content query*, expressed in SPARQL syntax and based on our encoding scheme:

```
SELECT ?map ... WHERE {
  ?map maps:represents ?g .
  GRAPH ?g { ...   }
}
```

This query can be translated as: *which maps represent content which contains statements of the following form* (i.e., *answering the graph pattern*)...? The historian's question to be answered by the map is posed in terms of the graph pattern. The encoded content (as far as it was made explicit) can also be retrieved along with the map. Note that map content graphs may be intensional, and thus do not have to contain an answer. Still, a triple store can deliver a map which contains an answer if the map's intensional description (including blank nodes) satisfies the pattern. That is, even though the answer might not be explicitly encoded, knowledge about whether the map contains the answer can be used to automate map selection. We illustrate this mechanism in the following by our examples. All of these queries were tested on a standard triple store using a standard reasoner.[29]

Where was Hildesheim in 1840?

```
SELECT DISTINCT ?map ?where WHERE {
    ?map maps:represents ?g ;
         maps:mapsTime "1840"^^xsd:gYear .
    GRAPH ?g {{dbp:Hildesheim ?p ?o}UNION{?a ?d dbp:Hildesheim}
    }
    dbp:Hildesheim phen:where ?where .
}
```

This query simply retrieves maps which represent the location of Hildesheim in 1840. The `UNION` keyword, a logical or, allows detecting Hildesheim in a map's content regardless of whether it is subject or object of an assertion. As mentioned in Sect. 4.1, all maps which represent Hildesheim link to it directly, and by Eq. 12, Hildesheim therefore must have some location. Thus, the query delivers a meaningful result, even though the geometry of Hildesheim was never explicitly encoded.

What were the types of landcover around Hildesheim in 1840?

```
SELECT DISTINCT ?map ?class WHERE {
   ?map maps:represents ?g ;
        maps:mapsTime "1840"^^xsd:gYear .
   GRAPH ?g {{dbp:Hildesheim ?p ?o}UNION{?a ?d dbp:Hildesheim}
      ?instance a ?cl .
   }
    ?instance a ?class .
    ?class rdfs:subClassOf phen:Landcover.
}
```

In this query, the intended types of landcover (`?class`) are those subclasses of `phen: Landcover` for which we know that there is some instance of this class depicted in a map which shows Hildesheim in 1840. We are not interested in these instances as such, only in their classes. And in fact, for this query to deliver results, the content graph does not have to contain any actual instances. In fact, they are all

---

[29] OWLIM-Lite 4.0 with OWL-Horst reasoner (optimized) and OpenRDF workbench version 2.7.7. The endpoint is available at http://data.uni-muenster.de:8080/openrdf-sesame/repositories/mapcontents.

blank nodes. Note that since reasoning is done based on the ontology, the landcover classification pattern needs to be located outside the content graph. In this way, an RDFS reasoner will be able to automatically classify all instances which are landcover, regardless of whether this was made explicit in the content graph.

How many people did Napoleon's army have when soldiers arrived in Smolensk during his 1812 campaign?

```
SELECT ?map ?soldiers WHERE {
   ?map maps:represents ?g .
   GRAPH ?g {
      ?event phen:partOfEvent dbp:French_invasion_of_Russia ;
             phen:happensAt dbp:Smolensk ;
             phen:participantsNumber ?soldiers .
   }
}
```

In order to answer this question, we need to find the respective subevent. Only if this subevent and the corresponding property is depicted in the map, we can be sure that the map serves to answer that question.

Very similarly, this query allows to retrieve a minimal property of all subevents. *What were the lowest temperatures during Napoleon's campaign*?[30]

```
SELECT (MIN(?temperature) AS ?temp) WHERE {
   ?map maps:represents ?g .
   GRAPH ?g {
      ?event phen:partOfEvent dbp:French_invasion_of_Russia ;
             phen:temperature ?temperature .
   }
}
```

Which places did Napoleon's army come across during the 1812 campaign?

```
SELECT DISTINCT ?map ?places WHERE {
   ?map maps:represents ?g .
   GRAPH ?g {
      ?event phen:partOfEvent dbp:French_invasion_of_Russia ;
             phen:happensAt ?places. }
}
```

For the Prussia map, we need to make use of object parthood relations in order find answers:

Which territories were part of Prussia in 1783?

```
SELECT ?map ?parts WHERE {
   ?map maps:represents ?g ;
        maps:mapsTime "1783"^^xsd:gYear .
```

---

[30] Note: Due the fact that the temperature uses a literal dbp-de:Reaumur-Skala, the SPARQL function MIN does not work without additional effort.

```
   GRAPH ?g {
      ?parts phen:partOfObject dbp:Prussia .
   }
}
```

Which Prussian territories were acquired by Friedrich Wilhelm, the great elector?

```
SELECT ?map ?parts WHERE {
   ?map maps:represents ?g .
   GRAPH ?g {
      ?parts phen:partOfObject dbp:Prussia .
      ?parts phen:wasAcquiredBy dbp:Friedrich_of_Brandenburg .
   }
}.
```

7 Conclusion

In this chapter, we argued that the contents of historic maps can be encoded and efficiently queried in terms of named RDF graphs with blank nodes. This allows to formally represent map content in terms of a set of triple assertions and to link the description of the map as a document with its content in an explicit way. It furthermore makes possible intensional descriptions of map content, in which parts of the content are not made explicit while still being useful to inform about the types and relations depicted in a map. This solution allows to logically express explicit contents to different degrees: Complex spatio-temporal content can be encoded in great detail, going well beyond simple links from maps to historic entities. Named phenomena can be linked to external knowledge sources on the Web, while nameless or implicit content can be encoded in an intensional manner using blank nodes. All this may help libraries in encoding map content, and it allows to retrieve historic maps based on specific content-related questions that historians have when searching for maps, instead of keywords or simple map properties.

We discussed corresponding encoding schemes, vocabularies and tools. We furthermore tested encodings as well as map content queries on a standard triple store, based on a list of competency questions about three historic map examples.

Future work should focus on the following aspects: Which tools would support users in encoding complex map contents without deeper knowledge of Semantic Web technology? The georeferencing tool presented in this chapter needs to be extended in order to generate complex content descriptions such as named graphs. And correspondingly, which tools would allow library users to formulate arbitrary content queries and hence go beyond browsing and exploring (Simon et al. 2012)? Annotation quality and usability of these tools needs to be evaluated in a library context. It will require a community of interested volunteers to generate content descriptions for a significant portion of a given map collection. Furthermore, the proposed solution for encoding and reasoning needs to be further developed, based on actual information needs of historians. It would, e.g., be useful to hide blank nodes from users and to provide frequent query patterns as templates. It remains also unclear which sort of reasoning is needed to deal with intensional map contents in general.

Acknowledgments This work has been funded by the German Research Foundation (DFG) through the *Linked Data for eScience Services* (LIFE) project (KU 1368/11-1). We also would like to thank our project partners, the Münster University Library (ULB) and the Institute for comparative urban history (ISTG) for their constant support of this work.

References

Arteaga MG,(2013) Historical map polygon and feature extractor. In: MAPINTERACT'13, Orlando, FL, USA. ACM, New York, NY, 05–08 Nov 2013

Battle R, Kolas D (2012) Enabling the geospatial semantic web with parliament and GeoSPARQL. Semantic Web 3(4):355–370

Bizer C, Heath T, Berners-Lee T (2009) Linked data: the story so far. Int J Semantic Web Inf Syst 5(3):1–22

Carral D, Scheider S, Janowicz K, Vardeman C, Krisnadhi A, Hitzler P (2013) An ontology design pattern for cartographic map scaling. In: Cimiano P, Corcho O, Presutti V, Hollink L, Rudolph S (eds) The semantic web: semantics and big data. Lecture notes in computer science, vol 7882. Springer, Berlin, pp 76–93

Gangemi A, Presutti V (2009) Ontology design patterns. In: Staab S, Studer R (eds) Handbook on ontologies, international handbooks on information systems. Springer, Berlin, pp 221–243. doi:10.1007/978-3-540-92673-3_10

Gkadolou E, Stefanakis E (2013) A formal ontology for historical maps. In: 26th international cartographic conference

Gkadolou E, Tomai E, Stefanakis E, Kritikos G (2013) Ontological standardization for historical map collections: studying the Greek borderlines of 1881. In: ISPRS annals of the photogrammetry, remote sensing and spatial, information sciences, vol I-2

Grossner K (2010) Representing historical knowledge in geographic information systems. Ph.D. thesis, University of California, Santa Barbara

Hart G, Dolbear C (2013) Linked data: a geographic perspective. CRC Press, Boca Raton

Haslhofer B, Robitza W, Guimbretiere F, Lagoze C (2013) Semantic tagging on historical maps. In: Proceedings of the 5th annual ACM web science conference, WebSci '13. ACM, New York, NY, USA, pp 148–157

Hyvönen E, Tuominen J, Kauppinen T, Väätäinen J (2011) Representing and utilizing changing historical places as an ontology time series. In: Ashish N, Sheth AP (eds) Geospatial semantics and the semantic web, semantic web and beyond, vol 12. Springer, New York, pp 1–25

Kottmann C, Reed C (2009) The OpenGIS abstract specification. Topic 5: features

Kraak MJ (2003) Geovisualization illustrated. ISPRS J Photogrammetry Remote Sens 57(56): 390–399

Krötzsch M, Simancik F, Horrocks I (2012) A description logic primer. CoRR abs/1201.4089

MacEachren AM (2004) How maps work: representation, visualization, and design, 2nd edn. The Guilford Press, New York

Montello D (1993) Scale and multiple psychologies of space. In: Frank AU, Campari I (eds) Spatial information theory a theoretical basis for GIS. Lecture notes in computer science, vol 716. Springer, Berlin, pp 312–321

Ruotsalo T, Haav K, Stoyanov A, Roche S, Fani E, Deliai R, Mäkelä E, Kauppinen T, Hyvönen E (2013) SMARTMUSEUM: a mobile recommender system for the web of data. Web Semant Sci Serv Agents World Wide Web 20:50–67

Simon R, Haslhofer B, Robitza W, Momeni E (2011) Semantically augmented annotations in digitized map collections. In: Proceedings of the 11th annual international ACM/IEEE joint conference on digital libraries, JCDL '11. ACM, New York, NY, USA, pp 199–202

Simon R, Barker E, Isaksen L (2012) Exploring Pelagios: a visual browser for geo-tagged datasets. In: International workshop on supporting users' exploration of digital libraries. Paphos, Cyprus, 23–27 Sept 2012

Smith B, Mark D (2001) Geographic categories: an ontological investigation. Int J Geog Inf Sci 15(7):591–612

Trame J, Keßler C, Kuhn W (2013) Linked data and time: modeling researcher life lines by events. In: Tenbrink T, Stell J, Galton A, Wood Z (eds) Spatial information theory. Lecture notes in computer science, vol 8116. Springer International Publishing, pp 205–223. doi:10.1007/978-3-319-01790-7_12