Designing a Map Legend Ontology For Searching Map Content

Song Gao[1], Krzysztof Janowicz[1], Dingwen Zhang[2]

[1] STKO Lab, University of California, Santa Barbara, USA
[2] Harbin Institute of Technology Shenzhen Graduate School, China

Abstract. Map legends are keys to the understanding of symbols used on maps. Without such legends and the knowledge to interpret them, maps are reduced to mere pictures. While it is possible to semi-automatically extract text, contour lines, major transportation infrastructure, and so forth from maps, more complex aspects and the relation between symbols get lost. Even if these features get accurately extracted, they are mostly used to derive digital representations of (old) analog maps. From an information retrieval perspective, facts such as that a certain map contains transportation features organized in a hierarchy of highways, streets, trails, and so forth, remain hidden and therefore can neither be used by machines nor humans to enable a richer search for map contents. In this work, we formalize a first version of a map legend ontology (MLO) that can be used to semantically annotate and query map contents via their legend in a machine-readable manner using Semantic Web and Linked Data technologies. To demonstrate our results, we introduce three different examples of real-world map legends and show how to represent them by using the map legend ontology.

Keywords: Map Legend, Symbology, Ontology Design Pattern

1 Introduction

When most people think about maps, they have topographic maps in mind or even merely navigation-oriented mapping services such as Google Maps. However, there are dozens of map types that can be structured and organized in different ways. Common map types include, topological maps, thematic maps, topographic maps, cadastral maps, navigational charts, isochrone maps, you-are-here maps, geological maps, pictorial maps, flow maps, and even cartograms. In fact, most of the impactful maps that we see today are not topographic but various kinds of thematic maps. Maps depict (attributed of) places, events, and objects, as well as their relationships by using symbols, including images, point markers, lines, polygons, colors, textures, and so forth. The used symbology varies greatly across maps even if they depict similar information and similar areas. Consequently, a dictionary is needed to understand what is meant by the used shapes, textures, sizes, and so on. This is where map legends come into play. Simply speaking, legends associate symbols (and variables) with terms.

They also provide additional structure, e.g., by grouping symbols into hierarchies. Without a legend and the knowledge to interpret it, maps become mere pictures and thereby lose most of their value. While some semi-automated techniques can be used to extract text from maps or to trace lines and features, e.g., to digitize old paper maps [13], a fully automated semantic annotation of maps has not been achieved to date. As a result, maps are not searchable, i.e., they have to be retrieved by contextual clues such as surrounding text, instead of being able to query their content directly. Although recent work by Scheider et al. [15] investigates how to encode and query a collection of historical maps on the Semantic Web, to the best of our knowledge, no ontology design patterns have been designed that would enable an extendible description of map legends. Such a map legend ontology design pattern should be able to answer queries such as:

- What are the common symbols to represent highway transportation systems?
- Which maps show places with a population density larger than 1000 people per square miles?
- Which maps contain both Ski Areas and Camping Areas?
- How many thematic maps depict crime rates in US cities?

2   Related Work

Several efforts have been made to annotate and publish the content of paper maps on the Web. Traditional methods employ simple metadata scheme including titles, authors, keywords and (standardized) topics, the year of production, a bounding box of the covered area, and so forth. Non of these enable to query map content directly. Georeferenced libraries such as the Alexandria Digital Library (ADL) support geospatial indexing and searching for maps, imagery, and other library resources by using digital gazetteers to identify geographic location by coordinates, place names, and feature types [4,8]. Hardy and Durante [5] designed a metadata scheme to support geospatial resource discovery and facilitate inter-institutional sharing by reusing well-established schemata such as *Dublin Core* and *GeoRSS*.

More recently, researchers started to investigate how to semantically annotate both paper maps and digital map collections [7,16]. For example, Scheider et al. [15] presented an approach using Semantic Web technologies to encoding and querying historic maps in a machine-readable form. They also linked named phenomena in those historical maps to external knowledge sources on the Web, e.g., DBpedia. Haslhofer et al. [6] demonstrated how to provide annotations on historic maps and augment content with links to contextually relevant resources on the Web. Gkadolou and Stefanakis [3] proposed a geo-ontology and demonstrated how historical maps could be represented via the CIDOC Conceptual Reference Model. Carral et al. [2] developed an ontology design pattern to describe cartographic map scaling with regard to semantic relationships among geometric representations, phenomena, and map-scale levels. Recently, Hu et al. [10,11] combined data-driven techniques with theory-informed approaches to

enable semantic search and knowledge discovery for the leading Web GIS portal — Esri's ArcGIS Online. More specifically, they developed an ontology for ArcGIS Online maps, converted the metadata into Linked Data, and enriched the metadata by learning ISO topics and geographic entities from titles and nature language descriptions using Labeled Latent Dirichlet Allocation (LLDA).

Very few studies have focused on map legend ontologies. One exception is the work done by Roula et al. [14]. They proposed to encode information about symbols including icons, color, texture, font style, orientation, and so forth using *CartOWL*. Our research has a different focus, namely on map *content* and not *visual* aspects.

3 Map Legend Ontology Design Pattern

In this section, we introduce the proposed ontology design pattern for modeling map legends[3].

3.1 Conceptualization

As depicted in Fig.1, a map legend ontology (MLO) can be conceptualized via a few classes and relations between them. Our ontology design pattern for map legends can also be integrated and re-used by mapping or content search ontologies.

- A *Map* has at least one legend *MapLegend*.
- A *MapLegend* consists of at least one *LegendItem*. These items can also be nested, i.e., a legend item may contain another legend item.
- A *LegendItem* consists of at least one symbol and is used to describes how a specific group of map features or a feature type is shown on the map.
- A *Symbol* can be an image/figure, textual label, scalebar, and so forth that depicts a specific map element or a feature type (such as mountain, park, and city). Note that in a map showing population by county, e.g., by using a color ramp, *County* would be the feature type. Symbols are labeled.
- Symbols can be further subdivided into *GraduatedColorSymbol*, *GraduatedSizeSymbol*, and *CategorizedSymbol*. Each of these subclasses may consist of multiple symbols as well.
- A *GraduatedColorSymbol* usually describes a choropleth map, i.e., a thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map, such as population density or per-capita income (See example in Fig.2).
- A *GraduatedSizeSymbol* usually quantifies an attribute/variable of the data in which the sizes of shapes, e.g., circled, are scaled in proportion to the measurement being displayed on the map, such as population size or disease cases (See example in Fig.3).

---

[3] Our map legend ontology axiomatization can be downloaded at `http://stko-exp.geog.ucsb.edu/mlo/map_legend_ontology.owl`
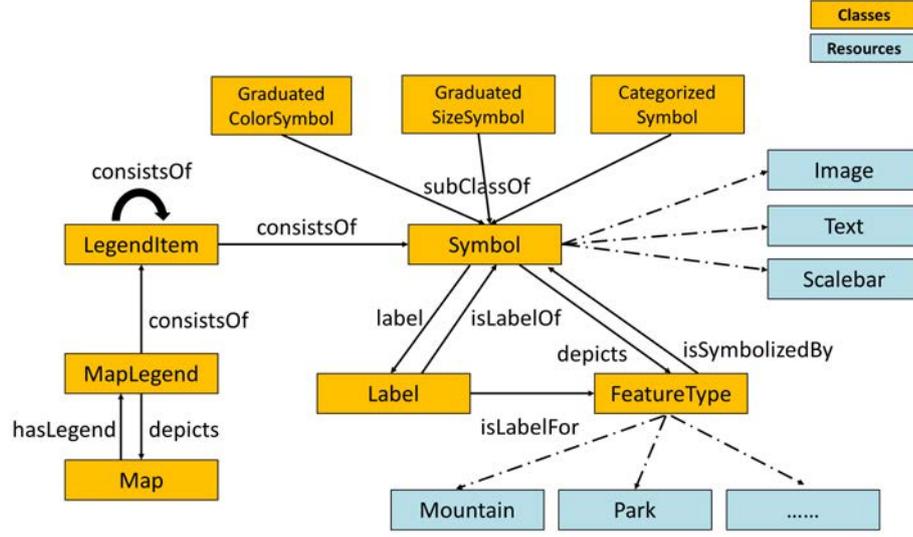
Fig. 1. Concept map of the map legend ontology (MLO).

- A *CategorizedSymbol* usually contains different subclasses of a geographic feature type, such as a road hierarchy. (See example in Fig.4).
- Maps can depict multiple *FeatureTypes*. These can include types such as river, mountain, administrative unit, and so forth. These types are not defined in MLO but can be defined or imported from other sources, e.g., the ADL feature type thesaurus.

3.2   Axiomatization

In this section we present selected axiomatization of the Map Legend Ontology. Based on the outlined conceptualization given above, we formally define the classes N_C and the relations N_R among these classes.

$$N_C = \{Map, MapLegend, LegendItem, Symbol, Label, FeatureType,$$
$$GraduatedColorSymbol, GraduatedSizeSymbol, CategorizedSymbol\} \quad (1)$$

$$N_R = \{consistsOf, hasLegend, isLabelOf, isLabelFor, depicts,$$
$$isSymbolizedBy, showsFeaturesOfType\} \quad (2)$$

Legend items are composed of symbols and/or other legend items that in turn may structure additional symbols.

$$LegendItem \sqsubseteq \exists consistsOf.Symbol \sqcup \exists consistsOf.LegendItem \quad (3)$$
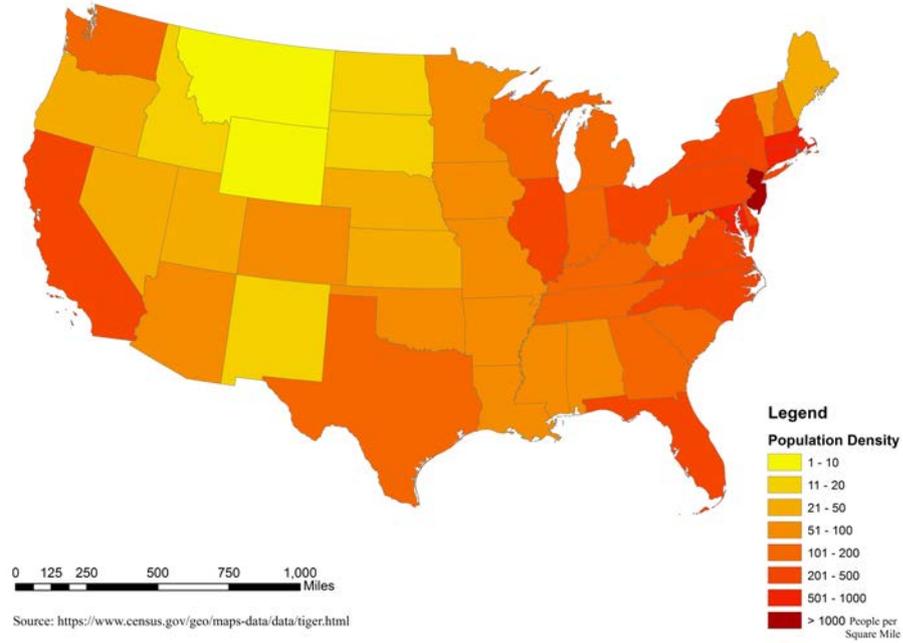
Fig. 2. An example of GraduatedColorSymbol for representing the population density in the contiguous United States.

Feature types are symbolized by symbols, e.g., parking spaces can be symbolized by a parking icon.

$$FeatureType \sqsubseteq \exists SymbolizedBy.Symbol \sqcap \forall SymbolizedBy.Symbol \quad (4)$$

Labels provide textual description of symbols or name feature types.

$$Label \sqsubseteq \exists isLabelOf.Symbol \sqcup \exists isLabelFor.FeatureType \quad (5)$$

There are multiple types of symbols, three of them are shown here.

$$GraduatedColorSymbol \sqsubseteq Symbol \quad (6)$$

$$GraduatedSizeSymbol \sqsubseteq Symbol \quad (7)$$

$$CategorizedSymbol \sqsubseteq Symbol \quad (8)$$

Next, we enforce functionality for the properties *isLabelFor*, *isLabelOf*, and *SymbolizedBy*.

$$\top \sqsubseteq\, \leq 1.isLabelFor.\top \quad (9)$$

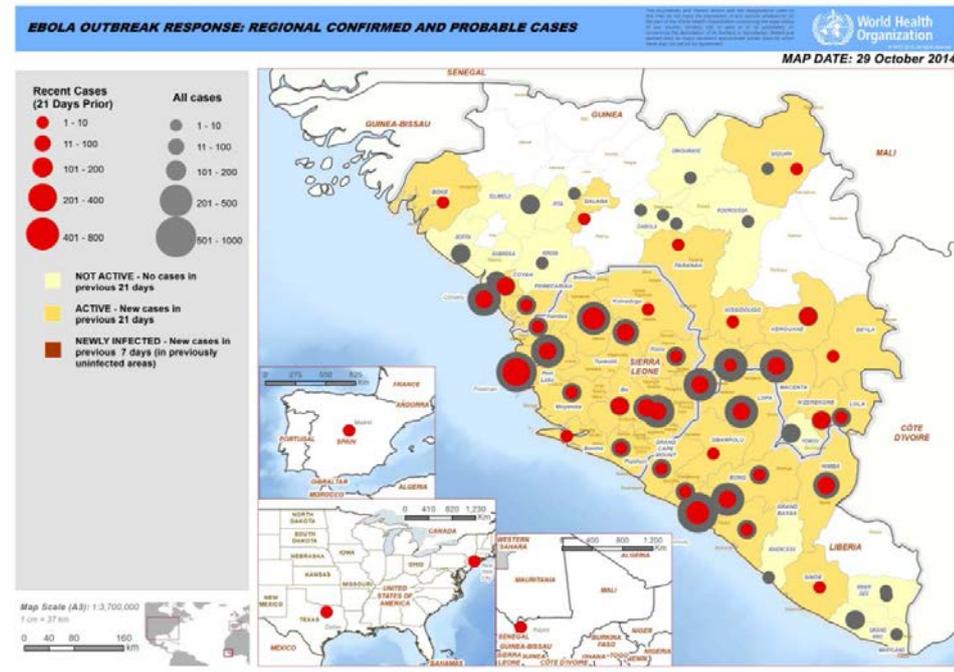$$\top \sqsubseteq\, \leq 1.isLabelOf.\top \quad (10)$$

Fig. 3. An example of GraduatedSizeSymbol for quantifying ebola disease cases on the map (Source: World Health Organization).

$$\top \sqsubseteq\ \leq 1.SymbolizedBy.\top \tag{11}$$

A legend is not contained by anything else (while legend items can be consist of other legend tiems).

$$\neg \exists consistsOf^- \sqsubseteq MapLegend \tag{12}$$

A legend consists of at least one legend item.

$$MapLegend \sqsubseteq \exists consistsOf.LegendItem \tag{13}$$

If a map has a legend that contains an item that shows a symbol for a certain type of feature, then the map contains features of the said type.

$$hasLegend \circ consistsOf \circ consistsOf \circ depitcs \sqsubseteq showsFeaturesOfType \tag{14}$$

4   Examples

In this section, we highlight three different types of map legends and show how to encode them with the proposed ontology.
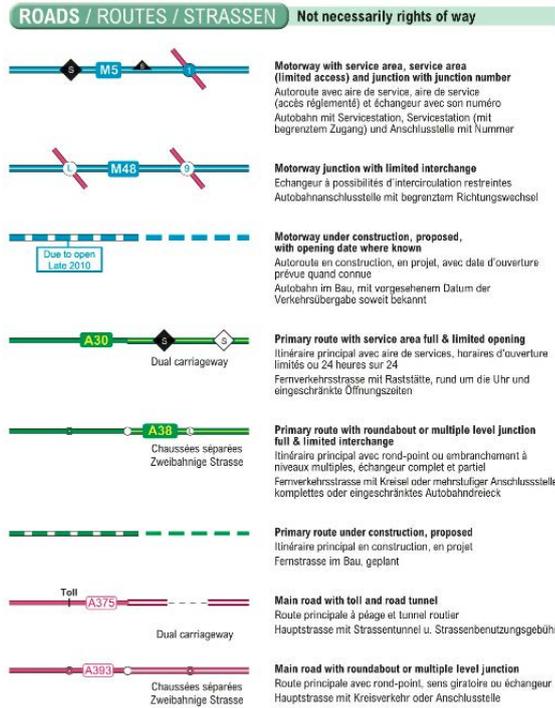
Fig. 4. An example of CategorizedSymbol for different road classes from Ordnance Survey Open Data. (Source: https://commons.wikimedia.org/wiki/File:Ordnance_Survey_1-250000_Legend_2010.tif)

4.1 Example 1

Let us start with a simple map legend as shown in Fig.5. This map legend only contains one legend item with multiple symbols. All symbols except Highways are single images with text labels and can be modeled using the hierarchical resource and property statements. A symbol which depicts a geographic feature type can also be linked to other SPARQL endpoints which contain these features types, such as DBpedia. For instance, the symbol for *Cities & Towns* can be represented as follows:

```
ex:Map1  mlo:hasLegend ex:Legend1.
ex:Legend1 mlo:consistsOf ex:LegendItem1.
ex:LegendItem1 mlo:consistsOf ex:CaTSymbol.
ex:CaTSymbol mlo:depicts ex:CityTown.
...
```

The feature type *Highway* should be encoded using the symbol subclass *CategorizedSymbol* which consists of three single symbols: *Interstate*, *State Highway*,

Fig. 5. Map legend example 1.

and *Federal Highway.* Each of these subclass symbols can be encoded in the same way as other single feature types with images and textual labels.

4.2   Example 2

In the second example (see Fig.6), the map legend contains multiple legend items which are used to describe different themes, including *road classifications*, *population of cities*, *speed limit*, *scale of miles*, *highway markers*, and *map symbols.* Each legend item consists of one symbol such as the *speed limit* or multi-symbols such as the *map symbols*, which consist of categorized symbols, e.g., four types of airports. Both legend items *road classifications* and *highway markers* can be represented using *CategorizedSymbol*, while the *population of cities* can be represented by *GraduatedSizedSymbol.* Note that our MLO allows to represent and structure the symbols in different ways and using multiple legend items but it does not reflect the order of the legend as this is a design and not a content issue.

4.3   Example 3

The third example (see Fig.7) is a geological map legend which also contains mutli-legend items. The difference of this example is that the primary part of the legend is actually showing three volcanic centers as subgroups and types of geologic composition structure. They can be modeled using *CategorizedSymbol* with varying colors. In addition, this map legend also has other items at the bottom of the legend frame, showing several single symbols to represent different geographic feature types (*FeatureType*) such as *mountain peak (Sommet), village, and international border (Frontière internationale).*

5   SPARQL Query

In this section, we briefly demonstrate how the presented map legend ontology can support interesting map content queries. The map legend example data for

Fig. 6. Map legend example 2. (Credit to Source: http://kids.britannica.com/comptons/art-53621)

the SPARQL queries shown below can be downloaded from `http://www.geog.ucsb.edu/~sgao/maplegend/data.rdf`.

Query 1: Which map legends contain highway information?

```
Select ?legend ?label
   where{
      ?legend mlo:consistsOf ?s .
      ?s mlo:label ?label .
      FILTER ( regex(?label, "Highway" ))
   }
  # Note that this query broader than a feature type query.
```

Query 2: Which maps contain both ski areas and camping areas?

```
 Select ?legend ?legenditem ?label
    where{
      ?legend mlo:consistsOf ?legenditem .
```
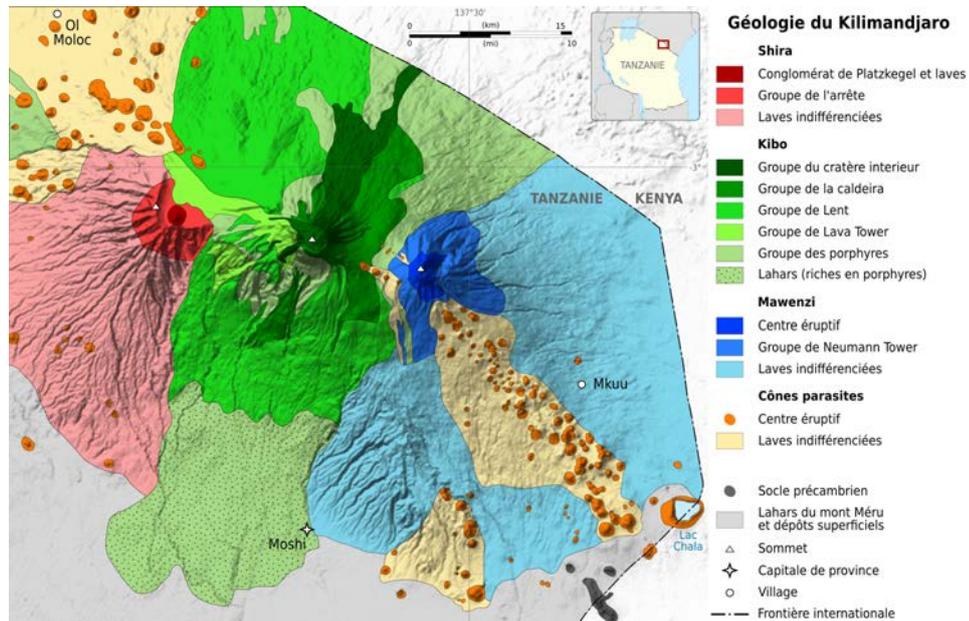
Fig. 7. Map legend example 3. (Credit to Wikimedia Commons user: Sémhur, Source: https://commons.wikimedia.org/wiki/File:Mount_Kilimanjaro_Geology_map-fr.svg)

```
    ?legenditem mlo:consistsOf ?s .
    ?s mlo:label ?label .
    FILTER (regex (?label, "Ski" )
    || regex (?label, "Camp" ))
}
```

Query 3: What are the common symbol resources to represent highway transportation systems?

```
select ?legend ?s ?imageURI
   where{
    ?legend mlo:consistsOf ?s .
    ?s mlo:depicts <http://dbpedia.org/page/Highway> .
    ?s mlo:hasImage ?imageURI
   }
   # This example uses feature types (from an external resource) directly.
```

6   Conclusion and Future Work

In this work, we designed a first version of a map legend ontology that can be used to annotate and query map *content* via legends in a machine-readable manner using Semantic Web and Linked Data technologies. We have shown multiple

examples demonstrating how the map legend ontology design pattern can be used. We have also shown test queries to demonstrate how the pattern can be queried. To ensure that the developed pattern is not overly specific and thus remains reusable, we have not introduced specific types of geographic features but leave this to external ontologies and vocabularies. We have also not introduced attributes/variables nor units although we realize that they may have to be introduced in a future version to support specific types of maps or richer queries.

In future work, we will apply this ontology for annotating more maps with legends. The challenging part of this work is to automate the process of annotating existing map legends. It might involve two approaches. One is to motivate the general public or map creators to contribute the structured content, i.e., user-generated content. The other is to integrate image recognition technology to extract the symbols with textual labels for the encoding purpose.

References

1. Bizer, C., Heath, T., Berners-Lee, T.: Linked data-the story so far. International journal on semantic web and information systems 5(3), 1–22 (2009)
2. Carral, D., Scheider, S., Janowicz, K., Vardeman, C., Krisnadhi, A.A., Hitzler, P.: An ontology design pattern for cartographic map scaling. In: Extended Semantic Web Conference. pp. 76–93. Springer (2013)
3. Gkadolou, E., Stefanakis, E.: A formal ontology for historical maps. In: Proceedings of the 26th international cartographic conference (2013)
4. Goodchild, M.F.: The alexandria digital library project review, assessment, and prospects. Trends in Information Management (TRIM) 1(1) (2012)
5. Hardy, D., Durante, K.: A metadata schema for geospatial resource discovery use cases. Code4lib Journal 25 (2014)
6. Haslhofer, B., Momeni, E., Gay, M., Simon, R.: Augmenting europeana content with linked data resources. In: Proceedings of the 6th International Conference on Semantic Systems. p. 40
7. Haslhofer, B., Robitza, W., Guimbretiere, F., Lagoze, C.: Semantic tagging on historical maps. In: Proceedings of the 5th Annual ACM Web Science Conference. pp. 148–157. ACM (2013)
8. Hill, L.L.: Georeferencing: The Geographic Associations of Information. The MIT Press (2006)
9. Hitzler, P., Krotzsch, M., Rudolph, S.: Foundations of semantic web technologies. CRC Press (2009)
10. Hu, Y., Janowicz, K., Prasad, S., Gao, S.: Enabling semantic search and knowledge discovery for arcgis online: A linked-data-driven approach. In: AGILE 2015, pp. 107–124. Springer (2015)
11. Hu, Y., Janowicz, K., Prasad, S., Gao, S.: Metadata topic harmonization and semantic search for linked-data-driven geoportals: A case study using arcgis online. Transactions in GIS 19(3), 398–416 (2015)
12. Janowicz, K., Scheider, S., Pehle, T., Hart, G.: Geospatial semantics and linked spatiotemporal data–past, present, and future. Semantic Web 3(4), 321–332 (2012)
13. Leyk, S., Boesch, R., Weibel, R.: Saliency and semantic processing: Extracting forest cover from historical topographic maps. Pattern recognition 39(5), 953–968 (2006)

14. Roula, K., Franck, F., Robert, L., Rima, C.: Integrating many legends through ontology for cartographic symbols. In: Proceedings of 6th International Conference SAGEO2010 (2010)
15. Scheider, S., Jones, J., Sánchez, A., Keßler, C.: Encoding and querying historic map content. In: Connecting a Digital Europe Through Location and Place, pp. 251–273. Springer (2014)
16. Simon, R., Haslhofer, B., Robitza, W., Momeni, E.: Semantically augmented annotations in digitized map collections. In: Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries. pp. 199–202. ACM (2011)