A Linked Open Data Platform for Historical Geographic Data

Görz, Günther
guenther.goerz@fau.de
FAU Erlangen-Nürnberg, Informatik, Deutschland; Bibliotheca Hertziana, Rom, Italien

Seidl, Chiara
chiara.seidl0510@gmail.com
FAU Erlangen-Nürnberg, Informatik, Deutschland; Bibliotheca Hertziana, Rom, Italien

Thiering, Martin
pd_dr_Martin_Thiering@posteo.de
TU Berlin, Deutschland; Bibliotheca Hertziana, Rom, Italien

The goal of Bibliotheca Hertziana's project "Historical spaces in texts and maps" is to investigate the relations between historic geographical texts and maps to reconstruct a historical understanding of space and the knowledge associated with it. Starting with a cognitive-semantic analysis of Flavio Biondo's "Italia Illustrata" (1474), first of all, toponyms, place descriptions and spatial relations are annotated in the text and Renaissance maps. Our contribution to Spatial Humanities is based on the conviction that all maps are cognitive maps, depicting culture-specific spatial knowledge and practices (Goerz et al., 2018, 2019).

In general, our research combines cognitive-semantic parameters such as toponyms, landmarks, spatial frames of reference, geometric relations, gestalt principles and different perspectives with computational linguistic analysis (Thiering, 2015). We designed a workflow comprising the steps of transcription, annotation, geographic verification, export and ontology-based semantic enrichment of these data, finally stored and published as Linked Open Data. We use Recogito (https://recogito.pelagios.org , 15.09.2019) as our main tool for static annotations of places and persons/peoples in text and maps. Toponyms are georeferenced with gazetteers, in our case primarily with Pleiades (https://pleiades.stoa.org , 15.09.2019), and the annotations can be exported in various formats, in particular, CSV, GeoJSON, and KML. Spatial relations in texts are annotated in terms of the cognitive-semantic parameters by means of the brat tool. These annotation data are mapped into triples encoding cognitive parameters, primarily in "figure-spatial_relation-ground" constructions. Furthermore, dependency parsing (http://ufal.mff.cuni.cz/udpipe , 15.09.2019) has been applied to the text for comparison. To achieve a generic semantic level for linguistic and map-related annotations, we perform a transition to an ontology-based representation. For this purpose, we defined a domain ontology *hmap* for historical maps and geographical texts based on the event-centered CIDOC Conceptual Reference Model (CRM, ISO standard 21127) and its spatio-temporal extension CRMgeo in OWL-DL (http://erlangen-crm.org , 15.09.2019). Using the CRM opens up a wide spectrum of interoperability and linking to many web resources. The domain ontology *hmap* for the description of historical maps and their content offers a framework for the general metadata of maps and geographical texts as well as for descriptions of their content.

As Linked Data platform we chose the Virtual Research Environment WissKI (Scholz et al., 2016; http://wiss-ki.eu , 15.09.2019), a semantic database extension of the CMS Drupal, in which we defined our data model in terms of so-called ontology paths. These are sequences of triples built from entities and properties of the ontology. As an example, in a map production event (*hmap:M9_Map_Production*) there is an actor, the Creator, defined by

```
hmap:M28_Map --> hmap:A3i_was_produced_by
--> hmap:M9_Map_Production --
> hmap:A4_carried_out_by_map_author
--> hmap:M1_Map_Author -->
ecrm:P131_is_identified_by
--> ecrm:E82_Actor_Appellation.
```

For each map we may have several images, in which depicted objects are annotated; so there is an analogous data model for images (*hmap:M34_Image*). What the image depicts, in our case annotated places, is specified by

```
hmap:M34_Image --> hmap:A43_depicts
--> hmap:M3_Annotated_Place --
> ecrm:P1_is_identified_by -->
ecrm:E42_Identifier.
```

For each annotated place (*hmap:M3_Annotated_Place* is a subclass of *crmgeo:SP6_Declarative_Place*) where the (geographical) contents of the annotations are encoded in the columns of the CSV tables, each column is transformed into a component for which similar ontology paths are defined. The annotated place is linked to the image by

```
hmap:M3_Annotated_Place -->
hmap:A43i_is_depicted by --> hmap:M34_Image
--> ecrm:P48_has_preferred_identifier -->
ecrm:E42_Identifier.
```

So, e.g., for *QUOTE_TRANSCRIPTION*, the path is

```
hmap:M3_Annotated_Place -->
ecrm:P87_is_identified_by -->
hmap:M42_Transcribed_Place_Appellation
```

Each annotation, represented as a row in the table, has a unique ID (UUID) and refers, if geographically verified with a gazetteer (Pleiades), via a URL to a graph containing various information such as e.g. sources, archeological data, images, etc. In some maps, annotated places are additionally represented by a visual item (E36) such as a church, a tower, or a wall. There are also further data models for image series and works like map collections or atlases. From these paths, WissKI generates

automatically input forms for map and text metadata and provides an interface for importing all table-formatted annotations and converting them into triples. Ontological enrichment of our data with CRM allows for a semantic interpretation of annotations such that, e.g., for each PlaceName, an instantiated CRM description in RDF/OWL triple format is generated and stored in a triple store. Using the semantically enriched geo-information from text (and map) annotations as CRM instances, spatial entities ("figure", "ground") and relations obtained by spatial role labeling as "figure-spatial_relation-ground" triples can now be upgraded to this rich semantic level by linking data. Due to the fundamental underlying triple structure for all kinds of annotations, the data are immediately ready for publication as standardized Linked (Open) Data; WissKI provides a SPARQL query interface. These triple data constitute a huge knowledge graph; they are the "raw material" for further research steps, i.e. the exploration of the historical understanding of spaces and the associated knowledge. Interpretation of the data has just begun: Actually, a study of Biondo's spatial language – comprising the whole text for the first time – by Berthele and Thiering is being finalized (2020) and a comparative overview of the toponyms in the Latium book with different maps (traditional and "modern" Ptolemaic maps, genuine maps of Italy and portolans) as well as a representation of the hodological dimension of the text is being prepared.

Bibliography

Blakemore, Michael / Harley, Brian J. (1980): Concepts in the History of Cartography - A Review and Perspective. In: Cartographica. International Publications on Cartography, 17/4, Monograph 26. University of Toronto Press: Toronto.

Bodenhamer, David J. / Corrigan, John / Harris, Trevor M. (eds.) (2010): The Spatial Humanities. GIS and the Future of Humanities Scholarship. Bloomington & Indianapolis: Indiana University Press.

Görz, Günther / Geus, Klaus / Michalsky, Tanja / Thiering, Martin (2018): Spatial Cognition in Historical Geographical Texts and Maps: Towards a cognitive-semantic analysis of Flavio Biondo's "Italia Illustrata", e-perimetron 13,4: 182-199.

Görz Günther / Seidl Chiara / Thiering, Martin (2019): Linked Biondo: Modelling Geographical Features in Renaissance Texts and Maps **.** In: Boutoura, Ch., Tsorlini, A., Livieratos, E. (Ed.): Proceedings 14th ICA Conference *Digital Approaches to Cartographic Heritage,* Thessaloniki, 8-10 May 2019.International Cartographic Association, Commission on Cartographic Heritage into the Digital. AUTH CartoGeoLab, ISSN 2459-3893, 124-140.

Michalsky, Tanja / Thiering, Martin (2020): Walking through history. An interdisciplinary approach to Flavio Biondo's spaces in the "Italia illustrata". Rome: Bibliotheca Hertziana.

Scholz, Martin / Merz, Dorian / Goerz, Günther (2016): Working with WissKI - A Virtual Research Environment for Object Documentation and Object-Based Research. Digital Humanities 2016, Conference Abstracts, Krakow, 11-16 July 2016. ISBN 978-83-942760-3-4, 944-945.

Thiering, Martin (2015): Spatial Semiotics and Spatial Mental Models: Figure-Ground Asymmetries in Language. Berlin: De Gruyter Mouton.