Inhalts- und Kontextbeschreibung - Neuartige Informationsdienste für Bibliotheken mit Linked Spatio-Temporal Data

S. Scheider, A. Degbelo, W. Kuhn, H. Przibytzin

Zusammenfassung: Raum, Zeit und thematische Inhalte sind wesentliche Dimensionen zum Beschreiben und Suchen von Informationsmedien in Bibliotheken. Letztere stellen nicht nur Dokumente und traditionelle Medien wie Papierkarten bereit, sondern in zunehmendem Maße auch wissenschaftliche Datensätze, sowie Metadaten zur Beschreibung dieser Dokumente und Daten bezüglich Inhalt und Herkunft. Wie können Bibliotheken in ihrer Rolle als „Information broker“ für diese diversen Medien unterstützt werden? In diesem Artikel diskutieren wir eine Reihe von Bibliotheksdiensten, deren Realisation eine Herausforderung darstellt im Hinblick auf die Verknüpfung von Medien mit raum-zeitlichen Inhalten. Linked Spatio-Temporal Data (LSTD) bieten eine Möglichkeit, diese Dienste zu verwirklichen und damit das momentane Angebot wesentlich zu erweitern. Wir verdeutlichen unsere Vorschläge anhand von Beispielen aus der Linked Open Data Initiative der Universität Münster (LODUM) und aus verwandter Forschung. Wir zeigen anhand einer Auswahl anschaulicher Werkzeuge, wie man diese Herausforderung mit LSTD angehen kann.

Schlüsselwörter: Linked Spatio-Temporal Data, Bibliotheken, LODUM

Content and context description - How linked spatio-temporal data enables novel information services for libraries

Abstract: Space, time and thematic content are essential dimensions that allow libraries and their users to efficiently describe, search and access information media. The latter include not only documents and traditional media, such as paper maps, but to an increasing extent also scientific data sets, as well as all kinds of metadata describing these documents and data sets, both content-wise and in terms of their provenance. How can libraries be supported in their role as information broker for these diverse media? In this paper, we discuss a number of library services which have been challenging or impossible to realize in the past, especially with respect to linking media with spatio-temporal content descriptions and descriptions of their spatio-temporal accessibility. We argue that linked spatio-temporal data (LSTD) provide a way of realizing these services, in a manner which may substantially broaden the current scope of library information services. We illustrate these services based on examples from the Linked Open Data initiative of the University of Münster (LODUM) and related research. We discuss (based on a variety of illustrative tools) how LSTD is suitable to tackle these challenges.

Keywords: linked spatio-temporal data, libraries, LODUM

Autoren

Dr. Simon Scheider[1]
M. Sc. Auriol Degbelo[1]
Prof. Dr. Werner Kuhn[1,2]
Dipl.-Geol. Holger Przibytzin[3]

[1]Institut für Geoinformatik - Universität Münster
D-48149 Münster

[2]Department of Geography – University of California, Santa Barbara
CA 93106-4060, Santa Barbara, Kalifornien

[3]Universitätsbibliothek Münster
D-48143 Münster

E: {simon.scheider, degbelo, kuhn, holger.przibytzin}@ uni-muenster.de

1 Introduction: space, time and thematic content

In the age of printing, libraries were mainly providers of metadata catalogues for documents. The focus was on managing and organizing retrieval and physical access to these documents, not so much on supporting the process of scholarly work which unfolds around the document. Since documents were seen as information containers irrespective of their content, metadata focused on attributes of those containers, such as shelf numbers. Gradmann (2014) argues that in the age of the Web, libraries face a paradigm shift:

- From documents, understood as mere *containers* of information, towards their *content*, i.e., the things a document talks about; and
- From document containers to their *context*, i.e., the way background knowledge documents provenance, and external information shapes the interpretation of a document.

In the Web, the notion of a monolithic document dissolves into hypertext (Gradmann, 2014), i.e., into a network of references interlinking content with context. The distinction between metadata describing data and the data itself becomes blurred. In consequence, the role of libraries changes from metadata providers to information brokers. This change implies a number of principal challenges.

First, the explicit description of contents and the embedding of documents in their context require careful scholarly work. Second, this work amounts to more than what the staff of a single institution, such as a library, can afford. And third, this new kind of metadata needs to be made available to users who search for particular media, in a way which allows exploiting the various links, searching through and accessing specific content in space and time. This requires the support of query possibilities beyond simple text indices as well as of navigation through media.

These principal challenges directly translate into a number of *technical challenges*:

1. How can libraries support users in *finding media* based on the *relation between content and context*? Users are neither database experts capable of formulating complex queries, nor do they know about the database schemas necessary for this purpose. Furthermore, without access to the semantics hidden in these schemas, it is difficult to assess whether a document may be relevant or not.

2. How can libraries *generate useful content and context descriptions*? Metadata either needs to be generated by external sources, e.g. by researchers, or automatically (Christoph, 2013). One option is to *crowdsource* metadata generation, i.e., to turn it over to document users and interested scientists. The challenge here lies in enabling the reuse of descriptions. That is, in order to contribute metadata, it is necessary to reuse metadata schemas as well as spatio-temporal references that are machine readable, instead of annotating media with free text. Another option is to automatically enrich document metadata with external sources, such as www.geonames.org, using a linking service (Christoph, 2013).

3. How can libraries *efficiently manage media* in terms of content and context? Since data schemas for content and context are necessarily very diverse, how can different data schemas, terminologies or languages be supported?

4. How can libraries efficiently *organize access* to media? This involves easing navigation through physical collections and making virtual resources accessible and searchable from anywhere in the world.

In this article, we argue that *linked spatio-temporal data* (LSTD) and corresponding technologies which integrate geographic information technology with the Semantic Web (Janowicz et al., 2012) provide a way to significantly enhance current library services and cope with these challenges. LSTD integrates space, time and semantics. All three are essential dimensions for organizing information (Janowicz, 2010). LSTD can therefore be used for linking content with context. It can be exploited for improved search capabilities as well as for easy-to-use and reusable metadata descriptions, and it provides a way to manage and enrich media information in an efficient manner. However, substantial research is required in order to realize this vision.

In the next sections, we give a brief overview of the state-of-the-art and the current linked data movement in libraries and e-Science, before we discuss in detail a number of library services that have been difficult or impossible to realize with current technology. We introduce LSTD as a technology, discuss its advantages over previous approaches, and present useful services and tools which address the challenges.

2 Geographic information and linked open data (LOD) in libraries and e-Science

As information brokers, libraries need to embed documents into the scholarly research cycle. Documents and research data come with a specific *context of provenance* and have a specific *content*. Both of these, in turn, have *spatial and temporal dimensions* (Kuhn et al., 2014). This requires spatio-temporal as well as semantic metadata handling. How do libraries currently deal with this challenge?

There is an *e-Science* movement in the library community (Hey and Hey, 2006), in which *Geographic Information Systems (GIS)* play a key role. For example, numerous libraries in the US provide data services for browsing and retrieving geospatial data sets, including MIT libraries[1], NCSU[2] and the UCSB library map and imagery laboratory[3]. In addition, a large number of libraries hold collections of (historic) map documents and provide spatial search and retrieval portals for them, including Harvard[4], the University of Georgia[5], the Library of Congress[6] (see also Tony Campbell's Website[7] and Kovarsky (2011)) and the SLUB Dresden[8]. All these portals, however, rely on map catalogue services[9] or traditional Web services. They do not offer resolvable links into the content of data sets, nor do they supply any machine readable description of the provenance of a document. Furthermore, many of these data sets are not openly available or accessible, which limits their reuse.

The realization of e-Science requires libraries to become more broadly involved (Hey and Hey, 2006), providing access to all research outputs, not only to official documents. How to record provenance and preserve information, how to crowdsource recording and annotation of research results, and how to maintain e-Science repositories (Hey and Hey, 2006) remain open questions. Labahn et al. (2012) recognize the quality of geocoding as well as the integration of spatial, temporal and textual access to library information as major future challenges. Kuhn et al. (2014) argue that linked data could be a paradigm shift for the theory and practice of (geographic) information science, offering a way to better handle such challenges. Kauppinen et al. (2013) recently suggested *Linked Science* as an approach to e-Science. The idea is to explicate documents, data, methods, tools, provenance as well as licenses in terms of linked open data and to publish everything in the cloud, such that other researchers are capable of reproducing research results. Realizing this vision would address major technical challenges of e-Science. As Pohl and Danowski (2013) emphasize, however, there are also organizational and legal challenges in realizing a linked e-Science.

According to Razum (2011), there are four types of metadata: *technical metadata* (e.g. data type, data size), *provenance metadata* (i.e. how the media was created, by who, when, and why), *descriptive metadata* (i.e. tied to a specific domain), and *license metadata*. In this article, we focus on library information services for descriptive data of documents or media, not on research data sets, and we do not discuss legal and organizational issues.

Libraries started using and publishing linked open data catalogues a while ago. Pohl and Danowski (2013) trace the development of linked data in libraries back to 2007, when the Library of Congress was criticized by the Open Knowledge Foundation for their restrictive license regarding their catalogue data. The issue of *data licensing*, the non-technical aspect of the "open" in linked open data, remains a central requirement in order to obtain the full advantage of linked data in libraries. As Pohl and Danowski (2013) argue, it is only through linked open data[10] that a process of mutual enrichment and added value among library institutions and users can be realized. If each party has full access to external metadata, it can enrich its data as well as reuse externally generated links in its services, making the data and services subject to free and continual improvement by external parties.

Since 2007, major institutions have followed the Library of Congress in publishing their catalogues as linked open data, such as the OCLC (WorldCat) in 2012 (Pohl and Danowski, 2013). In Europe, the Europeana project[11] provides data about the European cultural heritage from museums and libraries in many countries as linked data. CERN and the North Rhine-Westphalian Library Service Center (hbz)[12] - from Switzerland and Germany respectively - were among the first libraries to publish their catalogs under the CC0 Public Domain license. In 2010, the German National Library (DNB) started a linked data service for their authority data set (GND) (Geipel et al., 2013). The latter is a set of authorized names and standard identifiers for publication metadata, including persons, families, institutions, events and works. It is

[1] http://libguides.mit.edu/Geodata
[2] http://www.lib.ncsu.edu/gis/
[3] http://www.library.ucsb.edu/mil
[4] http://hcl.harvard.edu/libraries/maps/
[5] http://www.libs.uga.edu/magil/collections/maps.html
[6] http://memory.loc.gov/ammem/gmdhtml/
[7] http://www.maphistory.info
[8] http://www.deutschefotothek.de/cms/kartenforum.xml
[9] http://www.opengeospatial.org/standards/cat
[10] See http://opendefinition.org/od/ (last accessed: June 13, 2014) for the meaning of 'open' in this context.
[11] http://www.europeana.eu/
[12] Hochschulbibliothekszentrum des Landes Nordrhein Westfalen

also available under Creative Commons License (CC0) and comes with its own ontology, which is aligned to the Friend of a Friend (FOAF) ontology. In Germany, the "culture-graph"[13] platform started with the algorithmic linking of authority data (Normdaten) of the DNB with bibliographic catalogues from a range of German libraries, effectively allowing the query of catalogues across libraries.

3 Library information functions and corresponding challenges

In the following, we discuss some of the technical challenges for libraries in a more systematic way and with reference to various information functions and mandates that libraries have. In some sense, all libraries provide services for describing, exploring, accessing and archiving media. We focus here on those services which are *difficult to realize with current technology*.

1. *Describing media by space, time, content and context*. Libraries have always been playing the role of metadata providers for documents in terms of cataloguing. They now need to provide services that allow librarians as well as researchers to describe and access diverse kinds of media (images, videos, maps, text documents, data sets) by space, time, content and context. These services include:

(a) *Sharing vocabularies, reference systems and background knowledge*. Standard cataloguing formats such as MARC21 need to be openly accessible and complemented by other vocabularies (Pohl and Danowski, 2013). Different vocabularies need to be supported for different purposes (Janowicz et al., 2012). Furthermore, schema knowledge as well as factual (background) knowledge characterize the content and context of a medium, and thus need to be shared and published together with it. A particularly important kind of data schema for research are *reference systems* for space, time and measurements (Chrisman, 2001), which provide meaning (Kuhn, 2003). Current library catalogues are often based on technology which is unaware of semantic links and spatio-temporal data types. These catalogues tend to become opaque data silos if identifiers remain unpublished or non-unique and if closed schemas are used[14]. Support for diverse schemas, spatio-temporal reference systems, semantic content and shared background knowledge is often missing.

(b) *Enriching content and context descriptions*. Content and context descriptions go beyond cataloguing. Since librarians are not capable of providing this on their own, they need enrichment services. As discussed above, there are two possibilities, namely crowdsourcing and automation. Regarding the former, current metadata and catalogue systems lack easy access to metadata as well as an easy way to edit metadata by external users. Furthermore, user communities need to be developed. Regarding automation, data extraction and linking tools are needed which take the semantics of data (Christoph, 2013; Pfeffer, 2013) as well as space and time into account (Janowicz et al., 2012). A particular challenge is to disambiguate names of places, people and events. Space and time help in distinguishing entities with ambiguous names (Janowicz, 2010)[15].

(c) *Validating and curating descriptions*. Descriptions need to be validated and curated. Data curation by librarians as well as external users needs to be supported. For this purpose, unique identifiers for things need to be shared and reused in annotations and data improvements. Furthermore, data quality standards, trust and expertise levels need to be established and implemented inside user communities.

2. *Searching and exploring media by space, time and content*. Users need to benefit from the logical structure of space, time, content and context. State-of-the-art search engines such as SOLR[16] provide fast access to data (e.g. based on indexing of XML files), but do not support a user in exploiting the logical structure of data. The reason why current catalogue (as well as Web) search engines are mainly based on unstructured text is that the method is flexible and has the lowest entry barrier. It does not require users to deal with data structures. However, it is imprecise in terms of content and context. When a scientist chooses inadequate search terms or is interested in a specific content or a specific place or period, it is apparent that text search becomes ineffective. For example, how could a historian possibly search for maps which allow her to determine landcover around Hildesheim in the 19th century (Scheider et al., 2014)? This question could be answered based on the particular historic map shown in Figure 1, but it is nearly impossible to turn this question into a meaningful search operation based on document titles or keywords. Furthermore, searching manually is not scalable. What is needed instead is a way to help untrained users formulate precise content and context related

[13] http://www.culturegraph.org
[14] For example, a publication database such as *Allegro-C* requires entering text strings for authors.
[15] See (Degbelo et al., 2014) for a study about the usefulness of spatio-temporal information for identity reasoning in the context of a bibliographic dataset.
[16] http://lucene.apache.org/solr/

questions and to search for media by matching questions with thematic, spatial, and temporal descriptions. Scalable spatial query capacities are available in spatial database technology[17], but are not properly linked to the Web of Data. Current Geo Web service standards, such as catalogue services (CSW)[18] or web feature services (WFS)[19], allow querying by space, but do not allow searching through semantic content down to the level of singular statements (Kuhn et al., 2014). Three major challenges are related to this state-of-the-art:

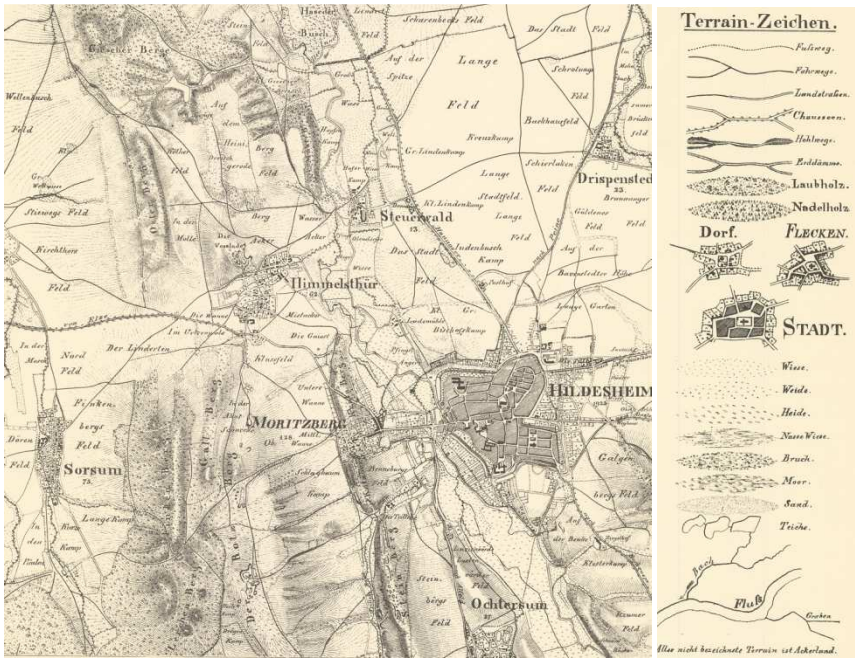

Figure 1: Excerpt of a map of Hildesheim of the "Gaußsche Landesaufnahme" from 1839. How can the content of similar documents be made accessible and searchable for library users? Source: Historische Kommission für Niedersachsen, Hannover 1963.

(a) *Exploring query vocabularies*. Users need to be informed about the available types, schemas and relations, i.e., the vocabularies available in a metadata repository which can be used for formulating queries. The latter may change from repository to repository. Space and time values, if available, need to be automatically detected, encapsulated and provided for search.

(b) *Visual querying based on space, time and theme*. Linked spatio-temporal datasets are currently accessible through SPARQL[20] endpoints pointing to triple stores. The exploration of these triple stores requires knowledge of (i) the SPARQL syntax and (ii) the vocabularies used to annotate the data stored in the triple stores. In the future, untrained users should be able to easily state queries without having to learn a query language. Space and time of media needs to be made queryable by digital map windows and time sliders.

(c) *Analyzing data by GIS clients*. GIS users may want to apply analysis and cartographic technology on library data. In this way, the spatial and temporal distribution of publications and contents could be analyzed and overlayed with other cartographic information (Keßler et al., 2012).

3. *Accessing resources in space and time*. Different kinds of media need to be analogically or digitally accessible. Furthermore, libraries need to support users in localizing related resources on a campus. How can one find a harmonized form of publication which is general enough to account for diverse media and data sets (published or unpublished) and can be reused across libraries? How can resources be spatially localized on a campus? Once a physical resource has been digitally identified, its location on a campus needs to be specified and users need to be navigated to it.

[17] http://en.wikipedia.org/wiki/Spatial_database
[18] http://www.opengeospatial.org/standards/cat
[19] http://www.opengeospatial.org/standards/wfs
[20] The query language for linked data, see http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/

4 Linked Spatio-Temporal Data (LSTD)

In this section, we present the main technical characteristics of LSTD, and discuss some advantages over traditional information technology. Linked open data evolved in the Semantic Web (Hitzler et al., 2009). It comes with a technology stack for semantically describing, publishing, linking and retrieving resources on the Web. Resources can be anything. They are identified by uniform resource identifiers (URI) and are connected by links to other resources or to data values (literals), forming a graph in the RDF (Resource Description Framework)[21] language. Links (properties) are likewise identified by URIs. URIs allow retrieving linked data by machines in the Web. An RDF graph can be published as a file in different serialization formats using content negotiation (RDF/XML, Turtle), it can be hidden in Web pages (RDFa), or it can be stored on database servers (triple stores). It has formal semantics like a logical language: resources (nodes) stand for any entities (e.g. "Goethe") and links stand for binary relations (e.g. "author of"), and if a link connects two nodes (forming a "triple"), then this stands for an assertion, such as "Goethe author-of Faust". This formal semantics allows the storage and inference of implicit knowledge by logical reasoning. Furthermore, the graph can be queried over *http* and a protocol called SPARQL using a RESTful Web interface called "endpoint". Various natural languages as well as data values (e.g. spatial and temporal references) are supported by *typed literals*, nodes that are strings associated with a data type.

From the viewpoint of a library, there are a number of advantages of linked open data compared to traditional *library catalog services* (see also Pohl and Danowski (2013)), as well as compared to traditional cartographic or *Geographic Information System (GIS) technology* (see also Kuhn et al. (2014)):

- *Content and context can be treated as graphs*. RDF provides a universal language to describe context and content assertions about a document in terms of a graph, where each assertion is a triple and thus a statement linking a subject with an object (Scheider et al., 2014).
- *Metadata are data*. RDF serves as a meta-language for other languages (Hitzler et al., 2009), covering diverse schemas, terminologies and natural languages. Metadata describing linked data are also linked data, published and shared in the same manner. In contrast, traditional database schemas hide metadata in the data structure. Linked data unifies the structure and separates it from data semantics, which becomes publishable as a vocabulary (Kuhn et al., 2014). A well documented and openly accessible vocabulary therefore serves the purpose of successful information sharing. Some useful vocabularies for describing documents in libraries are Dublin Core Terms (dct)[22], bibo[23], CIDOC CRM[24] and MARC 21[25]. A set of recommendations for encoding documents as linked data can be found in DINI (2014).
- *Publication on the Web is incorporated*. Data sets can be published and retrieved using *http*, which allows users to RESTfully access data over the Web and point to data without requiring any further technical service interface. This enables in particular *crowdsourcing* of data enrichment over the Web.
- *Retrieval is precise*. Since triple stores[26] enable querying and reasoning over graphs, explicit and implicit information about a document can be retrieved in a very precise manner, beating text search. For example, it is possible to query for a *bibo:document* which contains information about cities and automatically get back maps of type *bibo:map* about Hildesheim, using RDFS[27] inference. Many triple stores allow spatio-temporal queries, such as *which other places are contained in the map of Hildesheim from 1839?* (Scheider et al., 2014).
- *Data can be flexibly linked to space and time*. Space and time can be handled by linking arbitrary nodes in a graph to typed literals. This overcomes the restrictions of modeling space and time in GIS as snapshots of geometries with attributes. It allows for modeling snapshots as well as dynamic entities (e.g. events and trajectories) in space and time (Hu et al., 2013). Literals denote spatial or temporal data values, while types denote reference systems. Space can be handled in a precise manner based on the GeoSPARQL ontology[28] (Battle and Kolas, 2012), which encodes geometries (polygons, lines or points) in terms of OGC's[29] *Simple Features* standard, e.g., as WKT[30]. Geometry literals should always be typed by a *reference system URI*, using an EPSG identifier[31], which denotes a particular coordinate reference system (Chrisman, 2001). A simpler and more often used alternative is the W3C

[21] http://www.w3.org/RDF/
[22] http://dublincore.org/documents/dcmi-terms/
[23] http://purl.org/ontology/bibo/
[24] http://www.cidoc-crm.org/official_release_cidoc.html
[25] http://marc21rdf.info/elements/
[26] http://en.wikipedia.org/wiki/Triplestore
[27] http://www.w3.org/TR/rdf-schema/
[28] See http://www.opengis.net/ont/geosparql/1.0, prefix geo.
[29] Open Geospatial Consortium
[30] Well known text (http://en.wikipedia.org/wiki/Well-known_text), OGC's simplest text serialization of the Simple Features geometry standard.
[31] http://www.epsg.org/

basic geo vocabulary[32]. However, the latter is restricted to the representation of points in a single reference system (WGS84) which makes it unsuitable for general purpose spatial representation. Time can be encoded by linking nodes to intervals or points in the OWLTime ontology[33], which draws on *xsd:DateTime* literals. Many triple stores, such as OWLIM[34] can handle spatial queries with basic coordinates. Allegro Graph[35] can handle simple polygons, while only Parliament[36] so far supports spatial queries with simple features and thus is comparable to the functionality of a spatial database. However, by combining a spatial database such as PostGIS[37] with triple stores through *http* interfaces, it becomes possible to use geospatial processing on linked data stores (Scharrenbach et al., 2012; Jones et al., 2014). One approach is to wrap linked data into OGC service interfaces such that GI processing can directly be applied to it[38].

In order to illustrate the description of spatial content, we analyze a data record from the German National Library (DNB) about the village Zavala in Bosnia and Herzegovina, encoded in MARC 21:

```
006 http://d-nb.info/gnd/7739653-4
034 agx$dE 017 58 39$fN 042 50 50$uhttp://sws.geonames.org/3186727$2geonames
034 dgx$dE017.977579$fN042.847490$uhttp://sws.geonames.org/3186727$2geonames
043 XA-BA
151 Zavala
670 Imenikmestau Jugoslaviji
679 Ort in Bosnien-Herzegowina, Serbische Republik, bei Trebinje
```

Using linked spatio-temporal data best practices as discussed above, the place should be encoded (in Turtle[39] format) as:

```
@prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix geo:<http://www.opengis.net/ont/geosparql#>.
@prefix owl:<http://www.w3.org/2002/07/owl#>.
@prefix gndo:<http://d-nb.info/standards/elementset/gnd#>.
@prefix gnd:<http://d-nb.info/gnd/>.
@prefix gndac:<http://d-nb.info/standards/vocab/gnd/geographic-area-code#>.

gnd:7739653-4 gndo:definition "Ort in Bosnien-Herzegowina, Serbische Republik, bei
                                Trebinje"@de;
        gndo:geographicAreaCode gndac:XA-BA;
        gndo:preferredNameForThePlaceOrGeographicName "Zavala";
        geo:hasGeometry
        [a geo:Point;
        geo:asWKT "<http://www.opengis.net/def/crs/EPSG/0/4326> POINT(17.977579
        42.847490)"^^geo:wktLiteral];
        rdf:type gndo:TerritorialCorporateBodyOrAdministrativeUnit;
        owl:sameAs <http://sws.geonames.org/3186727>
```

Zavala is identified by the URI gnd:7739653-4. Just as this place is linked to its preferred name and natural language definition, which is typed by the German language (@de), it is also linked via *geo:asWKT* to a location encoded as a WKT point, which is additionally typed by an EPSG number (as part of the literal), denoting the reference system WGS84 (the reference system underlying GPS). Without knowing this reference system, the spatial interpretation of a coordinate pair remains uncertain by several kilometers (Chrisman, 2001). Note also that the place is easily linked via *owl:sameAs* to equivalent entities in external databases, such as Geonames, and that every element of the data set is a resolvable Web address which can be looked up for further information. For example, one can click on the Geonames URI for a map of the place and on the corresponding link for learning about GND area codes or EPSG reference systems.

In order to illustrate the linking of documents with context and of content and with space and time, we list the encoding of the Map of Hildesheim shown in Figure 1 by two graphs, one denoting the map's context, the other denoting its content (for details, see (Scheider et al., 2014)). A SPARQL query can be used to retrieve map documents based on a particular kind of semantic content.

---

[32] http://www.w3.org/2003/01/geo/

[33] See http://www.w3.org/2006/time; prefix time.

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

[35] http://franz.com/agraph/allegrograph/

[36] http://parliament.semwebcentral.org/

[37] http://en.wikipedia.org/wiki/PostGIS

[38] See https://github.com/jimjonesbr/lod4wfs and (Jones et al., 2014).

[39] Terse RDF Triple Language http://www.w3.org/TR/turtle/

```
@prefix maps:<http://geographicknowledge.de/vocab/maps#>.
@prefix phen:<http://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 time:<http://www.w3.org/2006/time#>.
```

Map document:

```
: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))"^^geo:wktLiteral.
```

Content graph (:hildesheim):

```
_: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.
…
```

Query: *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.
          ?instance a ?cl.
         }
          ?instance a ?class.
          ?class rdfs:subClassOf phen:Landcover.
}
```

Note that the WKT geometry in this case is a polygon encoded as a literal string of five points, each denoted by a pair of coordinates, instead of a single point. Thus the map area covers a complex region. Georeferenced map images in general do not have to be rectangular because georeferencing may transform a map's own reference system in a non-linear manner (compare Figures 2 and 4).

How these advantages of LSTD can be translated into useful services that help address the challenges discussed in Section 3 will be discussed next.

5 How LSTD enables novel library services

Section 3 identified a number of library functions whose realization is challenging using current technology. This section illustrates how LSTD helps in realizing these functions by presenting services from LODUM[40] and from related work and suggesting further developments. LODUM is the open data initiative of the University of Münster, initiated by Werner Kuhn in 2010. Its goal is to make public information about the University of Münster available as linked data for both research and administrative matters. A brief introduction to LODUM's technical infrastructure is given in (Keßler and Kauppinen, 2012).

[40] http://lodum.de

5.1 Describing media by space, time, content and context

LSTD is particularly suited to generate metadata in a collaborative fashion, because metadata is linked data and thus easily accessible, and because data is easily linked to space, time and background knowledge.

5.1.1 Sharing vocabularies, reference systems and background knowledge

There are many linked data vocabularies available for the description of library content. Linked Open Vocabularies[41] is an entry point which collects and classifies some of these vocabularies by vocabulary spaces including library, media, metadata and entertainment. Klee (2013) provided a review of vocabularies which are relevant to the annotation of bibliographic data. There is a lot of work on publishing provenance as well as technical and license metadata through linked data (see for example Moreau et al., 2011). The Dublin Core vocabulary (through the terms *MediaType* and *license*) can be used to annotate media with respect to technical characteristics and license. There is less work on descriptive metadata. LODUM has focused on descriptive vocabularies for metadata about *spatio-temporal content*, based on external knowledge as well as spatial and temporal references. The approach suggested in (Scheider et al., 2014) illustrates how historic maps can be encoded as named graphs together with their spatio-temporal extent. One could reuse this approach for the description of *all documents* of libraries. For example, one could describe the content of books, dissertations and scientific publications as named graphs, for the purposes of precise semantic search. What services are needed to make content description vocabularies available for librarians and non-expert users? Tools such as DBpedia Spotlight[42], which help annotate occurrences of concepts in text, can assist librarians in this task. Corresponding *content vocabulary exploration and recommendation* tools help users find and understand content-related concepts for describing a given document, including its spatio-temporal references, taking advantage of the fact that metadata is linked data as well and thus can be easily queried and retrieved.

5.1.2 Enriching content descriptions

In order to ease (semi-)automatic enrichment of external links, tools such as SILK (Volz et al., 2009) are already being applied in the library community (Christoph, 2013). *Georeferencing* of place names using a gazetteer such as Geonames[43] is currently done by the German National Library (DNB). Yet, how can *document content descriptions* be crowdsourced or offered as a service to librarians? In LODUM, we developed a *georeferencing Web client*[44] (Sanchez, 2013) which helps describe historic maps by space, time, and content in linked data (Figure 2). Besides georeferencing historic maps based on historic places mapped in Open Street Map, the tool suggests links from maps to DBpedia (e.g., links from historic map contents to modern places). The outcomes (metadata) of the description can be exported as KML (Keyhole Markup Language) files for display in Google Earth or as RDF data that can be loaded into a triple store. A semantic map tagging tool based on Dbpedia suggestions was also proposed in (Haslhofer et al., 2013). What has not been developed so far is a tool that lets users precisely describe *arbitrary content as assertions* in terms of a graph (compare content graph in Section 4), i.e., a *generic content description tool* based on LSTD, as suggested in Scheider et al. (2014). Such a tool would let users express and manipulate content descriptions as visual graphs, similar to the visual query tool SPEX (see below).

5.1.3 Validating and curating descriptions

Validation and curation of descriptions can be done manually by a community of experts or volunteers. In LODUM, annotation tools such as the georeferencer allow users to reload and edit metadata. Since linked data handles provenance naturally in terms of statements made by somebody somewhere at some time (Kuhn et al., 2014), it is also possible to manage a community of editors. LSTD provides the necessary technical basis for (semi-)automatic validation and curation of descriptions in a library context. DeFacto[45] (Lehmann et al., 2012) is an example of a tool for validating RDF statements by confirming sources for them on the web; it takes a statement as input and tries to find evidence for the validity of that statement by searching textual information on the web. The tool is still a prototype[46], but one can imagine its extension to support library media descriptions.

---

[41] See http://lov.okfn.org/dataset/lov/index.html (last accessed: June 14, 2014).

[42] DBpedia Spotlight (see https://github.com/dbpedia-spotlight/dbpedia-spotlight/ wiki/Web-service, last accessed: June 14, 2014) helps annotating mentions of DBpedia concepts in plain text.

[43] http://www.geonames.org/

[44] See http://ubsvirt116.uni-muenster.de/georeferencer/georef.html, and https://github.com/lodum/georef for the source code.

[45] See http://defacto.aksw.org/defacto-ui/ (last accessed: June 14, 2014).

[46] For example, the user cannot enter all kinds of triples; she is at the moment only provided with exemplary statements to validate.

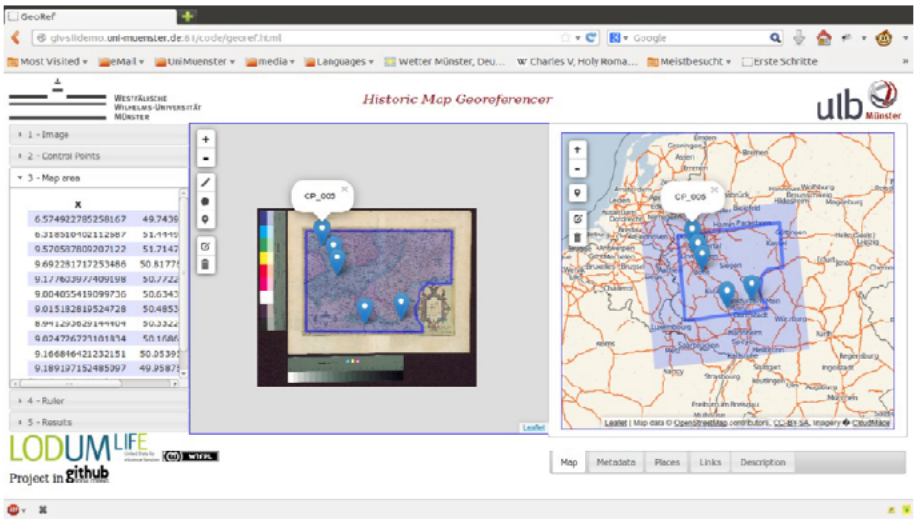

Figure 2: A georeferencing tool which allows combined spatial, temporal and semantic descriptions of historic maps.

5.2 Searching media by space, time, content and context

LSTD allows for rich descriptions of content and context of diverse media, but it also facilitates searches exploiting semantics and thus improving information retrieval. First of all, it enables search across different kinds of media (books, maps, postcards, articles) and different kinds of library databases, as illustrated by the meta-search tool of the Institute for urban history (ISTG) in Münster[47] (compare Figure 3). Another example for such an integration service is the *lobid* API of the hbz[48], which allows searching over organisations, persons, subjects as well as bibliographic resources from the North-Rhine Westphalia union catalogue, the German National Library and the German ISIL registry. As discussed in the following, LSTD furthermore enables multilingual search, browsing of data and metadata, automatic vocabulary suggestions, visual formulation of semantic queries on a graph, as well as constraining search by space and time.

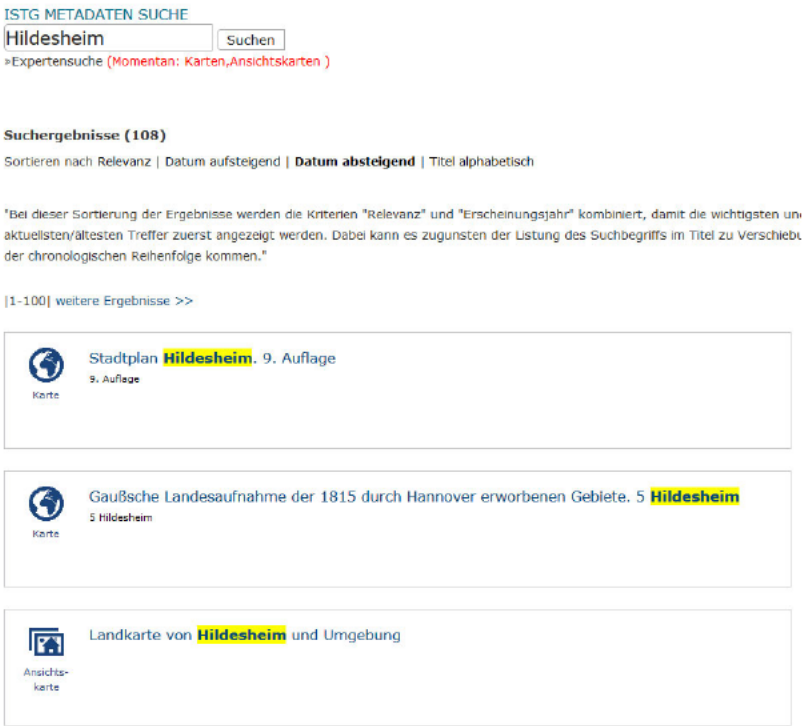

Figure 3: Meta-search over four urban databases (maps, postcards, literature, urban texts). In this example, we searched for resources about Hildesheim. Note that the map displayed in Figure 1 is among the results.

[47] http://www.uni-muenster.de/Staedtegeschichte/portal/datenbanken/mediensuche.html

[48] http://lobid.org/api

5.2.1 Exploring query vocabularies

Since LSTD treats metadata as data, the metadata needed for retrieving documents in a repository can be easily explored, such that users who do not know about the content of a repository can find it out. Alonen et al. (2013) introduced two tools for the exploration of datasets from the Linked University initiative[49]. The first tool, called Visualization Playground[50], supports the specification and creation of visualizations for the exploration and comparison of linked datasets. The second tool, named Vocabulary Visualizer[51], shows the usage of linked data vocabularies in multiple datasets. In LODUM, we are currently developing a *vocabulary suggester* tool along the lines of SPEX (see below). This tool allows users to search for vocabularies available in a triple store and suggests building blocks for visual queries.

5.2.2 Visually querying for space, time and semantic content

Visual query systems assisting library users in SPARQL queries and in stating spatio-temporal constraints further exploit the advantages of LSTD. Examples of tools assisting the user in formulating SPARQL queries are visualSPARQL[52], Konduit (Moller et al., 2008) and NITELIGHT (Smart et al., 2008). However, these tools do not inform about content available in a repository and do not support spatio-temporal queries. LODUM is currently developing a visual query system for triple stores, called Spatio-temporal content explorer (SPEX), which is open[53], targeted at novice users, and reusable across triple stores. It suggests classes, properties and instances available in a triple store, automatically handles spatial or temporal information from nodes, displays it in maps and sliders, and lets users construct visual graph patterns with space-time constraints for querying content (see Figure 4). It furthermore automatically labels all information and hides the linked data syntax from users.

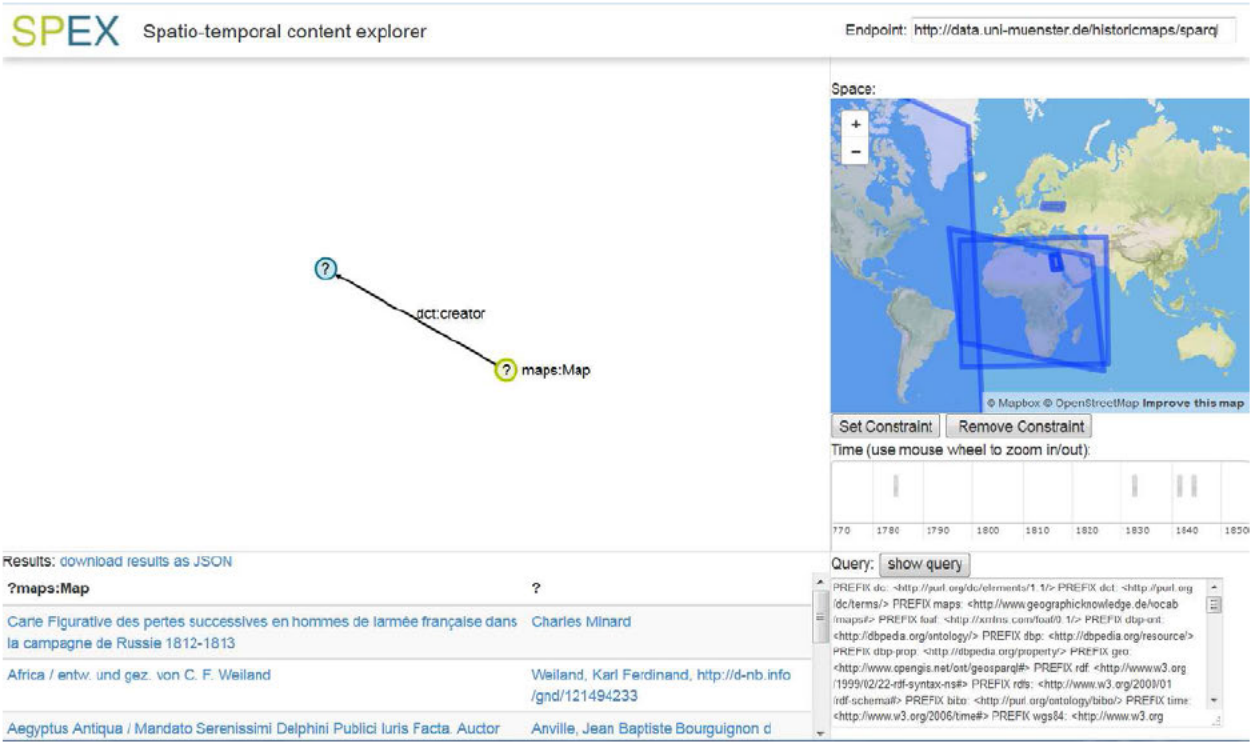

Figure 4: SPEX (at http://lodum.de) lets users formulate spatio-temporal and thematic queries in terms of a visual graph using auto-suggested classes and properties contained in a SPARQL endpoint. For example, it enables querying for maps about a certain region, from a certain time and from certain creators. SPEX automatically detects the presence of classes, properties, as well as spatial and temporal information, displays the latter in maps and time sliders, and allows for using it in queries.

[49] Linked University is an alliance of european universities engaged into exposing their public data as linked data (see http://linkeduniversities.org/lu/; last accessed: June 14, 2014).
[50] See http://data.aalto.fi/visu (last accessed: June 14, 2014).
[51] See http://data.aalto.fi/V2 (last accessed: June 14, 2014).
[52] http://graves.cl/visualSparql/
[53] See https://github.com/lodum/SPEX

5.2.3 Analyzing meta-data by GIS clients

In LODUM, we developed an adapter[54] which (a) listens to WFS requests and translates them into SPARQL queries; and (b) transforms RDF result sets (received after the execution of SPARQL queries) into WFS documents (Jones et al., 2014). This adapter provides the missing link between Web Feature Services (standardized by the Open Geospatial Consortium, OGC) and the Web of linked open data. Using this adapter, ordinary (OGC-conformant) GIS clients can be directly used on library data exposed as LSTD (see Figure 5). We provided a performance-tested reference implementation Web service written in Java, which allows WFS layers to be customized for arbitrary triple stores with a Web interface.

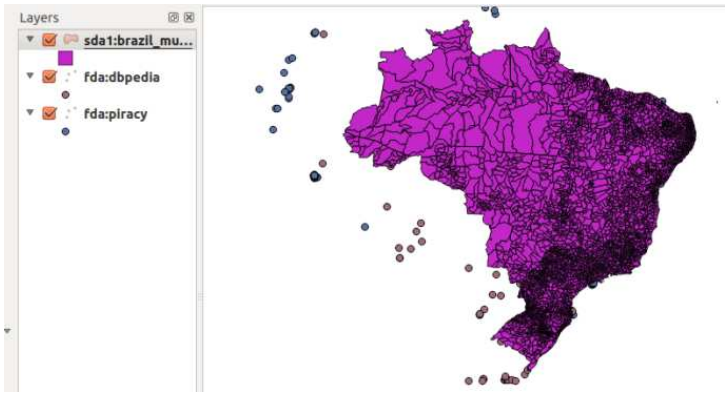

Figure 5: The LOD4WFS adapter enables GIS analysis directly on LSTD.

5.3 Accessing resources in space and time

At the moment of writing this article, the LODUM triple store contains campus data about people (e.g. faculty), activities (e.g. lectures), facilities (e.g. auditoriums), works (e.g. research papers), and locations (e.g. of a department). An official Campusplan App[55] has been developed at the University of Münster using this data. The application provides students, staff, administrators, and visitors with navigation instructions to university buildings, an overview map of all university buildings (spread across town), and practical information such as cafeteria menus on mobile phones or Web browsers (Figure 6). We are currently developing a *navigation support tool* for disabled people which allows users in wheelchairs to retrieve accessibility information about rooms and buildings.

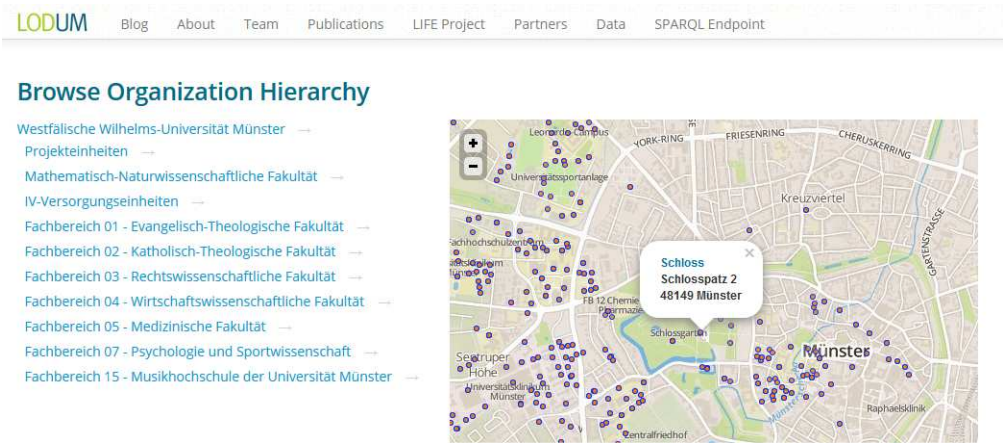

Figure 6: In LODUM, people are linked to buildings (displayed on a map) via the organizations to which they belong. Information about access to buildings and rooms can easily be added.

[54] See https://github.com/jimjonesbr/lod4wfs and (Jones et al., 2014).
[55] See http://app.uni-muenster.de/.

6 Conclusion

We have presented examples of tools illustrating how linked spatio-temporal data can be used to address challenges in developing library services. In their role as information brokers in the age of the World Wide Web, libraries need to embed documents and media into descriptions of their content and context, including spatio-temporal coverage. This requires a flexible way of representing spatio-temporal and thematic information. As this goes substantially beyond traditional cataloguing, libraries need to either crowdsource or automate content and context descriptions. The former requires researchers and other library users to be able to easily contribute descriptions, the latter requires machine access to structured external information. In order to let users benefit from context and content descriptions, libraries need to enable search through and access to media based on space, time, content and context.

The linked data movement has already gained ground in library institutions. Spatio-temporal information, however, is currently handled separately from linked data in terms of geospatial services. For sharing vocabularies, enriching and validating content and context descriptions, and searching, exploring and accessing resources by untrained users, linked spatio-temporal data offers many advantages compared to traditional catalogues as well as geospatial search. Content and context can be treated as a single linked data graph, including metadata, published on the Web, allowing for more precise retrieval, and referenced to space and time in a very flexible manner.

This integration of linked data with spatio-temporal referencing and reasoning enables, among other things, the development of vocabulary exploration and recommendation tools, tools that combine georeferencing with semantic descriptions and linking to external data repositories, as well as metadata validation tools. Furthermore, LSTD enables search tools that autonomously determine the content of a given repository, suggest search categories, and allow untrained users to construct precise queries in terms of visual graphs, using maps and sliders to deal with spatio-temporal constraints. By encapsulating LSTD as geospatial service standards, it becomes possible to analyze metadata published as LSTD by ordinary GIS tools. Furthermore, using LSTD, it is straightforward to support spatial navigation to a physical resource once its description has been found.

In future work, we will focus on developing exploration and search tools for library repositories, in particular for historic map collections, which demonstrate the advantages of linked data and space and time for researchers and library users. Furthermore, tools which let users easily describe content and context of media in a precise way as LSTD are needed. Another future challenge lies in interlinking and maintaining links across data repositories, both with respect to the process of crowdsourcing and with respect to external knowledge sources. These challenges have important organizational aspects and thus require more than the kinds of technical solutions suggested here.

Acknowledgements

The work of LODUM is funded by the German Research Foundation under the LIFE project (DFG KU 1368/11-1).

References

Alonen, M., Kauppinen, T., Suominen, O., and Hyvonen, E. (2013). Exploring the Linked University Data with visualization tools. In Cimiano, P., Fernandez, M., Lopez, V., Schlobach, S., and Völker, J., editors, The Semantic Web: ESWC 2013 Satellite Events, pages 204–208, Montpellier, France. Springer.

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

Chrisman, N. (2001). Reference systems for measurement. In Exploring geographical information systems, pages 15–35.

Christoph, P. (2013). Datenanreicherung auf LOD Basis. In (Open) Linked Data in Bibliotheken, pages 140–167.

Degbelo, A., Birregah, B., and Keßler, C. (2014). Using spatial and temporal information for identity reasoning: the case of a bibliographic dataset. In Proceedings of the 16th International Symposium on Spatial Data Handling (SDH 2014). Toronto, Canada. (Forthcoming).

DINI, A. K. G. T. (2014). Empfehlungen zur RDF-Repräsentation bibliografischer Daten. Number 14 in DINI-Schriften. Deutsche Initiative für Netzwerkinformation (DINI).

Geipel, M. M., Bohme, C., Hauser, J., and Haffner, A. (2013). Herausforderung Wissensvernetzung. In (Open) Linked Data in Bibliotheken, pages 168–185.

Gradmann, S. (2014). From containers to content to context: The changing role of libraries in eScience and eScholarship. Journal of Documentation, 70(2):241–260.

Haslhofer, B., Robitza, W., Guimbretiere, F., and Lagoze, C. (2013). Semantic tagging on historical maps. In Proceedings of the 5th Annual ACM Web Science Conference, WebSci ’13, pages 148–157, New York, NY, USA. ACM.

Hey, T. and Hey, J. (2006). E-science and its implications for the library community. Library Hi Tech, 24(4):515–528.

Hitzler, P., Krotzsch, M., and Rudolph, S. (2009). Foundations of Semantic Web Technologies. Chapman & Hall/CRC.

Hu, Y., Janowicz, K., Carral, D., Scheider, S., Kuhn, W., Berg-Cross, G., Hitzler, P., Dean, M., and Kolas, D. (2013). A geo-ontology design pattern for semantic trajectories. In Tenbrink, T., Stell, J., Galton, A., and Wood, Z., editors, Spatial Information Theory, volume 8116 of Lecture Notes in Computer Science, pages 438–456. Springer International Publishing.

Janowicz, K. (2010). The role of space and time for knowledge organization on the semantic web. Semantic Web, 1(1-2):25–32.

Janowicz, K., Scheider, S., Pehle, T., and Hart, G. (2012). Geospatial semantics and linked spatiotemporal data - past, present, and future. Semantic Web, 3(4):321–332.

Jones, J., Kuhn, W., Keßler, C., and Scheider, S. (2014). Making the web of data available via web feature services. In Huerta, J., Schade, S., and Granell, C., editors, Connecting a Digital Europe Through Location and Place, Lecture Notes in Geoinformation and Cartography, pages 341–361. Springer International Publishing.

Kauppinen, T., Baglatzi, A., and Keßler, C. (2013). Linked Science: Interconnecting scientific assets. In Critchlow, T. and Kleese-Van Dam, K., editors, Data Intensive Science. CRC Press, USA.

Keßler, C., Janowicz, K., and Kauppinen, T. (2012). spatial@linkedscience – Exploring the research field of GIScience with linked data. In Xiao, N., Kwan, M.-P., Goodchild, M., and Shekhar, S., editors, Geographic Information Science, volume 7478 of Lecture Notes in Computer Science, pages 102–115. Springer Berlin Heidelberg.

Keßler, C. and Kauppinen, T. (2012). Linked Open Data University of Muenster - Infrastructure and applications. In Demos of the Extended Semantic Web Conference 2012 (ESWC2012), Heraklion, Crete, Greece.

Klee, C. (2013). Vokabulare für bibliographische Daten. In Danowski, P. and Pohl, A., editors, (Open) Linked Data in Bibliotheken, pages 45–63. De Gruyter Saur.

Kovarsky, J. (2011). Searching for early maps: Use of online library catalogs. ACMLA Bulletin, (138).

Kuhn, W. (2003). Semantic reference systems. International Journal of Geographical Information Science, 17(5):405–409.

Kuhn, W., Kauppinen, T., and Janowicz, K. (2014). Linked Data – A paradigm shift for Geographic Information Science. In Proceedings of the Eighth International Conference on Geographic Information Science (GIScience2014), Vienna University of Technology, Austria. (Forthcoming).

Labahn, K., Stephan, R., and Zepf, R. (2012). Digitale Bibliothek und Virtuelle Forschungsumgebung – Raumbasierte Verknüpfung und Recherche von bibliothekarischen Datenbeständen. In Bill, R., editor, Virtuelle Forschungsumgebung für die Kulturlandschaftsforschung auf Basis von Internet-GIS-Technologien, pages 59–82, Wichmann Berlin.

Lehmann, J., Gerber, D., Morsey, M., and Ngonga Ngomo, A. (2012). DeFacto - Deep fact validation. In Cudré-Mauroux, P., Heflin, J., Sirin, E., Tudorache, T., Euzenat, J., Hauswirth, M., Parreira, J., Hendler, J., Schreiber, G., Bernstein, A., and Blomqvist, E., editors, The Semantic Web–ISWC 2012, pages 312–327, Boston, Massachusetts, USA. Springer Berlin Heidelberg.

Moller, K., Dragan, L., Ambrus, O., and Handschuh, S. (2008). A Visual interface for building SPARQL queries in Konduit. In Bizer, C. and Joshi, A., editors, Proceedings of the Poster and Demonstration Session at the 7th International Semantic Web Conference (ISWC-2008), Karlsruhe, Germany. CEUR-WS.org.

Moreau, L., Clifford, B., Freire, J., Futrelle, J., Gil, Y., Groth, P., Kwasnikowska, N., Miles, S., Missier, P., Myers, J., Plale, B., Simmhan, Y., Stephan, E., and Van den Bussche, J. (2011). The open provenance model core specification (v1.1). Future Generation Computer Systems, 27(6):743–756.

Pfeffer, M. (2013). Eine Infrastruktur für "Big Metadata". b.i.t. online, 16(1):11-16.

Pohl, A. and Danowski, P. (2013). Linked Open Data in der Bibliothekswelt: Grundlagen und Überblick. In (Open) Linked Data in Bibliotheken, pages 1–44.

Razum, M. (2011). Systeme und Systemarchitekturen für das Datenmanagement. In Büttner, S., Hobohm, H., and Müller, L., editors, Handbuch Forschungsdatenmanagement, pages 123–138. Bock + Herchen Verlag.

Sanchez, A. (2013). A LOD-based georeferencing tool for historic maps. Masters thesis, University of Münster.

Scharrenbach, T., Bischof, S., Fleischli, S., and Weibel, R. (2012). Linked raster data. Seventh International Conference on Geographic Information Science.

Scheider, S., Jones, J., Sanchez, A., and Keßler, C. (2014). Encoding and querying historic map content. In Huerta, J., Schade, S., and Granell, C., editors, Connecting a Digital Europe Through Location and Place, Lecture Notes in Geoinformation and Cartography, pages 251–273. Springer International Publishing.

Smart, P. R., Russell, A., Braines, D., Kalfoglou, Y., Bao, J., and Shadbolt, N. R. (2008). A visual approach to semantic query design using a web-based graphical query designer. In Gangemi, A. and Euzenat, J., editors, Knowledge Engineering: Practice and Patterns (EKAW 2008), pages 275–291, Acitrezza, Italy. Springer-Verlag Berlin Heidelberg.

Volz, J., Bizer, C., Gaedke, M., and Kobilarov, G. (2009). Silk - a link discovery framework for the web of data. In LDOW.