Deutsche Nationalbibliothek
Adickesallee 1
60322 Frankfurt am Main

Enriching the German National Library's Linked Data Service with Geographic Coordinates : Approach and Lessons Learned (so far)

Lars G. Svensson

Abstract
The German National Library (Deutsche Nationalbibliothek, DNB) publishes geographic coordinates for approx. 40,000 geographic entities and 20,000 maps and charts. The data is available in several formats, including MARC 21, CSV and RDF. This paper discusses the design issues considered when choosing which RDF vocabulary to use and why the preliminary choice is to use geosparql and simple features. The paper is concluded with some examples of the current implementation and an outlook at some future work.

Introduction

Online services like GoogleMaps or Open StreetMap have changed the way users interact with geographic information. In order to improve the integration of library resources within those information environments, libraries need to provide hooks in the form of geographic coordinates for two kinds of resources: maps/charts and geographic entities.
Effective from January, 2014, the DNB publishes geographic information about maps, charts and geographic entities curated by the library. This service will be continually developed and augmented in order to make use and re-use of the data as efficient as possible.
The rest of this paper is structured as follows:

The first section is an introduction to the data curated by the DNB, the metadata services provided and to the amount of geographic information available. That section is followed by information about the DNB's linked data service. The third part is a survey of the RDF vocabularies evaluated as part of modelling of the geographic information in RDF. The paper is concluded with some examples of geographic data in RDF, some use cases interested in consuming this data and a look at some future work.

DNB Metadata Services

The data the DNB curates is of two kinds: bibliographic information and authority data. While the bibliographic information is data describing the approx. 18 million books stored in the library's stacks, the authority data describes the entities surrounding the book, such as persons (as authors or as subjects of scientific discourse), places, subject, geographic entities etc. Currently, the DNB publishes the entities from both of those information sets as structured data with various levels of completeness:

- MARC (in the flavours MARC 21 and MARC-XML) – data formats specific to the library domain – feature complete bibliographic and authority records including all internal cross-links and external references

- DNB Casual – title data available in CSV and XML – is a structured format with all content presented as literals, i. e. there is no cross-linking and also no external references to related data

- RDF – as the linked data format – contains *almost* the complete information contained in the authority and bibliographic records, except for information specifically geared toward libraries (the RDF data is seen as a way to connect to non-library organisations).

All data – i. e. bibliographic and authority data – published in DNB Casual and RDF is available under a CC0 public domain dedication license and thus open data. The same applies to authority data in MARC. For bibliographic data in MARC, the DNB uses a moving wall: This moving wall is based on the so-called "volume number" of the Deutsche Nationalbibliografie (the former number

of the weekly index) and is currently set at 31 December 2012. Data indexed after this date is available against a fee; data from earlier bibliographic years is available under CC0. Exceptions to this rule are titles in the so-called Series O (online publications) which are also available under CC0. It is anticipated that this moving wall will remain until mid-2015, after that all data in all formats will be under CC0.

Maps and Charts

As part of its legal deposit, the German National Library collects maps and charts published in or covering Germany. Currently, the DNB has slightly more than a quarter of a million maps in its collection. Since 2010 the descriptive cataloguing of maps and charts also contains coordinates describing the geographic area the publication covers; today about 20,000 maps have this information. For smaller areas, e. g. street-maps, we use midpoint coordinates, for larger areas the coordinates are catalogued as a bounding box expressed with westbound and eastbound longitudes, and southbound and northbound latitudes.

Geographic Entities

In the German-speaking countries, information about geographic entities is part of the Integrated Authority File (Gemeinsame Normdatei, GND), a collection of data records representing persons, corporate bodies, congresses, geographic entities, topics and works. The GND contains information about approximately 300,000 geographic entities, divided into several subtypes, including administrative units, religious territories, borders, roads and fictitious places. Until recently, the records for those entities contained mainly textual– name, description – and hierarchical – part-of – information.

As a first step to enrich this information, we added geographic midpoint coordinates for approximately 40,000 places which we derived from GeoNames. This was achieved by matching the names in the GND with information from GeoNames. In order to increase precision, the matching algorithm considered not only the common name for the entity, but also other factors such as geographic neighbours, hierarchies and the country code.

It is planned to continuously improve the quality of this information, both through the integration of new data sources, such as the Getty Thesaurus of Geographic Names (TGN), and the extension of the data to include not only midpoint coordinates, but also polygons.

The DNB Linked Data Service

The DNB publishes both bibliographic and authority data as linked data. The service went live in 2010 with an experimental modelling of the authority data. In early 2012 bibliographic data was added to the service and in April of the same year, the modelling of the authority data was changed due to the introduction of the GND.

The GND data is modelled using the GND ontology,[1] an OWL vocabulary engineered to mirror the GND data model in RDF. This has the advantage that the transformation of the data into RDF is fairly straight-forward, since the internal and the external data models are closely related and any specific features of the original data can be seamlessly integrated into the RDF representation. Potential drawbacks are that data consumers need to engage with the vocabulary before accessing the data; in order to reduce this effort, parts of the vocabulary has been mapped to well-known element-sets such as dcterms, foaf and SKOS.

The bibliographic data is modelled using an application profile based on a mixture of vocabularies.[2] This dataset contains the DNB's main holdings (currently excluded are printed music and the holdings of the German Exile Collections) and the serial publications (journals, newspapers and periodicals in the German Union Catalogue of Serials (ZDB)). The full complexity of the bibliographic records is not represented in RDF; rather those elements that are required for the identification of the resource have been selected for the linked data representation. The modelling follows the core element set recommendations of the DINI WG KIM Bibliographic Data group V.1.0.[3]

The RDF data is available through URI dereferencing – currently available in RDF/XML with Turtle and JSON-LD coming later this year – and through bulk download.[4]

---

[1] http://d-nb.info/standards/elementset/gnd

[2] For more information on the RDF modelling of the bibliographic data cf. http://www.dnb.de/SharedDocs/Downloads/EN/DNB/service/linkedDataModellierungTiteldaten.pdf?__blob=publicationFile

[3] Empfehlungen zur RDF-Repräsentation bibliografischer Daten: urn:nbn:de:kobv:11-100212769 (only available in German)

[4] For more information cf. http://dnb.de/EN/lds

Vocabularies for Representing Geographic Information in RDF
Initially, we examined existing ontologies and vocabularies for the representation of geographic coordinates in RDF. The list is probably not exhaustive but should contain the most important ones for the cultural heritage domain.
Examples are shown in Turtle and use the following set of prefixes:

```
@prefix gndo: <http://d-nb.info/standards/elementset/gnd#>
@prefix dc: <http://purl.org/dc/elements/1.1/>
@prefix dct: <http://purl.org/dc/terms/>
@prefix rdaGr1: <http://rdvocab.info/Elements/>
@prefix wgs84_pos: <http://www.w3.org/2003/01/geo/wgs84_pos#>
@prefix geo: <http://www.opengis.net/ont/geosparql#>
@prefix geom: <http://geovocab.org/geometry#>
@prefix sf: <http://www.opengis.net/ont/sf#>
@prefix marcrel: <http://id.loc.gov/vocabulary/relators/>
@prefix bnf-onto: <http://data.bnf.fr/ontology/bnf-onto/>
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
@prefix rdarelationships: <http://rdvocab.info/RDARelationshipsWEMI/>
@prefix gn: <http://www.geonames.org/ontology#>
@prefix lgdo: <http://linkedgeodata.org/ontology/>
@prefix spatial:  <http://geovocab.org/spatial#>
```

GND Ontology (GNDO)

The GND-Ontology contains properties for the designation of geographic coordinates for places: gndo:easternmostLongitude, gndo:westernmostLongitude, gndo:northernmostLatitude and gndo:southernmostLatitude. All those properties are subproperties of gndo:coordinates and have as their domain gndo:PlaceOrGeographicName. This implies that the geographic coordinates represent the real-world place and not a map or a metadata record representing the place. In order to use the gndo-properties to describe the extent of a map or a chart, an intermediate object representing the geographic feature is needed. Representation of point coordinates is done by displaying the longitude and latitude twice. The property gndo:typeOfCoordinates can be used to specify if the coordinates are analogue or decimal. Technically, the GND format supports polygons; the GNDO – however – does not and would need to be extended to allow the representation of polygons. There are no known RDF implementations using the geographic features of the GNDO.

RDA (Resource Description and Access) Element set

The upcoming standard for bibliographic descriptions Resource Description and Access (RDA)[5] is accompanied by an RDF vocabulary mirroring the semantics of the cataloguing rules. The RDA group one elements (rdaGr1)[6] contains, among others, properties for the description of maps, such as rdaGr1:projectionOfCartographicContent to describe the projection (e. g. Mercator or Hammer-Aitov) and rdaGr1:coordinatesOfCartographicContent for the exact coordinates. The RDA cataloguing rules suggest the recording of the coordinates as a string using westernmost, easternmost, northernmost and southernmost boundaries. Alternatively, a string of coordinate pairs can be entered to describe a polygon. The RDA instructions give a preference for the use of analogue coordinates.
The description of geographic entities is part of the RDA Group 3 Elements.[7] That element set contains no properties for geographic coordinates.
The BnF linked data service uses rdaGr1 to publish coordinates for maps in RDF.

Example
Map of Angoulême from the BnF:[8]

```
<http://data.bnf.fr/ark:/12148/cb40860726q>
      a <http://rdvocab.info/uri/schema/FRBRentitiesRDA/Manifestation> ;
      dct:date "1815" ;
      bnf-onto:FRBNF "40860726"^^<http://www.w3.org/2001/XMLSchema#integer> ;
      dct:publisher "1815" ;
      rdaGr1:scale "[1:86400] ou 1 ligne pour 100 toises" ;
      rdaGr1:dateOfPublicationManifestation <http://data.bnf.fr/date/1815/> ;
```

[5] http://www.rda-jsc.org/rda.html
[6] http://rdvocab.info/Elements/ Note that the rdaGr1 elements are in the process of moving to a new domain and as part of that action the new vocabularies will use opaque URIs. The old classes and properties will be deprecated and there will be redirects in place. The underlying semantics will not change.
[7] http://rdvocab.info/ElementsGr3/
[8] http://data.bnf.fr/ark:/12148/cb40860726q

```
        dct:description "1 carte : 60,5 x 94 cm" ;
        rdaGr1:Note
              "Note : Carte levée entre 1766 et 1768 par Louis Capitaine. Gravée par
        Capitaine et par Bourgoin pour la lettre" ;
        rdaGr1:coordinatesOfCartographicContent
              "W 13' - E 49' / N 45°58' - N 45°33'" ;
        dct:title "[Angoulême. Nouv. éd.]. N°69" ;
        rdfs:seeAlso <http://catalogue.bnf.fr/ark:/12148/cb40860726q> ;
        rdarelationships:electronicReproduction
              <http://gallica.bnf.fr/ark:/12148/btv1b7711743b> .
```

wgs84_pos

wgs84_pos is a W3C recommendation for the publication of geographic coordinates in RDF.[9] The vocabulary is designed to be simple and has properties for longitude, latitude and altitude above the reference ellipsoid. The vocabulary is limited to point coordinates and does not allow the description of rectangles or polygons. Longitude and latitude are explicitly to be described using decimal degrees. Domain of the properties is wgs84_pos:SpatialThing; this implies that the coordinates are those of the real world object and not those of a map or a data record describing the place.
wgs84_pos is widely used. Among the users are GeoNames[10] and BnF (for geographic entities)[11]. The Conference of European Research Libraries (CERL) refers to wgs84_pos in their metadata profile.[12]

Example
Description of Singapore from GeoNames:[13]

```
<http://sws.geonames.org/1880251/> a gn:Feature ;
        rdfs:isDefinedBy <http://sws.geonames.org/1880251/about.rdf> ;
        gn:name "Singapore" ;
        # omitting several names here...
        gn:featureClass gn:A ;
        gn:featureCode <http://www.geonames.org/ontology#A.PCLI> ;
        gn:countryCode "SG" ;
        gn:population "4701069" ;
        wgs84_pos:lat "1.36667" ;
        wgs84_pos:long "103.8" ;
        gn:parentFeature <http://sws.geonames.org/6255147/> ;
        gn:childrenFeatures <http://sws.geonames.org/1880251/contains.rdf> ;
        gn:neighbouringFeatures <http://sws.geonames.org/1880251/neighbours.rdf> ;
        gn:locationMap <http://www.geonames.org/1880251/republic-of-singapore.html> ;
        gn:wikipediaArticle <http://en.wikipedia.org/wiki/Singapore> ;
        rdfs:seeAlso <http://dbpedia.org/resource/Singapore> .
```

DCMI Point and DCMI Box

The DCMI Point[14] and Box[15] Encoding Schemes are syntax encoding schemes that can be used to encode a geographic point or the extent of a geographic feature. Those two schemes were created before the DC element set was refactored to RDF and can probably be considered obsolete today. The BnF uses the non-existing properties dct:NorthLimit etc. that are probably derived from DCMI Box to publish coordinates for geographic entities.

Geosparql and WKT

In order to allow for complex SPARQL queries over geographic information sets, the Open Geospatial Consortium (OGC) developed the SPARQL Extension geosparql.[16] Part of the reference implementation was an RDF vocabulary for the description of geographic features on the basis of Simple Features. geosparql supports different serialisations of the coordinates, e. g. Well Known Text [WKT] specified in Simple Features or Geografy Markup Language [GML] specified in OGC 07-036, with WKT[17] being the most commonly used serialization. Users of geosparql include the

---

[9] http://www.w3.org/2003/01/geo/
[10] http://www.geonames.org/ontology/
[11] Cf. e. g. http://data.bnf.fr/15248454/angouleme__charente__france_/rdf.xml
[12] http://www.cerl.org/resources/cerl_thesaurus/editing/format/123
[13] http://sws.geonames.org/1880251/
[14] http://dublincore.org/documents/dcmi-point/
[15] http://dublincore.org/documents/dcmi-box/
[16] http://www.opengeospatial.org/standards/geosparql
[17] http://en.wikipedia.org/wiki/Well-known_text

National Geographic Institute of Spain and the LinkedGeoData Project; the UK Ordnance Survey has mapped their own vocabulary to geosparql and might gradually migrate to geosparql. Several database systems and search engines (e. g. MySQL und SolR 4) support WKT natively.

Example
Description of Liebigstraße 24, Dresden, from LinkedGeoData[18]

```
<http://linkedgeodata.org/triplify/node264695865>
       a lgdo:Pub , spatial:Feature , lgdo:Amenity , lgdm:Node ;
       rdfs:label "B'liebig" ;
       geom:geometry lgd-geom:node264695865 ;
       lgdo:addr%3Acity "Dresden" ;
       lgdo:addr%3Acountry "DE" ;
       lgdo:addr%3Ahousenumber "24" ;
       lgdo:addr%3Astreet "Liebigstraße" ;
       # omitting some information
       dct:modified "2011-02-19T00:29:37"^^xsd:dateTime ;
       geo:lat 5.10328E1 ;
       geo:long 1.3721598100000001E1 .

lgd-geom:node264695865
       a geom:Geometry ;
       geo:asWKT "POINT(13.721598100000001 51.0328)"^^ogc:wktLiteral .
```

Evaluation

The main reason the DNB publishes linked is to make the data easier accessible for users outside of the library domain (libraries will probably be better served with MARC 21). Thus the key requirements were broad acceptance – and ideally curated by a recognised standards organisation – and the possibility to use the same vocabulary for both geographic entities and for maps (represented by points and/or polygons) so that a data consumer can query both data sets with one query.

GNDO
The features for the description of geographic entities in the GNDO are not widely used. The advantage of the GNDO is that we have full control over the vocabulary and that it is easy to represent the GND data using it. In order to ensure interoperability it would be necessary to align the relevant GNDO terms with other vocabularies. The vocabulary is geared towards authority data, its usefulness for maps and charts needs to be assessed.

RDA
The RDA elements for geographic descriptions are currently used by the BnF who is one of the major players in linked library data; there are no known implementations outside of the library domain. Currently, the RDA element set has detailed terms for the description of maps, but none for geographic features.

wgs84_pos
wgs84_pos is curated by the W3C and widely used. It only supports point coordinates, and cannot be used for e. g. polygons.

DCMI Point/Box
Those syntax encoding schemes do not seem to be widely used and they can probably be considered obsolete.

geosparql
geosparql/simple features is an ISO and OGC standard developed to support the description of all kinds of geographic features, so that it can be used for both point coordinates and polygons. The vocabulary is used in large-scale implementations outside of the library domain and offers seamless database and search-engine integration.

Implementation
The preliminary decision was to use geosparql/simple features for both geographic entities and bibliographic descriptions. The main motivation was the support for both points and polygons – and thus the possibility to use the same vocabulary for geographic features and for maps and charts – the wide-spread adoption and the potentially easy integration of the data into geographic information systems.

[18] http://linkedgeodata.org/page/triplify/node264695865

For the authority data this decision is a move away from the current policy to only use the GNDO to describe data from the GND. In that respect, the decision to use geosparql is not final and might be revised in favour of the use of the GNDO. This would require the addition of new terms to the vocabulary and the alignment of those new terms to e. g. geosparql and wgs84_pos.

Examples

The following two examples (a chart and a geographic entity) show how the geographic coordinates are displayed using WKT: the chart extent as a WKT polygon and the midpoints coordinates of the town as a WKT point. For the geographic entity, an owl:sameAs link to the corresponding GeoNames entity is provided as well.

A chart of the Baltic Sea:[19]

```
<http://d-nb.info/1043451617> a bibo:Map , bibo:Document ;
	owl:sameAs <http://hub.culturegraph.org/resource/DNB-1043451617> ;
	bibo:isbn13 "9783869876023" ;
	bibo:gtin14 "9783869876023" ;
	dcterms:language <http://id.loc.gov/vocabulary/iso639-2/ger> ;
	dcterms:isPartOf <http://d-nb.info/551113464> ;
	dc:title "[Deutsche Seekarte]" ;
	isbd:P1053 "2 Kt. auf Vorder- und Rücks." ;
	dcterms:spatial [
		a sf:Polygon ;
		geo:asWKT "Polygon (( +009.333333 +055.916666,
			+009.333333 +053.416666,
			+015.200000 +053.416666,
			+015.200000 +055.916666,
			+009.333333 +055.916666 ))"^^geo:wktLiteral
	] ;
	dcterms:subject <http://d-nb.info/gnd/4044107-6> ,
		<http://d-nb.info/gnd/4179589-1> , <http://d-nb.info/gnd/4182476-3> ;
	dcterms:issued "2014" ;
	rda:otherTitleInformation
		"3002. Planungskarte für die Klein- und Sportschifffahrt Ostsee" .
```

The description of the city Aalen:[20]

```
<http://d-nb.info/gnd/4000015-1> foaf:page <http://de.wikipedia.org/wiki/Aalen> ;
	gndo:gndIdentifier "4000015-1" ;
	gndo:oldAuthorityNumber "(DE-588)2003538-X" ,
		"(DE-588b)2003538-X" , "(DE-588c)4000015-1" ;
	owl:sameAs <http://sws.geonames.org/2959927> ;
	geo:hasGeometry [
		a sf:Point ;
		geo:asWKT "Point ( +010.093299 +048.837769 )"^^geo:wktLiteral
	];
	gndo:geographicAreaCode
		<http://d-nb.info/standards/vocab/gnd/geographic-area-code#XA-DE-BW> ,
		<http://d-nb.info/standards/vocab/gnd/geographic-area-code#XA-DXDE> ;
	gndo:homepage
		<http://www.aalen.de/sixcms/detail.php?template=d_aa_gl_startseite&_be
	reich=6> ;
	gndo:definition
		"Kreisstadt des Ostalbkreises, im Osten Baden-Württembergs"@de ;
	gndo:variantNameForThePlaceOrGeographicName "Stadt Aalen" ;
	gndo:preferredNameForThePlaceOrGeographicName "Aalen" ;
	a gndo:TerritorialCorporateBodyOrAdministrativeUnit .
```

Use Cases

Deutsche Digitale Bibliothek

The Deutsche Digitale Bibliothek (German Digital Library, DDB)[21] is the national platform for German digital cultural heritage. It collects metadata and derivative media files for digitized or born-digital cultural heritage objects from all cultural domains and offers central access via its web portal and its API. The DDB uses the Europeana Data Model with a DDB-specific application profile. Today only few collections provided by DDB's partner institutions contain geographic coordinates. A greater percentage uses URIs for geographic entities, with those URIs mostly linking to the GND, but also to other authority files. A third group of collections contains street address information

[19] http://d-nb.info/1043451617
[20] http://d-nb.info/gnd/4000015-1
[21] http://www.deutsche-digitale-bibliothek.de/

(literals) which DDB plans – in collaboration with the DNB - to match with official registry data from the federal states' survey offices to enrich the existing data with URIs and/or geographic coordinates.
Based on the GND URIs, the DDB will introduce a search facet for location role. The role will be differentiated into (current) location, point of origin, and several eventPlace types from the LIDO specification. Another application will be entity pages for locations as exploratory entry points for end users, where information relating to the specified location both from inside the DDB as well as from external sources will be presented. The external URLs are provided by a dedicated web service called "entity facts" operated by the DNB.
In a showcase demonstrating the features of the API, the DDB will implement a map-based search for the collection of monuments and historic buildings in the state of Hesse. Objects that fit both the entered search term and a chosen bounding box will be displayed by a web based mapping client, optimized for mobile devices. With further utilization of the coordinates provided by the GND it is planned to extend this service to other collections and to incorporate the spatial search into the DDB portal.

Further development of the ZDB

The Zeitschriftendatenbank (German National Union Catalogue of Serials, ZDB)[22] is one of the world's largest specialized databases for serial titles (journals, annuals, newspapers etc.). It contains more than 1.6 million bibliographic records of serials from the 16th century onwards, from all countries, in all languages, held in 4.400 German and Austrian libraries, with 12.9 million holdings information. It is maintained and further developed by the Berlin State Library (Staatsbibliothek zu Berlin – Preußischer Kulturbesitz) and the DNB. Within the frame of a one-year project funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), the ZDB is further developed as control tool for digitization projects on the one hand and as research tool for newspapers on the other hand. A new interface for the catalogue is developed that especially supports spatial and temporal queries. The coordinates in the GND are intended to be used for a geographic visualization of the distribution places of newspapers that are linked to the GND data. Furthermore, coordinates stored in the directory of addresses of libraries and related organizations (ISIL- und Sigelverzeichnis) are intended to be utilized for visualizing the location of all those institutions that have a specific title.

Future work

Parallel to the evaluation of the use of the geographic information in the DNB datasets, we will continue to improve the alignment of the geographic features in the GND with GeoNames and to start aligning them with other geographic datasets such as the TGN or the German State Survey Offices. Part of this process is to evaluate the possibilities to store not only points and rectangles, but also true polygons. Further we will decide if we use the GNDO to describe geographic features from the GND and what changes will need to be done.

Conclusion

Cross-domain information reuse requires that the data is supplied in a format allowing a broad variety of data consumers to process the information contained. A corollary of this is to move away from vocabularies originating in a specific supplier domain and instead focus on what kind of data is described and use a vocabulary suitable for that piece of information. Together with the possibility to encode both points and polygons using the same vocabulary, this was one of the major drivers behind the choice to use geosparql and simple features.

[22] http://www.zeitschriftendatenbank.de/