The Linked Data Service of the German National Library

Version 2.1

September 28th, 2010

Introduction

Developments in the area of the semantic web are aimed at improving the usability and accessibility of data. The idea of the Semantic web also allows links to be created between data from heterogeneous sources, leading in turn to the establishment of new services. As a result of the linked-data movement, many providers (mostly non-profit organisations, universities or public institutions) are already offering their data in a form which is semantic-web-compatible. Above all, this includes data which is of general use within the public domain. Examples include geographical information, thesauri, encyclopaedias as well as bibliographic and authority data.

Libraries, too, have recognised the great potential offered by this form of data publication. The first institutions are already actively offering their information as linked data, or are planning to do so. The German National Library is committed to making a significant contribution to ensuring the stability and reliability of the "linked-data-cloud" by providing data which has largely been generated and maintained by trained professionals. The German National Library with its high quality data intends to become one of the mainstays of the semantic web.

In the long term the German National Library is planning to offer a linked data service which will permit the semantic web community to use the entire stock of its national bibliographic data, including all authority data. A suitable data service needs to be created to distribute the new data format alongside the already established access channels (OAI, SRU etc.).

One of the aims of the service will be to attract new target groups and accordingly it is important for the project to analyse their requirements in detail, making contact with them in order to identify their precise needs. The proposal, therefore, is to launch a beta service which is based primarily on past experience and users requirements. The beta service described in this documentation is aimed at establishing an initial partnership with this new clientele as a means of sounding out each other's views. In the medium-term, the target groups will be expanded to include commercial service providers such as operators of search engines and knowledge management systems alongside research institutions and non-profit organisations. The German National Library is endeavouring to make a significant contribution to the global information infrastructure with its new data service by laying the foundations for modern commercial and non-commercial web services.

Note regarding access

The German National Library has been making parts of its knowledge base available via interfaces (OIA, SRU) for some time now. As part of its linked data activities, the German National Library is also aiming to provide RDF data via these interfaces (see Figure 1). Initially, however, we are only offering HTTP resolving via the German National Library’ web portal.

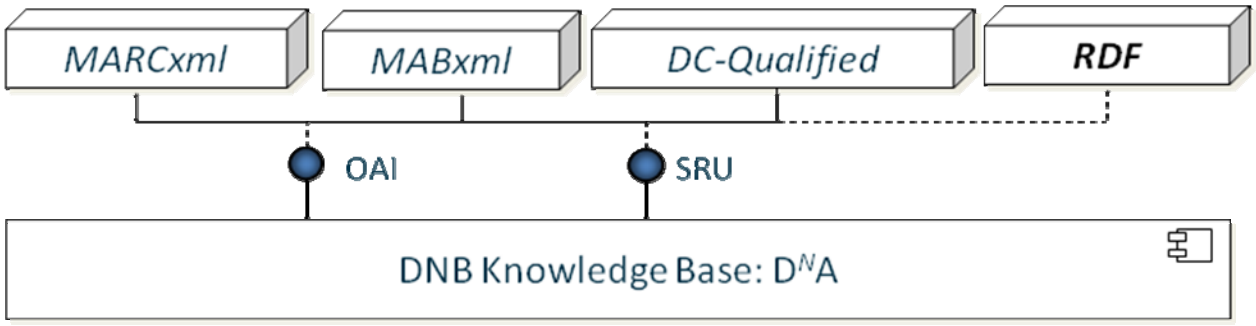

Figure 1: The system interfaces of the German National Library (DNB)

The Linked Data Service is already incorporated in the web portal and therefore publicly accessible. There are two basic methods of accessing data in RDF/XML: with and without redirection (see below). FTP downloads of the entire data are also offered.

Scope, release date and use of the data pool

Scope

The data pool currently represents 1,797,911 differentiated persons from the PND (Personal Name Authority File), 187,743 subject headings from the SWD (Subject Headings Authority File) and 1,320,711 corporate bodies from the GKD (Corporate Body Authority File). In addition, 51,748 classes and 110 subject headings of the German Dewey Decimal Classification (DDC) are represented. 40,182,561 triples were needed internally to represent the data.

Release date

The data pool represents the German National Library's data as it was on 23.08.2010 (SWD, DDC) and 28.4.10 (PND, GKD), respectively. This means that data records added or changed as of this date are *not* represented correctly.

The commencement of this BETA service saw the data and its modelling reach a level of stability. The Linked Data Service is, however, work-in-progress, meaning that adjustments and additions are possible in the longer term. Changes to the data and the modelling will then be documented here, or will be apparent from a comparison of the respective PICA/RDF mappings.

Version history

- 2010-08-30 (current Version, BETA service)
    - Inclusion of the German DDC
    - Links SWD -> DDC (with 4 degrees of equivalency, based on results of the CrissCross Project) included
    - Update of the SWD data (from 23.08.2010); no update of PND and GKD at this point
- 2010-04-28 (BETA service)

    - Improved modelling of PND and SWD (including adjustment of GKD properties), initial modelling of GKD. See below.
    - Links to LCSH, RAMEAU added
- 2010-03-31 (prototype)
    - Modelling of PND and SWD (subject headings). See PICA/RDF mappings (prototype).
    - Links to Wikipedia, DBPedia, VIAF established

Use of data

The non-commercial use of this linked data is free of charge. Commercial use is also possible, but the German National Library wants to be informed of such uses. A detailed version of the conditions of use is currently being drawn up.

Access without redirect

The RDF/XML representation of an authority record is stored with its identifier <IDN> at http://d-nb.info/gnd/<IDN>/about.

Example: The author, theatre director and dramatist Bertolt Brecht (IDN: 118514768) has the URI http://d-nb.info/gnd/118514768. Currently, this is also the URL for the HTML representation of the relevant data about the person. The corresponding RDF/XML full display can be found at http://d-nb.info/gnd/118514768/about

Individual persons can be found by searching for them in the portal and then selecting the appropriate listing. There is a link on the right of the full display of individualised persons ("RDF/XML-Repräsentation dieses Datensatzes") which can be clicked to obtain the RDF/XML representation.

Subject headings and corporate bodies can be found in the same way. In the portal search, the search filters “Schlagwörter” ("Subject Headings") and “Organisationen” ("Organisations") can be used correspondingly under “Alle Normdaten” ("All authority files") on the right of the portal search results.

Example: The job title “Wirtschaftsinformatiker” ("Business IT specialist", IDN: 4190055-8) can be found at http://d-nb.info/gnd/4190055-8, the corresponding RDF/XML representation at http://d-nb.info/gnd/4190055-8/about

Example: The organisation “Deutsche Nationalbibliothek” ("German National Library", IDN: 10140798-1) can be found at http://d-nb.info/gnd/10140798-1, the corresponding RDF/XML full version at http://d-nb.info/gnd/10140798-1/about

Access with redirect

Clients can signal that they would like to receive the answer in RDF/XML and not e.g. in HTML by using an HTTP request an accept value of "application/rdf+xml". Request headers can be "modified" for test purposes e.g. with Firefox plug-ins such as "Modify Headers".

The accept header information is assessed in the resolver. If "application/rdf+xml" is found, it is redirected from http://d-nb.info/gnd/<IDN> to the address http://d-nb.info/gnd/<IDN>/about and the RDF/XML representation shown.

Example: The data record for the composer “Johann Sebastian Bach” has the following URI: http://d-nb.info/gnd/11850553X. A request containing "application/rdf+xml" in the Accept Header is automatically redirected to http://d-nb.info/gnd/11850553X/about where the RDF/XML representation of the person is held.

FTP access

The entire data pool is available in N-Triple format on the FTP server of the German National Library. It is not possible to offer other formats at present. It should be noted that the data pool is very large and that the voluminous XML representation adds further to its bulk. The entire uncompressed data dump is just over 4.1 GB. Older file systems or ZIP programs may not be able to handle files of this size, which can lead to error messages during unpacking.

- Example (extract from data pool, approx. 500 KB)
- Entire data pool (N-Triple format, approx. 455 MB ZIP file, date 28.4.10)
- SWD (update) and DDC data (N-Triple format, about 24 MB ZIP file, date 23.08.2010)

Please note the above remarks on using the data.

Ontology modelling

One of the key criteria in ontology modelling for the authority data of the German National Library was the ability to make extensive use of appropriate elements of ontologies already registered. The following ontologies were examined as the basis for selecting suitable ontology components:

- ABC-Ontology[1]
- Bibliographic Ontology Specification[2]
- BIO: A vocabulary for biographical information[3]
- DCMI Metadata Terms[4]
- FOAF Vocabulary Specification[5]
- GeoNames Ontology[6]
- MarcOnt[7]
- Music Ontology[8]
- Ontologie der Wissenschaftsdisziplinen[9]

1 http://metadata.net/harmony/Results.htm
2 http://bibliontology.com
3 http://vocab.org/bio/0.1/.html
4 http://dublincore.org/documents/dcmi-terms/
5 http://xmlns.com/foaf/spec/
6 http://www.geonames.org/ontology/
7 http://www.marcont.org/ontology/index.html#term_documentation/marcont.html
8 http://musicontology.com
9 http://owd.hu-berlin.de/ontologie.php#modell

- Provenance Vocabulary Core Ontology[10]
- RDA Element Sets[11]
- Relationship Vocabulary[12]

Given the high likelihood that both the ontology and the RDA (Resource Description and Access) cataloguing rules will be used in the future by libraries, the modelling of the ontology being presented here is oriented primarily towards the RDA element set. The RDA element set permit FRBR-compliant entity representation, which is the target for the German National Library's Linked Data Service. The FRBR entities are currently specified by the "FRBR Entities for RDA". This elements do not, however, represent a conclusive solution because the FRBR Review Group has not yet released an official definition for FRBR entities and the FRBR relationships. Equivalent elements in the RDA element set will be replaced on publication of the FRBR ontology.

The potential for practical use and exclusion criteria for the use of all listed ontologies are discussed below. The discussion examines in particular the RDA Group 2 entities (person, family, corporate body), as these are to be realized as a semantic web compliant representation at the start of the Linked Data Service project.

The ABC ontology is based on FRBR concepts but does not, however, permit a finely granular description of person, family or corporate body entities. Therefore, the ontology was not taken into consideration in the modelling.

Non-model compliant approaches (FRBR and FRAD were ignored) were chosen to describe bibliographic metadata in the Bibliographic Ontology Specification. This ontology is therefore not suitable for defining Group 2 FRBR entities.

The Bio-Ontology is currently in a rudimentary and relatively unstable state, which would also prohibit practical use. Also, most of the elements specified in this ontology are covered by the RDA element sets. The Bio-Ontology does, however, contain an element - "Marriage" - which needs to be examined more closely for a later version, as no statements can be made concerning marriages between two persons, or their duration, in the current ontology.

The DCMI Metadata Terms have not yet been considered for ontology modelling. They could, however, be useful for specifying information about metadata (meta-metadata). The extent to which the DCMI Metadata Terms can be utilised in implementing FRBR-ised bibliographic data should also be investigated. This should be checked during the course of the project.

One of the most widespread ontologies in the semantic web is the FOAF vocabulary. Despite its widespread distribution, it emerged during the ontology modelling that the vocabulary is only partially suitable for use by libraries. The entities are defined in a similar way to FRBR, however the conceptions for naming are less than ideal. Similar to the RDA ontology, in FOAF it is possible to give a name as a literal, however FOAF

10 http://trdf.sourceforge.net/provenance/ns.html

11 http://metadataregistry.org/rdabrowse.htm

12 http://purl.org/vocab/relationship

provides no possibility to distinguish between a person's preferred and variant names. FOAF has certain advantages over RDA with regard to the granularity of name forms. Its use of "familyName", "givenName", "firstName" and "lastName" elements facilitates differentiated identification of name parts and permits a large number of international name types. The disadvantage of such differentiation is that the elements have been specified with the domain "person". This means that, when identifying a number of different name forms for a person, it is not possible to assign these parts to a specific name.

More than 6.2 million places are currently described geographically by the GeoNames Ontology. This ontology has not yet been considered for the current service, although its potential for linking geographical information has been recognised. In the future, referencing could permit existing literals (names of places, etc.) to be replaced to allow the non-ambiguous identification of the referenced geographical information.

MarcOnt is based on the classic MARC format. This means that the ontology does not include any FRBR concepts. For this reason it was not included for the implementation of the authority data. Whether or not the ontology can supplement the RDA element sets in certain cases in the modelling of bibliographic data should be investigated more thoroughly in the future.

The Music Ontology has not been included, either, in the modelling for the time being. It is highly likely to be utilised alongside the RDA elements sets for representing the collections of the German Music Archive. The suitability of incorporating the Music Ontology in the authority data will be thoroughly investigated during the ontology modelling for representing the bibliographic data of the German Music Archive.

The academic disciplines ontology has not yet been included. The ontology is to be scrutinised more carefully for a future version as it may be able to offer controlled representation of the academic degrees which an individual has successfully completed. The ontology could also play a role in representing university publications (theses, etc.) in the future.

The service currently published does not yet identify any meta-metadata. The integration of meta-metadata is being considered for a subsequent version and will investigate the incorporation of the Provenance Vocabulary alongside the DCMI metadata terms.

The Relationship Vocabulary allows relationships between persons, families and organisations to be expressed in more granular terms than is currently possible using the RDA relationships within Group 2. The data of the German National Library already reflects this differentiation (e.g. the kind of family relationship between two persons), meaning that it would be desirable to maintain this labelling of characteristics. The Relationship Vocabulary specifies a range of basic relationship definitions for utilisation. Apart from this it should be examined whether the vocabulary meets all the relevant demands or whether it would be advisable to include further elements to permit greater precision in describing relationships.

Transformation of authority data

In order to transform the German National Library's data into RDF it was first necessary to express the existing data structures in semantic-web-compliant vocabulary. Consequently, an essential consideration for the Linked Data Service project is how to

map the German National Library's authority data onto approproriate ontologies. The authority data concerned:

- Gemeinsame Körperschaftsdatei (Corporate Body Authority File, GKD)
- Personennamendatei (Name Authority File, PND) and
- Schlagwortnormdatei (Subject Headings Authority File, SWD)

The German National Library's internal format (Pica+) is used as the basis for generating concordances between the authority data and a semantic-web-compliant representation. The description of the relevant internal formats can be found in the corresponding field directories (GKD[13], PND[14], SWD[15]).

The new service includes transformation of the GKD, the PND and parts of the SWD. Particular importance was attached here to crosslinks within the given authority file and between the individual authority files. In addition, external links from the PND to DBpedia, Wikipedia and the Virtual International Authority File (VIAF), and from SWD subject headings to the corresponding equivalents in the Library of Congress Subject Headings (LCSH), RAMEAU and the German Dewey Decimal Classification (DDC) have been included.

The modelling of person and corporate body descriptions and the representation of relationships existing between the entities were based on the

- RDA Group 2 elements[16] and the
- RDA Relationships for Persons, Corporate Bodies, Families[17]

Elements of additional vocabularies such as FOAF and the Relationship Vocabulary were incorporated for data which could not be adequately represented by RDA. Consequently, the type of described resources of the person or corporate entity is based on the defined domain of the properties used.

Elements of the following vocabularies were used to model the Subject Headings Authority File (SWD):

- Simple Knowledge Organisation System (Skos)[18]
- Dublin Core[19]

Moreover, while working out the mappings, characteristics were identified which could not be represented by any of the ontologies presented. Special elements were specified under the GND (Gemeinsame Normdatei) namespace for the representation of these characteristics. These elements are not yet registered. Consequently, the German National Library reserves the right to adjust the way in which these are used. In the longer term, however, the intention is to anchor the library's own registered elements in

13 http://support.d-nb.de/iltis/feldverzeichnis/Normdaten_GKD_endf.pdf
14 http://support.d-nb.de/iltis/feldverzeichnis/Normdaten_PND_endf.pdf
15 http://support.d-nb.de/iltis/feldverzeichnis/Normdaten_SWD_Cross.pdf
16 http://metadataregistry.org/schema/show/id/15.html
17 http://metadataregistry.org/schema/show/id/22.html
18 http://www.w3.org/TR/skos-reference/skos.html
19 http://dublincore.org/documents/dcmi-terms/

the ontology modelling to ensure maximum quality and interoperability amongst the data records.

Implementation of the GKD

The elements for representing a corporate body description as implemented in the service are shown in Table 1. The ontology source of the properties used is indicated by the prefix. Registered elements are also linked to their description in the Web. The table contains extracts of the GKD field directory.

Table 1: Elements used to describe a corporate body

PICA3	PICA+	Ind.	Field content	RDF element	Remarks
020	007P	$0	**GKD-Nummer (GKD number)**	rdaGr2:identifierForTheCorporateBody	(DE-588b)…
022	007T	$0	**LoC-Nummer (LoC number)**	rdaGr2:identifierForTheCorporateBody	(DLC)…
023	007W	$0	**SWD-Nummer (SWD number)**	rdaGr2:identifierForTheCorporateBody	(DE-588c)…
026	007G	$0	**GKD-Nummer umgelenkter Datensätze (GKD number of redirected data records)**	gnd:invalidIdentifierForTheCorporateBody	(DE-588b)…
027	007H		**Identnummer aus altem System (Identifier from old system)**		
	007H	$S	"|m|" Deutsches Musikarchiv		
	007H	$0	IDN	gnd:invalidIdentifierForTheCorporateBody	(DE-101c)…
150	029A		**Körperschaftsname in Ansetzungsform (Authorized form of name of** the **corporate body)**	gnd:preferredNameForTheCorporateBody	$a <$c> / $b <$x> / $b <$x>…
		$a	Hauptkörperschaft (Main corporate body)		
		$c	Ordnungshilfe zu Hauptkörperschaft (Qualifier for main corporate body)		
		$b	Abteilung(en) (Department(s))		

		$x	Ordnungshilfe zu Abteilung(en) (Qualifier for department(s))		
151	029B	$a	**Abkürzung der Ansetzungsform des Körperschaftsnamens (Acronym for authorized form of corporate body name)**	gnd:preferredNameAcronymForTheCorporateBody	
155	029N	$a	**Offizieller Name der Körperschaft (Official name of corporate body)**	rdaGr2:nameOfTheCorporateBody	
250	029@		**Körperschaftsname (Vorlage- bzw. Verweisungsform, Abkürzung zur Vorlage- bzw. Verweisungsform) (Name of corporate body (variant name, acronym for variant name))**	gnd:variantNameForTheCorporateBody	$a <$c> / $b <$x> / $b <$x>...
		$a	Hauptkörperschaft (Main corporate body)		
		$c	Ordnungshilfe zu Hauptkörperschaft (Qualifier for main corporate body)		
		$b	Abteilung(en) (Department(s))		
		$x	Ordnungshilfe zu Abteilung(en) (Qualifier for department(s))		
		$d	Akcronym (Acronym)	gnd:variantNameAcronymForTheCorporateBody	
300	032A	$a	**Daten der Körperschaft (Dates of corporate body)**	rdaGr2:dateOfEstablishment and rdaGr2:dateOfTermination	
325	032K		**Allgemeine Funktionsbezeichnung (General role designation)**		
		$S	"|r|" Indikator für Rolle (Role indicator)	gnd:functionOfTheCorporateBody	
		Plus			

		$9	IDN SWD-Satz (IDN SWD record)		Link to URI of SWD set by expanding the NID[20] from $0
		or			
		$a	Normierte Funktionsbezeichnung (Ansetzungsform SWD-Satz) (controlled role designation) (authorized form of SWD record)		Use literal if no IDN available
440	038F	$9	**GKD-Nummer der übergeordneten Körperschaft (GKD number of hierarchical superior)**	rdaRelGr2:hierarchicalSuperior	Link to URI of GKD set by expanding the NID from $0
450	038C		**Verknüpfung "früherer, späterer, zeitweiser Name" (Link "predecessor, successor, temporary name")**		
		$S	"|a|" früherer Name (reziproke Beziehung) ("|a|" earlier name (reciprocal relationship))	rdaRelGr2:predecessor	
			"|c|" späterer Name (reziproke Beziehung) ("|c|" later name (reciprocal relationship))	rdaRelGr2:successor	
			"|d|" früherer und späterer Name ("|d|" earlier and later name)	rdaRelGr2:predecessor and rdaRelGr2:successor	
			"|m|" früherer Name (nicht reziproke Beziehung) ("|m|" earlier name (non-reciprocal relationship))	gnd:predecessorWithoutSuccessor	
			"|o|" späterer Name (nicht reziproke Beziehung) ("|o|" later name (non-reciprocal relationship))	gnd:successorWithoutPredecessor	
			"|s|" siehe auch ("|s|" see also)	rdfs:seeAlso	

20 (international) authority data identifier

		$9	IDN GKD-Satz (IDN GKD record)		Link to URI of GKD set by expanding the NID from $0
485	009Q		**Verlinkte Ressource (Linked resource)**		
		$u	URL (Uniform Resource Locator)	foaf:homepage	The URL is associated as a resource if the entry starts with http://
		$3	führt zu: … (Homepage, Quelle, Provenienzmerkmal, …) (leads to: … (homepage, source, provenance characteristic, …)		If the field contains "Homepage", the value is from $u foaf:homepage
797	003.	$0	**Identifikationsnummer (IDN) (Identifier (IDN))**	rdaGr2:identifierForTheCorporateBody	(DE-588)…
811	042B	$a	**Ländercode nach ISO 3166** ***ersetzt 007*** (**Country code as per ISO 3166 *replaces 007***)	gnd:countyCodeForTheCorporateBody	
812	042C	$a	**Sprachencode nach ISO/TC46/SC4-N350 (Language code as per ISO/TC46/SC4-N350)**	rdaGr2:languageOfTheCorporateBody	
892	039I	$9	**ID-Nummer und Ansetzungsform des Zieldatensatzes bei** **Umlenkung von Datensätzen (Identifier and authorized form of name of target data record for redirected data records)**	owl:sameAs	

All corporate bodies are converted to a semantic-web-compatible representation as a result of the transformation of the GKD.

The integrated identifiers (*rdaGr2:identifierForTheCorporateBody*) of a corporate body serve in particular to identify the corporate bodies through external systems. Even if a corporate body is not known in these systems by its URI, the identifiers permit a

comparison between the systems and allow the creation of a corresponding link. Besides valid identifiers, redirected identifiers and identifiers from old systems (*gnd:invalidIdentifierForTheCorporateBody*) are also provided to offer a greater potential for comparing corporate body data.

Literals are used to specify corporate body names. Each corporate body has just one preferred name for the corporate body (*gnd:preferredNameForTheCorporateBody*) and an arbitrary number of variant names (*gnd:variantNameForTheCorporateBody*). The authorized form of name shown in Table 1 is used to represent the corporate body names. With the publication of the RDA rules, the name representation was specified more precisely and the rules for RDA-compliant names became known. This will allow literal-based name representations for corporate bodies to be replaced by the *rdaGr2:preferredNameForTheCorporateBody* and *rdaGr2:variantNameForTheCorporateBody* elements in a future revision of the linked data service.

In addition, GND elements have been introduced for the acronym of the preferred name of the corporate body (*gnd:preferredNameAcronymForTheCorporateBody*) and for acronyms of the variant name of the corporate body (*gnd:variantNameAcronymForTheCorporateBody*).

The official name of a corporate body, by contrast, is represented by an RDA element (*rdaGr2:nameOfTheCorporateBody*), but without indicating whether this is a preferred or variant form of the name. With the publication of the RDA rules, this representation can also be specified more precisely in the future.

The date of establishment (*rdaGr2:dateOfEstablishment*) and the date of termination (*rdaGr2:dateOfTermination*) of a corporate body can be specified. The dates found in these elements are not given in a controlled form. In some cases a specific date is given, whereas in other cases periods of time are given.

Functions (roles) which a corporate body carried out during the production process of a publication are shown by means of the function (*gnd:functionOfTheCorporateBody*) labels. At present, no bibliographic data is linked to the descriptions of the corporate body. It is therefore sensible to include this information as an additional search criterion in the description of the corporate body. A description of the associated roles is provided in the form of references to corresponding SWD entries. The introduction of functional labels is a relatively new tool for the German National Library's corporate body data records and the number of instances of application is currently still correspondingly low.

Predecessors (*rdaRelGr2:predecessor*), successors (*rdaRelGr2:successor*) and hierarchical superiors (*rdaRelGr2:hierarchicalSuperior*) can be referenced as relationships of one corporate body to another (link to separate data records). "See-also" relationships can also be identified (*rdfs:seeAlso*). The given relationships between corporate bodies are always reciprocal. However, non-reciprocal relationships between corporate bodies are also defined in the GKD. New GND elements (*gnd:predecessorWithoutSuccessor*, *gnd:successorWithoutPredecessor*) have been specified for the non-reciprocal relationships. In a subsequent project the intention is to replace the non-reciprocal relationships by RDA elements, using reasoning processes (*rdaRelGr2:mergee*, *rdaRelGr2:productOfAMerger*, *rdaRelGr2:productOfASplit*).

A resource which is linked to a corporate body is represented as the homepage of the corporate body (*foaf:homepage*) if the corresponding sub-field specifies it as a homepage.

Countries in which the corporate body was or is active can be identified by recording the country code (*gnd:countryCodeForTheCorporateBody*). The identification of the language of a corporate body (*rdaGr2:languageOfTheCorporateBody*) shows which language it has published in.

For redirected data records, i.e. data records of corporate body descriptions which have been replaced by a different data record, an equivalence relationship (*owl:sameAs*) is indicated for the corporate body description currently used. The redirected data records, however, are then scheduled for deletion shortly afterwards. Consequently, this relationship between two entities will only exist temporarily. A task for a potential synchroniser (not yet established) will be to remove the old data records and the existing links from the semantic web compliant representation when deleting data in the German National Library's internal system.

Implementation of the PND

The elements implemented in the service for representing a person's description are listed in the table below. The ontology source of the properties used is indicated by the prefix. Registered elements are also linked to their description in the Web. The table shows extracts of the PND field directory.

Table 2: Elements used to describe a person

PICA3	PICA+	Ind.	Field content	RDF element	Remarks
021	007Q	$0	**SWD-Nummer** (m) (**SWD number** (m))	rdaGr2:identifierForThePerson	(DE-588c)...
022	007T	$0	**LoC-Nummer (LoC number)**	rdaGr2:identifierForThePerson	(DLC)...
026	007G	$0	**Identifikationsnummern umgelenkter Datensätze (Identifiers of redirected data records)**	gnd:invalidIdentifierForThePerson	(DE-588a)...
027	007H		**Identnummer aus altem System (Identifier from old system)**		
	007H	$S	"|m|" Deutsches Musikarchiv		
	007H	$0	IDN	gnd:invalidIdentifierForThePerson	(DE-101c)...
100	028A		**Einheitliche Ansetzungsform oder Ansetzungsform nach RAK (Authorized form of name or authorized form of name as per RAK)**	gnd:preferredNameForThePerson	$a, $d $c <$I> or if $a empty: $5 <$I>

				gnd:usedRules	Representation of a name entity
		$5	Persönlicher Name (Personal name)	gnd:personalName	
		$a	Familienname (Surname)	gnd:surname	
		$d	Vorname(n) (Forename(s))	gnd:forename	
		$c	Präfix (Prefix)	gnd:prefix	
		$l	Ordnungshilfe (Qualifier)	gnd:qualifier	
120	028B		**Ansetzungsform nach RSWK (Authorized form of name as per RSWK)**	gnd:preferredNameForThePerson	$a, $d $c <$l> or if $a empty: $5 <$l>
			Feldinterne Steuerzeichen wie 100 (Internal field control characters - as 100)		Representation of a name entity
140	028C		**Ansetzungsform der LoC (Authorized form of name as per LoC)**	gnd:variantNameForThePerson	$a, $d $c <$l> or if $a empty: $5 <$l>
			Feldinterne Steuerzeichen wie 100 (Internal field control characters - as 100)		Representation of a name entity
		$p	"¬%" Bemerkungen / Qualifier für USMARC-Unterfelder ("¬%" remarks / Qualifier for USMARC sub-fields)	gnd:locQualifier	
145	028D		**Alternative Ansetzungsform / Abweichende Namensform der Person (Variant form of name for the person)**	gnd:variantNameForThePerson	$a, $d $c <$l> or if $a empty: $5 <$l>
			Feldinterne Steuerzeichen wie 100 (Internal field control characters - as 100)		Representation of a name entity
		$S	"|b|" Alternative Ansetzungsform nach RAK-ÖB (Variant form of name as per RAK-ÖB) "|c|" Alternative Ansetzungsform nach RAK-WB (Variant form of name as per RAK-WB) "|d|" Alternative Ansetzungsform einer sonstigen RAK Anwendung (Variant form of name of any	gnd:usedRules	

			other RAK application) "|g|" Alternative Ansetzungsform nach PI (Variant form of name as per PI) "|h|" Alternative Ansetzungform HEBIS (Variant form of name as per HEBIS) "|p|" Ansetzungsform nach der Polnischen Nationalbibliografie (Variant form of name as per Polish national bibliography)		
200	028.		**Verweisungsformen für alle Namensverweisungen (Variant forms of names for all name references)**	gnd:variantNameForThePerson	$a, $d $c <$I> or if $a empty: $5 <$I>
			Feldinterne Steuerzeichen wie 100 (Internal field control characters - as 100)		Representation of a name entity
300	032A		**Lebens- und Wirkungsdaten der Person (Dates of birth and death and periods of activity)**		
		$S	"|a|" Lebensdaten in normierter Form (Dates of birth and death in controlled form)	rdaGr2:dateOfBirth and rdaGr2:dateOfDeath	
			"|b|" Wirkungsdaten in normierter Form (Period of activity in controlled form)	rdaGr2:periodOfActivityOfThePerson	
			"|c|" ungefähre Zeitangaben (Approximate dates)	rdaGr2:periodOfActivityOfThePerson	
			"|e|" exakte Wirkungsdaten in normierter Form (Angabe TT.MM.JJJJ) (Exact period of activity in controlled form (DD.MM.YYYY))	rdaGr2:periodOfActivityOfThePerson	
		$a	Angabe von Jahreszahlen gemäß Indikator (Years as per indicator)		Transfer of indicator content to literal of respective element
310	032B.		**Identifizierende Angaben zur Person (Identifying attributes of person)**		

		$S	"|c|" Geburtsort (Place of birth)	rdaGr2:placeOfBirth	
			"|d|" Sterbeort (Place of death)	rdaGr2:placeOfDeath	
			"|e|" Wirkungsort(e) (Place of work)	rdaGr2:placeOfResidence	
			"|f|" Exilland/Exilländer (Country/countries of exile)	rdaGr2:placeOfResidence	
			"|q|" Adelstitel (Title of nobility)	rdaGr2:titleOfThePerson	
			"|t|" Akademischer Titel (Academic title)	gnd:academicTitleOfThePerson	
			"|u|" nichtnormierte sonstige identifizierende Angaben (Other biographical information, uncontrolled)	rdaGr2:biographicalInformation	
			"|w|" Geschlechtsangabe 'm' bzw. 'f' (Gender 'm' or 'f')	rdaGr2:gender	
			"|z|" Studienfächer (Fields of study)	gnd:studyPathsOfThePerson	
		Plus			
		$9	IDN SWD-Satz (IDN SWD record)		Transfer of expansion ($g) from indicator into literal of respective element
		or			
		$a	Text gemäß Indikator (Text as in indicator)		Transfer of indicator content to literal of respective element
315	032F		**Angabe Beruf und/oder Funktion (Profession, occupation and/or function)**	rdaGr2:professionOrOccupation	
		$S	"|a|" Weite, normierte Berufsbezeichnung aus der Liste der normierten Berufsbezeichnungen (Broadly defined, controlled job titles from the list of controlled job		

			titles)		
			"|b|" Engere, normierte Berufsbezeichnung aus der SWD (Narrowly defined, controlled job titles from the SWD)		
		$9	"!...!" IDN SWD-Satz (IDN SWD record)		Used to construct referenced URI
		or			
		$S	"|c|" Nicht normierte Berufsbezeichnung (Uncontrolled job title)		
		$a	- ohne - Normierte Berufs- und/oder Funktionsangabe (- without - uncontrolled job title and/or role)		Transfer to literal
320	032G	$a	**Codes für weite Funktionsbezeichnung (Codes for general role designation)**	gnd:functionOfThePerson	
321	032H	$a	**Abkürzungen für Angaben zu Interpreten (Instrumente der E- und U-Musik) (nach RAK-Musik Anlage M 4) (Abbreviations for indications for interpreters (instruments used in classical and popular music) (as per RAK music appendix M 4))**	gnd:functionOfThePerson	
322	032I	$a	**Abkürzungen für Angaben zu Sonstigen Funktionen (nach RAK-Musik Anlage M 4) (Abbreviations for other roles (as per RAK music appendix M 4))**	gnd:functionOfThePerson	
325	032K	$a	**Allgemeine Funktionsbezeichnung (General role designation)**	gnd:functionOfThePerson	
410	038M		**Beziehungen (Relationships)**		
		$S	"|a|" familiäre Beziehungen (Family relationships)	rdaRelGr2:relatedPersonPerson	
			"|c|" sonstige Beziehungen (Other relationships)	rdaRelGr2:relatedPersonPerson	

		$9	IDN PND-Satz (IDN PND record)		Used to construct referenced URI
		$p	Bemerkungen (Remarks)	rel:spouseOf rel:parentOf rel:siblingOf rel:childOf rel:grandparentOf rel:grandchildOf	Interpretation of literals to identify type of relationship
450	038C	$9	**Körperschaft, zu der eine Person in Beziehung steht (Related corporate body)**	rdaRelGr2:relatedCorporateBodyPerson	Link to URI of GKD set by expanding the NID from $0
485	009Q	$u	**URL (Uniform Ressource Locator) der vCard oder Homepage der Person (URL (Uniform Resource Locator) of the vCard or the person's homepage)**	foaf:homepage	
797	003@	$0	**Identifikationsnummer (IDN) (Identifier (IDN))**	rdaGr2:identifierForThePerson	(DE-588)...
811	042B	$a	**Ländercode nach DIN EN 23166 (ISO 3166) (Country code as per DIN EN 23166 (ISO 3166))**	gnd:countryCodeForThePerson	
812	042C		**Sprachencode nach ISO/TC46/SC4-N350 (Language code as per ISO/TC46/SC4-N350)**	rdaGr2:languageOfThePerson	
892	039I	$9	**PND-Nummer und Ansetzungsform des Zieldatensatzes bei Umlenkung von Datensätzen (PND number and authorized form of name of target data record for redirected data records)**	owl:sameAs	Link to URI of GKD set by expanding the NID from $0
899	046G		**Titelangaben (titles of publications of the person)**	gnd:publicationOfThePerson	

To date, the transformation of the PND has only been used for differentiated persons (Tp* data records). The reason for this is that these data records represent actual entities in the form of a person. From a modelling viewpoint there is a great deal of

interest in treating the non-differentiated persons (Tn* records) as persons in order to avoid conflicts arising from incorrect representation.

The integrated identifiers (*rdaGr2:identifierForThePerson*) for a person serve in particular to identify the persons through external systems. If a person is not known in these systems through his/her URI, the identifiers still permit a comparison to be made between the systems and the creation of a corresponding link. Besides valid identifiers, redirected identifiers and identifiers from old systems (*gnd:invalidIdentifierForThePerson*) are given to provide greater potential for comparing person-related data.

The same principle is used to represent the names of a person occurring in the PND (authorized form of the name, or RAK-based form, RSWK form, LoC form, variant names, and reference forms for all name references).

It is desirable to represent a person's name as a literal; the element *gnd:preferredNameForThePerson* or *gnd:variantNameForThePerson* is therefore used at present for this purpose. The rules for RDA-compliant authorized forms for names became known through the publication of the RDA . This will allow literal-based name representations for persons using *rdaGr2:preferredNameForThePerson* and *rdaGr2:variantNameForThePerson* elements in a future release. The representation rules shown in the table are currently used.

It was also deemed highly desirable to be able to represent the name parts in separate elements. The elements *gnd:preferredNameForThePerson* and *gnd:variantNameForThePerson* are therefore also to be used as references for a name entity (Range: *gnd:NameOfThePerson*) (cf. FRAD model concepts). This permits simple searches for persons using parts of the name, e.g. the surname. In the German National Library's internal system this granular representation takes the form of sub-fields in the PICA data format (cf. Table 2, PICA3 field 100). A name entity obtains an exact representation of the current status through the use of individual properties. A further element - *gnd:usedRules* - was introduced to indicate which set of name rules has been used. Figure 2 highlights the resulting relationships.

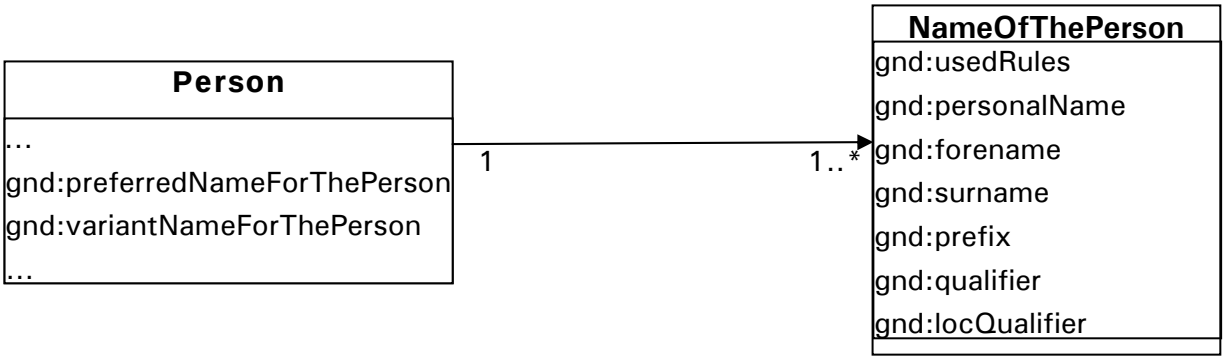

Figure 2: Use of name entities

The figure shows that each person must have at least one name. The cataloguing rules used can be determined from the respective PICA fields. The form of the LoC carries an additional qualifier for which a further element was introduced.

The *gnd:NameOfThePerson* entity is not represented by a URI only by blank nodes.

Dates pertaining to the person's life (*rdaGr2:dateOfBirth*, *rdaGr2:dateOfDeath*) and period of activity (*rdaGr2:periodOfActivityOfThePerson*) data of a person are defined as literals to the corresponding RDA elements.

The places associated with a person (*rdaGr2:placeOfBirth*, *rdaGr2:placeOfDeath*, *rdaGr2:placeOfResidence*) are also given as literals in the representation. In some cases there are also references to the corresponding SWD records. Geographical information is not included in the current SWD transformation, meaning that the expanded literals have to be shown in these elements for the time being.

Besides the RDA element for representing a person's title (*rdaGr2:titleOfThePerson*), an additional GND element (*gnd:academicTitleOfThePerson*) is used for the exclusive representation academic titles.

The gender of the person is defined by the corresponding RDA element (*rdaGr2:gender*) and the relevant controlled vocabulary[21]. The controlled vocabulary allows a distinction to be made between male or female, or to signify that the gender is not known.

Further biographical information (*rdaGr2:biographicalInformation*) and academic degrees taken by the person (*gnd:studyPathsOfThePerson*) are recorded as literals.

Details of a person's occupation (*rdaGr2:professionOrOccupation*), if such a reference exists, are indicated by a reference to the relevant SWD entry. If no SWD entry is linked as an occupation, the given literal is used.

Function labels for persons (*gnd:functionOfThePerson*) indicate functions/roles which persons carry out in relation to bibliographical records. At present, no bibliographic data is linked to the descriptions of the person. It is therefore sensible to include this information as an additional search criterion in the description of the person. The function/role descriptions are represented by codes in the PND. The transformation algorithm assesses these codes and, if possible, resolves them into a reference to the corresponding SWD entry. If no link to a SWD entry is possible, the code is incorporated as a literal in the element.

Relationships between the person described and another person are represented by links using the element *rdaRelGr2:relatedPersonPerson*. Exceptions are family relationships which permit more precisely interpretation. The labelling of the type of relationship has not yet been controlled in the person data records. The type of relationship has to be deduced from the comments (a literal). Parent-child relationships, (marital) partners, siblings and grandparent-grandchild relationships are currently indicated using the Relationship Vocabulary.

A corporate body to which a person is related is rendered by the RDA element *rdaRelGr2:relatedCorporateBodyPerson*. Similar to relationships between persons, it would be possible to evaluate the corresponding remarks field in order to interpret uncontrolled types of relationship. This service is not being offered initially.

The FOAF vocabulary is used to represent a person's homepage (*foaf:homepage*).

The country code (gnd:countryCodeForThePerson) can be used to identify countries in which the person was or is active. The language of a person

21 http://metadataregistry.org/vocabulary/show/id/101.html

(*rdaGr2:languageOfTheCorporateBody*) indicates which languages the person has published works in.

An equivalence relationship (*owl:sameAs*) to the person's description to be used in the future is indicated for redirected data records (i.e. similar to redirected GKD data records).

The publication titles identify publications published by the person (*gnd:publicationOfThePerson*). Most of the information comes from the conversion of old data. Some of these listed titles are not included in the Deutsche Nationalbibliografie. Consequently, the element remains relevant even after being associated with semantic web compliant bibliographic data of the German National Library.

Implementation of the SWD

The elements for representing SWD subject headings implemented in the service are listed in the table below. The ontology source of the properties used is indicated by the prefix. Registered elements are also linked to their description in the Web. The table is an extract of the SWD field directory.

Table 3: Elements used to describe a subject heading

PICA3	PICA+	Ind.	Field content	RDF element	Remarks
021	007Q	$0	**SWD-Nummer** (m) (**SWD number** (m))	dcterms:identifier	(DE-588c)…
026	007G	$0	**Identifikationsnummern umgelenkter Datensätze** (nur bei Hinweissätzen) (**Identifiers of redirected data records** (only for info sets))	gnd:invalidIdentifierForTheSubject	(DE-588c)…
601	041G	$s	**Nicht Deskriptor (Non-descriptor)**	rdfs:label	
606	041O	$0	**Zu verknüpfende Deskriptoren (Descriptors to be linked)**	gnd:useInsteadSWD	Link to URI of SWD record by expanding the NID from $0
800	041A.	$s	**Hauptschlagwort (Main subject heading)**	skos:prefLabel	With language attribute "de"
808			**Erläuterungen zum Schlagwort (Additional information to subject heading)**		
	046A	$S	"|b|" Indikator für Definition (Indicator for definition)	skos:definition	With language attribute "de"
	046A	$S	"|c|" Indikator für Benutzungshinweise (Indicator for instructions for use)	skos:scopeNote	With language attribute "de"

	046A	$a	Text gemäß Indikator (Text as per indicator)		Transfer of indicator content to literal of respective element
811	042B.	$a	**Ländercode nach DIN EN 23166 (ISO 3166) (Country code as per DIN EN 23166 (ISO 3166))**	gnd:countryCodeForTheSubject	
812	042C	$a	**Sprachencode nach ISO/TC46/SC4-N350 (Language code as per ISO/TC46/SC4-N350)**	dcterms:language	
830	041F	$s	**Äquivalente Bezeichnung (Equivalent name)**	skos:altLabel	With language attribute "de"
845	041S		**Übergeordneter Begriff zu Individualbezeichnung (Broader term for individual name)**	skos:broader	With language attribute "de" Multi-part heading
850	039C	$s	**Übergeordnetes Schlagwort (Broader subject heading)**	skos:broader	With language attribute "de"
860	039D	$s	**Verwandtes Schlagwort (Related subject heading)**	skos:related	With language attribute "de"

The SKOS vocabulary is used to represent main subject headings (skos:prefLabel), alternative labels (*skos:altLabel*), related subject headings (*skos:related*) and broader subject headings (*skos:broader*). The sub-terms (*skos:narrower*) of a subject heading are also identified, in contrast to the original SWD representation in the data of the German National Library.

The subject heading annotations are represented in the semantic web compliant representation by the SKOS elements *skos:definition* (for the definition of a subject heading) and *skos:skopeNote* (for instructions for use).
The language code, which does not signify the language of the subject heading but a language-related peculiarity regarding the subject heading, is represented by *dcterms:language*.

Broader terms to individual names are represented in the SWD as multipart broader terms which are currently represented with *skos:broader* as a literal. Besides the topic-based subject headings, the literal can also consist of corporate body, person and publication title subject headings.

Similar to the PND implementation, valid identifiers (*dcterms:identifier*) but also redirected identifiers and identifiers from old systems (*gnd:invalidIdentifierForTheSubject*) are given. The country code, which is used for national subject headings, is represented by *gnd:countryCodeForTheSubject*.

A special feature of the SWD is the use of so-called info records where non-descriptors are defined. Non-descriptors are terms which may not be used when indexing.

Accordingly they do not conform to the basic idea behind a SKOS concept. For this reason, each non-descriptor is assigned a new RDF type (*gnd:NonDescriptor*). Like all GND elements, the *gnd:NonDescriptor* type has not yet been registered. However, it is planned to register *gnd:NonDescriptor* as a SKOS extension in its own right in the form of a sub-class of *skos:concept*. Non-descriptors are not preferred labels, and accordingly they are specified as *rdfs:label* and not as *skos:prefLabel*. An instruction record exists for each non-descriptor which references two or more subject headings, the combination of which corresponds to the meaning of the non-descriptor. The subject headings to be linked are referenced by the GND element *gnd:useInsteadSWD*. A blank node (*gnd:useConceptsInsteadSWD*) highlights the common bond between the descriptors. The cross-reference from non-descriptor to blank node is realised by the element gnd:useConceptsInsteadSWD. In addition, explicit details are given in an instruction for use (*gnd:useInsteadNoteSWD*) in the non-descriptor on how non-descriptors are to be handled.

Links to external data sources

Besides representing an organisation's own data, an important aspect of the *Linking Open Data* concept is to link it to external data sources. Currently, there are links from the person descriptions to DBpedia[22], Wikipedia[23], and VIAF[24], and from the subject headings to LCSH[25] and RAMEAU[26]: Table 4 and Table 5 contain the elements used for the links.

Table 4: PND links to external data sources

External data source	**Property**
DBpedia	owl:sameAs
Wikipedia	foaf:page
VIAF	owl:sameAs

Table 5: SWD links to external data sources

External data source	**Property**
Library of Congress Subject Headings (LCSH)	skos:closeMatch
RAMEAU	skos:closeMatch

Implementation of parts of the German Dewey Decimal Classification

This section is currently being translated.

[22] http://dbpedia.org/About

[23] http://wikipedia.org/

[24] http://www.viaf.org/

[25] http://authorities.loc.gov/

[26] http://rameau.bnf.fr/

Discussion and outlook

The purpose of this section is to draw attention to potential areas of optimisation in semantic web compliant representation. Possible solutions will also be discussed. A further intention is to stimulate discussion amongst external users of the service. The German National Library also hopes to obtain feedback regarding additional requirements, and an assessment of its use by libraries and those using it in the commercial domain.

The following contains a description of where the focus of further development work should lie, based on selected open points and referring to the single authority files.

Further development of the GKD transformation

It is recommended, for reasons of international interoperability, that corporate body names be represented uniformly in authorized form corresponding to the RDA rules for *rdaGr2:preferredNameForTheCorporateBody* and *rdaGr2:variantNameForTheCorporateBody* elements.

Non-reciprocal relationships between corporate bodies are currently being expressed by the GND elements (*gnd:predecessorWithoutSuccessor*, *gnd:successorWithoutPredecessor*). It is desirable to assess the practical significance of these relationships using the principles of logical reasoning. This would allow the GND elements to be replaced by more precise identifications of the relationship on the basis of RDA elements (*rdaRelGr2:mergee*, *rdaRelGr2:productOfAMerger*, *rdaRelGr2:productOfASplit*).

The incorporation of existing and future information on metadata (meta-metadata) has not been implemented to date. There are, however, two main approaches which form the main focus of its development. Firstly, meta-metadata can be assigned to entities. With regard to the authority data this means that a meta-metadata set is linked to each person, family or corporate body. The second approach involves assigning information on metadata to the individual statements in entities. Consequently, a further project should select suitable ontologies and include them in the model.

Another desirable goal would be to link descriptions of corporate bodies to descriptions in external data sources. The identification of equivalence relationships permits reciprocal data enrichment.

Further development of the PND transformation

Just as with literal-based name representations of corporate body names, it is recommended that the labelling literals for names of persons be represented uniformly according to the RDA cataloguing rules for *rdaGr2:preferredNameForThePerson* and *rdaGr2:variantNameForThePerson* elements.

Further development of the SWD transformation will enable links from the person descriptions to additional controlled and normalized values. A conceivable solution could, for example, be the referencing of associated countries based on the integrated country code using the *rdaGr2:countryAssociatedWithThePerson* element.

Another challenge which arose during the ontology modelling was the representation of aggregated information. An example here is the statement that a person lived at a particular place for a particular period of time. If several places are given with respective

time periods, currently, no clear assignment is possible using the existing RDA vocabulary. Figure 3 illustrates this problem.

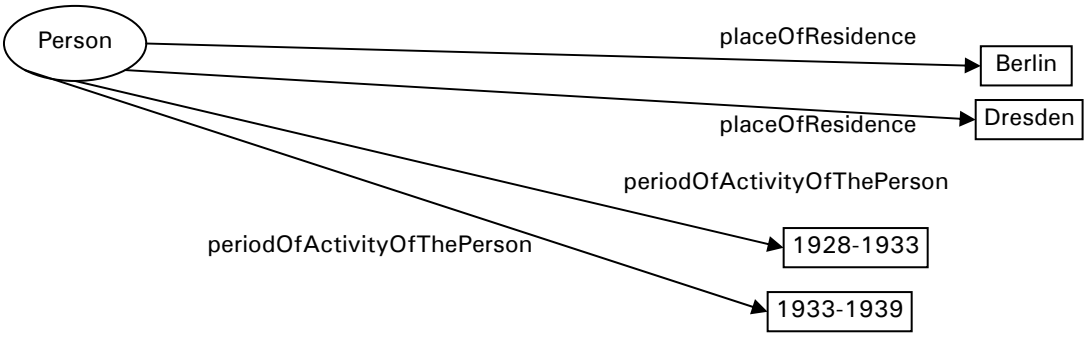

Figure 3: Illustration of problem in aggregation of data

This conflict could be resolved through the use of empty nodes (see Figure 4), however this contradicts the methods of use defined in the ontologies.

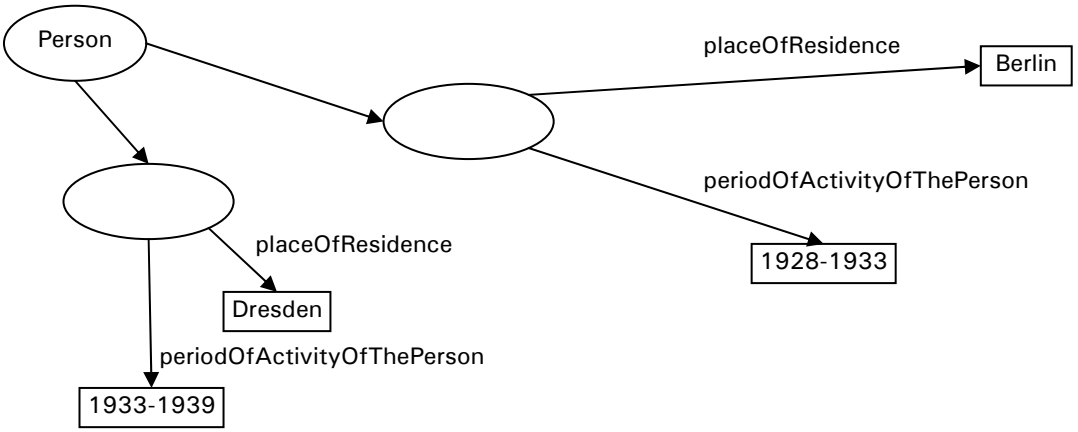

Figure 4: Potential solution for the aggregation of data

There is also the need for improvement in representing relationships between persons and between persons and corporate bodies. Statistical surveys are currently being carried out in the ways in which relationship descriptions have been issued to date. The aim for the future is to represent all relationships in a normalized manner on the basis of ontology elements. At present the relationships are only represented in exemplary form in the service.

There is a similar problem with corporate bodies in the linking of meta-metadata to person descriptions.

Besides the external data sources already linked, the aim is to include additional external data references in the description of persons.

There are currently roughly 10,000 family descriptions in the PND data. Families are currently ignored and are therefore not transformed in the semantic web compliant representation. As a further development it is therefore important to undertake specific modelling for family representations, to release these from the PND and to represent them in a manner which conforms to the model.

Further development of the SWD transformation

Multipart broader terms (field 041S) are currently represented by literals. It would, however, be preferable to represent the individual components as links to the respective subject heading URIs. Only select parts of the SWD subject headings are implemented in the present service, however. The multipart broader terms also contain other subject headings (geographical information, corporate bodies, etc.), so that it is not currently possible to reference these through links.

Special GND elements have been defined to represent the non-descriptors of the SWD, because there was no suitable representational possibility in existing vocabularies. This solution should, however, be regarded as provisional.

Furthermore, the SWD subject groups (042A) are to be skosified, thereby enabling SWD subject headings to be linked to them.

In terms of content, it would be desirable to include the subject headings for persons, corporate bodies, publications and geographical information in the SWD implementation.

Deutsche Nationalbibliothek (Leipzig, Frankfurt am Main, Berlin)
2010