



## Towards an Editable, Versionized LOD Service for Library Data

**Felix Ostrowski**

literarymachine.net, Germany  
[felix.ostrowski@gmail.com](mailto:felix.ostrowski@gmail.com)

**Adrian Pohl**

North Rhine-Westphalian Library Service Center, hbz, Germany  
[pohl@hbz-nrw.de](mailto:pohl@hbz-nrw.de)

### Abstract

The Northrhine-Westphalian Library Service Center (hbz) launched its LOD service *lobid.org* in August 2010 and has since then continuously been improving the underlying conversion processes, data models and software. The present paper first explains the background and motivation for developing *lobid.org*. It then describes the underlying software framework *Phresnel* which is written in PHP and which provides presentation and editing capabilities of RDF data based on the Fresnel Display Vocabulary for RDF. The paper gives an overview of the current state of the Phresnel development and discusses the technical challenges encountered. Finally, possible prospects for further developing Phresnel are outlined.

**Key Words:** Linked Open Data; Web application development; Cataloguing

## 1. Introduction

In the broader library world, Linked Open Data (LOD)<sup>1</sup> has gained a lot of attention over the last two years, with projects moving increasingly from theory to practice. The library domain is gearing more and more towards the technical and legal issues implied by this paradigm shift, with the announcement of 'A Bibliographic Framework for the Digital Age' by the Library of Congress (Marcum & Library of Congress, 2011) and the Conference of European National Librarians' affirmation of open licensing for their data (Conference of European National Librarians, 2011) being only two examples.

Since 2009 the North Rhine-Westphalian Library Service Center (hbz) has been exploring Linked Open Data and Semantic Web technologies, where both the legal and the technological aspects of this ongoing paradigm change in information representation and provision. The hbz launched its LOD service *lobid.org* — standing for "Linking Open Bibliographic Data" — in August 2010 and since then has continuously been improving it.

The overall goal is to develop the underlying software framework so that its read/write services can be run, including a web presentation of the underlying RDF data and online forms to create, update and delete the underlying information represented in RDF (Resource Description Framework). To keep track of and to be able to revoke changes, the system should also fully versionize the underlying data structures. This paper explains its motivation (section 2) and describes the LOD service *lobid.org* (section 3), as these developments were initiated to improve the service. In section 4 the Fresnel Display Vocabulary for RDF is introduced which serves as a generic way to configure the presentation of RDF data. Section 5 explains Phresnel, a free software framework for presenting and editing RDF data based on Fresnel and PHP. Finally, in section 6 prospects for further developing Phresnel are listed.

## 2. Motivation and expected benefits

The hbz has been in the business of cooperative cataloguing for some time now, running a union catalogue since 1973. From this perspective, Linked Open Data provides a very interesting approach for distributed cooperative cataloguing in a web environment.

## **2.1. Web integration**

Adherence to international and cross-domain web standards for Linked Data means web integration of library data, from which the following benefits are expected in the long term:

- *Increased discoverability.* Web-integrated data can easily be harvested by search engines and other discovery services.
- *Multiple usability.* RDF data stored in one data sink can easily be used as such by different services within the hbz and beyond.
- *Interoperability and re-usability.* Web standards facilitate reuse by reducing the need for conversion processes and post-processing.
- *Flexibility.* RDF and Triple Stores are very flexible regarding extensions and changes in the data model used.

## **2.2. Synergy effects**

Following the best practices of the Linked Open Data community and working towards a standardization of the data produced by the different services within the hbz, return of investments in form of intra- and inter-organizational synergy effects are expected. Within the organization, we already see some effects on hbz projects reusing each others' data. With LOD, this is possible in a straightforward way, whereas having to deal with proprietary interfaces and different formats often makes it quite labor-intensive for services to communicate effectively. Thus, standardization of services has the effect of liberating resources which then can be used for additional services or for improving existing services.

## **2.3. Less vendor dependencies**

The provision of many library services depends on technology, and there only exist a few different vendors in the library business. Often organizations are depending on products by one or two vendors, with significant costs for switching to a different vendor or product. In other words, the lock-in effect is very strong in the library domain.

The utilization of Linked Open Data best practices improves this situation in two ways:

1. Being able to easily get hold of your data in an open, cross-domain standard enables you to switch to another product with less effort.
2. With data being held in RDF rather than open but opaque formats like MARC, there will likely emerge many additional softwares of interest from other vendors. Growing competition leads to improved products and/or decreasing prices.

### 3. lobid.org

In August 2010 the hbz launched its experimental Linked (Open) Data service lobid.org<sup>2</sup>, which is comprised of two services: a “catalogue” of bibliographic resources and holding information (lobid-resources<sup>3</sup>) and an index of libraries and related organizations (lobid-organizations<sup>4</sup>). lobid.org fully employs Linked Data principles as well as — whenever possible — Open Data principles.

Since 2010 the two lobid.org services and their underlying data have been continuously improved:

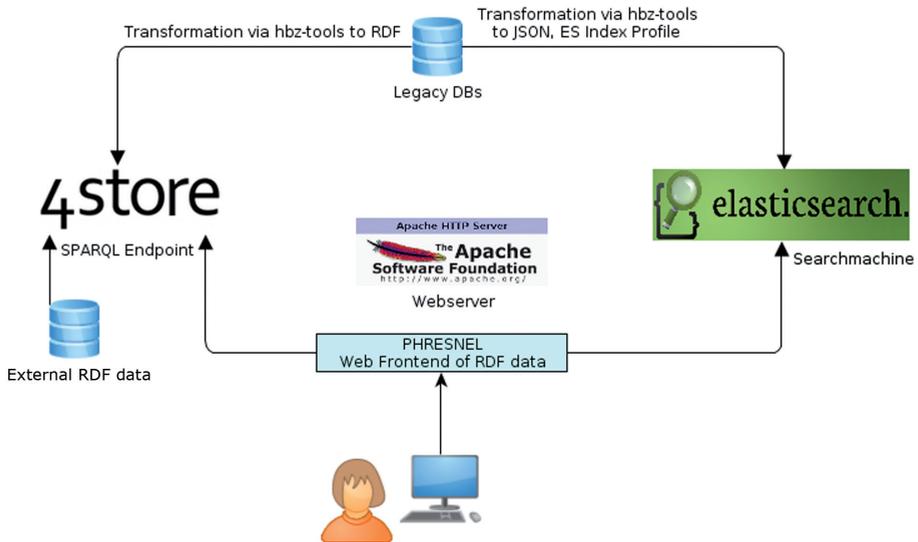
- Information is being extended by adding more fields from legacy data to the mapping and by revising vocabulary and property choices.
- Context is being added by linking resources to other Linked Data sets.

Interaction options for end users are being improved, e.g. by a search engine interface and by aligning the user interface of both sub-services (see 5).

Figure 1 gives an overview over the currently employed technology stack, data sources and conversion processes. As a triple store lobid.org employs Garlik’s 4store<sup>5</sup> for full-text indexing and elasticsearch<sup>6</sup> for searching. The web front end runs on an Apache server and is generated by the Phresnel framework described in more detail below.

As one can see, lobid.org is currently almost entirely based on legacy data dumps from existing systems that are converted to RDF using custom tools. The resulting RDF data are enriched with links to other datasets in the LOD

Fig. 1: lobid.org technology stack.



All marks mentioned may be trademarks or registered trademarks of their respective owners.

cloud. Some external LOD datasets are also indexed into the triple store: currently these are the ontologies used within lobid.org as well as the German national authority file *Gemeinsame Normdatei* (GND)<sup>7</sup> provided by the German National library. Until now, no possibilities exist for manually adding and editing the RDF data, e.g. to add new information (commonly called ‘cataloguing’) or to correct mistakes.

### 3.1. lobid organisations

When the hbz started to publish Linked Open Data, it became clear that the bibliographic records from the hbz union catalogue would just be the start. If you want to build useful services on top of Linked Open Data, you also need URIs for and descriptions of items, holding institutions and services. For example, a geo-based query which gives you back all items of a specific manifestation in a 5 km radius requires URIs for and RDF descriptions of at least three entities: There is a manifestation M that is exemplified by an item

I that is held by organisation O. In the RDF serialization turtle it reads as illustrated in Figure 2.

The corresponding graph looks as in Figure 3.

Since two years ago people and organizations wouldn't move very much if you asked them to provide Linked Open Data, the realization was made that one has to do things yourself. Thus, lobid.org was launched with lobid-organizations in July 2010 (Ostrowski, 2010).

Fig. 2: RDF example (turtle) relating manifestation, item and holding organization.

```

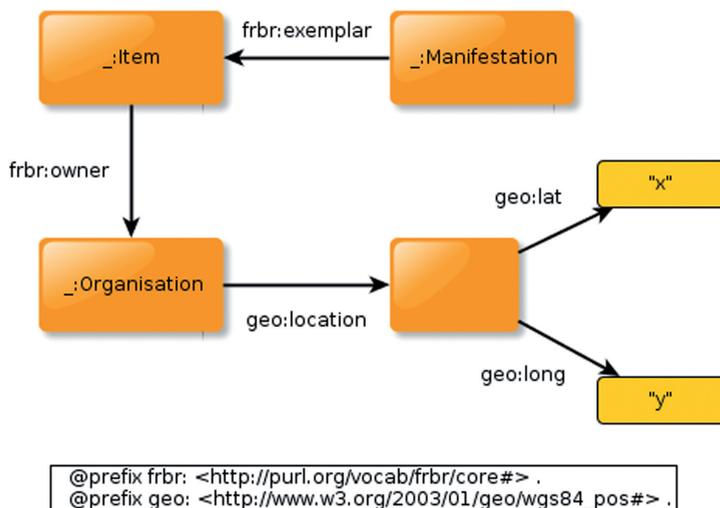
@prefix frbr: <http://purl.org/vocab/frbr/core#> .
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>

_:Manifestation frbr:exemplar _:Item .

_:Item frbr:owner _:organisation .

_:organisation geo:location [
  geo:lat "x" ;
  geo:long "y"
] .
    
```

Fig. 3: Graph showing relations between manifestation, item and holding organization.



The underlying data come from the German ISIL registry<sup>8</sup> and the MARC organization code database<sup>9</sup> maintained by the Library of Congress. By now, lobid.org has minted URIs for more than 40,000 institutions and provides basic RDF descriptions of them.<sup>10</sup> Currently there exists neither an openly available dump of the data nor is it openly licensed as we cannot decide on this, not having produced the data ourselves.

### **Enrichments**

In addition to the data obtained from the mentioned data sources, new links to other datasets in the LOD cloud are created. By now, links to DBpedia and Wikipedia (Christoph, 2012d) and to GeoNames (Christoph, 2012a) have been added. Furthermore, organization descriptions are enriched with a QR code which contains contact information (ibid.). Also, based on the geo coordinates for most of the libraries, we show their location on Open Street Map<sup>11</sup> embedded in the web page.

An example web page for the German National Library – generated from underlying RDF as described in 5.1 — is illustrated in Figure 4.

### **3.2. lobid resources**

*lobid-resources* is basically the LOD interface for Open Data from the hbz union catalogue. It offers URIs for and descriptions of bibliographic resources like monographs and multi-volume works on a FRBR-manifestation level as well as URIs for and descriptions of corresponding items held by hbz member libraries. Also, journals and serials are included that cannot be comprised under the FRBR WEMI (work-expression-manifestation-item) model (Pohl, 2011). Information on FRBR expression or especially on work-level is planned to be integrated in the future.

Since the first main open data publication in March 2010 (North Rhine-Westphalian Library Service Center, 2010) gradually more and more data from this catalogue has been published as open data in agreement with cooperating libraries. As of August 2012, the dataset comprises approximately 16 Million records published under a Creative Commons Zero

Fig. 4: lobid.org webpage describing the German National Library.<sup>12</sup>

<a href="http://lobid.org/organisation/DE-101">http://lobid.org/organisation/DE-101</a>									
Name	Deutsche Nationalbibliothek								
Geokoordinaten									
Anschrift	<table border="1"> <tr> <td>Straße</td> <td>Adickesallee 1</td> </tr> <tr> <td>PLZ</td> <td>60322</td> </tr> <tr> <td>Ort</td> <td>Frankfurt/Main</td> </tr> <tr> <td>Land</td> <td>Germany</td> </tr> </table>	Straße	Adickesallee 1	PLZ	60322	Ort	Frankfurt/Main	Land	Germany
Straße	Adickesallee 1								
PLZ	60322								
Ort	Frankfurt/Main								
Land	Germany								
ISIL	DE-101								
Homepage	<a href="http://www.dnb.de">http://www.dnb.de</a>								
E-Mail	<a href="mailto:info-f@dnb.de">mailto:info-f@dnb.de</a>								
Kontakt-QR									
Weitere Information	<a href="http://dispatch.opac.d-nb.de/DB=1,2/CMD?ACT=SRCHA&amp;IKT=8529&amp;TRM=DE-101">http://dispatch.opac.d-nb.de/DB=1,2/CMD?ACT=SRCHA&amp;IKT=8529&amp;TRM=DE-101</a>								
Sitz	<a href="http://sws.geonames.org/6553153/">http://sws.geonames.org/6553153/</a>								
Wikipedia-Link	<a href="http://de.wikipedia.org/wiki/Deutsche_Nationalbibliothek">http://de.wikipedia.org/wiki/Deutsche_Nationalbibliothek</a>								
DBpedia-Link	<a href="http://de.dbpedia.org/resource/Deutsche_Nationalbibliothek">http://de.dbpedia.org/resource/Deutsche_Nationalbibliothek</a>								

license<sup>13</sup>, which represents 85% of the hbz union catalogue (Christoph, 2012e). Using custom conversion tools, the data are generated based on an XML dump from the hbz Aleph system. The resulting RDF data can be queried via a public SPARQL endpoint<sup>14</sup>, and a full data dump is also available for download<sup>15</sup>.

### Enrichments

Because identifiers from the German-wide authority file for names, subject headings and corporate entities already existed in the legacy data, links to the Linked Data version of the *Gemeinsame Normdatei* (GND, first published in 2010) were easy to implement.

Existing language encodings were replaced by links to the ISO 639-2 Codes for the Representation of Names of Languages provided by the Library of Congress.<sup>16</sup>

*Fig. 5: Example bibliographic description in lobid.org with links to GND, Project Gutenberg (eBooks & audio books), Open Library and DBpedia.<sup>17</sup>*

<a href="http://lobid.org/resource/HT002189125">http://lobid.org/resource/HT002189125</a>		
Titel	Typee	
Titelzusatz	a peep at Polynesian life	
Autor	<a href="http://d-nb.info/gnd/118580604">http://d-nb.info/gnd/118580604</a>	
Erscheinungsjahr	1982	
Typ	< <a href="http://purl.org/ontology/bibo/Book">http://purl.org/ontology/bibo/Book</a> >	
Typ	< <a href="http://purl.org/vocab/frbr/core#Manifestation">http://purl.org/vocab/frbr/core#Manifestation</a> >	
Volltext	<a href="http://gutenberg.org/ebooks/1900">http://gutenberg.org/ebooks/1900</a>	
Volltext	<a href="http://gutenberg.org/ebooks/28656">http://gutenberg.org/ebooks/28656</a>	
Volltext	<a href="http://gutenberg.org/ebooks/9269">http://gutenberg.org/ebooks/9269</a>	
Volltext	<a href="http://gutenberg.org/ebooks/23969">http://gutenberg.org/ebooks/23969</a>	
Ausgabe	3. print.	
Erscheinungsort	New York, NY	
Verlag	Literary Classics of the United States [u.a.]	
Format	print	
Umfang	1333 S.	
ISBN-10	0521262194	
ISBN-10	0940450003	
ISBN-13	9780521262194	
ISBN-13	9780940450004	
Werktitel	<a href="http://openlibrary.org/works/OL14953734W">http://openlibrary.org/works/OL14953734W</a>	
Werktitel	<a href="http://dbpedia.org/resource/Typee">http://dbpedia.org/resource/Typee</a>	
Werktitel	<a href="http://de.dbpedia.org/resource/Typee">http://de.dbpedia.org/resource/Typee</a>	
Exemplar	<a href="http://lobid.org/item/HT002189125%3AU+MEL-11">http://lobid.org/item/HT002189125%3AU+MEL-11</a>	
	Typ	< <a href="http://purl.org/vocab/frbr/core#Item">http://purl.org/vocab/frbr/core#Item</a> >
	Zugehöriger Titel	<a href="http://lobid.org/resource/HT002189125">http://lobid.org/resource/HT002189125</a>
	Besitzer	<a href="#">Institut für Anglistik, Amerikanistik und Keltologie, Bibliothek</a>
	Signatur	U MEL-11
Weitere Information	<a href="http://193.30.112.134/F/?func=find-c&amp;ccl_term=IDN%3DHT002189125">http://193.30.112.134/F/?func=find-c&amp;ccl_term=IDN%3DHT002189125</a>	

Also, links to other datasets that include bibliographic data were added step by step. Using simple matching algorithms for ISBN and title string in combination with some post processing based on simple heuristics, links to Dbpedia (Christoph, 2012b), Open Library (Christoph, 2012c) and Project Gutenberg were added to a subset of resources. These links provide some kind of work-level bundling of resources, enabling for instance mutual enrichment of bundled resources with subject headings, links etc.

In the future the hbz aims at enhancing the data even more by adding subject headings and classification as well as by providing more links to other datasets and to full texts online. A simple API will be developed to enable easy use by libraries who want to re-use these enrichments

An example resource description is depicted in Figure 5.

## **4. Presenting RDF data using the Fresnel Display Vocabulary for RDF**

### **4.1. Rationale**

In the beginning, the converted legacy data for bibliographic resources was exposed using Pubby, “a Linked Data Frontend for SPARQL Endpoints”<sup>18</sup>. While being very easy to set up, the resulting views — and among those especially the human-readable HTML — that were generated by Pubby exposed too much of the underlying technology. The organizations data on the other hand was presented using a custom SPARQL query, PHP scripts and some HTML-templates in order to, e.g. include a map in the HTML view. This provided more flexibility but was not easily adapted to data other than that about organizations since each content model needed a manually created query and corresponding template. Besides that, there was the idea to enable libraries to easily create RDFa<sup>19</sup> descriptions of their organizations. In order to do so, the need for a simple, intuitive editor arose. Instead of exposing the underlying RDF model to content-creators, a browser based HTML form was aimed at, providing a familiar environment for anybody acquainted with the Web.

With these requirements in the back of the head, the search for a schema language from which such a front-end could be derived began. When dealing

with RDF data, RDF Schema (RDFS) or the Web Ontology Language (OWL) are the first candidates that spring to mind. Since ontologies expressed in these languages are usually designed to be application-independent, experiments in this direction were rather fruitless, because the resulting views were too generic to fulfil the requirements. Especially mixing classes and properties from several vocabularies in a concise and comprehensible way is nearly impossible. Luckily, the Fresnel Display Vocabulary for RDF came across. It is designed precisely to specify “*what* information contained in an RDF graph should be presented and *how* this information should be presented”<sup>20</sup> without interfering with the underlying ontologies. Similar to the ontology languages mentioned above, it is itself based on RDF, making it possible to stay within one data model all throughout the implementation.

#### 4.2. Lenses

Fresnel *lenses* address the first aspect mentioned above, namely *which* data should be displayed. A single lens can be related to instances in several ways, the simplest possibility being a reference to its class (i.e. its `rdf:type` values) as demonstrated in Figure 6. For the selected instances, an ordered list of properties is supplied, which is very easy and readable in turtle notation. In order to include in the output data about a related entity, another lens may be referred to. In the example in Figure 6, the author’s first and last name will be displayed in a document description and not only its URI as would be the case when listing `dc:creator` without referring to such a `:person` *sublens* as is done on the right hand side.

Fig. 6: Fresnel lenses and sublenses.

<pre> :person   a fresnel:Lens ;   fresnel:classLensDomain     foaf:Person ;   fresnel:showProperties (     foaf:familyname     foaf:givenname   ) . </pre>	<pre> :document   a fresnel:Lens ;   fresnel:classLensDomain     bibo:Document ;   fresnel:showProperties (     dc:title     bibo:isbn     [ a fresnel:propertyDescription ;       fresnel:property dc:creator ;       fresnel:sublens :person ]   ) . </pre>
---	---

Fig. 7: Data yielded by applying lenses.

```

<http://lobid.org/resource/HT002948556>
  a bibo:Document ;
  dc:title "With reference to reference"@en ;
  bibo:isbn "9780915145539" ;
  dc:creator <http://d-nb.info/gnd/135539897> .

<http://d-nb.info/gnd/135539897>
  a foaf:Person ;
  foaf:familyName "Elgin" ;
  foaf:givenName "Catherine" .

```

Applying the above lenses to an (imaginary) triple store should yield the triples in Figure 7, which should be displayed in that order:

All in all, Fresnel lenses allow for a very concise and declarative way to express which data to select and which order to display it in.

### 4.3. Formats

Fresnel *formats* deal with the second issue stated above: they express how the selected data should be displayed. Possibilities range from custom labels for properties that differ from the labels defined in an ontology, to styling hooks used to reference CSS classes. An example can be found in Figure 8.

In a way similar to Fresnel lenses as a way to select and order data, the format vocabulary allows to configure the way that data is displayed in a declarative, application independent way.

Fig. 8: Formats for custom property labels and applicable CSS classes.

```

[] a fresnel:Format ;
  fresnel:propertyFormatDomain foaf:familyName ;
  fresnel:label "Last name"@en ;
  fresnel:label "Nachname"@de .

[] a fresnel:Format ;
  fresnel:classFormatDomain foaf:Person ;
  fresnel:resourceStyle "withBorder"^^fresnel:styleClass .

```

## **5. Phresnel – Implementing Fresnel in PHP**

### **5.1. Alternatives**

Several implementations of Fresnel already existed<sup>21</sup> when the development of Phresnel began. Most of them are written in Java and none is implemented in PHP, upon which the pre-Phresnel version of lobid.org was based. Stand-alone applications supporting Fresnel, such as IsaViz<sup>22</sup>, do not fulfil our requirement of providing a classical browser-based user interface. JFresnel<sup>23</sup> as a low-level Fresnel API is an interesting library to implement Java-based Fresnel-aware applications, but it does not deliver any application logic. Reusing our existing PHP web-application code — such as request dispatching / URL routing — would not have been possible, thus we decided against the usage of this library. Longwell<sup>24</sup> is geared towards faceted browsing, which is indeed an important aspect of a system such as lobid.org. LENA<sup>25</sup> is yet another Linked-data viewer, but both of these solutions do not provide the means to alter data. Also, in both cases the demos are offline and development appears to have stalled. There are further Linked-data front-ends for SPARQL endpoints, such as Pubby<sup>26</sup> and Elda<sup>27</sup>, but these do not use Fresnel and once again do not provide editing capabilities.

Thus, a new, PHP-based editing-aware framework dubbed ‘Phresnel’<sup>28</sup> was implemented as a proof-of-concept. Currently only a small subset of lens and format features is implemented in Phresnel, limited to those absolutely necessary to get the prototype up and running.

### **5.2. Displaying**

In order to display data according to a Fresnel lens, the web application detects the lens to be used from the URL to which a GET request was issued, e.g. “document”. It then uses the Phresnel framework to generate a generic (X)HTML view (with embedded RDFa) of the requested data as shown in Figure 9. At this point it is assumed that an HTTP-303-redirect following the linked-data design pattern<sup>29</sup> has already occurred in a previous step.

Internally, Phresnel uses the lens definitions to generate a series of SPARQL CONSTRUCT queries such as those depicted in Figure 10 and then uses the lens and format definitions to order and style the resulting RDF.

Fig. 9: Phresnel generating (X)HTML views.

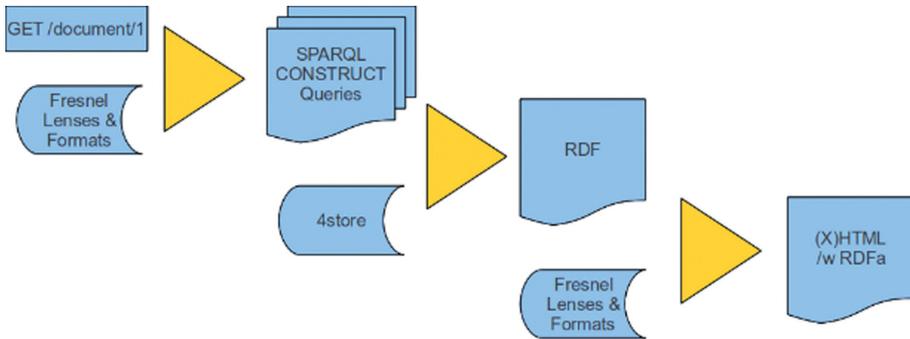


Fig. 10: SPARQL CONSTRUCT queries generated by Phresnel.

```

CONSTRUCT {
  <http://lobid.org/resource/HT002948556> ?p ?o .
} WHERE {
  <http://lobid.org/resource/HT002948556> a bibo:Document .
  <http://lobid.org/resource/HT002948556> ?p ?o .
  FILTER (?p=rdftype || ?p=dc:title || ?p=bibo:isbn || ?p=dc:creator) .
}

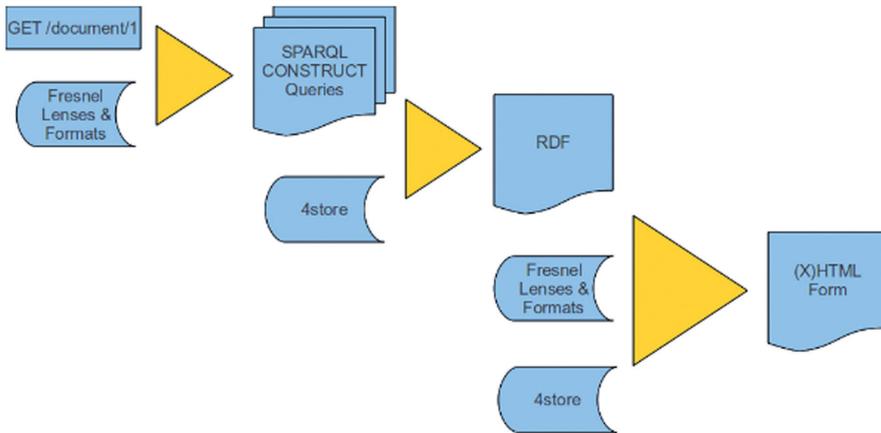
CONSTRUCT {
  ?s ?p ?o .
} WHERE {
  <http://lobid.org/resource/HT002948556> dc:creator ?s .
  ?s a foaf:Person .
  ?s ?p ?o .
  FILTER (?p=rdftype || ?p=foaf:givenName || ?p=foaf:familyName) .
}
  
```

Currently, only a hard-coded box model (Bizer, Lee, & Pietriga, 2005) based on nested tables is available. The ordering and transformation to (X)HTML can obviously be skipped when a pure RDF representation is requested via content-negotiation.

### 5.3. Editing

When assembling an editing the view of a resource, the steps are very similar to those when requesting a simple display representation, as can be seen in Figure 11. But there is one major difference. When the RDF resulting from the SPARQL queries is transformed to an (X)HTML form, all literals are simply

Fig. 11: Phresnel generating (X)HTML Forms.



converted to text input elements. Unfortunately this would result in an awkward interface for those cases where links to other entities are expected, since the URIs of those entities would have to be looked up and typed in manually. This is unacceptable from a usability point of view.

The current implementation solves this in an insufficient, but at least reasonably visually appealing way. Link targets, e.g. a list of possible authors in the above example, are displayed as drop-down lists populated from the triple store. Unfortunately this has two severe limitations: the list can easily get excessively long, and the entries in the list are limited to the own data source, making it hard to link to other sources on the web. For the future, it is planned to get around these limitations by means of auto-complete text-fields that are driven either by custom javascript-snippets for external data-sources or connected to the own search index discussed in the next section. At this point it should be noted again that while data can be edited using this web front-end, currently no work-flow for storing the resulting data is in place.

## 5.4. Search

While the Linked Data paradigm is great for navigation, search is vital to actually discover resources. Although SPARQL supports regular expressions<sup>30</sup> that can be used for search, this currently does not scale to the data volume

of lobid.org. This is why the search for organizations<sup>31</sup> as well as the one for resources<sup>32</sup> on lobid.org is currently backed by an elasticsearch<sup>33</sup> index, accessed via a custom web application that provides a CQL interface. This concrete setup has historical reasons. The resource index has existed long before lobid.org and is used by several hbx services, and it was easiest to simply integrate the organization index into the same infrastructure.

Due to this construction, search results — which are received as Atom feeds — have a data structure that does not match the Fresnel lens definitions driving lobid.org. Hence, currently only the identifiers are extracted from the search result, which are used to construct the URI of the discovered resource, using Phresnel to then receive and format data as described above. This works fairly well and is much more efficient than using native SPARQL queries for fulltext search but suffers from the fact that after the resource has been identified the data about it has to be retrieved again, this time from the triple store. This is only one of the problems that will have to be tackled next.

## 6. Prospects

The current proof-of-concept implementation for a read/write system for LOD-based library data uncovers interesting prospects for future data management. It is clear though that the efforts are still very much at the beginning. Further Phresnel developments will explore options in the following areas.

### 6.1. Purely JavaScript based editing

Since the most important part of the editor — the facility to look up resources and create links — will need to be reworked for reasons mentioned above, it is considered to switch to an editor implementation based solely on RDFa and JavaScript in a way similar to create.js<sup>34</sup>, abandoning server-generated forms all together. This would not only create a more fluid experience for the user but would also reduce load on the server and provide a better separation of concerns between the front-end and the back-end.

## **6.2. Data production and maintenance**

The results of editing data in the web front-end will have to become persistent. There are several non-trivial decisions that have to be made in this respect: in how many places should the data be stored (triple store, flat files, search index), how should the data be organized (named graphs (Dodds, 2009), ...), which provenance should be recorded, which authorization system should be used etc. Since an application like the one described in this paper naturally lives in a networked, decentralized environment, thought will also have to be put into a solution to inform other connected services about creation, updates and deletions of data.<sup>35</sup> One idea is to use a real-time, message based protocol such as IRC or XMPP.<sup>36</sup>

Another important feature regarding data management is obviously identification and authentication of the agents acting upon the data. Instead of implementing such a system from scratch, it should be based upon a standard and ideally it should also be based on Linked Data principles. Because of this, the most likely approach to be used is WebID (Sporny, Inkster, Story, Harbulot, & Bachmann-Gmür, 2011) which uses FOAF descriptions of agents in conjunction with SSL certificates. This results in a secure distributed identification and authentication mechanism that is reasonably easy to be used by humans as well as by machines.

## **6.3. Versioning**

Among the most important provenance information is a seamless history of changes made to the data, along with the identification of the agent (be it a system or person) that is responsible for these changes. While it is possible to express changes to RDF data as change-sets using an RDF<sup>37</sup> vocabulary, it is very likely that this is not the most efficient way to store them since the triple count in a store would explode (at least if it is the same store that holds the actual data). Alternatives that are on the map to be explored are using a versioning system that operates on flat files, such as git<sup>38</sup>, or using the versioning features of elasticsearch, which cannot only be considered as a search engine but also as a document store. In order to expose the different versions of the data in a standardized way, it is being considered to implement a Memento<sup>39</sup> interface for the selected versioning system.

## 6.4. JSON-LD in ES / Fresnel-based search engine indexing

The way in which search is currently tied into the application is not very generic, and it depends on external organizational and technical processes. Since look-up is a very important part of the system, a solution that ties in more naturally is preferred. Elasticsearch being schema-less should play nicely with RDF data. Since it indexes JSON data, an evaluation of several JSON-RDF-serialisations has begun. The most promising approach seems to be JSON-LD<sup>40</sup>, on the one hand because it is the one most likely to become a standard, and on the other hand because it structures data in a way that matches the key-value approach that elasticsearch is based upon. The Fresnel lenses driving the front-end could be reused to generate the JSON-structures for the index.

## References

- Berners-Lee, T. (2006, July 27). *Linked Data - Design Issues*. Retrieved January 9, 2013, from <http://www.w3.org/DesignIssues/LinkedData.html>.
- Bizer, C., Lee, R., & Pietriga, E. (2005, June 30). *Fresnel - Display Vocabulary for RDF [User Manual]*. Retrieved January 9, 2013, from <http://www.w3.org/2005/04/fresnel-info/manual/>.
- Christoph, P. (2012a, April 16). *lobid.org news [Blog post]*. Retrieved January 9, 2013, from <http://wiki1.hbz-nrw.de/x/mYQ7>.
- Christoph, P. (2012b, May 3). *First results using SILK to link to Dbpedia [Blog post]*. Retrieved January 9, 2013, from <https://wiki1.hbz-nrw.de/x/mAJH>.
- Christoph, P. (2012c, May 23). *1.2 M links to Open Library [Blog post]*. Retrieved January 9, 2013, from <http://wiki1.hbz-nrw.de/x/I4BO>.
- Christoph, P. (2012d, June 11). *lobid-organisation - new links to German Dbpedia [Blog post]*. Retrieved January 9, 2013, from <https://wiki1.hbz-nrw.de/x/6IBO>.
- Christoph, P. (2012e, August 16). *lobid-resources — much more records, authority URIs dereferencable, provenance [Blog post]*. Retrieved January 9, 2013, from <http://wiki1.hbz-nrw.de/x/cQJ5>.
- Conference of European National Librarians (2011). *Europe's national librarians support Open Data licensing [Press release]*. Retrieved January 9, 2013, from [http://is.gd/cenl\\_announcement](http://is.gd/cenl_announcement).

Dodds, L. (2009, November 5). *Managing RDF Using Named Graphs [Blog post]*. Retrieved January 9, 2013, from <http://www.ldodds.com/blog/2009/11/managing-rdf-using-named-graphs/>.

Marcum, D., & Library of Congress (2011, October 31). *A Bibliographic Framework for the Digital Age*. Retrieved January 9, 2013, from <http://www.loc.gov/marc/transition/news/framework-103111.html>.

North Rhine-Westphalian Library Service Center (2010). *March 2010: Releasing catalogue data: Cologne-based libraries to pioneer Open Data practices [Press Release]*. Retrieved January 9, 2013, from [http://www.hbz-nrw.de/dokumentencenter/presse/pm/datenfreigabe\\_engl](http://www.hbz-nrw.de/dokumentencenter/presse/pm/datenfreigabe_engl).

Open Knowledge Foundation (2009). *Open Knowledge Definition, v1.1*. Retrieved January 9, 2013, from <http://opendefinition.org/okd/>.

Ostrowski, F. (2010, July 20). *Building a Linked Data based index of library institutions [Blog post]*. Retrieved January 9, 2013, from <http://blog.lobid.org/2010/07/building-linked-data-based-index-of.html>.

Pohl, A. (2011, October 17). *Serials and FRBR [Blog post]*. Retrieved January 9, 2013, from <https://wiki1.hbz-nrw.de/display/SEM/2011/10/17/Serials+and+FRBR>.

Sporny, M., Inkster, T., Story, H., Harbulot, B., & Bachmann-Gmür, R. (2011, December 12). *WebID 1.0 - Web Identification and Discovery [W3C Editor's Draft]*. Retrieved January 9, 2013, from <http://www.w3.org/2005/Incubator/webid/spec/>

## Notes

---

<sup>1</sup> To clarify our notion of “Linked Open Data”: We call data “Linked Data” if it is published following the requirements in Berners-Lee (2006). We call data “Linked Open Data” if the data additionally meets Open Knowledge Foundation (2009).

<sup>2</sup> <http://lobid.org/>.

<sup>3</sup> <http://lobid.org/resource>.

<sup>4</sup> <http://lobid.org/organisation>.

<sup>5</sup> <http://4store.org/>.

<sup>6</sup> <http://www.elasticsearch.org/>.

<sup>7</sup> See the GND-entry on the Data Hub: <http://thedatahub.org/dataset/dnb-gemeinsame-normdatei>.

<sup>8</sup> <http://dispatch.opac.d-nb.de/DB=1.1/DB=1.2/>.

<sup>9</sup> <http://www.loc.gov/marc/organisations/>.

- <sup>10</sup> To get an overview over the RDF properties used in lobid organizations see <http://lobid.org/dataset/organisation/properties.txt>.
- <sup>11</sup> <http://www.openstreetmap.org/>.
- <sup>12</sup> Retrieved from <http://lobid.org/organisation/DE-101/about>.
- <sup>13</sup> <http://creativecommons.org/publicdomain/zero/1.0/>.
- <sup>14</sup> <http://lobid.org/sparql/>.
- <sup>15</sup> See the description at the Open Data registry the Data Hub: <http://thedatahub.org/dataset/lobid-resources>.
- <sup>16</sup> <http://id.loc.gov/vocabulary/iso639-2.html>.
- <sup>17</sup> Retrieved from <http://lobid.org/resource/HT002189125/about>.
- <sup>18</sup> <http://www4.wiwiss.fu-berlin.de/pubby/>.
- <sup>19</sup> RDFa (or Resource Description Framework – in – attributes) is a W3C Recommendation that makes it possible to embed RDF statements in web documents.
- <sup>20</sup> <http://www.w3.org/2005/04/fresnel-info/>.
- <sup>21</sup> <http://www.w3.org/2005/04/fresnel-info/#implementation>.
- <sup>22</sup> <http://www.w3.org/2001/11/IsaViz/>.
- <sup>23</sup> <http://jfresnel.gforge.inria.fr/>.
- <sup>24</sup> <http://simile.mit.edu/wiki/Longwell>.
- <sup>25</sup> <http://www.uni-koblenz-landau.de/koblenz/fb4/institute/IFI/AGStaab/Research/systeme/lena>.
- <sup>26</sup> <http://www4.wiwiss.fu-berlin.de/pubby/>.
- <sup>27</sup> <http://elda.googlecode.com/hg/deliver-elda/src/main/docs/index.html>.
- <sup>28</sup> <https://github.com/lobid/Phresnel>.
- <sup>29</sup> <http://www.w3.org/TR/cooluris/#r303gendocument>.
- <sup>30</sup> <http://www.w3.org/TR/rdf-sparql-query/#funcex-regex>.
- <sup>31</sup> <http://lobid.org/organisation/search>.
- <sup>32</sup> <http://lobid.org/resource/search>.
- <sup>33</sup> <http://www.elasticsearch.org/>.
- <sup>34</sup> <http://createjs.org/>.

<sup>35</sup> An initiative for creating a synchronization standard for web content, called “ResourceSync”, is already underway at the NISO, see <http://www.niso.org/workrooms/resourcesync/>. The resulting standard would be quite relevant for this context.

<sup>36</sup> <http://xmpp.org/>.

<sup>37</sup> <http://docs.api.talis.com/getting-started/changesets>.

<sup>38</sup> <http://git-scm.com/>.

<sup>39</sup> <http://www.mementoweb.org/>.

<sup>40</sup> <http://json-ld.org/>.