

Zur Entwicklung eines Linked-Open-Data-Dienstes für Bibliotheksdaten

Felix Ostrowski (literarymachine.net)

Adrian Pohl (Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen, hbz)



Lizenz: <http://creativecommons.org/licenses/by/3.0/de/>

Abstract (Deutsch)

Das Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz) startete seinen Linked-Open-Data-Service lobid.org im August 2010 und verbessert seitdem kontinuierlich die dem Dienst zugrundeliegenden Konvertierungsprozesse, Datenmodelle und Software. Dieser Beitrag erläutert zunächst Hintergrund und Motivation für die Entwicklung von lobid.org. Er beschreibt dann das zugrundeliegende Software-Framework *Phresnel*, das in PHP geschrieben ist und Präsentations- wie Editierungsmöglichkeiten von RDF-Daten auf Basis des Fresnel Display Vocabulary for RDF bereitstellt. Das Papier gibt einen Überblick über den derzeitigen Stand der *Phresnel*-Entwicklung und diskutiert die technischen Herausforderungen. Schließlich werden mögliche Aussichten für die weitere *Phresnel*-Entwicklung skizziert.

Abstract (Englisch)

The Northrhine-Westphalian Library Service Center (hbz) launched its LOD service lobid.org in August 2010 and since then has been continuously improving the underlying conversion processes, data models and software. This 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 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 *Phresnel* development and discusses the technical challenges encountered. Finally, possible prospects for further developing *Phresnel* are outlined.

Einleitung

Das Thema Linked Open Data (LOD)¹ hat in den letzten Jahren zunehmende Aufmerksamkeit erlangt, und es lassen sich vermehrt entsprechende bibliothekarische Projekte ausmachen. Seit 2009 beschäftigt sich das hzb sowohl mit den rechtlichen als auch den technischen Aspekten von Linked Open Data. Im August 2010 startete das hzb seinen LOD-Service lobid.org – das Akronym steht für “Linking Open Bibliographic Data” – und verbessert diesen Dienst seitdem kontinuierlich.

Im Kontext von lobid.org wird auf eine Weiterentwicklung des zugrundeliegenden Software-Frameworks abgezielt, so dass es read/write-fähig wird, das heißt sowohl die HTML-Präsentation der RDF-Daten als auch die Generierung von Webformularen zum Erstellen, Aktualisieren und Löschen von in RDF repräsentierten Informationen abdeckt. Dieser Beitrag erläutert den Hintergrund und die Motivation der Entwicklung, erläutert technische Herausforderungen und gibt einen Überblick über den gegenwärtigen Stand des Projekts.

Motivation und erwarteter Nutzen

Aus Sicht des hzb als Verbundzentrale stellt Linked Open Data einen sehr interessanten Ansatz für die zukünftige verteilte, kooperative Katalogisierung im Web dar. Einige Vorteile seien im Folgenden genannt.

Webintegration

Die Berücksichtigung internationaler domänenübergreifender Web-Standards bei der Erstellung und Publikation von Daten bringt es mit sich, dass die resultierenden Daten in das Web integriert sind. Das bedeutet konkret:

- *Erhöhte Auffindbarkeit.* In das Web integrierte Daten können leicht von Suchmaschinen, Portalen wie Europeana oder Die Deutsche Bibliothek sowie anderen Aggregatoren eingesammelt werden.
- *Vereinfachte Mehrfachnutzung.* Publierte RDF-Daten können ohne großen Aufwand von verschiedenen Services parallel genutzt werden – inner- wie außerhalb des hzb.

1 Zur Klarstellung unserer Verwendung des Ausdrucks “Linked Open Data”: Wir bezeichnen eine Datenpublikation als “Linked Data” wenn sie den Anforderungen in Berners-Lee, 2006 genüge trägt. Als “Linked Open Data” bezeichnen wir publizierte Daten, wenn sie zusätzlich zu den genannten Anforderungen Open Knowledge Foundation, 2009 entsprechen. Für mehr Informationen zu Open Data siehe Pohl, 2010. Linked Data wird in Pohl, 2011 detailliert erläutert.

- *Maximale Interoperabilität und Nachnutzbarkeit.* Webstandards erleichtern die Nachnutzung von Daten, indem sie notwendige Konversionsprozesse wie auch das Post-Processing minimieren.
- *Flexibilität.* RDF und Triple Stores sind sehr flexibel im Hinblick auf Erweiterungen und Änderungen des zugrundeliegenden Datenmodells.

Synergieeffekte

Die Anwendung von Best Practices aus der Linked-Open-Data-Community und der sukzessiven Standardisierung der von verschiedenen hbz-Services produzierten Daten bewirkt intra- wie interinstitutionelle Synergieeffekte. Bei der gegenseitigen Nachnutzung von Daten verschiedener hbz-Projekte sehen wir dies bereits jetzt. LOD macht dies auf unkomplizierte Weise möglich, wohingegen die Berücksichtigung proprietärer Schnittstellen und verschiedener Formate eine Kommunikation verschiedener Services recht ressourcenaufwändig macht. Somit hat eine Standardisierung von Services den Effekt, Ressourcen freizusetzen, die für die Optimierung bestehender Dienstleistungen oder den Aufbau zusätzlicher Angebote genutzt werden können.

Größere Herstellerunabhängigkeit

Die Bereitstellung vieler bibliothekarischer Dienstleistungen ist technologieabhängig und es gibt immer weniger Anbieter und damit auch weniger Konkurrenz in dem Geschäft mit bibliothekarischer Software. Viele Institutionen sind stark abhängig von den Produkten nur eines Anbieters und ein Wechsel zu einem anderen Anbieter ist mit signifikanten Kosten verbunden, mit anderen Worten: Es existiert ein starker Lock-In-Effekt.

Die Anwendung von Best Practices der Linked-Open-Data-Community verbessert diese Situation für die Kunden (Bibliotheken und Verbünde) auf zweierlei Weise:

1. Ist man in der Lage, die eigenen Daten leicht in einem offenen, domänenübergreifenden Standard aus einem System zu exportieren, erleichtert dies einen Umstieg auf ein anderes System erheblich, wodurch ein Lock-In-Effekt verhindert wird.
2. Werden bibliothekarische Daten in RDF und nicht mehr in – zwar offenen aber opaken – Formaten wie MARC oder MAB gespeichert und ausgetauscht, dann wird auch andere, generische Software von anderen Anbietern interessant bzw. es wird für andere Anbieter leichter den Markt zu betreten. Ein stärkerer Wettbewerb führt zur Verbesserung der Produkte und/oder zu sinkenden Preisen.

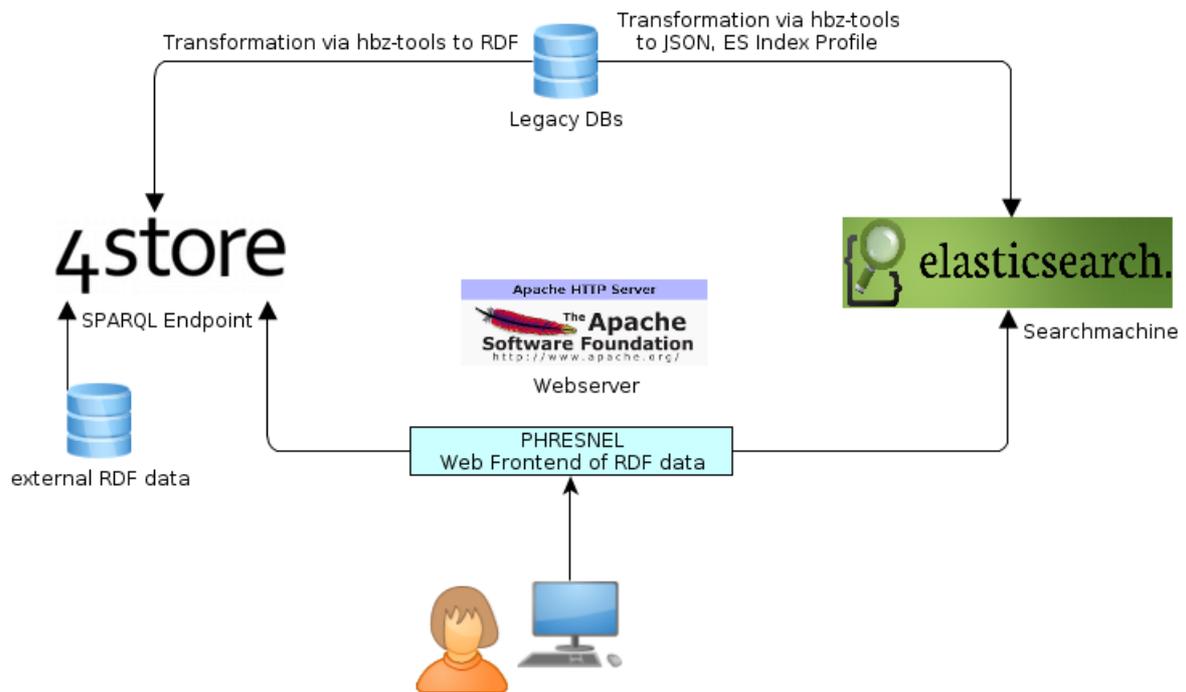
lobid.org

Im August 2010 startete das hbz den experimentellen Linked-(Open)-Data-Dienst lobid.org², der aus zwei Subdiensten besteht: bibliographische Ressourcen (lobid-resources) und einem Verzeichnis von Bibliotheken, Museen und verwandten Organisationen (lobid-organisations). lobid.org folgt den Linked-Data-Prinzipien und geht auch – wann immer möglich – konform mit der Open Definition (Open Knowledge Foundation, 2009).

Seit 2010 wurden und werden die beiden lobid.org-Services und die ihnen zugrundeliegenden RDF-Daten kontinuierlich verbessert:

- Die Daten wurden durch mehrfache Überarbeitung von Konvertierung und Vokabularauswahl verbessert.
- Die vorhandene Information wurde zunehmend kontextualisiert durch Links zu anderen Datenquellen.
- Interaktionsmöglichkeiten für Endnutzer wurden verbessert, z.B. durch eine Suchmaschinenanbindung und die Vereinheitlichung der Oberfläche beider Subdienste (siehe Abschnitt 5).

2 <http://lobid.org/>



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

Abbildung 1: lobid.org Technologie-Stack

Abbildung 1 gibt einen allgemeinen Überblick über den gegenwärtig in lobid.org verwendeten Technologie-Stack, die Datenquellen und Konversionsprozesse. Es wird ausschließlich Open-Source-Software eingesetzt. Der verwendete Triple Store ist 4store³ der Firma Garlik, für die Volltextindexierung und die Suchschnittstelle wird elasticsearch⁴ eingesetzt. Das Webfrontend läuft auf einem Apache-Server und wird mit dem Phresnel-Framework generiert, das unten detailliert beschrieben wird.

Wie in Abbildung 1 zu sehen ist, basiert lobid.org fast ausschließlich auf Daten, die aus existierenden Legacy-Systemen exportiert und – unter Nutzung hbz-spezifischer Tools – nach RDF konvertiert werden. Die resultierenden RDF-Daten werden angereichert mit Links zu anderen Datenquellen in der LOD-Cloud. Darüber hinaus werden auch einige externe LOD-Datasets in den Triple Store geladen: dies sind derzeit die in lobid.org genutzten Ontologien sowie die von der deutschen Nationalbibliothek (DNB) bereitgestellte Gemeinsame Normdatei (GND)⁵. Bisher gibt

3 <http://4store.org/>

4 <http://www.elasticsearch.org/>

5 Siehe den Eintrag für das LOD-GND-Dataset auf the Data Hub: <http://thedatahub.org/dataset/dnb-gemeinsame-normdatei>.

es keine Möglichkeit, auf einfache Weise manuell neue Informationen zu ergänzen oder Fehler in den Daten zu beheben.

lobid organisations

Als das hbz mit der Publikation von Linked Open Data begann, wurde schnell klar, dass die bibliographischen Daten und Bestandsinformationen aus dem hbz-Verbundkatalog nur einen Anfang darstellen würden. Um nützliche Services auf der Basis von Linked Open Data aufzubauen, ist es ebenso notwendig URIs für und RDF-Beschreibungen von Exemplaren, Bibliotheken und Dienstleistungen zu haben. Zum Beispiel braucht man für eine geobasierte Suche nach Exemplaren einer bestimmten Ausgabe im Umkreis von 5 km URIs und Beschreibungen von mindestens drei Dingen: einer Manifestation M, die exemplifiziert wird durch ein Exemplar E, das im Bestand einer Organisation O ist. Als RDF-Graph sieht das beispielhaft wie folgt aus (Abbildung 2):

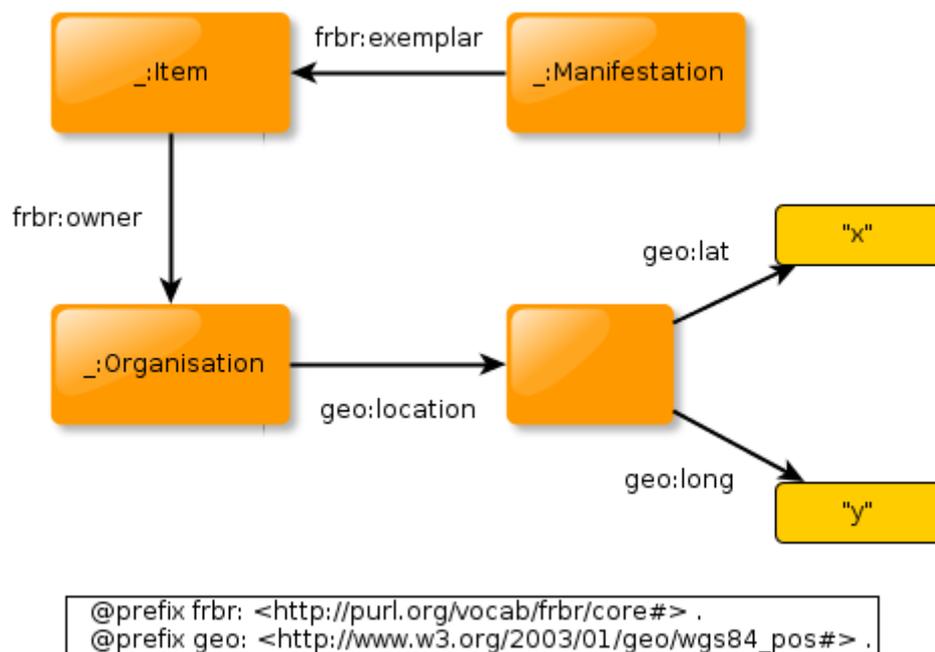


Abbildung 2: Graph der Relationen zwischen Manifestation, Exemplar und Organisation

Da sich vor zwei Jahren wenig bewegte, wenn man Menschen und/oder Organisationen darum bat, Linked Data bereitzustellen, wurde kurzerhand beschlossen, selbst URIs für bibliothekarische Institutionen zu prägen und

Beschreibungen bereitzustellen. Im Juli 2010 startete schließlich lobid.org mit dem Subdienst lobid-organisations (Ostrowski, 2010).

Die dem Organisationsverzeichnis zugrundeliegenden Datenquellen sind das deutsche ISI- und Sigelverzeichnis⁶ und die MARC Organisation Code Database⁷. Bislang wurden in lobid.org URIs für mehr als 40.000 Institutionen geprägt und mit entsprechenden Informationen⁸ versehen. Allerdings wird weder ein Vollabzug der Daten zum Download bereitgestellt noch sind die Daten offen lizenziert, weil die Daten nicht vom hbz produziert wurden und es somit nicht berechtigt ist eine Nutzungslizenz für Dritte zu bestimmen.

Zusätzlich zu den Daten aus den erwähnten Datenquellen werden Verlinkungen zu anderen Datenquellen in der LOD-Cloud generiert. Bis jetzt wurden Links zur DBpedia und Wikipedia (Christoph, 2012d) sowie zu GeoNames (Christoph, 2012a) ergänzt. Darüberhinaus wurden die Organisationsbeschreibungen mit QR-Codes angereichert, in denen Kontaktinformationen kodiert sind (ebd.). Außerdem existieren Geokoordinaten für die meisten Institutionen, so dass etwa ein entsprechender Kartenausschnitt in lobid.org eingebettet werden kann.⁹ Eine Beispielbeschreibung (die - wie in Abschnitt 5.1 beschrieben - auf Basis der RDF-Daten generiert wurde) wird in Abbildung 3 gezeigt.

http://lobid.org/organisation/DE-101									
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	http://www.dnb.de								
E-Mail	mailto:info-f@dnb.de								
Kontakt-QR									
Weitere Information	http://dispatch.opac.d-nb.de/DB=1.2/CMD?ACT=SRCHA&IKT=8529&TRM=DE-101								
Sitz	http://sws.geonames.org/6553153/								
Wikipedia-Link	http://de.wikipedia.org/wiki/Deutsche_Nationalbibliothek								
DBpedia-Link	http://de.dbpedia.org/resource/Deutsche_Nationalbibliothek								

Abbildung 3: lobid.org-Beschreibung der Deutschen Nationalbibliothek¹⁰

lobid resources

lobid-resources stellt eine Linked-Open-Data-Schnittstelle für Open Data aus der hbz-Verbunddatenbank dar. Publiziert werden URIs für und Beschreibungen von bibliographischen Ressourcen wie Monographien, Zeitschriften, Serien, Mehrbändigen Werken usw. Seit der ersten Publikation von Open Data im hbz-Verbund im März 2010 (Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen, 2010) wurden immer mehr Daten des Verbundkatalogs unter einer offenen Lizenz veröffentlicht. Mit Stand August 2012 kooperieren bis auf zwei Bibliotheken alle hbz-Verbundbibliotheken mit dem hbz bei der Freigabe der Daten (Christoph, 2012). Damit umfassen die Daten mittlerweile knapp 16 Millionen Titeldatensätze, die unter einer Creative-Commons-Zero-Lizenz¹¹ veröffentlicht werden, das sind etwa 85% des gesamten Verbundkatalogs.¹² Die Daten können über eine Suchschnittstelle¹³ sowie einen SPARQL-Endpoint¹⁴ abgefragt werden. Außerdem steht ein Vollabzug der RDF-Daten zum Download¹⁵ bereit.

Auch die lobid-resources Daten wurden mit Verlinkungen zu anderen Datenquellen angereichert. Mittels einfacher ISBN- und Titel-Matchingalgorithmen und ein wenig Postprozessierung wurden für einige Ressourcen Links zur Dbpedia (Christoph, 2012b), zur Open Library (Christoph, 2012c) und zu Projekt Gutenberg ergänzt. Diese Links sind über die einfache Verlinkung hinausgehend nützlich. Beispielsweise wird eine wechselseitige Anreicherung verschiedener Titel mit Schlagwörtern, Links etc. ermöglicht.

Abbildung 4 zeigt die Beispielbeschreibung einer bibliographischen Ressource in lobid.org inklusive Links zu anderen Datenquellen.

10 URL: <http://lobid.org/organisation/DE-101/about>

11 <http://creativecommons.org/publicdomain/zero/1.0/>

12 Neben Daten der Bestände der zwei Bibliotheken, die nicht mit dem hbz kooperieren, werden auch die Metadaten, zu den über die Nationallizenzen erworbenen Inhalten nicht als Open Data publiziert. Der Grund ist, dass die von der DFG festgelegten "Grundsätze für den Erwerb DFG-geförderter überregionaler Lizenzen", denen die einzelnen mit den Verlagen im Rahmen der Nationallizenzen geschlossenen Verträgen entsprechen müssen, eine Publikation der Metadaten als Open Data verhindern (vgl. Kreutzer, 2011, S. 22f.).

13 <http://lobid.org/resource/search>

14 <http://lobid.org/sparql/>

15 Siehe die Beschreibung im Open-Data-Verzeichnis the Data Hub: <http://thedatahub.org/dataset/lobid-resources>.

http://lobid.org/resource/HT002189125									
Titel	Typee								
Titelzusatz	a peep at Polynesian life								
Autor	http://d-nb.info/gnd/118580604								
Erscheinungsjahr	1982								
Typ	< http://purl.org/ontology/bibo/Book >								
Typ	< http://purl.org/vocab/frbr/core#Manifestation >								
Volltext	http://gutenberg.org/ebooks/1900								
Volltext	http://gutenberg.org/ebooks/28656								
Volltext	http://gutenberg.org/ebooks/9269								
Volltext	http://gutenberg.org/ebooks/23969								
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	http://openlibrary.org/works/OL14953734W								
Werktitel	http://dbpedia.org/resource/Typee								
Werktitel	http://de.dbpedia.org/resource/Typee								
Exemplar	http://lobid.org/item/HT002189125%3AU+MEL-11								
	<table border="1"> <tr> <td>Typ</td> <td><http://purl.org/vocab/frbr/core#Item></td> </tr> <tr> <td>Zugehöriger Titel</td> <td>http://lobid.org/resource/HT002189125</td> </tr> <tr> <td>Besitzer</td> <td>Institut für Anglistik, Amerikanistik und Keltologie, Bibliothek</td> </tr> <tr> <td>Signatur</td> <td>U MEL-11</td> </tr> </table>	Typ	< http://purl.org/vocab/frbr/core#Item >	Zugehöriger Titel	http://lobid.org/resource/HT002189125	Besitzer	Institut für Anglistik, Amerikanistik und Keltologie, Bibliothek	Signatur	U MEL-11
Typ	< http://purl.org/vocab/frbr/core#Item >								
Zugehöriger Titel	http://lobid.org/resource/HT002189125								
Besitzer	Institut für Anglistik, Amerikanistik und Keltologie, Bibliothek								
Signatur	U MEL-11								
Weitere Information	http://193.30.112.134/F/?func=find-c&ccl term=IDN%3DHT002189125								

Abbildung 4: Beispielbeschreibung einer bibliographischen Ressource in lobid.org¹⁶

Fresnel Display Vocabulary for RDF

¹⁶ URL: <http://lobid.org/resource/HT002189125/about>

Hintergrund

Anfangs wurden in lobid.org die konvertierten Daten anhand von Pubby, einem Linked Data Frontend für SPARQL Endpoints¹⁷, präsentiert. Wenngleich die Einrichtung von Pubby sich sehr einfach gestaltet, so exponieren die resultierenden Ansichten - insbesondere die für Menschen gedachte HTML-Darstellung - zu viele Details der zugrundeliegenden Technologien. Die Präsentation der Organisationsdaten in lobid.org hingegen wurde anhand von selbsterstellten SPARQL-Anfragen, PHP-Skripten und HTML-Templates umgesetzt. Diese Variante verhalf zwar zu größerer Flexibilität bei der Darstellung, allerdings war sie nicht ohne weiteres auf andere Daten übertragbar, da dafür jeweils eigene SPARQL-Anfragen und HTML-Templates benötigt wurden. Darüber hinaus stand bereits der Plan im Raum, Bibliotheken eine einfache Möglichkeit zu bieten, Informationen in RDFa zu erstellen. Um dies zu ermöglichen, wurde ein einfacher, intuitiver Editor benötigt. Anstelle direkt die zugrunde liegenden Daten zu bearbeiten, sollte also eine HTML-Formular-basierte Lösung eingesetzt werden, um so die Arbeit in einer vertrauten Web-Umgebung zu ermöglichen.

Mit diesen Anforderungen im Hinterkopf begann die Suche nach einer Schema-Sprache, anhand derer ein Datenmodell formuliert werden kann, aus welchem wiederum das Frontend automatisch hervorgeht. Beim Umgang mit RDF-Daten sind die naheliegendsten Lösungen RDF-Schema (RDFS) und die Web Ontology Language (OWL). Da aber die Ontologien, die in diesen Sprachen erstellt werden, meist anwendungsunabhängig sind, waren Experimente in diese Richtung eher unfruchtbar, da auch die resultierenden Editor-Ansichten zu generisch waren, um die Anforderungen zu erfüllen. Insbesondere die Verwendung von Klassen und Eigenschaften aus verschiedenen Vokabularen ist mit RDFS bzw. OWL nahezu unmöglich auf eine präzise und eingängige Art und Weise zu gestalten. Anders ist dies im *Fresnel Display Vocabulary for RDF*. Anhand dieses Vokabulares ist es möglich zu spezifizieren, *welche* Informationen aus einem RDF-Graphen angezeigt, und *wie* diese dargestellt werden sollen¹⁸, und zwar ohne mit den zugrunde liegenden Ontologien zu interferieren. Ähnlich wie auch die oben erwähnten Ontologie-Sprachen basiert auch Fresnel auf RDF. Dies ermöglicht es, innerhalb der gesamten Implementation in einem Datenmodell zu verbleiben.

Fresnel-Linsen

Fresnel-Linsen dienen der Spezifizierung, *welche* Daten angezeigt werden sollen. Eine einzelne Linse kann auf verschiedene Weisen mit Instanzen verknüpft werden.

17 <http://www4.wiwiss.fu-berlin.de/pubby/>

18 <http://www.w3.org/2005/04/fresnel-info/>

Die einfachste Möglichkeit ist dabei die Referenzierung der Klasse, also über die `rdf:type`-Eigenschaft, wie unten veranschaulicht. Für die so ausgewählten Instanzen wird dann eine geordnete Liste von Prädikaten angegeben, was dank der Turtle-Notation sehr einfach und gut lesbar ist. Um auch Daten zu einer verknüpften Entität mit anzuzeigen, kann anstelle eines einfachen Prädikates auch auf eine andere Linse verwiesen werden. Im Beispiel (Abbildung 5) etwa werden Vor- und Nachname des Autors ebenfalls angezeigt, und nicht nur dessen URI. Letzteres wäre der Fall, wenn nur das Prädikat `dc:creator` in der Liste aufgeführt würde, ohne Verweis auf die entsprechende `:person`-Sub-Linse wie im Beispiel auf der rechten Seite.

<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>
---	---

Abbildung 5: Fresnel-Linsen und Sub-Linsen

Die Anwendung obiger Linse auf einen Triplestore, der die entsprechenden Daten enthält, sollte die in Abbildung 7 dargestellten Tripel hervorbringen, welche genau in dieser Reihenfolge dargestellt werden sollen:

<pre> <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" . </pre>

Abbildung 6: Auf Basis einer Fresnel-Linse gewonnene RDF-Daten

Zusammenfassend lässt sich festhalten, dass Fresnel-Linsen eine sehr präzise und deklarative Möglichkeit bieten, um auszudrücken, welche Daten ausgewählt und in welcher Reihenfolge sie angezeigt werden sollen.

Formate

Fresnel-Formate erlauben es auszudrücken, wie die anhand von Linsen ausgewählten Daten dargestellt werden sollen. Die Möglichkeiten erstrecken sich dabei von benutzerdefinierten Bezeichnungen von Prädikaten bis hin zu Referenzen auf externe CSS-Klassen. Ein Beispiel findet sich in Abbildung 7.

```
[] 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 .
```

Abbildung 7: Formate für benutzerdefinierte Prädikat-Bezeichnungen und anzuwendende CSS-Klassen

Ähnlich wie es Fresnel-Linsen erlauben, Daten auszuwählen und zu ordnen, bieten Fresnel-Formate also die Möglichkeit, die Darstellung auf eine deklarative, anwendungsunabhängige Art und Weise zu konfigurieren.

Implementation

Unter den verschiedenen existierenden Implementierungen¹⁹ von Fresnel gab es keine mit einer Bearbeitungsansicht und keine die – wie lobid.org – in PHP implementiert ist. Außerdem schienen die meisten Projekte inaktiv. Daher wurde prototypisch ein neues PHP-basiertes Framework namens “Phresnel” implementiert, das auch eine Datenbearbeitung ermöglicht. Derzeit wird nur eine kleine Untermenge der Linsen- und Formatfeatures unterstützt, der Umfang beschränkt sich hierbei auf die für einen Prototypen absolut notwendigen Funktionen.

¹⁹ <http://www.w3.org/2005/04/fresnel-info/#implementation>

Anzeige

Um Daten entsprechend einer Fresnel-Linse anzuzeigen, ermittelt die Web-Anwendung zunächst anhand der URL die zu verwendende Linse, z.B. "Dokument". Anschließend wird mithilfe des Phresnel-Frameworks eine generische (X)HTML-Ansicht (mit eingebettetem RDFa) der angeforderten Daten erzeugt (Abbildung 8). Zu diesem Zeitpunkt wird davon ausgegangen, dass ein HTTP-303-Redirect entsprechend des Linked-Data-Design-Patterns²⁰ bereits stattgefunden hat.

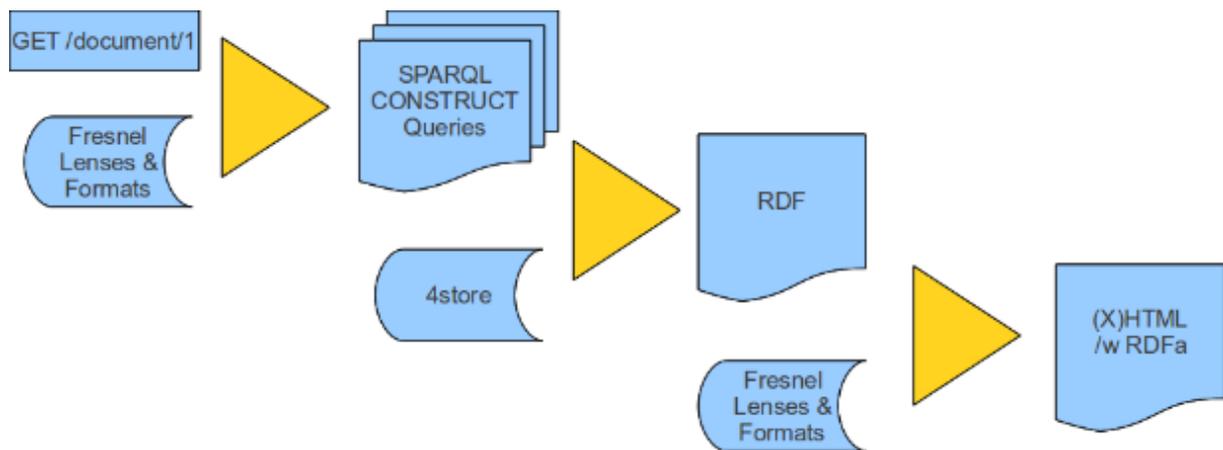


Abbildung 8: Generierung von (X)HTML-Ansichten mit Phresnel

Intern verwendet Phresnel die Linsen-Definitionen, um wie in Abbildung 9 dargestellt eine Reihe von SPARQL-CONSTRUCT-Abfragen zu generieren. Danach wird entsprechend der Linsen- und Format-Definitionen die eigentliche Ausgabe erzeugt.

```
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=rdf:type || ?p=dc:title || ?p=bibo:isbn || ?p=dc:creator) .
}

CONSTRUCT {
  ?s ?p ?o .
} WHERE {
```

²⁰ <http://www.w3.org/TR/cooluris/#r303gendocument>

```

<http://lobid.org/resource/HT002948556> dc:creator ?s .
?s a foaf:Person .
?s ?p ?o .
FILTER (?p=rdf:type || ?p=foaf:givenName || ?p=foaf:familyName) .
}

```

Abbildung 9: Von Phresnel generierte SPARQL-CONSTRUCT-Abfragen

Derzeit ist nur ein hartcodiertes Box-Model (Bizer et al., 2005) auf Basis von verschachtelten Tabellen implementiert. Die Transformation zu (X)HTML kann natürlich übersprungen werden, wenn anhand von Content Negotiation eine reine RDF-Darstellung angefragt wurde.

Bearbeiten

Die Schritte für die Erstellung der Bearbeitungsansicht einer Ressource sind den obigen, zur reinen Darstellung notwendigen, sehr ähnlich – mit einem großen Unterschied. Prinzipiell wird das RDF, welches sich durch die Auswertung der erzeugten SPARQL-Abfragen ergibt, zu einem (X)HTML-Formular umgewandelt, indem alle Literale einfach durch Texteingabelemente ersetzt werden. Dies führt allerdings in den Fällen, in denen Links zu anderen Entitäten erzeugt werden sollen, zu einem unschönen Interface, da hier URIs eingegeben werden müssten. Mit Hinsicht auf die Benutzbarkeit ist dies inakzeptabel.

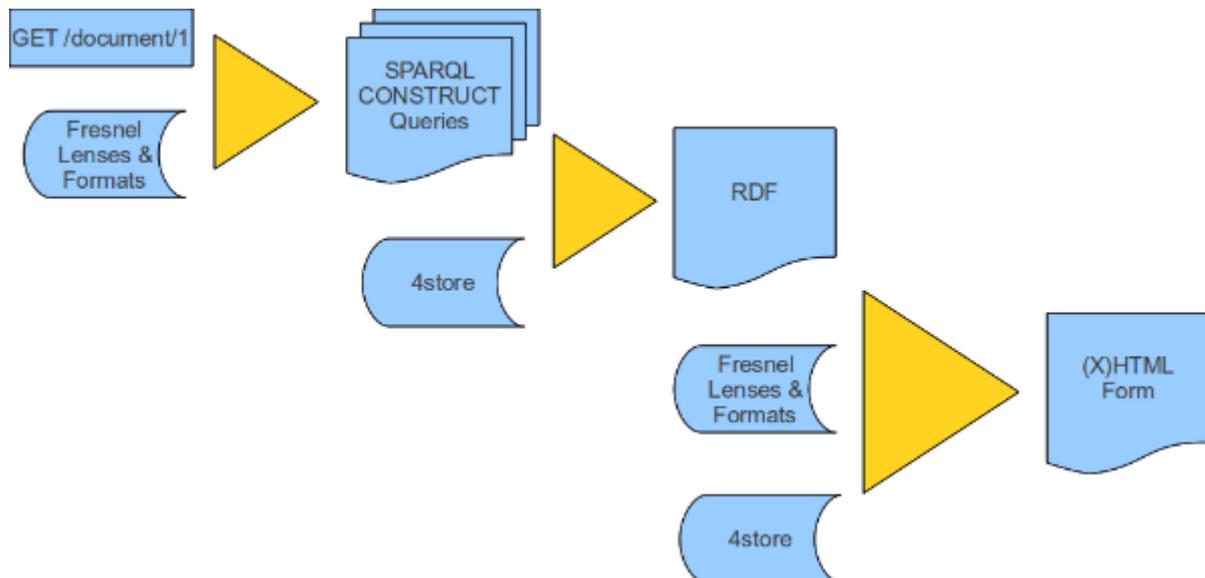


Abbildung 10: Generierung von (X)HTML-Formularen mit Phresnel

Die aktuelle Implementierung löst dieses Problem auf eine nur ausreichende, aber zumindest einigermaßen optisch ansprechende Art und Weise. Verfügbare Link-Ziele aus dem Triple Store, z. B. dem obigen Beispiel folgend eine Liste der möglichen Autoren, werden als Dropdown-Listen angezeigt. Leider bringt diese Lösung zwei schwerwiegende Einschränkungen mit sich: Die Liste kann (1) zu lang werden und (2) sind die Listeneinträge auf die eigene Datenquelle beschränkt, was es schwierig macht, andere Quellen im Web zu verknüpfen. Für die Zukunft ist geplant, diese beiden Einschränkungen mit Hilfe von Autovervollständigung zu beheben. Diese wird anhand von Javascript gesteuert und soll es ermöglichen, sowohl für interne als auch externe Datenquellen die Eingabe des Nutzers zu vervollständigen, so dass eine komplette Liste der möglichen Verknüpfungen nicht mehr notwendig ist. Für das Verlinken auf interne Daten kann hierbei der im nächsten Abschnitt erläuterte Suchindex verwendet werden. An dieser Stelle sei nochmals darauf hingewiesen, dass zwar Daten mit diesem Web-Frontend bearbeitet werden können, zur Zeit aber kein Workflow für die Speicherung implementiert ist.

Suche

Während das Linked-Data-Paradigma für die Navigation hervorragend geeignet ist, so sind die Möglichkeiten einer Volltextsuche für das Auffinden von Ressourcen unabdingbar. Zwar unterstützt SPARQL reguläre Ausdrücke²¹, die für die Volltextsuche verwendet werden können. Dies skaliert allerdings derzeit nicht auf die Datenmenge von lobid.org.

Deshalb basiert sowohl die Suche für Organisationen²² als auch die für Ressourcen²³ derzeit auf einem elasticsearch-Index²⁴, der wiederum über eine API, die eine CQL-Schnittstelle bereitstellt, angesprochen wird. Dieses konkrete Setup hat historische Gründe.²⁵

Wegen dieses Aufbaus haben die aus der CQL-Anfrage resultierenden Suchergebnisse, die als Atom-Feeds ausgeliefert werden, eine von den Linsen-Definition abweichende Datenstruktur. Daher werden derzeit nur Identifikatoren aus den Suchergebnissen extrahiert, um die entsprechenden Ressourcen dann anhand von Phresnel aus dem Triplestore zu laden und darzustellen. Dies funktioniert recht gut und ist deutlich leistungsfähiger als die Verwendung von nativen SPARQL-

21 <http://www.w3.org/TR/rdf-sparql-query/#funcex-regex>

22 <http://lobid.org/organisation/search>

23 <http://lobid.org/resource/search>

24 <http://www.elasticsearch.org/>

25 Der Titeldaten-Index hat lange vor lobid.org existiert und wird von mehreren hbz-Diensten verwendet. Es war am einfachsten, die Index für Organisationen einfach in die gleiche Infrastruktur zu integrieren.

Abfragen für die Volltextsuche. Die Tatsache, dass, nachdem die Ressource identifiziert worden ist, die entsprechenden Daten noch aus dem Triple Store geholt werden müssen, zeigt, dass dieser Aufbau aber noch nicht optimal ist. Besser wäre es, die Daten direkt aus dem Suchergebnis zu übernehmen, was aber eben eine Harmonisierung der Datenstrukturen zwischen Volltextindex und Triplestore voraussetzt. Dies ist nur eines der Probleme, die als nächstes in Angriff genommen werden sollen.

Ausblick

Die derzeitige Proof-of-Concept-Implementierung eines LOD-basierten System für den Lese- und Schreibzugriff auf Bibliotheksdaten zeigt interessante Perspektiven für das zukünftige Datenmanagement auf. Unbestreitbar sind die Bemühungen noch ganz am Anfang. Möglichkeiten zur Weiterentwicklung des Systems werden im folgenden aufgezeigt.

Rein Javascript-basierter Editor

Da der wichtigste Teil des Editors - die Komponente zum Auffinden und Verknüpfen von Ressourcen - aus oben genannten Gründen überarbeitet werden muss, wird in Betracht gezogen, den Editor insgesamt auf Basis von RDFa und Javascript zu implementieren. In einer create.js²⁶ ähnlichen Art würde damit auf serverseitig generierte Formulare zugunsten eines clientseitigen Editors verzichtet. Dies würde nicht nur die Benutzerfreundlichkeit verbessern, sondern auch die Belastung des Servers minimieren und für eine bessere Trennung zwischen Front- und Backend sorgen.

Datenproduktion und Pflege

Die Ergebnisse einer Bearbeitung von Daten anhand des Web-Frontends müssen persistent gemacht werden. Es gibt einige nicht-triviale Entscheidungen, die in dieser Hinsicht getroffen werden müssen. So gilt es etwa zu klären, wo die Daten gespeichert werden (Triple Store, Dateisystem, Suchindex), wie sie organisiert werden (Named Graphs (Dodds, 2009), ...), welche Provenienzdaten benötigt werden und wie das Berechtigungssystem aufgebaut sein muss. Da eine Anwendung wie die in diesem Artikel beschriebene in einer vernetzten, dezentralen Umgebung verwendet wird, muss auch eine Lösung dafür angedacht werden, wie andere angeschlossene Dienste über die Erstellung, Aktualisierung und Löschung von Daten

26 <http://createjs.org/>

informiert werden können²⁷. Eine Idee in dieser Hinsicht ist es, ein nachrichtenbasiertes Protokoll wie IRC oder XMPP²⁸ zu verwenden.

Offensichtlich wichtige Aspekte der Datenverwaltung sind die Identifikation und Authentifizierung derjenigen Akteure, die die Daten bearbeiten. Anstatt ein solches System von Grund auf neu zu konzipieren und implementieren, sollte möglichst auf bestehende Standards zurückgegriffen werden – im Idealfall sollten diese ebenfalls auf Linked-Data-Prinzipien fußen. Aus diesem Grund ist der vielversprechendste Ansatz der Einsatz von WebID (Sporny et al., 2011). Hierbei werden FOAF-Beschreibungen in Verbindung mit SSL-Zertifikaten verwendet, was zu einem sicheren verteilten Identifizierungs- und Authentifizierungsmechanismus führt, der sowohl für Menschen als auch für Maschinen nachvollziehbar ist.

Versionierung

Zu den wichtigsten Provenienzangaben gehört ein lückenloser Verlauf der an den Daten vollzogenen Änderungen zusammen mit der Identifikation des Akteurs (sei es ein System oder eine Person), der für diese Änderungen verantwortlich ist. Während es durchaus möglich ist, Änderungen an RDF-Daten als Changesets anhand eines RDF-Vokabulars²⁹ auszudrücken, ist dies sehr wahrscheinlich nicht der effizienteste Weg, da für eine Änderung sehr viele Tripel benötigt werden. Zu den Alternativen, die in Erwägung gezogen werden, gehören der Einsatz eines Systems zur Versionsverwaltung das auf einfachen Dateien basiert (z.B. git³⁰) und die Verwendung der Versionskontrollfeatures von elasticsearch. Um die verschiedenen Versionen der Daten in einer standardisierten Art und Weise ausliefern zu können ist darüber hinaus angedacht, eine Memento-Schnittstelle³¹ für das ausgewählte Versionierungssystem zu implementieren.

JSON-LD in ES / Fresnel-basierter Suchindex

Die Art und Weise, auf die die Suche derzeit in die Anwendung eingebunden ist, ist nicht sehr generisch und hängt von externen organisatorischen und technischen Prozessen ab. Da aber der Lookup ein sehr wichtiger Teil des Gesamtsystems ist, wird eine besser eingebundene Lösung angestrebt. Elasticsearch als schemaloser Suchindex sollte mit RDF-Daten gut verträglich sein. Da es Daten auf Basis von

27 Eine Initiative zur Entwicklung eines Synchronisierungsstandards für Webinhalte mit dem Namen "ResourceSync" gibt es bereits bei der NISO, siehe <http://www.niso.org/workrooms/resourcesync/>. Ein daraus resultierender Standard wäre sehr relevant für diesen Kontext.

28 <http://xmpp.org/>

29 Siehe etwa <http://docs.api.talis.com/getting-started/changesets>.

30 <http://git-scm.com/>

31 <http://www.mementoweb.org/>

JSON indiziert, werden verschiedene JSON-RDF-Serialisierungen evaluiert. Der vielversprechendste Ansatz scheint momentan JSON-LD³² sein. Zum einen wird diese Serialisierung derzeit zum Standard für die Codierung von RDF in JSON, zum anderen – und dies ist aus technischer Sicht wesentlich wichtiger – verwendet JSON-LD einen Key-Value-Ansatz, der mit elasticsearch sehr gut zusammenspielt. Die bereits für die Darstellung und Bearbeitung von Ressourcen verwendeten Fresnel-Linsen könnten verwendet werden, um die JSON-LD-Strukturen für den Index zu erzeugen.

Referenzen

[Für alle URLs gilt: Letzter Zugriff am 30.6.2012.]

T. Berners-Lee (2006, 27. Juli). Linked Data - Design Issues. URL: <http://www.w3.org/DesignIssues/LinkedData.html>

C. Bizer, R. Lee & E. Pietriga (2005, 30. Juni). Fresnel - Display Vocabulary for RDF [User Manual]. URL: <http://www.w3.org/2005/04/fresnel-info/manual/>

P. Christoph (2012a, 16. April). lobid.org news [Blogbeitrag]. URL: <http://wiki1.hbz-nrw.de/x/mYQ7>

P. Christoph (2012b, 3. Mai). First results using SILK to link to Dbpedia [Blogbeitrag]. URL: <http://wiki1.hbz-nrw.de/x/mAJH>

P. Christoph (2012c, 23. Mai). 1.2 M links to Open Library [Blogbeitrag]. URL: <http://wiki1.hbz-nrw.de/x/l4BO>

P. Christoph (2012d, 4. Juni). lobid-organisation - new links to German Dbpedia [Blogbeitrag]. URL: <http://wiki1.hbz-nrw.de/x/6lBO>

P. Christoph (2012e, 16. August). lobid-resources - much more records, authority URIs dereferencable, provenance [Blogbeitrag]. URL: <http://wiki1.hbz-nrw.de/x/cQJ5>

Conference of European National Librarians (2011, 28. September). Europe's national librarians support Open Data licensing [Pressemitteilung]. URL:

32 <http://json-ld.org/>

<https://app.e2ma.net/app/view:CampaignPublic/id:1403149.7214447972/rid:48e64615892ac6adde9a4066e88c736c>

L. Dodds (2009, 5. November). Managing RDF Using Named Graphs [Blogbeitrag]. URL: <http://www.ldodds.com/blog/2009/11/managing-rdf-using-named-graphs/>

D. Marcum & Library of Congress (2011, 31. Oktober). A Bibliographic Framework for the Digital Age. URL: <http://www.loc.gov/marc/transition/news/framework-103111.html>

Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen (2010). Köln, im März 2010: Freigabe der Katalogdaten: Kölner Bibliotheken leisten Pionierarbeit [Pressemitteilung]. URL: <http://www.hbz-nrw.de/dokumentencenter/presse/pm/datenfreigabe>

T. Kreuzer (2011). Open Data – Freigabe von Daten aus Bibliothekskatalogen. Ein Leitfaden. Hg. v. Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen. URL: <http://www.hbz-nrw.de/dokumentencenter/veroeffentlichungen/open-data-leitfaden.pdf>

Open Knowledge Foundation (2009, November). Open Knowledge Definition, v1.1. URL: <http://opendefinition.org/okd/>

F. Ostrowski (2010, 20. Juli). Building a Linked Data based index of library institutions [Blogbeitrag]. URL: <http://blog.lobid.org/2010/07/building-linked-data-based-index-of.html>

A. Pohl (2010): Open Data im hbz-Verbund. ProLibris, 3, S. 109-113. Preprint u.a. einsehbar unter <http://hdl.handle.net/10760/14838>

A. Pohl (2011): Linked Data und die Bibliothekswelt. Erscheint im Konferenzband zur ODOK 2010. Preprint einsehbar unter <http://hdl.handle.net/10760/15324>.

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