

## ■ LIBRARY LINKED DATA — TECHNOLOGIEN, PROJEKTE, POTENTIALE

von *Ulrike Krabo und Markus Knitel*

### Inhalt

1. Einleitung
2. Das Semantic Web
3. Technologien und Standards
4. Linked Data
5. Library Linked Data: Projekte und Erwartungen
6. Herausforderungen
7. LLD-Anwendungen in Österreich
8. Fazit

**Zusammenfassung:** *Das Semantic Web und seine Auswirkungen auf Bibliotheken rücken immer mehr in den Fokus informationswissenschaftlicher Forschung. Dieser Artikel erläutert grundlegende funktionale wie technische Konzepte des Semantic Web, um darauf aufbauend in das Thema Library Linked Data einzuführen. Dafür werden einige kürzlich entstandene Projekte und Initiativen vorgestellt. Neben den Visionen und Zielen der jeweiligen Initiatoren, wie bessere Sichtbarkeit von bibliographischen Daten und Entwicklung neuer Applikationen, werden auch offene technische und rechtliche Fragestellungen bzw. Probleme kurz angerissen. In einem letzten Punkt werden mögliche praktische Linked Data-Anwendungsfälle für den österreichischen Kontext vorgestellt.*

**Schlagwörter:** *Linked Library Data, Semantic Web*

**Abstract:** *The Semantic Web, Library Linked (Open) Data and its impact on libraries has recently become a popular and much discussed topic. The aim of this article is to provide a concise overview of these concepts and to point out some use cases for the Austrian context. The first part is dedicated to an introduction to the fundamental concepts of the Semantic Web. Then some recent Linked Library Data-projects and initiatives, especially throughout Germany, their goals and challenges are being presented. Finally some scenarios and chances for the Austrian context, like enhanced search engines or catalogue enrichment, are discussed.*

**Keywords:** *Linked Library Data, Semantic Web*

## 1. Einleitung

Das Semantische Web ist keine Erfindung der jüngsten Vergangenheit. Seit vielen Jahren wird bereits, in unterschiedlicher Intensität und mit schwankendem Interesse der medialen Öffentlichkeit, am neuen, „intelligenteren“ World Wide Web gearbeitet. Seit Beginn des Jahrtausends wird diese Vision auch vom einflussreichen W3C<sup>1</sup> vorangetrieben und unterstützt. Während Bibliotheken nur in Ausnahmefällen zu den sog. early adopters zählten<sup>2</sup>, scheint das Thema Semantic Web inzwischen immer mehr in das Blickfeld der bibliothekarischen Fachwelt zu geraten. Gut besuchte Konferenzen wie die vom Hochschulbibliothekszentrum Nordrhein-Westfalen (hbz) mitorganisierte Fachtagung *Semantic Web in Bibliotheken*<sup>3</sup> sowie zahlreiche jüngere Publikationen zu diesem Thema<sup>4</sup> legen davon beredtes Zeugnis ab. Die sog. semantischen Technologien, von denen weiter unten noch zu sprechen sein wird, integrieren dabei zahlreiche, dem Bibliothekar wohl bekannte Konzepte und finden auch in offiziellen Berichten über die Lage des Bibliothekswesens Erwähnung.<sup>5</sup> Linked Data, das Zur-Verfügung-Stellen von bibliographischen Metadaten in einem semantischen Datenformat, stellt den ersten Schritt hin zu einem voll ausgebauten Semantic Web dar, bildet gleichsam dessen Grundlage. Ziel dieses Beitrages ist es, in konziser Form in das Thema Bibliotheken und Linked Data einzuführen. Dafür wird in einem ersten Schritt etwas ausführlicher auf das Semantic Web und dessen grundlegende Technologien eingegangen, um sodann konkreter das Konzept der Library Linked Data herauszuarbeiten und kurz einige Initiativen und Problembereiche vorzustellen. Schlussendlich werden mögliche Anwendungsfälle für den österreichischen Kontext präsentiert.

## 2. Das Semantic Web

Unter dem Begriff **Semantic Web** (SW) verbergen sich eine ganze Reihe von Technologien und Konzepten, die zusammengenommen eine neue Ära in der maschinellen Datenverarbeitung des Internet<sup>6</sup> einläuten könnten. Der Begriff wurde vom Erfinder des **World Wide Web** (Web), Tim Berners-Lee, geprägt und mehrmals ausführlich beschrieben.<sup>7</sup> Zum besseren Verständnis der Möglichkeiten und Funktionsweise des SW sei im Folgenden auf in unserem Kontext relevante, fundamentale Konzepte des derzeitigen, uns bekannten Web eingegangen.

Das Web kennt, ganz wie auch die Welt der gedruckten Schrift, Wege und Mittel, den zu transportierenden Inhalt anschaulich und übersichtlich

zu gestalten. Was der Schrift allgemein die Abstände, Punkte, Beistriche, Strichpunkte, Doppelpunkte oder Rufzeichen, das sind dem Web *h1*, *div* oder *li*. *h1*, *div* und *li* sind Teil der Auszeichnungssprache **HTML** (Hypertext Markup Language), deren hauptsächliche Funktion eine basale Strukturierung des Webseiten-Inhaltes ist. Sie erlaubt u.a. eine Unterteilung des Textes in Überschriften, Absätze und Aufzählungen sowie das Setzen von Links zu bestimmten Textstellen oder anderen Webseiten.<sup>8</sup>

<pre>&lt;h1&gt;Dies ist eine Überschrift&lt;/h1&gt; Unformatierter Text Text Text &lt;ul&gt;   &lt;li&gt;Punkt1&lt;/li&gt;   &lt;li&gt;Punkt2&lt;/li&gt;   &lt;li&gt;Punkt3&lt;/li&gt; &lt;/ul&gt;</pre>	<p><b>Dies ist eine Überschrift</b></p> <p>Unformatierter Text Text Text</p> <ul style="list-style-type: none"> <li>• Punkt1</li> <li>• Punkt2</li> <li>• Punkt3</li> </ul>
--	---

Abbildung 1: Links die HTML-Auszeichnungselemente, rechts der dadurch dargestellte Text.

Der Mensch ist in der Lage, dem Web Sinn abzugewinnen, die Bedeutung des Gelesenen zu erfassen, zu erkennen, dass bspw. die Zeichenfolge „Willard Van Orman Quine“ einen Eigennamen darstellt, der Gegenstand bzw. das Thema einer Abhandlung ist und nicht deren Autor. Weiters können wir diesen Namen zu anderen Dingen oder Sachverhalten in Beziehung setzen wie: Quine war Sprachphilosoph, lernte u.a. bei Rudolf Carnap, las Frege und war Bürger der USA, ein großer Staat in Nordamerika, der traditionell gute Beziehung zu Israel, einem Staat in der östlichen Mittelmeerregion, pflegt etc.

Hinzu kommt die Fähigkeit, logische Schlüsse zu ziehen: Wenn Quine bei Carnap gelernt hat und „bei jemandem lernen“ u.a. bedeutet, dass man zu einem bestimmten Zeitpunkt gleichzeitig lebt, dann haben Quine und Carnap zu einem bestimmten Zeitpunkt gleichzeitig gelebt.

Während wir in den meisten Fällen in der Lage wären, den Sinn des Gelesenen auch ohne derartige Strukturierung zu erfassen, genauso wie man auch Sätze ohne Punktierung und Groß- bzw. Kleinschreibung verstehen kann, ist leicht ersichtlich, dass sich eine Sprache wie HTML kaum für die maschinelle Verarbeitung eignet. Die Verwendung der Elemente ist nicht standardisiert und der Inhalt völlig beliebig. In einer *li*-Liste kann sich von der Aufzählung von Personennamen bis hin zu einer Einkaufsliste alles verbergen. Suchmaschinen wie Google werden gerade deswegen so geschätzt,

weil sie in der Lage sind, im semi-strukturierten Web die Spreu vom Weizen, inhaltstragende von inhaltsleeren Wörtern zu trennen und die Relevanz eines Dokumentes im Verhältnis zu einem Suchbegriff verblüffend gut einzuschätzen. Die komplexen Algorithmen sind nicht von ungefähr gut gehütete Geheimnisse.

Die Idee des SW besteht nun genau darin, das bestehende Web auch für die maschinelle Verarbeitung zu optimieren, die Inhalte zu kategorisieren (hier im Sinn von auszeichnen bzw. annotieren) und zueinander in Beziehung zu setzen – gleichsam die Ausweitung der bibliothekarischen Methode auf das gesamte Web. **Semantik**<sup>9</sup> wird oftmals als die mittlere Stufe in einer Pyramide, bestehend aus Syntax, Semantik und Pragmatik, dargestellt. Während die Syntax Auskunft über die Verkettung und Beziehung der einzelnen Zeichen gibt, ist die Semantik die Lehre von der Bedeutung eines Begriffes. Die Pragmatik schließlich nimmt den Verwendungszusammenhang in den Blick. Man sagt: „Ja, ich will.“ und nicht „Ich ja will.“ (Syntax). „Ja, ich will.“ bedeutet so viel wie „seine Zustimmung geben“ oder „mit etwas einverstanden sein“ (Bedeutung). Und schließlich hat die Aussage „Ja, ich will.“ ganz andere Auswirkungen, je nach dem ob man sie in der Betriebskantine oder in Anwesenheit seiner langjährigen Lebensgefährtin und eines Priesters tätigt (Pragmatik).

Das Web semantisch zu machen, bedeutet zunächst, dass man an Stelle eines *h1* (=Überschrift) eine feinere Granulierung wählt und jedes/n für die Weiterverarbeitung interessierende Wort, Satz, Dokument, Sachverhalt etc. eindeutig und explizit als solche/s/n deklariert:

<pre>&lt;h1&gt;Carnap, Quine und das Problem der Bedeutung&lt;/h1&gt; &lt;h2&gt;von Max Mustermann&lt;/h2&gt;</pre>	<pre>&lt;person&gt;Carnap&lt;/person&gt; &lt;person&gt;Quine&lt;/person&gt; und das Problem der &lt;thema&gt;Bedeutung&lt;/thema&gt; von &lt;autor_vorname&gt;Max&lt;/autor_vorname&gt; &lt;autor_nachname&gt;Mustermann&lt;/autor_nachname&gt;</pre>
---	---

Abbildung 2: Links die Auszeichnungselemente von HTML, rechts die viel detaillierte Version in XML.

Sind die Inhalte erst einmal solcherart ausreichend mit **Metadaten** (Meta-Tags oder auch Annotationen) versehen, ist es Computerprogrammen bzw. Software-Agenten bereits möglich, gezielte Anfragen zu beantworten: Welche Personen kommen im Text vor? Welches Konzept wird thematisiert? Wer ist der Autor des Artikels? Derartige Informationen wären unmöglich aus *h1* oder *h2* zu ermitteln – Tags, die lediglich besagen, dass es sich um eine Überschrift handelt, nicht aber worum es in dieser Überschrift geht.

Wirklich gewinnbringend verwertbar werden die annotierten Inhalte aber erst durch deren Platzierung in einem Wissensraum, in einem Kontext mit anderen Konzepten und Dingen. Sog. **Ontologien** sorgen für die Organisation und Verbindung der einzelnen Termini. Als Fortführung bekannter Konzepte wie Taxonomien (hierarchische Einordnung in Klassen) und Thesauri (relationiertes, genormtes Vokabular) bilden sie das Rückgrat des Semantischen Webs. Sie bestimmen:

- a) worüber überhaupt etwas ausgesagt werden darf: Fische, Atome, Bücher
- b) was darüber gesagt werden darf: etwas ist ein Titel, eine Fußnote, ein Name
- c) wie Dinge in Verbindung stehen: Etwas ist Teil von ..., das Gleiche wie ...
- d) woraus etwas folgt: Wenn C A kennt, dann kennt C auch B.

So wie es Fachterminologien für einzelne Gebiete der Wissenschaft gibt, so gibt es auch Ontologien für Domänen, Subdomänen und Sub-subdomänen des Wissens: für menschliche Beziehungen, Pizza und Bücher, für Kraftfahrzeuge, Krebsforschung und Werkzeuge. Heißt das nun, dass Computer die Bedeutung des Verarbeiteten wirklich verstehen? Natürlich nicht. Sie werden nur mit einer ganzen Menge neuer Algorithmen und mühsam angereicherter Daten gefüttert – in der Hoffnung, dass diese Algorithmen Ergebnisse zeitigen, die nicht vorhersehbar waren. Mit den Worten von Berners-Lee:

„The concept of machine-understandable documents does not imply some magical artificial intelligence allowing machines to comprehend human mumblings. It relies solely on a machine’s ability to solve well-defined problems by performing well-defined operations on well-defined data. So, instead of asking machines to understand people’s language, the new technology, like the old, involves asking people to make some extra effort, in repayment for which they will get substantial new functionality – just as the extra effort of producing HTML markup (Hypertext Markup Language) is outweighed by the benefit of having content searchable on the web.“<sup>10</sup>

### 3. Technologien und Standards

Das SW wird heute im Wesentlichen durch Standards, definiert durch das W3C, getragen und entwickelt. Die wichtigsten sind: **XML**, **RDF**, **RDFS**, **OWL** und **SPARQL**.

Die **Extensible Markup Language** (XML) prägt schon seit einigen Jahren die Infrastruktur des Internets. Es handelt sich bei ihr um eine Art Meta-Metasprache, mit deren Hilfe Auszeichnungssprachen wie z.B. HTML modelliert werden können. In XML lässt sich beispielsweise eine private Auszeichnungssprache für die eigene CD-Sammlung entwerfen.

```
<cd>
  <titel>...</titel>
  <interpret>...</interpret>
  <bewertung>...</bewertung>
</cd>
```

Abbildung 3: XML erlaubt den Entwurf eigener Metadaten-Formate

In seinem Aufbau folgt XML einer Baumstruktur, d.h. durch hierarchische Verschachtelung der einzelnen Elemente lassen sich Abhängigkeiten und Beziehungen darstellen. Man könnte meinen, dass XML für die Anreicherung des Webs mit Metadaten bereits ausreichend ist, bezogen auf die semantische Annotierung hat XML jedoch gewichtige Nachteile: So lässt sich ein und derselbe Sachverhalt auf unterschiedliche Art und Weise in XML codieren, was bei maschineller Verarbeitung zu Interpretationsfehlern führen könnte. Außerdem lässt die hierarchische Struktur nur in sehr begrenztem Maß die Darstellung komplexer Beziehung zu und schließlich ist auch bei XML, genauso wie bei HTML, die Belegung und Bedeutung der einzelnen Metadaten-Felder vom jeweiligen Verfasser abhängig und, wenn überhaupt, nur über Schemata oder eine Dokumentation interpretierbar.

Das **Resource Description Framework** (RDF) soll diese Mängel kompensieren und bildet gleichsam die Grammatik des semantischen Netzes. Es legt in sehr abstrakter Form fest, wie man zunächst einmal Aussagen über den Inhalt oder das Format einer bestimmten Ressource formulieren soll. Eine RDF-Aussage, ein sog. Triple, setzt sich aus Subjekt (über das etwas ausgesagt wird), Prädikat (was über dieses Etwas ausgesagt wird) und Objekt (der Inhalt des Prädikates) zusammen. „Cäsar bewundert Napoleon“ oder „Wien liegt in Österreich“ sind syntaktisch korrekte Triple.

„[...] RDF is an open-world framework that allows anyone to say anything about anything.“<sup>11</sup>

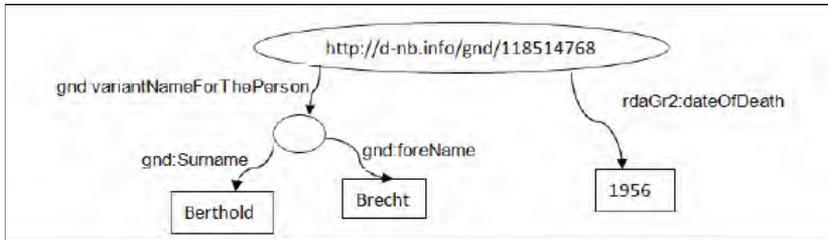


Abbildung 4: RDF-Aussagen grafisch dargestellt. Die Kreise sind Subjekt bzw. Objekt, während die Pfeile als Prädikate fungieren.

Im Gegensatz zu XML bzw. XML-Schemata wird die Relationierung unmittelbar und direkt durch RDF zum Ausdruck gebracht und in keiner Dokumentation über den Zusammenhang der einzelnen Tags versteckt.<sup>12</sup> „Liegt in“ oder „bewundert“ sind gleichsam auf der gleichen Ebene angesiedelt wie Subjekt und Objekt. RDF-Tripel können in unterschiedlichen Formaten wie Turtle, N3 und (verwirrenderweise) auch in XML codiert werden:

```
<rdf:Description rdf:about="http://d-nb.info/gnd/118514768">
  <gnd:variantNameForThePerson rdf:parseType="Resource">
    <gnd:foreName>Berthold</gnd:foreName>
    <gnd:surname>Brecht</gnd:surname>
    <gnd:usedRules>RAK-WB</gnd:usedRules>
  </gnd:variantNameForThePerson>
</rdf:Description>
```

Abbildung 5: Der GND-Normdatensatz zu Bertolt Brecht in RDF-XML. Im Tag Description verbirgt sich die Ressource, über die etwas ausgesagt werden soll. Die darunter geschachtelten Tags wie surname oder foreName stellen Zuschreibungen (Prädikate) dar.

Ein sehr beliebtes RDF-Format im Web ist RDFa<sup>13</sup>. Es ermöglicht, herkömmliche Webseiten mit RDF-Statements anzureichern – beispielsweise eine Adresse auf einer Webseite semantisch auch als diese auszuweisen. Suchmaschinen wie Google können diese Informationen auswerten und so bessere Suchergebnisse produzieren.

Wie Abbildung 5 zu entnehmen, ist ein weiteres Prinzip von RDF die Verwendung von URIs (Uniform Resource Identifier). Wenn möglich sollten zur Bezeichnung der Subjekte, der Objekte, aber auch der Prädikate URIs verwendet werden, worunter man vereinfacht einen Bezeichner bzw. Zeiger verstehen kann. In seiner bekanntesten Form ist ein URI ein URL (Uniform Resource Locator), d.h. eine Webadresse, wie sie im täglichen Gebrauch verwendet wird. Wie ebenfalls gut ersichtlich, handelt es sich konkret meist um einen Link, der im Idealfall zu einem aussagekräftigen, das Subjekt, Prädikat oder Objekt beschreibenden Dokument führt, wobei keinesfalls der Eindruck entstehen sollte, dass mit URIs ausschließlich Webseiten referenziert werden können. Im Vordergrund steht die eindeutige Identifikation. Ob es sich dabei um eine Webpräsenz oder eine natürliche Person handelt, ist im Grunde zweitrangig. Es können auch mehrere URIs angegeben werden, die auf ein und dasselbe Ding verweisen.

Um die Interoperabilität zu gewährleisten, müssen gleichzeitig das in einer RDF-Aussage verwendete Vokabular und die möglichen Beziehungen genormt werden: Flugzeuge haben keine Ehepartner, Ehepartner können aber Flugzeuge haben. Das Wort „babig“ beschreibt keinen Aggregatzustand. Vielmehr sollen die Worte „flüssig“, „fest“ oder „gasförmig“ Verwendung finden. Nur so kann es einem Programm gelingen, Verweisungen zu fremden Daten zu folgen und relevante Informationen zu extrahieren. Die erlaubten Relationen sowie die verwendete Terminologie werden der Maschine über **Schemata** bzw. **Ontologien** mitgeteilt.

**RDF-Schema** (RDFS) stellt eine basale Möglichkeit der Ordnung von Sachverhalten dar. Mit seiner Hilfe können für einen bestimmten Wissensbereich Klassen, Unterklassen, Instanzen sowie Attribute oder erlaubte Datentypen definiert werden. RDF-Schema kann auch wieder in RDF ausgedrückt werden kann – immerhin sagt man ja nur etwas über etwas. Mächtiger als RDFS ist die **Web Ontology Language** (OWL, nicht WOL), die ihrerseits in mehreren Komplexitätsstufen vorliegt. Ontologien stellen im Bereich des SW den höchsten Grad von Formalisierung dar. Sie sind in ihrer Tiefe umfassender als Thesauri und erlauben eine Vielzahl von logischen Operationen auf das durch sie beschriebene Datenset. Eine ausführliche Beschreibung würde an dieser Stelle zu weit führen.

Die reiche und mühevoll semantische Bearbeitung der Webinhalte wäre sinnlos, würde es nicht auch geeignete Abfragesprachen dafür geben. Mit **SPARQL** (SPARQL Protocol and RDF Query Language) verwaltet das W3C eine solche Abfragesprache. In ihrem Aussehen erinnert sie stark an die Datenbankabfragesprache SQL.

```

<owlx:Class owlx:name="Vintage" owlx:complete="false">
  <owlx:ObjectRestriction owlx:property="#vintageOf">
    <owlx:minCardinality owlx:value="1" />
  </owlx:ObjectRestriction>
</owlx:Class>

```

Abbildung 6: Das Beispiel zeigt die Definition einer Restriktion für eine Eigenschaft eines Objekts einer bestimmten Klasse. Die Eigenschaft `vintageOf` muss mindestens einmal vorkommen. OWL kann wie RDF in XML dargestellt werden.

```

SELECT DISTINCT ?uri, ?subject, ?benennung
WHERE {
  ?uri <http://purl.org/dc/terms/subject> ?subject .
  ?subject <http://www.w3.org/2004/02/skos/core#prefLabel> ?benennung .
}
LIMIT 100

```

Abbildung 7: Abfragebeispiel mit SPARQL.

Abbildung 7 zeigt eine SPARQL-Abfrage ausgeführt im Mannheimer SPARQL-Interface<sup>14</sup>. Sie liefert URIs, d.h. Datensätze die eine thematische Angabe (SWD-Schlagwort, RVK-Notation) enthalten. Es wird der URI des Titeldatensatzes, der URI des Themas und deren Benennung ausgegeben. Dieses Beispiel zeigt den Vorteil der Verlinkung via URIs. Die Benennung der RVK-Notation ist nicht in den Titeldaten vorhanden. Sie wird vielmehr direkt aus der RVK-Beschreibung gelesen und beim Titeldatensatz angezeigt.

Würden nun für Autorenangaben in Katalogen auch URIs verwendet werden, wären natürlichsprachliche Anfragen wie: „Welche Bücher wurden in Deutschland in der Zwischenkriegszeit von jüdischen Autoren veröffentlicht, die ab 1938 emigriert sind?“ durchaus möglich.

#### 4. Linked Data

Vom SW muss der Begriff **Linked Data** (LD oder auch LOD für Linked Open Data, wenn keine lizenzrechtlichen Restriktionen mit den Daten verbunden sind) unterschieden werden. Umfasst das SW als ein übergreifendes Konzept eine Vielzahl von Technologien, so bilden die miteinander verknüpften Daten gleichsam die Grundlage für darauf aufbauende seman-

tische Anwendungen. Der Begriff Linked Data subsumiert unterschiedliche Standards und Regeln für das Publizieren semantisch annotierter Daten. Neben der Darstellung der Ressourcen in RDF lauten die vier wichtigsten Regeln:

1. „Use URIs as names for things.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF\*, SPARQL).
4. Include links to other URIs, so that they can discover more things.”<sup>15</sup>

Nach diesen Grundsätzen publizierte bibliographische Metadaten werden als **Library Linked Data** (LLD) bezeichnet. Wie man sich das konkret vorstellen kann, hat das hbz vor kurzem eindrucksvoll gezeigt<sup>16</sup>: 4.8 Mio. Datensätze wurden in 82 Mio. RDF-Triple (RDF-Aussagen) überführt. Als Ontologie, also Vokabular und Relationspool, wurde u.a. die bereits bestehende Ontologie **Bibliographic Ontology**<sup>17</sup> herangezogen. Fremde Quellen bzw. Informationsfragmente werden nicht in den eigenen Datenbestand aufgenommen, sondern lediglich via URI referenziert.

Das Verbinden der Daten sowie deren Befreiung aus den viel zitierten Datensilos, schaffen die Grundlage für komplexe, darauf aufbauende Applikationen und viele weitere Anwendungsfälle, die im Folgenden näher beleuchtet werden sollen. Eine übersichtliche Liste der bereits freigegebenen Daten findet sich im CKAN (Comprehensive Knowledge Archive Network)-Verzeichnis<sup>19</sup>.

## 5. Library Linked Data: Projekte und Erwartungen

Bei der ersten SWIB-Konferenz 2009 ging es noch darum, das Bewusstsein dafür zu schärfen, dass Bibliotheken ihre Daten als Linked Data bereitstellen sollten. Nur ein Jahr später, auf der Folgekonferenz, diskutierte man bereits über konkrete Erfahrungen bei der Bereitstellung von Daten als Linked Data in den ersten Projekten und den entstandenen Diensten.

In Deutschland gibt es dank der Kooperation zwischen der Deutschen Nationalbibliothek (DNB) und dem hbz viele Fortschritte im Bereich der Bereitstellung bibliografischer Daten als Linked Data. Die Normdateien PND und SWD der DNB sind seit dem Frühjahr 2010 zum Download verfügbar.<sup>20</sup> Aufgrund von technischen Problemen wird noch kein Abfrage-Interface für SPARQL angeboten. Einzelne Datensätze können auch über den URI abgerufen werden.<sup>21</sup> Derzeit ist nur die nicht-kommerzielle Nach-



nutzung gestattet. Wie bereits angedeutet, gab es auch entsprechende Bemühungen einiger Bibliotheken<sup>22</sup> im Umfeld des hzb. Diese Daten unterliegen im Gegensatz zur GND keiner Restriktion im Hinblick auf eine Nachnutzung. Ebenso haben Bibliotheken des SWB<sup>23</sup> im Rahmen eines Prototyps Daten als LOD veröffentlicht.

Auch international gibt es zahlreiche Initiativen: LIBRIS, der schwedische Verbundkatalog, war im Jahr 2008 einer der ersten Bibliothekskataloge, der die Daten auch als Linked Open Data anbot. Es folgten die Library of Congress mit dem Bereitgestellten der Library of Congress Subject Headings<sup>24</sup> (LCSH) und der Library of Congress Control Numbers<sup>25</sup> (LCCN), sowie im Laufe des letzten Jahres immer mehr Einrichtungen in europäischen Ländern wie Deutschland, Großbritannien, Frankreich, Ungarn und Tschechien. Eine Übersicht der bereitgestellten Datensets mit bibliografischen Informationen ist im CKAN in den Gruppen *LLD*<sup>26</sup> und *Bibliographic Data*<sup>27</sup> zu finden.

Nicht unerwähnt bleiben soll das Virtual International Internet Authority File<sup>28</sup> (VIAF), das Normdaten unterschiedlicher Nationalbibliotheken zu einer großen, virtuellen Datenbank zusammenführt und auch als Linked Data veröffentlicht. Mindestens ebenso ambitioniert ist das LD-Projekt von Europeana, das sowohl im internen Datenmodell als auch bei der Verlinkung und Nachnutzung auf Linked Data baut.<sup>29</sup>

Diese Projekte sind mit einigem Aufwand an Personal und Infrastruktur verbunden. Es stellt sich die Frage, was sich die jeweiligen Initiatoren konkret davon erhoffen. Schomburg<sup>30</sup> z.B. sieht Verlage und Bibliotheken im Wissenskreislauf oft außen vor, denn bibliografische Daten werden im Internet durch viele Akteure bereitgestellt – z.B. LibraryThing Tags oder Amazon Reviews. Derartige Metadaten sind gefragt, aber Bibliothekskataloge werden oft nur von einem kleinen Prozentsatz der Informationssuchenden genutzt. Das SW bietet die Chance, die eigenen Daten erfolgreich in andere gesellschaftliche Bereiche hineinzutragen, sie einem größeren Nutzerkreis zugänglich zu machen. Weitere Gründe nennen Ostrowski und Pohl<sup>31</sup>:

1. Bibliotheken sind der ideale Partner, um bibliografische Daten ins LOD-Netz zu bringen, diese zu pflegen und so die Vertrauenswürdigkeit und Stabilität zu sichern.
2. Bei der Verbreitung und Produktion wissenschaftlicher Ergebnisse spielen zunehmend die semantische Annotierung und Verlinkung eine wesentliche Rolle.
3. Als Teil des LOD-Netzes kann man die Sichtbarkeit im Netz maximieren und die eigenen Bestände im Web leichter auffindbar machen.

4. LOD stellt die Nachnutzbarkeit von bibliografischen Daten durch Personen oder Programme außerhalb des Bibliothekswesens sicher.
5. Auch für Anreicherungs-zwecke der Bibliotheks- und Verbundkataloge ist diese Technologie bestens geeignet.
6. Und, last but not least, könnten neue und bessere Rechercheinter-faces entstehen.

Aus technischer Sicht bieten die Linked Data-Standards, verglichen mit der Vielzahl anderer Technologien für den Austausch bibliographischer Daten, den großen Vorteil einer einheitlichen Schnittstelle. Deren Verwendung erlaubt die einfache, niedrighschwellige Nachnutzung in z.B. Mashups (s.u.). Da es sich bei LD um eine Sammlung offener Standards handelt, die auch außerhalb der Bibliothekswelt anerkannt und weit verbreitet sind, könnten fruchtbringende Kooperationen mit Experten entstehen, die nicht unmittelbar in der Domäne der bibliografischen Formate beheimatet sind.

## 6. Herausforderungen

Wie nicht anders zu erwarten, bringen neue Chancen neue Herausforderungen mit sich, die sowohl rechtliche, wie auch technische Fragestellungen umfassen.

### *Lizenzen*

Bereits kurz angeklungen sind die unterschiedlichen rechtlichen Formen der Bereitstellung bibliographischer Metadaten. Dem semantischen Netz als Ganzem liegt die Vision von miteinander im Austausch stehenden maschinellen Agenten zu Grunde, die aus bereits bestehender Information neue Verknüpfungen ableiten bzw. sie in neuer Form zusammenführen und präsentieren. Diese Bemühungen werden durch das Überstülpen rechtlicher Restriktionen konterkariert, wenn nicht zerstört. Daher finden hauptsächlich die Creative Commons Zero<sup>32</sup> (CC0) und Open Data Commons-Lizenzen Verwendung. Das hzb hat sich für die CC0-Lizenz entschieden (jede Form der Nachnutzung erlaubt, auch kommerzielle). Bedenken, dass die Daten einen Wertverlust erleiden, werden nicht geteilt. Vielmehr gilt sinngemäß: „Bibliografische Daten, die nicht frei zugänglich sind, sind nicht relevant“<sup>33</sup>.

## ***Dezentralität der Daten***

Die per definitionem dezentrale Organisation des LD-Web bringt architektonische Probleme mit sich. So besteht z.B. durch das massenhafte Publizieren von Metadaten zu ein und demselben Objekt, sei es ein Buch oder eine Person, die Gefahr, dass jeder Anbieter seine eigenen URIs definiert, die dann erst wieder nicht miteinander in Beziehung gesetzt werden. Das Verlinken zu anderen Daten der LOD-Cloud hat sich als nicht-triviales Problem herausgestellt. Mapping-Dienste wie sameas.org oder, für den Bereich bibliographischer Metadaten, CultureGraph.org, sollen dem entgegen wirken.

Aus technischer Sicht ist eine Abfrage mit SPARQL bei weitem noch nicht so performant wie entsprechende Abfragen in relationalen Datenbanken mittels SQL. Experten gehen gar davon aus, dass relationale Datenbanken 5 bis 50 Mal schneller sind als entsprechende RDF-Stores.<sup>34</sup> Abfragen über gleich mehrere RDF-Quellen sind dementsprechend aussichtslos. Unter anderem aus diesem Grund wird derzeit keine Abfragemöglichkeit für die gesamte GND angeboten.

## ***Zuverlässigkeit und Datenqualität***

Analog zum „alten“ Web sind die Themen Zuverlässigkeit und Aktualität der Daten auch im SW ein wichtiger Aspekt. Im Fall von SPARQL-Interfaces sind diese Schnittstellen oft langsam oder schlicht nicht erreichbar. Das macht es derzeit unmöglich, diese Dienste in Applikationen erst auf Anforderung des Nutzers zu laden. Daher werden Daten oft schon vorab abgefragt und im lokalen Cache gehalten. Um das Problem zu umgehen, werden die als LD bereitgestellten Daten meist (auch) als Gesamt-Export bereitgestellt. Solche recht umfangreichen Export-Dateien können in der eigenen Datenbank gespeichert und für Abfragen optimiert werden. Neben der eigentlich unerwünschten redundanten Datenhaltung bringt das den Nachteil der Nicht-Aktualität der Daten mit sich.

Wenn Datensets zusätzlich eine Beschreibung ihrer selbst anbieten, können potenzielle Nutzer leichter entscheiden, ob dieses Datenset ihren Anforderungen entspricht. Viele Datensets stellen mittlerweile derartige Metadaten via void<sup>35</sup> bereit.

## ***Inkompatibilität der Ontologien***

Zur Beschreibung der bibliografischen Metadaten stehen mehrere Vokabularien zur Verfügung: Dublin Core, Bibliographic Ontology, RDA Ontology,

um nur einige zu nennen. Um Abfragen effektiv über die Grenzen von Datensets hinaus durchführen zu können, müssen diese teilweise stark divergierenden Ontologien miteinander verknüpft werden. Wird die Ontologie zu komplex, gefährdet das die Nachnutzung durch andere Domänen des Web of Data. Wird sie zu einfach, geht vorhandene Information verloren. Während das hzb hauptsächlich auf die Bibliographic Ontology setzt<sup>36</sup>, verwendet die Ungarische Nationalbibliothek Dublin Core und die Deutsche Nationalbibliothek sah sich gar gezwungen, für die Darstellung einzelner Teile der Normdatensätze der GND eine eigene Ontologie zu entwerfen (siehe die gnd-Tags in Abbildung 5). Bis jetzt ist nicht absehbar, welche Form der Repräsentation sich im bibliothekarischen Bereich durchsetzen wird.

## 7. LLD-Anwendungen für Österreich

Die Library Linked Data Incubator Group<sup>37</sup> (LLD XG) ist eine Arbeitsgruppe, die vom W3C und einigen namhaften Bibliotheken und Herstellerfirmen von Bibliothekssystemen ins Leben gerufen wurde. Diese Gruppe hat, mit dem Ziel der Identifikation „Killer-Applikation“<sup>38</sup> für das Semantische Web im Bereich von Bibliotheken, eine Reihe von Anwendungsfällen (use cases) gesammelt, die in folgende Bereiche gegliedert wurden: Bibliographic Data, Authority Data, Vocabulary Alignment, Archives and heterogeneous data, Citations, Digital objects, und Collections.

Einige der von dieser Gruppe angedachten Vorschläge wären technisch bereits machbar bzw. mit verhältnismäßig wenig Aufwand zu realisieren. Im Folgenden seien skizzenhaft ausgewählte Anwendungsbeispiele umrissen:

### *Kataloganreicherung*

Der eher bescheidene Umfang der im österreichischen Verbundkatalog klassifizierten Datensätze wurde bereits mehrfach beklagt<sup>39</sup> und führt in facettierten Suchmaschinen, wie dem kürzlich eingeführten Retrieval-System Primo, zu eher lückenhafter Facettenbildung. Facetten, die lediglich einige wenige Prozent des gesamten Angebotes umfassen, können das Auffinden von relevanten Dokumenten sogar erschweren, anstatt es zu erleichtern. LLD könnte in diesem Zusammenhang neue Möglichkeiten der redundanzfreien, zeitnahen Kataloganreicherung im Bereich der Klassifikationssysteme (RVK, BK, DDC) bieten. Das Gleiche gilt für andere Formen der maschinellen wie intellektuellen Indexierung. Das Spektrum reicht von der Einbindung Benutzer-erzeugter Daten wie Tags oder Reviews bis hin zur

automatischen semantischen Annotierung von unstrukturierten Volltextdokumenten à la OpenCalais<sup>40</sup>.

### **Mashups**

Als Mashup wird eine Neuordnung bestehender Inhalte bezeichnet. Vergleichbar mit einem Aggregator-Dienst, werden in einem Mashup unterschiedliche vorliegende Quellen in neuer Form aufbereitet und zusammengeführt. Ist diese Form der Re-Kombination im sog. Web 2.0 schon lange Usus, hinken Bibliotheken bzw. deren Kataloge diesbezüglich etwas hinterher und weisen zumeist nur Dokumente aus der eigenen Datenbank nach. Die Einbindung anderer Linked Data Sets könnte das eigene Angebot um einiges attraktiver machen. So wurden z.B. in der neuen Suchoberfläche des Österreichischen Bibliothekenverbundes<sup>41</sup> die in Kategorie 100f., Subfeld 9 katalogisierte PND des/r Autors/sonstigen beteiligten Person verwendet, um gefundene Treffer mit zusätzlichen Informationen attraktiver zu gestalten und gleichsam in einen Kontext zu betten.



Abbildung 9: Angaben aus der DBPedia in der Verbundsuchmaschine.

Dieser Dienst basiert auf dem Projekt DBPedia<sup>42</sup>, der semantischen Version von Wikipedia. Auf die gleiche Art und Weise könnten vielerlei andere Daten integriert werden: Mehr Details zu Normdatei-Einträgen durch die RDF-Version der GND, Reviews und Tags aus offenen bibliographischen Systemen oder aber Einbindung geographischer Informationen zum Standort oder Inhalt eines Buches.

### **Bessere Recherche-Interfaces**

Wenn die Daten in RDF vorliegen, können mit SPARQL relativ komplexe Anfragen formuliert werden. Das ist vergleichbar mit einer Expertensuche in klassischen Bibliothekskatalogen und geht noch weit darüber hinaus.

Durch die komplexe Verschachtelung der Daten mit Hilfe von Ontologien können beliebige Sachverhalte in Beziehung gesetzt werden. An die Stelle von simplen Stichwortsuchen und dem Vertrauen auf gute Rankingalgorithmen würden neue Interfaces treten, die in der Lage wären, natürlichsprachliche Anfragen (s.o.) zu beantworten. Die Universität Mannheim stellt derzeit eine Applikation zum Testen derartiger Abfragen bereit und bietet Teile der Daten der DNB, des hzb, des SWB und des HEBIS in einer RDF-Datenbank zur Recherche an.<sup>43</sup> Visualisierungen basierend auf via LOD bereitgestellten Normdaten sind leicht zu realisieren. Suchbegriffe könnten in einem semantischen Kontext präsentiert und visualisiert werden. Inspirationen gibt es z.B. bei SWOC<sup>44</sup>, Timeline<sup>45</sup> oder Tag-Cloud-Darstellungen aller Art<sup>46</sup>. Die semantischen Rechercheinterfaces könnten den Wechsel von einem Nachweis-orientierten, Dokument-zentrierten Katalog, zu einem Informations-zentrierten Katalog, einem Portal des Wissens und der Vernetzung ermöglichen.<sup>47</sup>

### ***Institutionen-Verzeichnis***

Das Institutionen-Verzeichnis der Österreichischen Informationseinrichtungen<sup>48</sup> enthält Angaben über österreichische Einrichtungen im BID-Bereich (Adressen, ISIL, Sigel etc.). Die Aktualität der dortigen Angaben hält sich leider in Grenzen. Mit Hilfe semantischer Technologien wäre es möglich, die Angaben dort abzuholen, wo sie mit hoher Wahrscheinlichkeit immer aktuell gehalten werden: den Homepages der jeweiligen Einrichtungen. Viele CMS-Systeme unterstützen bereits jetzt RDFa. Öffnungszeiten, Adressen o.ä. könnten in RDFa kodiert und in regelmäßigen Abständen abgefragt werden. Die verteilte Datenbank des Web of Data würde redundante Datenhaltung und die aufwändige Pflege von Angaben an mehreren Orten obsolet machen. Immer aktuell, wäre das Institutionen-Verzeichnis somit wieder interessant für die Verwertung in oben genannten Recherche-Interfaces wie der Verbundsuchmaschine.

## **8. Fazit**

Der geschulte Bibliothekar, die geschulte Bibliothekarin mag sich von einigen vorgestellten Konzepten, ja überhaupt von der Idee, das Web mit semantischen Metadaten zu versehen, wenig beeindruckt zeigen. Immerhin sind entsprechende Techniken des Wissensmanagements seit langem bekannt.<sup>49</sup> Das Semantic Web – alter Wein in neuen Schläuchen? Nicht ganz. Schon

alleine der Einsatz von weltweit anerkannten Repräsentationsformaten und somit die Öffnung hin zur ganzen Gemeinschaft des Web stellt einen qualitativen Sprung dar.<sup>50</sup> In Zukunft werden nicht nur Bibliotheksserver Titeldaten austauschen. Der Katalog der Zukunft könnte mit unterschiedlichsten Diensten des Web in Verbindung stehen: LeserInnen-Communities, Online-Buchhandel, Lexika, Statistik-Datenbanken oder öffentliche Ämter.

Die vorgestellten Projekte und Ausblicke stellen nur einen kleinen Ausschnitt denkbarer Szenarien und Nutzungsmöglichkeiten von Library Linked Data dar. Sie sind bewusst eher konkret als visionär gehalten, um den Aspekt der Machbarkeit in den Vordergrund zu stellen. Das Semantische Web ist, trotz der zahlreichen offenen Fragen und technologischen Schwächen, in Teilen bereits Realität.

Dennoch wartet der Großteil der inzwischen zahlreich bereitgestellten Daten auf seine effektive und innovative Nutzung. LD-Browser wie Tabulator<sup>51</sup> oder Marbles<sup>52</sup> haben sich nicht wirklich durchgesetzt. Um via LD bereitgestellte Daten in Applikationen zu nutzen, werden diese oft auf eigene Server kopiert und zwischengespeichert, weil die Methoden für eine just-in-time-Nutzung noch nicht ausreichend entwickelt sind. Semantische Suchmaschinen könnten diese Lücke füllen.

Die einfache und für jedermann nachvollziehbare semantische Auszeichnung bestehender oder neu erstellter Inhalte wird für das Semantische Web erfolgskritisch sein. Zahlreiche Programme unterstützen bereits jetzt die Überführung unstrukturierter Daten in RDF-Triple und manche Content Management Systeme beherrschen die semantische Annotierung auf Knopfdruck. Trotzdem wird der Großteil der heute online publizierten Information nach wie vor für den menschlichen Leser produziert. Auch die Bibliothekswelt ist in hohem Maß von nicht-semantischen Technologien und Softwareprodukten abhängig. Ohne massiven Druck von Seiten der Anwender wird sich daran wenig ändern.

Mit Linked Data bzw. dem Semantic Web verbinden sich viele Hoffnungen auf eine umfassende Erneuerung der bibliographischen Datenerfassung sowie auf eine zukunftsorientierte Art und Weise der Informationsvermittlung in einem Umfeld, das in immer größerem Maß vom Internet und einer daran gewöhnten Klientel abhängt. Der deutsche Wissenschaftsrat sieht die Bibliothekswelt gar an einem Scheideweg zwischen einer globalen, zentralen Datenbank wie WorldCat und einer verteilten, offenen Datenbank, wie sie das Web der Linked Data darstellt.<sup>53</sup> Man kann diese Einschätzung teilen oder nicht, das Semantic Web bleibt in jedem Fall eine spannende und vielleicht auch bedrohliche Herausforderung für die Zukunft der Informationsvermittler.

Ulrike Krabo, BSc  
Die Österreichische Bibliothekenverbund  
und Service GmbH (OBVSG)  
E-Mail: [ulrike.krabo@obvsg.at](mailto:ulrike.krabo@obvsg.at)

Mag. Markus Knitel  
Die Österreichische Bibliothekenverbund  
und Service GmbH (OBVSG)  
E-Mail: [markus.knitel@obvsg.at](mailto:markus.knitel@obvsg.at)

- \* Wir danken Otto Oberhauser für die kritische Durchsicht und wertvollen Anmerkungen.
- 1 Beim W3C handelt es sich um eine Standardisierungsorganisation, die Spezifikationen und Standards rund um das World Wide Web veröffentlicht. Die Standards werden in Zusammenarbeit der offiziellen Mitglieder und der interessierten Öffentlichkeit erarbeitet. Mehr unter: <http://www.w3.org/> [aufgerufen am 16.02.2011].
  - 2 Siehe exemplarisch das Wachstum der einzelnen Sektoren im Semantic Web unter <http://richard.cyganiak.de/2007/10/lod/> [aufgerufen am 18.02.2011].
  - 3 <http://swib.org> [aufgerufen am 16.02.2011].
  - 4 Vgl. Pohl, A. und Ostrowski, F. (2010). Linked Data – und warum wir uns im hbz-Verbund damit beschäftigen!. B.I.T. online 3/2010, 259–268.
  - 5 Vgl. Wissenschaftsrat. Empfehlungen zur Zukunft des bibliothekarischen Verbundsystems in Deutschland. Berlin, S. 10. Online: <http://www.wissenschaftsrat.de/download/archiv/10463-11.pdf> [aufgerufen am 16.02.2011].
  - 6 Die folgenden Ausführungen beziehen sich großteils auf das Web der Textdateien, weniger auf das der audiovisuellen Medien.
  - 7 Vgl. Berners-Lee, Tim (2001). The Semantic Web. Scientific American 05/2001, S. 35–43.
  - 8 Diese Darstellung von HTML ist bewusst verkürzend um den Unterschied zum Semantic Web besser akzentuieren zu können.
  - 9 Wie immer eignet sich Wikipedia exzellent für einen schnellen Einstieg in die Materie: [http://de.wikipedia.org/wiki/Bedeutung\\_%28Sprachphilosophie%29](http://de.wikipedia.org/wiki/Bedeutung_%28Sprachphilosophie%29) [aufgerufen am 16.02.2011]. Für Freunde gedruckter Werke vgl. Zoglauer, Thomas (2008). Einführung in die formale Logik für Philosophen. Göttingen: UTB.
  - 10 Berners-Lee, Tim und Hendler, James. Scientific publishing on the Se-

- mantic Web. Online: <http://www.nature.com/nature/debates/e-access/Articles/bernerslee.htm> [aufgerufen am 16.02.2011].
- 11 <http://www.w3.org/TR/2002/WD-rdf-concepts-20020829/#xtocid48014> [aufgerufen am 18.02.2011].
  - 12 Mehr zum Unterschied zwischen XML und RDF vgl.: <http://www.w3.org/DesignIssues/RDF-XML.html> [aufgerufen am 16.02.2011].
  - 13 <http://www.w3.org/TR/xhtml-rdfa-primer/> [aufgerufen am 17.02.2011].
  - 14 <http://data.bib.uni-mannheim.de/sparql> [aufgerufen am 22.02.2011].
  - 15 <http://www.w3.org/DesignIssues/LinkedData.html> [aufgerufen am 16.02.2011].
  - 16 <http://lobid.org/de/resource.html> [aufgerufen am 16.02.2011].
  - 17 <http://bibliontology.com/> [aufgerufen am 21.02.2011].
  - 18 <http://richard.cyganiak.de/2007/10/lod/> [aufgerufen am 14.02.2011].  
Ein Vergleich mit früheren Versionen der Wolke offenbart das rasante Wachstum der publizierten Daten innerhalb der letzten Jahre.
  - 19 <http://ckan.net/> [aufgerufen am 16.02.2011].
  - 20 <https://wiki.d-nb.de/display/LDS/> [aufgerufen am 16.02.2011].
  - 21 Siehe Endnote 13.
  - 22 Das sind u.a.: Universitäts- und Stadtbibliothek Köln, Hochschulbibliothek der Fachhochschule Köln, Stadtbibliothek Köln, Bibliothek/Mediathek der Kunsthochschule für Medien Köln, Landesbibliothekszentrums Rheinland-Pfalz und die Hochschulbibliothek RWTH Aachen.
  - 23 Die Universitätsbibliotheken Konstanz und Tübingen.
  - 24 <http://lcssubjects.org/> [aufgerufen am 17.02.2011].
  - 25 <http://lccn.lcssubjects.org/> [aufgerufen am 17.02.2011].
  - 26 <http://ckan.net/group/lld> [aufgerufen am 17.02.2011].
  - 27 <http://ckan.net/group/bibliographic> [aufgerufen am 17.02.2011].
  - 28 <http://viaf.org> [aufgerufen am 17.02.2011].
  - 29 <http://version1.europeana.eu/web/guest/news/-/blogs/open-linked-data-and-europeana>.
  - 30 Vgl. im Folgenden den Vortrag auf der SWIB 2010: <http://dx.doi.org/10.4016/27098.01> [aufgerufen am 22.02.2011].
  - 31 Siehe Endnote 4.
  - 32 <http://creativecommons.org/about/cc0> [aufgerufen am 22.02.2011].
  - 33 Vgl. die stv. Direktorin des hzbz Silke Schomburg und ihr Vortrag auf der SWIB 2010. Video: DOI: 10.4016/27098.01, Präsentationsfolien: [http://swib.org/swib10/vortraege/swib10\\_schomburg.PPS](http://swib.org/swib10/vortraege/swib10_schomburg.PPS) [beide aufgerufen am 17.02.2011].
  - 34 Als RDF- oder auch Triple-Store wird eine Datenbank für RDF-Daten-

- sätze bezeichnet. Zu den Performanzproblemen vgl. Sören Auer und sein Vortrag auf der SWIB 2010. DOI: 10.4016/27091.01 [aufgerufen am 18.02.2011].
- 35 void ist eine Beschreibungssprache für LD-Datensets. Vgl.: <http://semanticweb.org/wiki/Void> [aufgerufen 17.02.2011].
- 36 <https://wiki1.hbz-nrw.de/display/SEM/2010/08/31/Linked-Open-hbz-Data> [aufgerufen am 22.02.2011].
- 37 <http://www.w3.org/2005/Incubator/ld/>.
- 38 Als KillerApp wird eine Anwendung bezeichnet, die einer bestimmten Technologie zum Durchbruch verhilft. Diesbezüglich meint Berners-Lee: „The Semantic Web is the killer app.“ (a.a.O. S. 42).
- 39 Vgl. Oberhauser, Otto (2008). Sachliche Erschließung im Österreichischen Verbundkatalog: Status und Perspektiven. Mitteilungen der Vereinigung Österreichischer Bibliothekare & Bibliothekarinnen 3/2008, S. 59–77.
- 40 <http://www.opencalais.com/> [aufgerufen am 17.02.2011].
- 41 <http://search.obvsg.at/OBV> [aufgerufen am 17.02.2011]
- 42 <http://dbpedia.org> [aufgerufen am 17.02.2011].
- 43 <http://data.bib.uni-mannheim.de/> [aufgerufen am 17.02.2011].
- 44 <http://sisinflab.poliba.it/publications/2010/Mir10/gii-2010.pdf> [aufgerufen 17.02.2011].
- 45 <http://www.simile-widgets.org/timeline/> [aufgerufen 17.02.2011].
- 46 Z.B. <http://sisinflab.poliba.it/semantic-wonder-cloud/index/> [aufgerufen 17.02.2011]
- 47 [http://swib.org/swib10/vortraege/swib10\\_coyle.ppt](http://swib.org/swib10/vortraege/swib10_coyle.ppt) [aufgerufen am 17.02.2011].
- 48 <http://opac.obvsg.at/acc09> [aufgerufen am 17.02.2011].
- 49 Schon alleine deshalb kann die Frage „Library and Linked Data, the perfect match?“ (so formuliert von Antoine Isaac, einem Mitglied der LLD XG des W3C) nur rhetorisch gemeint sein. Vgl.: [http://swib.org/swib10/vortraege/swib10\\_isaac.ppt](http://swib.org/swib10/vortraege/swib10_isaac.ppt) [aufgerufen am 17.02.2011].
- 50 Die deutsche Zentralbibliothek für Wirtschaftswissenschaften (ZBW) hat z.B. vor kurzem den Standard-Thesaurus Wirtschaftswissenschaften (STW) im LD-kompatiblen SKOS-Format herausgegeben: <http://www.w3.org/2001/sw/sweo/public/UseCases/ZBW/> [aufgerufen am 17.02.2011].
- 51 <http://www.w3.org/2005/ajar/tab> [aufgerufen 17.02.2011].
- 52 <http://marbles.sourceforge.net/> [aufgerufen 17.02.2011].
- 53 Wissenschaftsrat a.a.O., S. 32.