

Linked Data und die Bibliothekswelt

Adrian Pohl (Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen)

<pohl@hbz-nrw.de>

Hinweis Dieser Text ist eine aktualisierte und überarbeitete Fassung des Beitrags „‘Linked Data’ – und warum wir uns im hbz-Verbund damit beschäftigen“, den ich gemeinsam mit Felix Ostrowski verfasst habe.¹

Abstract

Der Themenkomplex „Linked Open Data“ findet zunehmend Einzug in die bibliothekarische Sphäre. Während „Open Data“ sich auf die rechtliche Situation von Daten aller Couleur bezieht, ist es das Ziel von „Linked Data“, die Interoperabilität von Daten durch einheitliche (Web-)Standards zu verbessern.

In diesem Beitrag wird zunächst erläutert, warum „Linked Data“ auch für die Bibliothekswelt relevant ist. Daraufhin werden in einer knappen Einführung die wichtigsten Linked Data-Standards vorgestellt sowie einige Gedanken zur Migration von bibliographischen Daten in eben diese Standards skizziert. Zuletzt werden die Linked-Data-Aktivitäten des hbz vorgestellt, etwa die experimentelle Plattform *lobid.org* und die beiden dort laufenden Linked(-Open)-Data-Dienste *lobid-resources* und *lobid-organisations*.

¹ Pohl, Adrian / Ostrowski, Felix: ‚Linked Data‘ - und warum wir uns im hbz-Verbund damit beschäftigen. B.I.T. Online 13, 3, S. 259-268. Preprint online einsehbar u.a. unter http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/pohl_ostrowski_2010_linked-data.pdf.

1 Einleitung

Mit dem Internet und – als dessen wichtigstem Bestandteil – dem World Wide Web formt sich seit einigen Jahrzehnten eine umfassende Publikations- und Kommunikationsplattform aus, auf der zukünftig der Großteil aller Publikation und Kommunikation stattfinden wird. Als eine Erweiterung des bestehenden Webs unter Berücksichtigung der urheberrechtlichen Implikationen lässt sich Linked Open Data verstehen. Mit Linked Open Data werden zwei Standards bezeichnet, die die Funktionalität eines Netzes von Daten sichern sollen, indem sie die rechtliche und technische Kompatibilität von Daten im Web garantieren:

- *Open-Data-Standards* sorgen für die *rechtliche* Basis der Nutzung und Kombination verteilter Daten im Netz.
- *Linked-Data-Standards* sorgen für die *technische* Kompatibilität zwischen verteilt vorliegenden Daten.

In diesem Text geht es in erster Linie um die technische Seite von Linked Open Data, also um *Linked Data*. Für eine Einführung in die rechtliche Dimension und damit in *Open Data* siehe den Artikel *Open Data im hbz-Verbund*² und als Orientierung bei einer Freigabe bibliographischer Daten die *Principles on Open Bibliographic Data*³.

2 Semantic Web und Linked Data

Was steckt aber nun konkret hinter ‚Linked Data‘?⁴ Warum beschäftigen sich Bibliothekarinnen und Bibliothekare, Bibliotheken und Verbünde damit? Linked Data ist der erste Schritt zum Semantic Web. Die Entwicklung des Semantic Web ist ein langfristiges Projekt, vorangetrieben vom Erfinder des World Wide Web Tim Berners-Lee⁵ und dem W3C (World Wide Web Consortium). Während ein großer Teil des World Wide Web aus Texten besteht, die sich an *Menschen* richten, ist das Semantic Web der Versuch, Informationen so im Web bereitzustellen, dass sie von *Maschinen* leicht verarbeitet werden können, um nützliche Dinge zu tun. Mit anderen Worten: Das Semantic Web ist die Erweiterung des World Wide Webs als eines Netzes von Dokumenten zu einem Netz von Daten, von maschinenlesbaren, einzelnen Aussagen. Durch maschinelles logisches Schlussfolgern soll aus diesen Daten zusätzliches Wissen generiert werden können.

² Pohl, Adrian (2010): Open Data im hbz-Verbund, ProLibris 3/2010, S.109-113. Preprint online einsehbar u.a. unter http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/pohl_2010_open-data.pdf.

³ <http://openbiblio.net/principles/>

⁴ Verwiesen sei hier zur Beantwortung dieser Frage auch auf die sehr gute, knappe einführende Übersicht mit weiterführenden Links Campbell, Lorna M. / MacNeill, Sheila (2010): The Semantic Web, Linked and Open Data. A Briefing Paper. Einsehbar unter http://wiki.cetis.ac.uk/images/1/1a/The_Semantic_Web.pdf.

⁵ Eine der ersten breit rezipierten Publikationen zum Semantic Web ist Berners-Lee, Tim / Hender, James / Lassila, Ora (2001): The Semantic Web, Scientific American 284, no. 5: 34. Frei zugänglich unter <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>.

Auf den Seiten des W3C findet sich folgende Definition des Semantic Web:

*The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF).*⁶

Die Vision des Semantic Web könnte so formuliert werden: Im Web werden immer mehr veröffentlichte Daten in RDF (s.u.) vorliegen, d.h. sie folgen denselben Standards und können daher frei miteinander kombiniert und verlinkt werden. So entsteht das Semantic Web als eine verteilte Datenbank, die komplexe Abfragen und Schlussfolgerungen über dezentral vorliegende Daten ermöglicht.

Linked Data ist der erste Schritt und das Herzstück des Semantic Webs. Auf den Webseiten des W3C heißt es:

*[T]o make the Web of Data a reality, it is important to have the huge amount of data on the Web available in a standard format, reachable and manageable by Semantic Web tools. Furthermore, not only does the Semantic Web need access to data, but relationships among data should be made available, too, to create a Web of Data (as opposed to a mere collection of datasets). This collection of interrelated datasets on the Web can also be referred to as Linked Data.*⁷

Linked Data legt also das Augenmerk auf die ersten Schritte, auf das Veröffentlichen und Verlinken von Daten, während im Semantic Web Anwendungen, die auf maschinell durchführbaren logischen Schlussfolgerungen basieren, die entscheidende Rolle spielen. Es ist aber zu konstatieren, dass im gegenwärtigen Sprachgebrauch ‚Semantic Web‘ und ‚Linked Data‘ häufig synonym verwendet werden.

3 Warum Linked Library Data?

Warum sollen Bibliotheken ihre Daten als Linked Data publizieren? Im Prinzip verlinken Bibliothekare Daten spätestens seit der Etablierung des Kartenkatalogs, indem sie etwa Verweise von Schlagwörtern und Personen auf Titeldaten dokumentieren. Bibliotheken produzieren also schon lange verlinkte Daten. Außerdem sind sie vertrauenswürdige, öffentliche Institutionen, die tief in unserer Kultur verwurzelt sind und Vertrauenswürdigkeit ist eine wichtige Eigenschaft, die Linked Open Data haben sollte. Bibliotheken sind also geradezu prädestiniert dazu, das Linked-Data-Netz mit aufzubauen, zu pflegen und seine Vertrauenswürdigkeit und Stabilität zu sichern.

Ein weiterer Grund, Bibliotheksdaten in das Semantic Web zu migrieren, sind die stattfindenden Veränderungen in der Produktion, Publikation und Verbreitung wissenschaftlicher Ergebnisse, seien dies Daten oder Texte. Semantisches Markup und Verlinkung werden eine wichtige Rolle in der Weiterentwicklung der Wissensproduktion spielen und damit großen Einfluss auch auf Wissenschaftliche

⁶ <http://www.w3.org/2001/sw/>

⁷ <http://www.w3.org/standards/semanticweb/data>

Bibliotheken haben.⁸ Dies macht deutlich, dass Semantic-Web-Technologien in Zukunft ohnehin eine große Rolle im Bibliothekswesen spielen werden.

Darüber hinaus gibt es eine Menge weiterer Gründe, Linked Library Data zu produzieren, von denen hier einige genannt seien:

- Um endlich im Web anzukommen⁹
- Um die eigene Sichtbarkeit im Web zu maximieren und die eigenen Bestände leichter auffindbar zu machen.
- Um die Nachnutzbarkeit bibliothekarischer Daten durch andere – insbesondere nichtbibliothekarische Akteure – sicherzustellen
- Um die Informationen aus Bibliotheks- und Verbunddatenbanken durch die Verknüpfung mit anderen Informationen anzureichern
- Um die Recherchemöglichkeiten über die Daten zu optimieren

Illustriert sei das hier Geschilderte an einem Beispiel: Eine interessierte Kölnerin hat von Nelson Goodmans Symboltheorie gehört und möchte mehr darüber erfahren. Sie gibt eine entsprechende Suche bei Google ein und landet beim Wikipedia-Eintrag zu Nelson Goodman (siehe Abbildung 1).

⁸ Für einen Einblick in Experimente und Entwicklungen, die im Bereich des sogenannten *Semantic Publishing* stattfinden, vgl. Shotton, David (2009): *Semantic publishing: the coming revolution in scientific journal publishing*, *Learned Publishing* 22 (April 2009): 85-94. DOI: 10.1087/2009202.

⁹ Die meisten Bibliothekskataloge genügen bis heute nicht den gängigsten Web-Konventionen: Einträge sind nicht verlinkbar und die Inhalte können nicht in Suchmaschinen indexiert werden. Vgl. dazu Blogbeitrag und Kommentare: Hauschke, Christian (2010): *Permalinks für Katalogisate*, einsehbar unter <http://infobib.de/blog/2009/10/27/permalinks-fur-katalogisate/>.

Nelson Goodman

Nelson Goodman (* 7. August 1906 in Somerville, Massachusetts; † 25. November 1998 in Needham, Massachusetts) war ein amerikanischer Philosoph.

Goodman war Schüler Alfred North Whiteheads. Nach seiner Graduation 1928 in Harvard leitete er von 1929 bis 1940 eine Kunstgalerie in Boston. 1941 erhielt er seinen Ph.D. und diente danach bis 1945 in der US-Infanterie.

Von 1946 bis 1964 lehrte er an der University of Pennsylvania, wo er 1951 eine Professur erhielt. Ab 1964 lehrte er zusätzlich an der Brandeis University und Tufts University. 1968 erhielt er eine Professur in Harvard.

Werk [Bearbeiten]

Er war stark vom Empirismus und von Rudolf Carnaps Phänomenalismus beeinflusst. Bekannt wurde er vor allem durch sein *Neues Rätsel der Induktion* und den Entwurf einer Symboltheorie zu notationalen und nicht-notationalen Zeichensystemen. Goodman vertritt philosophisch Positionen des relativistischen Pluralismus.

In der europäischen Philosophie weitgehend unbeachtet geblieben, spielte Goodman in der amerikanischen Analytischen Philosophie eine wichtige Rolle. Seiner streng extensionalen Symboltheorie gelingt es metaphysische Konstruktionen der Objekt-Zeichenverknüpfung, die Charles S. Peirce für seine Symboltheorie noch benötigte, zu vermeiden.

In den "Sprachen der Kunst" legt Goodman dar, dass der Unterschied zwischen einem Bild und einer Beschreibung eines Gegenstandes nicht darin besteht, dass das Bild dem Gegenstand ähnlicher ist als seine Beschreibung. Er verdeutlicht, dass Ähnlichkeit weder hinreichende noch notwendige Bedingung für Repräsentation ist, da Ähnlichkeit eine symmetrische Beziehung ist, während dies bei Repräsentation nicht der Fall ist. Wenn XY ähnlich ist, dann ist YX auch ähnlich. Aber wenn XY repräsentiert, dann repräsentiert YX in der Regel nicht. D.h.: Wenn ein Zwilling dem anderen ähnlich ist, so gilt dies auch andersherum. Dennoch wäre die Behauptung, ein Zwilling repräsentiere den anderen, absurd. Hingegen repräsentiert mich ein Bild von mir aber nicht umgekehrt. Auch zu behaupten, ich sähe dem Bild ähnlich ist wiederum nicht haltbar. Goodman analysiert den Unterschied zwischen pikturaler Repräsentation und Beschreibung letztlich als einen syntaktischen Unterschied in der Darstellungsweise. Einem Bild schreibt er die Eigenschaft zu, (im logischen Sinne) analog zu sein, während ein sprachliches Symbolschema digital ist. Als analog gilt für Goodman ein Symbolschema, das syntaktisch dicht ist. Dies bedeutet, dass alle Elemente des Symbolschemas von Bedeutung sind, man mithin überhaupt keine Grenzen zwischen den einzelnen Elementen ziehen kann. Ein digitales Schema ist hingegen disjunkt und endlich differenziert. Es besteht aus Inskriptionen oder Tokens, die auf einen Typ oder Charakter verweisen. Ein Token ist disjunkt, wenn er nicht zugleich dem Typ1 und dem Typ2 zugeordnet werden kann. Er ist endlich differenziert, wenn in einer endlichen Zeitspanne entschieden werden kann, ob er Typ1 oder Typ2 zugerechnet werden muss. So existiert das Wort "Ente" milliardenfach als Inskription in Form von Kopien. Eine einzelne Inskription von "Ente" ist disjunkt, da feststeht, dass sie nicht zugleich "Ente" und "Ende" darstellen kann. Sie ist endlich differenziert, wenn man diese Entscheidung in endlich vielen Schritten fällen kann. Bei pikturaler Darstellung existiert hingegen kein Typ. Entsprechend lassen sich die Kategorien disjunkt und endlich differenziert gar nicht erst anwenden. Auf dieser Grundlage entwirft Goodman ein breites Spektrum von gänzlich analogen Darstellungen, bis zu rein digitalen. Ein Bild ist gänzlich analog, während eine Partitur rein digital ist. Die Sprache hingegen ist eine Mischung, da sie zwar syntaktisch digital ist, aber nicht semantisch. Das Wort "Ball" kann sowohl ein Sportgerät als auch eine Tanzveranstaltung bezeichnen, es ist also semantisch nicht disjunkt.

In den "*Weisen der Weltzerzeugung*" wendet sich Goodman der Ontologie zu. Er löst das logische Dilemma, wonach zwei Weltbeschreibungen in sich widerspruchslös und in diesem Sinne wahr sein können, sich aber einander widersprechen, dahingehend auf, dass diese beiden Beschreibungen nicht die gleiche, sondern zwei verschiedene Welten beschreiben. Beispielsweise sind sowohl: "die Erde bewegt sich", als auch: "die Erde steht still" beide wahr, abhängig vom jeweiligen Bezugsrahmen. Ein Astronom, der kosmische Bewegung untersucht, untersucht eine Erde, die sich bewegt. Ein Wächter mit dem Befehl, Gefangene zu erschießen, sobald sie sich bewegen, wird dies eher nicht tun, "weil sie sich mit hoher Geschwindigkeit um die Sonne bewegt haben". Es handelt sich um zwei verschiedene Weltversionen.

In den "*Revisionen - Philosophie und andere Künste und Wissenschaften*" fasst Goodman schließlich gemeinsam mit Catherine Z. Elgin seine Forschungsergebnisse der Wissenschaftstheorie, der Symboltheorie und der Erkenntnistheorie zusammen und schlägt im dritten Teil des Buches eine "Neufassung der Philosophie" vor. Im Kern dieser Neufassung wird der Begriff der "Wahrheit" durch den der "Richtigkeit" abgelöst, da sich mit letzterem nicht der "berühmte philosophische Morass" auftut und er zudem auch auf nicht verbale beziehungsweise proportionale Symbolsysteme anwendbar ist. So kann man davon sprechen, dass es richtig ist, dass ein Bild Trauer ausdrückt, während man kaum sagen kann, dass dies wahr ist.

Schriften [Bearbeiten]

- *The Structure of Appearance*. Indianapolis. Hackett 1966.
- *Languages of Art*. London. Oxford University Press. 1969. dt.: *Sprachen der Kunst*. Suhrkamp 1998. ISBN 3-518-28904-7
- *Problems and Projects*. Indianapolis. Hackett 1972.
- *Ways of Worldmaking*. Indianapolis. Hackett 1978. dt.: *Weisen der Weltzerzeugung*. Frankfurt a.M. Suhrkamp 1984.
- *Of Mind and other Matters*. Cambridge, Massachusetts, London. 1984. dt.: *Vom Denken und anderen Dingen*. Frankfurt a.M. Suhrkamp 1987.
- *Tatsache, Fiktion, Voraussage*. Frankfurt a.M. Suhrkamp 1988. ISBN 3-518-28332-4
- zusammen mit Catherine Z. Elgin: *Revisionen*. Frankfurt a.M. Suhrkamp 1993.

Weblinks [Bearbeiten]

- Literatur von und über Nelson Goodman  im Katalog der Deutschen Nationalbibliothek
- Alessandro Giovannelli: „Goodman's Aesthetic“  in der *Stanford Encyclopedia of Philosophy* (englisch, inklusive Literaturangaben)

Normdaten: PND: 118696432  – weitere Informationen | LCCN: n50037322  | VIAF: 22426 

Kategorien: Hochschullehrer (Philadelphia) | Hochschullehrer (Waltham, Massachusetts) | Hochschullehrer (Harvard) | Philosoph (20. Jahrhundert) | Analytischer Philosoph 

Abbildung 1: Wikipedia-Eintrag zu Nelson Goodman

Der Wikipedia-Eintrag liefert rudimentäre Informationen zu Goodman und seinem Werk sowie eine Liste seiner Schriften. Auch haben fleißige Wikipedianer bereits eine Verlinkung in den Katalog der Deutschen Nationalbibliothek (DNB) geschaffen, so dass mit einem Klick Literatur von und über Nelson Goodman aus dem Bestand der DNB angezeigt werden kann.¹⁰

Für einen ersten Eindruck hat die Kölnerin also bereits einige Informationen erhalten und sie weiß auch, welche Bücher sie sich näher anschauen könnte, wenn sie tiefer in das Thema einsteigen will. Allerdings gibt es für sie keine Gelegenheit direkt herauszufinden, welche dieser Bücher in Kölner Bibliothekskatalogen verzeichnet sind und von ihr ausgeliehen werden können. Kennt sie die Rechercheangebote wie KölnBib¹¹ oder einzelne Kataloge Kölner Bibliotheken nicht, so wird sie sich das Buch im Internet bestellen oder einfach auf eine weitergehende Lektüre verzichten.

Was aber könnte passieren, wenn die Daten aus deutschen Bibliothekskatalogen als Linked Open Data vorlägen? Wie die Verlinkung in den Katalog der DNB zeigt,

¹⁰ Diese Art der Verlinkung ist nur möglich, weil die Titel- und Normdaten der DNB permanente URLs – sogenannte Permalinks – haben, im Unterschied zu vielen anderen Katalogen. Vgl. hierzu Fußnote 9.

¹¹ <http://www.koelnbib.de/>

arbeiten Wikipedianer seit einiger Zeit daran, auf Angebote von Bibliotheken zu verlinken. Mit Linked Open Bibliographic Data wäre es technisch nicht schwierig, Informationssuchenden im Zusammenhang eines Wikipedia-Artikels eine Liste von Bibliotheken aus ihrer Umgebung anzeigen zu lassen, die das jeweilige Buch im Bestand haben.¹² Damit könnten zusätzliche Informationen über Ort, Öffnungszeiten und Ausleihmöglichkeiten verbunden sein, so dass die Auswahl einer geeigneten Bibliothek leicht fällt.

Was vorerst Zukunftsmusik ist, könnte in nicht allzu ferner Zeit Realität werden, sobald Bibliotheken und Verbünde in großem Maßstab ihre Katalogdaten in das Linked-Open-Data-Netz migrieren. Was das bedeutet und wie die Migration vonstattengehen könnte, wird im weiteren Verlauf des Textes erläutert.

4 Linked-Data-Prinzipien

Die von Tim Berners-Lee aufgestellten Linked-Data-Prinzipien dienen als anerkannte Konventionen, an die sich halten sollte, wer Linked Data publizieren möchte:

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.*¹³

Es wurde bereits eine große Anzahl Daten entsprechend dieser Linked-Data-Prinzipien publiziert. Abbildung 2 zeigt die Linked-Data-Cloud vom Juli 2009, eine Darstellung der zu diesem Zeitpunkt gemäß Linked-Data-Prinzipien veröffentlichten Datensammlungen.¹⁴ Linked Data aus Bibliotheken befinden sich im rechten oberen Teil der Cloud.¹⁵ Dort finden sich neben LIBRIS, dem schwedischen Verbundkatalog¹⁶, den Subject Headings der Library of Congress (LCSH)¹⁷, der Gemeinsamen Normdatei (GND) der Deutschen Nationalbibliothek¹⁸,

¹² Die Wikipedia scheint mittlerweile auch darauf zu setzen, das in ihr gesammelte Wissen mit RDF anzureichern, vgl. Simonite, Tom (2010): Wikipedia to Add Meaning to Its Pages. Einsehbar unter <http://www.technologyreview.com/web/25728/?a=f>.

¹³ <http://www.w3.org/DesignIssues/LinkedData.html>

¹⁴ Wohlgemerkt handelt es sich häufig nur um Linked Data, weil eine Lizenzierung der Daten mit offenen Lizenzen häufig (noch) nicht stattgefunden hat.

¹⁵ Mit der Version vom September 2010 ist die Linked-Data-Cloud so groß geworden, dass sie nicht mehr für die Darstellung in einer Print-Publikation geeignet ist, weshalb empfohlen wird, den Link zur Quelldatei in den Browser einzugeben.

¹⁶ <http://libris.kb.se/> Die Daten des Libris liegen allerdings nicht als Komplettabzug zum Download bereit, weshalb hier nicht von Linked Open Data, sondern von Linked Data gesprochen werden kann.

¹⁷ Die LoC hat im Mai 2009 begonnen, ihre Normdaten in RDF zu veröffentlichen, vgl. <http://id.loc.gov/authorities/>.

¹⁸ Auch die DNB hat die Normdaten als Linked Data veröffentlicht, vgl. die Dokumentation unter <https://wiki.d-nb.de/display/LDS/Dokumentation+des+Linked+Data+Services+der+DNB>. Allerdings sind die Daten (noch) nicht unter einer offenen Lizenz veröffentlicht.

der Virtuellen Internationalen Normdatei VIAF¹⁹, der Open Library und den Daten aus der Universitätsbibliothek Mannheim²⁰ auch die beiden (experimentellen) Dienste lobid-resources²¹ und lobid-organisations²², die vom hbz betrieben werden.²³

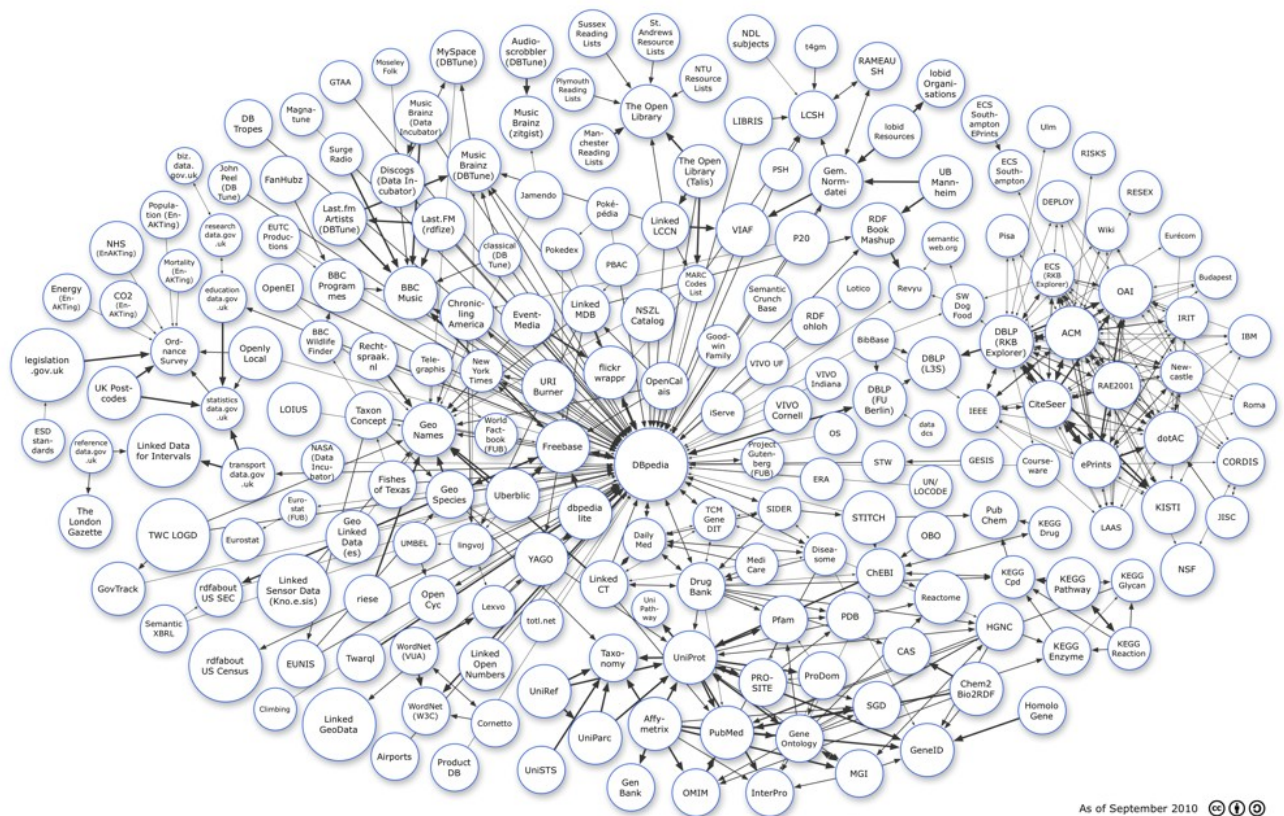


Abbildung 2: Die Linked-Open-Data-Cloud 09/2010, Quelle <http://richard.cyganiak.de/2007/10/lod/>.

5 Linked Data: Standards und Grundfunktion

Welche Technologie steckt hinter Linked Data? Welche sind die Prinzipien, auf deren Fundament das Linked-Data-Netz wächst? Bereits bestehende Web-Standards bilden auch für Linked Data die Basis: Das *Hypertext Transfer Protocol* (HTTP) als Datenaustauschprotokoll und HTTP-URIs (Uniform Resource Identifier)²⁴ zur Identifikation von Dingen²⁵ sind Eckpfeiler von Linked Data. Darüber hinaus kommen weitere Standards hinzu. Im Folgenden wird ein knapper

¹⁹ <http://viaf.org/>

²⁰ <http://data.bib.uni-mannheim.de/>

²¹ <http://lobid.org/de/resource.html>

²² <http://lobid.org/de/organisation.html>

²³ Vgl. Abschnitt 7.2.

²⁴ URIs kennen wir alle. Der Link in der vorherigen Fußnote ist eine URI, aber auch andere Identifikatoren folgen dem URI-Schema, z.B. URNs oder auch Mailboxangaben der Form <mailto:pohl@hbz-nrw.de>. Zum URI-Schema und weiteren Beispielen vgl. Wikipedia, The Free Encyclopedia (2010): Artikel „URI scheme“. Bearbeitungsstand: 9. Juli. URL: http://en.wikipedia.org/w/index.php?title=URI_scheme&oldid=372490112

²⁵ ‚Ding‘ wird hier – sehr allgemein – im Sinne von ‚Entität‘ verwendet und umfasst auch abstrakte Konzepte.

Überblick über die grundlegenden Standards geliefert.

5.1 URIs

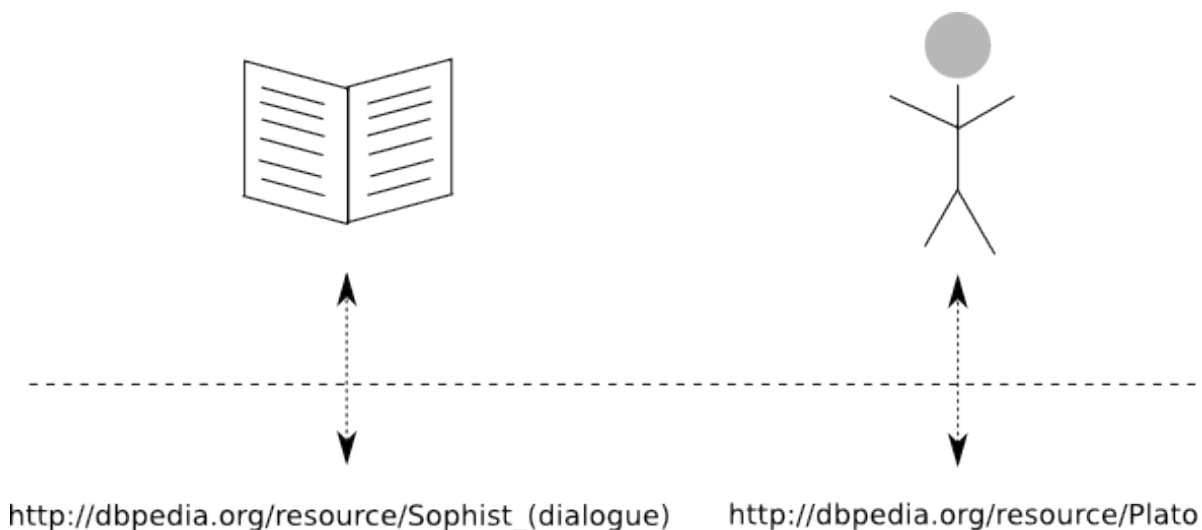


Abbildung 3: URIs als Namen von Dingen

Der Unterschied zum aktuellen Web besteht darin, dass HTTP-URIs nicht mehr allein zur Identifikation und Adressierung von abrufbaren Webressourcen (HTML-Seiten, Dateien), sondern auch zur Bezugnahme auf Nicht-Web-Ressourcen wie Personen, Bauwerke, Orte etc. benutzt werden. Abbildung 3 zeigt, wie URIs als Namen von Dingen wie einem Buch oder einer Person genutzt werden.²⁶

Das besondere an URIs ist: Sie sind öffentliche Identifikatoren. Jeder kann sie verwenden. Das heißt, jeder Mensch kann überall RDF-Aussagen mit jedem URI veröffentlichen und dadurch das Wissen über das referenzierte Ding erweitern. URIs können auch als Dubletten verknüpft werden: Auch das ist eine Aussage.²⁷

5.2 RDF-Tripel

Als wichtigste Ergänzung zu bestehenden Standards kommt das Resource Description Framework (RDF) hinzu, ein Modell zur Repräsentation einzelner Aussagen. Es ist so konzipiert, dass Maschinen optimal über RDF-Daten operieren können und ist somit gewissermaßen für Linked Data das, was HTML für das – in erster Linie menschenlesbare – World Wide Web ist.²⁸

Hier ein Beispiel zur Veranschaulichung: Die Aussage „Platon ist der Autor des *Sophistes*.“ ist für jeden Menschen verständlich, der Deutsch spricht und ein gewisses Hintergrundwissen mitbringt, der etwa weiß, dass ‚Platon‘ auf einen

²⁶ In diesem und den folgenden Beispielen benutzen wir dbpedia-URIs aus dem Namensraum <http://dbpedia.org>. DBpedia ist ein Projekt der Freien Universität Berlin und der Universität Leipzig, das strukturierte Information aus der Wikipedia als Linked Data bereitstellt.

²⁷ Wenn auch die Eineindeutigkeit von URIs relativ leicht herzustellen sein mag, so lässt sich Redundanz, die durch die Bezeichnung eines Dings durch mehrere URIs entsteht, nicht immer verhindern. Dann wird es nötig, die URIs als gleichwertig zu bestimmen.

²⁸ Darüber hinaus gibt es mit RDFa einen Standard, der HTML und RDF verschmilzt, indem er die Einbettung von RDF in HTML ermöglicht. Siehe hierzu etwa W3C (2008): RDFa Primer - Bridging the Human and Data Webs. Einsehbar unter <http://www.w3.org/TR/xhtml-rdfa-primer/>.

antiken Philosophen bezugnimmt und ‚Sophistes‘ auf ein in Dialogform geschriebenes philosophisches Werk. Maschinen können mit diesem Satz nicht allzu viel anfangen. Dies ändert sich, wenn diese Aussage in Form einer RDF-Aussage vorliegt und mit anderen Aussagen verknüpft ist. RDF-Aussagen bestehen aus drei Teilen, aus diesem Grund spricht man auch von einem RDF-*Tripel*. Die drei Teile werden ‚Subjekt‘, ‚Prädikat‘ und ‚Objekt‘ genannt. Schauen wir uns einmal eine solche Aussage an:

<[http://dbpedia.org/resource/Sophist_\(dialogue\)](http://dbpedia.org/resource/Sophist_(dialogue))>
 <<http://purl.org/dc/elements/1.1/creator>> <<http://dbpedia.org/resource/Plato>> .²⁹

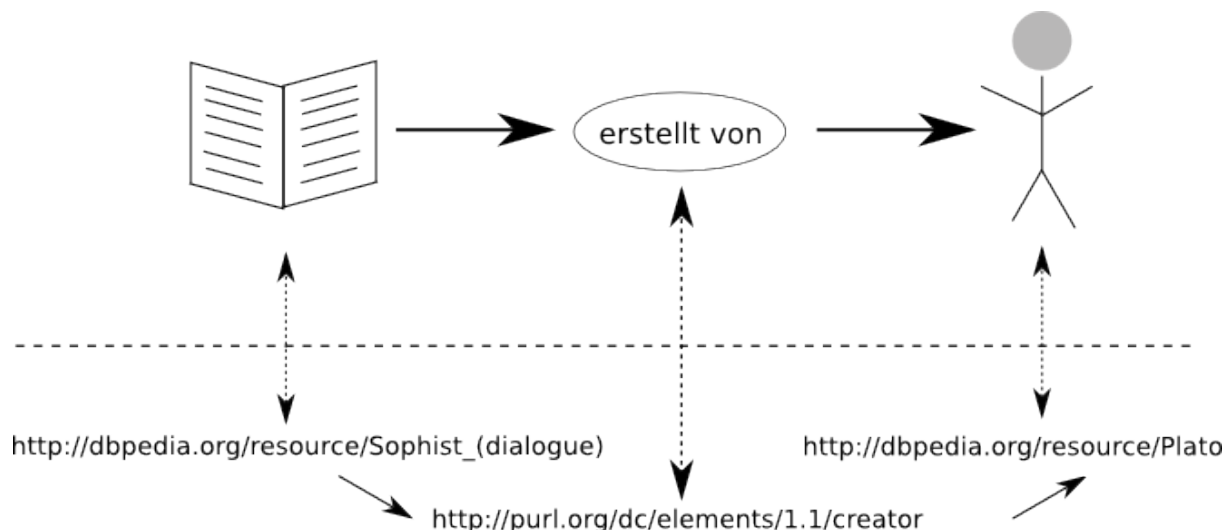


Abbildung 4: RDF-Tripel

Wie wir sehen, werden die drei Teile der RDF-Aussage (Subjekt, Prädikat und Objekt) durch HTTP-URIs repräsentiert. Warum ist das so? Was für Menschen nur schwer lesbar erscheint, erweist sich für eine verteilte Datenbank im Web als sehr vorteilhaft. Zum einen sind URIs – idealerweise – eindeutige Identifier, d.h. sie werden zur Referenz auf genau ein Ding benutzt, denn die Regel ist, einen URI nicht als Name für zwei verschiedene Dinge zu verwenden. Abbildung 4 illustriert, auf was die Bestandteile des Tripels bezugnehmen. Zum anderen ermöglichen HTTP-URIs es Maschinen, die Namen auf Basis des http-Protokolls aufzulösen, d.h. bei Eingabe in einen Browser Informationen über das referenzierte Ding zu liefern.³⁰

5.3 Typisierte Verlinkungen

Wie wir in unserem Beispiel sehen, sind nicht nur die referenzierten Dinge durch einen URI identifiziert, sondern auch die Verbindung zwischen beiden: das Prädikat. Im konkreten Fall handelt es sich um den URI³¹ für das Dublin-Core-

²⁹ Das Prädikat `dc:creator` und seine Nutzung können – wie auch andere Prädikate – für manche verwirrend sein. Hier hilft es, RDF wie ein Formular zu lesen: Das Subjekt ist das Ding, über das im Formular geredet wird. Dazu gibt es ein Element „creator“, dessen Inhalt durch das Objekt angegeben wird.

³⁰ Das Ding selbst kann freilich nicht geliefert werden, solange es sich nicht um eine Web-Ressource handelt. Eine Eingabe von <http://dbpedia.org/resource/Plato> in meinen Browser wird mir nicht Platon in mein Zimmer liefern, sondern leitet mich auf eine andere Seite (<http://dbpedia.org/page/Plato>) weiter, die Informationen über Platon enthält.

³¹ Die Verwendung von URIs für Prädikate macht es möglich, das Prädikat selbst im Web

Metadatenelement „creator“. Hier wird ein grundlegender Unterschied im Vergleich zum World Wide Web deutlich. Zwar basiert auch das Web auf Links, allerdings sind diese nicht typisiert, d.h. in einem HTML-Dokument steht zwar, dass dieses Dokument auf ein anderes verlinkt, welcher Art diese Verlinkung ist, bleibt allerdings allenfalls für Menschen nachvollziehbar. Ob auf das Dokument zum Zweck eines Hinweises, einer Rezension etc. verlinkt wird, ist nicht explizit angegeben. Im Linked-Data-Netz ist dies anders, hier ist *jede* Verlinkung typisiert, in unserem Fall als eine Werk-Autor-Beziehung.

5.4 Literale

Natürlich müssen die URIs in RDF mit natürlicher Sprache, Bildern usw. in Verbindung gesetzt werden, um letztlich uns Menschen nutzen zu bringen. Deshalb ist es möglich an die Objektstelle eines Tripels Freitext (in RDF-Sprech ein ‚Literal‘) zu schreiben, während an Subjekt- und Prädikatstelle immer ein URI stehen muss.³² Literale spielen zum Beispiel eine Rolle, wenn es darum geht, eine Aussage über den Titel eines Werkes zu machen. Ein Beispiel:

```
<http://dbpedia.org/resource/Sophist_(dialogue)>  
<http://purl.org/dc/elements/1.1/title> „Sophistes“ .
```

RDF verfügt über zwei Möglichkeiten, um solche Literale näher zu charakterisieren: die Angabe der Sprache, in der das Literal verfasst ist und die Angabe eines Datentyps wie etwa „Ganzzahl“ oder „Datum“. Es ist dabei je Literal nur eine Art der Markierung erlaubt, zum Beispiel:

```
<http://dbpedia.org/resource/Sophist_(dialogue)>  
<http://purl.org/dc/elements/1.1/title> „Sophistes“@de .
```

```
<http://dbpedia.org/resource/Sophist_(dialogue)>  
<http://purl.org/dc/elements/1.1/title> „Sophist“@en .
```

```
<http://dbpedia.org/resource/Sophist_(dialogue)>  
<http://purl.org/dc/elements/1.1/title> „Σοφιστής“@el .
```

```
<http://dbpedia.org/resource/Sophist_(dialogue)>  
<http://purl.org/dc/elements/1.1/created> „-360“^^xsd:gYear .
```

Die Markierung von Literalen mit Attributen ist die eine Form der in RDF vorgesehenen Möglichkeiten zur Typisierung. Um auch Aussagen über den Typ von Dingen treffen zu können, muss man sich auf das Gebiet der Ontologien wagen.

5.5 Ontologien

Wie bereits erläutert, handelt es sich bei RDF um ein abstraktes Modell, in dem Aussagen in Form von Tripeln gemacht werden können. Im Prinzip können dabei

nachschlagen zu können, wenn z.B. Der Verwendungszweck nicht klar ist. Siehe hierzu auch den folgenden Abschnitt über Ontologien.

³² Die abstrakte RDF-Syntax, die bestimmt, was alles RDF genannt werden darf, ist hier zu finden: <http://www.w3.org/TR/rdf-concepts/#section-Graph-syntax>. Der Einfachheit halber lassen wir Blank Nodes in der vorliegenden Darstellung außen vor, siehe dazu http://en.wikipedia.org/wiki/Blank_node.

einem Subjekt über beliebige Prädikate beliebige Objekte zugeordnet werden. Um diese Beliebigkeit einzuschränken, kann anhand von Ontologie-Sprachen wie RDF-Schema (RDFS) und der Web Ontology Language (OWL) festgelegt werden, welche Klassen von Entitäten im Kontext der jeweiligen Ontologie als sinnvoll erachtet werden, welche Beziehungen – also Prädikate – zwischen Entitäten herrschen können und welche Schlüsse man aus diesen Beziehungen ableiten kann.

Das folgende Ontologie-Fragment definiert beispielhaft die Klassen *Person*, *Dokument* und *Buch*, wobei *Buch* als Unterklasse von *Dokument* modelliert wird. Darüber hinaus wird die Relation bzw. das Prädikat *autor* definiert³³:

```
@prefix :      <#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

:Person rdf:type rdfs:Class .
:Dokument rdf:type rdfs:Class .
:autor rdf:type rdf:Property .
:Buch rdfs:subClassOf :Dokument .
:autor rdfs:domain :Dokument .
:autor rdfs:range :Person .
```

In dieser Ontologie stecken folgende explizite Definitionen:

- Es gibt die Klassen *Person* und *Dokument*.
- Es gibt das Prädikat *autor*.

Darüber hinaus werden drei Regeln definiert:

- Wenn in einem Tripel das Prädikat *autor* verwendet wird, dann ist das Subjekt notwendigerweise ein *Dokument*.
- Wenn in einem Tripel das Prädikat *autor* verwendet wird, dann ist das Objekt notwendigerweise eine *Person*.
- Wenn etwas ein *Buch* ist, dann ist es immer auch ein *Dokument* (aus diesem Verhältnis zwischen Ober- und Unterklasse geht auch hervor, dass es sich bei *Buch* ebenfalls um eine Klasse handeln muss).

Anhand dieser Regeln wird zum einen deutlich, wie die Instanzen einer Klasse zueinander in Beziehung stehen können. Darüber hinaus ermöglicht es die Ontologie, implizite Aussagen inferieren zu können. So kann aus der Aussage

`:Sophistes :autor :Plato`

gefolgert werden, dass

³³ Aufgrund der Abstraktheit der Syntax gibt es für RDF mehrere Notationssysteme. In diesem Beispiel wird zur Steigerung der Lesbarkeit die Turtle-Notation verwendet. Vgl. <http://www.w3.org/2007/02/turtle/primer/http://www.w3.org/2007/02/turtle/primer/>.

- :Sophistes vom Typ :Buch ist,
- :Sophistes damit auch vom Typ :Dokument ist und dass
- :Plato vom Typ :Person ist.

Die Kernaufgaben einer Ontologie sind also die Definition von Klassen, möglichen Beziehungen und möglichen Schlüssen, die aus diesen Beziehungen hervorgehen. Ontologien sind damit ein sehr mächtiges Werkzeug für die Modellierung beliebiger Weltausschnitte.

Ontologien werden für bestimmte Zwecke geschaffen, idealerweise von Fachleuten des Gebietes, das die jeweilige Ontologie strukturieren soll. Es ist sinnvoll, wenn möglich auf bestehende Ontologien oder Teile derer zurückzugreifen, so dass nicht verschiedene Ontologien für ein und dieselbe Domäne entstehen. Im Umkehrschluss heißt dies für die Veröffentlichung von Ontologien, deren Nachnutzbarkeit sicherzustellen. Dies schließt eine ausführliche Dokumentation sowie die Lizenzierung der Ontologie unter einer offenen Lizenz mit ein.³⁴

5.6 SKOS

SKOS steht für "Simple Knowledge Organization System"³⁵. Es ist entwickelt worden, um kontrollierte Vokabulare – also Klassifikationen, Thesauri und Taxonomien – Linked-Data-konform zu kodieren. Dies ermöglicht es, solche Dokumentationssprachen (1) modular aufzubauen, um die Nachnutzbarkeit zu erhöhen, und somit (2) miteinander vernetzen und in Beziehung setzen zu können.

SKOS ist demnach eine Ontologie für kontrollierte Vokabulare. Der Aufbau von SKOS ist denkbar einfach. Grundbausteine sind *Konzepte*, denen Notationen und Bezeichnungen³⁶ zugeordnet werden³⁷:

³⁴ Beispielhaft sei hier die Dokumentation der DC-Terms genannt, wo allerdings (bisher) eine offene Lizenzierung versäumt wurde, vgl. <http://dublincore.org/documents/dcmi-terms/>.

³⁵ Es folgt an dieser Stelle nur eine sehr knappe Skizze. Für eine ausführliche Einführung siehe <http://www.w3.org/TR/skos-primer/>.

³⁶ Es sind grundsätzlich zwei Arten von Bezeichnungen vorgesehen: bevorzugte Bezeichnungen, von denen jedes Konzept für jede Sprache nur eine haben darf, und eine beliebige Menge von alternativen Bezeichnungen.

³⁷ Dieses Beispiel stützt sich auf die Linked-Data-Variante der Dewey Decimal Classification, die die ersten drei Ebenen umfasst und welche unter <http://dewey.info/> zu finden ist.

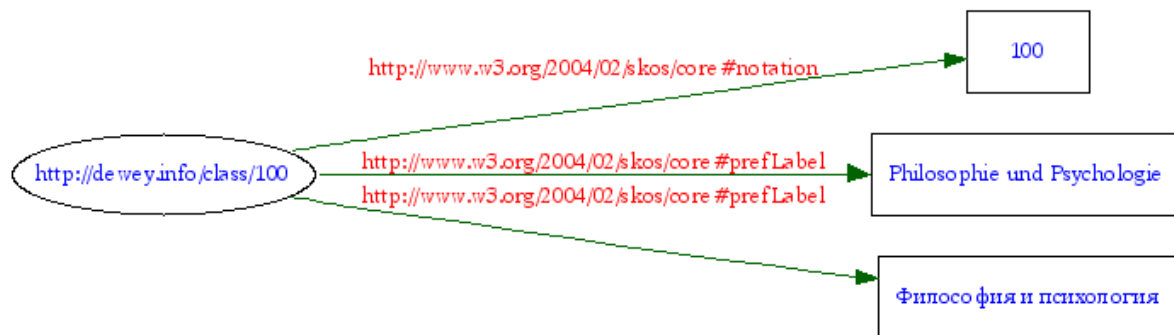


Abbildung 5: Darstellung einer DDC-Systemstelle in SKOS³⁸

Einzelne Konzepte können anhand der hierarchischen Relationen *narrower* und *broader* miteinander verknüpft werden³⁹:

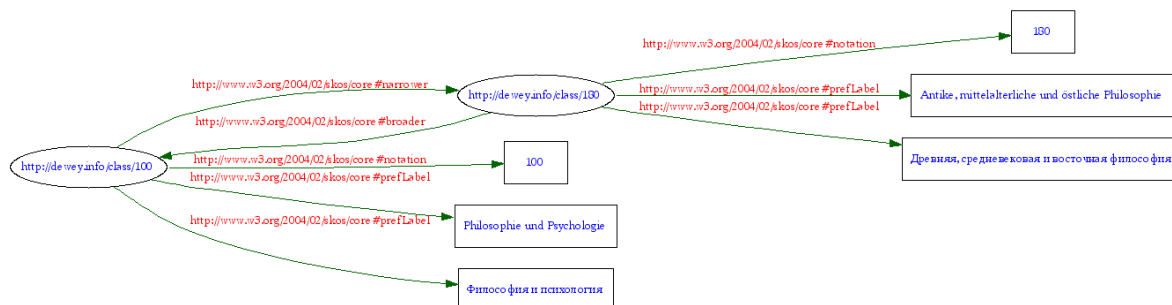


Abbildung 6: Darstellung der Verknüpfung von DDC-Systemstellen in SKOS

Neben den Relationen, die innerhalb eines kontrollierten Vokabulars bestehen, können darüber hinaus Verbindungen zu anderen Vokabularen etabliert werden. So können Konzepte der Dewey Decimal Classification (DDC) beispielsweise wie folgt anhand des Prädikats *closeMatch* auf die Konzepte der Library of Congress Subject Headings (LCSH)⁴⁰ abgebildet werden:

```
<http://dewey.info/class/100>
<http://www.w3.org/2004/02/skos/core#closeMatch>
<http://id.loc.gov/authorities/sh99005065#concept> .
```

Weitere bereits in SKOS veröffentlichte Dokumentationsprachen sind der Standard Thesaurus Wirtschaft⁴¹, der Thesaurus Sozialwissenschaften⁴² und die

³⁸ Die Abbildungen 5 und 6 exemplifizieren eine weitere Weise der Darstellung von RDF-Tripeln: die Darstellung als gerichteter Graph mit Knoten und Kanten. URI-Knoten werden als Oval dargestellt, Literale als Rechteck. Unter <http://www.w3.org/RDF/Validator/> können aus RDF-Beschreibungen solche Grafiken generiert werden.

³⁹ Darüber hinaus gibt es mit *related* auch eine nicht-hierarchische Relation, die Querverweise ermöglicht.

⁴⁰ Wie bereits erwähnt, stellt die Library of Congress die LCSH unter <http://id.loc.gov/authorities/> auch als Linked Data bereit.

⁴¹ <http://zbw.eu/stw/>, vgl. auch Borst, Timo / Neubert, Joachim (2009): Case Study: Publishing STW Thesaurus for Economics as Linked Open Data, <http://www.w3.org/2001/sw/sweo/public/UseCases/ZBW/>.

⁴² <http://www.gesis.org/dienstleistungen/tools-standards/thesaurus-sozialwissenschaften/>, vgl. auch Benjamin Zapilko und York Sure (2009): Converting TheSoz to SKOS. Einsehbar unter http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_met_hodenberichte/2009/technicalreport_09_07.pdf.

Normdaten der Deutschen Nationalbibliothek⁴³.

Bibliographische Ressourcen können durch RDF-Tripel mit den in SKOS kodierten kontrollierten Vokabularen verknüpft werden:

```
<http://dbpedia.org/resource/Sophist_(dialogue)>  
<http://purl.org/dc/elements/1.1/subject> <http://dewey.info/class/180> .
```

Doch welchen Wert haben all diese Verknüpfungen? Sie ermöglichen es zum einen, über das Linked-Data-Netz zu navigieren und zu neuen relevanten Ressourcen aus anderen Kontexten zu gelangen. So nützlich und sinnvoll allein schon das Navigieren über Linked Data sein kann, so notwendig ist natürlich zum anderen auch die Möglichkeit der Recherche. Im nächsten Abschnitt wird der Standard skizziert, der Abfragen über RDF-Tripeln dient.

5.7 SPARQL

Ein großer Vorteil von Linked Data ist, dass komplexe Abfragen über Daten gemacht werden können, sobald eine hinreichend große und komplexe Datenbasis vorliegt. SPARQL ist die Abfragesprache für RDF und das Akronym wird aufgelöst in „SPARQL Protocol And RDF Query Language“. Sie dient Entwicklern dazu, Anwendungen auf dieser Basis aufzubauen und ist nicht für den Einsatz durch Laien gedacht.

Eine einfache SPARQL-Anfrage an die DBpedia kann etwa so aussehen:

```
Select ?x  
Where {  
?x <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>  
<http://xmlns.com/foaf/0.1/Person> .  
?x <http://dbpedia.org/ontology/era>  
<http://dbpedia.org/resource/Ancient_philosophy> .  
?x <http://dbpedia.org/property/mainInterests>  
<http://dbpedia.org/resource/Epistemology> .  
}
```

Die Antwort auf diese Abfrage liefert uns sämtliche in DBpedia verzeichnete Personen, die der Antiken Philosophie zugeordnet werden und deren Hauptinteresse der Epistemologie galt.⁴⁴

SPARQL-Abfragen können beliebig komplex sein, d.h. es lassen sich auch Anfragen der Art bilden: *Liefere mir alle weiblichen Autoren, die während des ersten Weltkriegs in Wien publiziert haben und vor 1880 geboren sind.*

⁴³ Vgl. Fußnote 18.

⁴⁴ Die Antwort auf diese Anfrage lässt sich hier einsehen: http://dbpedia.org/snorql/?query=Select+%3F%3Fx%0D%0AWhere+%7B+%0D%0A%3F%3Fx+%3Chttp%3A%2F%2Fwww.w3.org%2F1999%2F02%2F22-rdf-syntax-ns%23type%3E+%3Chttp%3A%2F%2Fxmlns.com%2Ffoaf%2F0.1%2FPerson%3E+.%0D%0A%3F%3Fx+%3Chttp%3A%2F%2Fdbpedia.org%2Fontology%2Fera%3E+%3Chttp%3A%2F%2Fdbpedia.org%2Fresource%2FAncient_philosophy%3E+.%0D%0A%3F%3Fx+%3Chttp%3A%2F%2Fdbpedia.org%2Fproperty%2FmainInterests%3E+%3Chttp%3A%2F%2Fdbpedia.org%2Fresource%2FEpistemology%3E+.%0D%0A+%7D%0D%0A. Die geringe Trefferzahl spricht dafür, dass die Daten in der Wikipedia bisher nicht besonders sorgfältig erfasst bzw. nicht in RDF umgewandelt worden sind.

Voraussetzung für die Durchführung komplexer Anfragen sind natürlich immer große Mengen sorgfältig erfasster, standardisierter und komplexer Daten. Der Aufwand zur Erstellung und Pflege von Daten wird einem durch Linked Data nicht abgenommen.

6 Der Migrationspfad ins Linked-Data-Netz

Wie gelangen die Daten aus Bibliotheks- und Verbundkatalogen nun in das Linked-Data-Netz? Was müssen wir dafür tun? Im Folgenden werden sechs grundlegende Schritte erläutert.

6.1 Daten analysieren

Wir müssen unsere Daten kennen und müssen wissen, welche Regelmäßigkeiten in den Katalogisaten zu finden sind. Die grundlegenden Fragen sind:

- Welche Dinge beschreiben unsere Daten?
- Wie geschieht dies?

Die erste Frage mag leicht erscheinen (Antwort: Personen und Medien), aber man muss sich nur einmal eine FRBR-Diskussion anschauen, um zu merken, dass es gar nicht so klar ist, welche Dinge in Katalogen eigentlich beschrieben werden (sollten).⁴⁵

Die zweite Frage ist noch kniffliger. Zwar gibt es niedergeschriebene Katalogisierungsregeln und standardisierte Datenformate, allerdings kann man sich nicht darauf verlassen, dass bei der tatsächlichen Katalogisierung streng diesen Regeln gefolgt wurde bzw. die Regeln auf dieselbe Art und Weise interpretiert wurden. Es gibt in jeder Bibliothek und jedem Verbund individuelle Katalogisierungspraktiken.

Daten analysieren heißt also nicht nur Regelwerke analysieren, sondern auch statistische Daten über die tatsächliche Benutzung von MAB-/MARC-Feldern und deren Relationen untereinander zu erheben.

6.2 URIs prägen vs. URIS nachnutzen

Bevor RDF-Tripel aus Katalogdaten generiert werden können, müssen die beschriebenen Dinge erst einmal mit einem URI „getauft“ sein. Es gibt zwei Wege, an URIs für katalogisierte Dinge zu kommen:

1. Ich präge in einem von mir kontrollierten Namensraum⁴⁶ für jedes zu beschreibende Ding neue URIs.
2. Ich ermittle, ob für das beschriebene Ding bereits ein URI existiert und benutze diesen URI zur Beschreibung des Dings.

⁴⁵ Ein Beispiel für eine solche Diskussion ist jene vom Mai/Juni 2010 auf der Mailingliste der Open Bibliographic Data Group, siehe <http://lists.okfn.org/pipermail/open-bibliography/2010-May/000115.html>.

⁴⁶ ‚Namensraum‘ wird hier im Prinzip synonym mit ‚Domain‘ verwendet. Beispiele für Domains sind etwa wikipedia.org, hbz-nrw.de, w3.org.

Für die erste Variante spricht die Einfachheit. Es ist momentan nämlich alles andere als trivial, existierende URIs zu finden. Allerdings spricht etwas dafür, diese Mühen, die die zweite Variante mit sich bringt, auf sich zu nehmen: es ergibt sich dadurch eine höhere Auffindbarkeit der Daten, da der Name des referenzierten Dinges bereits außerhalb der eigenen Daten bekannt ist. Darüber hinaus wächst durch die Nutzung bestehender URIs das Wissen über ein Ding sofort an, was die Abfragemöglichkeiten verbessert. Ein Mittelweg ist es, eigene URIs zu prägen und diese über das Prädikat owl:sameas nach und nach mit weiteren existierenden URIs für dasselbe Ding zu verknüpfen.⁴⁷

6.3 Ontologien evaluieren, nachnutzen, kreieren

Wenn die Frage nach URIs für die beschriebenen Dinge geklärt ist, werden passende Ontologien oder Vokabulare benötigt. Fragen, die es hier zu beantworten gilt, sind: Wie werden die Dinge in unseren Katalogen klassifiziert, und welche Eigenschaften werden ihnen zugeschrieben? Gibt es bereits Ontologien, die die entsprechenden Klassen und Eigenschaften vorweisen? Dieser Schritt knüpft demnach nahtlos an den ersten Schritt der Datenanalyse an. Um die Interoperabilität der Daten zu maximieren, sollten bestehende Ontologien genau betrachtet werden.

Werden geeignete bereits existierende Klassen und Prädikate ausgemacht, so sollten diese auch verwendet werden. Ist dies nicht der Fall, so müssen eigene Ontologien aufgestellt werden. Häufig gibt es bestehende Ontologien, die aber nicht alle relevanten Klassen und Beziehungen aufweisen. In diesem Fall ist es günstig, die eigene Ontologie auf der bestehenden aufzubauen und neue Prädikate und Klassen zu ergänzen. Ein Beispiel für das Aufbauen auf relevanten Ontologien, die aber nicht spezifisch genug sind, ist die Bibliographic Ontology (Bibo)⁴⁸: Sie baut auf den DCMI Metadata Terms⁴⁹ sowie der FOAF⁵⁰- und der Event-Ontologie⁵¹ auf und ergänzt diese, wo es nötig ist, um eigene Klassen und Prädikate.

Für bibliographische Daten ist neben den bereits genannten Ontologien (Bibo und DCMI Metadata Terms) natürlich auch RDA (Resource Description and Access) interessant. Es gibt bereits erste Anfänge von RDA in RDF, wenn auch in einer schlecht konsumierbaren Version.⁵²

⁴⁷ In der Arbeitsgemeinschaft „Kooperative Verbundanwendungen“ wird derzeit nach Mitteln und Wegen gesucht, die verhindern, dass in der deutschen Bibliothekswelt von verschiedenen Verbänden und Bibliotheken unzählige URIs für ein und dasselbe Ding geprägt werden. Zu diesem Zweck müsste so etwas wie ein Verzeichnis entstehen, in dem bestehende URIs mit Identifikatoren wie der ISBN, der OCLC-Nummer oder der Erstkatalogisierungs-ID (EKI) in Beziehung gesetzt werden, so dass auf diese Weise bereits geprägte URIs ermittelt werden können.

⁴⁸ <http://bibliontology.com/>

⁴⁹ <http://dublincore.org/documents/dcmi-terms/>. DCMI steht für ‚Dublin Core Metadata Initiative‘.

⁵⁰ <http://www.foaf-project.org/>. Das Akronym FOAF steht für ‚Friend of a Friend‘.

⁵¹ <http://motools.sf.net/event/event.html>.

⁵² Siehe RDA in der Metadata Registry der National Science Digital Library: <http://metadataregistry.org/rdabrowse.htm>. Vgl. auch Hillmann, Diane / Coyle, Karen / Phipps, John / Dunsire, Gordon: RDA Vocabularies: Process, outcome, Use. D-Lib Magazine 16, Nr. 1 (1, 2010). Einsehbar unter <http://www.dlib.org/dlib/january10/hillmann/01hillmann.html>.

6.4 Daten (ad-hoc) konvertieren

Steht fest, welche URIs und Ontologien verwendet werden sollen, kann mit der Konvertierung der Daten nach RDF begonnen werden. Diese Konvertierung ist in der Übergangsphase hin zum Linked-Data-Netz ein regelmäßig zu wiederholender Prozess, weil Anwendungen RDF und Ontologien noch nicht nativ unterstützen. Da die Daten in der jeweiligen Anwendung – etwa einem Katalogisierungssystem – gepflegt werden, müssen die regelmäßig vorgenommenen Änderungen aus dem verwendeten Format – MAB2, MARC21 etc. – in die RDF-Repräsentation übernommen werden.

Statt Daten in großen Paketen regelmäßig nach RDF zu konvertieren, können auch Schnittstellen geschaffen werden, die die Daten ad-hoc zu dem Zeitpunkt überführen, zu dem sie angefordert werden. Ein Beispiel für dieses Szenario ist der bereits angesprochene Linked-Data-Service der Deutschen Nationalbibliothek.⁵³

6.5 Daten verlinken

Wie der Name „Linked Data“ sagt, ist es essentiell, die eigenen Daten mit anderen zu vernetzen; nur so können die Vorteile von Linked Data voll ausgenutzt werden. Das heißt, es sollten Literale wirklich nur dann in Tripeln vorkommen, wenn das Objekt nicht als Verweis auf ein anderes Ding aufgefasst werden kann – etwa bei dem Titel einer Publikation. Für viele Dinge existieren bereits geeignete Verknüpfungsmöglichkeiten; neben den bereits erwähnten SKOS-kodierten kontrollierten Vokabularen sei hier beispielhaft auf Geonames⁵⁴ für Ortsangaben verwiesen. Es ist beispielsweise wünschenswert, Geburts- und Verlagsorte nicht als Literale zu kodieren, sondern auf den URI des jeweiligen Ortes zu verlinken.

Nach der Veröffentlichung der Daten muss also daran gearbeitet werden, die Daten zunehmend mit anderen Daten zu verlinken. Es ist erstrebenswert, dass auf die eigenen URIs auch von außen verlinkt wird. Dies lässt sich nicht erzwingen. Eine offene Publikation von kompletten Datenabzügen kann sich hier aber positiv auswirken, da es die maschinelle Erzeugung von Links auf diese Daten stark vereinfacht.⁵⁵

6.6 Im Web bereitstellen

Während ein Gesamtabzug der Daten für gewisse Zwecke sehr nützlich ist, ist dies nicht die einzige Form, in der die Daten publiziert werden sollten. So wie bei der Eingabe einer URL die entsprechende HTML-Seite von einem Webserver ausgeliefert wird, muss auch das Abrufen der RDF-Beschreibung eines Dinges ermöglicht werden. In vielen Fällen, wenn es sich etwa um Personen oder Orte handelt, kann das Ding selbst nicht ausgeliefert werden.⁵⁶ Für diese Fälle muss der Webserver so konfiguriert werden, dass er auf ein beschreibendes Dokument weiterleitet, wenn ein solches Ding angefordert wird⁵⁷. Schließlich bietet es sich

⁵³ Vgl. Fußnote 18.

⁵⁴ <http://www.geonames.org/>.

⁵⁵ Vgl. zur offenen Lizenzierung von Daten Pohl (2010).

⁵⁶ Vgl. Fußnote 30.

⁵⁷ Eine ausführliche Erläuterung dieser Problematik würde den Rahmen dieses Artikels

auch an, die Daten über eine SPARQL-Schnittstelle zugänglich zu machen, um etwa Entwicklern eigene SPARQL-Abfragen zu ermöglichen.

7 Aktivitäten im hbz-Verbund

Es ist unschwer zu sehen, dass eine Migration bibliographischer Daten in das Linked-Data-Netz alles andere als trivial ist. In diesem Bereich muss noch eine Menge experimentiert werden, ehe diese Aufgabe befriedigend bewältigt worden ist. Welche Linked-Data-Aktivitäten im hbz stattfinden, soll nun abschließend knapp dargestellt werden.

7.1 Vernetzung und Kooperation

Da auf dem Weg zur erfolgreichen Implementierung von Linked Library Data noch eine Menge Erfahrungen gemacht werden müssen, ist es unerlässlich, sich mit anderen Akteuren auf diesem Gebiet auszutauschen. Im hbz findet deshalb alle zwei Wochen ein regelmäßiges internes Treffen statt, bei dem sich Kolleginnen und Kollegen über Fragen der Datenmodellierung und Semantic-Web-Aktivitäten austauschen. Darüber hinaus gibt es ein regionales Treffen, das dem Austausch von Vertretern aus dem hbz mit den Verbundbibliotheken über Aktivitäten in diesem Bereich dient.

Auch sucht das hbz regelmäßig den Erfahrungsaustausch in Arbeitstreffen und auf Tagungen, auf Mailinglisten und insbesondere durch die Ausrichtung der Konferenz *Semantic Web in Bibliotheken* (SWIB), die das hbz gemeinsam mit der Deutschen Zentralbibliothek für Wirtschaftswissenschaften (ZBW) im Jahr 2010 zum zweiten Mal veranstaltete.⁵⁸ Auf nationaler Ebene koordiniert das hbz seine Projekte zudem mit der DNB und arbeitet im Projekt *Resolving- und Lookup-Dienst fuer bibliothekarische Identifier in culturegraph.org* eng mit der DNB zusammen.⁵⁹

International engagiert sich das hbz vor allem in der *Working Group on Open Bibliographic Data* der Open Knowledge Foundation, deren Koordination der Autor dieses Artikels übernommen hat.⁶⁰ Seit Mai 2010 gibt es beim W3C eine *Library Linked Data Incubator Group*⁶¹, in der das hbz ebenfalls aktiv ist.

7.2 lobid.org

lobid.org ist die (zurzeit noch experimentelle) Plattform zur Entwicklung von Linked-Open-Data-Diensten des hbz. Das Akronym 'lobid' steht für "Linking Open Bibliographic Data". Bisher beinhaltet die Plattform zwei Dienste, die im Folgenden knapp beschrieben werden sollen.

sprengen, daher sei an dieser Stelle nur auf <http://www.w3.org/TR/cooluris/#solutions> verwiesen.

⁵⁸ Siehe <http://swib.org>.

⁵⁹ Für weitere Informationen zu dem Projekt siehe <https://wiki1.hbz-nrw.de/display/SEM/Resolving-+und+Lookup-Dienst+fuer+bibliothekarische+Identifier+in+culturegraph.org>

⁶⁰ Für mehr Informationen siehe die Wiki-Seite: <http://wiki.okfn.org/wg/bibliography>.

⁶¹ Siehe <http://www.w3.org/2005/Incubator/lld/>.

7.2.1 lobid-resources

lobid-resources ist ein Dienst, der Beschreibungen bibliographischer Ressourcen (Bücher, Artikel etc.) als Linked Open Data bereitstellt. Zurzeit finden sich in lobid-resources die bisher freigegebenen Daten aus dem hbz-Verbundkatalog⁶² als Linked Open Data. Dazu werden die offenen Daten, die bereits in einem „RDFMAB“-Format⁶³ vorliegen, unter Nutzung verschiedener Vokabulare, v.a. aber der Bibliographic Ontology (Bibo)⁶⁴, in RDF umgewandelt.⁶⁵

Die Daten sind für Entwickler/innen über einen SPARQL-Endpoint abfragbar.⁶⁶ Eine Suchschnittstelle für Endnutzer/innen befindet sich in der Entwicklung. Außerdem arbeitet das hbz daran, die Daten als Linked Open Data auch unter Nutzung des RDA-Vokabulars⁶⁷ zu publizieren.

7.2.2 lobid-organisations

Es sind nicht nur die Daten zur Beschreibung bibliographischer Ressourcen, die für Bibliotheken und verwandte Institutionen relevant sind. Beschreibungen von einzelnen Institutionen sowie auch von Sammlungen spielen eine ebenso wichtige Rolle. In Deutschland werden seit über 100 Jahren Bibliothekssigel vergeben und Adresdaten sowie andere Informationen über Bibliotheken gesammelt.⁶⁸ Es ist also nur konsequent auch Informationen über bibliothekarische Institutionen und bestenfalls auch Sammlungsbeschreibungen in RDF zu repräsentieren und als Linked Open Data im Web verfügbar zu machen. lobid.org verfolgt dieses Ziel mit lobid-organisations, dem Organisationsdienst von lobid.org und mit der geplanten Etablierung des Dienstes lobid-collections.

Institutionen-URIs sind beispielweise die Grundlage dafür, Bestandsangaben in RDF auszudrücken oder RDF-Aussagen über Öffnungszeiten, Zugriffsmöglichkeiten, Sammelschwerpunkte usw. zu tätigen. Aus diesem Grund hat das hbz begonnen, URIs für Institutionen zu prägen und Informationen über sie bereitzustellen. lobid-organisations prägt für bibliothekarische Institutionen URIs der Form <http://lobid.org/organisation/{ISIL/MARC Organization Code}>. Zum Beispiel hat die Deutsche Nationalbibliothek die ISIL "DE-101" und damit den URI <http://lobid.org/organisation/DE-101>. Die lobid-Institutions-URIs leiten auf Webseiten⁶⁹ mit rudimentären Informationen (ISIL, Adresse, Telefon, E-Mail, ...) über die jeweilige Institution weiter – in menschenlesbarem HTML mit eingebettetem maschinenlesbarem RDF.

⁶² Siehe <https://wiki1.hbz-nrw.de/display/SEM/Aktuelle+Open-Data-Exporte>.

⁶³ Siehe zur Dokumentation dieser RDF-Serialisierung für ISO 2709-basierte bibliografische Formate: <https://wiki1.hbz-nrw.de/pages/viewpage.action?pageId=1802465>.

⁶⁴ <http://bibliontology.com/>

⁶⁵ Eine Dokumentation dieses Transformationsprozesses findet sich hier: <https://wiki1.hbz-nrw.de/display/SEM/Converting+the+Open+Data+from+the+hbz+to+BIBO>.

⁶⁶ <http://lobid.org/sparql/>

⁶⁷ <http://metadataregistry.org/rdabrowse.htm>

⁶⁸ Vgl. Andreas M. Heise (2004): Gedruckt und online - Bemerkungen über Sigelverzeichnisse und Bibliothekssigel anlässlich des 100. Geburtstags des früheren Auskunftsbüros der Deutschen Bibliotheken. Zeitschrift für Bibliothekswesen und Bibliographie (ZfBB) 51, Nr. 5, S. 305-315.

⁶⁹ Beim Beispiel DNB ist dies die Seite <http://lobid.org/organisation/DE-101/about.html>.

Somit dient lobid-organisations als ein internationales Adressverzeichnis von Bibliotheken und verwandten Organisationen, das Linked-Data-Standards folgt. Momentan sind mehr als 36.000 Institutionen verzeichnet, davon 6.827 aus Deutschland und 26.257 aus den USA. Die Datengrundlage stammt zurzeit aus dem Online-Sigelverzeichnis⁷⁰ und der MARC Organization Codes Database⁷¹. Aus diesem Grund ist die Datenbasis – im Gegensatz zu lobid-resources – nicht offen lizenziert. Zudem findet keine regelmäßige Aktualisierung der Daten statt, u.a. weil mit innovativen Möglichkeiten der Datenpflege jenseits von zentral geführten Verzeichnissen experimentiert werden soll. Perspektivisch sollen nämlich Organisations- und Sammlungsbeschreibungen dezentral auf den Webseiten der Bibliotheken publiziert werden und von lobid.org nachträglich aggregiert und als Linked Open Data bereitgestellt werden.⁷²

⁷⁰ <http://dispatch.opac.ddb.de/DB=1.2/SRT=YOP/>

⁷¹ Siehe <http://www.loc.gov/marc/organizations/>.

⁷² Für weitere Informationen zu diesem Projekt siehe <https://wiki1.hbz-nrw.de/display/SEM/Aggregation+dezentraler+Organisations-+und+Sammlungsbeschreibungen>.

8 Quellen

Für alle Webressourcen gilt: zuletzt eingesehen am 15.7.2010.

Berners-Lee, Tim / Hendler, James / Lassila, Ora (2001): The Semantic Web, *Scientific American* 284, no. 5: 34. Frei zugänglich unter <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>.

Borst, Timo / Neubert, Joachim (2009): Case Study: Publishing STW Thesaurus for Economics as Linked Open Data, <http://www.w3.org/2001/sw/sweo/public/UseCases/ZBW/>.

Campbell, Lorna M. / MacNeill, Sheila (2010): The Semantic Web, Linked and Open Data. A Briefing Paper. Einsehbar unter http://wiki.cetis.ac.uk/images/1/1a/The_Semantic_Web.pdf.

Hauschke, Christian (2010): Permalinks für Katalogisate. Blogbeitrag einsehbar unter <http://infobib.de/blog/2009/10/27/permalinks-fur-katalogisate/>.

Heise, Andreas M. (2004): Gedruckt und online - Bemerkungen über Sigelverzeichnisse und Bibliothekssigel anlässlich des 100. Geburtstags des früheren Auskunftsbüros der Deutschen Bibliotheken. *Zeitschrift für Bibliothekswesen und Bibliographie (ZfBB)* 51, Nr. 5, S. 305-315.

Hillmann, Diane / Coyle, Karen / Phipps, John / Dunsire, Gordon: RDA Vocabularies: Process, outcome, Use. *D-Lib Magazine* 16, Nr. 1 (1, 2010). Einsehbar unter <http://www.dlib.org/dlib/january10/hillmann/01hillmann.html>.

Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (2010): Freigabe der Katalogdaten: Kölner Bibliotheken leisten Pionierarbeit. Einsehbar unter: <http://www.hbz-nrw.de/dokumentencenter/presse/pm/datenfreigabe>.

JISC (2010): One to Many; Many to One: The resource discovery taskforce vision. Einsehbar unter <http://ie-repository.jisc.ac.uk/475/>.

Pohl, Adrian (2010): Open Data im hbz-Verbund. : Was, Wie und Warum. Preprint online einsehbar u.a. unter <http://www.hbz-nrw.de/dokumentencenter/produkte/lod/>.

Pohl, Adrian / Ostrowski, Felix: ‚Linked Data‘ - und warum wir uns im hbz-Verbund damit beschäftigen. *B.I.T. Online* 13, 3, S. 259-268. Preprint online einsehbar u.a. unter http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/pohl_ostrowski_2010_linked-data.pdf.

Shotton, David (2009): Semantic publishing: the coming revolution in scientific journal publishing, *Learned Publishing* 22 (April 2009): 85-94. DOI: 10.1087/2009202.

Simonite, Tom (2010): Wikipedia to Add Meaning to Its Pages. Einsehbar unter <http://www.technologyreview.com/web/25728/?a=f>.

Styles, Rob / Ayers, Danny / Shapir, Nadeem (2008): Semantic MARC, MARC21

and the Semantic Web. Einsehbar unter
<http://events.linkeddata.org/ldow2008/papers/02-styles-ayers-semantic-marc.pdf>.

W3C (2008): RDFa Primer. Bridging the Human and Data Webs. Einsehbar unter
<http://www.w3.org/TR/xhtml-rdfa-primer/>.

Wikipedia, The Free Encyclopedia (2010): Artikel "URI scheme".
Bearbeitungsstand: 9. Juli. URL: http://en.wikipedia.org/w/index.php?title=URI_scheme&oldid=372490112.



Der Text ist unter einer Creative-Commons-Namensnennung-Lizenz 3.0
Deutschland veröffentlicht: <http://creativecommons.org/licenses/by/3.0/de/>.

Abbildung 2 steht unter einer Creative Commons BY-SA-Lizenz:
<http://creativecommons.org/licenses/by-sa/3.0/>.