

Linked Open Data in der Bibliothekswelt: Grundlagen und Überblick

Adrian Pohl
Patrick Danowski

Inhalt

Linked Open Data in der Bibliothekswelt: Grundlagen und Überblick	1
1 Einleitung	3
2 Konzepte und Entwicklungen	6
2.1 Semantic Web & Linked Data	6
2.2 Open Everything	9
2.3 Open Everywhere	10
2.4 Offene bibliographische Daten	12
2.4.1 Begriffsklärung	12
2.4.2 Die Geburtsstunde von Open Data in der Bibliothekswelt	13
2.4.3 Erste Linked-Library-Data-Experimente	14
2.4.4 Die WorldCat-Metadaten-Kontroverse	15
2.4.5 Erste Linked/Open-Data-Aktivitäten in der deutschsprachigen Bibliothekswelt	15
2.4.6 Arbeitsgruppen und Empfehlungen	16
2.4.7 Offene bibliographische Daten von nicht-bibliothekarischen Organisationen	18
2.4.8 Übersicht der Entwicklungen von 2007 bis 2012	19
3 Der Linked-Open-Data-Stack	22
3.1 HTTP	24
3.2 Uniform Resource Identifiers (URIs)	24
3.3 RDF	25
3.4 Ontologien/Vokabulare	28
3.5 SPARQL	28
3.6 Speicherung und Publikation	29
3.6.1 RDF-Dateien publizieren	29
3.6.2 Content Negotiation und RDF	30
3.6.3 RDF in Echtzeit generieren	31
3.6.4 Triple Store mit SPARQL-Endpoint	31
3.6.5 RDFa – Linked Data als Markup von Webseiten	32
4 Open Data und Lizenzierung	33
5 Herausforderungen beim Aufbau eines Linked-Open-Data-Ökosystems	37
5.1 Provenienzzangaben	37
5.2 Mapping und Konvertierung von Altdaten	38
5.3 Schnittstellen und Datensynchronisierung	39

5.4	Linked-Data-fähige Bibliothekssysteme	39
	Offenheit als Standard.....	41
5.5	Die bibliographische Allmende aufbauen	41
6	Quellenverzeichnis.....	42

1 Einleitung

Etwa fünf Jahre sind vergangen seitdem erste bibliothekarische Einrichtungen mit Linked-Data-Technologien experimentiert haben und erste bibliothekarische Daten unter offenen Lizenzen veröffentlicht wurden. Es hat sich eine Menge getan in dieser Zeit. Eine Menge Akteure in der Bibliothekswelt – darunter wichtige Institutionen wie OCLC, die Library of Congress und die Deutsche Nationalbibliothek – haben Linked-Data-Initiativen gestartet. Gleichwohl lässt sich konstatieren, dass die Entwicklung von Linked Open Data in Bibliotheken immer noch in ihren Anfängen ist.

Bedenken, dass es sich bei Linked Data in der Bibliothekswelt bloß um einen Hype handelt, erwiesen sich allerdings spätestens im Oktober 2011 als grundlos. Zu diesem Zeitpunkt gab die Library of Congress die Grundzüge ihrer Initiative „Bibliographic Framework for the Digital Age“ zur Ablösung der überkommenen Standards MARC 21 und Z39.50 bekannt. Darin heißt es, dass die zu entwickelnde Infrastruktur sich auf „das Web, Linked-Data-Prinzipien und -Mechanismen und das Resource Description Framework (RDF) als ein grundlegendes Datenmodell“ (Library of Congress, 2011) fokussieren würde.¹ Im allgemeinen Webkontext haben Facebook mit dem OpenGraph Protocol² und Google mit dem Knowledge Graph³ gezeigt, dass verlinkte strukturierte – allerdings hier nicht unbedingt offene – Daten eine zentrale Rolle im Web spielen werden.

Allein die zunehmende Verbreitung und wachsende Akzeptanz sind sicher gute Gründe, sich mit Linked Open Data auseinanderzusetzen, um diese laufende Entwicklung zu verstehen. Darüber hinaus stellen sich Fragen nach dem Hintergrund dieser Entwicklung: Welche Vorteile verspricht Linked Data Bibliotheken? Welche Motivation treibt Individuen und Organisationen an, Ressourcen in Linked-Open-Data-Projekte zu stecken? Einige wesentliche Punkte seien hier kurz aufgeführt.

Auffindbarkeit der Daten: Bibliothekskataloge sind bis heute häufig Teil des ‚Deep Webs‘, d.h. sie können nicht von Suchmaschinen gesammelt und indiziert werden und sind damit über den gängigen Einstieg in das Web – eine

¹ Im englischen Original heißt es: “The new bibliographic framework project will be focused on the Web environment, Linked Data principles and mechanisms, and the Resource Description Framework (RDF) as a basic data model.”

² The Open Graph protocol: <http://ogp.me/>

³ Introducing the Knowledge Graph: things, not strings
<http://googleblog.blogspot.co.uk/2012/05/introducing-knowledge-graph-things-not.html>

Suche mit Google oder einer anderen Suchmaschine – nicht auffindbar. Linked Data sind – wie jede HTML-Seite auch – Teil des Webs und können problemlos gecrawlt und indexiert werden.

Verlinkbarkeit: Trefferlisten und Einzeltrefferanzeigen eines OPACs lassen sich häufig nicht mit anderen teilen, weil URLs sessionbasiert sind und damit nur für kurze Zeit funktionieren. Linked Data ändert dies und macht die Informationen verlinkbar.

Flexibles Datenmodell: RDF ist ein flexibles Datenmodell. Im Gegensatz zu MARC kann es beliebig erweitert und lokal ergänzt werden.

Interoperabilität: MARC 21 und Z39.50 sind opake Standards, die nur in der Bibliothekswelt bekannt sind und von wenigen Entwicklern angewendet werden. Linked Data dagegen basiert auf offenen Webstandards, die über das gesamte Web verbreitet von Menschen benutzt und verstanden werden. Werden Bibliotheksdaten im Web als Linked Open Data angeboten, können automatisch viele Menschen und Dienste diese weiterverwenden, und es können nicht nur bibliothekarisch vorgeprägte Personen damit etwas anfangen. Dies bedeutet auch, dass es leichter wird, Entwickler als Quereinsteiger in Bibliotheksprojekte zu finden, weil sie sich nicht zunächst mühsam mit bibliothekarischen Formaten vertraut machen müssen, sondern die verwendeten Standards sind ihnen bekannt.

Weiterverwendung: Interoperabilität allein reicht oft nicht aus, um Daten optimal nutzbar zu machen. Hier kommt das „Open“ in Linked Open Data ins Spiel, das eben sicherstellt, dass die Daten auch von Dritten einfach weiterverwendet werden können, indem sie mit offenen Lizenzen versehen werden (vgl. Abschnitt 4).

Bibliotheken können von diesen Weiterverwendungsmöglichkeiten und der Interoperabilität von Linked Data doppelt profitieren:

1. als *Konsumenten*, die Daten aus anderen Kontexten verlinken, in ihre Angebote integrieren und diese somit anreichern;⁴
2. als *Produzenten*, die Daten zur Verlinkung und Weiterverwendung publizieren und damit den Wert der Daten und das Prestige der Institution steigern.

⁴ Vgl. hierzu den Beitrag von Pascal Christoph in diesem Band. Lins/Becker zeigen in ihrem Artikel sehr schön, welchen Nutzen offene Daten für die Erstellung einer Bibliographie bringen können, indem sie beschreiben, wie Open-Data-Publikationen im Kontext eines Projekts nachgenutzt werden.

Die Bereitstellung von Normdaten als Linked Open Data sei hier als ein Beispiel genannt. Es sind insbesondere Normdaten, deren Einsatz in nicht-bibliothekarischen Projekten und Diensten einen enormen Gewinn darstellen kann und gleichzeitig Nutzer zu den Angeboten von Bibliotheken führen können. Bereits ohne Linked Open Data hat die deutschsprachige Wikipedia einen Anwendungsfall aufgezeigt, wie Normdateien auch außerhalb von Bibliotheken sinnvoll verwendet werden können. So wird die Personennormdatei (PND), die inzwischen in die Gemeinsame Normdatei (GND) aufgegangen ist, seit 2005 verwendet, um weiterführende Literatur aus Wikipedia-Artikeln zu verlinken⁵. Durch das Vorliegen von Normdateien als Linked Open Data kann ihre Weiterverwendung in anderen Projekten und für weitere Zwecke ermöglicht werden. Viele Normdatenberesteller haben dies bereits verstanden und entsprechend gehandelt. Inzwischen liegen neben den Library of Congress Subject Headings (LCSH), auch die GND und weitere Normdateien als Linked Data vor. Ebenso wurden die Ergebnisse des Projekts VIAF als Linked Data veröffentlicht und unter der ODC-BY Lizenz freigegeben⁶. Bruno Racine, Präsident der Bibliothèque nationale de France, sagte zur Freigabe der VIAF-Daten: „Wir erwarten, dass diese starke Erweiterung des Zugangs zu VIAF die Mehrsprachigkeit sowie die Entstehung neuer Dienste außerhalb der Bibliothekswelt, einschließlich der Unterstützung von Data Mining, Verwaltung geistiger Eigentumsrechte usw. fördern wird.“⁷

Interessante Nutzungsmöglichkeiten bietet Linked Data auch im Bereich der Forschungsdaten, wo sich Bibliotheken zunehmend engagieren. Wir verweisen hier auf den Beitrag von Dominique Ritze, Kai Eckert und Magnus Pfeffer in diesem Sammelband.

Im Folgenden möchten wir zunächst die grundlegenden Konzepte und geschichtlichen Entwicklungen von Linked Data und Open Data näher darstellen, wobei wir den Fokus auf die Bibliothekswelt legen. Anschließend werden die technischen Grundlagen von *Linked Data* erläutert, und sodann die rechtlichen Rahmenbedingungen einer *offenen* Lizenzierung skizziert. Abschließend werden einige Herausforderungen genannt, die sich in den nächsten Jahren auf dem Weg zu einem Linked-Open-Data-Ökosystem für Bibliotheken stellen.

⁵ Vgl. Hengel/Pfeifer (2005), wo die Autorinnen bereits davon sprechen, dass „[m]it der Einführung von Webtechnologien und -standards in die Normdateien ... die Normdateien zur Basis eines semantischen Netzes für Recherche und Retrieval im Internet werden [können].“ Sowie Danowski/Pfeifer (2007). Siehe auch den Beitrag von Geipel et al. in diesem Sammelband.

⁶ OCLC (2012a).

⁷ Ebd. im Original heißt es: “We expect that this broader opening of access to VIAF will encourage multilingualism and the creation of new services beyond the library world, including for data mining, intellectual property rights management, etc”.

2 Konzepte und Entwicklungen

Dieser Abschnitt möchte zunächst die geschichtlichen Wurzeln von Linked Open Data in Bibliotheken aufzeigen und erste Erläuterungen der zugrundeliegenden Konzepte geben. Hierzu werden wir zunächst auf die Entwicklungen des Semantic Web, von Linked Data und Open Data eingehen. Danach wird der Rahmen beschrieben, vor dem die Entwicklung zu größerer Offenheit in der Bibliothekswelt stattfindet – zum einen im Hinblick auf die Dinge, die geöffnet werden, und zum anderen im Hinblick auf konkrete Open-Initiativen und -Bewegungen. Am Ende wird ein geschichtlicher Abriss der wichtigsten Entwicklungen im Hinblick auf Linked/Open Data in Bibliotheken gegeben.

2.1 Semantic Web & Linked Data

Häufig werden die Ausdrücke „Semantic Web“ und „Linked Data“ synonym verwendet und in der Tat sind die dahinterliegenden Konzepte und Aktivitäten eng miteinander verwoben. Gemeinsam ist beiden, dass die damit verbundenen Entwicklungen maßgeblich vom World Wide Web Consortium (W3C) – dem Standardisierungsgremium für die dem World Wide Web (WWW) zugrundeliegenden Technikstandards – und seinem Vorsitzenden Tim Berners-Lee, dem Erfinder des WWW, vorangetrieben werden. Schaut man sich den Text auf der offiziellen Semantic-Web-Webseite des W3C an, wird klar, wie eng die verschiedenen Begriffe verzahnt sind:

*"Das W3C hilft, in Ergänzung zum klassischen ‚Web der Dokumente‘ einen Technologiestack aufzubauen, der das ‚Web der Daten‘ stützt – solcher Daten, die man auch in Datenbanken findet. Das höchste Ziel ist, Computern zu ermöglichen, mehr nützliche Aufgaben zu erledigen, und Systeme zu entwickeln, die vertrauliche, netzbasierte Interaktionen unterstützen. Der Ausdruck ‚Semantic Web‘ bezieht sich auf die W3C-Vision eines Netzes verlinkter Daten [orig.: ‚Web of linked data‘]. Semantic-Web-Technologien ermöglichen es Menschen, Datenspeicher im Web aufzubauen, Vokabulare zu erstellen und Regeln für die Handhabung von Daten zu verfassen. Verlinkte Daten (orig.: ‚linked data‘) werden ermöglicht durch Technologien wie RDF, SPARQL, OWL und SKOS."*⁸

⁸ Semantic Web – W3C <http://www.w3.org/standards/semanticweb/>, im Original: „In addition to the classic ‘Web of documents’ W3C is helping to build a technology stack to support a ‘Web of data,’ the sort of data you find in databases. The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network. The term ‘Semantic Web’ refers to W3C’s vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF,

Anfangs war allein vom Semantic Web die Rede. Die Idee des *Semantic Web* wurde erstmals in Berners-Lee et al. (2001) (in dem der Ausdruck ‚linked data‘ nicht ein einziges Mal auftaucht) der breiten Öffentlichkeit vorgestellt. Der grundsätzliche Ansatz war, das zu dieser Zeit seit zehn Jahren bestehende World Wide Web als Basis für die Publikation und Vernetzung strukturierter Daten zu nutzen. Die in diesem Artikel dargestellte Vision orientiert sich noch stark an Ansätzen und Zielen der Künstlichen-Intelligenz-Forschung, die den Hintergrund für die Entwicklung des Semantic Web bildet. So gehen die Autoren in ihrem Artikel recht weit, wenn sie eine Zukunft skizzieren, in der Semantic-Web-Agenten uns ständig vollautomatisiert bei alltäglichen Aufgaben unterstützen – beispielsweise bei der Recherche nach einem passenden Arzt in einem nahegelegenen Krankenhaus, inklusive Verabreden eines Behandlungstermins.

Die Idee, alle Inhalte des World Wide Web semantisch angereichert darzustellen, stellte sich jedoch schnell als nur schwer zu realisierender Wunsch heraus, und der Hype um das Semantic Web verebbte in den Jahren nach 2001. Auch heute sind wir – nach mehr als zehn Jahren – von der in Berners-Lee et al. (2001) dargestellten Welt noch weit entfernt. Allerdings muss man sagen, dass die Zukunft mittlerweile in der Regel weniger utopisch dargestellt wird.

Mit „Linked Data“ wurde 2006 ein neues Etikett geschaffen, um die Aktivitäten im Bereich Semantic Web wieder neu zu beleben und zu popularisieren. Dabei wurde auf einen pragmatischen Ansatz gesetzt, der die Publikation und Verlinkung strukturierter Daten in den Vordergrund stellte, womit Aktivitäten automatisierter Semantic-Web-Agenten aus dem Fokus rückten.

Am 27. Juli 2006 publizierte Tim Berners-Lee eine persönliche Sicht, wie mit Hilfe von Semantic-Web-Technologie Linked Data realisiert werden kann (Berners-Lee, 2006). In diesem Text finden sich auch die weithin bekannten vier „Linked-Data-Prinzipien“ (vgl. Abschnitt 3). Berners-Lee fordert im Text insbesondere dazu auf, sich auf die Publikation von Daten als Linked Data zu konzentrieren, die bereits strukturiert vorliegen.

Es folgten verschiedene entsprechende Aktivitäten, Linked Data im Netz zu publizieren. Als wichtiger Meilenstein in der Frühgeschichte der Linked-Data-Bewegung kann sicherlich das Projekt DBpedia bezeichnet werden, das am 23. Januar 2007 online ging. In diesem Projekt werden strukturierte Daten aus der Wikipedia extrahiert und als Linked Data publiziert. Da die Wikipedia eigentlich aus Fließtext besteht, konzentriert man sich im Projekt auf Daten in Infoboxen, Kategorien oder Vorlagen (Templates), da diese strukturiert sind.

SPARQL, OWL, and SKOS.“ Übersetzung von A.P. Für eine Erläuterung der technischen Grundlagen von Linked Data siehe Abschnitt 3.

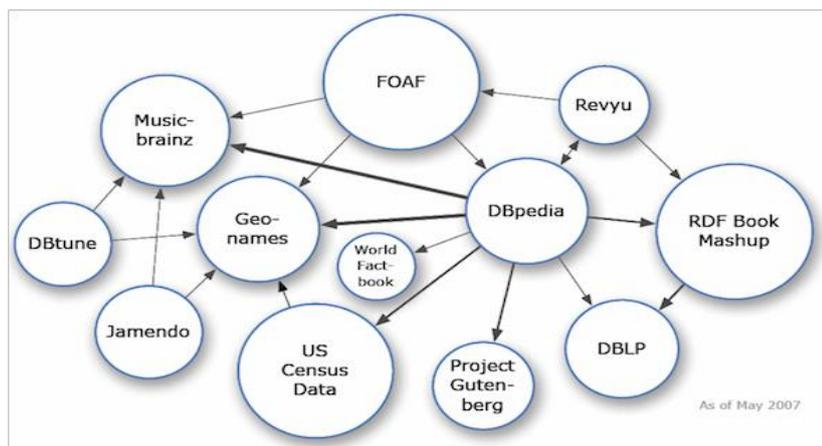


Abbildung 1: Erste Version des Linked Data Graphs

Seit 2007 hat die Anzahl der Daten und Datenquellen, die als Linked Data zur Verfügung gestellt werden, stetig zugenommen. Dies illustriert die *Linking Open Data Cloud*, die im Mai 2007 in der ersten Version veröffentlicht wurde (Abbildung 1).⁹

Die letzte Version des LOD-Cloud-Diagramms ist vom September 2011 (Abbildung 2): In vier Jahren ist die Cloud so sehr gewachsen, dass sie in einem Druckwerk wie diesem nicht mehr lesbar dargestellt werden kann.

⁹ Quelle: <http://lod-cloud.net/> von Richard Cyganiak und Anja Jentzsch, publiziert unter einer CC-BY-SA-Lizenz.

Im wissenschaftlichen Paper Stephens et al. (2005) wurde eine ähnliche Illustration veröffentlicht, die von Tim Berners-Lee in vielen Vorträgen benutzt wurde und als Vorbild für die bekannte Linked-Data-Cloud fungierte.

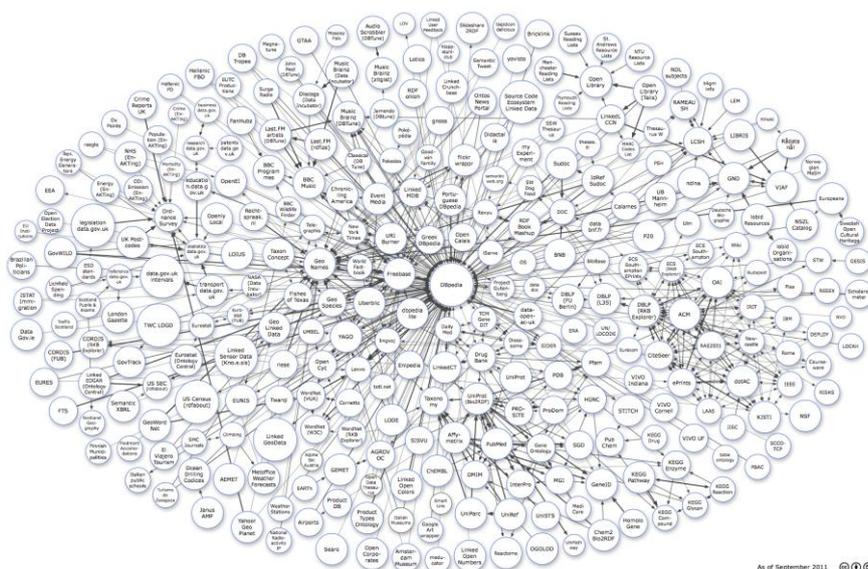


Abbildung 2: Letzte Version der Linked Data Cloud

Anfang 2013 sind bereits über 330 Projekte in der LOD-Cloud vertreten.¹⁰

2.2 Open Everything

Viele Akteure haben sich Offenheit¹¹ auf die Fahnen geschrieben und verschiedene Bewegungen lassen sich in diesem Bereich ausmachen. Manchmal scheint es für fast alle Forderungen nach und Konzepten für Offenheit zu geben, was sich in dem Ausdruck „Open Everything“ widerspiegelt. Bevor wir näher auf die historische Entwicklung eingehen, wollen wir genauer schauen, welche Arten menschlicher Artefakte unter entsprechenden Bedingungen und Lizenzen offen publiziert werden. Auch wenn sich diese Liste nicht auf digitale Artefakte beschränkt, ist Offenheit meist an Digitalisierbarkeit gekoppelt. In Abbildung 3] werden die unterschiedlichen

¹⁰ Für aktuelle Zahlen siehe die der graphischen Darstellung zugrundeliegende und von Richard Cyganiak und Anja Jentsch verwaltete „Linking Open Data Cloud“-Gruppe im Dataset-Verzeichnis *the Data Hub*: <http://thedatahub.org/group/locloud>

¹¹ In Abschnitt 4 wird näher erläutert, was unter „Offenheit“ in Bezug auf Daten und Inhalten zu verstehen ist. Eine allgemeine Kurzdefinition auch für andere Dinge lautet: Etwas ist offen, wenn es selbst bzw. seine Spezifikation/Baupläne a) frei zugänglich sind, b) in öffentlich dokumentierten Formaten vorliegen und c) es jedem erlaubt ist, es zu benutzen, weiterzuverwenden und -zuverbreiten.

Arten von Artefakten genannt, die frei oder offen sein können. Es lassen sich hier grob vier Arten unterscheiden:

- *Open Standards* sind Standards, die für alle Interessierten „leicht zugänglich, weiterentwickelbar und einsetzbar sind“¹², d.h. vor allem, dass die dazugehörigen Spezifikationen – idealerweise über das Internet – für jeden zugänglich und frei verwendbar sind.
- *Open Source Hardware* oder „Freie Hardware“ ist „Hardware, welche nach lizenzkostenfreien Bauplänen hergestellt wird“.¹³ Dabei geht es neben Computerhardware auch um die Hardware sonstiger Geräte und Maschinen.
- *Free/Open Content* sind sämtliche Produkte kreativer menschlicher Akte, die sich in digitaler Form abbilden lassen, und mit einer offenen Lizenz publiziert wurden.
- *Open Data* sind frei verfügbare Daten, die unter einer offenen Lizenz publiziert sind.

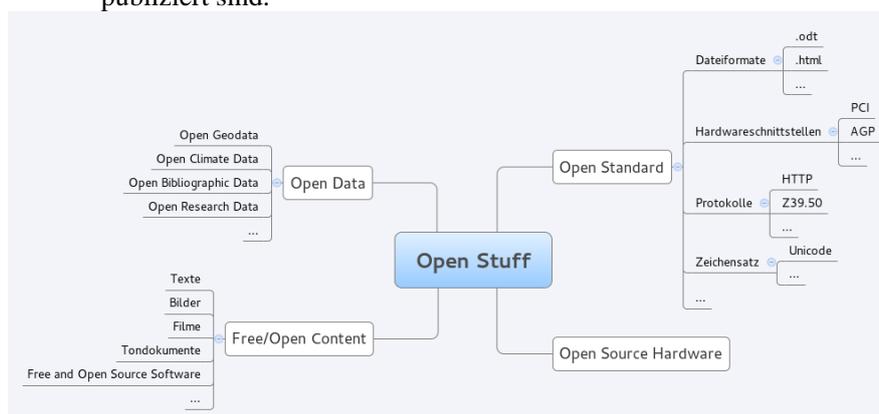


Abbildung 3: Arten freier Artefakte

2.3 Open Everywhere

Verschiedene *Open*-Bewegungen, die sich unter Etiketten wie „Open Culture“ oder „Open Knowledge“ subsumieren lassen, entstanden etwa zur gleichen Zeit wie die Semantic-Web-Bestrebungen. Zwischen 1999 und 2002 erwachte

¹² Seite „Offener Standard“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 16. Februar 2013, 00:09 UTC. URL: http://de.wikipedia.org/w/index.php?title=Offener_Standard&oldid=114259950 (Abgerufen: 27. Februar 2013, 18:14 UTC)

¹³ Seite „Freie Hardware“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 3. Februar 2013, 18:25 UTC. URL: http://de.wikipedia.org/w/index.php?title=Freie_Hardware&oldid=113756493 (Abgerufen: 27. Februar 2013, 18:15 UTC)

die *Open-Access-Bewegung* in der Wissenschaft¹⁴, die zum Ziel hat, dass „originäre wissenschaftliche Forschungsergebnisse ebenso wie Ursprungsdaten, Metadaten, Quellenmaterial, digitale Darstellungen von Bild- und Graphik-Material und wissenschaftliches Material in multimedialer Form“ im Internet frei zugänglich publiziert und mit entsprechenden Lizenzen versehen werden, die allen Nutzern gewähren, die Publikationen „zu kopieren, zu nutzen, zu verbreiten, zu übertragen und öffentlich wiederzugeben sowie Bearbeitungen davon zu erstellen und zu verbreiten“.¹⁵ In erster Linie wird Open Access jedoch mit Publikationen in wissenschaftlichen Zeitschriften in Verbindung gebracht. Die *Open-Science-Bewegung* setzt sich darüber hinaus dafür ein, dass neben Publikationen eben auch Forschungsdaten, bibliographische Daten, Source Code und andere Produkte wissenschaftlicher Arbeit unter offenen Lizenzen publiziert werden.

Die Wikipedia¹⁶ wurde 2001 gegründet. Ebenso *Creative Commons* (CC)¹⁷, das jedem verschiedene Lizenzmodelle bereitstellt, mit denen er seine Inhalte zur Nachnutzung im Netz bereitstellen kann. Beide Projekte lassen sich einer *Free-Culture-Bewegung* zurechnen, die etwa Anfang des 21. Jahrhunderts begann und durch den Creative Commons und seinen Gründer Lawrence Lessig, die Wikipedia-Community und anderen getragen wurde und wird.¹⁸ *Open Data* „steht für die Idee, Daten öffentlich frei verfügbar und nutzbar zu machen.“¹⁹ Open-Data-Konzepte gehen bereits auf die 50er Jahre zurück.²⁰ Der Ausdruck selbst fand – nach Yu/Robinson (2012), S. 189 – zum ersten Mal in den 1970er Jahren Verwendung. Allerdings kann von einer *Open-Data-Bewegung* frühestens seit Beginn des 21. Jahrhunderts die Rede sein. Open

¹⁴ Seite „Open Access“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 21. Februar 2013, 15:50 UTC. URL: http://de.wikipedia.org/w/index.php?title=Open_Access&oldid=113465354 (Abgerufen: 27. Februar 2013, 18:15 UTC) Kapitel Geschichte.

¹⁵ Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen (2003). Deutsche Version einsehbar unter http://oa.mpg.de/openaccess-berlin/Berliner_Erklaerung_dt_Version_07-2006.pdf. Die Berliner Erklärung wurde im Nachhinein stark verwässert, weil auch restriktiv lizenzierte Inhalte, die im Internet zum Download zur Verfügung stehen, als Open Access bezeichnet werden.

¹⁶ Seite „Wikipedia“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 25. Februar 2013, 03:32 UTC. URL: <http://de.wikipedia.org/w/index.php?title=Wikipedia&oldid=114651575> (Abgerufen: 27. Februar 2013, 18:18 UTC).

¹⁷ Seite „Creative Commons“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 20. Februar 2013, 18:07 UTC. URL: http://de.wikipedia.org/w/index.php?title=Creative_Commons&oldid=114463827 (Abgerufen: 27. Februar 2013, 18:19 UTC).

¹⁸ Vgl. auch das 2004 erschienene – und gleichzeitig online unter einer Creative-Commons-Lizenz veröffentlichte – Buch Lessig (2004).

¹⁹ Dossier Open Data der Bundeszentrale für politische Bildung <http://www.bpb.de/gesellschaft/medien/opendata/>.

²⁰ Vgl. etwa Keith G. Jefferys Kommentar zu einem Blogbeitrag von Peter Murray Rust: <http://blogs.ch.cam.ac.uk/pmr/2006/09/12/open-data-the-time-has-come/#comment-2236>

Data erstreckt sich dabei über verschiedene Bereiche. Große Fortschritte gab es vor allem bei den Daten aus Politik und Verwaltung (*Open Government Data*²¹), wo insbesondere zivilgesellschaftliche Einflussnahme einige Erfolge verzeichnen konnte. Dies zeigt sich nicht zuletzt an den verschiedenen Open-Data-Portalen, die in den letzten Jahren entstanden sind. Viele Aktivitäten gibt es in der Wissenschaft (*Open Science Data* und *Open Research Data*), aber eben auch in der Bibliothekswelt (*Open Library Data*), wo Bibliotheken, Bibliotheksverbände und Projekte wie Europeana im Bereich der freien bibliographischen Daten (*Open Bibliographic Data*) eine wichtige Rolle neben Verlagen, Wissenschaftlern und Universitäten spielen (siehe Abschnitt 2.4).

Offene Lizenzen wurden auch im Bereich Linked Data explizit eingefordert, nachdem Tim Berners-Lee die zentrale Bedeutung erkannte, die Open-Data-Prinzipien für die weitere Entwicklung der Linked-Data-Bewegung zukommen. Er begann 2009 für die Publikation von Daten unter offenen Lizenzen zu werben – wobei er sich insbesondere an öffentliche Einrichtungen wandte. Dies geschah erstmals am 4. Februar 2009 in dem TED Talk „Tim Berners-Lee on the next Web“²², wo Tim Berners-Lee den „Raw Data Now!“-Ruf anstimmte. 2010 ergänzte Berners-Lee dann die Linked-Data-Prinzipien um ein Fünf-Sterne-Schema für Linked Open Data.²³

2.4 Offene bibliographische Daten

In seinem einflussreichen Aufsatz „What is Web 2.0“ aus dem Jahre 2005 (O’Reilly, 2005) prophezeite Tim O’Reilly: "Data is the next Intel Inside" und stellte sogleich die Schlüsselfrage: Wem gehören die Daten? Diese Frage stellte sich in den folgenden Jahren auch in Bezug auf Bibliotheksdaten und bibliographische Daten im allgemeinen als eine zentrale Frage heraus und verschiedene Akteure auf der ganzen Welt begannen, sich damit auseinanderzusetzen. In diesem Abschnitt werden wichtige Meilensteine in der Entwicklung von Open Bibliographic Data genannt.

2.4.1 Begriffsklärung

Offene bibliographische Daten bzw. Open Bibliographic Data ist das Schlagwort, unter dem häufig die Freigabe von Bibliotheksdaten gefordert wird. Dies ist nicht falsch, jedoch beschränkt sich die Open-Bibliographic-Data-

²¹ Die – im Laufe der Zeit immer mehr eingeebnete – Unterscheidung zwischen „Open Government“ und „Open Data“ wird in Yu/Robinson (2012) detailliert betrachtet.

²² Tim Berners-Lee on the next Web (TED Talk)
http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html.

²³ Berners-Lee (2006).

Bewegung nicht auf die Bibliothekswelt. Es gibt eine Vielzahl von Akteuren, die bibliographische Daten produzieren und publizieren und damit auch freigeben können. Auch muss es sich bei offenen Daten aus einer Bibliothek nicht immer um bibliographische Daten handeln. So könnte es sich auch um Erwerbungs- und Ausleihstatistiken etc. handeln.²⁴ Es gibt zwar eine große Schnittmenge zwischen bibliographischen Daten und Bibliotheksdaten, sie sind aber nicht gleichzusetzen (vgl. Abbildung 4).

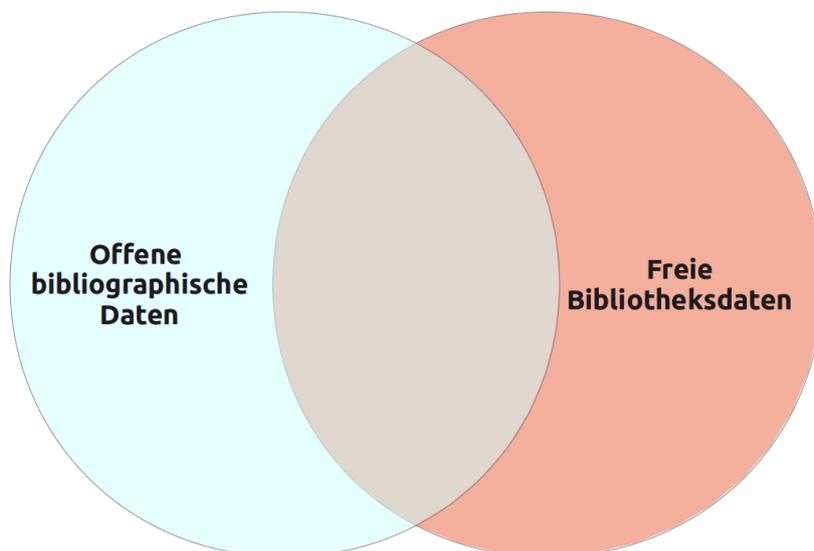


Abbildung 4: "Offene bibliographische Daten" vs "Freie Bibliotheksdaten"

2.4.2 Die Geburtsstunde von Open Data in der Bibliothekswelt

Einer der wesentlichen Auslöser, sich mit den Fragen auseinanderzusetzen, wer die Rechte an Katalogdaten hat und welche Nutzungsrechte daran anderen bereitgestellt werden sollten, war der Draft Report der *Working Group on Future of Bibliographic Control* der Library of Congress (Library of Congress Working Group on the Future of Bibliographic Control, 2007), der am 13. November 2007 veröffentlicht wurde. Dieser Bericht wurde insbesondere von der Open Knowledge Foundation stark kritisiert (Gray, 2007a), weil das Dokument der Library of Congress keinerlei Bezug auf die Nutzungsrechte von Katalogdaten enthält. Die Open Knowledge Foundation entwarf gemeinsam

²⁴ Die University of Huddersfield hat beispielsweise Ausleihdaten unter einer offenen Lizenz veröffentlicht. Siehe Pattern (2008). Vgl. die Empfehlungen zur Öffnung bibliographischer Daten im Anhang.

mit Aaron Swartz eine Antwort auf den Bericht, die von 150 Gruppen und Individuen unterzeichnet wurde.²⁵ Die bibliothekarische Gemeinschaft wurde also von außen auf dieses Thema aufmerksam gemacht und hat sich nicht aus eigener Motivation mit Open Data auseinandergesetzt.²⁶

Der Bericht der Library of Congress und die durch ihn hervorgerufene Kritik führten dazu, dass das Thema ab 2008 in der bibliothekarischen Gemeinschaft diskutiert wurde, so etwa in einem Workshop auf der Konferenz der European Library Automation Group (ELAG)²⁷ und in einer Podiumsdiskussion bei der IFLA-Konferenz in Quebec (vgl. Calhoun, 2008). Seitdem finden Diskussionen zu dieser Problematik regelmäßig statt. Bereits 2008 wendete die University of Huddersfield die Prinzipien der Offenheit auf Ausleih- und Empfehlungsdaten an und publizierte entsprechende Daten, die im Laufe von 13 Jahren gesammelt worden waren. Seitdem haben sich die Aktivitäten allerdings auf Titeldaten konzentriert und keine weitere Bibliothek ist diesem Beispiel gefolgt.

2.4.3 Erste Linked-Library-Data-Experimente

Im April 2008 ist der erste bibliothekarische Linked-Data-Service gestartet worden. Ed Summers – ein Entwickler an der Library of Congress – veröffentlichte eine inoffizielle Version der Library of Congress Subject Headings (LCSH) als Linked Data. Es handelte sich allerdings nicht um Open Data, da Ed Summers nicht die Rechte besaß, um die Daten offen zu lizenzieren. Dieser Service wurde dann auch am 18. Dezember 2008 auf Bitte der Library of Congress eingestellt.²⁸ Die Library of Congress veröffentlichte einige Monate später eine verbesserte, offizielle Version der LCSH unter <http://id.loc.gov/authorities/subjects>.

Ebenfalls 2008 wurden erstmals bibliothekarische Titeldaten als Linked Data veröffentlicht. Pionierarbeit leistete hier die Schwedische Nationalbibliothek mit der Publikation des LIBRIS-Verbundkatalogs als Linked Data.²⁹

²⁵ Open Knowledge Foundation Wiki: „Future of Bibliographic Control“ http://wiki.okfn.org/Future_Of_Bibliographic_Control, siehe auch Gray (2007b).

²⁶ Die Deutsche Nationalbibliothek (DNB) wurde schon vorher durch Wikipedia-Vertreter auf das Thema offene Lizenzierung der Katalogdaten gestoßen. Nach Informationen von Mathias Schindler (Wikimedia) wendeten sich Wikimedia-Vertreter erstmals 2006 an die DNB mit der Bitte um eine offene Lizenz. Die anschließende jahrelange Kommunikation zwischen DNB und Wikimedia verlief nicht-öffentlich, sodass hier auf keine entsprechenden Dokumente verwiesen werden kann.

²⁷ ELAG 2008 Rethinking Bibliographic Data“ <http://library.wur.nl/WebQuery/profielagws/full/117229>

²⁸ Vgl. Library of Congress Subject Headings. (2013, February 18). In: Wikipedia, The Free Encyclopedia. Abgerufen 18:38, 27. Februar 2013, von http://en.wikipedia.org/w/index.php?title=Library_of_Congress_Subject_Headings&oldid=538914921 Kapitel Data access.

²⁹ Malmsten (2008).

2.4.4 Die WorldCat-Metadaten-Kontroverse

Im November 2008 wurde die „Policy for Use and Transfer of WorldCat Records“ veröffentlicht, womit alle Hoffnungen auf die zügige Durchsetzung eines freien Lizenzmodells für den weltweit größten Verbundkatalog zerschlagen wurden. Stattdessen sah die erste Version der Policy vor, dass an jedem einzelnen Datensatz ein Copyright-Vermerk angebracht werde und die Nutzung der Daten bestimmten Restriktionen unterliege. Dieses Modell, wie auch sein Zustandekommen, wurde in der Folge von Bibliothekaren, Institutionen und anderen Akteuren in verschiedenen Punkten stark kritisiert.³⁰

Schließlich zog OCLC am 28. Juni 2009 diese Version der Nutzungsbedingungen zurück und setzte ein Gremium ein, das eine neue Policy erarbeiten sollte. Die neue, im April 2010 veröffentlichte Version mit dem Titel „WorldCat Rights and Responsibilities for the OCLC Cooperative“ (OCLC, 2010) war in vielerlei Hinsicht fortschrittlicher als der umstrittene erste Ansatz, allerdings wurde auch damit kein deutlicher Schritt in Richtung Open Data getan.³¹ Im Jahr 2012 machte OCLC dann einen deutlichen Schritt in Richtung Open Data als die in worldcat.org eingebetteten Linked Data unter einer ODC-BY-Lizenz gestellt wurden.³²

2.4.5 Erste Linked/Open-Data-Aktivitäten in der deutschsprachigen Bibliothekswelt

Das erste deutschsprachige Linked-Data-Projekt war die Publikation der ersten Version des Standard Thesaurus Wirtschaft (STW) als Linked Data durch die Zentralbibliothek für Wirtschaftswissenschaften (ZBW).³³ Im November 2009 veranstalteten ZBW und das Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz) erstmals die Konferenz *Semantic Web in Bibliotheken* (SWIB), die – obwohl sie sich primär an deutschsprachige Zuhörer wendete – von Anfang an auf internationales Interesse gestoßen ist. Patrick Danowski kündigte auf dieser Konferenz an, dass das CERN seine bibliographischen Daten unter der Creative Commons Public Domain Dedication (CC0) freigeben werde.³⁴ Dies wurde im Januar 2010 offiziell in die Tat umgesetzt und das CERN veröffentlichte als erste Bibliothek einen kompletten

³⁰ Für eine detaillierte Darstellung der Entwicklung und Kritik an der Policy bis zum Februar 2009 siehe Pohl (2009a).

³¹ Vgl. Pohl (2010).

³² OCLC (2012b).

³³ Vgl. Borst/Neubert (2009).

³⁴ Für eine Erläuterung der verschiedenen Open-Data-Lizenzen siehe Abschnitt 4.

Abzug der Daten (MARC) unter CC0 und der Public Domain Dedication Licence (PDDL). Am 12. März 2010 veröffentlichten mehrere Bibliotheken in Kooperation mit dem Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen (hbz) ihre Daten unter CC0. Auch diese Daten waren zunächst ein reiner Abzug der Katalogdaten. Das hbz kündigte aber an, die Daten für das Semantic Web aufzubereiten, was im Juli 2010 mit dem Linked-Open-Data-Dienst lobid.org realisiert wurde. Im Mai 2010 startete auch die Deutsche Nationalbibliothek (DNB) ihre Semantic-Web-Aktivitäten mit einem Linked-Data-Service für Normdaten, zur Lizenz gab es aber am Anfang von der DNB noch keine klare Aussage. Nachdem zunächst eigene Nutzungsbedingungen festgelegt wurden, hat man sich im Rahmen von Open-Data-Verlautbarungen der Conference of European National Librarians (CENL) (Conference of European National Librarians, 2011) für eine CC0 Public Domain Dedication entschieden, deren Nutzung seitdem auch stark von der DNB beworben wird (vgl. Richt, 2012). Im August 2010 startete die Universitätsbibliothek Mannheim den ersten Linked Data Service für bibliographische Titeldaten, der auch Daten aus dem SWB und HeBIS umfasste.³⁵ Parallel dazu publizierten mit den Universitätsbibliotheken in Konstanz und Tübingen die ersten beiden Bibliotheken im Südwestdeutschen Bibliotheksverbund (SWB) ihre Katalogdaten im MARC-Format unter einer offenen Lizenz.

2.4.6 Arbeitsgruppen und Empfehlungen

Im Laufe der Zeit haben sich einige Arbeitsgruppen und andere Kooperationen im Bereich Linked/Open Bibliographic Data inner- wie außerhalb der Bibliothekswelt gebildet.

Im Februar 2010 gründete die *Open Knowledge Foundation (OKFN)* eine *Arbeitsgruppe zu Open Bibliographic Data*. Es handelt sich dabei um einen losen Zusammenschluss von Wissenschaftlern, Technikern und Bibliothekaren, die nicht nur daran arbeiten, dass mehr bibliographische Daten freigegeben werden, sondern auch Software und Dienste für den Umgang mit offenen bibliographischen Daten entwickeln (siehe nächster Abschnitt). Die Arbeitsgruppe verfasste zudem die *Principles on Open Bibliographic Data*, die im Januar 2011 im englischen Original und fortan auch in verschiedenen Übersetzungen publiziert wurden.³⁶ Diese Prinzipien liefern sowohl eine Definition von bibliographischen Daten als auch von deren „Offenheit“.

³⁵ Vgl. E-Mail von Kai Eckert via INETBIB (09.08.2010) <http://www.ub.uni-dortmund.de/listen/inetbib/msg42859.html>.

³⁶ Coyle et al. (2011). Die deutsche Übersetzung „Prinzipien für Offene Bibliographische Daten“ findet sich im Anhang.

Eine wichtige Rolle bei der Förderung von Linked-Data-Standards spielte die *Library Linked Data Incubator Group*, die von Mai bis Oktober 2011 innerhalb des World Wide Web Consortiums (W3C) einen Bericht verfasste, der helfen sollte, „die globale Interoperabilität von Bibliotheksdaten im Web zu verbessern“ und Bereiche für eine zukünftige Zusammenarbeit zu identifizieren.³⁷ Zu diesem Zweck kamen in der Arbeitsgruppe verschiedene Fachpersonen zusammen, die bereits Linked-Data-Aktivitäten unternommen hatten.

Im deutschsprachigen Raum findet die Vernetzung im Bereich Linked Open Data seit April 2011 vor allem im Kontext der Arbeitsgruppe KIM (Kompetenzzentrum Interoperable Metadaten) im Rahmen der Deutschen Initiative für Netzwerkinformation (DINI) statt. Dort existieren Arbeitsgruppen zu verschiedenen Themen, wobei sich ein Großteil der Gruppen auch mit Linked-Open-Data-Standards auseinandersetzt. Die *Arbeitsgruppe Lizenzen* innerhalb der DINI KIM AG veröffentlichte im Oktober 2011 *Empfehlungen zur Öffnung bibliothekarischer Daten* (Gruppe Lizenzen der DINI-AG KIM, 2011).³⁸ Die *Gruppe Titeldaten* innerhalb der DINI-AG KIM arbeitet seit Januar 2012 an einer Empfehlung zur einheitlichen Nutzung bestehender RDF-Vokabulare für die Repräsentation von Titeldaten als Linked Data.³⁹

Im europäischen Kontext hat vor allem die Europeana – eine virtuelle Bibliothek, die das wissenschaftliche und kulturelle Erbe Europas in Form digitalisierter Bild-, Text-, Ton- und Video-Dateien zugänglich machen soll – eine Menge beigetragen zu einer Entwicklung der Bibliothekswelt in Richtung Offenheit und Linked Data. Ein Meilenstein dieser Entwicklung war sicherlich die Verabschiedung des *Europeana Data Exchange Agreement* im September 2011, in dem festgelegt wird, dass sämtliche in Europeana verwendeten Metadaten gemeinfrei sind, d.h. keinerlei Nutzungseinschränkungen unterliegen.⁴⁰

Auf internationaler Ebene bildeten sich Ende 2011 zwei internationale Arbeitsgruppen im Kontext von Linked Data und Semantic Web: Im September 2011 entstand eine *Linked Open Data Special Interest Group* innerhalb der *International Group of Ex Libris Users* (IGeLU)⁴¹ und im Dezember 2011 gründete sich schließlich die *IFLA Semantic Web Special Interest Group*.⁴²

³⁷ Vgl. die Charter der Library Linked Data Incubator Group: <http://www.w3.org/2005/Incubator/lld/charter>

³⁸ Diese finden sich gedruckt im Anhang.

³⁹ Der aktuelle Entwurf der Empfehlungen Der Gruppe Titeldaten innerhalb der DINI AG KIM findet sich unter <https://wiki.d-nb.de/x/cYMOB>.

⁴⁰ Europeana Foundation (2011).

⁴¹ Linked Open Data SIWG | IGeLU <http://igelu.org/special-interests/lod>.

⁴² Semantic Web Special Interest Group | IFLA <http://www.ifla.org/swsig>.

Zudem findet nach 2011 im Jahre 2013 der zweite *LOD-LAM-Summit*⁴³ statt, der internationale Akteure aus Bibliotheken, Archiven und Museen zusammenbringt, um sich über Projekte, Software, Lizenzmodelle und Methoden für eine Linked-Open-Data-Publikation in entsprechenden Einrichtungen auszutauschen. LOD-LAM steht für „Linked Open Data in Libraries, Archives and Museums“ und kann als ein loser Zusammenschluss internationaler Akteure in diesem Bereich verstanden werden, mit Schwerpunkt auf der angelsächsischen Welt. Auch auf einer Mailingliste⁴⁴ und unter dem Hashtag #lodlam⁴⁵ auf Twitter tauschen sich interessierte Personen zum Thema aus.

2.4.7 Offene bibliographische Daten von nicht-bibliothekarischen Organisationen

Es wurde bereits darauf hingewiesen, dass nicht nur Bibliotheken im Bereich Open Bibliographic Data aktiv sind. Auch von Wissenschaftlern und Verlagen werden wichtige Initiativen, Projekte und Dienste aufgebaut. Innerhalb der Open Knowledge Foundation arbeiteten seit 2011 vor allem Wissenschaftler und wissenschaftsnahe Entwickler am Aufbau einer offenen Software zum Erstellen und Teilen von Bibliographien. Das erste Softwareprojekt mit dem Namen *Bibliographica/openbiblio* wurde bereits 2011 eingestellt⁴⁶ und stattdessen die *BibServer*-Software weiterentwickelt, die bis heute stetig verbessert wird.⁴⁷

Im Dezember 2011 wurde mit der computerwissenschaftlichen Bibliographie *DBLP* eine umfangreiche, über viele Jahre gepflegte und von vielen Wissenschaftlerinnen und anderen Interessierten genutzte Bibliographie als Open Data freigegeben. Die Universität Trier und Schloss Dagstuhl - Leibniz-Zentrum für Informatik stellten die Daten unter eine ODC-BY-Lizenz (Ackermann, 2011).

Im März 2012 startete das Wikimedia-Projekt *Wikidata*.⁴⁸ In diesem Projekt werden strukturierte Daten aus der Wikipedia, wie Links zwischen Sprachversionen und Daten aus den Wikipedia-Infoboxen, in einer gemeinsamen Datenbasis für sämtliche Sprachversionen der Wikipedia zusammengeführt. Diese Datenbank ist – wie die Wikipedia selbst – frei editierbar, die Daten werden

⁴³ LODLAM – Linked Open Data in Libraries, Archives and Museums <http://lodlam.net/>

⁴⁴ Forum der Google Group Linked Open Data in Libraries, Archives, & Museums <https://groups.google.com/forum/?fromgroups#!forum/lod-lam>.

⁴⁵ Tweets mit dem hashtag #lodlam: <https://twitter.com/search?q=%23lodlam>.

⁴⁶ Beschreibung Bibliographica Projekt der Open Knowledge Foundation <http://openbiblio.net/p/bibliographica/>.

⁴⁷ Homepage von BibServer: <http://bibserver.org/>.

⁴⁸ Wikidata <http://www.wikidata.org/>.

unter einer CC0 Public Domain Dedication bereitgestellt. Es ist sehr wahrscheinlich, dass Wikidata in Zukunft eine wichtige Rolle im Web verlinkter Daten spielen wird, und für bibliothekarische Organisationen bestehen perspektivisch vielfältige Möglichkeiten, zu Wikidata beizutragen und Wikidata zu nutzen.

Mit *Nature Publishing* veröffentlichte im April 2012 erstmals ein Verlag bibliographische Daten als Linked Open Data (Nature Publishing Group, 2012). Die Linked-Open-Data-Plattform der Nature Publishing Group⁴⁹ wird beständig mit Informationen zu Artikeln aktualisiert und bietet verschiedene Möglichkeiten, auf die Daten zuzugreifen. Die Daten stehen unter einer CC0 Lizenz.

2.4.8 Übersicht der Entwicklungen von 2007 bis 2012

Die folgende tabellarische Übersicht zeigt in kondensierter Form die Entwicklungen im Bereich offene/verlinkte bibliographische Daten. Der Fokus mag – aufgrund des Hintergrunds der Autoren – auf bibliothekarischen Einrichtungen im deutschsprachigen Raum liegen, allerdings wird versucht, internationale Entwicklungen so weit es geht zu berücksichtigen.

Juli 2007	Launch der Beta-Version der <i>Open Library</i> ⁵⁰
13. November 2007	Veröffentlichung des Draft Reports der <i>Working Group on Future of Bibliographic Control</i> der Library of Congress
6. Dezember 2007	Kritik der Open Knowledge Foundation an dem Draft Report
April 2008	lsh.info. Inoffizielle Version der Library of Congress Subject Headings als Linked Data publiziert von Ed Summers. Eingestellt im Dezember 2008
2008	Schwedischer Verbundkatalog LIBRIS als Linked Data veröffentlicht
Dezember 2008	Bibliothek der University of Huddersfield veröffentlicht Ausleih- und Vorschlagsdaten unter CC0 ⁵¹
November 2008 bis Juni 2009	Die Veröffentlichung einer neuen restriktiven Policy für WorldCat-Daten durch OCLC führt zu einer langen und breiten Diskussion innerhalb der Bibliothekswelt. ⁵²
Februar 2009	Erste Version des Standard Thesaurus Wirtschaft (STW) von der ZBW als Linked Data veröffentlicht ⁵³

⁴⁹ Linked-Data-Plattform der Nature Publishing Group: <http://data.nature.com/>.

⁵⁰ Brantley (2007).

⁵¹ Pattern (2008).

⁵² Vgl. (Pohl, 2009a).

⁵³ Vgl. Borst/Neubert (2009).

28. Juni 2009	Die WorldCat-Metadaten-Policy wird offiziell zurückgezogen. ⁵⁴
November 2009	Erste SWIB-Konferenz „Semantic Web in Bibliotheken“ ⁵⁵
29. Januar 2010	CERN veröffentlicht Katalogdaten unter CC0. ⁵⁶
Februar 2010	Gründung der Working Group on Open Bibliographic Data innerhalb der OKFN
12. März 2010	Freigabe von Katalogdaten durch Kölner Bibliotheken und das Landesbibliothekszenrum Rheinland-Pfalz (LBZ) in Kooperation mit dem hbz ⁵⁷
Mai 2010	DNB startet Linked-Data-Service für Normdaten. ⁵⁸
Juli 2010	hbz startet lobid.org ⁵⁹
1. August 2010	„WorldCat Rights and Responsibilities for the OCLC Cooperative“ tritt in Kraft ⁶⁰
August 2010	Universitätsbibliothek Mannheim startet Linked-Data-Service für Titeldaten. ⁶¹
August 2010	Die ersten beiden SWB-Bibliotheken publizieren MARC-Katalogdaten als Open Data in Kooperation mit dem BSZ. ⁶²
23. August 2010	Als erste Nationalbibliothek in Europa veröffentlicht die British Library Katalogdaten unter CC0. ⁶³
November 2010	Start des culturegraph-Prototypen ⁶⁴
Januar 2011	Veröffentlichung der <i>Principles on Open Bibliographic Data</i> ⁶⁵
2./3. Juni 2011	1. LOD-LAM Summit in San Francisco ⁶⁶
September 2011	Gründung einer Linked Open Data Special Interest Group

⁵⁴ Vgl. Pohl (2009b).

⁵⁵ Homepage der 1. Semantic Web In Bibliotheken (SWIB) Konferenz: <http://www.swib09.de/>

⁵⁶ CERN Library (2010).

⁵⁷ Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen (2010).

⁵⁸ Vgl. E-Mail von Verena Binz an INETBIB “Linked Data Service der Deutschen Nationalbibliothek” <http://www.ub.uni-dortmund.de/listen/inetbib/msg42161.html>.

⁵⁹ Vgl. lobid.org? <http://blog.lobid.org/2010/07/lobidorg.html>.

⁶⁰ Vgl. History [WorldCat Rights and Responsibilities for the OCLC Cooperative] <http://www.oclc.org/us/en/worldcat/recorduse/policy/history.htm>.

⁶¹ Vgl. E-Mail von Kai Eckert via INETBIB (09.08.2010) <http://www.ub.uni-dortmund.de/listen/inetbib/msg42859.html>.

⁶² Vgl. <https://wiki.bsz-bw.de/doku.php?id=v-team:daten:openaccess:start&rev=1280922224>.

⁶³ British Library (2010).

⁶⁴ Deutsche Nationalbibliothek/Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen (2010).

⁶⁵ Pohl (2011a).

⁶⁶ Webseite des International Linked Open Data in Libraries Archives and Museums Summit June 2-3, 2011 San Francisco, CA, USA <http://lod-lam.net/summit/>.

	der International Group of Ex Libris Users (Igelu) ⁶⁷
September 2011	Linked-Data-Service der Bibliotheque Nationale de France unter offener Lizenz ⁶⁸ Europeana Data Exchange Agreement legt CC0 für Metadaten fest. ⁶⁹
21. September 2011	Schwedische Nationalbibliographie unter CC0 veröffentlicht ⁷⁰
29. September 2011	Die Conference of European National Librarians setzt auf mit CC0 auf Open Data ⁷¹
25. Oktober 2011	Der Abschlussbericht der W3C Library Linked Data Incubator Group wird veröffentlicht ⁷²
31. Oktober 2011	Library of Congress präsentiert Pläne für ein „Bibliographic Framework for the Digital Age“ ⁷³ .
3. November 2011	Das hbz veröffentlicht rechtlichen Open-Data-Leitfaden. ⁷⁴
17. November 2011	Empfehlungen der DINI AG KIM zur Öffnung bibliothekarischer Daten veröffentlicht ⁷⁵
9. Dezember 2011	DBLP als Open Data unter ODC-BY veröffentlicht ⁷⁶
Dezember 2011	Gründung einer Semantic Web Special Interest Group bei der IFLA
Dezember 2011	Linked-Open-Data-Service der spanischen Nationalbibliothek startet ⁷⁷
Dezember 2011	Publikation des B3Kat von BVB und KOBV als Open Data unter CC0 ⁷⁸
Januar 2012	DNB ergänzt Titeldaten und wechselt zu CC0. ⁷⁹

⁶⁷ Igelu: Background

Linked Open Data SIWG <http://igelu.org/special-interests/lod/background>.

⁶⁸ Vgl. Pohl (2011c).

⁶⁹ Europeana Foundation (2011).

⁷⁰ Malmsten (2011).

⁷¹ Conference of European National Librarians (2011).

⁷² W3C Library Linked Data Incubator Group (2011).

⁷³ Library of Congress (2011).

⁷⁴ Kreuzer (2011).

⁷⁵ DINI/KIM: DINI-AG KIM veröffentlicht Open-Data-Empfehlungen <https://wiki.d-nb.de/pages/viewpage.action?pageId=48765557> (siehe den Text der Empfehlungen im Anhang).

⁷⁶ Ackermann (2011).

⁷⁷ Vgl. Biblioteca Nacional de España und Ontology Engineering Group (2012).

⁷⁸ Siehe Pohl (2011c) und Kooperativer Bibliotheksverbund Berlin-Brandenburg (2011). Für Details siehe den Beitrag von Ceynowa et al. in diesem Sammelband.

⁷⁹ Mail von Julia Hauser an die DINI AG KIM Linked Library Data Mailingliste: „Die Deutsche Nationalbibliografie als Linked Open Data“ <http://lists.d-nb.de/pipermail/dini-ag-kim-llid/2012-January/000056.html>.

30.03.2012	Wikidata-Projekt zum Aufbau einer gemeinsamen Datenbasis für verschiedene Wikipedia-Sprachversionen startet ⁸⁰
Februar 2012	BibSoup gelauncht, das auf der Software BibServer basiert
April 2012	VIAF wird OCLC-Service und VIAF-Daten unter ODC-BY-Lizenz veröffentlicht ⁸¹
4. April 2012	Nature Publishing Group startet Linked-Open-Data-Dienst und veröffentlicht mehr als 400.000 Artikelbeschreibungen unter CC0. ⁸²
24. April 2012	Katalog der Harvard Library unter CC0 veröffentlicht ⁸³
Mai 2012	Freigabe der viaf.org-Daten unter ODC-BY-Lizenz ⁸⁴
20. Juni 2012	OCLC ergänzt Linked-Data-Markup unter offener Lizenz in worldcat.org ⁸⁵
September 2012	Die Schema Bib Extend Community Group nimmt ihre Arbeit auf. ⁸⁶
21. November 2012	Erster Entwurf des Bibframe-Datenmodells wird veröffentlicht. ⁸⁷

3 Der Linked-Open-Data-Stack

Die technischen – und seit 2010 auch die rechtlichen – Grundlagen von Linked Open Data werden im bereits erwähnten Dokument Berners-Lee (2006) erläutert. In diesem Abschnitt werden wir uns auf die technischen Grundlagen konzentrieren, die rechtlichen Aspekte werden in Open Data und Lizenzierung behandelt.

Die sogenannten Linked-Data-Prinzipien besagen:

1. Benutze URIs als Namen für Dinge.
2. Benutze HTTP-URIs, damit Menschen die Namen nachschlagen können.

⁸⁰ Schoneville (2012).

⁸¹ OCLC (2012a).

⁸² Nature Publishing Group (2012).

⁸³ Harvard Library (2012).

⁸⁴ Siehe Hickery (2012).

⁸⁵ OCLC (2012b).

⁸⁶ Vgl. Schema Bib Extend Community Group <http://www.w3.org/community/schemabibex/>.

⁸⁷ Library of Congress (2012).

3. Wenn jemand eine URI nachschlägt, liefere nützliche Informationen auf Basis der Standards (RDF, SPARQL).
4. Verlinke zu anderen URIs, so dass mehr Dinge entdeckt werden können.⁸⁸

Im Folgenden sollen diese Grundlagen näher erläutert werden. Abbildung 2 bietet einen Überblick über den „Linked-Open-Data-Stack“, d.h. über die Linked Open Data zugrundeliegenden Technologien. Die Basis-Technologien werden dabei im unteren Teil der Grafik angezeigt und darauf aufbauende Technologien jeweils darüber.

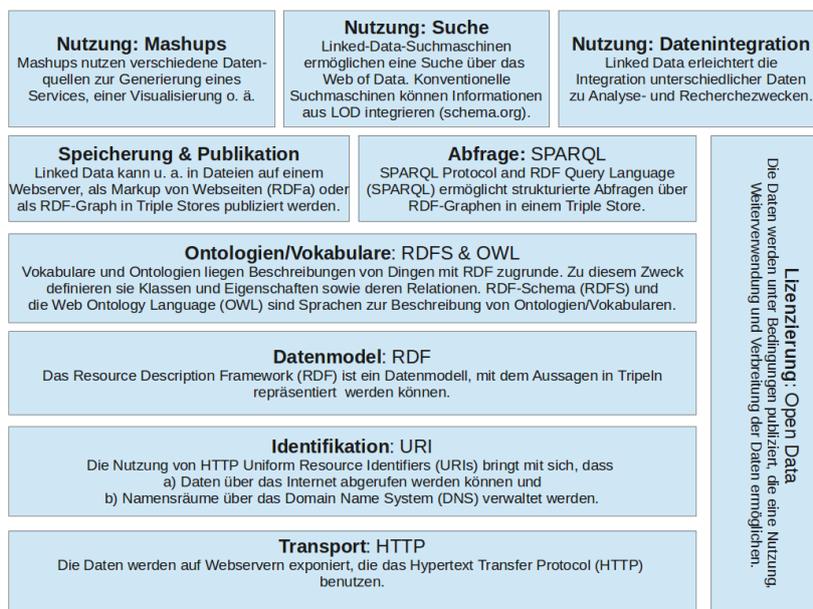


Abbildung 5: Der Linked-Open-Data Stack⁸⁹

In den folgenden Abschnitten werden die einzelnen Bestandteile im Detail betrachtet.

⁸⁸ Im Original: “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.” (Berners-Lee, 2006).

⁸⁹ Diese Illustration ist eine Übersetzung und leichte Anpassung des „Linked Open Data Stack“ von Tim Davies, siehe: <http://www.opendataimpacts.net/2011/05/whats-in-the-linked-open-data-stack/> Diese Abbildung ist – in Absprache mit Tim Davies – mit CC0 (<http://creativecommons.org/publicdomain/zero/1.0/>) lizenziert.

3.1 HTTP

Das Hypertext Transfer Protocol (HTTP) ist das Protokoll zur Datenübertragung, das dem World Wide Web zugrundeliegt. Linked Data und das Semantic Web basieren ebenfalls auf diesem Protokoll und sind aus diesem Grund als eine Erweiterung des WWW zu verstehen. Sämtliche Funktionen von HTTP können so auch für Linked Data genutzt werden.

3.2 Uniform Resource Identifiers (URIs)

Der Begriff Uniform Resource Identifier (URI) wurde 1994 von Tim-Berners Lee im RFC 1630 eingeführt.⁹⁰ Inzwischen ist der RFC 3986 die gültige Referenz für den URI.

Der URI-Syntax entsprechen verschiedene Typen von Identifikatoren, zum Beispiel:

- [doi:10.1371/journal.pbio.0020449.g001](https://doi.org/10.1371/journal.pbio.0020449.g001)
- [urn:nbn:de:gbv:089-3321752945](https://nbn-resolving.org/urn:nbn:de:gbv:089-3321752945)
- <mailto:pohl@hbz-nrw.de>
- http://de.wikipedia.org/wiki/Uniform_Resource_Identifier#Query

Der Anfangsteil der URI – „Schema“ genannt – gibt an, wie der URI interpretiert werden muss. In den Beispielen sind dies *doi*, *urn*, *mailto* und *http*.⁹¹

Es gibt also verschiedene Sorten von URIs, allerdings ist für Linked Data – dem zweiten Linked-Data-„Prinzip“ von Tim Berners-Lee entsprechend – die Verwendung von HTTP-URIs als Identifikatoren notwendig, weil sie bereits Bestandteil der Web-Architektur sind.

HTTP-URIs sind Webnutzern bereits bekannt, weil sie seit Beginn des Webs als Uniform Resource Locators (URLs) von Webseiten eingesetzt werden. Die vielleicht grundlegendste Erweiterung des WWW durch Linked Data besteht darin, dass HTTP-URIs nicht mehr allein als Adressen von abrufbaren Webressourcen (HTML-Seiten, Dateien, Diensten) benutzt werden, sondern auch zur Bezugnahme auf in Raum und Zeit vorhandene Dinge wie Personen, Bauwerke, Ortschaften, Bücher etc. verwendet werden (siehe das Beispiel in Abbildung 6).

⁹⁰ Dort wird das Akronym „URI“ aber noch mit „Universal Resource Identifier“ aufgelöst.

⁹¹ Während *urn* und *http* offizielle URI-Schema sind, die bei IANA registriert ist, ist *doi* zwar weit verbreitet jedoch nur inoffiziell ein URI.

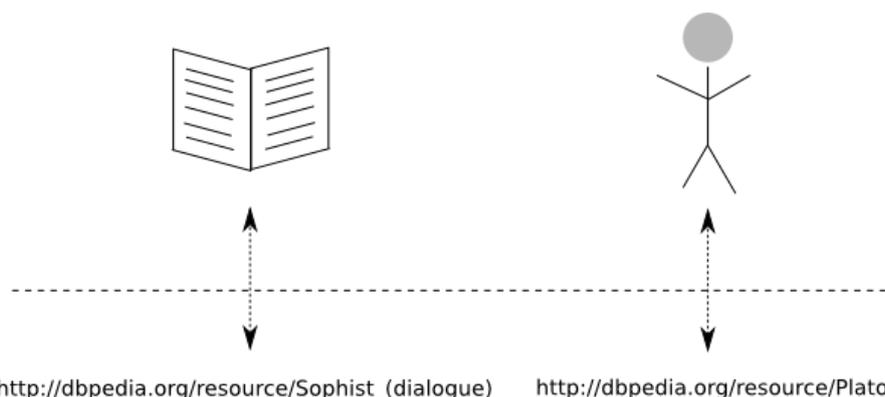


Abbildung 6: HTTP-URIs als Identifikatoren raumzeitlicher Dinge (Dialogtext, Autor)⁹²

3.3 RDF

Die wichtigste Ergänzung zu bestehenden Web-Standards im Kontext von Linked Data ist das Resource Description Framework (RDF), 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.⁹³

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 antiken Philosophen Bezug nimmt und ‚Sophistes‘ auf ein 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:

⁹² Diese, die folgende Abbildung sowie einige Formulierungen in diesem und den drei folgenden Abschnitten wurden übernommen aus Adrian Pohl und Felix Ostrowski (2010): 'Linked Data' - und warum wir uns im hbz-Verbund damit beschäftigen. In B.I.T. Online, 13/3, S. 259-268.

⁹³ 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).

```
<http://dbpedia.org/resource/Sophist_(dialogue)>
<http://purl.org/dc/elements/1.1/creator>
<http://dbpedia.org/resource/Plato>.94
```

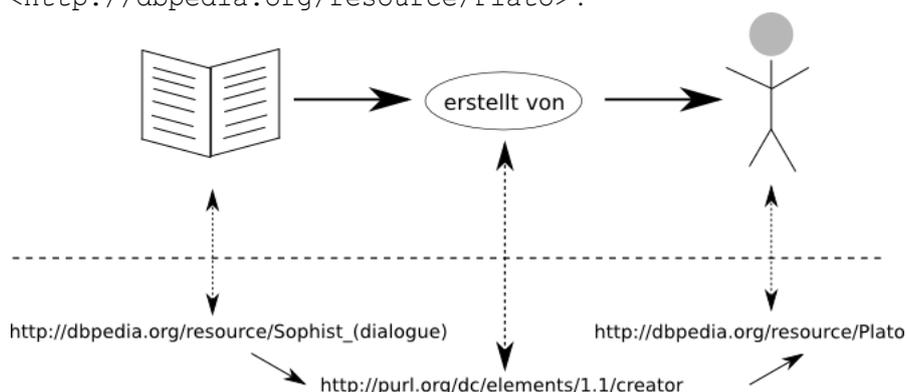


Abbildung 7: : RDF-Tripel

Wie wir sehen, werden die drei Teile der RDF-Aussage (Subjekt, Prädikat und Objekt) durch HTTP-URIs repräsentiert. Warum ist das so? 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 7 illustriert, auf was die Bestandteile des Tripels Bezug nehmen. 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.⁹⁵

Wie wir in unserem Beispiel sehen, sind nicht nur die referenzierten Dinge durch einen URI identifiziert, sondern auch die Verbindung zwischen beiden: das Prädikat des Tripels. Im konkreten Fall handelt es sich um den URI⁹⁶ für das Dublin-Core-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

⁹⁴ 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.

⁹⁵ 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.

⁹⁶ Die Verwendung von URIs für Prädikate macht es möglich, das Prädikat selbst im Web nachschlagen zu können, wenn z.B. der Verwendungszweck nicht klar ist. Siehe hierzu auch den folgenden Abschnitt über Ontologien.

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. Das Ergebnis der Verlinkung einer Dinges mit Literalen oder anderen Dingen ist ein *RDF-Graph*. Graphen sind mehr oder weniger komplexe Netze von Knoten (Dingen) und Kanten (Properties), die je nach Anzahl von verlinkten Dingen an Größe und Komplexität zunehmen.⁹⁷

Natürlich müssen die URIs in RDF mit natürlicher Sprache, Bildern usw. in Verbindung gesetzt werden, um letztlich für Menschen von Nutzen zu sein. 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.⁹⁸ 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

⁹⁷ Die Linked-Open-Data-Cloud selbst ist auch ein Graph. Siehe auch den Eintrag „Graph“ im Glossar am Ende dieses Sammelbands.

⁹⁸ 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 den entsprechenden Eintrag im Glossar am Ende des Sammelbands.

Typ von Dingen treffen zu können, muss man sich auf das Gebiet der Vokabulare und Ontologien begeben.

3.4 Ontologien/Vokabulare

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 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 zwischen Entitäten herrschen können und welche Schlüsse man aus diesen Beziehungen ableiten kann.

Eine knappe Definition von „Ontologie“ auf den Seiten des W3C lautet wie folgt:

“Ontologien sind formalisierte Vokabulare von Termen. Häufig decken sie eine bestimmte Domäne ab und werden von einer Nutzergemeinschaft geteilt. Ontologien definieren Terme, indem sie ihre Beziehungen mit anderen Termen der Ontologie beschreiben.”⁹⁹

Für eine detaillierte Erläuterung sowie eine Auflistung für Bibliotheken relevanter Vokabulare verweisen wir auf den Beitrag von Carsten Klee in diesem Band.

3.5 SPARQL

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

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

⁹⁹ Aus der OWL 2 Web Ontology Language Document Overview (Second Edition, einsehbar unter <http://www.w3.org/TR/owl2-overview/>), Übersetzung A.P. Im Original heißt es: “Ontologies are formalized vocabularies of terms, often covering a specific domain and shared by a community of users. They specify the definitions of terms by describing their relationships with other terms in the ontology.”

```

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 verzeichneten Personen, die der Antiken Philosophie zugeordnet werden und deren Hauptinteresse der Epistemologie galt.¹⁰⁰

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

3.6 Speicherung und Publikation

Es gibt verschiedene Möglichkeiten, RDF-Daten zu speichern und zu publizieren. Im Folgenden werden die gängigsten Methoden erläutert.

3.6.1 RDF-Dateien publizieren

Am leichtesten lässt sich RDF in einfachen Dateien auf einem Webserver publizieren. Diese Methode kann bereits jede Person wählen, die Zugriff auf einen Webserver oder ein Konto bei einem Cloud-Dienst wie Dropbox, github, Google Drive o.ä. hat. Sodann muss einfach ein RDF-Dokument erstellt werden, was sich auch manuell recht einfach bewerkstelligen lässt, wenn

¹⁰⁰ Die Antwort auf diese Anfrage lässt sich hier einsehen: http://dbpedia.org/snorql/?query=Select+%3Fx%0D%0AWhere+%7B+%0D%0A%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%3Fx+%3Chttp%3A%2F%2Fdbpedia.org%2Fontology%2Fera%3E+%3Chttp%3A%2F%2Fdbpedia.org%2Fresource%2FAncient_philosophy%3E+.%0D%0A%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 nötigen Informationen in der Wikipedia bisher nicht als strukturierte Daten erfasst wurden und somit nicht in RDF umgewandelt worden sind.

man die RDF-Serialisierung Turtle verwendet.¹⁰¹ Dieses Dokument kann dann auf einen Webserver geladen werden, so dass andere darauf verlinken können.

3.6.2 Content Negotiation und RDF

In den meisten Fällen möchte man allerdings Menschen, die einen URL in die Adressleiste eines Browsers eingeben, nicht mit RDF konfrontieren, weil es eben nicht für menschliche Kommunikation erfunden wurde. In diesem Fall möchte man doch eher eine HTML-Seite ausliefern. Die Frage ist, wie man unterschiedliche Anfragen (von Webbrowsern/Menschen oder Anwendungen/Maschinen) mit verschiedenen Antworten (HTML oder RDF) bedienen kann? Dies ist technisch ohne weiteres durch die im HTTP-Protokoll eingebaute *Content Negotiation* möglich. Wie funktioniert Content Negotiation?

Wenn ein URI für eine bestimmte *Ressource* steht, können wir mit Content Negotiation verschiedene *Repräsentationen* einer Ressource abrufen. Dies tut z.B. unserer Browser mit seinen Spracheinstellungen. Haben wir Deutsch als bevorzugte Sprache eingestellt, so fordert der Browser die deutsche Version einer Seite an und bekommt diese auch geliefert, wenn es eine solche gibt. Welche Arten von Repräsentation ein Browser oder ein anderer Software-Agent bevorzugt, wird im HTTP-Access-Header angegeben. Wird z. B. die deutsche Version einer Webseite bevorzugt, so steht bei einer Anfrage einer HTTP-URI im HTTP-Header: `Accept-Language: de`.

Eine Ressource kann nicht nur in verschiedenen Sprachen, sondern auch in verschiedenen Formaten angefordert werden. Dies geschieht, indem der Browser oder ein anderer Web-Agent den gewünschten MIME-Typ¹⁰² in der HTTP-Anfrage spezifiziert. Mit `Accept: text/html` wird beispielsweise ein HTML-Dokument angefordert, mit `Accept: application/pdf` ein PDF-Dokument. Neben diesen und vielen anderen Formaten lassen sich im HTTP-Header die verschiedenen RDF-Serialisierungen anfordern:

- `Accept: application/rdf+xml` für RDF/XML
- `Accept: text/turtle` für die Turtle-Notation
- `Accept: text/plain` für N-Triples
- ...

Ein gutes Beispiel für den Einsatz von Content Negotiation im Kontext bibliographischer Daten bietet CrossRef, das es ermöglicht, Beschreibungen einer durch einen Digital Object Identifier (DOI) identifizierten Ressource via Con-

¹⁰¹ Vgl. etwa Idehen (2012) und siehe auch den Eintrag "Turtle" im Glossar. <https://plus.google.com/112399767740508618350/posts/RwvAFV2cAe2>.

¹⁰² Die Liste der derzeit verzeichneten MIME-Types lässt sich unter <http://www.iana.org/assignments/media-types> einsehen.

tent Negotiation in vielen verschiedenen Formaten (darunter auch die verschiedenen RDF-Formate) und verschiedenen Zitierstilen zu bekommen.¹⁰³

3.6.3 RDF in Echtzeit generieren

Es ist nicht nötig, RDF-Daten überhaupt vorzuhalten. Sie können auch auf Basis der in einer Datenbank gehaltenen Daten zum Zeitpunkt einer Abfrage generiert und ausgeliefert werden. Dies lässt sich etwa bewerkstelligen, indem ein sogenannter „Wrapper“ zu einer bestehenden Web-API oder Datenbankschnittstelle geschrieben wird, der die vorgehaltenen Daten in RDF umwandelt, sobald bestimmte Informationen abgefragt werden.

Ein Beispiel für diese Praxis sind die Linked-Open-Data-Dienste der Deutschen Nationalbibliothek (DNB), deren Daten in einem Pica-System erfasst und gespeichert werden und im Falle und zum Zeitpunkt einer entsprechenden Anfrage in RDF umgewandelt werden.¹⁰⁴ Gebe ich zum Beispiel die GND-URI für Platon (<http://d-nb.info/gnd/118594893>) in den Browser ein, so werde ich automatisch auf die HTML-Seite mit einer Beschreibung Platons (<http://d-nb.info/gnd/118594893/about/html>) weitergeleitet. Fordere ich hingegen RDF an – via Content Negotiation oder indem ich direkt die URL <http://d-nb.info/gnd/118594893/about/rdf> aufrufe–, so wird in Echtzeit das RDF aus dem zugrundeliegenden Pica-Datensatz generiert. Dabei findet die im Rahmen von Culturegraph entwickelte offene Konvertierungssoftware Verwendung.¹⁰⁵

3.6.4 Triple Store mit SPARQL-Endpoint

Eine anspruchsvollere Variante der Speicherung und Publikation von RDF ist das Aufsetzen einer RDF-Datenbank – eines sogenannten Triple Stores – mit einem SPARQL-Endpoint. Triple Stores sind optimal an das RDF-Datenmodell angepasste Datenspeicher: Gespeichert wird ein RDF-Graph, der wiederum in verschiedene Subgraphen gegliedert werden kann. Es gibt mittlerweile ein größeres Angebot an Triple Stores, darunter sowohl proprietäre als auch Open-Source-Produkte.¹⁰⁶

¹⁰³ Vgl. http://crosstech.crossref.org/2011/11/turning_dois_into_formatted_ci.html. Es ist allerdings darauf hinzuweisen, dass CrossRef entgegen der guten Praxis zum Großteil MIME Types verwendet, die bisher nicht durch die Internet Assigned Numbers Authority (IANA) definiert wurden.

¹⁰⁴ Darüber hinaus bietet die DNB auch Gesamtabzüge der Daten in RDF an, die allerdings separat generiert werden.

¹⁰⁵ Culturegraph-Software <http://github.com/culturegraph>.

¹⁰⁶ Vgl. Triplestore. (2013, February 24). In: Wikipedia, The Free Encyclopedia. Abgerufen am 27. Februar 2013, 21:15 von <http://en.wikipedia.org/w/index.php?title=Triplestore&oldid=533770944> Abschnitt List of

Ein SPARQL-Endpoint ermöglicht komplexe Abfragen über die in einem Triple Store enthaltenen RDF-Daten.

3.6.5 RDFa – Linked Data als Markup von Webseiten

Eine weitere Methode zur Publikation von RDF ist die Einbettung von RDF in existierende Webseiten. Genau für diesen Zweck gibt es das RDF-Format Resource Description Framework in Attributes, kurz: RDFa.

Hinter RDFa steckt die Idee einer Verschmelzung von menschenlesbaren HTML-Seiten mit maschinenlesbaren Daten in RDF. HTML ist die Markup-Sprache des WWW und wurde in erster Linie im Hinblick auf eine menschenlesbare Präsentation von Inhalten geschaffen. Damit liegt der Großteil der im Web vorhandenen Informationen in einer Form vor, die eine maschinelle Verarbeitung erschwert, wenn nicht unmöglich macht. RDFa ist semantisches Markup für HTML-Dokumente. Mit RDFa lassen sich für maschinelle Verarbeitung strukturierte Daten in menschenlesbare HTML-Seiten einbetten. Somit können Maschinen ohne Probleme Daten aus mit RDFa angereicherten Webseiten gewinnen.

Ein einfaches Beispiel soll zeigen, wie RDFa funktioniert. Angenommen wir haben folgende rudimentäre Beschreibung eines literarischen Werks in HTML:

```
<div>
Titel: <span>Der Sophist</span><br />
Autor: <span>Plato</span><br />
Datum: <span>365 v. Chr.</span>
</div>
```

Ein Anreicherung dieser Daten mit RDFa kann wie folgt aussehen:

```
<div xmlns:dc="http://purl.org/dc/elements/1.1/"
about="http://openlibrary.org/works/OL51935W/">
Titel: <span property="dc:title">Der Sophist</span><br />
Autor: <span property="dc:creator" resource=
"http://viaf.org/viaf/104718382">Plato</span><br />
Datum: <span property="dc:date">365 v. Chr.</span>
</div>
```

implementations

http://en.wikipedia.org/w/index.php?title=Triplestore&oldid=533770944#List_of_implementations.

Das Beispiel zeigt, wie die grundlegenden Funktionen von RDF in RDFa ausgedrückt werden: Subjekt-URIs mit dem `about`-Attribut, Properties mit dem `property`-Attribut sowie Objekt-URIs mit dem `resource`-Attribut. Insbesondere im Kontext von `schema.org` wird mit Microdata¹⁰⁷ ein weiterer Ansatz zum Einbetten strukturierter Daten in HTML genutzt. Microdata lässt sich zwar eins zu eins nach RDF überführen, ist aber – aufgrund seiner geringeren Expressivität – selbst kein RDF. Deshalb soll in diesem Kontext auf eine detaillierte Betrachtung verzichtet werden.

4 Open Data und Lizenzierung

Die Beantwortung einer grundlegenden Frage wurde nur gestreift: Wann handelt es sich bei Daten um Open Data? An dieser Stelle soll das Konzept der Offenheit näher erläutert werden.¹⁰⁸

Eine weithin anerkannte Definition von Offenheit in Bezug auf Inhalte und Daten bietet die *Open Knowledge Definition* (OKD)¹⁰⁹, die innerhalb der Open Knowledge Foundation erarbeitet und 2006 in Version 1.0 veröffentlicht wurde (vgl. Open Knowledge Foundation, 2006a). Die Kurzform der Open Knowledge Definition lautet:

„Daten oder Inhalte sind offen, wenn es einer jeden Person freisteht, sie zu benutzen, weiterzuverwenden und -zuverbreiten – eingeschränkt höchstens durch die Bedingung der Attribution und/oder der Weitergabe unter gleichen Bedingungen (Share Alike).“¹¹⁰

Es lassen sich folgende zentrale Bedingungen extrahieren, an die hier die Vergabe des Etiketts „offen“ geknüpft wird:

- *Offener Zugang* zum Wissen, d.h. es muss offen und kostenlos als Gesamtheit (vorzugsweise im Internet) zugänglich sein.
- *Offene Standards*, d.h. das Wissen muss in einem offen dokumentierten und nicht-proprietären Format vorliegen.

¹⁰⁷ HTML5 Microdata Aktueller Diskussionsstand <http://www.whatwg.org/specs/web-apps/current-work/multipage/microdata.html#microdata>.

¹⁰⁸ Die Autoren dieses Textes sind keine Juristen. Sie versuchen lediglich, die wichtigsten Punkte in einer einfachen Form zusammenzufassen und verweisen für eine detaillierte Betrachtung der rechtlichen Fragen durch einen Juristen auf Kreuzer (2011).

¹⁰⁹ Open Knowledge Foundation (2006b). Die deutsche Übersetzung der Open Knowledge Definition findet sich im Anhang.

¹¹⁰ Vgl. <http://opendefinition.org/>, Übersetzung A.P. Im Original heißt es: „A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike.“

- *Offene Lizenzen*, d.h. die Daten müssen (als Einzeldatum und als Sammlung) unter einer offenen Lizenz publiziert werden.¹¹¹ Als einzige Einschränkungen sind erlaubt: die Forderung nach einer Attribution der Quelle sowie Share Alike, womit eine Lizenzierung abgeleiteter Werke unter denselben Bedingungen gefordert wird.¹¹²

Speziell auf Linked Open Data zugeschnitten gibt es ein Fünf-Sterne-Schema der Offenheit von Daten, mit dem Tim Berners-Lee 2010 die Linked-Data-Prinzipien ergänzt hat (siehe Abbildung 8).

- ★ Available on the web (whatever format)
but with an open licence, to be Open Data
- ★★ Available as machine-readable structured data
(e.g. excel instead of image scan of a table)
- ★★★ as (2) plus non-proprietary format (e.g. CSV instead of excel)
- ★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
- ★★★★★ All the above, plus:
Link your data to other people's data to provide context

Abbildung 8: Star System of Openness (Berners-Lee, 2006)

Wie man sieht, bekommen alle Daten, die frei lizenziert sind, einen Stern. Zwei Sterne erhalten Daten, die zusätzlich noch in einer strukturierten Form (also Excel-Format oder Endnote Files o. ä.) vorliegen. Wenn diese strukturierte Form noch ein offenes Format ist (wie beispielweise CSV oder auch MARC 21), erhalten die Daten drei Sterne. Sobald die offen lizenzierten Daten mit Hilfe von W3C-Standards wie RDF oder SPARQL zur Verfügung gestellt werden, erhalten die Daten vier Sterne. Wenn diese Daten auch noch auf andere Datenquellen verlinken, werden fünf Sterne vergeben. Zur Klarstel-

¹¹¹ Lizenzen sind rechtliche Werkzeuge, mit denen der Inhaber bestimmter Rechte (etwa dem Urheber eines Werkes) Dritten bestimmte Nutzungsrechte übertragen kann.

¹¹² Eine Liste von Lizenzen, die mit der Knowledge Definition kompatibel sind, lässt sich unter <http://licenses.opendefinition.org/> einsehen. Datenlizenzen, die OKD-konform sind, finden sich unter <http://opendefinition.org/licenses/#Data>.

lung: In diesem Schema sind offene Lizenzen Voraussetzung für alle weiteren Sterne, d.h. ohne eine offene Lizenz ist es überhaupt nicht möglich, überhaupt einen Stern zu erreichen.

Warum sind offene Lizenzen überhaupt wichtig? Die im Urheberrecht und verwandten Rechten festgeschriebenen Regeln gelten automatisch ab dem Zeitpunkt der Veröffentlichung eines Werkes. Bei bibliographischen Daten im Allgemeinen und Bibliotheksdaten im Besonderen unterliegt zwar meist der einzelne Datensatz selbst keinem Urheberrecht, weil die nötige Schöpfungshöhe hierfür nicht gegeben ist.¹¹³ Jedoch kann – zumindest in Europa auf Basis der Datenbankrichtlinie 96/9/EG – für eine Sammlung von Daten sowie für wesentliche Teile einer Sammlung ein Datenbankschutz bestehen, wenn für die Erstellung der Datenbank eine wesentliche Investition erbracht wurde.¹¹⁴

Eine Weiterverwendung von im Netz zugänglichen Daten kann demnach rechtlich problematisch sein, wenn die Daten ohne Angabe einer Lizenz zur Verfügung gestellt werden. Insbesondere kann eine Verwendung eines gesamten Abzugs der Daten (Datenbank-Dump) oder großer Teile einer Datenbank rechtlich unzulässig sein. Um das volle Potential von Linked Data zu entfalten – etwa indem verschiedene Datenquellen automatisiert miteinander verlinkt werden –, ist es allerdings häufig nötig, die gesamten Daten einer Datenquelle lokal vorliegen zu haben.

Ein Beispiel soll die Problematik fehlender Lizenzierung verdeutlichen: Nachdem – wie im einleitenden Abschnitt erwähnt – einige Jahre eine Kooperation zwischen der deutschsprachigen Wikipedia und der Deutschen Nationalbibliothek im Bezug auf Normdaten bestand, wurde Ende 2008 in der Wikipedia Community darüber nachgedacht, Titeldaten aus Bibliotheks- und Verbundkatalogen in der Wikipedia weiterzuverwenden. Dabei stellte sich schnell die Frage nach den Nutzungsbedingungen der Daten, nach Urheberrecht und Datenbankrechten.¹¹⁵ Da Rechtssicherheit in Bezug auf weiterverwendete Daten unerlässlich schien, leisteten Vertreter von Wikimedia jahrelange Überzeugungsarbeit insbesondere bei der Deutschen Nationalbibliothek, aber auch bei anderen bibliothekarischen Institutionen, damit diese Rechtssicherheit durch die Anwendung offener Lizenzen hergestellt würde. Es hat Jahre gedauert, bis das Ziel erreicht wurde und die Daten der DNB unter einer offenen Lizenz bereitgestellt wurden.

Dieses Beispiel veranschaulicht, wie wichtig es für Dritte sein kann, dass Datenabzüge mit einer Lizenz versehen werden, die potentiellen Weiterverwen-

¹¹³ Vgl. Kreuzer (2011), Abschnitt 2.3.1.

¹¹⁴ Vgl. ebd., 3.2.

¹¹⁵ Vgl. Ideen für verbesserte Unterstützung bibliographischer Informationen in Wikipedia: <http://de.wikipedia.org/w/index.php?title=Benutzer:Duesentrieb/Biblio&oldid=52914746>. Wie oben (Fußnote 26) bereits erläutert, geht die Kommunikation zwischen Wikimedia und der DNB über offene Lizenzierung der DNB-Katalogdaten sogar zurück bis auf das Jahr 2006.

dern deutlich macht, unter welchen Bedingungen die Daten genutzt werden können. Nicht jede potentielle Nachnutzerin hat die Möglichkeit, in jahrelanger Kommunikation eine Öffnung der Daten zu verhandeln oder bloß die Erlaubnis für eine bestimmte Nutzung einzuholen.

Was sind offene Lizenzen? Die Open Knowledge Definition setzt einen klaren Rahmen für offene Lizenzen: Eine Lizenz ist offen, wenn sie die Nutzung, Weiterverwendung und Kombination mit anderen Daten/Inhalten sowie die Weiterverbreitung erlaubt und höchstens folgende Einschränkungen beinhaltet: Angabe der Quelle (Attribution) sowie die Weitergabe von Derivaten unter den gleichen Lizenzbedingungen (Share Alike).

In Bezug auf Daten ist die Wahl der Lizenzen derzeit nicht ganz einfach, weil die Creative-Commons-Lizenzen, deren Nutzung im Web-Kontext zunächst naheliegt, in der derzeitigen Form nicht für Daten geeignet sind. Das Problem ist, dass CC-Lizenzen bis zur Version 3 zwar das Urheberrecht aber nicht den Datenbankschutz abdecken, der – wie erwähnt – gerade in Europa eine Rolle spielt.¹¹⁶ Daher eignen sich für das Bereitstellen von Open Data derzeit nur wenige Lizenzen:

- Die *CC0 Public Domain Dedication*¹¹⁷ sowie die Open Data Commons *Public Domain Dedication and License* (PDDL)¹¹⁸ sind beides Werkzeuge, um Daten ohne jegliche Einschränkungen zur Nutzung, Weiterverwendung und -verbreitung zu publizieren. Damit werden die Daten – soweit in der jeweils geltenden Rechtsgebung möglich – gemeinfrei.
- Die *Open Data Commons Attribution License* (ODC-BY)¹¹⁹ ist eine Lizenz für Daten, die es Weiterverwendern auferlegt, die Datenquelle zu nennen.¹²⁰
- Die *Open Data Commons Open Database License* (ODbL)¹²¹ fordert neben der Attribution auch eine Veröffentlichung von Derivaten unter den gleichen Bedingungen.

¹¹⁶ Dies wird sich mit Version 4.0 der CC-Lizenzen ändern, so dass in Zukunft auch in Europa CC-Lizenzen ohne Probleme sowohl für Inhalte als auch für Daten verwendet werden können.

¹¹⁷ Creative Commons – CC 1.0 Universal <http://creativecommons.org/publicdomain/zero/1.0/>

¹¹⁸ Open Data Commons Public Domain Dedication and License (PDDL) | Open Data Commons <http://opendatacommons.org/licenses/pddl/>

¹¹⁹ Open Data Commons Attribution License | Open Data Commons <http://opendatacommons.org/licenses/by/>

¹²⁰ Dies bezieht sich wohlgermerkt nicht auf einen einzelnen Datensatz, sondern auf die Datensammlung als Ganzes. So reicht es etwa aus, die Datenquelle auf der about-Seite eines Services zu nennen, die Namensnennung muss aber nicht am einzelnen Datensatz angebracht werden.

¹²¹ Open Data Commons Open Database License | Open Data Commons <http://opendatacommons.org/licenses/odbl/>.

Während alle vier Lizenzen gemäß der Open Knowledge Definition offen sind, gibt es dennoch Diskussionen darüber, welche Lizenzen für Daten am besten verwendet werden sollten.¹²² Creative Commons¹²³, die Open Knowledge Foundation¹²⁴ sowie die DINI AG KIM¹²⁵ empfehlen die Verwendung von CC0. Auch Wikimedia hat in Bezug auf die Deutsche Digitale Bibliothek immer wieder die Publikation der Metadaten unter CC0 gefordert.¹²⁶ Im deutschsprachigen Raum scheint sich die Verwendung von CC0 bei der Freigabe von Daten als Standard durchzusetzen. Bisher haben alle deutschsprachigen Institutionen, die sich zu einer Datenfreigabe entschlossen haben, CC0 gewählt.

5 Herausforderungen beim Aufbau eines Linked-Open-Data-Ökosystems

Im Jahr 2012 hatte die Konferenz Semantic Web in Bibliotheken (SWIB) den Untertitel „Towards an international LOD library ecosystem“. Dementsprechend wurde dort auch diskutiert, welche Herausforderungen und Arbeiten auf dem Weg zu einem internationalen Linked-Open-Data-Ökosystem warten. Im Folgenden werden verschiedene dieser Herausforderungen skizziert, denen sich die Linked-Data-Gemeinschaft im Allgemeinen wie die Bibliothekswelt im Besonderen für eine erfolgreiche und gewinnbringende Anwendung von Linked Data stellen muss.

5.1 Provenienzzangaben

Da Linked Open Data eine dezentrale kooperative Form der Datenhaltung und Erstellung impliziert, sind sicherlich Informationen zur Herkunft (Provenienz) von Daten ein wichtiger Aspekt, um Vertrauenswürdigkeit zu gewährleisten. Dass in dieser Hinsicht zwar einige Ansätze bestehen, allerdings noch einiges in Richtung Standardisierung getan werden muss, zeigt Kai Eckerts Artikel in diesem Sammelband im Detail auf.

¹²² Siehe etwa Tochtermann (2012) und die dortige Diskussion in den Kommentaren.

¹²³ Vollmer et al. (2012).

¹²⁴ Coyle et al. (2011).

¹²⁵ Gruppe Lizenzen der DINI-AG KIM (2011).

¹²⁶ Siehe etwa Schindler (2012).

5.2 Mapping und Konvertierung von Altdaten

Bibliothekarische Daten liegen bisher in bibliothekarischen Austauschformaten vor, d.h. dem im internationalen Kontext und zunehmend auch in Deutschland verwendeten MARC21-Format (MACHINE-Readable Cataloging) und dem im deutschsprachigen Raum entwickelten Maschinenlesbaren Austauschformat für Bibliotheken (MAB) oder aber auch in diesen sehr ähnlichen Formaten. Beide Formate wurden ursprünglich dazu entwickelt, Informationen für gedruckte Katalogkarten auszutauschen. Diese bibliographischen Austauschformate wurden in den 1960er und 1970er Jahren entwickelt und stellen einen Anachronismus dar. Vor dem Hintergrund der aktuellen Möglichkeiten elektronischer Datenverarbeitung sind sie ein Hindernis für innovative Entwicklungen. Dafür gibt es verschiedene Gründe, die etwa 2002 von Roy Tennant in seinem bekannten Text „MARC Must Die“ genannt wurden.¹²⁷ Den Daten fehlt häufig die nötige *Granularität*, *Erweiterbarkeit* und damit die Nachhaltigkeit von MARC haben ihre Grenzen. Außerdem sind MARC/MAB *obskure Standards*, die nur einige wenige Menschen in der Bibliothekswelt verstehen und verarbeiten können. Es gibt eine Vielzahl verschiedener Möglichkeiten, diese Formate auf das RDF-Datenmodell zu mappen und zu konvertieren. Dies wird erkennbar an der großen Anzahl von Vokabularen, die Carsten Klee in diesem Band vorstellt.¹²⁸ Im Kontext der DINI-KIM-AG „Titeldaten“ wird derzeit insbesondere für den deutschsprachigen Raum eine Empfehlung zur Repräsentation bibliographischer Daten in RDF entwickelt, wobei verschiedene Vokabulare Verwendung finden.¹²⁹ Darüber hinaus sind sicher auch die Ansätze im Rahmen der Library-of-Congress-Initiative „Bibliographic Framework for a Digital Age“¹³⁰ wie auch der Arbeitsgruppe zur Erweiterung des schema.org-Vokabulars¹³¹ für bibliographische Daten nennenswert. Was die Softwarewerkzeuge zur Überführung bibliographischer Daten angeht, gibt es verschiedene freie Software für die Überführung von MARC nach RDF – gerade auch im Bibframe-Kontext versucht die Library of Congress, frühzeitig Konvertierungstools zu teilen und gemeinsam mit der bibliothekarischen Gemeinschaft zu entwickeln. Im Zusammenhang von Culturegraph wurden darüber hinaus generische Tools zur Datenkonvertierung entwickelt,

¹²⁷ Tennant, Roy MARC must Die in Library Journal 15.10.02

<http://www.libraryjournal.com/article/CA250046.html>.

¹²⁸ Auch Geipel/Böhme/Hauser/Haffner stellen in ihrem Artikel die Schwierigkeiten dar, die eine Überführung von klassischen Bibliotheksdaten nach RDF mit sich bringt. Hans-Georg Beckers Beitrag zur Repräsentation fortlaufender Sammelwerke in RDF zeigt an einem konkreten Beispiel auf, dass eine solche Arbeit alles andere als trivial ist.

¹²⁹ Der aktuelle Entwurf der Empfehlungen kann hier eingesehen werden: <https://wiki.d-nb.de/x/cYMOB>.

¹³⁰ Bibframe.org :: New Bibliographic Framework <http://bibframe.org/>

¹³¹ Schema Bib Extend Community Group <http://www.w3.org/community/schemabibex/>

die in verschiedenen Kontexten Verwendung finden und weiterentwickelt werden.¹³²

Es gibt also durchaus vielversprechende Ansätze zur kooperativen Entwicklung von Tools und Mappings, um Katalogdaten nach Linked Data zu überführen. Um den Weg zu einem internationalen Linked-Open-Data-Ökosystem erfolgreich zu beschreiten, müssen sich allerdings erst Standards bei der Konversion der Daten und vor allem Standards der RDF-Repräsentation von Daten entwickeln.

5.3 Schnittstellen und Datensynchronisierung

Wie bereits mehrfach erläutert bedeutet Linked Data eine Erweiterung des Webs, indem dessen Architektur auch für die Publikation von Daten genutzt wird. Das Web ist ein in höchstem Maße verteiltes System, das erfolgreich die Illusion schafft, man bewege sich in einem geteilten Informationsraum.¹³³

Anwendungen über Linked Open Data basieren u. U. auf Daten aus vielen verschiedenen Quellen und häufig müssen lokale Kopien (Caches) der genutzten Daten gemacht werden, um die Performanz einer Anwendung zu gewährleisten.

Wie der Abschnitt zur Lizenzierung bereits deutlich gemacht hat, ist eine wichtige Voraussetzung für den Aufbau von Linked-Data-basierten Anwendungen die Verwendung offener Lizenzen. Auf der technischen Seite besteht die Herausforderung, die Datenbasis einer Anwendung mit den von ihr genutzten verteilten Datenquellen zu synchronisieren und die Daten aktuell zu halten. Diese Problematik wurde bereits von verschiedenen Akteuren adressiert, harrt aber noch einer allgemein anerkannten Lösung.¹³⁴

5.4 Linked-Data-fähige Bibliothekssysteme

In Bibliotheken und Bibliotheksverbänden wird die Erzeugung und Pflege bibliographischer Daten in der Regel mit einer speziellen Software (Bibliothekssystem) durchgeführt, die darüber hinaus für die Verwaltung von Nutzerdaten, Erwerbungen und Ausleihen benutzt wird. Bisher unterstützt kein bekanntes Bibliothekssystem die Erzeugung und Publikation oder den Konsum und die Integration von Linked Data. Für den Aufbau eines Linked-Data-Ökosystems ist dies allerdings unumgänglich, denn gerade kleineren Biblio-

¹³² Culturegraph <http://culturegraph.github.io/>. Vgl. den Beitrag von Geipel et al. in diesem Sammelband.

¹³³ Dan Connolly hat dies 2002 wie folgt ausgedrückt: "The point of the Web arch[itecture] is that it builds the illusion of a shared information space." Zitiert nach Summers (2013).

¹³⁴ Eine wichtige Initiative ist sicher das ResourceSync-Projekt der Open Archive Initiative und der NISO, siehe <http://www.niso.org/workrooms/resourcesync/>.

theiken fehlen die Ressourcen, um die technischen Herausforderungen selbst zu schultern. Derzeit übernehmen in erster Linie Bibliotheksverbände (wie das hbz oder der BVB) das Umwandeln und die Publikation der Daten.¹³⁵ Die zukünftige Verbreitung und Akzeptanz von Linked Data wird also in nicht geringem Maße davon abhängen, inwiefern Bibliothekssysteme in Zukunft Linked Data unterstützen. Erste Aktivitäten von Bibliothekssoftwareherstellern zeigen, dass sie durchaus Interesse an Linked Data haben und es gibt auch bei Open-Source-Bibliothekssystemen erste Experimente, um Linked-Data-Technologien nutzbar zu machen.¹³⁶

Innerhalb der International Group of Ex Libris Users (Igelu) hat sich wie erwähnt bereits im September 2011 eine Linked Open Data Special Interest Group gegründet, mit dem erklärten Ziel, dass, „– wo angemessen – essentielle Linked-Open-Data-Funktionen in allen Ex-Libris-Produkten“ aufgenommen werden, „sowohl von der Perspektive der Datenpublikation als auch des Datenkonsums und der Datenintegration“.¹³⁷ Ein Ergebnis dieser Arbeitsgruppe ist derzeit noch nicht veröffentlicht. Es ist unklar, in wie weit Exlibris die Forderungen der Arbeitsgruppe berücksichtigen wird.

OCLC, das verschiedene Bibliothekssysteme anbietet und mit WorldShare Management Services eine cloudbasierte Lösung entwickelt¹³⁸, hat bereits selbst begonnen, experimentell Linked Data – teilweise auch unter offenen Lizenzen – zu veröffentlichen: Sowohl die Dewey Decimal Classification¹³⁹ als auch WorldCat¹⁴⁰ bieten mittlerweile auch RDF-Repräsentationen der Daten an, wobei nur die RDF-Daten von worldcat.org mit einer offenen Lizenz (ODC-BY) versehen sind. Darüber hinaus betreibt OCLC mit viaf.org (ebenfalls lizenziert unter ODC-BY) einen zentralen Linked-Open-Data-Dienst für Personendaten. In welchem Maße OCLC allerdings Linked-Data-Konzepte im WorldShare-System integrieren wird oder gar in bestehenden Bibliothekssystemen wie LBS oder Sunrise unterstützen wird, ist unklar. Bibliotheken sind hier gefordert, Linked Data als eine wesentliche Anforderung an zukunftssichere Bibliothekssysteme zu formulieren. Nur dann ist zu

¹³⁵ Siehe in diesem Sammelband den Artikel von Ceynowa et al., der das Verfahren am Beispiel des durch B3Kat erläutert.

¹³⁶ Vgl. etwa SemantiKoha (<http://libriotech.no/blogs/semantikoha>).

¹³⁷ Siehe <http://igelu.org/special-interests/lod>. Im Original heißt es: “Objective of the Linked Open Data Special Interest Working Group is to achieve essential linked open data features in all Ex Libris products where appropriate, both from the data publishing, the data consuming and the data integration perspective.”

¹³⁸ Zum OCLC-Angebot von Bibliothekssystemen siehe <http://www.librarytechnology.org/oclc.pl>.

¹³⁹ Linked DDC data (experimental) <http://dewey.info/>.

¹⁴⁰ Vgl. OCLC: OCLC adds Linked Data to WorldCat.org (20.06.12) <http://www.oclc.org/news/releases/2012/201238.htm>.

hoffen, dass Bibliothekssoftwarehersteller die Ergebnisse von Linked-Data-Experimenten schließlich in ihre Produkte einfließen lassen.

Offenheit als Standard

Neben den technischen Herausforderungen ist es nicht zuletzt die konkrete Ausgestaltung der rechtlichen Bedingungen, die die zukünftige Verbreitung und Akzeptanz von Linked Open Data beeinflussen. In Deutschland hat sich die Freigabe von offenen Daten unter CC0 durchgesetzt – mittlerweile haben alle Bibliotheksverbände entweder Open Data veröffentlicht oder arbeiten intensiv an einer Freigabe. Es ist allerdings noch einiges zu tun, bis dieses Thema auch im internationalen Kontext erledigt ist. Erst wenn die Lizenzierung bibliothekarischer Daten kein Problem mehr ist, mit dem man sich beim Aufbau von Dienstleistungen regelmäßig auseinandersetzen muss, dann wird man in der Lage sein, sich vollständig auf die Entwicklung neuer Kooperationsmodelle, neuer Dienste und Crowdsourcing-Lösungen zu konzentrieren.

5.5 Die bibliographische Allmende aufbauen

Es wurde in Abschnitt 2.1.1 schon verdeutlicht: Die Daten aus Bibliothekskatalogen machen nur einen Teil der für Wissenschaftler und die Allgemeinheit relevanten Daten aus, denn Bibliotheken erfassen in der Regel nur „selbstständige Werke“. Während in Deutschland mittlerweile ein großer Teil der Katalogdaten unter offenen Lizenzen verfügbar ist, sind Artikelmetadaten bisher nicht offen verfügbar. Für viele Zwecke – etwa den Aufbau einer offenen Bibliographie oder eines offenen Suchindexes für ein Fach oder ein Thema – stellt dies ein Problem dar.

Bibliotheken haben in den letzten Jahren begonnen, nicht unbeträchtliche Summen an Anbietern von sogenannten Discovery Services zu bezahlen, um den eigenen Nutzern die Suche über einen einheitlichen, umfassenden Suchindex zu ermöglichen. Wäre es gängige Praxis, aktuelle Metadaten unter offenen Lizenzen zur Verfügung zu stellen, ließe sich an offenen Lösungen arbeiten, die von vielen genutzt werden können.

Eine Herausforderung für die Zukunft wird es sein, die Menge der geteilten, offenen bibliographischen Daten zu vergrößern, insbesondere, indem vermehrt auch Artikelmetadaten frei zur Verfügung gestellt werden. So könnte man in Zukunft, die in der Vergangenheit von Bibliotheken, Dokumentaren, Verlagen und Bibliographen getrennt erstellten und gepflegten Daten – zumindest innerhalb von Endnutzeranwendungen – zusammenzuführen.

Einzelne Gruppen und Projekte haben schon angefangen, dieses Problem – zumindest teilweise – anzugehen, zum Beispiel die Arbeitsgruppe zu Open Bibliographic Data der Open Knowledge Foundation oder jüngst das Quali

Ole Projekt und JISC. Diese beiden Organisationen entwickeln mit der Global Open Knowledgebase (GOKb) an einer offenen, community-basierten, internationalen Datenbank, die Bibliotheken mit Metadaten über elektronische Ressourcen versorgen und sie damit bei der Verwaltung von Subskriptionen unterstützen soll.¹⁴¹

Relevant sind hier auch Metadaten aus Repositorysystemen. Während bisher Bibliothekskataloge im Fokus der Bemühungen standen, bieten auch diese Daten ein großes Potenzial. Die Daten werden zwar oft über OAI-PMH zum Harvesten zur Verfügung gestellt, jedoch sind die zugrundeliegenden Daten meist deutlich umfangreicher und komplexer als die über OAI-PMH verteilten Daten. Linked Open Data hat hier das Potenzial, bessere Daten zur Entwicklung neuer Services zur Verfügung zu stellen.

Von einer breiten Praxis und allgemeinen Akzeptanz, bibliographische Daten zu öffnen, kann erst die Rede sein, wenn nicht nur bibliothekarische Organisationen, sondern auch andere Produzenten bibliographischer Daten - Verlage, Universitäten, Wissenschaftler usw. – ihre Daten öffnen.

6 Quellenverzeichnis

Ackermann, Marcel R. (2011, 9. Dezember): DBLP releases its 1.8 million bibliographic records as open data. URL: <http://openbiblio.net/2011/12/09/dblp-releases-its-1-8-million-bibliographic-records-as-open-data/>.

Berners-Lee, Tim / Hendler, James / Lassila, Ora (2001): The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. In: Scientific American, 284 (5), S. 34–43, Mai 2001 (dt.: Mein Computer versteht mich. In: Spektrum der Wissenschaft, August 2001, S. 42–49).

Berners-Lee, Tim (2006, 27. Juli): Linked Data - Design Issues. (Letzte Änderung 2010.) URL: <http://www.w3.org/DesignIssues/LinkedData.html>.

Biblioteca Nacional de España und Ontology Engineering Group (2012, 2. Februar): Linked Data at the Biblioteca Nacional de España. URL:

¹⁴¹ Webseite: <http://gokb.org>. In der Pressemitteilung „International collaboration to help transform the way libraries manage their resources“ (<http://gokb.org/post/25021222983/gokbpressrelease>) heißt es: “GOKb will be an open, community-based, international data repository that will provide libraries with publication information about electronic resources.”

<http://openbiblio.net/2012/02/02/linked-data-at-the-biblioteca-nacional-de-espana/>.

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

Brantley, Peter (2007, 16. Juli): Open Library launches in demo. URL: <http://toc.oreilly.com/2007/07/open-library-launches-in-demo.html>.

British Library (2010, 23. August): British Library to share millions of catalogue records. URL: <http://pressandpolicy.bl.uk/Press-Releases/British-Library-to-share-millions-of-catalogue-records-43b.aspx>.

Calhoun, Karen (2008, 16. August): Free the Data: Discussion Panel at IFLA 2008. URL: <http://community.oclc.org/metalogue/archives/2008/08/free-the-data-discussion-panel.html>.

CERN Library (2010): The CERN Library publishes its book catalog as Open Data. Pressemitteilung einsehbar unter <http://library.web.cern.ch/library/Library/announcement.html>.

Coyle, Karen / MacGillivray, Mark / Murray-Rust, Peter / O'stehen, Ben / Pitman, Jim / Pohl, Adrian / Pollock, Rufus / Waites, William (2011, 17. Januar): Principles on Open Bibliographic Data. URL: <http://openbiblio.net/principles/>.

Conference of European National Librarians (2011). Europe's national librarians support Open Data licensing. Pressemitteilung einsehbar unter: <http://www.libereurope.eu/sites/default/files/CENL%20adopts%20CC0.pdf>.

Danowski, Patrick, und Pfeifer, Barbara (2007). „Wikipedia und Normdateien: Wege der Vernetzung am Beispiel der Kooperation mit der Personennamendatei“. Bibliothek - Forschung und Praxis 31, Nr. 2 (2007): 149–155. Online zugänglich unter http://www.b2i.de/fileadmin/dokumente/BFP_Bestand_2007/Jg_31-Nr_2/Jg_31-Nr_2_Aufsaeetze/Jg_31-2007-Nr_2-S_149-156.pdf.

Deutsche Nationalbibliothek / Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (2010, November): culturegraph.org - Basisinfrastruktur für Gedächtnisinstitutionen im Semantic Web. Pressemitteilung einsehbar unter: www.hbz-nrw.de/dokumentencenter/presse/pm/culturegraph_de.

Europeana Foundation (2011): Europeana Data Exchange Agreement. URL: http://pro.europeana.eu/c/document_library/get_file?uuid=deb216a5-24a9-4259-9d7c-b76262e4ce55&groupId=10602

Gray, Jonathan (2007a, 6. Dezember): 'The Future of Bibliographic Control' and Licensing Policies for Bibliographic Data. URL: <http://blog.okfn.org/2007/12/06/the-future-of-bibliographic-control-and-licensing-policies-for-bibliographic-data/>.

Gray, Jonathan (2007b, 19. Dezember): Response to 'The Future of Bibliographic Control' draft from the Library of Congress. URL: <http://blog.okfn.org/2007/12/19/response-to-the-future-of-bibliographic-control-draft-from-the-library-of-congress/>.

Gruppe Lizenzen der DINI-AG KIM (2011, 31. Oktober): Empfehlungen zur Öffnung bibliothekarischer Daten. URL: <https://wiki.d-nb.de/x/zA21Ag>.

Harvard Library (2012, 24. April): Millions of Harvard Library Catalog Records Publicly Available. URL: <http://isites.harvard.edu/icb/icb.do?keyword=k77982&pageid=icb.page498373>

Hengel, Christel / Pfeifer, Barbara (2005): Kooperation der Personennamendatei (PND) mit Wikipedia.. In: Dialog mit Bibliotheken 17/3, S. 18-24.

Hickey, Thom (2012, 4. Mai): VIAF Dataset. URL: <http://outgoing.typepad.com/outgoing/2012/05/viaf-dataset.html>.

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

Idehen, Kingsley (2012): How To Describe Stuff You Like via a Turtle Document. URL: <https://plus.google.com/112399767740508618350/posts/RwvAFV2cAe2>.

Kooperativer Bibliotheksverbund Berlin-Brandenburg (2011, 30.12): Verbundkatalog B3Kat als (Linked) Open Data veröffentlicht. URL: <http://www.kobv.de/aktuelles/information/datum/2011/12/30/verbundkatalog-b3kat-als-linked-open-data-veroeffentlicht/>.

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

Lessig, Lawrence (2004): Free Culture – How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity. New York: The Penguin Press. Online zugänglich unter:
<http://www.free-culture.cc/freecontent/>.

Library of Congress Working Group on the Future of Bibliographic Control (2007): Report on the Future of Bibliographic Control. Draft for Public Comment. <http://www.loc.gov/bibliographic-future/news/lcwg-report-draft-11-30-07-final.pdf>.

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

Library of Congress (2012, 21. November): Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services: URL: <http://www.loc.gov/marc/transition/news/bibframe-112312.html>.

Malmsten, Martin (2008, 3. Dezember): LIBRIS available as Linked Data. URL: <http://librisbloggen.kb.se/2008/12/03/libris-available-as-linked-data/>.

Malmsten, Martin (2011, 21. September): Swedish National Bibliography and authority data released with open license. URL: <http://librisbloggen.kb.se/2011/09/21/swedish-national-bibliography-and-authority-data-released-with-open-license/>.

Nature Publishing Group (2012, 4. April): Nature Publishing Group releases linked data platform. URL: http://www.nature.com/press_releases/linkeddata.html?WT.mc_id=TWT_npg_news.

OCLC (2010): WorldCat Rights and Responsibilities for the OCLC Cooperative. URL: <http://www.oclc.org/worldcat/recorduse/policy/default.htm>.

OCLC (2012a, 4. April): Virtual International Authority File service transitions to OCLC; contributing institutions continue to shape direction through VIAF Council. Presseerklärung einsehbar unter <http://www.oclc.org/news/releases/2012/201224.htm>.

OCLC (2012b, 20. Juni): OCLC adds Linked Data to WorldCat.org. Presseerklärung einsehbar unter <http://www.oclc.org/en-US/news/releases/2012/201238.html>.

Open Knowledge Foundation (2006a, 6. September): Version 1.0 of the Open Knowledge Definition Released. URL:

<http://blog.okfn.org/2006/09/06/version-10-of-the-open-knowledge-definition-released/>.

Open Knowledge Foundation (2006b): Open Definition, Version 1.1. URL:

<http://opendefinition.org/okd/>.

Open Knowledge Foundation (2011): Definition: Offenes Wissen, Version

1.1. URL: <http://opendefinition.org/okd/deutsch/>.

O'Reilly, Tim (2005): What Is Web 2.0 - Design Patterns and Business Models for the Next Generation of Software. <http://oreilly.com/web2/archive/what-is-web-20.html>.

Pattern, Dave (2008, 12. Dezember): Free book usage data from the University of Huddersfield. URL: <http://www.daveyp.com/blog/archives/528>.

Pohl, Adrian (2009a): OCLC, WorldCat und die Metadaten-Kontroverse. In: Bibliotheksdienst, 43 (2009), Nr. 3, S. 274.290. URL:

http://www.zlb.de/aktivitaeten/bd_neu/heftinhalte2009/Erschliessung020309BD.pdf.

Pohl, Adrian (2009b): OCLC: Policy zurückgezogen. URL:

<http://www.uebertext.org/2009/06/oclc-policy-zuruckgezogen.html>.

Pohl, Adrian (2010): OCLC und die Public Domain. URL:

<http://www.uebertext.org/2010/04/oclc-policy-und-die-public-domain.html>.

Pohl, Adrian (2011a): Launch of the Principles on Open Bibliographic Data.

URL: <http://blog.okfn.org/2011/01/18/launch-of-the-principles-on-open-bibliographic-data/>.

Pohl, Adrian (2011b, 21. September): LOD at Bibliothèque nationale de

France. URL: <http://openbiblio.net/2011/09/21/lo-d-at-bibliotheque-nationale-de-france/>.

Pohl, Adrian (2011c): German Library Networks BVB and KOBV release 23 Million Records. URL: <http://openbiblio.net/2011/12/08/bvb-kobv-open-data/>.

Richt, Susanne (2012): „Unsere ‚Abgrenzung‘ zu anderen Einrichtungen wird in Bewegung kommen“ / Generaldirektorin Elisabeth Niggemann über verschiedene Projekte der Deutschen Nationalbibliothek, freie Daten und die Zukunft der Bibliotheken. BuB Forum Buch und Bibliothek 10/2012 S. 691-695.

Schindler, Mathias (2012, 28. November): Ein Meilenstein erreicht, 99 vor uns: Die Deutsche Digitale Bibliothek. URL: <http://blog.wikimedia.de/2012/11/28/deutsche-digitale-bibliothek-beta-ddb/>.

Schoneville, Catrin (2012, 30. März): 3, 2, 1, Wikidata! URL: <http://blog.wikimedia.de/2012/03/30/3-2-1-wikidata/>.

Stephens, Susie / LaVigna, David / DiLascio, Mike und Luciano, Joanne (2006): Aggregation of bioinformatics data using Semantic Web technology. Web Semant. 4, 3 (September 2006), 216-221. URL: <http://dx.doi.org/10.1016/j.websem.2006.05.004>.

Tochtermann, Klaus (2012, 3. August): Thoughts beyond Boundaries: CC0 for Library Data – Publish then Perish. URL: <http://www.zbw-mediatalk.eu/2012/08/thoughts-beyond-boundaries-cc0-for-library-data-publish-then-perish/>.

Vollmer, Timothy / Smith, MacKenzie / Keller, Paul / Peters, Diane (2012): Library catalog metadata: Open licensing or public domain? URL: <http://creativecommons.org/weblog/entry/33768>.

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

W3C Library Linked Data Incubator Group (2011, 25. Oktober): Library Linked Data Incubator Group Final Report. URL: <http://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/>.

Yu, Harlan und David G. Robinson (2012): The New Ambiguity of ‘Open Government’. 59 UCLA L. Rev. Disc. 178 (2012). <http://dx.doi.org/10.2139/ssrn.2012489>.