

Dezentral, offen, vernetzt - Überlegungen zum Aufbau eines LOD-basierten FID-Fachinformationssystems



Dies ist ein Preprint eines in "Bibliothek Forschung und Praxis" erscheinenden Artikels.
Diese HTML-Fassung des Beitrags wurde anhand [dieser Version](#) generiert.



Dieses Werk ist unter der CC0 Public Domain Dedication veröffentlicht.

English:

Title: Decentral, open, linked - reflections on building a sustainable FID metadata infrastructure

Keywords: Linked Open Data, libraries, metadata infrastructure, science

Abstract:

This article reflects the "Richtlinien für das DFG geförderte System der Fachinformationsdienste für die Wissenschaft" and shows how Linked Open Data (LOD) can help with rebuilding the Sondersammelgebiete (SSG) to Fachinformationsdienste (FID). In particular, the aspect of (meta-)data infrastructure is discussed. The thesis of this article is that LOD is predestined to build the base of a sustainable metadata infrastructure for science.

Authors:

Pascal Christoph and Adrian Pohl Jülicher Str. 6 D-50674 Cologne
E-Mail: christoph@hbz-nrw.de
E-Mail: pohl@hbz-nrw.de

Deutsch:

Schlüsselwörter: Linked Open Data, Fachinformationsdienste , Metadateninfrastruktur, Wissenschaft

Abstract:

Dieser Artikel reflektiert die "Richtlinien für das DFG geförderte System der Fachinformationsdienste für die Wissenschaft" und zeigt dabei auf, wie Linked Open Data (LOD) beim Umbau der Sondersammelgebiete (SSG) zu Fachinformationsdiensten (FID) helfen kann. Es wird dabei insbesondere auf den Aspekt (Meta-)Dateninfrastruktur eingegangen. Die These dieses Artikels ist, dass LOD prädestiniert ist, einen wichtigen Eckpfeiler einer nachhaltigen Metadateninfrastruktur für die Wissenschaft zu bilden.

Pascal Christoph kennt ViFas und SSGs aus seiner Arbeit von 2008-2011 am Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz) als technischer Umsetzer des Projekts "vascoda". Seit 2010 arbeitet er am Linked-Open-Data-Dienst "lobid" des hbz .

Adrian Pohl arbeitet seit 2008 im Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz). Sein Fokus liegt auf dem Bereich Linked Open Library Data und dabei insbesondere auf dem Aufbau und der Pflege des LOD-Service lobid. Seit Juni 2010 ist Adrian Koordinator der OKFN Working Group on Open Bibliographic Data.

Inhaltsverzeichnis

1. Einleitung
2. Was ist Linked Open Data?
 - 2.1 Offene Daten als grundlegende Anforderung
 - 2.2 Linked Data - Best Practices
3. Erfahrungen aus dem vascoda-Projekt
4. Cooperare necesse est
5. Konkrete Kooperationsmöglichkeiten bei Querschnittsaufgaben
 - 5.1 Metadatenmapping, Datentransformation und automatische Datenanreicherung
 - 5.2 Aufbau von Rechercheindizes und Web-APIs
 - 5.3 Hubs zur intellektuellen Datenanreicherung
 - 5.4 Flexible Arbeitsteilung
6. Fazit

7. Referenzen

Einleitung

Die "Richtlinien für das DFG geförderte System der Fachinformationsdienste für die Wissenschaft"¹ der Deutschen Forschungsgemeinschaft (DFG) und das damit verbundene Förderprogramm haben den Aufbau sogenannter "Fachinformationsdienste für die Wissenschaft" (FID) zum Ziel. Diese FIDs sollen die bestehenden Strukturen der Sondersammelgebiete (SSG) und Virtuellen Fachbibliotheken (ViFa) zusammenführen und schließlich ersetzen.

Diese von der DFG geförderten Strukturen dienen der wissenschaftlichen Informationsversorgung. Die DFG verbindet dementsprechend mit "der Verabschiedung des neu ausgerichteten Förderprogramms ... die Hoffnung, dass die 'Fachinformationsdienste für die Wissenschaft' als ein flexibles und zukunftsfähiges System der Informationsversorgung der Wissenschaft dienen können".²

Konkret nennen die Richtlinien der DFG folgende Kernaufgaben von Fachinformationsdiensten:

"Die Fachinformationsdienste gewährleisten eine optimale Versorgung - durch die überregionale Bereitstellung und Archivierung relevanter gedruckter und digitaler Medien und Fachinformationen, - durch den Aufbau und die Pflege komfortabler Nachweis- und Recherchesysteme, die technisch auf dem neuesten Stand gehalten werden und - durch vorausschauendes Engagement in der Entwicklung allgemeiner und fachspezifischer Informationsdienstleistungen (darunter beispielsweise Rechercheaufträge, Beratungsleistungen, Digitalisierung mit Mehrwertdiensten, Tiefenerorschließung)."³

Dieser Text befasst sich nur mit der Umsetzung der zweiten genannte Kernaufgabe "Aufbau und die Pflege komfortabler Nachweis- und Recherchesysteme" und den damit verbundenen Prozessen der Datenaggregation und des Datenmanagements. LOD mag in den beiden anderen genannten Aufgabengebieten auch einige Vorteile mit sich bringen, weil Metadaten auch bei der Auswahl von Ressourcen bzw. deren Archivierung eine wichtige Rolle spielen - den Autoren fehlt in diesen Bereichen aber die nötige Expertise, so dass dieser Beitrag dazu schweigt.

Die Grundannahmen dieses Artikels sind:

1. Wissenschaftler wollen auf einfache Weise über neue für sie relevante Veröffentlichungen erfahren.
2. Wissenschaftler wollen möglichst einfachen Zugriff auf alle für sie relevanten Informationen.
3. Wissenschaftler wollen die von ihnen erzeugten Texte und Daten für alle Interessierten auffindbar machen.
4. Die in 1. - 3. genannten Prozesse der Informationsversorgung geschehen idealerweise über das Internet.

Die Hauptthese dieses Artikels ist: Linked Open Data (LOD) kann einen wichtigen Beitrag dazu liefern, den Anforderungen eines flexiblen und zukunftsfähigen Systems der Informationsversorgung gerecht zu werden.⁴ Im Folgenden werden zunächst knapp die Grundlagen von Linked Open Data vermittelt (Abschnitt 2). Sodann werden Erfahrungen aus dem vascoda-Projekt geschildert und Probleme des Projekts aufgezeigt, die mit einem LOD-Ansatz hätten vermieden werden können (Abschnitt 3).

In Abschnitt 4 werden Entwicklungs-, Kommunikations- und Organisationsprinzipien behandelt, die in der LOD-Gemeinschaft gepflegt werden und den kollaborativen Aufbau einer nachhaltigen und flexiblen Informationsinfrastruktur begünstigen. Abschnitt 5 setzt sich mit konkreten Aspekten des Metadatenmanagements auseinander, in denen eine Kooperation besonders sinnvoll ist. Mit einem abschließenden Fazit in Abschnitt 6 endet der Beitrag.

Was ist Linked Open Data?

Die grundlegende Idee von LOD ist Offenheit. "Offenheit" findet sich dabei in mehreren Dimensionen:

- Die *Lizenz* ist offen, d.h. die Daten dürfen von jedem verändert und nachgenutzt werden.
- Das *Datenmodell* ist offen, d.h. es nicht starr, sondern kann flexibel erweitert und ergänzt werden.
- Die zugrundeliegenden *Standards* sind offen. Ihre Entwicklung wird vom World Wide Web Consortium ([W3C](#)) koordiniert.
- Die *Software* rund um LOD ist - im Vergleich zu anderen Bereichen - bestimmt von einer großen Zahl an Open-Source-Entwicklungen.⁵
- Die *Teilnahme* ist offen für jeden, da LOD - wie das World Wide Web - dezentral ist. Es gilt das Motto: "Anybody can say anything about anything".⁶

Eine knappe, eher technische Beschreibung von LOD lautet:

"Linked Open Data sind offene Daten (Open Data), die gemäß Linked-Data-Prinzipien (Linked Data) unter Nutzung der entsprechenden offenen W3C-Standards bereitgestellt werden. Während Open Data keine Datenformate vorgibt, solange es sich um offen dokumentierte Formate handelt, geht es bei Linked Data um die Etablierung von Best Practices für die Integration von Daten in das WWW auf Basis von Standards des World Wide Web Consortium (W3C). Die vier von Tim Berners-Lee⁷ formulierten Linked-Data-Prinzipien lauten:

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

Linked Data baut also auf den bestehenden Web-Standards Uniform Resource Identifiers (URI) und Hypertext Transfer Protocol (HTTP) auf und ergänzt diese durch das Datenmodell RDF (Resource Description Framework) zur Repräsentation von Information, SPARQL (SPARQL Protocol And RDF Query Language) zu Abfrage von RDF-Daten und RDFS sowie OWL (Web

Dieser Text wird die genannten Linked-Data-Technologien nicht näher erläutern. Für weitergehende Informationen sei verwiesen auf das Glossar der DINI AG KIM⁹ und den kürzlich erschienenen Beitrag von Adrian Pohl und Patrick Danowski "Linked Open Data in der Bibliothekswelt: Grundlagen und Überblick".¹⁰

Offene Daten als grundlegende Anforderung

Grundlage eines jeden Fachinformationssystems sind Aufbau und Pflege des *Bestands*, um Zugriff auf relevante Ressourcen ermöglichen zu können sowie die *Metadaten* für Nachweis- und Recherchesysteme, so dass eine Ressource überhaupt aufgefunden werden kann. Um technisch in der Lage zu sein, optimale Recherche- und Nachweissysteme zu bauen, sollte der Zugriff auf die nötigen Metadaten möglichst einfach und reibungslos vor sich gehen. Idealerweise stellen Metadatenproduzenten ihre Daten unter eine offene Lizenz.¹¹

In Deutschland haben die Deutsche Nationalbibliothek (DNB) und die deutschen Bibliotheksverbünde bereits mit der Freigabe von Daten begonnen.¹² Der Bibliotheksverbund Bayern (BVB) und der Kooperative Bibliotheksverbund Berlin-Brandenburg (KOBV) sowie das hbz stellen ihre Daten sowohl als MARC oder MAB und auch als LOD bereit und bieten darüber hinaus regelmäßige Aktualisierungen an. Die DNB bietet den größten Teil ihrer Daten unter einer offenen Lizenz an mit einjähriger "Moving Wall". Für bestimmte Fächer lässt sich mit diesem Angebot an offenen Daten u. U. bereits ein guter FID-Recherchedienst aufbauen. Dabei sind Monographiedaten ganz gut abgedeckt, oft fehlen aber Artikelmetadaten und graue Literatur. Damit das "web of data" Realität wird, müssen weitere Daten geöffnet werden. Und dies geschieht am besten dort, wo diese Daten originär anfallen.

Möglichst viele Verlage, Ersteller von Fachbibliographien, die FIDs selbst und andere Metadatenprovider sollten ebenfalls dazu übergehen, ihre Metadaten sowie andere für die Recherche nützliche Daten (wie z. B. Inhaltsverzeichnisse und Abstracts) als Open Data unter CC0 zur Verfügung zu stellen. Ein weiterer wichtiger Schritt bestünde in der Überlassung einer kompletten Kopie der digitalen Veröffentlichung, z.B. um diese in Volltextsuchmaschinen zu indexieren oder um sie zur automatischen Anreicherung der Metadaten (Klassifizierung, Verschlagwortung usw.) zu nutzen oder/und um die Ressource langzeitverfügbar zu machen. Die weitere Verbreitung von Open Access wäre auch für diesen Zweck sehr erfreulich, ja sogar notwendig. Leider liegen zur Zeit oft nicht einmal Hochschulschriften als Open Access vor.¹³

Linked Data - Best Practices

Linked Open Data bedeutet - wie soeben ausgeführt - einerseits die offene Lizenzierung von Daten ("Open Data"). Andererseits bezieht sich das "Linked Data" in LOD auf eine Menge von Best Practices zur Datenpublikation, die auf Standards des W3C basieren.

LOD ist somit eng mit dem World Wide Web (WWW) verzahnt; es baut auf den gleichen Technologien und Standards auf. Oft wird das WWW auch als "web of documents", also als "Netz von Dokumenten", bezeichnet. LOD hingegen kann als "web of data", als "Netz von Daten", bezeichnet werden. Während im WWW also Dokumente miteinander verknüpft sind, so sind im LOD-Netz Daten miteinander verknüpft. Ein wesentlicher Unterschied von LOD und den Hyperlinks zwischen HTML-Dokumenten (also den "Links" in z.B. Webseiten) ist die Art der Verknüpfung: Das LOD Paradigma erzwingt bedeutungsvolle Beziehungen von Subjekt (also das, was verlinkt wird) und dem Objekt (also das, wohin verlinkt wird). Das ist im Vergleich zu bibliothekarischen Datenbanken nichts Neues, denn dort werden stets Aussagen gemacht wie "Buch X hat den Autor Y" und "Autor Y wurde geboren im Jahr Z".

Anders als bei Datenbanken, die im "deep web" versteckt sind, kann bei LOD jeder zum gemeinsamen Datenpool etwas beitragen - denn, wer eine Webseite anlegen kann, kann auch LOD herstellen und Informationen dem "web of data" hinzufügen. So wächst die LOD-Cloud, und zwar vor allem durch Verlinkung. Ein Beispiel: Wissenschaftlerin A schreibt einen Artikel B der Artikel C referenziert und Wissenschaftler D widerspricht. Diese Referenzen sind nichts anderes als Verlinkungen zwischen A und B, B und C und C und D, jeweils mit einer spezifizierten semantischen Relation. Werden diese Beziehungen in dem Artikel als LOD hinterlegt¹⁴ und erhält der Artikel einen HTTP-URI und ist also Teil des WWW, dann ist damit die LOD Datenbank gewachsen.¹⁵

Auf Basis einer großen Menge dezentral gepflegerter, untereinander vernetzter Daten sind Abfragen wie "gib mir alle Dokumente, die von Personen geschrieben wurden, die im 16 Jhd. geboren sind und mit 'Chemie' verschlagwortet sind" einfach realisierbar. Dabei ist es ausreichend, einem Metadatensatz lediglich zwei Aussagen ("RDF-Tripel") über ein Dokument hinzuzufügen: Der eine Eintrag verknüpft das Dokument mit einer LOD-fähigen Schlagwortdatei, z.B. der Dewey Decimal Classification oder der GND, der zweite Eintrag verknüpft das Dokument mit der Personen ID. Alle weiteren Daten (multilinguale natürlichsprachige Bezeichnung des Schlagwortes, Lebensdaten des Autors usw.) müssen nicht mit in das Dokument aufgenommen werden, sondern sind, da sie im WWW bereits vorliegen, von dort automatisiert abrufbar.¹⁶ In einem LOD-"Graphen" gibt es per se keinen von vornherein abgeschlossenen Datensatz, da alles mit allem verknüpft sein kann und somit weder Anfang noch Ende hat. Datenproduzenten und -konsumenten müssen selbst entscheiden, wo ein Datensatz anfangen und wo er sinnvollerweise aufhören soll, welche Daten als vertrauenswürdig angesehen werden und welche etwa Spam sind.¹⁷ Das Datenmodell von LOD lässt also unbegrenzte neue Aussagen zu, während bei herkömmlichen Austauschformaten die maximale Größe eines Titeldatensatzes festgelegt ist.¹⁸ Für LOD gibt es auch keine Beschränkung der möglichen Informationsdichte eines Metadatensatzes, also dessen, was in den herkömmlichen Datenmodellen die Metadatenfelder sind. Zwar lassen sich auch den überkommenen, für Magnetbänder optimierten Datenstrukturen aus den 1960er/1970er Jahren neue Felder teilweise durch mühselige Standardisierungsprozesse hinzufügen und teilweise auch beliebig lokal definieren (was zu Inkompatibilität der Kataloge führt), doch sind diese "Format"-Beschreibungen nicht offen und funktionieren deshalb im besten Fall nur in ihrem begrenzten System: der Bibliothekswelt. Die Metadaten der Bibliothekswelt sind aber viel zu wichtig, um sie nicht in das WWW zu integrieren. Sie sind dafür auch besonders gut geeignet: Tatsächlich existiert bereits in den älteren Austauschformaten MAB/MARC/PICA eine große Anzahl von Verlinkungen (zwischen Titeldatensätzen, zwischen Titel- und Normdatensätzen) gerade im deutschsprachigen Raum.

So, wie das WWW dezentral aufgebaut ist und es viele verschiedene Webseitenbetreiber gibt, so liegen auch die verlinkten offenen Daten verteilt im Netz. Dieses LOD wird von verschiedenen Interessenten aggregiert und integriert, sei es um in eigenen internen Datenbanken "zu verschwinden", um neue Webanwendungen herzustellen oder/und um wiederum als LOD veröffentlicht zu werden. Im letzteren Fall wurde den ursprünglichen Daten neue Daten hinzugefügt, die rekonsumiert werden können.¹⁹ So kann z.B. die DBpedia (also die LOD-Variante der

Wikipedia) den auf sie zeigenden Links folgen und die eigenen Daten mit z.B. Metadaten aus Bibliothekskatalogen anreichern.²⁰ Dies eröffnet eine interessante Perspektive für Bibliotheken: So könnte man sich vorstellen, dass - sobald diese Daten in die Wikipedia gelangt sind - in einzelnen Artikeln die Verfügbarkeit einer referenzierten Ressource in den (durch die IP des Nutzers ermittelten) nächstgelegenen Bibliotheken angezeigt wird.

Erfahrungen aus dem vascoda-Projekt

Von 2008 bis Ende 2010 war der Autor Pascal Christoph beim hbz mitverantwortlich für Datentransformation und Suchmaschinenadministration für vascoda. Für das vascoda-Portal wurden Daten aus mehr als 40 verschiedenen Quellen (SSGs, ViFas, Fachportalen und auch Verbunddatenbanken) in verschiedensten Formaten für die Suchmaschine aufbereitet. Auch wenn in den DFG- Richtlinien eine Portallösung für FIDs nicht als zwingend vorausgesetzt wird, ist es sinnvoll, an dieser Stelle knapp auf die Erfahrungen mit vascoda, vor allem dem vascoda-Portal, einzugehen. Mit der Anwendung der LOD-Best-Practices durch FIDs wäre die Gefahr der Wiederholung vieler Probleme von vascoda nämlich von vornherein ausgeschlossen. Einige konkrete Probleme, die sich durch einen LOD-Einsatz vermeiden lassen, seien im Folgenden genannt.

Vascoda wurde als bloße Backlink-Maschine zu den jeweils eigenen Portalen konzipiert, d. h. viele wichtige Metadaten (u. a. Bestandsdaten) waren in vascoda gar nicht drin, sodass die an Bestandsnachweisen interessierte Benutzerin gezwungen war, in das jeweilige fachspezifische Portal zu wechseln. Solche unnötigen Portalbrüche werden aber als verwirrend und bestenfalls als umständlich empfunden - zumal, wenn das vascoda Portal so beworben wurde, als ob der Zugriff auf die Ressourcen direkt möglich gewesen wäre. Ebenso durften teilweise Schlagwörter zwar in vascoda indexiert und also gesucht, nicht aber zur Anzeige gebracht werden. Benutzer, die die eingegebenen Suchwörter nicht auch in der Trefferliste wiederfinden, begegnen den Ergebnissen eher skeptisch, da sie an das Hervorheben von Suchbegriffen in der Kurztrefferanzeige gewöhnt sind, das in fast allen Internetsuchmaschinen üblich ist. LOD setzt den Willen zur Bereitstellung eigener Daten für eine freie Nutzung u. a. durch die offene Lizenzierung der Daten voraus. Dadurch wird Datenaggregatoren größtmögliche Flexibilität bei der Einbindung der Daten gegeben.²¹.

Durch die Konzentration auf das vascoda-Portal als Mensch-Maschinenschnittstelle wurden Maschine-Maschineschnittstellen vernachlässigt. Und doch war das Interesse an den Daten groß genug, um vascoda in das "WorldWideScience"-Portal einzubinden. Dies geschah mittels des sogenannten HTML-Scrapings²² Erst am Ende des Projekts wurde die Einbindung in Dienste der USB Köln über eine Schnittstelle testweise realisiert. Gemäß dem LOD-Paradigma hingegen sind die Daten immer schon maschinenlesbar. D.h. die Daten liegen per definitionem strukturiert und maschinenlesbar vor, sodass die Einbettung der Daten in andere Applikationen sehr viel einfacher und stabiler ist.

Die Softwareentwicklung rund um vascoda geschah nicht als Open Source. Dadurch konnte sich keine Technikergemeinschaft ("Community") rund um das Projekt aufbauen, wodurch eine nachhaltige Entwicklung verhindert wurde. Ganz anders verhält es sich bei Werkzeugen, die im Umkreis von LOD entstehen, sei es [culturegraph.org](#), die LOD-Graphdatenbanken wie Virtuoso, 4store, LOD-kompatible Suchmaschinen wie elasticsearch, Datenreicherungswerzeuge wie SILK, oder die Werkzeuge der von der DFG geförderten Forschungsgruppe "Agile Knowledge Engineering and Semantic Web" ([AKSW](#)) in Leipzig. Natürlich gibt es auch einige proprietäre Angebote, doch existieren immer auch leistungsfähige Open-Source-Software-Alternativen, die sogar meist kostenlos verteilt werden, dafür aber auch meist mit kostenpflichtigen Serviceangeboten daherkommen.

Anders als im vascoda-Projekt, anders als wohl generell früher im Umfeld von bibliothekarischen Institutionen, begünstigt LOD eine weniger hierarchisch und zentralistisch strukturierte Arbeitsweise. Es eignet sich hervorragend als Basis zum Aufbau einer verteilten Community mit flachen Hierarchien. Ein gutes Beispiel liefert die Entstehung eines gemeinsamen Metadatenschemas für bibliographische Daten im deutschsprachigen Raum: Den verschiedenen LOD-Produzenten (DNB, hbz, BVB, Hebis u. a.) wurde zu keinem Zeitpunkt ein gemeinsames Metadatenschema oktroyiert, sondern es hat sich durch "Best Practices"-Konventionen herausgebildet, die mittlerweile - organisiert innerhalb der DINI-AG-KIM-Gruppe "Titeldaten" - in eine erste Version der "Empfehlung für die RDF-Repräsentation bibliografischer Daten"²³ gemündet sind. Auch hat sich im LOD-Bereich eine Arbeitsweise etabliert, die es erlaubt, auf neue Entwicklungen und Anforderungen flexibel zu reagieren. Diese Arbeitsweise ist vergleichbar mit der "[agilen Softwareentwicklung](#)", deren Motto des "release early, release often" ein ständiges Testen und Anpassen von Software an konkrete Bedürfnisse erlaubt. Komplizierte Organisationsprozesse und bürokratische Aufwände können so vermieden werden bzw. stehen der Arbeitskette nicht unbedingt im Weg - die Abhängigkeit von Entscheidungen auf oberer Ebene (z.B. wie das Datenapplikationsprofil aussehen soll) verhindern nicht, dass Daten bereits transformiert werden und über Nachweissysteme recherchierbar sind. Damit sind evolutionäre, oder auch "iterative", Anpassungen möglich - es ist nicht notwendig das "perfekte" Datenapplikationsprofil zu erstellen. Diese Arbeitsweisen, die nicht nur - aber auch - im LOD-Umfeld vorherrschen, sind somit höchst flexibel.

Wir haben in diesem und dem letzten beiden Kapiteln gesehen: LOD unterstützt auf der einen Seite durch den dezentralen Ansatz, die Flexibilität des Datenmodells und die offene Lizenzierung von Daten direkt Aufbau und Pflege einer Informationsinfrastruktur. Auf der anderen Seite fördert LOD indirekt den Aufbau einer solchen Infrastruktur, weil etwa bestimmte Arbeitsweisen (agile Entwicklung), ein offener Austausch und Kooperation oder die Nutzung von Open-Source-Software im LOD-Kontext weit verbreitet sind. Da im Vergleich zu den konkreten Technologien die "weichen Faktoren" der Kommunikation und Organisation von grundlegenderer Wichtigkeit sind, wird im folgenden zunächst auf diese näher eingegangen, ehe die konkrete Aufgabenumsetzung unter Nutzung von LOD angesprochen wird.

Cooperare necesse est

Die umfangreichen und teilweise komplexen Aufgaben, die FIDs zu bewältigen haben, und die Knappheit der dafür zur Verfügung stehenden Ressourcen legen eine Kooperation der betroffenen Einrichtungen nahe - nicht umsonst weist die DFG auf das Potential der Bündelung von Querschnittsaufgaben in Kompetenzzentren hin. Gute Bedingungen herzustellen für den Aufbau und die Pflege eines so nachhaltigen wie innovativen Systems der Informationsversorgung ist allerdings keine leichte Aufgabe. Konkrete Technologien spielen beim Aufbau eines solchen Systems eine untergeordnete Rolle, sie können seine Entstehung allenfalls unterstützen. Wichtiger sind bestimmte Formen der Zusammenarbeit, der Kommunikation, des Teilens von Erfahrungen und Wissen. Als pointierte Darstellung der wichtigsten Aspekte sei an dieser Stelle der im Linked-Open-Data-Bereich des hbz entstandene Entwurf eines Manifests zitiert, das der Beförderung einer entsprechenden kollaborativen Arbeitsweise für den Aufbau einer zukunftsfähigen, innovativen und durch öffentliche Einrichtungen kontrollierten Bibliotheks-IT-Infrastruktur dienen soll:

Libraries Empowerment Manifesto

We support the formation and development of an international library infrastructure that is:

- * future-proof: it supports the development of sustainable solutions;
- * progressive: it enables rapid development and quick adaptation to upcoming challenges;
- * empowering: it empowers libraries to control to the maximum the infrastructure underlying their services for collection, indexing and dissemination of published knowledge.

In order to reach the goal of a future-proof, progressive infrastructure that empowers libraries to control their future to the maximum, we endorse the following principles:

- Openness. To the extent possible, we share our data, content, tools and ideas on the web, according to the requirements of the [Free Software Definition](#) and the [Open Definition](#) and the best practices of the wider web community.
- Transparency. Rules of conduct, project plans, minutes, plans for new features, and other artifacts are open, public, and easily accessible.
- Inclusion. Our projects are open to all; we provide the same opportunity to all. Everyone participates with the same rules; there are no rules to exclude any potential contributors which include, of course, direct competitors in the marketplace.
- Meritocracy. The more you contribute the more responsibility you will earn. Leadership roles are also merit-based and earned by peer acclaim.
- Reuse. We actively seek for solutions others have developed for a specific problem and are happy to reuse and adapt them for our context.
- Credit. We give credit where credit is due.²⁴

We know that libraries, archives, museums, and related organizations can't change from one day to the next. Nonetheless, we seek to organize our work according to these principles, even if we might start imperfectly and/or only within parts of the organization.

Das Manifest ist auf Englisch verfasst, weil Kollaboration an Landesgrenzen nicht Halt macht und Erfahrungen und Lösungen, die auch für nicht-deutsche Institutionen von Nutzen sein könnten, im besten Fall auch diesen zugänglich gemacht werden sollten.

Die einzelnen Punkte des Manifests seien hier kurz auf Deutsch erläutert:

- Offenheit ('Openness') im Hinblick auf die rechtliche Seite der Lizenzierung von Software, Daten und Inhalten ermöglicht erst die reibungslose Übernahme und Anpassung andernorts entstandener Lösungen.²⁵
- Transparenz ('Transparency') ermöglicht es Außenstehenden, einen detaillierten Überblick über ein Projekt, die dabei entstandenen Probleme und betretenen Irrwege sowie über seine Erfolge und Errungenschaften zu bekommen und die verantwortlichen Personen als mögliche Kontakt- und Kooperationspartner zu finden.
- Inklusion ('Inclusion') steht für die soziale Offenheit, die interessierten Personen Partizipationsmöglichkeiten in einem Projekt verschafft.
- Meritokratie ('Meritocracy') bedeutet, dass Entscheidungen nicht auf der Basis von Hierarchien, sondern von Erfahrung und Expertise getroffen werden.
- Wiederverwendung ('reuse') als leitendes Prinzip stellt sicher, dass die Nutzung und Anpassung andernorts entwickelter Lösungen Priorität besitzt vor der Eigenentwicklung, wodurch das weit verbreitete [Not-invented-here-Syndrom](#)²⁶ überwunden werden soll.
- Namensnennung ('credit') stellt sicher, dass bei der Wiederverwendung andernorts entstandener Daten, Inhalte oder Software die Urheber entsprechend aufgeführt werden.²⁷

Wir denken, dass dieses Manifest die wichtigsten Eigenschaften einer kooperativen Arbeitsweise für Aufbau und Betrieb einer innovativen, nachhaltigen und unabhängigen Metadatenökologie nennt. Wie bereits angemerkt handelt es sich bei dem Manifest aber um einen Entwurf, um einen Vorschlag, der mit anderen zu einer Version überarbeitet werden soll, die dann erste Institutionen unterzeichnen können. Alle Interessierten sind herzlich eingeladen, sich an der weiteren Entwicklung zu beteiligen.²⁸

Konkrete Kooperationsmöglichkeiten bei Querschnittsaufgaben

Den FIDs wird ziemlich freie Hand dabei gelassen, die Informationsbedürfnisse des Wissenschaftlers zu befriedigen. Das bedeutet unter anderem, dass FIDs die Wahl haben, technische Infrastrukturen selbst aufzubauen und zu pflegen oder aber lediglich als Vermittler zwischen den Dienstleistern (Datenerzeugern auf der einen, Datenverarbeiter auf der anderen Seite) aufzutreten. In den DFG-Richtlinien steht dazu im Kapitel "Querschnittsaufgaben":

"Im System der Fachinformationsdienste sind für jene technisch-organisatorischen Arbeiten, die für die einzelnen Fachgebiete gleichzeitig durchzuführen sind und zugleich einen hohen Arbeitsaufwand und besondere Expertise erfordern, Querschnittsbereiche vorgesehen, in denen die Betreuung dieser Aufgaben gebündelt wahrgenommen wird. Dies betrifft vor allem den Umgang mit digitalen Medien. Dabei ist es ein ausdrückliches Ziel, durch die Aufgabenbündelung eine Entlastung zugunsten der fachlichen Arbeit in den einzelnen Fachinformationsdiensten zu erreichen und durch Synergieeffekte auch das Potential für Einsparungen optimal auszunutzen."²⁹

Dieser Anstoß der DFG zur Kooperation und zum Aufbau von Kompetenzzentren für "Querschnittsaufgaben" ist zu begrüßen. Einige Möglichkeiten zur Herstellung von Synergieeffekten sollen im Folgenden näher betrachtet werden insbesondere mit Blick auf die Arbeit mit und Bereitstellung von Daten. Es geht also vor allem um folgende Aufgaben:

- Metadatenmapping, Datentransformation und automatische Datenanreicherung
- Aufbau von Rechercheindizes und Web-APIs
- Intellektuelle Erstellung, Verbesserung und Anreicherung von Metadaten

Metadatenmapping, Datentransformation und automatische Datenanreicherung

Bei jedem Dienst, der eine Recherche über Daten aus verschiedenen Quellen ermöglicht, fallen regelmäßig und langfristig Aufgaben im Kontext von Datenaggregation, Metadatenmapping, Datentransformation und -anreicherung an. Bisher haben meist unterschiedliche Institutionen jeweils eigene Expertisen und eigene Lösungen für diese Aufgaben entwickelt. Häufig findet auch eine gegenseitige Unterstützung und Zusammenarbeit einiger Institutionen statt, indem etwa entwickelte Software-Anwendungen weitergegeben werden. Kooperationen finden aber bisher eher unter der Hand und nicht transparent im Web statt.

Diese Situation ändert sich seit einiger Zeit. Mittlerweile sind mächtige Werkzeuge zur Metadatentransformation als freie Software im Web verfügbar. Beispiele sind das im culturegraph-Projekt entwickelte Metafacture³⁰ und das zentrale Tool des LibreCat, nämlich die Open-Source-Software Catmandu³¹(<https://github.com/LibreCat/Catmandu>).

Bei der Entwicklung des Datentransformationswerkzeugs "Metafacture", das auch im hbz verwendet wird, war Modularität und somit Flexibilität und Wiederverwendbarkeit ein zentrales Anliegen. So sind die für die Datentransformationen notwendigen Beschreibungen von Feldüberführungen ("Mapping") in Konfigurationsdateien hinterlegt.³² Diese Mappingdateien werden über eine Workflowdatei³³ gesteuert. Die Adaption dieser zwei Dateien erlaubt es, den kompletten Transformationsprozess zu kontrollieren.³⁴ So können auch ohne Programmierkenntnisse Datentransformationen definiert und durchgeführt werden.

Auch für die automatische Datenanreicherung gibt es freie Software, die diese Aufgabe unterstützt. Z. B. können mit dem im EU-Projekt "LOD2" entwickelten Silk und auch mit Limes Zusammenführungen von LOD-Daten durchgeführt werden. Auch "culturegraph" bietet eine Plattform für Datenzusammenführung. Synergien entstehen hier ebenfalls durch die transparente, offene Zusammenarbeit im Bereich Algorithmenentwicklung. So sind z.B. erste Algorithmen zur Bündelung bibliographischer Ressourcen - etwa zum Zwecke der Deduplizierung oder FRBRisierung - offen publiziert.³⁵ Alle an automatischer bibliographischer Datenanreicherung Interessierten können auf diese Algorithmen zugreifen und sie für eigene Zwecke nutzen und weiterentwickeln.

Aufbau von Rechercheindizes und Web-APIs

Zum Aufbau von Rechercheindizes und APIs³⁶ eignet sich bereits vorhandene leistungsfähige Open-Source-Software wie etwa die Suchmaschinen Solr oder elasticsearch. Technische Dienstleister - dazu gehören auch Verbundzentralen und große Bibliotheken - können die nötige Infrastruktur bereitstellen und andere bei der Nutzung dieser Infrastruktur unterstützen. Dabei wird teilweise das gesamte Leistungsspektrum angeboten, von Beratung über Installation und Hosting bis zur (Weiter-)Entwicklung. Das hbz bietet beispielsweise seit Kurzem seinen Verbundkatalog über eine selbstentwickelte LOD-API an.³⁷ Bestandteil dieser API, die sich sehr einfach in eigene Anwendungen über das WWW einbinden lässt, sind auch Daten der Gemeinsamen Normdatei (GND), auf die mittels Vorschlagsuche nach Autoren gesucht werden kann, um auf diese Weise die für die Person eindeutige GND-ID zu bekommen. Diese API steht jedem offen und kann z.B. in Katalogisierungsklienten eingebunden werden. Dadurch entstehen hohe Synergieeffekte, da diese Funktionalität prinzipiell von allen Katalogisierenden benötigt wird jedoch nicht von jeder Institution neu entwickelt zu werden braucht. Die zugrundeliegende modulare Technik erlaubt es zudem, weitere Datenquellen aufzunehmen und z.B. auch dafür eine Such vervollständigung anzubieten. Möchte ein anderer Dienstleister diese Funktionalität auf eigenen Servern selber anbieten, so kann er auf die komplette Software - wie auch auf die Daten - zugreifen, um die Dienste nachzubauen und die Software weiter (mit) zu entwickeln. Schön wäre es, wenn dabei Softwareverbesserungen auch zurückfließen. Erfahrungen mit anderen Open-Source-Projekten zeigen, dass genau dies geschieht.

Hubs zur intellektuellen Datenanreicherung

Neben der sog. "weißen Literatur" gibt es "graue Literatur", also alles das, was nicht durch den Buchhandel vertrieben wird. Es stellt sich die Frage, wie ein FID auch noch relevante graue Literatur integrieren kann.

Eine Sammlung fachrelevanter Ressourcen zu betreiben erfordert viel intellektuelle Arbeit. Besonders im Bereich der grauen Literatur, die teilweise gar nicht erschlossen oder nur mit unzureichenden Metadaten ausgestattet ist, ist deshalb der Einsatz von sog. Crowdsourcing³⁸ unverzichtbar, vor allem vor dem Hintergrund der rapiden Zunahme an solchen Quellen. Damit steht der Nutzer nicht einfach vor einer immer größeren Informationsflut, sondern sieht sich einer mit ihm interaktiv agierenden Gemeinschaft von Gleichgesinnten gegenüber, die bereitwillig und getragen durch ebenenes Medium "Internet" an der Organisation dieses Wissens mithelfen kann.³⁹ Wenn diese Wissensorganisation dann nach LOD Prinzipien geschieht, dann sind die Ergebnisse dieser Wissensorganisation maschinell komfortabel verarbeitbar - die Informationsflut lässt sich also tatsächlich beherrschen. Die Wikimedia Foundation hat das erkannt, und mit Wikidata⁴⁰ eine Infrastruktur geschaffen, mit derer Teile der Wikipedia nach LOD-Prinzipien strukturiert werden.⁴¹

Flexible Arbeitsteilung

Bei der oben beschriebenen Prozesskette kann jeder FID sich entscheiden, an welcher Stelle er die notwendigen Arbeiten selbst umsetzt und welche Leistungen als "Querschnittsdienste" an Dienstleister abgegeben werden. Dies betrifft Auswahl der Daten, Datentransformation und (halb)automatische Datenanreicherung, Veröffentlichung der Daten als LOD, die Indexierung in spezielle Datenhaltungssysteme, die Bereitstellung von APIs über diese Datenhaltungssysteme, und die diese APIs einbindenden Nachweis- und Recherchesysteme und Portale.

Durch diese Modularität und strikte Trennung von Daten, Datenhaltung und Datenanzeige⁴² lassen sich die einzelnen Komponenten resp. die Dienstleister leichter austauschen. Der FID behält die Kontrolle und begibt sich weniger stark in alternativlose Abhängigkeiten. Verbesserungen

der Daten bleiben erhalten und werden nicht "vergessen" wenn ein Portal, ein Nachweissystem oder eine API abgeschaltet oder ausgetauscht wird, da die Daten nicht zu eng an diese Komponenten gekoppelt sind: "Data is stable – functions are not". Durch diese Entkoppelung der Daten von den sie verarbeitenden Systemen wird erst Nachhaltigkeit im Umgang mit Daten sichergestellt. Mit dieser Praxis wird eine Lehre aus James Governors Bonmot "Data matures like Wine, Applications like Fish" ("Daten altern wie Wein, Anwendungen wie Fisch")⁴³ gezogen.

Fazit

Es wurde gezeigt, warum eine LOD-Metadateninfrastruktur nachhaltig ist und wie diese zum Aufbau von FIDs-Dienstleistungen genutzt werden kann um z. B. bessere Recherche- und Nachweissysteme zu ermöglichen und die Daten in das WWW resp. in das "web of data" zu integrieren. Zudem wurde erläutert, dass große Teile dieser Infrastruktur bereits vorhanden oder in Entwicklung sind. Entscheidend für diese Metadateninfrastruktur ist, dass die Daten samt zeitnahen Updates offen lizenziert zur Verfügung stehen. Die Daten sind die Grundlage der Infrastruktur. Die Technik zur Datenhaltung, -transformation, -publikation und Recherche ist das geringere Problem - die Datenbeschaffung bleibt die größte Herausforderung. Hier können und sollten die FIDs eine entscheidende Rolle übernehmen.

Der Ansatz birgt aber auch ein Risiko, da die DFG die Investition in eigenentwickelte Gemeinschaftslösungen für eine Informationsinfrastruktur als kritisch für einen späteren nachhaltigen Betrieb gesehen hat und stattdessen eher kommerzielle und proprietäre Lösungen bevorzugte.⁴⁴ Zwar unterstützt die DFG lediglich die [Migration der Verbünde in vorhandene proprietäre Datenbanken wie OCLCs WorldShare und ExLibris' Alma](#)⁴⁵ gleichzeitig veröffentlicht OCLC die WorldCat-Daten auch als Linked Data und auch bei ExLibris gibt es seit längerem Bewegung in diese Richtung.⁴⁶ Es zeigt sich, dass sich viele zusätzliche Dienste und Dienstleistungen, wie z.B. die Integration in das WWW oder in Internetsuchmaschinen ohne Linked Data gar nicht oder nur umständlich bewerkstelligen lassen.⁴⁷

Dementsprechend würde der Aufbau einer LOD-basierten Metadateninfrastruktur durchaus bisherigen Weichenstellungen der DFG und die Nutzung von Plattformen wie OCLCs WorldShare und ExLibris' Alma ergänzen und könnte dabei viel zur Lösung der an die FIDs gestellten Aufgaben beitragen.

Referenzen

- Berners-Lee, Tim (2009): On the next web. URL: http://www.ted.com/talks/tim_bernies_lee_on_the_next_web.html.
- Christoph, Pascal (2013): Datenanreicherung auf LOD-Basis. In: (Open) Linked Data in Bibliotheken S. 139-167. URL: http://www.dr0i.de/lib/pages/Datenanreicherung_auf_LOD_Basis.html.
- Danowski, Patrick / Pohl, Adrian (Hg.) (2013): (Open) Linked Data in Bibliotheken. Walter de Gruyter GmbH, Berlin/Boston. URL: <http://dx.doi.org/10.1515/9783110278736>.
- DFG (2013a): DFG bewilligt drei Projekte zur Neuausrichtung überregionaler Informationsservices. In: Information für die Wissenschaft Nr. 11 | 15. März 2013. URL: http://dfg.de/foerderung/info_wissenschaft/archiv/2013/info_wissenschaft_13_11/index.html.
- DFG (2013b): Fachinformationsdienste für die Wissenschaft. Richtlinien für das DFG geförderte System der Fachinformationsdienste für die Wissenschaft. DFG-Vordruck 12.102 - 03/13. URL: http://www.dfg.de/formulare/12_102/12_102_de.pdf.
- DINI AG KIM (2013a): Empfehlung für die RDF-Repräsentation bibliografischer Daten (Textressourcen). URL: <https://wiki.dnb.de/x/cYMOB>.
- DINI AG KIM (2013b): Glossar. URL: <http://git.io/2CPvQ>.
- Diverse (2013): Massively-Multiplayer_Online_Bibliography. https://meta.wikimedia.org/wiki/Massively-Multiplayer_Online_Bibliography.
- Governor, James (2007): Why Applications Are Like Fish and Data is Like Wine. URL: <http://redmonk.com/jgovernor/2007/04/05/why-applications-are-like-fish-and-data-is-like-wine/>.
- Graf, Klaus (2013): Schlechte Bücher? Publikationsmöglichkeiten im 21. Jahrhundert als Herausforderung für Bibliotheken <http://archiv.twoday.net/stories/472713645>.
- Grant, Carl / Pohl, Adrian (2011): Ex Libris, Alma and Open Data . URL: <http://openbiblio.net/2011/08/11/ex-libris-alma-and-open-data/>.
- 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>.
- Pohl, Adrian (2013): Mit der DFG und CIB nach WorldShare und Alma. URL: <http://www.uebertext.org/2013/04/mit-der-dfg-und-cib-nach-wordshare-und.html>.
- Pohl, Adrian / Danowski, Patrick (2013): Linked Open Data in der Bibliothekswelt: Grundlagen und Überblick. In: Dies. (Hg.) (2013): (Open) Linked Data in-Bibliotheken, Berlin/Boston, S.1-44.
- Tochtermann, Klaus (2013): 10 Thesen zum zukünftigen Profil von wissenschaftlichen Informations-Infrastruktureinrichtungen mit überregionaler Bedeutung. URL: <http://www.zbw-mediatalk.eu/2013/08/klaus-tochtermann-zehn-thesen-zum-zukunftigen-profil-von-wissenschaftlichen-informationsinfrastruktureinri>

1. DFG (2013b): Fachinformationsdienste für die Wissenschaft. Richtlinien für das DFG geförderte System der Fachinformationsdienste für

- die Wissenschaft. DFG-Vordruck 12.102 - 03/13. URL: http://www.dfg.de/formulare/12_102/12_102_de.pdf.
2. DFG: Fachinformationsdienste für die Wissenschaft. Richtlinien für das DFG geförderte System der Fachinformationsdienste für die Wissenschaft. 2013, S. 4.
 3. Fachinformationsdienste für die Wissenschaft. Richtlinien für das DFG geförderte System der Fachinformationsdienste für die Wissenschaft. 2013, S. 5-6 .
 4. Siehe dazu auch Tochtermann, Klaus (2013): 10 Thesen zum zukünftigen Profil von wissenschaftlichen Informations-Infrastruktureinrichtungen mit überregionaler Bedeutung. URL:
<http://www.zbw-mediatalk.eu/2013/08/klaus-tochtermann-zehn-thesen-zum-zukunftigen-profil-von-wissenschaftlichen-informationsinfrastrukturen-insbesondere-6-7-und-9-Der-in-Tochtermanns-Thesen-verwendete-Begriff-des-Semantic-Web-ist-synonym-zu-Linked-Data-verwendet>
 5. Es ist dennoch zu betonen, dass es auch eine große Auswahl an proprietären Softwarelösungen gibt - insbesondere bei sogenannten "Triple Stores" oder Graphdatenbanken.
 6. Zitiert nach Tim Berners Lee, 1999: <http://lists.w3.org/Archives/Public/www-rdf-interest/1999Dec/0113.html>
 7. Berners-Lee ist ein britischer Physiker und Informatiker, Erfinder von HTML und Begründer des WWW, siehe https://de.wikipedia.org/wiki/Tim_Berners_Lee
 8. DINI AG KIM (2013b): Glossar. URL:<http://git.io/YWNx8w>
 9. Ebd.
 10. Pohl, Adrian / Danowski, Patrick (2013): Linked Open Data in der Bibliothekswelt: Grundlagen und Überblick. In: Dies. (Hg.) (2013): (Open) Linked Data in Bibliotheken. Walter de Gruyter GmbH, Berlin/Boston, S.1-44.
 11. D. h. es wird eine Lizenz verwendet, die kompatibel ist mit der [Open Definition](#). Für eine Definition von offenen *bibliographischen* Daten siehe die "Prinzipien zu offenen bibliographischen Daten" unter <http://openbiblio.net/principles/de/>. Ein umfangreicher rechtlicher Ratgeber zur Freigabe von Daten aus Bibliotheks-katalogen ist Kreutzer, Till (2011): Open Data – Freigabe von Daten aus Bibliotheks-katalogen. Ein Leitfaden. Hg. v. Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen. URL:
<http://www.hbz-nrw.de/dokumentencenter/veroeffentlichungen/open-data-leitfaden.pdf>.
 12. Für eine Übersicht über offene Datensets im internationalen Bibliothekskontext siehe die Gruppe "Bibliographic Data" im Open-Data-Verzeichnis "The Data Hub": <http://datahub.io/group/bibliographic> .
 13. Siehe dazu z.B. Graf, Klaus (2013): [Schlechte Bücher? Publikationsmöglichkeiten im 21. Jahrhundert als Herausforderung für Bibliotheken]<http://archiv.twoday.net/stories/472713645>.
 14. Z. B. durch RDFa in HTML oder XMP in PDF/A.
 15. Auf diese Weise ließe sich auch recht einfach aussagekräftigere Metriken berechnen wie z. B.: Wie oft wurde Artikel A referenziert und mit welcher Absicht (Widerspruch/Kritik, Zustimmung)? Wie oft wurden die Artikel, die Artikel A referenzierten, referenziert? Mit welchen Schlagworten sind die Ressourcen versehen?
 16. Sollte der verlinkte Service - aufgrund von Server- oder sonstigen Probleme - nicht mehr am Netz sein, heißt das natürlich nicht, dass diese verlinkten Daten verschwunden sind und nicht angezeigt werden können. Auch bei Linked Data verhindern etablierte Technologien wie das Caching oder auch eine lokale Indexierung solche Probleme.
 17. Da die Daten global verfügbar sind wird zukünftig immer weniger redundant mehrfachkatalogisiert. Die Arbeit wird aber nicht weniger, sondern verlagert sich in Richtung Moderation und Selektion von Daten aus dem "web of data".
 18. Bei MAB2 und MARC sind das 9999 Bytes.
 19. Wenn ein Datensatz dadurch größer als 9999 Bytes wird, so stellt dies, anders als bei MARC/MAB, kein Problem dar. Tatsächlich ist die Größe eines Datensatzes beliebig.
 20. Der Weg dahin wurde bereits eingeschlagen, siehe <https://github.com/dbpedia/dbpedia-links/>.
 21. Siehe dazu Christoph, Pascal (2013): Datenanreicherung auf LOD-Basis. In: Danowski, Patrick / Pohl, Adrian (Hg.): (Open) Linked Data in Bibliotheken, S. 139-167. Preprint: http://www.dr0i.de/lib/pages/Datenanreicherung_auf_LOD_Basis.html
 22. Das ist eine krude Methode, um aus einfachen, nicht extra zur Datenextraktion hergestellten HTML-Seiten strukturierte Daten zu gewinnen. Da der Aufbau von Webseiten sich ändern kann ist diese Technik instabil und von deren Einsatz abzuraten.
 23. DINI KIM AG (2013): Empfehlung für die RDF-Repräsentation bibliografischer Daten (Textressourcen). URL: <https://wiki.dnb.de/x/cYMOB>
 24. <http://blog.ninapaley.com/2011/06/27/credit-is-due/>
 25. Eine Entscheidung für Open-Source-Software bedeutet wohlgerne nicht, dass sich eine Institution eine eigene Softwareentwicklungsabteilung zulegen muss. Produkte wie [Kuali OLE²⁵](#) oder [edu-sharing²⁶](#) zeigen, wie öffentliche Einrichtungen private Unternehmen beauftragen können, um Open-Source-Software zu entwickeln.

26. Seite „Not-invented-here-Syndrom“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 1. April 2013, 02:23 UTC. URL: <http://de.wikipedia.org/w/index.php?title=Not-invented-here-Syndrom&oldid=116429841> (Abgerufen: 27. September 2013, 13:05 UTC).
27. Dies entspricht der wissenschaftlichen Praxis, deren Entstehungsgeschichte und Sanktionsmechanismen unabhängig vom Urheberrecht sind.
28. Wer das Manifest mit weiterentwickeln möchte, kann dies leicht auf einem Etherpad tun unter <http://etherpad.lobid.org/p/LEM>.
29. *DFG: Fachinformationsdienste für die Wissenschaft. Richtlinien für das DFG geförderte System der Fachinformationsdienste für die Wissenschaft. S. 10.*
30. (<https://github.com/culturegraph/metafacture-core>)
31. <https://github.com/LibreCat/Catmandu>
32. Ein Beispiel zur Überführung der ZDB-ISIL Datei in LOD ist: <http://git.io/6mgZrA>.
33. Beispiel einer sog. flux-Datei: <http://git.io/2nV4ng>.
34. Sollte einmal eine Funktionalität fehlen, so sind die Module einfach zu programmieren und der Software zufügbar. So entstanden bisher viele Module, z.B. für das Einlesen von MARC, MAB, Pica, csv, xml, ntriples.
35. Siehe dazu <http://hub.culturegraph.org/statistics/alg>.
36. Eine API ist eine Programmierschnittstelle, die es ermöglicht, von einem Softwaresystem auf ein anderes zuzugreifen. Bei einer Web-API geschieht dies über das WWW, siehe https://en.wikipedia.org/wiki/Web_API.
37. Siehe <http://api.lobid.org/>
38. ...also dem Einsatz von freiwilligen Benutzern.
39. Beispiele von funktionierendem Crowdsourcing sind u.a. Wikipedia, Amazon, IMDB. Früher war z. B. auch dmoz recht beliebt.
40. Siehe <https://de.wikipedia.org/wiki/Wikidata>
41. Im Wikimedia-Kontext ist mit der "Massively-Multiplayer_Online_Bibliography" eine recht neue Projektidee entstanden. Dabei geht es darum, mit Hilfe Freiwilliger (der "Crowd") Millionen von frei verfügbaren Essays und Artikeln usw. nach LOD Prinzipien zu annotieren. Auch wenn dieses Projekt vielleicht gar nicht starten wird oder auch, wenn es dabei bleibt, nur nicht-akademische, nicht-fiktionale Texte zu katalogisieren, so ist dieses Projekt doch eine interessante Blaupause für ein Modell, wie graue akademische Literatur gesammelt werden kann. In dieses Projektmodell passen nämlich hervorragend Bibliothekare oder eben FIDs mit ihrem Wissen. Sie können z.B. dabei helfen, Taxonomien für die Texte bereitzustellen oder, bei fehlenden Taxonomien, diese erstellen. Andere Menschen in dem Projekt werden Katalogisierungswerzeuge zur Verfügung stellen, in denen diese Taxonomien eingebunden sind, sodass sie vom Katalogisierer (der "Crowd") per Drop-Down-Menü und durch multilinguale Vorschlagssuchen kategorisiert werden können. Danach können die so kategorisierten Ressourcen automatisch gefiltert und die Metadaten automatisch an den jeweiligen FID gesendet werden, um sie der fachspezifischen Sammlung nach einem Moderationsprozess hinzuzufügen. Diese Information, also das Kategorisieren einer Ressource zu einem FID, ist wiederum eine interessante, aussagekräftiges Information, die der ursprünglichen Ressourcenbeschreibung zugefügt wird. Somit fließt die moderierende, qualitätssichernde Kontrolle der Fachprofis in die Ausgangsdaten zurück, wovon alle anderen, die diese Daten konsumieren, profitieren. Im LOD Paradigma erzeugt also die Nutzung einer Leistung potentiell einen Mehrwert für die Leistung. Es handelt sich um eine sog. Selbstverstärkung.
42. Softwareentwickler nennen dieses Verfahren das "Model-View-Controller" (MVC) Prinzip.
43. Governor, James (2007): Why Applications Are Like Fish and Data is Like Wine. URL: <http://redmonk.com/igovernor/2007/04/05/why-applications-are-like-fish-and-data-is-like-wine/>.
44. Vgl. Pohl, Adrian (2013): Mit der DFG und CIB nach WorldShare und Alma. URL: <http://www.uebertext.org/2013/04/mit-der-dfg-und-cib-nach-wordshare-und.html>.
45. Siehe DFG (2013a): DFG bewilligt drei Projekte zur Neuausrichtung überregionaler Informationsservices. In: Information für die Wissenschaft Nr. 11 | 15. März 2013. URL: http://dfg.de/foerderung/info_wissenschaft/archiv/2013/info_wissenschaft_13_11/index.html.
46. Siehe Grant, Carl / Pohl, Adrian (2011): Ex Libris, Alma and Open Data . URL: <http://openbiblio.net/2011/08/11/ex-libris-alma-and-open-data/> und die Linked-Open-Data-Arbeitsgruppe von Ex-Libris-Kunden: Joint IGELU/ELUNA Special Interest Working Group on Linked Open Data, URL: <http://igelu.org/special-interests/lod>.
47. Siehe dazu z.B. Schema.org, dass von den großen Internetsuchmaschinenbetreibern initiiert wurde, um die Webseiten für (Such-)Maschinen mittels Linked Data besser verstehbar zu machen. Einige ViFas haben schon begonnen, Schema.org in ihre Portale zu integrieren, z.B. <http://edoc.vifapol.de>. Econstor, ein Dokumentenserver der ViFas, bietet aus diesem Grund schon seit geraumer Zeit Linked Data an.