

Hausarbeit zum Seminar
Gesellschaftliche Strukturen im digitalen Wandel
im Wintersemester 2014/15

Einsatz von Semantic Web Technologien
und digitaler Wandel in Bibliotheken

Franz Teichmann

Leipzig, September 2015

Inhaltsverzeichnis

1	Einleitung	2
2	Theoretische Grundlagen	2
2.1	Vertrauensbegriff und Vertrauen in Technik	2
2.2	Paradigmenwechsel nach Thomas S. Kuhn	3
2.3	Lange Wellen	5
2.4	Digitaler Wandel und Open Culture	7
3	Digitale Infrastruktur von Bibliotheken	8
3.1	Datentypen in der Bibliothek	8
3.2	Dateiformate im Bibliotheksumfeld	9
3.3	Linked Data und der Semantic Web Stack	10
4	Entwicklungen im Bibliotheksstandard	14
4.1	Von MARC zu MARC 21	14
4.2	MAB und MAB 2	15
4.3	Die gemeinsame Normdatei	16
4.4	Der Weg zum Semantic Web	17
5	Auswertung	19

1 Einleitung

Die Bibliothekswelt ist neben anderen Bereichen des alltäglichen Lebens eines der Felder, welches stark vom digitalen Wandel der Moderne beeinflusst wird. Im Zuge meiner Bachelorarbeit vom Dezember 2013 zum Thema „Datenaggregation verschiedener Bibliotheksinformationssysteme auf Basis von Semantic Web Technologien“ konnte ich diese Entwicklung lokal an der Universitätsbibliothek Leipzig erleben und begleiten. Es ging zunächst darum, ein Datenmodell zu entwerfen, welches den neuen Anforderungen genügt und anschließend die vorhandenen Einträge effizient in das neue Datenformat umzuwandeln. Im letzten Schritt wurde die Verlinkung zwischen den einzelnen, separat umgewandelten Datensätzen hergestellt. So wurden z.B. die Indizes der Titeldaten mit den Identifikatoren der einzelnen, in der Bibliothek vorhandenen Exemplare verbunden.

Im Seminar zum interdisziplinären Modul „Gesellschaftliche Strukturen im digitalen Wandel“ konnte ich die Erfahrungen aus dieser Arbeit einbringen, um meinen Kommilitonen aus Fachbereichen fernab der Informatik die Schwierigkeiten, Intentionen und Besonderheiten der Datenmodellierung näherzubringen. Zusätzlich wurde im Seminar abstrahiert vom lokalen Beispiel der Leipziger Universitätsbibliothek über die internationale Entwicklung der Bibliotheken diskutiert.

Diese Hausarbeit soll die Arbeit aus einem anderen Blickwinkel betrachten und zusammenfassend die praktischen Ergebnisse mit grundlegenden formalen Begriffen aus der Theorie technischen und digitalen Wandels verbinden.

Dazu werde ich zunächst in einem kurzen Grundlagenkapitel die Begriffe des digitalen Wandels, den Vertrauensbegriff und den Begriff des Paradigmenwechsels zusammenfassen und vergleichend die wirtschaftliche Theorie langer Wellen in Bezug auf technische Entwicklung einbeziehen.

Anschließend soll ein kurzer Einblick in die gängigen historischen und modernen Dateiformate in der digitalen Infrastruktur von Bibliotheken und die Entwicklung diesbezüglicher Formate sowie die dazu in der jüngeren Geschichte erfolgten Diskussionen gegeben werden.

In den letzten Kapiteln werde ich die aktuelle Entwicklung genauer evaluieren und die Verbindung zu den im Grundlagenkapitel genannten Konzepten erläutern. Abschließend sollen die Fragen nach der Bewertung dieses digitalen Wandels, den neuen Möglichkeiten der Technologie und nach Prognosen für den weiteren Verlauf der Entwicklung diskutiert werden.

2 Theoretische Grundlagen

In diesem Kapitel werden die theoretischen Grundlagen für die späteren Erläuterungen gelegt. Es sollen dabei relevante Begriffe in Bezug auf den digitalen Wandel der Bibliotheken als spezielle Form des technischen Wandels aufgegriffen und so knapp wie möglich zusammengefasst werden. Es sei hierbei auch auf die Literatur zum Thema verwiesen.

2.1 Vertrauensbegriff und Vertrauen in Technik

Vertrauen nach Luhmann [15] findet zunächst auf sozialer Ebene zwischen Menschen statt und beinhaltet die Überzeugung von der „Richtigkeit oder Redlichkeit“ [15] zwischen Personen trotz einer Komponente der Ungewissheit. Es werden auf Basis von begründeten Erwartun-

gen Hypothesen über eintreffende Ereignisse zum Zweck des Handlungsvollzuges getroffen, um genau diese Komponente der Ungewissheit zu reduzieren und eine praktikable Handlungsfähigkeit herzustellen.

Das Vertrauen in Technik als spezifische Form dieses Vertrauensverhältnisses hat in unserem Kontext eine zentrale Bedeutung, denn es nimmt auf den ersten Blick eine Sonderstellung im alltäglichen Umgang ein. Trotzdem lässt sich, da Technik stets von Menschen gemacht ist, diese Form der Erwartungshaltung bzw. dieses Verhältnis gegenüber einem Artefakt oder stellvertretend einer technischen Methodik wie z.B. einem Verschlüsselungsverfahren auf das Verhältnis zu einem anderen Menschen herunterbrechen. Es werden dabei eigene und fremde Erfahrungswerte einbezogen und auf diese Weise Erwartungshaltungen zu begründeten Erwartungshaltungen gegenüber der Technik stellvertretend für einen anderen Menschen verfeinert. Umgangssprachlich wird diese gestellte Erwartung oft als der „Stand der Technik“ formuliert.

Aufgrund dieses Prozesses ist Vertrauen stets individuell geprägt und lässt sich weder sinnvoll objektiv messen noch vom Individuum auf eine Menschengruppe abstrahieren. Trotzdem spielt Vertrauen auch für den Handlungsvollzug von Gruppen eine Rolle. Dies geschieht auf einer höheren Ebene auf Basis von Kommunikation und Abstimmung über die individuellen Erwartungen, welche kooperativem Handeln vorausgeht. Auf Basis der Ergebnisse des kooperativen Handelns findet in Form einer kontinuierlichen Reflexion eine Re-Evaluierung der begründeten Erwartungen statt.

Dieser Prozess der Anpassung des Vertrauensbegriffs findet kontinuierlich auch bei der Weiterentwicklung bzw. dem Voranschreiten der Technik statt. Dies soll hier der Einstiegspunkt in die Betrachtungen zum technischen bzw. digitalen Wandel sein.

In der Informatik existieren kaum umfassende Theorien zur Evaluierung technischen Wandels, weshalb in den folgenden Abschnitten auf der Suche nach einer möglicherweise anpassbaren Theorie ein Exkurs in die Betrachtungen zum Paradigmenwechsel nach Kuhn in der Wissenschaftstheorie bzw. Sozialwissenschaft sowie die Theorie der langen Wellen aus den Wirtschaftswissenschaften stattfinden soll.

2.2 Paradigmenwechsel nach Thomas S. Kuhn

In seinem Buch „Die Struktur wissenschaftlicher Revolutionen“ [12] beschreibt Kuhn umfassend die Begriffe Paradigma und Paradigmenwechsel und setzte damit einen Meilenstein in der Wissenschaftsgeschichte, der Philosophie und der Soziologie.

Am Beispiel der kopernikanischen Wende zeigt Kuhn auf, dass zwar die zu dieser Zeit gängigen Modelle in der Lage waren, die neuen Beobachtungen am Sternenhimmel in ihre Berechnungen im Ptolemäischen Weltbild einzubeziehen, aber die Bahnberechnungen immer komplexer wurden. Zwar bot sein Paradigma eine Vereinfachung dieser Berechnungen, aber aufgrund der gewohnten Formeln aus dem alten Weltbild gelang ihm kein greifbarer Mehrwert in der Berechnungen der Planetenlaufbahnen und die Theorie wurde von seinen Kollegen abgelehnt. Erst später wurde mit den Berechnungen von Galileo Galilei ein tatsächlicher Paradigmenwechsel möglich.

Auf Basis dieser und weiterer Analysen von Paradigmenwechseln aus der Chemie stellt Kuhn die These auf, dass der Übergang zu einem neuen Paradigma weniger eine Frage besserer Argumente oder empirischer Ergebnisse ist als vielmehr ein soziologischer Prozess auf Ba-

sis von Überzeugungen und der Option auf zukünftigen wissenschaftlichen Fortschritt. Neue Hypothesen in der Wissenschaft stehen für ihn immer im Zusammenhang mit dem aktuellen Wissenschaftsparadigma, weil sie nicht außerhalb eines Paradigmas miteinander verglichen werden können. Dies erklärt auf einfache Weise die mit einem Paradigmenwechsel einhergehende Überzeugungsarbeit und den erlebten Widerstand beim Herausreißen der Menschen aus deren gewohnten Denkmustern.

Diese Theorie erntete verständlicherweise starke Kritik, da sie den Wechsel auf ein neues Paradigma nicht als logische Schlussfolgerung auf empirische Beobachtungen betrachtet, sondern eine menschliche Komponente in Form des Diskurses in wissenschaftliche Revolutionen hereinbringt und so scheinbar wissenschaftliche Errungenschaften abwertet. Er argumentiert anhand von Beispielen, dass sich nach einem unkämpften Paradigmenwechsel die allgemein als richtig erachtete Betrachtungsweise des Sachverhaltes ebenso verändert wie die Betrachtung des Weges vom Alten zum Neuen. Aus dieser neuen Perspektive schließlich wird der erfolgte Paradigmenwechsel fälschlicherweise als rein logische Konsequenz gesehen und der Diskurs ausgeblendet.

Es gibt laut Kuhn 5 Phasen des Paradigmenwechsels:

1. Prä-Paradigmenphase

Diese findet nur ein einziges Mal statt und bezeichnet die Phase, in der keine allgemein anerkannte Theorie zu einem bestimmten Sachverhalt vorherrscht, sondern lediglich Hypothesen und unzusammenhängende kleine Theorien existieren. Es findet ein wissenschaftlicher Austausch über Möglichkeiten von Experimenten oder empirischen Erhebungen statt.

2. Normal Science

Dieser von ihm geprägte Begriff bezeichnet die Wissenschaftsumgebung nach der Bildung eines allgemeinen Konsens. Es werden Beobachtungen und Experimente des „normalen“ Wissenschaftsalltages angestrebt, welche das nun existierende Paradigma stützen. Es findet eine Akkumulation dieser Beobachtungen statt, ohne das Paradigma zu hinterfragen oder anzugreifen.

3. Krise

Normalerweise lassen sich Beobachtungen und Widersprüche, welche nicht in das aktuell anerkannte Muster des Paradigmas passen, in der vorangegangenen Phase des alltäglichen Wissenschaftsbetriebes auflösen. Nur, wenn diese das Paradigma widerlegen oder in starkem Widerspruch stehen, befindet sich die Wissenschaftsgemeinde nach starken Anstrengungen in dieser Phase. Sie ist gekennzeichnet von vielen mitunter stark voneinander abweichenden Hypothesen und intensiver Diskussion.

4. Wissenschaftliche Revolution

Erst in dieser Phase werden die dem Paradigma zugrunde liegenden Axiome ernsthaft hinterfragt und analysiert. Danach treten an deren Stelle neue Axiome und ein neues Paradigma wird aufgebaut, welches erst nach und nach das alte Paradigma ablöst.

5. Post-revolutionäre Phase

Das neue Paradigma ist allgemein von der Wissenschaftsgemeinde anerkannt und diese geht wieder in Phase 2 über, indem kleine Unstimmigkeiten innerhalb des neuen Paradigmas durch Anpassung aufgelöst werden. Es wird versucht, dieses Paradigma weiter zu stärken.

Diese Theorie und die 4 Phasen des Paradigmenwechsels (die Prä-Paradigmenphase sei außen vor gelassen) lassen sich meiner Meinung nach mit entsprechenden Anpassungen auch an den Eigenschaften des technischen Wandels erkennen. Betrachtet man diesen nicht als kontinuierliches Voranschreiten, sondern als zyklischen Prozess in Form von neuen „Meilensteinen“ oder „major releases“, sind die Phasen zwei bis fünf erkennbar, in welchen die Technologie innerhalb ihrer Möglichkeiten voranschreitet, an ihre Grenzen stößt und anschließend grundlegend überarbeitet in neuer Version veröffentlicht wird. Nach der Re-Definition des „aktuellen Standes des Technik“ muss auch das Vertrauen in diese bestimmt werden nach dem in Kapitel 2.1 beschriebenen Prozess.

Stärker als beim Wissenschaftsprozess wird diese Re-Definition in vielen Anwendungsbeispielen wie Alltagsgeräten von wirtschaftlichen Prozessen begleitet, weshalb im kommenden Abschnitt eine Betrachtung der Theorie langer Wellen in der Wirtschaft und deren Verzahnung mit technischer Entwicklung erfolgen soll. Auch dort spielen zyklische Entwicklungsphasen eine entscheidende Rolle.

2.3 Lange Wellen

In seinem Aufsatz „Die Langen Wellen der Konjunktur“ von 1926 stellte Kondratjew anhand empirischer Daten aus verschiedenen Ländern die Existenz langer Konjunkturwellen mit einer Dauer von mindestens 40 Jahren fest. Die Ursache dieser langen Wellen liegen für ihn in den Gesetzmäßigkeiten des Kapitalismus und neue technische Entwicklungen sind nur die Folgen solcher wirtschaftlicher Umschwünge.

Diese seitdem heiß diskutierte Theorie wurde unter anderem ab 1933 und 1939 von Joseph Schumpeter in seiner Arbeit über Konjunkturzyklen aufgegriffen und der Begriff der Kondratjew-Zyklen geprägt, in deren Fortschreiten Innovationen und technische Entwicklung in Form von so genannten Basisinnovationen eine maßgebliche Rolle spielen.

Auch der genaue Ablauf der Wellen ist umstritten, doch allgemein wird von vier wiederkehrenden Phasen ausgegangen, siehe auch [8].

1. Wachstumsphase

Ungefähr 25 Jahre lange Phase wirtschaftlichen Wachstums und begleitet von Investition in Forschung und Verfeinerung existierender Technologie.

2. Primäre Rezession

Bezeichnet einen Einschnitt in die Wirtschaft, nachdem diese an ihre Beschränkungen gestoßen ist und einen Wandel bzw. eine Re-Evaluierung der bestehenden Situation erfordert.

3. Plateau-Periode

Bezeichnet eine ungefähr 10-jährige Phase flachen Wachstums, in welcher ein struktureller Wandel stattfindet und sich die Ökonomie stärker in Richtung Konsum richtet und in einer ungefähr 3-jährigen allgemeinen Krise mündet.

4. Sekundäre Depression

Auch genannt der Kondratjew-Winter handelt es sich hierbei um eine etwa 15-jährige deflationäre Phase, in der eine Neuausrichtung und das Finden einer neuen Basis für den kommenden Zyklus stattfinden.

Trotz zahlreicher wissenschaftlicher Beiträge zur Weiterentwicklung der empirischen Argumentation hin zur Formulierung einer konkreten Theorie langer Wellen ist die Existenz dieser bis heute umstritten. Hauptkritikpunkte sind die stark veränderten wirtschaftlichen Rahmenbedingungen und Prozesse im Laufe der Jahrhunderte, die das Aufstellen einer allgemein gültigen Theorie der langen Wellen erschweren. Eine weitere Problematik ist die grundlegend kurzlebige Dynamik der Wirtschaft, in der sich viele zyklische Prozesse überlagern und mitunter global zeitverzögert stattfinden.

Soll nun die Rolle, die technische Innovationen bzw. Paradigmenwechsel in der Wissenschaft für diese langen Wirtschaftszyklen spielen, und die Verbindung dieser Zyklen mit den Phasen des Paradigmenwechsels nach Kuhn übertragen auf technischen Wandel untersucht werden, fällt es zunächst einmal schwer, eine feste kausale Richtung der Argumentation festzulegen. Laut Schumpeter spielen so genannte Basisinnovationen und damit technische Entwicklung eine treibende Rolle, um nach einer Depression einen neuen Kondratjew-Zyklus anzustoßen.

Es ist also klar, dass die Wissenschaft durch Innovationen und die Entwicklung neuer Technologien Einfluss auf das wirtschaftliche System nimmt. Auf der anderen Seite sind wissenschaftliche Arbeit und Entwicklung von Technologie natürlich von Fördergeldern abhängig und gerade in Zeiten wirtschaftlicher Krisen sind auch Investitionen in Forschungsprojekte schwieriger.

Auch die öffentliche Meinung, welche unter anderem von Wirtschaftskrisen beeinflusst wird, trägt nach Kuhn zum Durchsetzen eines Paradigmenwechsels bei, wie im vorherigen Kapitel erläutert.

Klar ist, dass beide Systeme in bedingter Abhängigkeit zueinander stehen und sie sich, da beide ein Teil des gesellschaftlichen öffentlichen Raumes sind, beeinflussen müssen. Gleichzeitig findet der technische Wandel parallel in vielen verschiedenen Bereichen zeitversetzt statt und somit kann nicht jede Innovation bzw. jeder dieser Innovationszyklen einen nachweisbaren Einfluss auf das System der langen Wellen nehmen.

Meiner Meinung nach stellt sich die bedingte Abhängigkeit beider Systeme so dar, dass in kritischen Zeiten das jeweilige System anfälliger für äußeren Einfluss ist. So ist in Phase 4 der langen Wellen die Depression mit dem Einläuten der nächsten Wachstumsphase beendet, welches z.B. durch das Einbringen eines neuen Wissenschafts- oder Wirtschaftsparadigmas bzw. einer neuen Basistechnologie erfolgen kann, wie in Phase 4 der Innovationszyklen nach Kuhn beschrieben. Ebenso kann Phase 3 der Innovationszyklen nach Kuhn, die Krise, bei der die Wissenschaft mit dem etablierten Modell an Grenzen stößt, mit Phase 2 der langen Wellen, der primären Rezession, verglichen werden. Ich vermute, dass diese wissenschaftliche Krise gegeben einer entsprechenden Tragweite auch dazu führen kann, die Wachstumsphase zeitversetzt in eine Rezession zu überführen.

Es ist schwer abzuschätzen, ob und inwiefern die aktuelle Diskussion über den digitalen Wandel und die Entwicklungen der Informationsgesellschaft, einschließlich der Bibliotheken, genügend Tragweite für einen derartigen Einfluss auf das wirtschaftliche Gefüge hat.

Die Charakterisierung des digitalen Wandels als speziellen Fall technischer Entwicklung soll im kommenden Kapitel beschrieben werden.

2.4 Digitaler Wandel und Open Culture

Der digitale Wandel bezeichnet die Entwicklung von Technologie und Strömungen in der digitalen Welt im Gegensatz zum technischen Wandel als Begriff für die Entwicklung von technischen Geräten. Da jedoch auch die digitale Welt und damit der digitale Wandel nicht im übertragenen Sinne im „luftleeren Raum“ stattfindet, sondern stets an Geräte gebunden ist, kann man beide Begriffe für die Beschreibung einer technologischen Entwicklung nicht strikt voneinander trennen. Vielmehr setzt jeder digitale Wandel eine entsprechende, auf Geräten basierende Technik in Form einer stabilen Infrastruktur voraus und enthält, wie schon beim Vertrauen in Technik, eine übertragbare menschliche Komponente durch das Entwicklerteam jeweiliger technischer Komponenten.

Die Eigenheiten des digitalen Wandels liegen insbesondere in der Veröffentlichung von Standards, bei deren zyklischer Entwicklung die 5 Phasen des Paradigmenwechsels nach Kuhn ebenfalls erkennbar sind. So gibt es zunächst kein allgemein anerkanntes und international kompatibles Format, bis nach intensiver Diskussion und Entwicklungszeit ein solches analog zu einem Paradigma eingeführt wird. Dieses bleibt so lange bestehen, bis die äußeren Umstände eine Änderung und Anpassung erzwingen. Im Kapitel 3 sollen zunächst relevante Datentypen und danach Dateiformate vorgestellt werden, um diesen Gedanken am Beispiel von Bibliotheksformaten weiter zu verfolgen.

Wie eingangs erwähnt, hat der digitale Wandel einen großen Einfluss auf die Entwicklung der Moderne, die Lebenswelt des Menschen und damit auch auf das Wirtschaftssystem. Da sich dieser Einfluss quer über so gut wie alle Teile sowohl privater als auch öffentlicher Lebenswelt erstreckt und damit eine genaue Analyse unmöglich im Rahmen einer Seminararbeit erfolgen kann, sollen in dieser Arbeit nur die für Bibliotheken, hier als einen Teil des öffentlichen Raumes, relevanten Aspekte als Rahmenbedingungen des konkreten Wandels betrachtet werden.

Im aktuellen Gespräch über den digitalen Wandel ist einer der am häufigsten genannten Schlüsselbegriffe die so genannte Open Culture Bewegung, welche auch im Seminar eine zentrale Rolle spielte und deren einzelne Aspekte näher beleuchtet wurden. Open Culture ist ein Sammelbegriff für verschiedene Ideen zur Gestaltung der Rahmenbedingungen für eine offene, freie und gemeinschaftliche Informationsgesellschaft. Beispielfhaft sollen nachfolgend die für die Bibliothekswelt relevanten Konzepte zusammengefasst werden.

Open Access. Als Open Access wird der freie Zugang zu Medien, insbesondere Literatur und wissenschaftliche Publikationen bezeichnet. Dieser bezieht sich ursprünglich auf Medien, welche über das Internet erreichbar sind und beinhaltet unter anderem das Recht, diese Materialien zu lesen, lokal zu speichern und ohne Entgelt zu nutzen. Geboren in den 1990er Jahren aus der Notwendigkeit, mit geringeren Bibliotheksetats und steigenden Publikationskosten umzugehen, argumentiert diese international vertretene Bewegung, dass öffentlich geförderte

Forschung auch ihre Ergebnisse der Öffentlichkeit frei zur Verfügung stellen sollte, um eine systematische doppelte Finanzierung über Verlage zu verhindern. Um jedoch weiterhin der Lizenzierung von Inhalten gerecht zu werden, ist einer der Hauptpunkte das Finden einer rechtlichen Form, in der diese Inhalte publiziert werden können.

Einer der wichtigsten Meilensteine in der Diskussion über Open Access ist die „Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen“ [16], welche 2003 verfasst und bisher von rund 500 Institutionen unterzeichnet wurde. Neben der Forderung nach freier Zugänglichkeit wissenschaftlicher Zeitschriften im Internet beinhaltet diese Erklärung auch die Einbeziehung des Kulturguts aus Archiven, Bibliotheken und Museen. Das vorgeschlagene Lizenzmodell sieht unter anderem die freie Verbreitung, Verwendung und Bearbeitung eines Textes vor und hat in Form der Creative-Commons-Lizenz eine mögliche juristische Umsetzung der Berliner Erklärung gefunden.

Die Bibliothek als reiner, meist staatlich geförderter Informationsdienstleister hat von jeher die Aufgabe, der Öffentlichkeit abgesehen von marginalen Beiträgen zur Verwaltung freien Zugang zu akademischen Wissen zu gewährleisten. Folglich fügen sich die Entwicklungen des Open Access gut in die Konzepte der Bibliotheken ein und die Herausforderungen liegen neben der juristischen Klärung der Lizenzfragen vor allem im Aufbau einer technischen und digitalen Infrastruktur für die Publikation im Web sowie in der Entwicklung eines Webtauglichen Standards, in welchem die in der Bibliothek vorhandenen Daten sowie insbesondere deren Metadaten angeboten werden können. Die Arbeit daran soll in Kapitel 3 vorgestellt werden.

Open Source. Open Source bezeichnet einen Typ Software, deren Quelltext frei verfügbar ist bzw. offen angeboten wird. Dazu muss das Programm einer der anerkannten Open-Source-Softwarelizenzen [17] unterliegen. Diese legen grob zusammengefasst fest, dass das Programm in einer von Menschen lesbaren Hochsprache veröffentlicht ist, beliebig genutzt und verbreitet werden darf und dass die Software sowohl verändert als auch unter meist gleicher Lizenz erneut veröffentlicht werden darf. Diese Lizenzen lassen sich neben Quelltexten für Software auch auf Beschreibungen von Datenmodellen anwenden und sind Grundlage für eine freie Veröffentlichung und Diskussion über eben jene Datenmodelle und der Etablierung eines allgemein gültigen Standards für die Publikation der Daten. Folglich ist das Konzept des Open Source unumgänglich für die beschriebenen Bemühungen an der Universitätsbibliothek.

3 Digitale Infrastruktur von Bibliotheken

In diesem Kapitel soll ein kurzer Einblick in die gängigen Dateiformate im Bibliotheksumfeld gegeben werden. Zusätzlich dazu soll der Begriff des Semantic Web anhand des viel beschriebenen Technologiestacks und dessen Zusammenhang zu Linked Data und RDF umrissen werden.

3.1 Datentypen in der Bibliothek

Zu den immer wieder auftretenden Datentypen in einer Bibliothek zählen die Titeldaten, Exemplardaten, Stammdaten und Normdaten. Während die Titeldaten alle Informationen betreffend der abstrakten Entität eines Buches zusammenfassen (wie z.B. Titel, Autor, Ver-

lag, Erscheinungsjahr, ISBN), umfassen die Exemplardaten alle Informationen zu einem konkreten, in der Bibliothek vorhandenen Exemplar eines Buches (z.B. Barcode, Regalnummer, Ausleihstatus). In die Kategorie der Normdaten fallen alle sonstigen Daten zu einem Titel wie z.B. abweichende Schreibweisen des Titels oder Autors und sind zentral aus der Datenbank der deutschen Nationalbibliothek [1] abrufbar. Zuletzt sind mit den Stammdaten an der Universitätsbibliothek Leipzig noch die einzelnen Standorte und Öffnungszeiten erfasst.

3.2 Dateiformate im Bibliotheksumfeld

Historisch gesehen können alle digitalen Dateiformate im Bibliotheksbereich auf analoge Kataloge bzw. Karteikarten zurückgeführt werden und lassen den Wunsch erkennen, in übersichtlicher Form Metainformationen zum betreffenden Buch oder allgemein zum betreffenden Bibliotheksmedium zusammenzufassen. Logischerweise sind Karteikarten oder deren digitale Repräsentation durch einfaches Abtippen denkbar ungeeignet für den bibliotheksübergreifenden Informationsaustausch. Dieser gewünschte Informationsaustausch setzt einen allgemeinen Standard voraus, um Kompatibilität und Vergleichbarkeit der Daten zu gewährleisten. Von den zahlreichen bibliographischen Dateiformaten sollen im kommenden Kapitel zwei vorgestellt werden.

3.2.1 MAB

MAB als Kurzform für „Maschinelles Austauschformat für Bibliotheken“ ist ein Dateiformat für den Informationsaustausch zwischen Bibliotheken, welches überwiegend im deutschsprachigen Raum [9] von Bedeutung ist. Mit ihm lassen sich alle oben beschriebenen Datentypen einpflegen und austauschen.

Es handelt sich um ein feldorientiertes Format, d.h. zum Zweck der Maschinenlesbarkeit sind die einzelnen Elemente eines Datensatzes durch spezielle, im Standard definierten Zeichen getrennt. Sie sind mit Wegweisern vergleichbar und ermöglichen dem Computer die Navigation und Suche im Datensatz. Solche abgetrennten Elemente eines Datensatzes werden Felder genannt; jedes Feld hat seine spezielle Funktion. Es müssen dabei nicht alle Felder belegt sein.

Eine tabellarische Übersicht aller Felder und deren Funktion kann unter [2] nachgeschlagen werden.

3.2.2 MARC

Das internationale Pendant zu MAB ist das ebenfalls feldorientierte Datenformat MARC. Dies steht für „Machine-Readable Cataloging“ und baut auf dem ISO-Standard 2709 [10] auf. Es ähnelt damit strukturell MAB aus dem vorherigen Abschnitt.

Die Elemente des Datensatzes sind ebenfalls in Feldern strukturiert, jedes Feld beginnt mit einem Tag, welches stets aus einer dreistelligen Ziffer, gefolgt von bis zu zwei weiteren Ziffern oder Buchstaben als Indikatoren für Unterfelder, besteht. Eine vollständige Auflistung aller verwendeten Tags sowie deren Bedeutung ist auf den Webseiten der Library of Congress [14] zu finden.

Diese Übersicht mag auf den ersten Blick zu umfangreich für den praktischen Gebrauch erscheinen, doch finden nur ca. 10% aller Tags regelmäßige Verwendung. Die mit Abstand

10	Library of Congress Control Number (LCCN)
20	(ISBN)
100	Name des Autors
245	Informationen zum Titel
260	zusätzliche Informationen zur Publikation
300	Beschreibung des Buches
650	Fachgebiet

Tabelle 1: Meist verwendete MARC 21 Tags

am häufigsten verwendeten und damit für einen Umstieg in ein anderes Format relevantesten Tags sind in Tabelle 1 aufgeführt.

Diese Tags bestehen aus Elementen der natürlichen Zahlen von 1 bis 899 und jeder Hunderterbereich ist für bestimmte Teilmengen an bibliothekarischen Daten reserviert, wie der Tabelle 3.2 auf der Webseite [14] entnommen werden kann.

In der Informatik kennt man eine solche Datenstruktur als geordnete, indizierte Liste von Key-Value Paaren, auch Array genannt. Jeder Datensatz der Titeldaten umfasst eine solche Liste von geordneten Paaren und wird durch eine in der gesamten Datenbasis eindeutige Nummer identifiziert, hier die ISBN, welche innerhalb der Liste daher eine Sonderstellung einnimmt. Jedes Listenelement kann dabei selbst eine Liste von vordefinierten Unterfeldern sein. Eine in dieser Form geschachtelte Liste nennt man auch Trie.

Die Darstellungsform der Datensätze als Elemente und Unterelemente ist der grundlegende Unterschied zwischen MAB und MARC. So würden in MARC zusätzliche Informationen zu einem Datenfeld wie z.B. weitere Titel eines Werkes in einem Subfeld zusammengefasst werden, während dies bei MAB stets in einem separat definierten Datenfeld stehen müsste. Kurz gesagt hat der MAB-Trie eine geringere Tiefe als der MARC-Trie.

3.3 Linked Data und der Semantic Web Stack

Das Semantic Web ist als technologische Weiterführung des Web-Konzeptes zu verstehen. Es sollen dabei wie beim Web weltweit gültige Standards entwickelt werden, mit denen die Kommunikation und der Informationsaustausch zwischen Rechnern stellvertretend für Personen ermöglicht wird. Die Neuerung besteht darin, dass im Gegensatz zum Web das Konzept einen sehr hohen Strukturierungsgrad vorsieht.

Das bedeutet, dass alle Verknüpfungen zwischen den Informationen, welche sich sonst nur von Menschen durch den Kontext eines Begriffes im Text erschließen lassen, explizit mit in die Datenstruktur eingegeben werden und somit die maschinelle Informationsverarbeitung auf der entstandenen Wissensbasis auf eine neue Ebene gehoben werden kann. Daher der Name Semantic Web; es soll ein Netzwerk aus maschinenlesbaren, semantisch angereicherten Informationen aufgebaut werden.

Dazu wurden eine ganze Reihe von Standards eingeführt, welche in übersichtlicher Form in Abbildung 1 dargestellt sind.

Die Abbildung ist von unten nach oben mit ansteigender Komplexität aufgebaut. Ganz unten stehen mit URI die grundlegende Technologien zur Definition von Identifikatoren für

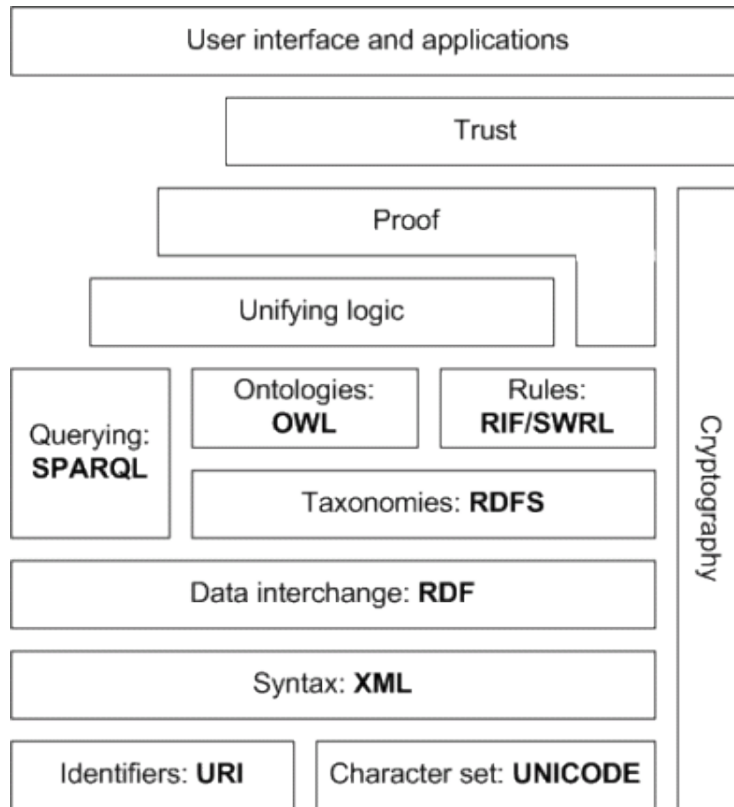


Abbildung 1: Der Semantic Web Technologie Stack (Quelle: [23])

einzelne Datenelemente wie z.B. Entitäten in der Wissensbasis und damit der textuellen Repräsentation real-weltlicher Bezüge. Diese sind eingebettet in ein XML-Dokument, welches wiederum der Form des Datenaustauschformates RDF genügen muss.¹ Taxonomien und allgemein gültige Regeln für das RDF-Dokument sind in RDFS geschrieben, für die Darstellung der Ontologie der Wissensbasis wird OWL verwendet. Die eigens für die Arbeit mit einer derartigen Wissensbasis entwickelte Abfragesprache heißt SPARQL.

Die Veröffentlichung einer nach diesen Standards aufgebauten Wissensbasis erfolgt in Form von Linked Data. Nach Tim Berners Lee [13] werden dabei vier Grundsätze verfolgt, welche im Kapitel 3.3.1 kurz vorgestellt werden sollen.

3.3.1 Linked Data Prinzipien

1. URIs als Bezeichner für Ressourcen

URI steht hierbei für Uniform Resource Identifier und ist ein Webstandard² für die Bezeichnung von Ressourcen. Jede URI folgt dabei einer speziellen Syntax.

`<scheme>:<hierarchical part>[?<query>][#<fragment>]`

¹Die aktuelle Diskussion über den Semantic Web Stack geht in die Richtung, XML als Datenformat gegen z.B. JSON oder Turtle auszutauschen, um die Lesbarkeit zu erhöhen und bei der Arbeit an der Wissensbasis auf unnötige Komplexität zu verzichten.

²URI Standard: <http://tools.ietf.org/html/std66>

Hierbei steht der Name des Schemas als Kürzel für das verwendete Protokoll, z.B. http, ftp, mailto, https. Der Hierarchieteil beschreibt die Lage der jeweiligen Ressource bzw. den Pfad, unter dem sie zu finden ist. Hier wird auch, wenn nötig, der Port mit „:port“ angefügt. Gefolgt wird die hiermit eigentlich abgeschlossene URI von eventuell übergebenen Parametern in der Form

...?variable1=[x]&variable2=[y]

sowie dem Verweis auf einen bestimmten Abschnitt der Ressource mit „#Abschnitt“.

Nach Definition, dass jede URI einzigartig ist, können auf diese Weise einzelne Entitäten unterschieden und eindeutig referenziert werden. In Bezug auf die im letzten Kapitel beschriebene digitale Infrastruktur würde dies bedeuten, das Wurzelement jedes Titeldaten-Tries mit einer URI zu versehen. Die schematische Definition dieser URIs, als Repräsentation der Ressourcen und damit realweltlicher Bezüge, ist dabei ein Teil der Modellierung einer Datenbasis.

2. Verwendung von HTTP als Schema

Die Wiederverwendung dieser etablierten Technologie hat zum Ziel, dass neben der eindeutigen Identifizierung einer Ressource eine URI, vorausgesetzt entsprechende Informationen liegen vor, auch im Browser eingegeben und eine dazugehörige Webseite angezeigt werden kann. Diesen Vorgang nennt man Dereferenzierung einer URI. Dabei wird eine lokale Kopie der Ressource erzeugt. Der Webserver schickt dafür eine HTML-Seite bzw. ein XML-Dokument, welches die angefragte Ressource bzw. ein Ressourcenfragment beinhaltet oder beschreibt, an den Client. Dies wird vor allem dann genutzt, wenn man über eine spezielle Ressource mehr Informationen erhalten möchte.

Um den Unterschied in der Anfrage zwischen der Ressource selbst und der Webseite, welche Informationen über eine Ressource enthält, zu modellieren, wird wie oben beschrieben, meist # verwendet³. So ist die Ressource selbst ein Fragment ihrer entsprechenden Seite, welche durch die URI referenziert wird und die Beschreibung der Ressource enthält. Sinnvoll für den Bibliothekskontext wäre hierbei eine Übersicht des Datensatzes (Trie-Struktur) mit Links zu assoziierten Ressourcen wie z.B. dem Autor als Personendatensatz.

3. Standards bei der Dereferenzierung von HTTP-URIs

Diese Standards sind z.B. die Verwendung von RDF⁴. RDF bezeichnet ein Framework zur Darstellung von Informationen. Alle in der Wissensbasis enthaltenen Fakten werden dabei als Tripel, also dreiteilige Informationsblöcke formuliert und ergeben auf diese Weise eine typische Baum- bzw. Graphstruktur mit Blättern als Ressourcen und Kanten als Verbindungen zwischen diesen. Die Tripel folgen dem Schema

[Subjekt] — [Praedikat] — [Objekt].

³Diese Verwendung des Symbols # in der URI stellt einen Spezialfall der im HTML-Standard beschriebenen Anwendungsmöglichkeiten dar. Allgemein wird auf diese Weise ein eindeutiger Zeiger auf ein spezielles Element mit einem id-Attribut versehenen Tag des HTML-DOMS spezifiziert.

⁴RDF – eine genauere und wohl die ausführlichste Beschreibung findet sich im umfangreichen Webangebot des W3C: <http://www.w3.org/standards/techs/rdf>

Subjekt und Prädikat sind dabei stets Ressourcen, im Falle des Subjektes eine dereferenzierbare HTTP-URI. Das Objekt ist entweder auch eine Ressource oder ein Literal, wie z.B. ein Text als String, eine Zahl o.Ä. Es besteht dabei auch die Möglichkeit, ungetypte Literale zu verwenden und mit „`@language`“ die Sprache anzugeben. Die Eignung dieses Modells für den Bibliothekskontext sollen in den kommenden Kapiteln diskutiert werden.

4. Darstellung von Zusammenhängen zwischen Ressourcen

Hier kommt es darauf an, die eigene Wissensbasis mit schon vorhandenen anderen Wissensbasen zu verlinken und so den Zusammenhang im Globalen Linked Open Data⁵ Graphen ersichtlich zu machen. Im Falle der Bibliothek wäre dies die Vernetzung mit Personen über die Normdaten mithilfe der owl:sameAs⁶ Beziehung.

3.3.2 Vertrauen im Semantic Web Kontext

In der Abbildung 1 zum Technologiestack des Semantic Web steht der Begriff „Trust“ als letzter Schritt auf dem Weg von der Basistechnologie zur fertigen Anwendung. Zwar wäre dieser mit „Vertrauen“ zu übersetzen, doch im Technologiestack steht „Trust“ für die Schlüssigkeit neuer, mit Algorithmen aus der Wissensbasis abgeleiteter Informationen und damit für eine noch in der Entwicklung befindliche Technik.

In diesem Kapitel soll statt dessen der Vertrauensbegriff aus Kapitel 2.1 als Relation im Kontext des Semantic Web betrachtet werden. Gemeint ist damit das Vertrauen, welches bei der Modellierung einer Wissensbasis durch Verlinkung zu einer anderen Wissensbasis ausgedrückt wird. Diese Verknüpfung ist, wie unter Punkt 4 in Kapitel 3.3.1 genannt, ausdrücklich erwünscht und führt zu einer höheren Qualität und Verwendbarkeit des Semantic Web in seiner Gesamtheit. Die Voraussetzung für einen Mehrwert dieser Verlinkung ist natürlich das Bestreben nach einer in sich fehler- und widerspruchsfreien Ontologie, doch die naive Forderung, alle Informationen im Semantic Web frei von Inkonsistenzen und unter Einhaltung höchster Qualitätsansprüche einzutragen, ist aufgrund der schiereren Menge an Informationen und Nutzern grundsätzlich nicht realistisch. Von daher sollte meiner Meinung nach das Ziel darin bestehen, die Inkonsistenzen auf ein dem praktischen Gebrauch angemessenes Maß zu reduzieren, welches dem Maß der existierenden Widersprüchen in der realen Welt nahekommt.

Durch die Verlinkung zu einer anderen Wissensbasis drückt der Modellierer Vertrauen in Form einer Erwartungshaltung an die andere Wissensbasis bzw. deren Modellierer aus. Er erwartet, dass die andere Wissensbasis von hoher Qualität, hohem Wahrheitsgehalt und konform mit den Inhalten der eigenen Wissensbasis ist. Nur bei der Erfüllung dieser Erwartungshaltung besteht ein Mehrwert in der Verlinkung zwischen den Wissensbasen und es müssen entsprechende Konsequenzen aus einem Vertrauensbruch gezogen werden.

Damit ist der Vertrauensbegriff im Semantic Web Kontext vergleichbar mit dem Vertrauensbegriff nach Luhmann, Kapitel 2.1. Wie im Paper [22] von Richardson argumentiert, besteht

⁵Linked Open Data: Der Unterschied zwischen *Linked Data* und *Linked Open Data* ist das Hauptaugenmerk auf der freien Nutzbarkeit der Informationen bzw. nach außen nicht offener Informationen bei Nutzung der gleichen Technologie. Die Ideen des Open Access sind hier wiederzufinden, vergleiche Kapitel 2.4.

⁶OWL – Web Ontology Language ist eine Sprache zur Definition von allgemeinen Regeln für die Wissensbasis sameAs stellt eine Verlinkung zu einer gleichwertigen Ressource, mitunter auch aus einer anderen Wissensbasis her und ist in besonderem Maße nützlich für die übergreifende Recherche.

die Option, dass sich der Umgang mit dem Vertrauen am Umgang in Bezug auf „normale“ Webseiten orientiert. Dort kann vom Anwender nicht erwartet werden, die Zuverlässigkeit jeder einzelnen Seite einzuschätzen und aufgrund der trotzdem subjektiven Natur von Vertrauen ist eine zentrale Vergabe eines „Vertrauenslevels“ für Webseiten nicht sinnvoll. Ebenso erscheint dies auch für Wissensbasen im Semantic Web weder praktisch noch sinnvoll.

Im Web wird das Vertrauen in die Qualität einer Webseite erfolgreich durch den Mechanismus des Pagerank⁷ gehandhabt. Analog dazu könnte zwar eine Wissensbasis als umso vertrauenswürdiger angesehen werden, je stärker sie mit anderen Wissensbasen vernetzt ist, doch dies würde der subjektiven Natur des Vertrauens und damit der grundlegenden Definition nach Luhmann nicht mehr gerecht werden. Zusätzlich dazu können im Gegensatz zum Web die Links im Semantic Web verschiedenste Semantiken ausdrücken, weshalb ein naives Zählen gewichteter Verbindungen ohne Berücksichtigung dieser Semantik keinen praktischen Mehrwert erbringt.

So definieren die Autoren des Papers Vertrauen als eine zwischenmenschliche Größe zwischen Nutzern des Netzwerkes. Demnach kann ein Nutzer eine kleine Anzahl an für ihn vertrauenswürdigen Nutzern benennen und der Algorithmus berechnet invers ein „Web Of Trusts“ zwischen allen Nutzern des Netzes. Wichtig dabei ist, dass dies immer individuell aus der Perspektive eines Beteiligten geschieht und so das Verhalten aus der realen Welt besser simuliert wird. So gibt es, anders als beim Pagerank, keinen globalen Vertrauenswert für einen Nutzer stellvertretend für die von ihm gepflegte Wissensbasis. Mit diesem einfachen Algorithmus erzielen die Autoren bereits einen guten Wert im praktischen Test mit Nutzern, auch weil die grundlegende Idee des „Web Of Trusts“ nicht neu, sondern bereits aus dem Gebiet der Kryptographie bekannt ist.

Für die Umsetzung dieser Idee im Bibliotheksumfeld wäre notwendig, in gemeinsamer und öffentlicher Diskussion an einer Beschreibungsform zu arbeiten und einzelne Wissensbasen nach dem obigen Modell zu verknüpfen, um ein „Web of Trusts“ aufzubauen. Dabei sollten nach dem Vertrauensmodell des Semantic Web weitere Wissensbasen aus anderen Wissenschaftsbereichen und öffentlichen Wissensbasen verknüpft werden. Diese Bemühungen und Argumentationen zielen zwar auf den Aufbau einer weltweiten, allgemein akzeptierten Ontologie ab, doch wie schon eingangs erwähnt ist diese Gestaltung aufgrund der Widersprüchlichkeit der realen Welt und der schiereren Menge der Daten und beteiligten Personen zumindest in naher Zukunft schwer vorstellbar. Die erfolgten Entwicklungen im Bibliotheksstandard sollen im kommenden Kapitel beleuchtet werden.

4 Entwicklungen im Bibliotheksstandard

4.1 Von MARC zu MARC 21

MARC entstand in den 1960er Jahren in den USA und wurde 1967 veröffentlicht. Federführend für die Entwicklung war die US-Amerikanische Library of Congress mit dem Ziel, ein maschinenlesbares Format zur Strukturierung von Bibliotheksdaten, auch genannt Records, einzuführen. 1971 wurde MARC in den USA zum landesweiten Standard erhoben und setzte

⁷Je häufiger eine Webseite auf anderen Seiten verlinkt ist, desto höher wird deren Qualität und Informationsgehalt eingeschätzt und desto höher werden von dieser Seite ausgehende Links gewichtet. Zusätzlich optimiert wurde dieser Algorithmus mit bestimmten, besonders vertrauenswürdigen Seiten, um dem Problem der Linkfarmen und gezielter Manipulation entgegenzuwirken. [18]

sich bis 1971 auch international durch. Es gab bei MARC als durchgehend weiterentwickelten Standard logischerweise viele Versionen und Erweiterungen, z.B. für die Erfassung von Bestands- und Lokaldaten, denn ursprünglich sollten nur Titeldaten erfasst werden. So entstanden im Laufe der Zeit viele nationale Ableger des Formates, unter anderem UKMARK (GB), CAN/MARC (Kanada) und UNIMARC (europäischer Raum). 1999 wurden das Kanadische und das US-Amerikanische Format aneinander angeglichen und es entstand ein neuer Standard unter dem Namen MARC 21. Ein Jahr später folgte darauf mit MARCXML noch eine auf XML basierte Version als alternative Notationsform des MARC 21 mit den Vorteilen von XML wie z.B. einer höheren Flexibilität, besserer Menschenlesbarkeit und der Möglichkeit, Validationstools zu verwenden.

Meiner Meinung nach sind, abgesehen von ein paar Besonderheiten, an dieser Entwicklung alle Züge des Paradigmenwechsels nach Kuhn abzulesen. Zunächst gibt es kein Datenmodell für den Austausch digitaler Titeldaten zwischen Bibliotheken, dies entspricht der Prä-Paradigmenphase. Zwar ist die Phase der „normal Science“ im Nachhinein nicht deutlich erkennbar, aber die erfolgte Diskussion und verschiedenen Einzelformate sind vergleichbar mit dem Prozess der untereinander abweichenden Hypothesen aus Phase 3 nach Kuhn. Der internationale Durchbruch und die einheitliche Etablierung als Standard 1999 gleicht stark der Durchsetzung eines Paradigmas in der Wissenschaftslandschaft. Auch hier geschieht dies durch Überzeugung der beteiligten Personen.

4.2 MAB und MAB 2

MAB wurde als Austauschformat für Bibliotheken im deutschsprachigen Raum 1973 eingeführt und nach einer umfangreichen Überarbeitung 1995 von MAB 2 abgelöst. Es existieren vergleichbar mit der Erweiterung von MARC für die Erfassung von zusätzlichen Datenbeständen neben den Titeldaten noch weitere Formate. Die Entwicklung und Pflege von MAB und allen damit assoziierten Formaten wurde 2006 in Vorbereitung auf den Umstieg zu MARC 21 geschlossen. Dieser Umstieg wurde 2013 mit dem offiziellen Ende der Auslieferung von MAB-Daten durch die DNB abgeschlossen. Die Gründe für den Umstieg sind umfassend in [21] beschrieben.

1. Gemeinsame Nutzung der Formate zum Informationsaustausch

Die Intention hinter der Entwicklung der Formate war der bibliotheksübergreifende Austausch von Titel- und Bestandsdaten. Dabei spielt auch die Übernahme von Daten aus internationalen Katalogen eine immer größere Rolle und würde logischerweise durch ein einheitliches internationales Format vorangetrieben werden. Mit MARC 21 existiert bereits ein global anerkanntes Format für diesen Zweck. Dabei muss die Kompatibilität der Formate untereinander untersucht werden, um sicherzustellen, dass beim Wechsel die aktuellen Bestandsdaten nicht gefährdet sind. Die Unterschiede und Möglichkeiten zur Konvertierung der Datensätze werden von Robert Porth in Kapitel 2.1 [21] genau dargestellt. Er kommt zu dem Ergebnis, dass auch für den kritischen Aspekt der mehrbändigen Werke diese zwar durch Generierung einer einstufigen Darstellung abbildbar sind, der dabei stattfindende Prozess hat jedoch mit hoher Wahrscheinlichkeit einen Verlust an Information zur Folge. Auf der anderen Seite steht der ausgleichende Mehrwert einer einfacheren Darstellung und einer besseren Analysierbarkeit des Gesamtdatensatzes.

2. Kostenfrage

Bei einem derartigen Umstieg auf eine neue Technologie ist die Kostenfrage von entscheidender Bedeutung. Bekannt aus der Softwaretechnik ist der Umstand, dass die Einführung eines neuen Systems sowohl initiale Umstiegskosten für Einrichtung des Systems und Schulung der Mitarbeiter zur Folge hat sowie eine nur allmählich anwachsende Produktivität zu verzeichnen ist. Erst nach einiger Zeit kann der tatsächliche Mehrwert eines technologischen Umstieges gemessen werden und muss daher vorher so genau und sorgfältig wie möglich geschätzt werden, um die Kosten und den Nutzen gegeneinander abzuwägen.

3. Höheres Potential

Porth schreibt dazu: „Vor dem Hintergrund, dass MARC 21 das am weitesten verbreitete – wenn auch natürlich nicht das einzige – internationale Format für bibliographische Daten ist, ist die Erhöhung des Anteils der Fremddatenübernahme bei fremdsprachigen Titelaufnahmen ein erklärtes Ziel des Formatumstiegs. In keinem anderen Format liegen mehr bibliographische Datensätze vor und aufgrund der Bedeutung von MARC 21 können i.d.R. auch Bibliotheken und Bibliotheksverbände außerhalb des MARC 21-Bereiches das Format als Austauschformat verwenden. Außer Bibliotheken und Verbänden im MARC 21-Raum verwendet beispielsweise auch die weltweit größte Datenbank für bibliographische Daten der WorldCat, MARC 21 in einer eigenen Interpretation (OCLC-MARC) als Austausch und Katalogisierungsformat. Welches Potential für eine einzelne Bibliothek in der Fremddatenübernahme von MARC 21 Daten liegt, hängt aber natürlich stark von deren Profil ab. Je internationaler das Erwerbungsprofil ausgerichtet ist, desto größer wird das Potential sein.“ [21, Seite 37]

Es ist offensichtlich, dass die Hauptgründe für den Umstieg von MAB 2 zu MARC 21 bei den Punkten 1 und 3 liegen und die Kostenfrage sich durch automatische Konvertierung mithilfe von Skripten und eine gewissen Ähnlichkeit beider Formate in einem akzeptablen Rahmen bewegte. Aktuell steht ein weiterer Wechsel in der Diskussion, nämlich der Wechsel vom bibliotheksinternen Austauschformat hin zu einem in vielen Bereichen der digitalen Welt verwendeten Framework, dem RDF aus dem Paradigma des Semantic Web und Linked Open Data. Die Gründe und Mehrwerte für diesen Umstieg sollen nach einer kurzen Vorstellung der gemeinsamen Normdatei im Kapitel 4.4 erläutert werden.

4.3 Die gemeinsame Normdatei

Ein weiterer wichtiger Schritt war der Aufbau einer zentralen und gemeinsam genutzten Normdatei für den unter 2.4 beschriebenen Datentyp der Normdaten. Diese existiert im Raum des MARC 21-Formates unter dem Namen LCNA (Library of Congress Name Authorities) bzw. LCSH (Library of Congress Subject Headings) und wird, wie der Name schon sagt, von der Library of Congress verwaltet. Im deutschsprachigen Raum gab es dafür zunächst drei Normdateien: die Personennamendatei, die gemeinsame Körperschaftsdatei und die Schlagwortnormdatei unter der Organisation der DNB. Diese und die Zeitschriftendatenbank wurden 2012 zusammengelegt zur Gemeinsamen Normdatei [25]. Sie ist frei online verfügbar sowohl im Format MARC 21 als auch im Linked Data Service als RDF-Repräsentation.

4.4 Der Weg zum Semantic Web

Die Diskussion über den Umstieg auf Semantic Web und Linked Open Data in der Bibliothekswelt ähnelt stark der Argumentation zum Thema Open Culture im Kapitel 2.4. Es geht um die Verbesserung der Arbeitsmöglichkeiten, Kreativität und die Erhöhung des Nutzens, den Technologie für den Anwender hat durch Offenheit dieser. Der Nutzer soll in den Prozess der Entwicklung von Technologie mit einbezogen werden, indem ihm die Rechte und Werkzeuge in die Hand gegeben werden, diese zu verändern und anzupassen. Nach der Philosophie der Vertreter der Open-Bewegung soll dadurch auch das Voranschreiten der technologischen Entwicklung verbessert werden.

Für die Bibliotheken würde dies bedeuten, die internen, historisch gewachsenen digitalen Datenbestände sowohl untereinander als auch der Öffentlichkeit, den Besuchern der Bibliothek, zugänglich zu machen und eine gemeinsame Weiterentwicklung dieser Beschreibungsformen anzustreben. Diese Zugänglichkeit der Daten hätte zur Folge, dass viele neue Anwendungen zur Auswertung und Generierung von Sichten auf diese Daten möglich wären und sich daher der Nutzen dieser hochqualitativen und gut gepflegten Daten stark erhöhen würde. Es wäre zudem eine Verbindung zu anderen frei verfügbaren Daten aus dem Netz des Linked Open Data⁸ möglich. Die Hauptanwendungen dieser Verknüpfung lägen im Bereich der vernetzten Recherche über viele Wissensbasen hinweg.

Neben der Vernetzung mit anderen Domänen ist auch die Nachnutzbarkeit der Daten ein wichtiger Aspekt. Auf der anderen Seite stellt das im Punkt 3.3.2 beschriebene Vertrauen im Semantic Web Kontext als Basis für die erfolgreiche und nachhaltige Verlinkung von Wissensbasen eine zusätzliche Hürde und einen Mehraufwand dar. Eine Hauptschwierigkeiten bei der Verlinkung ist die teilweise schlechte Kompatibilität der bestehenden Services. Dies liegt an unterschiedlicher Struktur der verwendeten Ontologien und ein wichtiges Ziel ist es daher, ein einheitliches und damit praktikables Modell zu entwerfen und etablieren.

Zur Diskussion dieser Fragen im deutschsprachigen Raum wurde das Kompetenzzentrum Interoperable Metadaten (KIM) [11] im Rahmen des gleichnamigen Projektes gegründet. Es beschäftigt in diversen, untereinander vernetzten Arbeitsgruppen mit nationalen und internationalen Metadatenstandards sowie der Integration von Metadaten im Semantic Web Kontext. Gefördert wird KIM von der Deutschen Forschungsgemeinschaft⁹ und arbeitet eng mit der Deutschen Nationalbibliothek und der Max Planck Digital Library¹⁰ zusammen. Das Ziel ist, neben der Verbesserung der Interoperabilität von Metadaten im deutschsprachigen Raum, der Aufbau einer Informationsplattform bzw. dem Aufbau eines Expertenteams zur Unterstützung und Beratung beim Aufbau von Wissensbasen für Metadaten oder der Entwicklung derartiger Beschreibungsformen im Semantic Web. Gleichzeitig soll der Austausch mit internationalen Organisationen in dem Bereich vorangebracht werden.

Wie bereits erwähnt, strukturiert sich die Arbeit von KIM in diverse Untergruppen, deren größte hier kurz vorgestellt werden sollen.

⁸Linked Open Data in der Lod-Cloud, eine Übersicht von 2014 über alle frei verfügbaren und online zugänglichen Wissensbasen vom Insight Centre for Data Analytics at NUI Galway mit Direktlink zur Beschreibung der jeweiligen Untercloud: <http://lod-cloud.net/>

⁹Deutsche Forschungsgemeinschaft: <http://dfg.de>

¹⁰Max Planck Digital Library: <http://mpdl.mpg.de>

- Linked Library Data Gruppe

Diese Gruppe hat es sich zur Aufgabe gemacht, zwischen internationalen Gruppen wie dem W3C oder der Open Knowledge Foundation zu vermitteln, im deutschsprachigen Raum Informationsmaterial zu den internationalen Linked Data Aktivitäten zu sammeln und durch Übersetzung der deutschsprachigen Community zur Verfügung zu stellen. Dazu gehört unter anderem die Übersetzung und Vermittlung von Dublin Core Dokumenten. Dublin Core bezeichnet eine von der Dublin Core Metadata Initiative herausgegebene Sammlung von Konventionen für die Beschreibung von Dokumenten und deren Metadaten. Es ist vergleichbar mit einem Framework für die Erstellung einer Ontologie im Semantic Web. Zu den Anwendungen gehören unter anderem die Serialisierung in RDF/XML, womit Dublin Core meiner Meinung nach zu einer der vielversprechendsten Meilensteine auf dem Weg zu einer Titeldaten-Ontologie im Semantic Web gehört.

Zu den weiteren Projekten der Linked Library Data Gruppe gehört z.B. auch die Mitarbeit an Europeana¹¹, einem Projekt, welches von der Europäischen Kommission gefördert wird, mit dem Ziel, der Öffentlichkeit in Form einer virtuellen Bibliothek das kulturelle Erbe Europas zugänglich zu machen. Das Projekt startete 2009 mit Version 1.0 und bietet neben einer REST-API unter europeana.ontotext.com/sparql einen erreichbaren SPARQL-Endpoint an.

- Lizenzen Gruppe

Die Untergruppe Lizenzen gab im Oktober 2011 die Empfehlungen zur Öffnung bibliothekarischer Daten [19] heraus. Darin enthalten ist die Aufforderung an Bibliotheken, speziell im deutschsprachigen Raum, alle nicht personenbezogenen oder anderweitig sensible Daten in der Bibliothek nach den Prinzipien des offenen Zugangs, offener Standards und offener Lizenzen im Internet verfügbar zu machen; die Parallelen zu den Forderungen der Open Access Bewegung sind offensichtlich. Das umschließt neben den Bestandsdaten und Titeldaten auch anonymisierte Ausleihdaten. Leider spezifiziert diese Empfehlung neben der Forderung nach Strukturiertheit und maschineller Lesbarkeit der Daten das Format nicht näher.

- Normdaten Gruppe

Gegründet 2011 arbeitet diese Gruppe an der Sammlung und Herausgabe von Standards für Normdaten verschiedener Domänen. Zur wichtigsten Sammlungen im deutschsprachigen Raum gehört die Gemeinsame Normdatei GND, welche wie im Kapitel 4.3 vorgestellt sowohl in MARC 21 als auch RDF/XML zur Verfügung steht.

- Titeldaten Gruppe

Diese Untergruppe arbeitet im Rahmen der Deutschen Initiative für Netzinformationen e.V.¹² sehr intensiv an den Beschreibungsformen zur Darstellung der Titeldaten und publiziert eine regelmäßig aktualisierte „Empfehlung für die RDF-Repräsentation bibliographischer Daten“ [4] auf den Seiten der DNB. Diese im Juli 2014 in Version 1.1 erschienene Empfehlung befindet sich momentan im Bearbeitungszustand, was eine genauere Analyse des aktuellen Arbeitsstandes erschwert. Die Publikation beschreibt

¹¹Europeana Wiki Seite: <http://europeana.eu>

¹²DINI: <http://dini.de>

ausgehend vom digitalen Wandel der deutschen Hochschullandschaft die Notwendigkeit der Förderung der Entwicklung der Informationsinfrastrukturen und Verbesserung des Dienstleistungsangebotes. Die Empfehlungen für die konkreten Beschreibungsformen orientieren sich dabei am Grad der Verbreitung, d.h. das bereits erwähnte Dublin Core als Framework, welches auch außerhalb der Bibliothekswelt relativ verbreitet ist, spielt dabei eine meiner Meinung nach positiv zu bewertende zentrale Rolle. Es werden ein Mapping der Datenmodellierung nach RDF ausgehend vom gemeinsam genutzten Format MARC 21 sowie eine möglichst umfassende Verlinkung zur Gemeinsamen Normdatei in Bezug auf die Personen und Körperschaften empfohlen. Weniger konkret ist die Empfehlung beim Unterpunkt der eindeutigen Identifikation realweltlicher Bezüge (URI-Schema) und der spezifischen Angabe der Medientypen eines Titeldatensatzes. Es werden hier nur Beispiellisten unter Verwendung verschiedenster Frameworks bzw. Namensräume angeboten und es wird auf zukünftige Arbeit verwiesen.

Schon aus dieser zusammengefassten Aufstellung der Untergruppen und deren Ergebnissen werden der hohe Arbeitsaufwand und die Bemühungen des KIM für eine Vereinheitlichung der Metadaten deutlich. Es wird auch ein hohes Vertrauen in die zukünftige Bedeutung der Technologien des Semantic Web ersichtlich. Leider beschränken sich die Herausgaben abgesehen von den Erfolgen im Bereich der Normdaten bisher nur auf vergleichsweise unverbindliche Empfehlungen. Diese enthalten zwar gut strukturierte und damit durchführbare Arbeitsvorschläge, nach denen wir das Linked Data Angebot an der Universitätsbibliothek Leipzig deutlich voranbringen konnten, doch ist dabei die Konkretisierung des Formats sowie der Beschreibungsformen unvollständig und es besteht ein meiner Meinung nach zu großer Freiraum beim Entwurf einer Ontologie für die Titeldaten. Auch ist der praktische Anwendungsbereich abgesehen von einer möglichen vernetzten Suche über mehrere Wissensbasen hinweg kaum erschlossen.

5 Auswertung

Dieser nicht oder noch nicht direkt erkennbare Mehrwert ist einer der Hauptkritikpunkte am Umstieg bzw. an Linked Data allgemein. Es ist denkbar und plausibel, dass mit der sehr hohen Strukturierung, gekoppelt mit der permanenten Verfügbarkeit der Daten und übergreifenden Suche über mehrere Wissensbasen eine Vielzahl nützlichen Anwendungen entstehen wird. So wäre eine auf der Abfragesprache SPARQL [24] aufbauende Suchmaschine in der Lage, die Wissensbasen übergreifend nach verknüpften Informationen zu durchlaufen und so dem Internetnutzer wesentlich präzisere Antworten zu liefern als herkömmliche Suchmaschinen. Es wäre dabei sogar möglich, nicht Dokumente als Antwort zu liefern, sondern korrekt artikulierte Sätze. Leider gibt es derartige Suchmaschinen beim aktuellen Stand der Technik nicht.

Die notwendigen Schritte für die Erstellung einer Anwendung unter Verwendung eines neuen technologischen Grundgerüsts sind in der Informatik wohlbekannt und wurden z.B. mit der wirtschaftlichen Erschließung des Internets unter Beweis gestellt. So zeigte die auch im Seminar diskutierte Entwicklung des 3D-Drucks ähnliche Eigenschaften: Mit der Einführung und Vorstellung des Gerätes, welcher vergleichbar mit der Einführung des Linked Data Standards ist, gab es nur sehr wenige Nutzer. Die Anwendung dieser Technologie fand in Nischen statt. Heute ist die Entwicklung voran geschritten und hat viele Interessenten und Mitwirkende gefunden, welche darauf aufbauend anwendungsbezogene Neuheiten entwickeln.

Demnach bin ich der Meinung, dass die Entwicklung im Bereich Linked Data und das momentan stattfindende Umsteigen der Bibliotheken sich mit einigen Besonderheiten in den im Abschnitt 2.3 erklärten Begriff des Paradigmenwechsels nach Kuhn einfügt. Standards unterliegen, anders als es mitunter bei wissenschaftlichen Grundparadigmen der Fall ist, einer kontinuierlichen Weiterentwicklung und eine Abwärtskompatibilität wird in den meisten Fällen gewährleistet. Auch bewegt sich die Diskussion stärker am Verbreitungsgrad eines Standards. Diese Besonderheiten beachtend wären wie in Kapitel 4.1 beschrieben die Jahre vor 1960 die Phase 1, in denen es keinen Standard für den digitalen Austausch von Daten gab, aber der Wunsch bestand, dieses Anwendungsfeld zu erschließen. In der Phase 2 und 3 wurden nach MARC mehrere Standards lokal bzw. national eingeführt und damit bereits erste Anwendungen geschaffen. Mit dem Wunsch nach stärkerer und internationaler Vernetzung wurden diese Standards nach und nach schrittweise vereinheitlicht und erneuert, bis im Jahre 2013 MARC 21 als internationales Format von den digitalen Schnittstellen so gut wie jeder Bibliothek akzeptiert wird. Die in Phase 3 beschriebene Krise bahnte sich schon lange vor diesem Schritt mit der abnehmenden Bedeutung der Bibliotheken für die akademische Recherche und dem Vormarsch des Internets an und erreicht momentan ihren Höhepunkt.

Selbstverständlich nutzen auch die Bibliotheken das Internet und haben Webangebote, mit denen auf ihre lokalen Kataloge zugegriffen werden kann. Doch diese Sichten sind eingeschränkt und stellen nur ein schwaches Abbild der zugrundeliegenden Daten dar. Zusätzlich dazu sind sich kaum bibliotheksübergreifend vernetzt oder mit anderen Wissensbasen verknüpft. Diese Krise forderte eine Reaktion durch das Öffnen der internen Daten und die Suche nach einem neuen Format, in dem diese angeboten werden können, denn sie ist nicht innerhalb des vorherrschenden Paradigmas lösbar. Dieses befindet sich, polemisch formuliert, mit MARC 21 noch auf dem Stand von 1999. Ein neues Paradigma in Form eines Standards ist möglich, mit den Linked Data Prinzipien wurde ein vielversprechender Weg gefunden und zahlreiche Gruppen arbeiten wie beschrieben flächendeckend an der Ausarbeitung und Verbreitung.

Auch der Charakter des Paradigmenwechsels nach Kuhn wird deutlich, denn der Umstieg ist nicht die logische Konsequenz der Krise, sondern wird allein durch Überzeugungsarbeit und die Kollision von persönlichen Ansichten zum Thema realisiert. Dadurch, dass die Phase 5 noch nicht eingetreten ist, also der Zustand der „Normal Science“, der im übertragenen Sinne Experimente bzw. hier die Erstellung von Anwendungen, ist mit selbigen noch nicht zu rechnen und das Argument der mangelnden „Nützlichkeit“ kann nicht als Kriterium für die Bewertung der Entwicklung dienen.

Statt dessen zielt die Entwicklung aktuell auf die Findung eines zukunftsweisenden Weges ab und kann momentan noch nicht abgesehen werden. Klar ist allerdings, dass im Zuge der technischen Entwicklung im Bereich der Hardware die Rechenleistung und Speicherfähigkeit von Computern sowohl für die Bibliotheken als auch für Endanwender quasi exponentiell wie bisher weitergehen wird und damit die aktuellen Probleme des zeiteffizienten Zugriffs auf große Wissensbasen bald keine Hürde mehr darstellen werden. Zusätzlich ist mit den bekannten Methoden der Softwareentwicklung sichergestellt, dass bei einem erfolgreichen Umstieg auf einen neuen Standard dieser auch genutzt und ausgebaut werden wird. Man kann behaupten, dass die Technik des Linked Data Webs zwar die konzeptuelle Prüfung überstanden hat, aber immer noch in den Kinderschuhen steckt und sich erst noch durchsetzen muss. Damit ist auch die Frage nach dem Einfluss, den dieser digitale Wandel auf das System der langen Wellen in der Wirtschaft hat oder haben wird, noch nicht eindeutig zu beantworten. Ich vermute, dass mit der schrittweisen Verbreitung von Linked Data und dem Anwachsen des

Semantic Web in den kommenden Jahren zunächst der Einfluss auf die akademische Welt, z.B. in den Bereichen der Publikationen, Forschung und später Lehre spürbar wird. Um den wirtschaftlichen Einfluss zu schätzen, könnte man in zukünftigen Arbeiten Linked Open Data mit anderen Projekten der Open-Initiative wie z.B. Wikipedia vergleichen.

Literatur

- [1] Deutsche Nationalbibliothek. Normdaten im Onlineangebot.
<http://www.dnb.de/gnd>.
- [2] Deutsche Nationalbibliothek. MAB-Mapping. [dnb.de/SharedDocs/Downloads/DE/DNB/standardisierung/mabTabelleDeutschEnglisch2006.pdf](http://www.dnb.de/SharedDocs/Downloads/DE/DNB/standardisierung/mabTabelleDeutschEnglisch2006.pdf).
- [3] Deutsche Nationalbibliothek. Empfehlung für MARC 21-RDF-mapping.
<http://wiki.dnb.de/display/DINIAGKIM/MARC+21-RDF-Mapping>. Zuletzt gesehen: 03/2015.
- [4] Deutsche Nationalbibliothek. Empfehlung für die RDF-Repräsentation bibliographischer Daten.
<http://wiki.dnb.de/pages/viewpage.action?pageId=68060017>. Zuletzt gesehen: 03/2015. Version 1.1 siehe <http://edoc.hu-berlin.de/series/dini-schriften/2014-14/PDF/14-de.pdf>.
- [5] Document Availability Information Ontology (DAIA).
<http://uri.gbv.de/ontology/daia>. Zuletzt gesehen: 03/2015.
- [6] Fabian M. Fürste. *Linked Open Library Data*. Wiesbaden, 2011.
- [7] Fabian Gandon. RDF/XML Source Declaration.
<http://www.w3.org/Submission/rdfsource/>. Zuletzt gesehen: 03/2015.
- [8] Hans Gert Gräbe. Lange Wellen und globale Krise. In: Sozialgeschichte online 11/2013. Siehe auch <http://hg-graebe.de/EigeneTexte/Roth-12.pdf>.
- [9] Reinhold Heuvelmann, Deutsche Nationalbibliothek. MAB.
http://www.dnb.de/DE/Standardisierung/Formate/MAB/mab_node.html.
Zuletzt gesehen: 03/2015.
- [10] ISO 2709: Datenaustauschformat zwischen Bibliotheken, unter anderem einst auf Magnetbändern. http://en.wikipedia.org/wiki/ISO_2709.
- [11] KIM-Forum – Kompetenzzentrum Interoperable Metadaten.
<http://kim-forum.org>.
- [12] Thomas S. Kuhn. *Die Struktur wissenschaftlicher Revolutionen*. Suhrkamp, Frankfurt/Main 1973.
- [13] Tim Berners Lee. Linked Data Design Issues.
<http://www.w3.org/DesignIssues/LinkedData.html>. Zuletzt gesehen: 03/2015.

- [14] Library of Congress. Understanding MARC Bibliography.
<http://www.loc.gov/marc/umb>. Zuletzt gesehen: 03/2015.
- [15] Niklas Luhmann. *Vertrauen. Ein Mechanismus der Reduktion sozialer Komplexität*. UTB, Stuttgart 1973. <http://www.derdieckmann.de/archives/46>.
- [16] Max-Planck-Gesellschaft. Berliner Erklärung, 22. Oktober 2003.
<http://openaccess.mpg.de/Berliner-Erklaerung>.
- [17] Open Source Initiative. Licences and Standards.
<http://opensource.org/licenses/>.
- [18] PageRank Algorithmus Patent. http://worldwide.espacenet.com/publicationDetails/biblio?locale=de_EP&CC=US&NR=6285999
- [19] Adrian Pohl. Empfehlungen zur Öffnung bibliothekarischer Daten.
<http://wiki.dnb.de/pages/viewpage.action?pageId=45419980>.
- [20] Adrian Pohl. *Linked Data und die Bibliothekswelt*. BIT Online, 3/2010.
<http://www.b-i-t-online.de>
- [21] Robert Porth. Betrachtung des Formatumstiegs von MAB 2 zu MARC 21 – potentielle Mehrwerte bei der Internationalisierung eines bibliographischen Formats. Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft, Heft 247. <http://edoc.hu-berlin.de/series/berliner-handreichungen/2009-247/PDF/247.pdf>. Zuletzt gesehen: 03/2015.
- [22] Matthew Richardson, Rakesh Agrawal, Pedro Domingos. Trust management for the semantic web.
<http://ml.cs.washington.edu/www/media/papers/richardson-a103a.pdf>.
- [23] Der Semantic Web Technologie Stack.
<http://upload.wikimedia.org/wikipedia/en/3/37/Semantic-web-stack.png>.
Zuletzt gesehen: 03/2015.
- [24] SPARQL update. <http://www.w3.org/Submission/SPARQL-Update/>. Zuletzt gesehen: 03/2015.
- [25] Brigitte Wiechmann, Deutsche Nationalbibliothek. Gemeinsame Normdatei.
http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html. Zuletzt gesehen: 03/2015.