

HUMBOLDT-UNIVERSITÄT ZU BERLIN
INSTITUT FÜR BIBLIOTHEKS- UND INFORMATIONSWISSENSCHAFT



BERLINER HANDREICHUNGEN
ZUR BIBLIOTHEKS- UND
INFORMATIONSWISSENSCHAFT

HEFT 338

BIBLIOGRAFISCHE DATEN, NORMDATEN UND METADATEN
IM SEMANTIC WEB

KONZEPTE DER BIBLIOGRAFISCHEN KONTROLLE
IM WANDEL

VON
KATHI WOITAS

BIBLIOGRAFISCHE DATEN, NORMDATEN UND METADATEN
IM SEMANTIC WEB

KONZEPTE DER BIBLIOGRAFISCHEN KONTROLLE
IM WANDEL

VON
KATHI WOITAS

Berliner Handreichungen zur
Bibliotheks- und Informationswissenschaft

Begründet von Peter Zahn
Herausgegeben von
Konrad Umlauf
Humboldt-Universität zu Berlin

Heft 338

Woitas, Kathi

Bibliografische Daten, Normdaten und Metadaten im Semantic Web :
Konzepte der Bibliografischen Kontrolle im Wandel / von Kathi Woitas. -
Berlin : Institut für Bibliotheks- und Informationswissenschaft der
Humboldt-Universität zu Berlin, 2013. - 141 S. : graph. Darst. - (Berliner
Handreichungen zur Bibliotheks- und Informationswissenschaft ; 338)

ISSN 14 38-76 62

Abstract:

Die Informationsräume Bibliothek und World Wide Web (WWW) stehen vor der gleichen Herausforderung: Eine korrekte, umfassende und nutzergerechte Erschließung ihrer Dokumente zu leisten. Eine Lösung aus beiden Perspektiven wäre eine einfache und universelle Ressourcenbeschreibung, die implizit generiert wird. Hieraus ergibt sich eine zentrale Frage für die Bibliothekswissenschaft: Inwieweit und auf welche Weise können solche Techniken auch für die Bibliografische Kontrolle nutzbar gemacht werden? Im ersten Teil der Arbeit werden die klassischen Konzepte der Bibliografischen Kontrolle analysiert. Als aktuelle Umformungen dieser Komponenten wird auf die Konzeptmodelle des "Bibliografischen Universums" FRBR, FRAD und FRISAD eingegangen. Das Begriffskontinuum Metadaten und Ressourcenbeschreibung wird erläutert und als Bibliografisches System redefiniert, bevor die gegenwärtigen Metadaten systemen abschließend kritisch betrachtet werden. Im zweiten Teil werden Prinzipien und Technologien beschrieben, die unter dem Begriff Semantic Web zusammengefasst werden und die Struktur des WWW durch maschinenverarbeitbare Informationen erweitern sollen. Ein Überblick über aktuelle Anwendungen in Form von Linked-Data-Projekten schließt den Teil ab. Im letzten Abschnitt wird erläutert, wie Bibliografische Metadaten systeme durch die vorgestellten Semantic-Web-Techniken reformuliert und integriert werden können. Dies wird danach anhand des DCAM und konkreter Darstellungen von bibliografischen Metadaten schemata, Normdaten und Kontrollierten Vokabularen mittels RDF(S)/SKOS veranschaulicht. Schließlich wird das neue Katalogisierungsregelwerk Resource Description and Access (RDA) vor diesem Hintergrund betrachtet. Die Arbeit endet mit der Schlussfolgerung, dass Semantic-Web-Techniken dazu geeignet erscheinen, die strukturelle und inhaltliche Komplexität von bibliografischen Metadaten einheitlich darzustellen sowie diese hochgradig anpassungsfähig und universell verarbeitbar zu gestalten. Im Gegenzug können qualitativ hochwertige bibliothekarische Daten als wichtige Referenzpunkte in der sich rasch entwickelnden Linked Data Cloud fungieren - und ihrerseits in ungeahntem Ausmaß verknüpft und somit angereichert werden.

Diese Veröffentlichung geht zurück auf eine Magisterarbeit (2010) im Direktstudiengang Bibliothekswissenschaft an der Humboldt-Universität zu Berlin.

Online-Version: <http://edoc.hu-berlin.de/series/berliner-handreichungen/2013-338>



Dieses Werk steht unter einer Creative Commons [Namensnennung-NichtKommerziell-KeineBearbeitung 3.0](https://creativecommons.org/licenses/by-nc-nd/3.0/de/) Deutschland-Lizenz.

Inhaltsverzeichnis

1	Einleitung	7
2	Entwicklung und Konzepte der Bibliografischen Kontrolle	13
2.1	Bibliografische Beschreibung	14
2.1.1	Functional Requirements for Bibliographic Records (FRBR)	17
2.2	Katalogisierung	24
2.2.1	Statement of International Cataloguing Principles (ICP)	26
2.3	Normenkontrolle	27
2.3.1	Functional Requirements for Authority Data (FRAD)	33
2.4	Kontrollierte Vokabulare (KOS)	38
2.4.1	Thesauri	41
2.4.2	Kontrollierte Vokabulare in der Normenkontrolle	44
2.4.3	Functional Requirements for Subject Authority Data (FRSAD)	46
2.5	Fazit: Bibliografische Kontrolle	50
3	Metadaten als Mittel der Bibliografischen Kontrolle	52
3.1	Bibliografische Kontrolle im elektronischen Kontext	52
3.2	Metadaten – Begriffsdiskussion	53
3.3	Metadaten als Bibliografisches System	56
3.4	XML-kodierte Metadaten	60
3.5	Dublin Core Metadata Element Set (DCMES)	63
3.6	Kritik: Bibliografische Metadatensysteme	66
4	Semantische Technologien im World Wide Web	70
4.1	Vom WWW zum Semantic Web	70
4.1.1	„Semantische“ Informationen	71
4.1.2	Ein universelles Beschreibungsmodell	74
4.1.3	Formale Wissensrepräsentation	75
4.2	Resource Description Framework (RDF)	78
4.3	Aktuelle Anwendung von RDF(S) – Linked Data	83
4.4	Wissensrepräsentation im Semantic Web	86
4.4.1	Ontologien	86
4.4.2	RDF Schema und OWL	89
4.4.3	Simple Knowledge Organization System (SKOS)	92

5	Bibliografische Kontrolle mit dem Semantic Web?	96
5.1	Metadaten und RDF(S) – das DCMI Abstract Model	96
5.2	Eine neue Grundstruktur der Bibliografische Kontrolle?	102
5.3	Bibliografische Metadatenschemata in RDF(S)	103
5.3.1	FRBR als bibliografischen Ontologie?	107
5.4	Datenwertstandards in RDF(S) und SKOS	110
5.4.1	Normdaten im Semantic Web	111
5.4.2	Kontrollierte Vokabulare (KOS) im Semantic Web	116
5.5	Ein neuer Erschließungsstandard: Resource Description and Access (RDA)	120
5.6	Fazit: Bibliografische Kontrolle mit dem Semantic Web?	125
6	Schlussbemerkung	130
7	Abbildungsverzeichnis	134
8	Literaturverzeichnis	135

1 Einleitung

*A semantic web: a more librarian web.*¹

Bibliografische Daten, Normdaten und Metadaten im Semantic Web – Konzepte der Bibliografischen Kontrolle im Wandel. Der Titel dieser Arbeit zielt in ein essentielles Feld der Bibliotheks- und Informationswissenschaft, die Bibliografische Kontrolle. Als zweites zentrales Konzept wird der in der Weiterentwicklung des World Wide Webs (WWW) bedeutsame Begriff des Semantic Webs genannt. Auf den ersten Blick handelt es sich hier um einen ungleichen Wettstreit. Auf der einen Seite die Bibliografische Kontrolle, welche die Methoden und Mittel zur Erschließung von bibliothekarischen Objekten umfasst und traditionell in Form von formal-inhaltlichen Surrogaten in Katalogen daherkommt. Auf der anderen Seite das Buzzword Semantic Web mit seinen hochtrabenden Konnotationen eines durch Selbstreferenzialität „bedeutungstragenden“, wenn nicht sogar „intelligenten“ Webs. Wie kamen also eine wissenschaftliche Bibliothekarin und ein Mitglied des World Wide Web Consortiums² 2007 dazu, gemeinsam einen Aufsatz zu publizieren und darin zu behaupten, das semantische Netz würde ein „bibliothekarischeres“ Netz sein?

Um sich dieser Frage zu nähern, soll zunächst kurz die historische Entwicklung der beiden Informationssphären Bibliothek und WWW gemeinsam betrachtet werden. Denn so oft – und völlig zurecht – die informationelle Revolution durch das Internet³ beschworen wird, so taucht auch immer wieder das Analogon einer weltweiten, virtuellen Bibliothek auf. Genauer gesagt, nahmen die theoretischen Überlegungen, die später zur Entwicklung des Internets führen sollten, ihren Ausgangspunkt (neben Kybernetik und entstehender Computertechnik) beim Konzept des Informationsspeichers Bibliothek. Hans Dieter Hellige schreibt über diese Visionen in den 1950er und 60er Jahren:

Die Bibliotheks-Metapher war somit in einer widersprüchlichen Weise konstitutiv für die Internet-Genese: Einerseits wurde die „Library of the Future“⁴ als ein die neuen Freiheiten der Digitaltechnik und Vernetzung ausschöpfender Gegenentwurf zur Bibliothek konzipiert, andererseits orientierten sich die Visionen eines weltumspannenden Dokumenten-Universums bis in die 90er Jahre hinein, ja teilweise bis heute an Architekturen, sozialen Rollen und Gebrauchsweisen der traditionellen Bibliotheksorganisation. [Hellige(2007):495]

1 [McCathieNevile/Méndez(2007):28]

2 Eva Méndez Rodríguez, siehe <http://www.bib.uc3m.es/~mendez/profesional/index.html> und Charles McCathieNevile, siehe <http://www.w3.org/People/Charles/brcv>

3 „Internet“ soll im Folgenden aus sprachlichen Gründen gleichbedeutend mit dem World Wide Web verwendet werden, obwohl dieses eigentlich nur eine besondere Anwendung des Internets darstellt.

4 „Library of the Future“ (1965) heißt die Studie von Joseph Carl Robnett Licklider, der die Konzepte des Time-Sharings und der Computervernetzung entscheidend mitentwickelt hat.

Wie kann man diese Orientierung begreifen? Wo sind Ähnlichkeiten zwischen klassischem Web und Bibliothek auszumachen? Und worin bestehen ihre fundamentalen Differenzen, die diese Informationsräume doch so unterschiedlich erscheinen lassen?

I.

Als zentrale Instanz gilt in beiden Sphären das Dokument oder Objekt, in Form des Buches, Aufsatzes, oder eben als Textdokument, dessen Darstellung mit einer Auszeichnungssprache festgelegt wird. Im ursprünglichen WWW waren Kenntnisse des HTML-Codes nötig und Websites relativ statisch. Der Begriff „Read-Only-Web“ zeigt deutlich, dass seine Nutzung für die breite Masse auf die schlichte Rezeption beschränkt war. Indem sie sich auf Suchmaschinen zur Öffnung der informationellen Black Box WWW verlassen müssen, setzt sich diese Ohnmächtigkeit der Nutzer auch noch gegenwärtig fort. Aus bibliothekarischer Sicht liegen ähnliche Machtverhältnisse gar nicht so weit zurück. Angewiesen auf die sachkundige Hilfe und den guten Willen einer Thekenkraft war der Zugang zu Informationen ebenfalls eher schicksalhaft als autonom. Ob der Wechsel zur eigenhändigen Recherche an alphabetischen und Sachkatalogen dann bessere Ergebnisse gebracht hat, sei dahin gestellt, komplexe Erschließungsregelwerke haben jedoch sicher nicht dazu beigetragen.

Bibliothekaren wurde schnell klar, dass das Konzept der klassischen Bibliografischen Kontrolle im Internet nicht funktionieren würde: Quantität, Internationalität, und auch Qualität von Websites waren nicht mit denen herkömmlicher Informationsobjekte vergleichbar. Die Internet-Community entwickelte derweil eigene Strategien zu dessen Erschließung. Was durch intellektuelle Katalogisierung in Bibliotheken geleistet wurde, wurde versucht durch automatische Volltext-Indexierung oder – eher kläglich – durch Web-Verzeichnisse zu ersetzen. Eine Erschließung, die bibliothekarischen oder dokumentarischen Ansprüchen nahe kommt, findet jedoch nur sporadisch statt und bleibt selbst dann oberflächlich. Die Erfassung von sogenannten Metadaten verzeichnet nur in Ansätzen formale, und schon mehr schlecht als recht inhaltliche Merkmale der Webdokumente.

Bibliotheken und das WWW sind mittlerweile längst nicht mehr in den beschriebenen nutzerfeindlichen und statischen Strukturen gefangen. Bibliotheken haben Bestände und Magazine für den Publikumsverkehr geöffnet, einfach bedienbare OPAC- und andere Service-Interfaces geschaffen, schließlich sogar Hauslieferung und Selbstverbuchung eingeführt. Anschaffungen werden zum großen Teil durch Kundenwünsche initiiert, und bezüglich Öffnungszeiten und Ausstattung versucht man ebenfalls auf diese einzugehen: Sonn- und Feiertagsöffnungen, Eltern-Kind-Bereiche, Multimedia-Arbeitsplätze sind dafür nur einige Beispiele. Das Internet hat unterdessen eine bedeutende Erweiterung im Sinne des „Read-Write-

Web“ erhalten. Durch neue Techniken können Internetnutzer Dokumente ungleich leichter und schneller erstellen, manipulieren und verbreiten; Code-Kenntnisse sind hierfür nicht mehr erforderlich. Webdokumente und ihre einzelnen Bestandteile können unkompliziert in neue Zusammenhänge gestellt werden, Objekte gemeinsam und interaktiv gestaltet werden. Die Informationsobjekte selbst werden dadurch zwangsläufig kurzlebiger und instabiler, sozusagen „liquider“. Abseits von Volltextsuche und Web-Portalen beginnen Internetnutzer eigene Konzepte von Erschließung und Aggregation zu entwickeln. Im World Wide Web vollzog sich also, vergleichbar zu Bibliotheken vorher, ein Wandel von einer dokumentenzentrierten, statischen, rezeptiven Informationsumgebung zu Einer, die den Nutzer ermächtigt und seine Bedürfnisse und Einschätzungen in den Mittelpunkt stellt.

Mit diesem Fortschritt sind aber längst nicht alle Probleme des Internets gelöst, insbesondere die Suche und der zielgerichtete Zugang zu Informationen bereitet den Nutzern weiter Schwierigkeiten. Eine genaue, nutzergerechte und vor allem in der Breite wirksame Erschließung ist ein immer noch ein Desiderat, das heutige Suchmaschinen nicht erfüllen können. Als weitgehend gescheitert kann man dagegen die Ansätze von intellektuell geführten Web-Katalogen und von Forderungen an Internetnutzer nennen, ihre Webdokumente eigenständig mit qualitativ hochwertigen Metadaten auszuzeichnen.

Doch auch in der bibliothekarischen Sphäre ist man sich ungelöster Probleme in der Erschließung gewahr, obwohl die Bibliografische Kontrolle über eine längere Entwicklungsgeschichte verfügt. In Bibliotheken brachten Computerkataloge und Metadaten zwar eine neue Qualität der Suchfunktionen und des Zugriffs, jedoch birgt die Vielzahl an inkompatiblen und schwer anpassbaren Formaten größer werdende Schwierigkeiten. Zudem bewirkt die Datenbanktechnologie, dass Bibliotheksdaten zumeist in einer Art Datensilo gefangen und nur über eine entsprechende Suchmaske zugänglich sind.

Die Probleme der beiden Informationsphären stellen zwei extreme Positionen dar. Während bibliothekarische Erschließung an seiner elaborierten Grundhaltung und komplizierter technischer Umsetzung krankt, besteht das Grundproblem im WWW darin, dass eine wirkliche Erschließung, abgesehen von einigen hehren Kämpfern für den Dublin Core, nicht stattfindet. Was aus beiden Perspektiven nötig wäre, ist ein goldener Mittelweg, eine einfache und universelle Metadatenmatrix. Die entscheidende Anforderung für einen stetig wachsenden und liquiden Informationsraum wäre aber eine einfache Umsetzbarkeit, am besten eine automatisierte Erstellung dieser Metadaten. Oder wie Elaine Svenonius es anschaulich fasst:

An important question today is whether the bibliographic universe can be organized both intelligently (that is, to meet the traditional bibliographic objectives) *and* automatically. [Svenonius(2001):26, Hervorhebung K.W.]

Wenn man von einer automatischen Indexierung spricht, sind damit in der Regel informationslinguistische Verfahren gemeint, deren Aufwand für den allgemeinen Einsatz im Internet allerdings ungeeignet erscheint. Die zentrale Fragestellung nach einer universellen, einfach zu realisierenden und qualitativ hochwertigen Erschließung von Inhalten im WWW bleibt also vorerst ungelöst. Tim Berners-Lee, James Hendler und Ora Lassila haben den Grund für dieses Problem in der derzeitigen Verfasstheit des World Wide Webs erkannt, indem sich das Netz zwar nun nutzerorientierter gibt, gegenüber einzelnen Informationen bzw. Daten jedoch weitestgehend „blind“ bleibt:

To date, the Web has developed most rapidly as a medium of documents for people rather than for data and information that can be processed automatically. [Berners-Lee/Hendler/Lassila(2001)]

II.

Die Informationssphären Bibliothek und World Wide Web verfügen nichtsdestotrotz über entscheidende strukturelle Differenzen. Wenn eben von der gemeinsamen „Dokumenten-zentriertheit“ Beider gesprochen wurde, blieb ein entscheidendes Kriterium außen vor. Es ist das, was den Entwurf Tim Berners-Lees im Kern ausmachte, die Verwandlung von rein linearem Text in einen „Über-Text“ und seinen weltweiten, unreglementierten Einsatz. Letztlich war es das Erfolgsgeheimnis von HTTP/HTML, eine jahrzehntelange Vision auf ihr Minimalmaß herunterzubrechen und mit möglichst einfachen Mitteln technisch umzusetzen. Mag sich auch Ted Nelson über den überragenden Erfolg dieser Ultra-Light-Version seines Xanadu-Projektes ärgern – und noch mehr über den geistigen Diebstahl des Begriffs Hypertext – wie wir wissen, hat auch er die grundlegende Idee einer bedeutungsvollen und technisch operablen Verknüpfung von Dokumenten nur übernommen. Zumeist wird Vannevar Bush mit seinem Artikel *As we may think* aus dem Jahr 1945 diese Leistung zugeschrieben. [Warnke(2008)] Wie wahrscheinlich schon Einige vor ihm, erkannte er in der assoziativen Funktionsweise des menschlichen Gehirns den Schlüssel zur Verarbeitung von Informationen und der Generierung von Wissen – im Unterschied zu den externalisierten medialen Darstellungen von Wissen, die durch Linearität gekennzeichnet waren:

The human mind does not work this way. It operates by association. [Bush(1945)]

Die Idee, die Bush daraus folgert, stellt dann die Geburtsstunde des Konzepts Hypertext dar:

Man cannot hope fully to duplicate this mental process artificially, but he certainly ought to be able to learn from it. [...] The first idea however, to be drawn from the analogy concerns selection. Selection by association, rather than indexing, may be mechanized. [Bush(1945)]

Und schließlich beschreibt er, wie ein Nutzer seines Memex das „associative indexing“, also das intellektuelle Verknüpfen von Dokumenten realisieren soll:

When the user is building a trail, he names it, inserts the name in his code book, and taps it out on his keyboard. Before him are the two items to be joined, projected onto adjacent viewing positions. At the bottom of each there are a number of blank code spaces, and a pointer is set to indicate one

of these on each item. The user taps a single key, and the items are permanently joined. In each code space appears the code word. Out of view, but also in the code space, is inserted a set of dots for photocell viewing; and on each item these dots by their positions designate the index number of the other item. [Bush(1945)]

Diese „trails“ erinnern unmittelbar an unsere heutigen WWW-Hyperlinks, sind im Vergleich zu diesen aber weitaus bemerkenswerter. Die Verknüpfungen, die sich Bush hier erträumt, sind nicht nur bidirektional, sondern bedeutungsvoll. Das zu wählende Codewort definiert diesen trail, legt also seine Bedeutung, um nicht zu sagen, seine Semantik, fest. Der Hypertext Berners-Lees, den wir heute kennen, wirkt daneben doch etwas enttäuschend.

Vor diesem Hintergrund kann man die avisierte Erweiterung des World Wide Webs, das unter dem Begriff Semantic Web firmiert, als eine Aufarbeitung des ursprünglichen Hypertext-Konzepts begreifen. 2001 vorgestellt von Tim Berners-Lee, James Hendler und Ora Lassila, wird mit diesem Konzept

Semantic Web [...] eine Erweiterung des bestehenden Internets [umschrieben], in der jedes Dokument (bzw. jede Informationseinheit) um Meta-Informationen angereichert ist, die Angaben zum Inhalt des Dokuments und zu seinem Kontext machen [...]. Die Meta-Informationen liegen als Metadaten in einem Format mit festgelegter Syntax und Semantik vor, sodass sie von rechnergestützten Informationssystemen automatisch verarbeitet werden können. [Strauch(2004):Semantisches Netz, S. 110]

Auch wenn diese Definition oberflächlich an eine Beschreibung des normalen Metadaten-Konzeptes erinnert, liegt in den Begriffen der allgemein definierten Syntax und Semantik und der automatischen Verarbeitbarkeit der Schlüssel zum Entwurf von semantischen Informationen. Semantic-Web-Technologien begründen insofern durchaus ein neues „Metadaten“-Paradigma, wie Karen Coyle andeutet:

The Semantic Web community began with a kind of metadata *tabula rasa* and a natural tendency to think about machine processing of data at a deep level. Its work began with a study of the basic nature of metadata, or at least the very nature of machine-actionable, networked, interoperable metadata. [Coyle(2010b):12, Hervorhebung Coyle]

Die grundlegende Idee der Semantic-Web-Vision besteht darin, die Selbstreferenzialität des Webs zu nutzen, um Metainformationen von Dokumenten, insbesondere aber von einzelnen Daten und Konzepten, automatisiert und definiert zu verknüpfen. Ein Datenmodell, das auf einfachster Aussagenstruktur basiert, bietet hierfür eine universelle Syntax und formale Wissensrepräsentationen leisten die eindeutige Identifizierung von Konzepten – definieren also ihre Semantik. Es geht im Kern darum, *bedeutungsvolle* Zusammenhänge im WWW herzustellen, die semantische Schwäche von konventionellen Hyperlinks auszugleichen und Links zwischen einzelnen Informationseinheiten schaffen, die eine spezifische Bedeutung tragen. Die Nähe zum trail-Konzept Vannevar Bushs ist offenbar.

Damit zeigen semantische Technologien einen vielversprechenden Weg auf, das essentielle Problem einer inhaltlichen Erschließung des World Wide Webs gleichermaßen automatisiert und qualitativ gesichert anzugehen. Die sich hieraus ergebende Frage ist: Inwieweit können die Methoden der Bibliografischen Kontrolle ebenfalls von diesen Techniken profitieren? Könnten damit auch den bibliotheksspezifischen Erschließungsproblemen begegnet werden? Welche Auswirkungen hätte dies auf die zukünftige Entwicklung der Bibliografischen Kontrolle und ihre zentralen Konzepte? Und schließlich, um auf das Eingangszitat zurückzukommen: Wo verlaufen eventuelle Parallelen, könnten Synergieeffekte zwischen Bibliografischer Kontrolle und Semantic Web genutzt werden?

Um diese Fragen zu beantworten, wird diese Arbeit in drei Teile gegliedert: Im ersten Abschnitt (Kapitel 2 und 3) sollen Konzepte der Bibliografischen Kontrolle grob chronologisch aufgerollt werden. Dabei werden die Entwicklung der traditionellen Katalogisierung hin zu einer metadatengestützten Bibliografischen Kontrolle dargestellt sowie neue Konzeptmodelle des Bibliografischen Feldes behandelt. In einem zweiten Teil (Kapitel 4) werden Prinzipien und Technologien beschrieben, die unter dem Begriff Semantic Web zusammengefasst werden und die Struktur des derzeitigen WWW durch maschinenverarbeitbare Informationen erweitern sollen. Im letzten großen Abschnitt (Kapitel 5) soll schließlich der zentralen Frage der Arbeit nachgegangen werden, in welcher Weise Konzepte der Bibliografischen Kontrolle wie bibliografische, Norm- und Metadaten durch Semantic-Web-Techniken weiterentwickelt werden können und inwieweit Bibliografische Kontrolle an sich vielleicht vor einem (weiteren) paradigmatischen Wandel steht.

2 Entwicklung und Konzepte der Bibliografischen Kontrolle

2008 erstellte die Working Group on the Future of Bibliographic Control der Library of Congress einen Bericht [On the record(2008)] über zukünftige Entwicklungslinien der Bibliografischen Kontrolle. Deren allgemeine Zielsetzung wird dort folgendermaßen umrissen:

...to facilitate discovery, management, identification, and access of and to library materials and other information products. [On the record(2008):1]

Bibliografische Kontrolle soll also ermöglichen, Bibliotheksmaterialien und andere Informationsprodukte aufzufinden, zu verwalten, eindeutig zu identifizieren und Zugang zu ihnen zu erlangen. In ihrem umfassendsten, internationalen Anspruch, einer *universal bibliographic control*, bedeutet dies

...to make universally and promptly available, in a form that is internationally acceptable, basic bibliographic data for all published resources in all countries. [ISBD(2007):vii, Introduction]

Bibliografische Daten aller Publikationen aus allen Ländern, sollen also möglichst zeit- und ortsunabhängig in international verständlicher Weise zur Verfügung gestellt werden. In einfachster Weise kann man bibliografische Daten als Referenzdaten von Publikationen begreifen, deren Erstellung und Struktur jedoch durchaus elaboriert ist. Lynne C. Howarth schreibt genauer:

Dedicated to the creation of surrogates – or bibliographic records – to represent actual physical – and more recently, virtual – items or objects, the theory and practice of bibliographic control has focused on *systematic, uniform, and consistent* approaches to *describing intellectual or artistic content and physical characteristics*. [Howarth(2005):39, Hervorhebungen K.W.]

Bibliografische Kontrolle ist demnach die Beschreibung von Objekten bezüglich formaler und inhaltlicher Merkmale, mittels derer Stellvertreter für diese Objekte erzeugt werden, die *bibliografischen Beschreibungen*. Dadurch, dass die Objekte möglichst systematisch, einheitlich und konsistent beschrieben werden, sind die bibliographic descriptions nachvollziehbar, gleich strukturiert (Syntax) und benutzen gleiche Begriffe (Semantik). Dies sind Voraussetzungen für einen Katalog, in dem die Beschreibungen vergleichbar zusammengeführt und Beziehungen zwischen ihnen hergestellt werden. Ein bibliografischer Katalog hat zwei grundsätzliche Aufgaben:

...on the one hand, its function as a tool for describing individual information objects and discriminating among them; on the other hand, its function of clustering information objects with one or several information characteristics in common. [Cordeiro(2003):3]

Es werden also bestimmte Charakteristika ausgewählt, nach denen die Zusammenführung bzw. Anordnung der einzelnen bibliografischen Beschreibungen erfolgt: die Festlegung der Ordnungselemente, die primärer (bei der Haupteintragung) oder sekundärer Art (bei Nebeneintragungen und Verweisungen) sein können und bei der Katalogbenutzung als *Sucheinstiege* fungieren. Für den alphabetischen Katalog sind dies Verfasser, Titel, Urheber (eine Körperschaft als Verfasser) sowie abweichende Namensformen dieser, für den

Sachkatalog Schlagwörter oder Notationen. Da die objektinternen Angaben oft differieren, obwohl es sich um denselben Autor, dieselbe Körperschaft, Ausgaben desselben Werkes handelt, normiert man deren Namen. Diese zumeist von nationalbibliografischen Zentren erbrachte *Normenkontrolle* in Form von Normdaten beinhaltet nicht nur die präferierte, vereinheitlichte Namensform, sondern auch Verweisungen von abweichenden Namensformen auf diese, eindeutige Identifikatoren wie z.B. die International Standard Serial Number (ISSN) bei Periodika, sowie Erläuterungen. Inhaltliche Sucheinstiege werden in ähnlicher Weise normiert indem *Kontrollierte Vokabulare* bzw. *Begriffssysteme* eingesetzt werden. Dabei meint der Begriff Vokabular nicht nur Schlagwortlisten und -systeme, sondern ebenso nicht natürlich-sprachlich basierte Systeme wie Klassifikationen. „Kontrolliert“ werden sie genannt, weil ihre Bezeichnungen (Schlagworte, Deskriptoren, Notationen) jeweils eindeutig auf Begriffe (Konzepte) bezogen wurden und Ambiguitäten der natürlichen Sprache wie Polysemie, Homonymie und Synonymie durch Definitionen oder Verweisungen aufgelöst wurden.

Bibliografische Kontrolle kann also konzeptionell in vier Bereiche geteilt werden, die sich notwendig aufeinander beziehen: Bibliografische Beschreibung, Katalogisierung, Normenkontrolle und Kontrollierte Vokabulare. Sie umfasst alle Prozesse und Methoden zur Erstellung von konsistenten, formal und inhaltlich deskriptiven Surrogaten, die ihre Objekte eindeutig identifizieren und lokalisieren, und deren systematische Anordnung in Katalogen. Im weiteren Sinn steht der Begriff auch für die Konzeption dieser Methoden und Mittel (z.B. bibliografische und Normdatenformate, Einsatzregeln für Kontrollierte Vokabulare, Katalogisierungsvorschriften). Im Folgenden sollen die vier Bereiche der Bibliografischen Kontrolle im Hinblick auf ihre internationale Entwicklung⁵ und ihre Prinzipien gesichtet werden.

2.1 Bibliografische Beschreibung

Grundlegend für die Bibliografische Kontrolle eines Objekts ist seine formale Beschreibung. Gemäß des Zieles einer Universellen Bibliografischen Kontrolle trieb die International Federation of Library Associations and Institutions (IFLA) seit 1969 die Ausarbeitung einer international verständlichen Form von bibliografischen Beschreibungen voran. Die International Standard Bibliographic Description for Monographic Publications [ISBD(M)] wurde erstmals 1974 verabschiedet, war ausschließlich auf Monographien ausgerichtet, legte aber die Grundstruktur für alle folgenden *ISBDs* fest. Weitere acht spezifische Versionen für unter-

⁵ Anstatt nationale Bibliografiegeschichte zu betrachten, ist es gewinnbringender, die internationalen Entwicklungslinien und die dahinter stehenden Konzepte zu untersuchen, da so einerseits ein allgemeinerer Überblick möglich ist; andererseits wird die zukünftige Entwicklung der Bibliografischen Kontrolle auf jeden Fall im internationalen Rahmen ablaufen. Aus diesem Grund wird auch eine Parallelität von deutschen und englischen Fachbegriffen und Eigennamen angestrebt.

schiedliche Materialien⁶ wurden geschaffen, sowie die ISBD(G) mit der übergreifenden Möglichkeit, alle Materialien zu beschreiben. Eingebettet waren diese Entwicklungen und stetigen Revisionen in das Programm Universal Bibliographic Control (UBC), später umbenannt in Universal Bibliographic Control and International MARC (UBCIM) und seit 2003 unter der Ägide der IFLA-CDNL Alliance for Bibliographic Standards (ICABS)⁷ [Payer(2009):2.1]. Bis heute bilden die ISBDs das international gültige Rahmenwerk für bibliografische Beschreibungen, sie führen die Bedingungen für Beschreibung und Identifizierung von Veröffentlichungen aus, die in Bibliotheken gesammelt werden.⁸ Die aktuelle, „fusionierte“ Ausgabe von 2007 (Preliminary Consolidated Edition) [ISBD(2007)] soll für sämtliche Materialien gelten, stellt allerdings nur eine vorläufige Fassung dar und wird i.M. überarbeitet. In dieser Fassung ist einleitend (noch) zu lesen:

The ISBD is the standard that determines the data elements to be recorded or transcribed in a specific *sequence as the basis of the description* of the resource being catalogued. In addition, it employs prescribed *punctuation* as a means of recognizing and displaying these data elements and making them understandable independently of the language of the description. [ISBD(2007):Introduction, vii, Hervorhebungen K.W.]

Dabei wird klar, dass diese Konzeption noch auf dem Einsatz in analogen, linearen Katalogen wie Band- und Kartenkatalogen basiert. Die ISBD ist hier auf textbasierte bibliografische Beschreibungen ausgerichtet, die für sich abgeschlossene Einheiten bilden. Um eine innere Struktur in dieser Textsequenz darzustellen, ist eine Kennzeichnung der einzelnen bibliografischen Elemente in Form von Deskriptionszeichen nötig. Richard P. Smiraglia hat darauf hingewiesen, dass es sich bei dieser „prescribed punctuation“ um eine „early form of mark up“ handelt. [Smiraglia(2005):5-6] Eine Bibliografische Beschreibung wird dadurch in acht Bereiche (sowie Unterbereiche) geteilt, welche die Veröffentlichung umfassend bezüglich formaler Gesichtspunkte beschreiben und (für mit der ISBD vertraute Nutzer) in ihrer einzelnen Bedeutung durch diese Anordnung verständlich sind. Diese acht Bereiche der ISBD sind: 1. Sachtitel- und Verfasserangabe, 2. Ausgabevermerk, 3. materialspezifische (oder die Veröffentlichungsart betreffende) Angaben, 4. Erscheinungsvermerk, 5. physische Beschreibung, 6. Gesamttitelangabe, 7. Fußnoten und 8. Standardnummer und ähnliche Nummern, Angaben zur Beschaffung und Preis. [Payer(2009)].

Die Beschreibung geht dabei von einem realen Objekt aus (*Autopsieprinzip*), gleichzeitig können aber aus der Beschreibung andere bibliografisch bedeutsame Einheiten abgeleitet werden: So aus der Verfasserangabe die Existenz des Verfassers oder Urhebers, aus dem

6 Eine Übersicht über die ehemaligen spezifischen ISBDs ist auf der IFLA-Internetseite zu finden, vgl. <http://www.ifla.org/en/isbd-rg/superseded-isbd-s>.

7 Der Nachfolger von ICABS ist seit August 2008 die IFLA-CDNL Alliance for Digital Strategies (ICADS), siehe <http://www.ifla.org/icads>

8 „The International Standard Bibliographic Description (ISBD) specifies the requirements for the description and identification of the most common types of published resources that are likely to appear in library collections.“ [ISBD(2007):0-1]

Hauptsachtitel die konzeptionelle Einheit des Werkes, aus dem Ausgabevermerk indirekt die Existenz anderer Ausgaben, gegebenenfalls aus der Gesamttitelangabe eine übergeordnete bibliografische Einheit und eventuell aus Fußnoten verwandte Werke. Analytisch betrachtet vereint die ISBD – in einer sehr eingeschränkten Weise – Ausgabe-, Werk- und Verfasserangaben und deutet Beziehungen zu über- und nebengeordneten bibliografischen Einheiten zumindest an. Zu erwähnen sind zudem die im letzten Bereich genannten Standard- und ähnliche Nummern, die der eindeutigen *Identifizierung* nicht des Objekts oder des Werks, sondern der spezifischen Ausgabe dienen.

An dieser grundlegenden Bibliografischen Beschreibung von Objekten wurde bisher nicht viel verändert. Auch die vor kurzem erschienene Erklärung zu den internationalen Katalogisierungsprinzipien [ICPde(2009)] (im Original: Statement of International Cataloging Principles, ICP), auf die im Weiteren noch einzugehen ist, beschränkt sich diesbezüglich auf einen Verweis auf die ISBD (Punkt 5). Sie wiederholt deren Ansatz, stützt sich allerdings in zentraler Weise⁹ auf die Konzeption und Terminologie des Functional Requirements for Bibliographic Records: Final Report (FRBR) [FRBR(1998)]. Aus dieser Sichtweise kann sie die klassische bibliografische Beschreibung genauer analysieren und differenziert die verschiedenen *bibliografischen Entitäten*¹⁰, auf die sich die ISBD bezieht:

5.2. Eine bibliografische Beschreibung soll typischerweise auf dem Exemplar als dem Repräsentant der Manifestation beruhen und kann Merkmale enthalten, die zu dem in ihr verkörperten Werk (den Werken) und zu der in ihr verkörperten Expression (den Expressionen) gehören. [ICPde(2009)]

Exemplare werden nun als Instanzen einer Klasse, der Manifestation angesehen (im herkömmlichen Sprachgebrauch die Ausgabe und/oder Auflage). Der bisherige Begriff des Werkes wird durch zwei Konzepte ersetzt, einerseits das Werk, das aber zu einer abstrakteren Einheit erhoben wird und andererseits die Expression, eine grundlegende, rezipierbare Umsetzung des Werkes¹¹. Die bibliografischen Entitäten, die in einer bibliografischen Beschreibung zu Grunde liegen, sind also vielgestaltig. Sie können abstrakten (Werke, Expressionen), realen (Exemplare, Personen), Klassen-Charakter¹² (Manifestation) haben, oder auch verschiedener Art sein (Körperschaften)¹³.

9 Neben den FRBR werden auch die Functional Requirements for Authority Data (FRAD) und die in Ausarbeitung befindlichen Functional Requirements for Subject Authority Data (FRSAD) genannt.

10 Der allgemeine Begriff Entität bzw. entity wird im Glossar der Erklärung zu den internationalen Katalogisierungsprinzipien (ICPde) mit Hilfe des „Webster's Third New International Dictionary of the English Language“ folgendermaßen definiert: „Etwas, das einen einheitlichen und in sich geschlossenen Charakter hat, etwas, das unabhängig ist oder gesondert existiert, eine Abstraktion, ein Begriff, ein Gedankengut oder ein metaphysisches Objekt.“ [ICPde(2009):13]

11 Vgl. die Erläuterungen von Werk und Expression in den FRBR, [FRBRde(2009):16,18].

12 Die FRAD weisen darauf hin, dass, abhängig von der Katalogisierungstradition, auch Personen und Körperschaften den Charakter einer Klasse annehmen können. [FRADdraft(2007)4-5]

13 Für die unterschiedlichen Sichtweisen auf Körperschaften, siehe ebenfalls [FRADdraft(2007), S. 4-5].

Zusammenfassend lässt sich sagen: Die traditionelle und immer noch aktuelle Bibliografische Beschreibung basiert auf einem konkreten Exemplar, beschreibt und identifiziert gezielt die Manifestation, dabei aber implizit auch die Expression bzw. das Werk – dies sind die bibliografischen Entitäten, die der FRBR-Report als Gruppe 1 zusammenfasst. Die Beschreibung benutzt (und verweist so auf) die Entitäten Personen beziehungsweise Körperschaften (die Gruppe-2-Entitäten¹⁴) und gegebenenfalls verwandte andere Entitäten der Gruppe 1. [FRBRde(2009):48] Die Beschreibung erfolgt durch Merkmalsnennung nach festgelegter Abfolge und syntaktischer Auszeichnung, zu Grunde liegt also ein *Attribut-Wert-Schema zur Beschreibung und Identifizierung der Manifestation*, wobei auch Merkmale von Expression, Werk und Exemplar einfließen und immanent Verweisungen zu verwandten bibliografischen Entitäten erfolgen. Da dies alles in einem Datensatz geschieht, ist keine systematische Trennung der entitätenspezifischen Merkmale bzw. der bibliografischen Gruppe-1-Entitäten überhaupt, gegeben. Dies wäre aber wiederum die erste Voraussetzung um sinnvolle Verknüpfungen zwischen diesen herstellen zu können, wie es zwischen Entitäten der Gruppe 1 und der Gruppe 2 in aktuellen Katalogen schon länger geschieht.

2.1.1 Functional Requirements for Bibliographic Records (FRBR)

Obwohl das ICP durch seine Verbindung von ISBD und FRBR also keine grundlegende Neuerung der bibliografischen Beschreibung gebracht hat, muss der Functional Requirements for Bibliographic Records: Final Report (FRBR) [FRBR(1998)] als ein entscheidender Wendepunkt in der Fortentwicklung der Bibliografischen Kontrolle gelten. Die grundlegende Analyse von bibliografischen Daten, die von 1992 bis 1995 im Auftrag des IFLA Universal Bibliographic Control and International MARC (UBCIM) Programme und der IFLA Division of Bibliographic Control entstand, hatte das Ziel,

...ein konzeptionelles Modell zu erstellen, das als Grundlage dient, um die spezifischen Merkmale und Beziehungen (die im Datensatz als einzelne Datenelemente widergespiegelt sind) mit den verschiedenen Anforderungen zu verknüpfen, welche die Benutzer stellen, wenn sie bibliografische Datensätze heranziehen. [FRBRde(2009):3]

Der 1997 verabschiedete FRBR-Report – die erste deutsche Fassung erschien 2006 als Funktionelle Anforderungen an bibliografische Datensätze: Abschlussbericht (FRBRde) – hat den Anspruch, „so weit wie möglich eine 'verallgemeinerte' Sicht des bibliografischen Universums“ [FRBRde(2009):5] aufzuzeigen, indem Objekte, Akteure und die Strukturen zwischen ihnen in systematischer Form, einem Entity-Relationship-Modell (ER-Modell)¹⁵, dargestellt werden. In vereinfachter Form besteht ein ER-Modell aus Entitäten mit spezifischen Merk-

¹⁴ Um genau zu bleiben, wird unter Gruppe-2-Elementen auch die Entität Familie genannt, gemäß dem Bezug auf die FRAD. Diese Entität ist in der klassischen ISBD-Konzeption nicht vorgesehen.

malen (Attributen), und den Beziehungen zwischen diesen Entitäten. In der FRBR-Studie nutzte man dies konkret zur

...Entwicklung eines Rahmens, der die Entitäten, die für Benutzer bibliografischer Datensätze von Interesse sind, die Merkmale jeder Entität und die Arten der Beziehungen zwischen den Entitäten identifiziert und klar definiert. [FRBRde(2009):3]

Die Study Group on the FRBR extrahierte diese grundlegenden bibliografischen Entitäten durch eine analytische Betrachtung von bibliografischen Datensätzen und ihren Funktionsanforderungen. Der FRBR-Bericht unterscheidet eine erste Gruppe von Entitäten, „Produkte von intellektuellen bzw. künstlerischen Anstrengungen“ (*Werk, Expression, Manifestation* und *Exemplar*), eine zweite Gruppe, „die für den intellektuellen bzw. künstlerischen Inhalt stehen, für die physische Produktion und Verbreitung oder den Schutz solcher Produkte“ (*Person* und *Körperschaft*) sowie eine dritte Gruppe, die zusätzlich zu den beiden genannten Gruppen „als Themen von intellektuellen oder künstlerischen Anstrengungen dienen“ (*Begriff, Gegenstand, Ereignis* und *Ort*). [FRBRde(2009):12]

Beziehungen zwischen den bibliografischen Entitäten

Die verschiedenen Beziehungen zwischen diesen Entitäten werden zunächst „auf der höchstmöglichen Verallgemeinerungsstufe“ definiert, und zwar innerhalb der Gruppe-1-Entitäten, zwischen diesen und den Gruppe-2-Entitäten sowie die Themen-Beziehung zwischen der Entität *Werk* und den Entitäten aller drei Gruppen. Am besten erschließen sich diese Beziehungen durch folgende grafische Darstellungen, die der FRBR-Studie entnommen sind. [FRBRde(2009):12-15, 49-54] Entitäten werden hier als Rechtecke, Beziehungen als beschriftete Linien dargestellt, die ihre Kardinalität¹⁶ in Form von verschiedenen Pfeilspitzen anzeigen. [FRADdraft(2007):3] Ein Beispiel im ersten Diagramm: Genau ein *Werk* (einfache Pfeilspitze) wird durch eine oder mehrere *Expressionen* (Doppelpfeilspitze) realisiert.

15 Das Entity-Relationship-Modell geht auf Peter Pin-Shan Chen zurück, der es 1976 in seinem Aufsatz *The Entity-Relationship Model - Toward a Unified View of Data* vorstellte, vgl. <http://csc.lsu.edu/news/erd.pdf> Abweichend von dem Chenschen Modell wird in den FRBR/FRAD/FRSAD-Modellen stets von Entitäten und Beziehungen gesprochen, obwohl eigentlich Entitätentypen und Beziehungstypen gemeint sind.

16 Die Kardinalität eines Beziehungstyps (Relationship-Typs) gibt an, mit wie viel anderen Entitäten eine Entität eines bestimmten Entitätentyps in einer konkreten Beziehung stehen kann bzw. muss. Vgl. http://de.wikipedia.org/wiki/Kardinalit%C3%A4t_%28Datenbanken%29

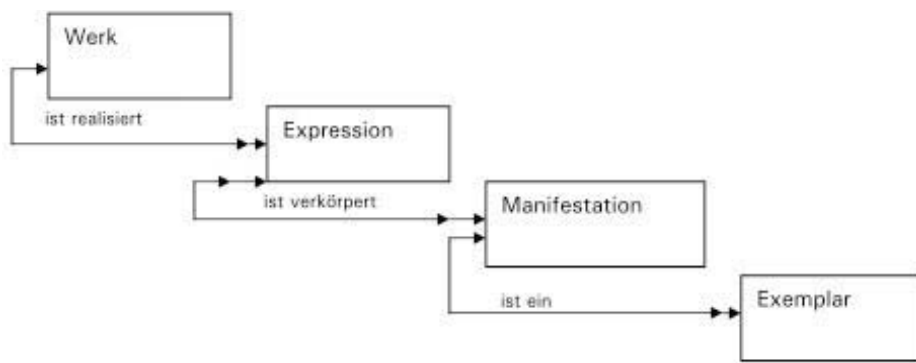


Abbildung 1: FRBR - Entitäten der Gruppe 1 und Primärbeziehungen [FRBRde(2009):13]

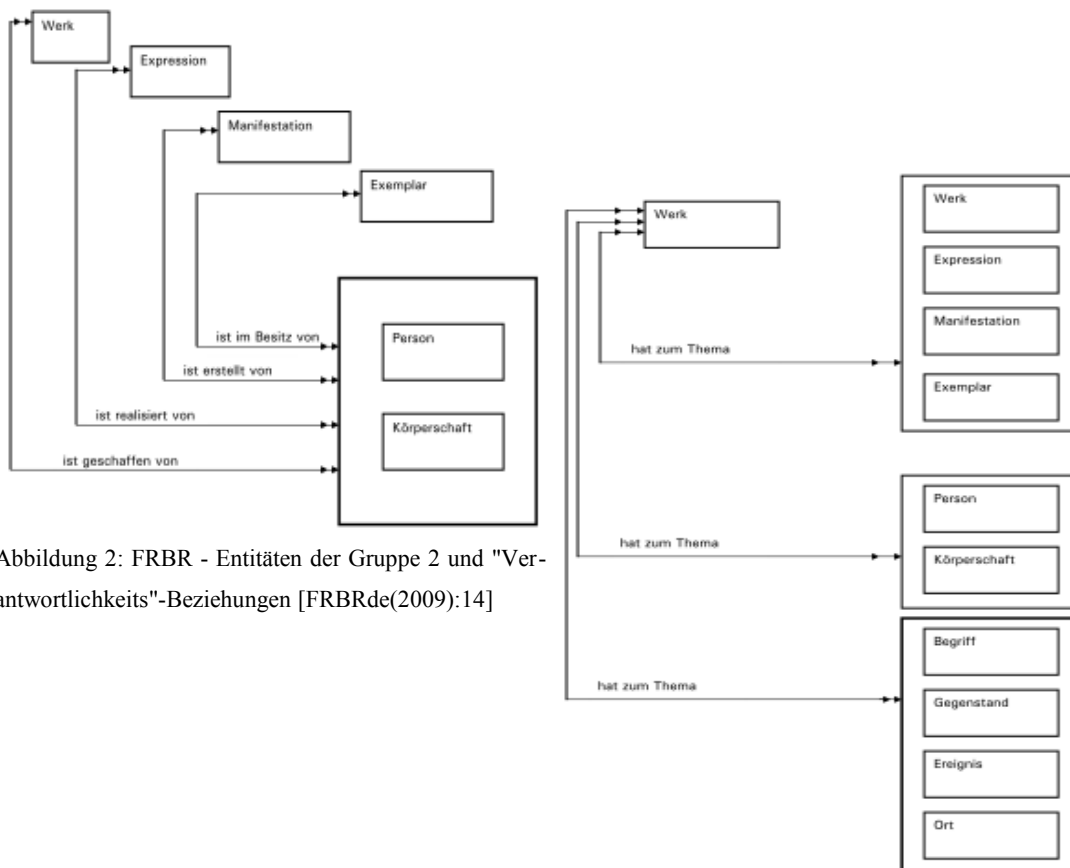


Abbildung 2: FRBR - Entitäten der Gruppe 2 und "Verantwortlichkeits"-Beziehungen [FRBRde(2009):14]

Abbildung 3: FRBR - Entitäten der Gruppe 3 und "Themen"-Beziehungen [FRBRde(2009):15]

Die Beziehungen in Abbildung 1, die als Abstrahierungen von spezifischeren Beziehungen anzusehen sind, benennen sozusagen die stufenweisen Ableitungen der konzeptionellen Ebenen eines Werkes bis hinunter zum Exemplar. Eine entscheidende Neuerung im Vergleich zu

den traditionellen bibliografischen Begriffen stellt dabei die Entität Expression dar, diese wird als

...die intellektuelle bzw. künstlerische Realisierung eines *Werkes* in Form von Buchstaben, Zahlen, Noten, Choreografien, Tönen, Bildern, Gegenständen, Bewegungen usw. oder einer Kombination dieser Formen [definiert].

Eine Expression ist die spezifische intellektuelle bzw. künstlerische Form, die ein *Werk* jedes Mal annimmt, wenn es realisiert wird. [FRBRde(2009):18, Hervorhebungen original]

Insofern nimmt sie eine diffizile Zwischenposition zwischen der vollends abstrakten Entität Werk und der bereits durch physische Merkmale beschreibbaren Manifestation ein:

Expressionen sind allerdings so definiert, dass Aspekte der physischen Form, wie z.B. Schriftart und Seitenlayout, die kein notwendiger Bestandteil der geistigen bzw. künstlerischen Realisierung des Werkes sind, ausgeschlossen sind. [FRBRde(2009):18]

Hintergrund für diese Neudefinition einer bibliografischen Entität ist, dass es so möglich wird, jede Änderungen der Form und an „intellektuellen Konventionen bzw. Instrumenten, [...] um ein Werk auszudrücken“ eigenständig zu erfassen. Dies betrifft z.B. Umsetzungen von Texten in eine gesprochene oder musische Darstellung oder Übersetzungen dieser. Prinzipiell wird hier der Brückenschlag von unterschiedlichen Werkformen, die sich auf eine gemeinsame Werk-Idee beziehen, möglich.

Zudem werden, wie in Abbildung 2 ersichtlich, die Verantwortlichkeitsbeziehungen zwischen den Gruppe-2-Entitäten Person und Körperschaft zu Werk, Expression, Manifestation und Exemplar definiert. Schließlich, wie in Abbildung 3 dargestellt, wird die Themenbeziehung zwischen der Entität Werk und allen definierten Entitäten kurz erläutert. [FRBRde(2009):52-54]

Dezidiert widmet sich die FRBR-Studie jedoch den Beziehungen zwischen den Gruppe-1-Entitäten und spezifiziert die Beziehungen, die quasi quer oder diagonal zwischen Gruppe-1-Entitäten gleicher oder verschiedener konzeptioneller Ebenen, sowie desselben oder verschiedener Werk-Exemplar-Bäume verlaufen. Damit sollen inhaltliche Beziehungen, die auch schon teilweise in der traditionellen Katalogisierung durch Verweisungen dargestellt wurden, begründet und systematisiert werden. Es werden explizit genannt und durch Beispiele erläutert [FRBRde(2009):54-71]:

1. Beziehungen zwischen Werken

Bei Nachfolgern, Beilagen und Ergänzungen unterscheidet man zwischen abhängigen und unabhängigen Werken und dies einzelfallspezifisch.

Zusammenfassung, Adaption, Umsetzung und Nachahmung dagegen stellen immer Beziehungen zwischen unabhängigen Werken dar.

2. Beziehungen zwischen dem Ganzen und Teilen auf der Werk-Ebene

Auch hier werden unabhängige (z.B. Monografie in einer Schriftenreihe, Aufsatz in einer Zeitschrift) und abhängige Teile (Verständnis nur im ganzen Werkzusammenhang möglich,

zumeist ohne Namen/Titel) unterschieden, und Letztere nochmals in Segmente (erkennbare Abschnitte wie Vorwort, Kapitel) und systemische Teile (notwendiger Aspekt, der mit der Werkgesamtheit verflochten ist, etwa eine Abbildung im Text) unterteilt.

3. Beziehungen zwischen Expressionen desselben Werkes

Kürzung, Überarbeitete Ausgabe, Übersetzung und Bearbeitung (Musik) werden als unabhängige Expressionen desselben Werkes definiert.

4. Beziehungen zwischen Expressionen unterschiedlicher Werke

Nachfolger, Beilage, Ergänzung können unabhängiger oder abhängiger Natur sein; Zusammenfassung, Adaption, Umsetzung und Nachahmung werden stets als Beziehungen zwischen unabhängigen Werken gefasst. Hier werden die Beziehungen, die auf der Werk-Ebene definiert werden, analogisiert.

5. Beziehungen zwischen dem Ganzen und Teilen auf der Expression-Ebene

Auch hier wird auf die bereits definierten Ganzes-Teil-Beziehungen auf der Werk-Ebene verwiesen mit dem Hinweis, dass es sich aber um expressionsspezifische Teile handelt, etwa Verzeichnisse und Indices.

6. Beziehungen zwischen Expression eines Werkes und anderen Werken

Diese entsprechen der Beziehungen zwischen Expressionen unterschiedlicher Werke. Sie werden gesondert definiert, da die Feststellung einer Expression-Werk-Beziehung oft einfacher ist als die der entsprechenden Expression-Expression-Beziehung.

7. Beziehungen zwischen Manifestationen derselben Expression

Reproduktion (z.B. Mikroreproduktion, Nachdruck, Mirror Site) und Alternative (z.B. gleichzeitig herausgegebene Ausgabe, alternatives Format) werden hier unterschieden.

8. Beziehungen zwischen dem Ganzen und Teilen auf der Manifestations-Ebene

Teile einer Manifestation werden nicht weiter typisiert, sie stellen sozusagen die Abstraktionen der realen Teile auf der Exemplar-Ebene dar, etwa von Bänden einer mehrbändigen Werkes, oder der Tonspur eines Films.

9. Beziehungen zwischen Manifestation und Exemplar

Reproduktionen dienen der Kennzeichnung einer Manifestation, die aus der Reproduktion eines bestimmten Exemplars hervorgegangen ist.

10. Beziehungen zwischen Exemplaren

Reproduktion und Umgestaltung werden unterschieden, Letztere insbesondere für den Fall des Zusammenbindens von verschiedenen Exemplaren oder fortlaufenden Sammelwerken.

11. Beziehungen zwischen dem Ganzen und Teilen auf der Exemplar-Ebene

Eigenständige Teile auf der Exemplar-Ebene sind etwa Bände eines mehrbändigen Exemplars („copy“). Integrale Teile eines Exemplars sind z.T. physisch untrennbar mit dem Exem-

plar verbunden und werden nicht separat betrachtet, etwa die Hülle einer CD oder der Einband eines Buches.

Merkmale der bibliografischen Entitäten

Schließlich werden im FRBR-Report Attribute (Merkmale) für die bibliografischen Entitäten definiert. Dazu wurden aus den einschlägigen internationalen Rahmenwerken bibliografische Merkmale herausgefiltert und diese eindeutig den definierten bibliografischen Entitäten zugeordnet. Da die FRBR-Studie auf einer Modellebene arbeitet, sind diese Attribute abstrakt konzipiert, deren Umsetzung in Datenelemente durchaus komplexer sein kann. Aus diesem Grund bietet der FRBR-Report eine Analyse und Abbildung der traditionellen Datenelemente der ISBD, GARE, GSARE¹⁷ und UNIMARC¹⁸ auf die Merkmale der FRBR-Entitäten im Anhang A. [FRBRde(2009):31-32]

Zudem werden diese Merkmale ausdrücklich als nicht abschließend und nicht notwendig für jede Instanz einer Entität konzipiert. Explizit werden die Attribute für die Entitäten der ersten, in abgeschwächter Form auch der zweiten Gruppe betrachtet. [FRBRde(2009):32-53] Die Entitäten der dritten Gruppe erhalten (vorerst) jeweils nur das Attribut „Bezeichnung“ der Entität. [FRBRde(2009):53-54] Die Merkmalsanalyse erfolgte hier eingeschränkt, nämlich nur in dem Maße, in dem sie für bibliografischen Datensätze von Bedeutung ist. Da die Beschreibung dieser Entitäten Anliegen der Normenkontrolle sind, wurde deren umfassende Behandlung auf eigene Studien, FRAD und FRSAD, vertagt. [FRBRde(2009):32]

Beispielsweise seien hier die Merkmale der Gruppe-1-Entität Werk aufgeführt:

- Titel des *Werkes*
- Form des *Werkes*
- Datum des *Werkes*
- andere unterscheidende Eigenschaft
- geplanter Abschluss
- Zielgruppe
- Kontext des *Werkes*
- Aufführungsmedium (musikalisches *Werk*)
- Numerische Kennzeichnung (musikalisches *Werk*)
- Tonart (musikalisches *Werk*)
- Koordinaten (kartografisches *Werk*)
- Äquinoktium (kartografisches *Werk*) [FRBRde(2009):32]

Da der Anspruch bestand, möglichst alle Veröffentlichungsformen in das Modell einzubeziehen, wurden zu den grundsätzlichen Merkmalen von Werk, Expression und Manifestation spezifische Merkmale für einige Werk- und Medienformen hinzu genommen. So werden spezifische Attribute für musikalische und kartografische Werke für die Entität Werk definiert;

17 GARE, Guidelines for Authority and Reference Entries, genauer erläutert im Kapitel zur Normenkontrolle; GSARE, Guidelines for Subject Authority and Reference Entries, vgl. [GSARE(1993)].

18 UNIMARC, siehe <http://www.ifla.org/en/publications/unimarc-formats-and-related-documentation>

ebenso wie für fortlaufende Sammelwerke, Musiknoten oder Tonaufnahmen, kartografische Bilder oder Gegenstände, Fernerkundungsbilder und grafische bzw. projizierte Bilder bei der Entität Expression. Für die Entität Manifestation werden neben den allgemeinen spezifische Attribute für Printmedien, handgedruckte Bücher, fortlaufende Sammelwerke, Tonaufnahmen, Bilder, Mikroformen, Projektionen, elektronische Ressourcen und elektronische Ressourcen im Fernzugriff definiert.

Nationalbibliografischer Basisdatensatz

Das zweite große Ziel der FRBR-Studie bestand in der Formulierung von „Basisfunktionalitäten von Datensätzen, die von nationalbibliografischen Agenturen erstellt werden“ – einem sogenannten *nationalbibliografischen Basisdatensatz*, der auf dem entwickelten ER-Modell basiert. [FRBRde(2009):19] Neu ist dabei der Bezug von bibliografischen Grundforderungen des Findens, Auswählens, Identifizierens, Zugänglichmachens auf die einzelnen Entitäten Werk, Expression und Manifestation. [FRBRde(2009):86] Auf dieser Grundlage werden die logischen Merkmale und Beziehungen zusammengetragen, die zur Erfüllung dieser entitätenbezogenen Anforderungen nötig sind, und die entsprechenden spezifischen Datenelemente, größtenteils auf den ISBDs und den GARE basierend, aufgeführt. [FRBRde(2009):89-100] Da die hier die Rückbeziehung auf die bibliografischen Strukturen der ISBDs (als beschreibende Elemente) und GARE (als ordnende Elemente) erfolgt, bleibt eine strukturelle Neuformulierung eines „nationalbibliografischen Datensatzes“ aus und es wird größtenteils nur den status quo von ISBD/GARE-konformen Titelaufnahmen bzw. -datensätzen bestätigt bzw. diese an das Modell der FRBR angeschlossen. [FRBRde(2009):101-105]

Die Vermengung der unterschiedlichen bibliografischen Ebenen von Werk, Expression, Manifestation¹⁹ und den differenzierten Beziehungen zwischen diesen bleibt hier vorerst bestehen. Eine grundlegende Neukonzeption mit der Definition von entitätenspezifischen Beschreibungen und der adäquaten Darstellung ihrer Entitätenbeziehungen wird hier nicht angestrebt – entgegen der erwähnten entitätenbezogenen Analyse der bibliografischen Funktionalitäten. Es scheint, als ob die gute analytische Arbeit, die mit dem FRBR-Modell geleistet wurde, an der entscheidenden Stelle, nämlich der praktischen Umsetzung in Form eines Dekonstruierens des Konzepts Bibliografische Beschreibung, abgebrochen wurde.

Zusammenfassend muss die besondere Leistung des FRBR-Report in der systematischen Klärung der differenzierten bibliografischen Merkmale und Beziehungen gesehen werden, d.h. in deren Zuordnung zu den konzeptionellen Ebenen von Werk, Expression, Manifestati-

¹⁹ Die Entität Exemplar und ihre Merkmale und Beziehungen wird aufgrund der Ausrichtung auf nationalbibliografische Daten dort außen vor gelassen.

on und Exemplar. Gleichzeitig ist diese komplexe Modellierung von bibliografischen Strukturen werkart- und medienübergreifend angelegt, sodass erstmals Beziehungen „zwischen den Künsten“ adäquat darstellbar sind.

Prinzipiell könnte die ER-modellierte Konzeption des bibliografischen Feldes den Grundstein für eine neue, strukturiertere und formalere Art der Bibliografischer Kontrolle legen, sofern Bibliografische Beschreibung und Katalogisierung daran konsequent neu ausgerichtet werden. Dies würde aber auch bedeuten, den bisherigen Ansatz von bibliografischen Datensätzen, die entitätenspezifische Angaben integrieren, in Frage zu stellen und ein neues System von entitätenspezifischen Beschreibungsdaten zu forcieren.

2.2 Katalogisierung

Die Bibliografische Beschreibung bildet wie bereits erwähnt nur das Grundgerüst einer funktionstüchtigen *Titelaufnahme* (bibliographic record²⁰) in einem Katalog, dies betont auch die aktuelle Fassung der ISBD und gibt gleichzeitig einen guten Überblick über die Komponenten einer Titelaufnahme:

The ISBD description forms *a part of a complete bibliographic record* and is not normally used by itself. The other elements that make up a complete bibliographic record, such as *headings and subject information*, are not included in the ISBD stipulations. The rules for such elements are normally given in cataloguing codes and other standards.

References from variations of the title proper or other references that may be called for in a national cataloguing code do not form a part of the bibliographic description and are not provided in the ISBD. They may be indispensable, however, in catalogues and bibliographies.

Informational elements relating to a particular collection (location of material, recording of holdings, etc.) are not provided by the ISBD; they are local elements that can be added to the general description.“ [ISBD(2007):S. 0-2, Hervorhebungen K.W.]

Katalogisierungsregelwerke legen in umfassender Weise fest, wie diese Titelaufnahmen inhaltlich und formell aussehen. Durch die Wahl von Ordnungselementen legen sie deren Abfolge, also die Struktur des Katalogs fest. Diese Ordnungselemente sind formale und sachliche *headings*²¹, also „Überschriften“ der Titelaufnahmen, die das Beieinander-Ordnen von formal oder sachlich ähnlichen Beschreibungen, die so genannte *Kollokation*, möglich machen. Neben Titelaufnahmen beziehen sich Katalogisierungsregelwerke auf Verweisungen²² und Nebeneintragen und haben auf die Verwendung von Lokaldaten (z.B. die Signatur) sowie von spezifischen Normdaten und kontrollierten Vokabularen Einfluss. Auch wenn ei-

20 Der Begriff Titelaufnahme wurde durch die Einführung von elektronische Kataloge mittlerweile durch eher durch den Begriff *Bibliografischer Datensatz* ersetzt, siehe die Definition in der ICPde: „Ein Set von Datenelementen, das eine bibliografische Ressource beschreibt und den Zugriff darauf bietet, sowie mit dieser verwandte Werke und Expressionen identifiziert.“ [ICPde(2009):12] Der entsprechende englische Begriff *bibliographic record* wird sowohl für traditionelle und elektronische Katalogisierung verwendet.

21 Mittlerweile spricht man eher von *access points* [ICP(2009):14] bzw. *Sucheinstiegen* [ICPde(2009):17].

22 Der Vollständigkeit halber soll hier erwähnt werden, dass auch die Begriffe *Verweisung/reference* prospektiv durch Abweichende Form des Namens bzw. *Variant Form of Name* ersetzt werden, vgl. [ICPde(2009):17] und [ICP(2009):14].

nige Entscheidungen zur Katalogisierung den anwendenden Institutionen überlassen werden, so haben (meist national gültige) Katalogisierungsregelwerke doch eine herausgehobene Stellung unter den vier Bereichen der Bibliografischen Kontrolle, da sie alle anderen Bereiche in der Praxis der Katalogisierung zusammenführen.

Um zu einer grundlegenden internationalen Übereinkunft von Mindestanforderungen an Formalkatalogen zu kommen, wurde bereits 1961 die International Conference on Cataloguing Principles einberufen, die als Ergebnis das Statement of Principles [PP(1961)] verabschiedete. Nach dem Ort dieser Konferenz spricht man auch von den Paris Principles, sie sollten zur Grundlage zahlreicher nationaler Erschließungsregelwerke werden, wie z.B. die Regeln für die alphabetische Katalogisierung (RAK) und Anglo-American Cataloguing Rules (AACR). Die Paris Principles waren auf die Beschreibung von Büchern (und ähnlichen Materialien, wie es in der Fußnote heißt) in Form von Titelaufnahmen in alphabetischer Abfolge eines Band- oder Kartenkatalogs ausgerichtet. Dabei beschränkten sie sich auf

...choice and form of headings and entry words – i.e. to the principal elements determining the order of entries – in catalogues of printed books in which entries under authors' names and [...] under the titles of works are combined in one alphabetical sequence. [PP(1961):1. Scope of statement]

Sie legten damit die Struktur von alphabetischen Katalogen fest sowie grundlegende Regeln zu den Eintragungen – etwa unter welchen Bedingungen die Bibliografische Beschreibung entweder unter dem einzigen Verfasser, dem ersten Verfasser, dem Titel oder dem korporativen Urheber eingeordnet werden soll. In den Paris Principles wurden damit Prinzipien der Formalerschließung festgeschrieben²³, die auch heute noch von Belang sind, wie die punktuelle Sammlung und damit Recherchierbarkeit 1. aller Werke eines Autors (eines persönlichen wie korporativen, auch wenn nicht explizit formuliert) sowie 2. aller Ausgaben eines Werkes im Katalog (Punkt 2 und 3), also die *formale Kollokation*. Mit der Unterscheidung von Haupteintragungen (main entries), Nebeneintragungen (added entries) und Verweisungen (references) (Punkt 4 und 5) wurden abweichende Sucheinstiege (headings) unterstützt und zugleich das Prinzip einer eindeutigen, festgelegten Form (uniform headings) von Titeln sowie Autoren- und Körperschaftsnamen (*Ansetzungs- bzw. Normierungsprinzip*) eingeführt. Zur Ansetzung von Personennamen wurde im letzten Punkt 12 schließlich das *Staatsbürgerprinzip* benannt. Zu bemerken ist, dass in den Paris Principles wenn auch nicht ausformuliert, aber implizit die Unterscheidung von book (das als Bibliotheksmaterial das Objekt der Formalerschließung ist) und work auftaucht (so z.B. in Punkt 2.2), sowie die Forderung nach *Individualisierung* von Personen- und Körperschaftsansetzungen (Punkt 8.22 und 9.46). Vor

²³ Diese Prinzipien wurden als internationale Richtlinie *festgeschrieben*, erfunden wurden sie bei weitem früher. Eine kurze Zusammenfassung zur Geschichte von Bibliothekskatalogen bietet [Haynes(2004):2-4].

dem deutschen RAK-Hintergrund ist zudem die der persönlichen gleichgestellte Urheberschaft von Körperschaften (corporate bodies) (Punkt 9) hervorzuheben.

Bis zu einem Neuentwurf eines mit den Paris Principles vergleichbaren, grundlegenden Rahmenwerks für Katalogisierungsregelwerke sollte einige Zeit vergehen. In der Zwischenzeit entstanden nicht nur viele nationale und internationale Standards und differenzierten sich angesichts neuer Medienformen aus, auch die technische Umsetzung der Katalogisierung durchlebte einen grundlegenden Wandel mit elektronischer Datenverarbeitung, Online-Katalogen und wachsendem Datenaustausch. [FRBR(1998):1]

2.2.1 Statement of International Cataloguing Principles (ICP)

Seit 2003 organisierte die IFLA Cataloguing Section internationale Treffen von Katalogisierungsexperten um die – sehr erfolgreichen – Paris Principles zu überarbeiten. 2009 wurde schließlich das neue Statement of International Cataloguing Principles (ICP) [ICP(2009)], in deutscher Fassung Erklärung zu den internationalen Katalogisierungsprinzipien [[ICPde(2009)]], veröffentlicht, dass die Tradition als „Leitlinie für die Entwicklung von Katalogisierungswerken“ weiterführt. Jedoch trifft das ICP weit über die Paris Principles hinausgehende Vorgaben für die Generierung und Nutzung von bibliografischen Daten, sowie als grundsätzliche Neuerung auch für Normdaten; und bezieht sich zudem ausdrücklich auf alle Materialarten. Im Gegensatz zu den Paris Principles benennt das ICP explizit allgemeine Prinzipien zur Katalogerstellung. Herausragendes Gebot ist der Benutzerkomfort, nachgeordnete und gleichgestellte weitere Prinzipien sind Allgemeine Gebräuchlichkeit, Wiedergabe, Ausführlichkeit und Notwendigkeit, Bedeutung, Ökonomie, Konsistenz und Standardisierung sowie Integration. [ICPde(2009):2-3] Hervorzuheben sind neben dem als „Wiedergabe“ benannten *Autopsieprinzip* vor allem die Forderung nach „soweit wie möglich“ *standardisierten bibliografischen Beschreibungen* und *Sucheinstiegen* unter dem Prinzip Konsistenz und Standardisierung (beides in Punkt 5). Normierte Sucheinstiege, also Ansatzungsformen und abweichende Namensformen, sollen in einem *Normdatensatz* zusammengefasst werden. Bestehend dabei: diese Normdaten sollen nicht mehr nur für Titel, Verfasser und Urheber geschaffen werden, sondern *für alle bibliografischen Entitäten* (Punkt 6.1. und 6.2.).

Wie bereits beschrieben, bezieht sich das ICP in zentraler Weise auf die FRBR-Studie. Zum konkreten Einsatz des FRBR-Modells sagt das ICP jedoch nicht sehr viel, nur dass „ein Katalogisierungswerk [...] die Entitäten, Merkmale und Beziehungen berücksichtigen“ soll. Im Einzelnen: „Merkmale, die jede Entität identifizieren“, sollen als Datenelemente verwendet werden und „[b]ibliografisch aussagekräftige Beziehungen“ zwischen den Entitäten dargestellt werden; vage bleibt die Aussage, dass die oben genannten Entitäten durch bibliografische Daten und Normdaten repräsentiert werden „können“. [ICPde(2009):3-4]

Schließlich gibt das ICP auch eine systematische und umfangreiche Aufstellung zu Funktionen eines Katalogs in Form des klassischen bibliothekarischen Aufgabenprofils des Findens, Identifizierens, Auswählens und des Zugänglichmachens von Informationen. Es vollzieht damit einerseits die Möglichkeiten nach, welche die computergestützte Katalogisierung bereits bietet und die auch schon länger genutzt werden wie nicht-normierte, zusätzliche Sucheinstiege (z.B. beitragende Personen, Erscheinungsorte, Medienformen) und faceted search bzw. browsing. [Tillett(2001):30]

Andererseits steckt die strukturelle Umsetzung des FRBR-Modells in den nationalen Katalogisierungsregelwerken meist noch in den Kinderschuhen. Dies wird besonders offensichtlich, wenn das ICP zusätzlich zu den genannten traditionellen Funktionalitäten eines Katalogs das Navigieren²⁴ über bibliografische Beziehungen „im Katalog und darüber hinaus“ einfordert. Um alle Beziehungen zwischen den bibliografischen Entitäten erfahrbar und nutzbar zu machen ist eine vorherige Abbildung der bibliografischen und Normdaten auf die FRBR-Konzepte dringend notwendig. Dies stellt eine große Aufgabe für Bibliotheken weltweit dar, verspricht aber auch bisher ungeahnte Möglichkeiten der Katalogisierung.

Das Kollokationsprinzip würde jenseits der minimalen Verfasser/Werk- und Werk/Exemplar-Beziehungen in eine vielseitige (formale wie inhaltliche) und vielschichtige Vernetzung von bibliografischen Entitäten weiterentwickelt. Umfangreiche Normierungen dieser würde die eindeutige Identifizierung von Personen, Körperschaften, Familien und Werken, von Begriffen, Gegenständen, Ereignissen und Orten voran treiben – damit natürlich Precision und Recall bei der Recherche entscheidend erhöhen – vor allem aber auch eine breite (internationale) Zusammenführung der Norm- und bibliografischen Daten ermöglichen. Insgesamt ergäbe sich nicht nur eine vielfach erweiterte Funktionalität von Online-Katalogen, sondern vor dem Hintergrund des internationalen Austauschs eine ganz neue Qualität von Bibliografischer Kontrolle. Über die Bibliothekswelt hinaus könnten deren Grundprinzipien wirksam werden:

The global information society, where the universal character of the Web claims to offer democratic access to information, needs more than ever the ideas of internalisation and interoperability behind the BUC (Bibliographic Universal Control), now as a *Universal Web Control*. [McCathieNevile/Méndez(2007):27]

2.3 Normenkontrolle

Die beiden Komponenten Normenkontrolle und Kontrollierte Vokabulare erfüllen im bibliografischen Gefüge prinzipiell den gleichen Zweck, die Standardisierung und Normierung von Bezeichnungen, die zur bibliografischen Beschreibung verwendet werden. Dabei unterschei-

²⁴ Das Navigieren als neue Funktion des Katalogs nennt Barbara Tillett bereits 2001, vgl. [Tillett(2001):32].

den sie sich nach landläufiger Meinung in Bezug auf ihren Anwendungsbereich: während *Normdaten* Namensformen von bibliografischen Entitäten wie Personen, Körperschaften und Werken vereinheitlichen und diese eindeutig identifizieren sollen²⁵, sind *Kontrollierte Vokabulare* (KOS) auf die sprachliche Kontingenz von Begriffen (Konzepten) ausgerichtet. Zudem kann man im Allgemeinen von einer geringer ausgeprägten Komplexität von Normdaten gegenüber Kontrollierten Vokabularen ausgehen. Dass diese Unterscheidung in der Praxis nicht einfach durchzuhalten ist, zeigt ein Beispiel. So erscheint in systematischer Sichtweise der Normdaten-Begriff bezüglich der deutschen Schlagwortnormdatei (SWD) und der Library of Congress Subject Headings (LCSH) verwirrend, da beide prinzipiell (Normierung von Konzepten) und strukturell (hierarchische Schlagwortsysteme) eher den Kontrollierten Vokabularen (KOS) zugerechnet werden können. Dass hier trotzdem von Normenkontrolle bzw. subject authority control gesprochen wird, erklärt sich durch die Erarbeitung durch die nationalbibliografischen Zentren als Standards für den allgemeinen Gebrauch als kontrollierte Sucheinstiege (analog zu den anderen Normdaten). Diese Normdaten zur sachlichen Erschließung basieren also zum großen Teil auf Kontrollierten Vokabularen; personen-, körperschafts-, und einheitssachtitelbezogene Normdaten sind dagegen Beschreibungen von bibliografischen Entitäten. Fasst man beide Ausprägungen der Normenkontrolle zusammen²⁶, vereint diese also sowohl beschreibende wie klassifizierende Merkmale, weist also Ähnlichkeiten mit der Bibliografischen Beschreibung und mit Kontrollierten Vokabularen auf. Normdaten können dann recht unterschiedlichen Charakter annehmen.

Folglich erscheint diese Sichtweise wenig sinnvoll, wenn es um die Analyse von Prinzipien der Bibliografischen Kontrolle gehen soll. Auch in Anbetracht der zukünftigen Rahmenwerke FRAD und FRAD ist eine getrennte Behandlung der beiden Bereiche der Normenkontrolle sinnvoll. In diesem Kapitel soll also vom engeren Normenkontrollbegriff, Normdaten, die sich auf bibliografischen Entitäten beziehen, ausgegangen werden.

Wie die Begriffe *Normdateien* und *Normlisten* (authority files oder lists) bereits andeuten, unterliegen Normdaten einem breiten formalen Spektrum. So herrschen im Bereich der Normierung von Personennamen (noch) eher Normdaten mit einfachem Listencharakter vor: das vorrangige Ziel besteht hier in der *Individualisierung* der einzelnen Personen. Als prominente Beispiele sind die deutsche Personennamendatei (PND), die Library of Congress Name Authority Files (LCNAF), oder die Union List of Artists (ULAN²⁷) zu nennen. In ähnlicher

25 „...the concept of 'authority control', which means both management of authorized forms and identification of the entities that are represented by those access points...“ [FRADdraft(2007):1]

26 So fasst z.B. Jane Greenberg aus der Sichtweise der Metadatendiskussion KOS und Normdateien unter dem Begriff „content value schemas“ zusammen, vgl. [Greenberg(2003)].

27 ULAN, siehe <http://www.getty.edu/research/tools/vocabulary/ulan/>

Weise kontextualisiert die deutsche Gemeinsame Körperschaftsdatei (GKD) in begrenztem Rahmen ihre normierten Körperschaftsnamen, verfügt aber zudem über hierarchische und chronologische Relationen zwischen ihnen. Es ist davon auszugehen, dass relevante Relationen auch zwischen Personennamen zukünftig eine größere Rolle spielen werden.

Wenn von der Normenkontrolle die Sprache ist, so sind die bibliografischen Entitäten gemeint, die eine reale (im Falle von beteiligten Personen und Körperschaften) oder auch metaphysische (im Fall eines wissenschaftlichen oder künstlerischen Werkes) Einzigartigkeit auszeichnet, die durch einen Eigennamen ausgedrückt wird. Im Einzelnen sind dies folgende Aufgaben:

1. die Identifizierung einzelner Entitäten angesichts variierender Namen und Namensformen,
2. die Aufzeichnung aller bedeutsamen Namensformen,
3. die konsistente Wahl von gleichbleibenden oder autorisierten Ordnungselementen für die Einzeltitelanzeige (i.e. die Ansetzungsform, K.W.)
4. die Feststellung und Unterscheidung von gleichen Namen, die auf verschiedene Entitäten verweisen. [Buizza(2004):127]

Erste Ansätze einer international ausgerichteten Normenkontrolle sind bereits in dem schon erwähnten Statement of Principles [PP(1961)] zu finden. Obwohl explizit auf die Auswahl und Ordnung von Titelaufnahmen im alphabetischen Katalog ausgelegt, wurde mit der Ansetzung von Titeln sowie Autoren- und Körperschaftsnamen in einer national festzulegenden, präferierten Form (uniform headings) auch eine implizite Normenkontrolle eingeführt. Die Paris Principles nennen als Kriterien, nach denen die Ansetzungsform gewählt werden soll: Originalausgabe, aktuell gebräuchliche Bezeichnung und sprachliche Provenienz. Für Personennamen schreiben sie das Staatsbürgerprinzip vor, sodass im Regelfall nur ein nationalbibliografisches Zentrum (national bibliographic agency) die Entscheidung über die Ansetzung innehat, diese Ansetzungsform begründet auch die international zu verwendende Form. Andere Titel- und Namensformen, also abweichende Sucheinstiege, werden als headings von Nebeneintragungen (added entries) oder Verweisungen (references) in den Katalog aufgenommen. Ohne das hier zu Grunde liegende *Prinzip der bibliografischen Identität* [Buizza(2004):118-119] von einzelnen Werken und Urhebern abseits ihrer variierenden Bezeichnungen, ist eine *Ansetzungs- bzw. Normierungspraxis* und damit auch das *Kollokationsprinzip* nicht denkbar.

Mit der Festlegung von Ansetzungskriterien und Hoheit der nationalbibliografischen Zentren 1961 sowie den daraufhin von der IFLA herausgegebenen Ansetzungslisten für Namen von Staaten, staatlichen und internationalen Körperschaften, klassischen griechischen und römi-

schen Verfassern und wichtigen anonymen Werken²⁸ sollte ein Grundstein zu internationaler Austauschbarkeit gelegt werden. Ein recht großer Interpretationsspielraum der Ansetzungskriterien und auch vorsätzliche Abweichungen davon in den nationalen Katalogisierungswerken schmälerten den Erfolg aber erheblich.

Das zentrale Manko der Paris Principles lag aber in der ausschließlichen Konzentration auf die Katalogeinträge – Normenkontrolle bleibt an diese gebunden und erfolgt nur als Nebeneffekt der als Ordnungselemente benötigten „Eintrags-Überschriften“. Diese Vernachlässigung der Normdaten gegenüber den bibliografischen Daten stellt im Vergleich zur historischen Praxis der Library of Congress einen Rückschritt dar. Wie Pino Buizza anmerkt, wurden dort seit 1899 „Personen-Aufnahmen“ (authority cards) für alle Verfasser geführt, mit Ansetzungsform, vollständigem Namen, abweichenden Namensvarianten, Quellangaben und Verweisungen zu den Titeln, bei denen diese Namensformen vorliegen. [Buizza(2004):121] „Normdatenaufnahmen“ sollten erst ihre Verbreitung finden, nachdem Kataloge in elektronische Datenbanken überführt worden waren. Zudem war durch wachsenden internationalen wissenschaftlichen und gesellschaftlichen Austausch das Problem von Namen in fremden Sprachen und Schriften und damit die Unzulänglichkeit der gängigen Normenkontrollpraxis offenbar geworden. Durch die Tabellenstruktur von Katalog-Datenbanken wurden automatisch Ansetzungsformen von Titelaufnahmen getrennt: ein heading als Ordnungselement wurde überflüssig, eine eigene Datentabelle mit Ansetzungs- und abweichenden Namensformen dagegen nötig, um mit diesen verknüpfte Titelaufnahmen zusammenzuführen. Mit eigenen Normdatensätzen (die weitere Informationen enthielten, vergleichbar zu denen, welche die Library of Congress in ihren authority cards verzeichnete) wurde damit die Normenkontrolle unabhängig von der Katalogisierung von bibliografischen Beschreibungen: als Vorreiter in den 1970er Jahren wird hierfür die New York Public Library genannt. [Buizza(2004):122-123]

Ähnlich wie bei der Fremddatenübernahme von Titelaufnahmen wurden in der Nachfolge Projekte zur Zusammenführung der Normdaten auf nationaler Ebene forciert, wie die Gemeinsame Körperschaftsdatei (GKD, 1981) und das amerikanische Name Authority Cooperative Project (NACO²⁹, 1977). Und schließlich verlangte auch die Normenkontrolle nach einem der ISBD vergleichbaren Regelwerk, das Form und grundlegende Inhalte der Normdatensätze in einer international verbindlichen Form beschreibt. [Buizza(2004):123] 1984 erschienen daher die Guidelines for Authority and Reference Entries (GARE) der Working Group on an International Authority System im Rahmen des UBC-Programms der IFLA. Die

28 Eine genaue Aufführung dieser Ansetzungslisten mit allen Ausgaben ist ebenfalls im Aufsatz von Pino Buizza zu finden, siehe [Buizza(2004):121].

29 NACO, siehe <http://www.loc.gov/catdir/pcc/naco/naco.html>

GARE entwarfen Normdatensätze in die Kategorien authority entry (Normdateneintragung oder Normdatensatz), reference entry (Verweisung) und general explanatory entry (Pauschalverweisung), wobei eine Normdateneintragung aus folgenden Komponenten bestehen sollte: Ansetzungsform, Informationen zur Namensform, Hinweise auf Verweisungen auf den Eintrag, Quellangaben und (eine prospektive) International Standard Authority Data Number (ISADN) zur Identifizierung jeder genormten Entität. Ein zu erwähnender Nachteil neben der thematischen Begrenzung auf Personennamen, Körperschaftsnamen sowie Titeln von klassischen Anonyma ist der Ausschluss von externen Daten und Informationen, die gerade in Bezug auf die angestrebte Individualisierung von Personen und Körperschaften sinnvoll sind. [Buizza(2004):123-124]

1998 wurde dann eine tiefgreifendere Analyse zur vereinheitlichten Erstellung von Normdaten unter Leitung der IFLA UBCIM Working Group on Minimal Level Authority Records and ISADN (MLAR) unternommen. Mit der Erfahrung, dass einige bi- und multilateral ausgerichtete Studien und Projekte³⁰ mit dem Ziel von einheitlichen Ansetzungsformen nicht sehr erfolgreich verliefen, vollzog die MLAR mit der Studie Mandatory Data Elements for Internationally Shared Resource Authority Records [MLARreport(1998)] ein Paradigmenwechsel. Vor dem Hintergrund der Möglichkeiten, die elektronische Datenverarbeitung und das internationale bibliothekarische Austauschprotokoll Z39.50³¹ boten, war die Beschränkung auf eine international gültige Ansetzungsform (also oftmals die Übernahme der originalsprachlichen Namen bzw. ihrer Transliterationen) nicht mehr unbedingt nötig: ohne die nationalbibliografische Hoheit zu verletzen war eine gleichberechtigte Verknüpfung der Ansetzungsformen in verschiedenen Ländern und Sprachräumen denkbar geworden. Der entscheidende Punkt ist aber der der Nutzerfreundlichkeit:

There are reasons to use the form of names familiar to our own users, in scripts they can read and in forms they most likely would look for in their library catalogue or national bibliography. Therefore, this Working Group recognizes the importance of allowing the preservation of national or rulebased differences in authorized forms for readings to be used in national bibliographies and library catalogues that best meet the language and cultural needs of the particular institution's users. [MLARreport(1998):Introduction and Recommendations]

Der Umsetzungsvorschlag der MLAR Working Group bestand in einer elektronischen Sammlung aller nationalen Normdaten unter dem Dach der IFLA, sodass Verknüpfungen der verschiedenen nationalen Datensätze einer Entität möglich werden, die Vision einer „virtual shared resource authority file“. Vor diesem Hintergrund erhielt auch das Konzept der ISADN, das schon seit den 1970er Jahren in Diskussion war³², einen neuen Sinn: als Primärschlüssel

30 Als Beispiele werden die Anglo-American-Authority File (AAAF) und das europäische AUTHOR-Projekt genannt [Buizza(2004):125].

31 ANSI/NISO Z39.50 bzw. ISO 23950, siehe <http://www.loc.gov/z3950/agency/>

32 Neben der ISADN gab es Vorschläge für eine International Standard Authority Number (ISAN) sowie den ISO International Standard Text Code (ISTC), siehe [Tillett(2004):30].

für alle nationalen Normdatensätze wäre sie der eindeutige Identifikator einer genormten Entität. Trotz der grundsätzlichen Befürwortung einer ISADN plädierte man in Anbetracht des Aufwands, der rasanten technologischen Entwicklung und der absehbaren Zusammenarbeit mit verwandten Bereichen vorerst auf weitere Untersuchungen zum Thema. Das konzeptionelle Umdenken im Bereich der internationalen Normenkontrolle lässt sich auch gut am Glossar der „Mandatory Data Elements“ ablesen: so werden jetzt bibliografische Begriffe genauer analysiert und wo es nötig wird, revidiert bzw. neuere, passendere Bezeichnungen neben die Traditionellen gestellt. Als Beispiel sei das bereits erwähnte Konzept der Bibliografischen Identität³³ angeführt, das von dem der bibliografischen Entität unterschieden wird:

A single entity (e.g., person, corporate body, or work) can have more than one bibliographic identity, for example: real name and pseudonym(s); current and former name(s); conventional name or uniform title for a work. [MLARreport(1998):Appendix A, Glossary]

Generell insistierte der Report auf eine Überarbeitung der GARE und des darauf abgestimmten UNIMARC/Authorities-Formats – was 2001 schließlich zur Veröffentlichung der zweiten Ausgabe der GARE, der Guidelines for Authority Records and References (GARR) [GARR(2001)] führte. Parallel dazu wurde das Datenformat UNIMARC/Authorities als UNIMARC Manual: Authorities Format³⁴ neu konzipiert. [Buizza(2004):125]

Die noch heute gültigen GARR folgen grundlegend den genannten Empfehlungen der „Mandatory Data Elements“-Studie und stützen sich auch auf deren Terminologie sowie die des FRBR-Reports. Als umfassende Neuerung schließen sie zudem alle Publikationsarten in allen Medienformen ein. [GARR(2001):vi-vii] Analog zu den ISBDs im Bereich der Bibliografischen Beschreibung besteht ihre zentrale Aufgabe darin,

...to facilitate the international exchange and use of information on authorised headings and references. To that end, the *Guidelines* define a set of elements to be included in the entry, assign an order to the elements, and specify a system of punctuation for the entry. The specifications, however, are confined to the overall structure and major functional components of the entry. [GARR(2001):1, Hervorhebung original]

Sie führen also die allgemeine Definition und Strukturierung von Normdatensätzen, Verweisungen und Pauschalverweisungen der GARE weiter und begrenzen sich auch weiterhin auf die Normierung von Personen- und Körperschaftsnamen, sowie Einheitssachtiteln. Gemäß der FRBR-Präzisierung beziehen sich Letztere jetzt aber auch auf Expressionen bzw. Werke („work/expression“). Ganz klar wird aber die Ausgestaltung dieser Richtlinien den nationalen Regelwerken überlassen: so benennen die GARR zwar beispielhaft Kategorien von Na-

33 Darauf, dass das Konzept der Bibliografische Identität bereits in den Paris Principles angelegt gewesen ist, weist eine Fußnote in den GARR hin, vgl. [GARR(2001):18].

34 Diese zweite, revidierte Auflage ist online zugänglich, siehe <http://www.ifla.org/VI/8/projects/UNIMARC-AuthoritiesFormat.pdf>. Seit 2009 gibt es zudem eine dritte, aktualisierte Auflage.

mensformen, Namensbestandteilen und Qualifikatoren, klammern deren konkrete Auswahl und Form aber explizit aus. [GARR(2001):11]

Prinzipiell scheinen die sprachlichen und kulturellen Traditionen von Eigennamen zu verschieden zu sein, als dass diesbezüglich formelle Vorgaben generalisiert werden können. Inwiefern die im Statement on International Cataloguing Principles (ICP) geäußerte Forderung einer weitestmöglichen, internationalen Standardisierung von Sucheinstiegen überhaupt gegenüber Eigennamen durchgesetzt werden kann, erscheint da fraglich.

2.3.1 Functional Requirements for Authority Data (FRAD)

Mit den GARR-Bestimmungen von 2001 hatte man mit allgemeinen „Normdatenbeschreibungen“ den Schlußschluss zur ISBD-Konzeption in der bibliografischen Beschreibung vollzogen, blieb sich aber bewusst, dass vor dem Hintergrund der FRBR-Konzeption eine grundlegendere Beschäftigung mit der Normenkontrolle nötig sein würde. Bereits 1999 war die Working Group on Functional Requirements and Numbering of Authority Records (FRANAR) von der IFLA Division of Bibliographic Control und dem IFLA Universal Bibliographic Control and International MARC (UBCIM) Programme³⁵ gegründet worden, um das FRBR-Rahmenwerk für den Bereich der Normenkontrolle bezüglich Eigennamen auszubauen. Mittlerweile liegt ein Entwurf (Draft 2007-04-01) für das Functional Requirements for Authority Data: A Conceptual Model (FRAD) vor, dessen Ziel es ist,

...[to] provide a clearly defined, structured frame of reference for relating the data that are recorded in authority records to the needs of the users of those records; [and to] assist in an assessment of the potential for international sharing and use for authority data both within the library sector and beyond. [FRADdraft(2007):1]

Es fällt auf, dass der Begriff *authority data* (Normdaten) hier nicht klar definiert, sondern lediglich umschrieben wird, und zwar als „das, was in Normdatensätzen (authority records) gespeichert wird“. Dies macht dann eine Klärung des Begriffs *authority record* nötig, welche wie folgt definiert wird:

...an aggregate of information about an instance of an entity whose name is used as a controlled access point for bibliographic citations or records in a library catalogue or bibliographic file. [FRADdraft(2007):1]

Dieser Rückbezug auf eher traditionelle Terminologie, auch im Weiteren, wird z.B. von der DNB kritisiert, da das FRAD eigentlich eine formunabhängige Beschreibung des Normdatenbereichs anstreben sollte und das terminologische Zusammenspiel mit FRBR und ICP unklar wird. [DNBfrad(2007):1]

³⁵ Nach dem Auslaufen des UBCIM 2003 fällt die Entwicklung der FRAD ebenfalls unter die Ägide der IFLA-CDNL Alliance for Bibliographic Standards (ICABS), sowie der British Library. [FRADdraft(2007):iii]

Allgemein unterscheidet das FRAD-Modell Normdaten für Namen (personal, family³⁶, corporate, and geographic names), Titel (uniform titles and collective uniform titles), Kombinationen von Namen und Titeln (name-title authority data), für die sachliche Erschließung (subject authority data) und für Identifikatoren (authority data for various types of identifiers). Explizit entwirft es aber nur ein ER-Modell für die Ansetzung von Personen-, Familien- und Körperschaftsnamen, sowie Titel- und Titel-Namen-Ansetzungen für Werke (und damit auch für die entsprechenden Expressionen, Manifestationen und Exemplare). [FRADdraft(2007):2] Außerdem werden in Erweiterung zum Ansetzungsprinzip Identifikatoren als separate Entitäten³⁷ in das Modell eingeführt.

Insbesondere analysiert der FRAD-Entwurf die Beziehungen, die zwischen den bibliografischen Entitäten Person, Familie, Körperschaft und Werk bestehen, die mit ihren Eigennamen eben Gegenstand der Normenkontrolle bilden. Zum einen werden dabei jeweils die Beziehungen zwischen Personen, zwischen Familien und zwischen Körperschaften; zum anderen die Beziehungen zwischen Personen und Familien sowie Körperschaften betrachtet. [FRADdraft(2007):31-38] Danach werden erneut die Beziehungen zwischen Werken betrachtet – zum Teil überschneidend mit den Ergebnissen der FRBR-Studie, zum Teil ergänzend dazu. [FRADdraft(2007):38-42]

Zudem werden die Beziehungen der bibliografischen Entitäten Person, Familie, Körperschaft und Werk zur neuen Entität Name, sowie die Beziehungen zwischen Normierten Sucheinstiegen analysiert. [FRADdraft(2007):42-47]

Grundsätzlich erweitert das FRAD-Modell die FRBR-Struktur dahingehend, dass alle *bibliografischen* Entitäten³⁸ mit den neuen Entitäten *Name*³⁹ (name) und *Identifikator* (identifier) verbunden werden, welche die Basis zur Festlegung von *Normierten Sucheinstiegen* (controlled access points) – ebenfalls eine neue Entität – bilden. Dies bietet denn auch die Grundlage für die Forderung des Statement on International Cataloguing Principles (ICP) nach Normierung *aller* bibliografischen Entitäten (Punkt 6.1.1.). [ICPde(2009):5] Ferner werden für die in den FRBR beschriebenen Gruppe-1- und Gruppe-2-Entitäten einige neue Attribute benannt und die Entitäten *rules*⁴⁰ und *agency*⁴¹ – die für Wahl der Normierten Sucheinstiege

36 Die Entitätengruppe 1 der FRBR wird im FRAD-Modell um die Entität Familie (family) erweitert, sicher auch im Hinblick auf Datenaustausch mit dem Archiv-Bereich.

37 Identifikatoren tauchten schon in den klassischen bibliografischen Beschreibungen und Normdatensätzen auf, sind dort jedoch direkt an diese gebunden, also eher ein Attribut, denn ein eigenständiges Konzept.

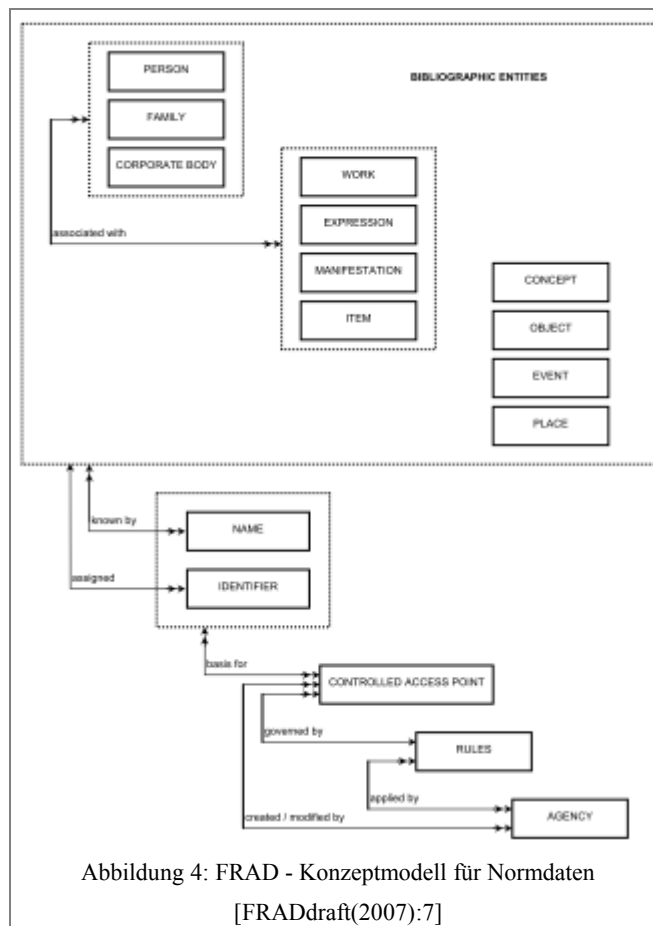
38 Bibliografische Entitäten nach FRBR/FRAD: Person, Familie, Körperschaft, Werk, Expression, Manifestation, Exemplar, Begriff, Gegenstand, Ereignis, Ort.

39 Soweit durch autorisierte Übersetzungen wie das ICPde (das sich auf FRAD-Terminologie stützt) belegt, verwende ich die deutschen Begriffe, ansonsten die Englischen. Auf die Problematik der Übersetzung von „name“ weisen die Autoren des ICPde hin. [ICPde(2009):2, Fußnote 6]

40 Rules: „A set of instructions relating to the formulation and/or recording of controlled access points (authorized or preferred forms, references, etc.).“ [FRADdraft(2007):15]

41 Agency: „An organization responsible for creating or modifying a controlled access point. The agency is responsible for application and interpretation of the rules it uses.“ [FRADdraft(2007):15]

verantwortlich sind – werden in das Modell eingeführt. Zur Veranschaulichung der Grundstruktur, die das FRAD-Modell enwirft, soll das überblicksartige Diagramm hier übernommen werden.



Die Entität Name wird in den FRAD ganz allgemein definiert als

A character or group of words and/or characters by which an entity is known. [FRADdraft(2007):12],

die Entität Identifikator als

A number, code, word, phrase, logo, device, etc. that is uniquely associated with an entity, and serves to differentiate that entity from other entities within the domain in which the identifier is assigned. [FRADdraft(2007):13]

Interessant ist die neue Modellierung des Namens als eine diskrete Entität, die in einer n:n-Beziehung zu den bibliografischen Entitäten steht – während name bzw. title in FRBR noch als Attribut dieser behandelt wird. Robert L. Maxwell bemerkt dazu kritisch:

It is not clear how practical considering name as a separate entity rather than as an attribute would be designing a bibliographic database...[Maxwell(2008):56]

Des Weiteren weist die DNB in ihrem Kommentar zum FRAD-Entwurf auf die ungenaue Spezifikation der Beziehung zwischen Sucheinstiegen und Namen bzw. Identifikatoren hin:

...the identifier always uniquely refers to one and only one specific instance of an entity whilst names may refer to multiple instances [...]. Accordingly access points based on names need to be specified by additions based on attributes or citations of related entities (type of relationship 'is based on'), whilst identifiers do not need any complements (type of relationship 'is used as'). [DNBfrad(2007):2]

Diese additions, also „identifizierende Merkmale“⁴², durch die Namen oftmals erst zu Normierten Sucheinstiegen werden können, sollten stärker im Modell berücksichtigt werden und nicht nur als Attribut des Sucheinstiegs behandelt werden. Die Gleichstellung von Namen und Identifikatoren durch die „basis-for“-Relation zu Sucheinstiegen führt nämlich auch zur Gleichbehandlung von Normierten Sucheinstiegen, die einerseits auf Ansetzungsformen oder Identifikatoren basieren und andererseits auf alternativen Namensformen⁴³. Die Definition des Normierten Sucheinstiegs (controlled access point), wie es das FRAD anbietet, mutet denn auch weniger systematisch begründet an, als es dem grundsätzlichen Ziel der FRAD entspräche:

A name, term, code, etc. under which a bibliographic or authority record or reference will be found. [FRADdraft(2007):14]

Dies stellt ein Rückbezug auf ältere Terminologie dar, der der konzeptionellen Stellung des Normierten Sucheinstiegs im FRAD-Modell schlichtweg nicht gerecht wird: der Bezug zu den Entitäten Name und Identifikator bleibt hier nicht nur un-unterschieden, sondern ganz außen vor.

Mit der Klärung der Beziehungen zwischen mehreren Namen bzw. Normierten Sucheinstiegen einer bibliografischen Entität liefert der FRAD-Entwurf allerdings die Grundlage für eine angestrebte Parallelität von Sucheinstiegen in verschiedenen Sprachen, Schriften und gemäß unterschiedlicher rules und agencies. Spezifiziert werden zudem die Beziehung zwischen Sucheinstiegen, die auf Eigennamen beruhen, und solche, die dieselbe Entität in einem Kontrollierten Vokabular repräsentieren sowie zwischen namensbasierten Sucheinstiegen und identifikatorbasierten Sucheinstiegen. [FRADdraft(2007):47-49] Eine prinzipielle Neuausrichtung der Normenkontrolle muss insgesamt darin gesehen werden, dass mehrere gleichgewichtete Sucheinstiege für eine Entität einer zentralen Ansetzungsform bzw. eines Normdatensatzes vorgezogen werden. Die Festlegung einer Ansetzungsform (preferred name) wird prospektiv wohl eher nebensächlich werden. Auch aus diesem Grund plädiert man im englischsprachigen Bereich für eine Umbenennung der authority control in access point control, also statt der Normen-Kontrolle eine Kontrolle der Sucheinstiege.

42 So genannt in der deutschsprachigen Version der ICP, siehe [ICPde(2009):9].

43 Die diesbezügliche Unterscheidung der Normierten Sucheinstiege, ob der Sucheinstieg auf der Ansetzungsform oder einer alternativen Namensform beruht, geschieht dann nur durch deren Attribut „Designated usage of controlled access point“, vgl. [FRADdraft(2007):26].

Der FRAD-Entwurf schlägt zudem das umfassende Konzept einer nationalbibliografischen Definition von Normierten Sucheinstiegen für *alle* bibliografische Entitäten vor (auch wenn er explizit nur die Objekte der „engeren“ Normenkontrolle: Personen, Familien, Körperschaften und Werke, behandelt). Dies stellt eine enorme Forderung dar, besonders für den Bereich der Gruppe-3-Entitäten.

Dass der Bereich der Normenkontrolle gegenüber den anderen Bereichen der universellen Bibliografischen Kontrolle zurückliegt, rührt sicher einerseits daher, dass reale und metaphysische Entitäten „einer gewissen Einzigartigkeit“ per se stark durch sprachliche und kulturelle Muster geprägt sind und dies durch ihre Eigennamen auch ausdrücken. Zudem musste die Eigenständigkeit der Normenkontrolle erst mühsam erarbeitet werden und reicht eben noch nicht lange zurück. Es wurde aber auch von bibliothekarischer Seite schon frühzeitig gegen die internationalen Übereinkünfte verstoßen, und davon waren insbesondere die Personen- und Körperschaftsansetzungen betroffen. [Buizza(2004):121-122] Deutschland gab hier mit der minimalen Anerkennung der korporativen Urheberschaft (und damit natürlich verbunden, der Körperschaftsansetzungen) ein unrühmliches Beispiel. So bereiten Forderungen, die bereits in den Paris Principles gestellt wurden, bis heute Probleme oder zumindest großen Arbeitsaufwand, hier wäre z.B. in Deutschland die umfassende Individualisierung von Personen- und Körperschaftsansetzungen, die erst seit 2006 bei der DNB realisiert wird, zu nennen. Ebenfalls lange hat es gedauert, dass Normdaten wie bibliografische Daten selbstverständlich für Katalognutzer⁴⁴ einsehbar sind.

Nicht nur aus Sicht einer universellen Bibliografischen Kontrolle ist das zu bedauern, vor dem Hintergrund, dass auch andere Institutionen und Organisationen aus dem BAM-Bereich⁴⁵ und dem Publikations- und Rechtemanagementssektor⁴⁶, dieselben Informationen benutzen, wären konsistente, verknüpfte Normdaten im höchsten Maße wünschenswert. Prinzipiell sollten einzigartige (reale und abstrakte) Entitäten wie Personen, Körperschaften und Expressionen doch leichter konsensfähig zu beschreiben und zu identifizieren sein als sachliche Konzepte.

44 So sind die LCNAF seit 2004 öffentlich recherchierbar, im deutschen Raum ist hier beispielhaft der Online-Katalog des Gemeinsamen Bibliotheks-Verbands (GBV) zu nennen.

45 Zum Beispiel wäre zur Zusammenarbeit mit Archiven eine Konkordanz zum XML-basierten Format Encoded Archival Context (EAC), vgl. <http://eac.staatsbibliothek-berlin.de/> wünschenswert.

46 Im Bereich Rechtemanagement/E-Commerce sind hier die <indecs>-Initiative (Interoperability of Data in E-Commerce Systems) und das InterParty-Projekt zu nennen; das LEAF-Projekt dagegen soll Normdaten bibliothekarischer Provenienz für Archive verfügbar machen. [Tillett(2004):31]

2.4 Kontrollierte Vokabulare (KOS)

Wie bereits erwähnt, werden Kontrollierte Vokabulare ebenfalls zur Festlegung von Sucheinstiegen im Rahmen der Normenkontrolle eingesetzt. Hier macht man sich die Normierung von Bezeichnungen zu Nutze, die Kontrollierte Vokabulare für Begriffe nahezu aller Fachgebiete bieten. Die Unterscheidung von Begriff (Konzept) und Benennung ist für diesen Bereich von zentraler Bedeutung. Wie in den DIN-Normen 2330 und 2342/1 formuliert, ist ein Begriff

...eine Denkeinheit (d.h. ein gedankliches Konstrukt, K.W.), die aus einer Menge von Gegenständen unter Ermittlung der diesen Gegenständen gemeinsamen Eigenschaften mittels Abstraktion gebildet wird. [Stock/Stock(2008):52]

Man spricht von Kontrollierten Vokabularen⁴⁷, da die verwendeten Bezeichnungen mittels einer *Terminologischen Kontrolle* von den Ambiguitäten der natürlichen Sprache (Synonyme, Homonyme und Polyseme) befreit und ein-eindeutig auf Begriffe bezogen wurden. Im Z39.19 Standard der NISO bzw. ANSI (Guidelines for the construction, format, and management of monolingual controlled vocabularies) wird dies so beschrieben:

A major goal of vocabulary control is to ensure that each distinct concept is described by a single linguistic form. If multiple forms exist, these should be controlled or regularized so that information or content that is provided to a user is not spread across the system under multiple access points, but is gathered together in one place. Eliminating ambiguity and compensating for synonymy through vocabulary control assures that each term has only one meaning and that only one term can be used to represent a given concept or entity. [ANSI/NISOZ39.19(2005):12]

Knowledge Organisation Systems (KOS) ist dagegen eine recht junge, übergreifende Bezeichnung für Thesauri, Schlagwort-, Klassifikations- und ähnliche Systeme, die inhaltliche Definitionen von Begriffen durch die Einordnung in ein Begriffssystem bieten und so zur (kontrollierten) sachlichen Erschließung von Objekten dienen. Diese englische Bezeichnung als „Wissensorganisationssysteme“⁴⁸ bezieht sich weniger auf die sprachnormierende Funktion denn auf die Modellierung von Wissensstrukturen. Dass diese Beschreibung nicht so übertrieben ist, wie sie auf den ersten Blick scheinen mag, erschließt sich, wenn man den Wissensbegriff betrachtet. Diesem liegt, wie Jiri Panyr bemerkt, die *ontologische Annahme* zu Grunde, dass

...sich alles für ein intelligentes Verhalten Wesentliche (d.h. im Prinzip das gesamte relevante Wissen der uns umgebenden Welt) in einer Menge festgelegter, unabhängiger Elemente (Begriffe, Fakten, Gesetze) zerlegen lässt (vgl. auch Dreyfus (1985:154ff) [Panyr(2006):139-140]

Wissen kann man so umgekehrt auch als Zusammenhang von einzelnen vernetzten Informationen begreifen, es entsteht eben

47 Wolfgang G Stock und Mechthild Stock sprechen dagegen von Wissensordnungen, vgl. [Stock/Stock(2008):39-40]. Die geläufige alternative Bezeichnung Dokumentationssprache soll hier ebenfalls erwähnt werden; da die Verwendung der kontrollierten Vokabulare aber längst nicht mehr nur Sache der wissenschaftlichen Information und Dokumentation (IuD) ist, ziehe ich allgemeinere Bezeichnungen vor. Ich werde in dieser Arbeit die Benennungen knowledge organisation systems (KOS), Begriffssysteme und Kontrollierte Vokabulare gleichbedeutend benutzen.

48 Barbara Geyer-Hayden benutzt z.B. diese wörtliche Übersetzung von KOS. [Geyer-Hayden(2009)]

...dadurch, dass Informationen in einen größeren kognitiven Kontext gestellt werden [...]. Wissen als Netzwerk von aus aktuellen Informationen gewonnenen Wissensstücken, die in größere Wissensobjekte relational eingebunden werden, ist aber [...] ein kohärentes Geflecht von Wissenseinheiten mit bestimmten semantischen Merkmalen... [Kuhlen(2004):160]

Dass Rainer Kuhlen Wissen als „kohärentes Geflecht [...] mit bestimmten semantischen Merkmalen“ umschreibt, verweist auf die bedeutungstragenden Beziehungen zwischen den Informationseinheiten. Angelehnt an diese Vorstellung versucht man das Wissen eines Bereiches, also Informationselemente und ihre Relationen, mit KOS modellhaft zu repräsentieren:

...systems, that organize ontological and lexical information (information about concepts and terms and their relationships), such systems are called *Knowledge Organization Systems (KOS)*. [Soergel(2009):27, Hervorhebung original]

Im umfangreichsten Fall vereinen KOS also die Beziehungen zwischen Begriff und Benennungen (lexical information) mit der Darstellung der semantischen Struktur zwischen den Begriffen (ontological information). Doch sind auch Systeme möglich, die sich auf ein der beiden Beziehungsarten beschränken: Schlagwortlisten einerseits und die meisten Ontologien andererseits.

Schlagwortlisten (bzw. Begriffslisten, [Geyer-Hayden(2009):130]) verfügen über die einfachste Struktur der kontrollierten Vokabulare, sie werden aber eher selten verwendet. Sie beschränken sich auf die Forderung der Terminologischen Kontrolle: der Definition einer Vorzugsbenennung für jeden Begriff, und der Zusammenfassung der übrigen Benennungen für diesen Begriff in einer *Äquivalenzklasse*, die auf die Vorzugsbenennung verweist. Stella G. Dextre Clarke gibt hierzu eine gute Übersicht über Kategorien von synonymen Benennungen, z.B. lexikalische Varianten, Invertierungen bei Mehr-Wort-Benennungen, Bezeichnungen unterschiedlicher Herkunft sowie zu Quasisynonymen, die kontextabhängig nahezu bedeutungsgleich sein können, wie antonyme Eigenschaften und spezifische Begriffe, die unter einem Allgemeineren subsumiert werden sollen. Zu erwähnen auch ist die Möglichkeit einer partiellen Äquivalenzbeziehung, also der Darstellung einer mehrgliedrigen Relation, etwa eines Begriffes (in Form eines Synonymes) durch zwei Vorzugsbenennungen. [Dextre Clarke(2001):40-41]

Schlagwortlisten verzichten somit auf Relationen zwischen den Begriffen (die in der Form der Vorzugsbenennungen vorliegen). Begriffs-Benennungsbeziehungen (concept-term relationships) sind aber auch struktureller Bestandteil vieler KOS, die dezidiert auf die Darstellung von Begriffsrelationen (conceptual structure [Soergel(2009):31]) abheben. Sie unterstützen hier die Auswahl und Identifikation von Konzepten bei Indexierung und Retrieval. Die *Äquivalenzrelation* dient also der Zuordnung von verschiedenen Bezeichnungen zu einer Vorzugsbenennung, die das Konzept repräsentiert.

Andererseits kann die Gleichordnung (Koordination) von Konzepten der *Bildung von Klassen* (also Gruppen) dienen. Diese spiegeln die Ähnlichkeit der Begriffe bezüglich ein oder

mehrerer Merkmale wieder – eine relative Gleichartigkeit also. Chowdhury/Chowdhury sehen darin das Grundprinzip des Klassifizierens:

Thus we may say that classification is the process of arranging things in order of their degrees of likeness, and therefore by analogy separating things according to their degrees of unlikeness. [Chowdhury/Chowdhury(2007):6]

Auch Jiri Panyr weist darauf hin, dass eben nicht Ober- und Unterordnung das erste Prinzip einer Klassifikation bzw. Taxonomie ist, sondern die Bildung von Klassen der Ähnlichkeit. Somit ist

...[d]ie Grundlage der numerischen Taxonomie (Clusteranalyse, automatischen Klassifikation) [...] die Auswahl von geeigneten Merkmalen. Ihre Ausprägungen charakterisieren die zu klassifizierenden Objekte. [Panyr(2006):141]

Erst dadurch können Beziehungen zwischen diesen gruppierten Konzepten abgeleitet werden: gemeinsame und unterschiedliche Merkmale begründen dann die Ober- und Unterordnung von Klassen. Durch die weitere Unterscheidung der Konzepte einer Klasse nach mindestens einem bestimmten Gesichtspunkt werden Unterklassen gebildet. Dadurch wird eine hierarchische Ordnung von allgemeinen Konzepten zu immer spezifischeren möglich, die *generische Relation* (Abstraktionsrelation, Hyponomie) genannt wird. Biologische Taxonomien sind hierfür ein gutes Beispiel – auch der Name der Beziehung ist von der Genus-Spezies-Struktur abgeleitet.

Eine andere Möglichkeit, eine hierarchische Ordnung zu etablieren ist die *partitive Relation* (Bestandsrelation, Meronymie), welche die Beziehung von Teilen zu einem Ganzen repräsentieren. Sie wird sie oft zur Unterscheidung von geografischen, anatomischen und sozialen Strukturen benutzt sowie zur Gliederung von wissenschaftlichen Disziplinen.

Eine dritte hierarchische Relationsart ist die *Instanzrelation*. Hier werden – im Regelfall auf der untersten Hierarchieebene – konkrete Instanzen (also individuelle Entitäten) einem Konzept untergeordnet, etwa dem Begriff Faltengebirge die Instanzen Alpen, Skanden, Kaukasus usw. [Dextre Clarke(2001):42-43]

Das Zusammenspiel von Klassenbildung und hierarchischen Relationen wird besonders in *Klassifikationssystemen*, aber auch in *Schlagwortsystemen* mit hierarchischer Struktur genutzt. Die grundsätzliche Unterscheidung zwischen diesen ist nicht etwa die künstlich sprachliche und natürlich sprachliche Basis (diese werden auch oft parallel verwendet). Entscheidend ist eher, dass in hierarchischen Klassifikationen eine einzige, vorgegebene Systemstelle, also ein Konzept zur Beschreibung der ganzen Objektes (also der Indexierung) dient, während mit Schlagwortsystemen zumeist eine synthetische Beschreibung mittels mehrerer Konzepte angestrebt wird. Diese gegensätzlichen Ansätze werden mit Präkoordinations-

tion und Postkoordination umschrieben. Per definitionem verfügen Klassifikationen⁴⁹ aber meistens über eine tiefere Ebenenstruktur während die Hierarchie bei Schlagwortsystemen⁵⁰ eher flach angelegt ist.

Zu diesen Ansätzen der Wissensorganisation kann das *Facettierungsprinzip* treten, das auf die Colon Classification (CC) von Shiyali R. Ranganathan zurück geht. Diese *Facettenklassifikation* ist eben nicht dazu geschaffen, alle Informationselemente, also Konzepte in einem System aus Ober-/Unter- und Gleichordnung anzuordnen (also präkoordinativ vorzugeben), sondern Informationselemente bereitzustellen, aus denen Konzepte zur Indexierung postkoordinativ synthetisiert werden können. [Stock/Stock(2008):273-277] Aus dieser, der *hierarchischen Klassifikation* entgegen gesetzten, Vorgehensweise ergibt sich automatisch eine „mehrdimensionalere“ Beschreibung von Begriffen. Reine facettierte System sind selten, oft werden Facetten aber als Erweiterung von hierarchischen Systemen eingesetzt. Zum einen als grobe Unterteilung von sehr umfangreichen kontrollierten Vokabularen in einzelne Teilkategorien wie etwa geografische, Personen- und andere Sachschlagwörter. Hier dienen sie zu entweder einfach zu Überblickszwecken, oder aber geben ein Schema vor, nach dem die Schlagwörter oder Notationen synthetisiert werden.

Zum anderen geht es um die Spezifikation der einzelnen Konzepte mittels Facetten, welche die eindimensionale Einordnung in ein hierarchisches System erweitern. [ANSI/NISO Z39.19(2005):27] So verfügen beispielsweise DDC und UDC über allgemeine und spezielle Anhängennummern, die eine weitere Beschreibung zusätzlich zur Systemstelle der Klassifikation ermöglichen.

2.4.1 Thesauri

Assoziationsrelationen als letzte große Gruppe von Begriffsrelationen werden meistens schlicht so beschrieben, dass sie den anderen beiden Gruppen nicht entsprechen. Ziel ist es, Verbindungen zwischen Konzepten herzustellen, die eine Vielzahl von Beziehungen sein können: Temporale, kausale, finale, antagonistische, komplementäre Beziehungen; Beziehungen zwischen einem Vorgang und seinen Akteuren, Methoden, Mitteln oder Ergebnissen usw. Diese sind

...zunächst einzig von besonderen, höchst zufälligen Bedingungen des Wissensstandes, der Kontexte und Kontexte, der situativen Umgebung, etc. gesteuert, d.h. weitgehend irregulär und

49 Beispiele für Universalklassifikationen sind die Dewey Decimal Classification (DDC), siehe <http://www.oclc.org/dewey/> bzw. <http://www.ddc-deutsch.de/> und die davon abgeleitete Universal Decimal Classification (UDC), siehe <http://www.udcc.org/about.htm>

50 Beispiele für universale hierarchische Schlagwortsysteme sind etwa die LCSH, siehe <http://www.loc.gov/cds/lcsh.html> und die SWD, siehe <http://www.d-nb.de/standardisierung/normdateien/swd.htm> Ein bekanntes disziplinäres Schlagwortsystem sind z.B. die Medical Subject Headings (MeSH), <http://www.nlm.nih.gov/mesh/meshhome.html>

deshalb nur sehr bedingt formal reproduzierbar. [...] [A]ssoziative Bedeutungsbeziehungen sind nicht nur zufällig, sondern überaus systembildend und für die primäre Strukturierung von Erfahrungswissen von Welt konstitutiv. [Panyr(2006):141]

Kontrollierte Vokabulare, die solche Assoziationsrelationen (neben Äquivalenz- und hierarchischen Beziehungen) darstellen, werden Thesauri genannt⁵¹. Thesauri bilden gewissermaßen eine Kombination der Vorteile von Schlagwortsystemen und von Klassifikationen, mit denen sie eine ausgeprägte hierarchische Struktur gemein haben – oft eine sogar darüber hinausgehende polyhierarchische⁵² und polydimensionale⁵³. Von den Schlagwortsystemen her rühren die natürlich sprachlichen Deskriptoren, die Einbeziehung ihrer Synonyme und die zumeist postkoordinative Ausrichtung der Indexierung.

Im Moment gelten noch der ISO-Standard 2788 für monolinguale bzw. ISO 5964⁵⁴ für multilinguale Thesauri, die aber prospektiv durch den derzeit erarbeiteten Standard ISO 25964-1 Thesauri and Interoperability with other Vocabularies. Part 1: Thesauri for Informational Retrieval⁵⁵ abgelöst werden sollen. Dieser geht neben dem Aufbau mono- und multilingualer Thesauri im zweiten Teil auf das Mapping mit anderen kontrollierten Vokabularen ein [Lindenthal/Scheven(2009):Folie2-5]. Diese Neuformulierung wird sich stark am britischen Standard BS8723 Structured Vocabularies for Information Retrieval – Guide⁵⁶ anlehnen, der Thesauri und andere Kontrollierte Vokabulare, ihre Interoperabilität und Formalisierung als Datenmodell systematisch beschreibt. Dieses Datenmodell, das auch dem ISO-Standard 25964-1 zu Grunde liegen wird, umfasst folgende Elemente:

- Begriffe eines Thesaurus (ThesaurusConcept)
- traditionelle hierarchische und assoziative Relationen sowie nutzerdefinierbare Relationstypen zwischen Begriffen
- textuelle Anmerkungen wie Erläuterungen (Scope notes), Änderungsvermerke, Definitionen, redaktionelle Anmerkungen ...
- bevorzugte Benennungen (Deskriptoren) und nichtbevorzugte Benennungen (Nichtdeskriptoren) sowie nichtbevorzugte Komposita (Verweisungen in Begriffskombinationen)
- Begriffsreihen (Arrays) und Node labels
- Begriffsgruppen, anwendungsspezifisch typisierbar [Lindenthal/Scheven(2009):Folie7]

Diese Elemente, ihre Attribute und Beziehungen werden im Datenmodell umfassend definiert und mittels Unified Modelling Language (UML) schematisch dargestellt.⁵⁷ Neben den erläuterten Relationen werden die Beziehung zu einem top concept (top level relationship)

51 Geyer-Hayden spricht hier von der Gruppe der Relationssysteme, unter die sie neben Thesauri auch Semantische Netze und Ontologien subsumiert, siehe [Geyer-Hayden(2009):133-134].

52 Von Polyhierarchie spricht man, wenn Konzepte Beziehungen zu zwei oder mehr Oberbegriffen – jeglicher hierarchischen Ausprägung – haben können. [Stock/Stock(2008):75]

53 Polydimensionalität meint die parallele Unterteilung nach verschiedenen Gesichtspunkten (Merkmalen). [Stock/Stock(2008):79-80] Ist dies in Klassifikationen manchmal unvermeidlich und wird als Fehler begriffen, nutzen einige Thesauri bewusst und sichtbar polydimensionale Strukturen.

54 ISO 2788:1986, siehe http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=7776 und ISO 5964:1985, siehe http://www.iso.org/iso/catalogue_detail.htm?csnumber=12159

55 ISO/DIS 25964-1, siehe http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=53657

56 BS8723, siehe <http://schemas.bs8723.org/> Hier insbesondere Part 5, Exchange formats and protocols for interoperability.

sowie einzelne Arten von notes wie definition, scope note, history note und editorial note definiert, ebenso Elemente und Beziehungen die eine Begriffskombination⁵⁸ (compound equivalence) konstituieren. Beziehungen an sich können spezifiziert werden mittels des role-Attributes: etwa die Äquivalenzbeziehung zwischen Deskriptor und Synonymen als „Abkürzung“, „ältere Form“ oder Assoziationsbeziehungen als kausal, Vorgang/Ergebnis usw. In vielerlei Hinsicht werden Thesaurusstrukturen, die schon länger praktisch gebraucht werden, systematisiert und formalisiert.

Ebenfalls präzisiert wird der Einsatz des *Facettierungsprinzips* in Form von Begriffsreihen und Begriffsgruppen. Begriffsgruppen (concept groups) gliedern Kontrollierte Vokabulare grob in Form von Teilthesauri, Hauptteile, Sachgruppen oder Kategorien.

Begriffsreihen (thesaurus arrays) sind gleichgeordnete Begriffe, die einem node label (eine „Knotenbenennung“, die nicht zur Indexierung, sondern nur zu Systematikzwecken dient) zugeordnet wurden: dies stellt das spezifische Merkmal dar, nach dem die Zusammenfassung der Begriffe erfolgt ist. Die Begriffsreihen brechen die hierarchische Ordnung auf, indem sie zusätzlich zu dieser weitere Gruppierungen nach zusätzlichen Merkmalen ermöglichen. [Lindenthal/Scheven(2009):Folie11-13] Im Zusammenhang mit Thesauri spricht man hier eher von Polydimensionalität der Begriffe. [Lindenthal(2005):Folie20-21]

Ein Beispiel zur Veranschaulichung: in einem Thesaurus zum Personalmanagement, der hierarchisch auf der Ausbildung und Qualifikation der Mitarbeiter aufbaut, wäre z.B. Begriffsreihen mit den node labels Mitarbeiter nach Dauer der Betriebszugehörigkeit und Mitarbeiter nach Familienstand sinnvoll. Die Konzepte der Begriffsreihe wären im ersten Fall zeitlich gestaffelt, im zweiten Fall etwa „ledig“, „verheiratet“, „Kind(er)“ usw. Durch diese Kategorien würden Konzepte (aus der Hierarchie) wie wissenschaftlicher Mitarbeiter, leitender Angestellter, Mitarbeiter im einfachen Dienst sinnvoll ergänzt werden. Dies ist vergleichbar zur facettierten Ergänzung durch Anhängeszahlen in Klassifikationssystemen.

In Klassifikationen und Schlagwortsystemen werden durch die Prinzipien Gleichordnung (Klassenbildung) und Unter-/Oberordnung (Hierarchierelationen) Konzepte in ein Begriffssystem eingeordnet; die Bedeutung (Semantik) der Begriffe wird dadurch mitbestimmt. Das lexikalische Vokabular zur Darstellung der Begriffe wird dabei normiert und es werden z.T. Verweisungen von unkontrollierten Benennungen hergestellt.

57 Der noch nicht als Standard anerkannte fünfte Teil des BS8723, DD 8723-5:2008 Exchange formats and protocols for interoperability, schlägt eine Formalisierung des Datenmodells als UML-Diagramm sowie als XML Schema vor, vgl. <http://schemas.bs8723.org/2007-06-01/Documentation/BS8723-5-Full.html> Erläuterungen dazu bietet Leonard Will in seiner Präsentation http://www.iskouk.org/presentations/will_21072008.pdf

58 Begriffskombination bezeichnet die Struktur in Thesauri, in der durch präkoordinative semantische Begriffszersetzung zur Indexierung und Retrieval eines Konzeptes auf zwei Deskriptoren verwiesen wird. Beispiel: Schienentransport (Kombinationsbegriff, Synonym) USE Transportmittel + Eisenbahn (Deskriptoren).

Das Facettierungsprinzip ermöglicht die Einbeziehung von mehrdimensionalen Aspekten der Konzepte. In Thesauri werden zudem (vielfältige und kontextabhängigere) assoziative Relationen, sowie polyhierarchische und polydimensionale Strukturen dargestellt. Ein differenziertes lexikalisches Vokabular von Vorzugsbenennungen, Synonymen und Begriffskombinationen wird hier verwendet:

Auf diese Art vermitteln die Beziehungen eines Deskriptors zu anderen Bezeichnungen (Deskriptoren oder Nicht-Deskriptoren) in gewisser Weise eine Definition des Deskriptors, da es seinen Ort im semantischen Gefüge aufzeigt. [DIN 1463-1(1987):5.1]

Laut diesem deutschen DIN-Standard bezüglich einsprachiger Thesauri sind es eben die im Vergleich zu anderen KOS qualitativ stark ausgeprägten Relationen, die eine implizite Definition der Konzepte ermöglicht. Das angestrebte Datenmodell des Standard BS8723 erweitert diese Eigenschaft noch, indem es individuelle Spezifikationen der Thesaurusbeziehungen ermöglicht. Mit dem UML-Schema des BS8723 und der Bereitstellung als XML Schema wird ein übergreifendes, formalisiertes Modell für Kontrollierte Vokabulare geschaffen, das der strukturellen Interoperabilität dieser dienen kann.

Entwickelt man dies generell zu freien Definitionsmöglichkeit von Relationen zwischen Konzepten weiter, gelangt man zur mächtigsten Form von knowledge organisation systems, Ontologien.

2.4.2 Kontrollierte Vokabulare in der Normenkontrolle

In den 1984 bzw. 2001 erschienenen GARE/GARR-Richtlinien für die internationale Normenkontrolle sind sachliche Normdaten (subject authority entries) explizit ausgeklammert worden, da diese über die Grenzen der bibliografischen Formalerschließung und ihrer Entitäten hinausgehen würde:

It was recognized that subjects posed additional problems of both a linguistic and conceptual nature, and that experts in subject headings would be needed to develop a system that would be universally useful, regardless of the subject access system used by any particular country or group of libraries. [GSARE(1993):1]

1993 erschienen daher die Guidelines for Subject Authority and Reference Entries (GSARE), welche die Working Group on „Guidelines for Subject Authority Files“ in der Section on Classification in der IFLA Division of Bibliographic Control erarbeitet hatte. Die Vorschriften der GSARE orientieren sich strukturell und inhaltlich stark an den der GARE und streben eine Kompatibilität mit diesen an, da sie sich bezüglich der Behandlung von Eigennamen als Normdaten notwendig überschneiden. Ziel ist hier ebenso die Struktur und Elemente der Normdatensätze (also authority entry bzw. Normdatensatz, reference entry bzw. Verweisung und general explanatory entry bzw. Pauschalverweisung) grob zu standardisieren ohne jedoch die konkrete Form und die Punctuation innerhalb der Elemente festzulegen. Zur Reali-

sierung dieser Richtlinien wird ebenfalls an die nationalbibliografischen Zentren verwiesen. [GSARE(1993):5-6]

Besondere Behandlung erfordert hier allerdings das Zusammenspiel mit den Kontrollierten Vokabularen, auf die sich die sachlichen Normdaten stützen. Subject headings – die sprachlichen Einheiten, auf die sich die Normierung bezieht – werden als normierter sachlicher Sucheinstieg definiert:

A uniform linguistic expression *representing a subject* and used to provide controlled access, in a catalogue or index, to the *content* of documents. It may consist of one indexing term [...] or several [...] combined in a string, according to the established syntax rules of the *indexing system* used in a catalogue, bibliography, etc. [GSARE(1993):15; Hervorhebungen K.W.]

Die subject headings werden darin unterschieden, ob sie Personen, Familien, Körperschaften, Werke, Geografika oder Themen (topical subject heading) bezeichnen. Die zu Grunde liegenden „indexing systems“ werden in den Richtlinien jedoch nur vage thematisiert.

Gemäß den GSARE erfolgt im Normdatensatz die Angabe der subject authority list, aus der das subject heading stammt, als Abkürzung. In Anlehnung an den Thesaurus-Standard ISO 2788 können des Weiteren die Relationen des kontrollierten Begriffs dargestellt werden: neben Äquivalenz-, hierarchischer, Assoziations-, generischer, partitiver und Instanzbeziehung die top level relationship („top term“) und als gesonderte assoziative Beziehungen die antonymische und chronologische Relation. [GSARE(1993):10-17] Diese Beziehungen werden in folgenden Elementen der Normdateneintragung dargestellt: Synonyme als Verweisungsreferenz (see reference tracing) mit dem Symbol UF (used for); Hierarchien, top level relationships, Assoziationsbeziehungen als see also reference tracing mit den Symbolen BT (broader term), NT (narrower term), TT (top term) RT (related term). Notationen von Klassifikationen, die dem kontrollierten Begriff zugeordnet sind, werden im information-notes-Bereich nach den scope notes (Definition oder Erläuterung des Konzepts) angesiedelt. Eine authority data number kann zur Identifizierung von Normdateneintragungen und Pauschalverweisungen eingesetzt werden. [GSARE(1993):18-46]

Die GSARE bilden die klassische Struktur von Normdaten und ihren Verweisungen ab, im Vordergrund steht die Verwendung von kontrollierten Begriffen als sachliche Sucheinstiege in bibliografischen Systemen. Aus diesem Grund erfolgte nur eine beiläufige Beschreibung von Konzepten aus Kontrollierten Vokabularen. Eine umfassende Einbeziehung der zu Grunde liegenden KOS war nicht vorgesehen bzw. konnte aus dieser eingeschränkten Perspektive nicht geleistet werden.

Maria Ines Cordeiro beklagt wahrscheinlich zu Recht, dass die konzeptionelle Entwicklung der sachlichen Normenkontrolle gegenüber der Normenkontrolle i.e.S. bisher vernachlässigt wurde:

Besides the lack of a wide and well established culture of authority work, the achievements in consensual practices and cooperation have been almost restricted to the field of name authority, closer to cataloguing; this has received much more attention than subject authority from bodies concerned with policies of bibliographic control, like IFLA and national bibliographic agencies. [Cordeiro(2003):3]

2.4.3 Functional Requirements for Subject Authority Data (FRSAD)

Die Zielsetzung der FRBR war ein übergreifendes Modell des bibliografischen Universums in Form von bibliografischen Entitäten mit Attributen und Relationen. Bereits dort wird die allgemeine Themenbeziehung zwischen der Entität Werk und allen bibliografischen Entitäten angedacht. Insbesondere werden hierfür die Gruppe-3-Entitäten Begriff, Gegenstand, Ereignis und Ort (concept, object, event, place) definiert, die in Ergänzung zu den Entitäten der ersten beiden Gruppen als Themen (subjects) eines Werkes fungieren können. Dies wird folgendermaßen begründet (hier am Beispiel der Entität Ereignis):

Die Definition der Entität *Ereignis* ermöglicht die konsequente Benennung bzw. die Identifizierung eines *Ereignisses*, unabhängig vom Vorliegen oder der Art des Namens des *Ereignisses* auf oder in einer bestimmten *Expression* bzw. *Manifestation* eines *Werkes*. [FRBRde(2009):26, Hervorhebungen original]

Damit wird bereits die grundlegende Funktion der Normenkontrolle (authority control) angesprochen, die Herstellung einer sprachlichen Kontingenz bzw. einer eindeutigen Identifizierung von Konzepten. In der darauf folgenden FRAD-Studie wurde eine neue Struktur der Normenkontrolle entworfen: Normierte Sucheinstiege (controlled access points) basieren auf Namen (names) oder Identifikatoren (identifiers), die in Beziehung zu den bibliografischen Entitäten stehen. Wichtig ist die Unterscheidung, dass Namen in einer n:n-Beziehung, Identifikatoren aber in einer n:1-Beziehung zu den bezeichneten Entitäten stehen, also eindeutig sein müssen. Diese Struktur war für die Normenkontrolle von Eigennamen (wie sie im FRAD dezidiert behandelt wird) wie für die der sachlichen Erschließung geplant. Zudem werden hier bereits Attribute für die Gruppe-3-Entitäten zu dem in den FRBR bereits genannten Attribut Name (term) ergänzt. [FRADdraft(2007):23-24][FRBR(1998):53-55]

Die explizite Auseinandersetzung mit dem Thema der sachlichen Normenkontrolle wurde schließlich in der 2005 gegründeten Working Group on Functional Requirements of Subject Authority Records (FRSAR), angesiedelt in der IFLA Division of Bibliographic Control, angegangen. Als zentrale Aufgaben für die FRSAR Working Group wurden benannt: Die Erstellung eines Konzeptmodells für die Gruppe-3-Entitäten der FRBR, das eine klar definierte und strukturierte Zuordnung der sachlichen Normdaten zu Nutzerbedürfnissen ermöglicht; sowie eine Einschätzung zum potentiellen internationalen Austausch von Daten der sachlichen Normenkontrolle. [FRSADdraft(2009):7] Als Ergebnis der Working Group liegt mo-

mentan ein zweiter Entwurf (2nd Draft 2009-06-10) der Functional Requirements for Subject Authority Data: A Conceptual Model (FSRAD) [FRSADdraft(2009)] vor.

Das Überraschende am derzeitigen FRAD-Entwurf ist, dass die bisherige Konzeption der vier Gruppe-3-Entitäten zur Darstellung sachlicher Themen verworfen wird. Vermutet wurde, dass vier Kategorien weder ausreichend noch adäquat dafür seien, sämtliche Konzepte einer sachlichen Erschließung zu erfassen bzw. sinnvoll zu gliedern. Aus diesem Grund initiierte die Working Group eine Pilotstudie, in der Begriffe vorhandener Kontrollierter Vokabulare von LIS-Studenten und -Fakultätsmitarbeitern in sechs Kategorien (concrete stuff, abstract stuff, event, time, place, others) eingeordnet werden sollten. Das Ergebnis war ernüchternd:

The results of this test indicate that it would be difficult for any user (end user, librarian or vocabulary developer) to conduct such a task when using subject authority data. These categories do not seem helpful or necessary to the end users either. [FRSADdraft(2009):21]

In einem nächsten Schritt wurden verschiedene alternative Konzeptionen für solche grundlegende Entitäten eruiert. Als einfachste Möglichkeiten ergaben sich, im Sinne der Fortführung des FRBR/FRAD, trotzdem die Gruppe-3-Entitäten beizubehalten bzw. diese um die Entität time zu erweitern. Weitere Szenarien waren die Anwendung der Grundfacetten der Colon Classification⁵⁹, der Grundentitäten des <indecs>-Metadatenmodells⁶⁰ oder eine neue Sammlung und Festlegung von entsprechenden Entitäten. [FRSADdraft(2009):21-24] Alle diese Möglichkeiten wurden vor allem aus folgendem Grund verworfen:

None of the scenarios 1-5 are ideal for all situations, while each may be a good solution for particular implementations. Any further categorization of Group 3 entities would prescribe a particular way of structuring the subject languages that are used to provide access to works. [FRSADdraft(2009):24]

Das FRSAD-Modell sollte demgegenüber ein möglichst allgemeingültiges, abstraktes Rahmenwerk sein, dass alle potentiellen domänenspezifischen Strukturen und Implementationen zulässt. Daran anschließend wurde auch Wert auf die Kompatibilität mit anderen, ähnlichen Modellen und Modellierungssprachen gelegt. [Zeng/Zumer(2009)] Die FRSAR-Working-Group entschied sich bei dieser Frage schließlich für eine sechste Option, nämlich die generelle Absage an eine Kategorisierung von Sach-Entitäten.

59 In der von Shiyali R. Ranganathan entwickelten Colon Classification sind die fünf Grundfacetten, anhand derer die Notation synthetisiert wird: personality (Personalität), matter (Stofflichkeit, Materie), energy (Energie, Dynamik), space (Raum) und time (Zeit) – abgekürzt wird auch von der PMEST-Formel gesprochen.

60 Die Hauptentitäten des <indecs> Metadata Framework sind: percept (perzept) mit der Unterscheidung in being (Iebewesen) und thing (ding), concept (konzept), relation (relation) mit der Unterscheidung in event (ereignis) und situation (situation), vgl. Godfrey Rust, Mark Bide, The <indecs> Metadata Framework, http://www.doi.org/topics/indecs/indecs_framework_2000.pdf bzw. die deutsche Übersetzung, http://mm-dra.de/datenmodell_02-01-21/AGMD_IndecsFrameworkDeutsch.pdf

Die grundsätzliche Struktur, wie sie das FRSAD entwirft, ist entsprechend einfach. Es werden mit Bezug auf das Semiotische Dreieck⁶¹ die Entitäten *thema* („any entity used as a subject of a *work*“) und *nomen* („any sign org sequence of signs [...] by which a *thema* is known, referred to or adressed as“) [FRSADdraft(2009):25] eingeführt, die in folgender Beziehung zur Entität Werk stehen:



Abbildung 5: FRSAD - Konzeptmodell [FRSADdraft(2009):25]

Zu beachten sind die n:n-Kardinalitäten der Beziehungen. Ein Werk kann mehrere themata behandeln, wie ein thema in vielen Werken behandelt werden kann. Ebenso verfügt ein thema über mehrere nomina, also Bezeichnungen, und ein nomen kann mehrere themata bezeichnen. Der FRSAD-Entwurf betont jedoch, dass in Kontrollierten Vokabularen eben Letzteres nicht gelten sollte, sondern ein nomen ein eindeutiger Bezeichner eines themas sein sollte. [FRSADdraft(2009):28]

Im Zusammenhang mit dem FRBR-Modell wird die Entität *thema* als eine Oberklasse der Gruppe-1- und Gruppe-2-Entitäten, sowie aller anderen Themen (subjects) von Werken aufgefasst. *Thema* erhält grundsätzlich das Attribut *type*, deren Werte allerdings von der konkreten Anwendung bestimmt werden. Explizit erwähnt der FRSAD-Entwurf Klasse oder Instanz als *type*-Wert. Darüber hinaus werden sonstige Attribute ebenfalls als implementationsabhängig, also frei definierbar beschrieben. Dies gilt ebenso für Attribute von *nomina*, hier werden allerdings die Gebräuchlichsten benannt: *type*, *schema*, *reference source*, *representation*, *language*, *script*, *transcription/transliteration*, *encoding*, *form*, *time of validity*, *audience* und *status*. [FRSADdraft(2009):27-36] Interessant ist, dass *identifier* (neben *controlled name*) hier als ein Wert des *nomen*-Attributs *type* definiert wird, also von der in den FRAD vorgestellten Struktur der nebengeordneten Entitäten *Name* und *Identifikator* abweicht.

Schließlich analysiert der FRSAD-Entwurf die Beziehungen zwischen *themata* sowie zwischen *nomina*. *Thema*-*thema*-Beziehungen werden analog zu bereits erläuterten Relationen in Begriffssystemen beschrieben: die hierarchischen Relationen (*generische*, *partitive*, *Instanzrelation*) inklusive *Begriffsreihen* (*Facettierung*) und *Polyhierarchie* sowie die *Assoziationsrelation* mit Beispielen. Als *nomen*-*nomen*-Beziehungen werden die *Äquivalenzrelation* und die *partitive Relation* zwischen Bestandteilen von *nomen* (die separate *nomen* sein können)

⁶¹ Bezüglich des Semiotischen Dreieck bzw. dem triadischen Zeichenbegriff wird im FRSAD-Entwurf neben ersten Ideen bei Philosophen der Antike auf Charles Kay Ogden und I. A. Richards, *The Meaning of Meaning : a study of the influence of language upon thought and of the science of symbolism*, London: Routledge & Kegan Paul, 1923, referiert. Siehe dazu auch [Stock/Stock(2008):51-52].

genannt. Zusätzliche, anwendungsbezogene Relationen zwischen themata sowie zwischen nomina sind möglich. [FRSADdraft(2009):37-43]

Das soweit entwickelte FRSAD-Modell wurde kritischer als die vorangegangenen FRBR und FRAD aufgenommen. Sicher ist der konkrete Bezug der praktischen Sacherschließung auf die sachliche Normenkontrolle bisher ein Desiderat gewesen. Und man hat durchaus die Herausforderung erkannt, die sachliche Normenkontrolle so neu auszurichten, dass man die Begriffssysteme, welche die Grundlage für die sachliche Erschließung liefern, angemessen einbeziehen und abbilden kann:

In addition to ensuring consistency in subject representation, a subject authority record also records and maintains semantic relationships among subject terms and /or their labels. Records in a subject authority file are connected through semantic relationships, which may be expressed statically in subject authority records or generated dynamically [...] of printed or online display of thesauri, subject headings lists, classification schemes, and other knowledge organization systems. [FRSADdraft(2009):5]

Allerdings erweist sich die minimal gehaltene Struktur des FRSAD eher als ein kleinster gemeinsamer Nenner der sachlichen Normenkontrolle. Die Konzeption von thema und nomen stellt prinzipiell keine neue Idee dar, sondern ist als Begriff-Benennung-Paradigma grundlegendes Prinzip der Sachlichen Erschließung. Mit der großen Freiheit, die man bezüglich Kardinalitäten der Beziehungen und Attributen der Entitäten lässt (und mit den großen Unterschieden in den vorhandenen KOS z.T. nachvollziehbar begründet), geht auch eine gewisse Beliebigkeit einher.

Die Schlussfolgerung aus der Pilotstudie, sachliche Kategorien oder Facetten grundsätzlich abzulehnen, wird von verschiedenen Seiten als überzogen angesehen. Sind doch viele Unterscheidungen, die FRBR und FRAD bisher einführten, auch für Nutzer nicht ohne Weiteres einsichtig, wie Jenn Riley anmerkt. [Riley(2009)] Insbesondere bemängelt Karen Coyle in ihrem Blog das Fehlen des Identifikator-Konzeptes:

Without an identifier, there is no way to represent an entity as metadata. The Work and the Thema (FRSAD's word for subject) have no existence in metadata without a machine-readable identity that allows them to have being. This is a basic rule of the Semantic Web, but it has always been a fact of metadata usage in machine-readable form. [Coyle(2009)]

Zwar taucht der identifier im FRSAD auf, jedoch nur als ein Attribut (von nomen), nicht als eigenständige Entität. Inwieweit dies mit dem genuinen Anliegen von Normenkontrolle vereinbar ist, ist fraglich. Zwar wird in der Einleitung der FRSAD die klassische Forderung der Normenkontrolle aufgestellt:

The purpose of authority control is to ensure consistency in representing a value – a name of a person, a place name, or a subject term – in the elements used as access points in information retrieval. [FRSADdraft(2009):5]

Aber eben diese geforderte Konsistenz wird – mehr noch als im FRAD-Modell – durch die oben beschriebenen Beliebigkeiten und dem Fehlen der eigenständigen Entität Identifikator

konterkariert. Noch unverständlicher erscheint, dass die für die Normenkontrolle zentrale Entität des Normierten Sucheinstiegs (controlled access point) keinerlei Erwähnung findet, geschweige denn eine Klärung der Beziehungen zu den FRSAD-Entitäten erfolgt. Insgesamt weicht der FRSAD-Entwurf entscheidend von den vorhergehenden Modellen ab und wirft die Frage auf, wie mit den unterschiedlichen Strukturen und Entitäten praktisch umgegangen werden soll. Jenn Riley äußert in ihrem Blog Unverständnis darüber, dass eine Abstimmung der drei Modelle offenbar nicht forciert wurde:

I'm concerned about the significant lack of cohesion between the FRBR, FRAD, and FRSAR [sic] reports. They show their nature of independently generated by different groups with different interests over a long span of time. [...] In fact, the thema/nomen structure in the FRSAD draft isn't really all that different than the (whatever entity)/name structure presented in FRAD. Much greater cohesion of the three reports could be made - what's written here seems to ignore FRAD in particular. [Riley(2009)]

2.5 Fazit: Bibliografische Kontrolle

Zum einen sollte in diesem strukturellen und chronologischen Rückblick die synthetische Funktionsweise der klassischen Bibliografischen Kontrolle erläutert werden. Nach dieser werden die Komponenten Bibliografische Beschreibung, Normenkontrolle und Kontrollierte Vokabulare durch die Vorschriften des vierten Teilbereichs Katalogisierung in der Praxis zusammengeführt. Deren Ergebnisse waren Titelaufnahmen, mehr oder weniger konsistente Surrogate, welche die Ergebnisse aller Teilbereiche, und dies nur schwach strukturiert, in sich vereinigten. Abstrahiert kann dieses Systems folgendermaßen dargestellt werden:

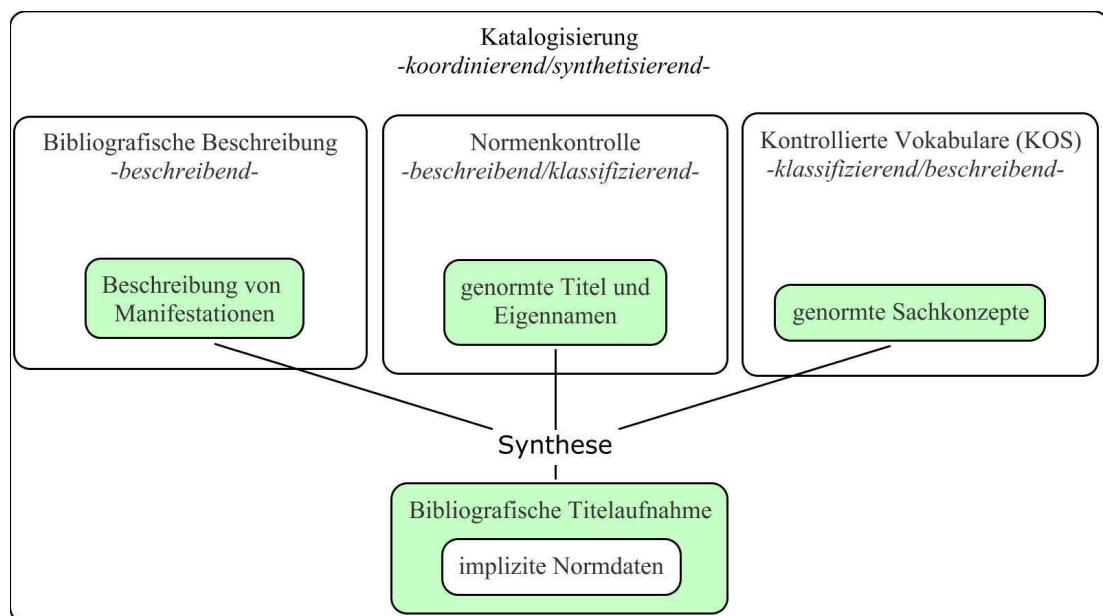


Abbildung 6: Klassische Bibliografische Kontrolle

Zum anderen sollten aktuelle theoretische Rahmenwerke (die sich weiter in Diskussion und Bearbeitung befinden) vorgestellt werden, die zwar auf dieser Gliederung aufbauen, jedoch eine strukturelle Neuordnung des „bibliografischen Universums“ anstreben. FRBR, FRAD und FRSAD beschäftigen sich nicht mit der praktischen Umsetzung der Bibliografischen Kontrolle, sondern versuchen eine grundlegende und einheitliche Modellierung durch bibliografische und zusätzlichen Entitäten, die über spezifische Merkmale und Beziehungen verfügen. Auf dieser Grundlage könnte eine wirklich gemeinsame und konsistente Struktur von Bibliografischer Kontrolle entstehen, die auf allgemeinere Datenmodelle leicht abbildbar ist. Darauf wird zurück zu kommen sein. Zunächst soll die praktische Weiterentwicklung, welche die klassische Bibliografische Kontrolle in Form von Metadaten systemen genommen hat, betrachtet werden.

3 Metadaten als Mittel der Bibliografischen Kontrolle

3.1 Bibliografische Kontrolle im elektronischen Kontext

Die früheren Standards der Paris Principles und der ISBDs wurden in diversen nationalen Katalogisierungsregelwerken in den 1960er und 1970er Jahren umgesetzt, sodass ein Mindestmaß an allgemeiner Verständlichkeit und Austauschbarkeit erreicht wurde. Als in den 60er Jahren mit den ersten Versuchen computergestützter Katalogisierung begonnen wurde, rückte diese potentielle Austauschbarkeit von bibliografischen Informationen in realistische Nähe. Waren diese Bestrebungen zunächst auf die schnelle und bibliothekenübergreifende Produktion von Katalogkarten ausgerichtet, erkannte man bald den Nutzen der Datenbanktechnologie für einen „Online-Katalog“. Die sequentielle Speicherung von Titelaufnahmen als abgeschlossene Einheiten in Buch- und Kartenkatalogen wurde aufgebrochen und machte der Darstellung der einzelnen bibliografischen Elemente in Tabellen von relationalen Datenbanken Platz. Mit der Abkopplung der als Ordnungselemente verwendeten Normierten Sucheinstiege von den Titelaufnahmen entstanden zudem eigenständige Normdatensätze. Ein neues Prinzip der Datenhaltung setzte sich durch: Datensätze mit vielen einzelnen Datenfeldern von Attribut-Wert-Paaren. Seit 1969 wurden die national verschiedenen Varianten von MACHINE READABLE CATALOGUING (MARC⁶²) [Haynes(2004):3-4] und 1973 das Maschinelle Austauschformat für Bibliotheken (MAB⁶³) für Deutschland entwickelt, die als Austauschformate konzipiert waren; später kamen diverse interne Formate von proprietären Katalogsystemen hinzu. Datensatzstrukturen und -inhalte dieser Formate, etwa Datenfelder, Unterfelder, Indikatoren und Indikatorwerte [Liu(2007):19], waren auf die ebenfalls in dieser Zeit entstandenen Katalogisierungsregelwerke ausgerichtet. Elektronische Katalogisierung und – zumindest mittelbare – internationale Standardisierung befruchteten sich dabei gegenseitig:

The cataloging rules of the nineteenth and twentieth centuries evolved from institutional-specific rules to a modern concept of a widely used standard for sharable data that would facilitate the exchange of catalog information between libraries. [Coyle(2010a):11]

Das zentrale Verdienst der datenbankbasierten Katalogisierung war es, Elemente der Bibliografischen Beschreibung und Normenkontrolle durch syntaktische Differenzierung einzeln sichtbar zu machen. Diese Elemente waren zwar schon lange einer konzeptionellen Analyse und Standardisierung unterworfen, jedoch verschwammen deren Ergebnisse in der komprimierten Form der klassischen Titelaufnahme. Die Auftrennung der starren Form der

62 Zur Übersicht über die MARC-Familie siehe <http://www.loc.gov/marc/> die aktuelle angloamerikanische Version MARC21 siehe <http://www.loc.gov/marc/96principl.html>

63 Zur Übersicht über die MAB2-Formate siehe <http://www.d-nb.de/standardisierung/formate/mab.htm>

bibliografischen Surrogate bewirkte ein Paradigmenwechsel, der retrospektiv als Übergang von der klassischen Katalogisierung zur *Metadaten*-Generierung verstanden werden kann. Der Ersatz von Titelaufnahmen durch Titeldatensätze und das Sichtbarwerden von Normdaten in Form von Normdatensätzen⁶⁴ ist nur ein oberflächlicher Ausdruck dessen. Das entscheidende Moment liegt aber in der Fragmentierung in einzelne bibliografische Elemente, die einen anderen Umgang mit, jetzt eher *bibliografischen Daten*, denn *Beschreibungen*, erforderte.

3.2 Metadaten – Begriffsdiskussion

Zur Klärung des Begriffs *Metadaten* wird häufig auf die prägnante und weitläufige Formel „data about data“ zurückgegriffen⁶⁵. Ein solche Definition ist nicht nur wenig aussagekräftig, da selbstreferenziell, sondern zumindest in Deutschland missverständlich, da der Datenbegriff und die damit zusammenhängenden Begriffe Information und Wissen in der informationstechnischen Theorie und dem pragmatischen informationswissenschaftlichen Ansatz stark differieren [Kuhlen(2004):158-159]. Weniger befangen geht man in englischsprachigen Raum mit den Begriffen Daten und Information um, denn obwohl man sich auch hier prinzipiellen Ambiguitäten bewusst ist [Greenberg(2003):1876], werden diese Begriffe oft gleichbedeutend zur Definition des Metadatenbegriffs eingesetzt⁶⁶. Aber die Umschreibung „data about data“ oder „information about information“ hat sich auch hinsichtlich der Objekte, die Metadaten beschreiben sollen, überlebt.

Wahrscheinlich entstand der Begriff im Umfeld der frühen Datenbanksysteme in den 1960er Jahren, auf jeden Fall hatte er sich in den 70er Jahren in diesem Forschungsfeld etabliert; später gewann der Begriff auch im Zusammenhang mit Daten aus Geoinformationssystemen (GIS) an Bedeutung: die Entwicklung von Metadatenstandards erleichterte den Austausch und gemeinsame Verarbeitung von GIS-Daten. In den 1990er Jahren fand der Begriff Verbreitung in immer mehr Fachcommunities zur Beschreibung von Dokumenten im WWW [McCathieNevile/Méndez(2007):29]. Ein erstes Konzept ganz allgemein gehaltener Metadaten schuf schließlich die Dublin Core Metadata Initiative (DCMI), die sich mit dem Ziel gründete, ein grundlegendes, disziplin- und formatübergreifendes Metadaten-set für Online-dokumente zu schaffen [Haynes(2004):4-5]. Ausrichter des ersten Metadata Workshops 1995 in Dublin, Ohio war neben dem National Center for Supercomputing Applications (NCSA)

64 Im Moment noch in eingeschränktem Maße, z.B. im Onlinekatalog des Südwestdeutschen Bibliotheksverbundes Baden-Württemberg, Saarland, Sachsen (SWB), vgl. <http://swb.bsz-bw.de/DB=2.1/>

65 Siehe z.B. die Sammlung von aktuellen Definitionen bei [Liu(2007), S. 5]

66 In den aktuellen Einführungswerken zu Metadaten im Bibliotheksbereich [Foulonneau/Riley(2008)], [Intner/Lazinger/Weihs(2006)], [Liu(2007)], [Haynes(2004)] finden sich zahlreiche Metadaten-Definitionen, das Nebeneinander von „data“ und „information“ als Grundbegriffe stellt aber anscheinend kein Problem für die Autoren dar.

das Online Computer Library Center (OCLC), das größte, international agierende Bibliotheksverbundsystem. Neben den Erfahrungen, die Datenbankentwickler und Vertreter von Fachcommunities einbrachten [Lagoze/Lynch/Daniel(1996):3], beruht die Entwicklung des Dublin Core Metadata Element Set (DCMES) also zum großen Teil auf bibliothekarischem Fachwissen. Insofern mutet es befremdlich an, dass Bibliothekare sich heute den Begriff Metadaten wieder aneignen müssen – und dies kritisch tun:

We can change our job titles, but being forced to do so points to the larger misconception that a cataloger and a “metadata specialist” are two completely different professions. [Miksa(2008):28-29]

Dadurch, dass der Begriff von unterschiedlichsten Disziplinen und Gesellschaftsbereichen schnell aufgenommen wurde und eine universelle Bedeutung erlangte, kann heute mit dem konzeptionellen Begriff Metadaten prinzipiell jede Art von Entität – physische (Bsp. Produkte, Personen), konzeptionelle (Bsp. Theorien, Beziehungen), prozessurale (Bsp. Vorgänge, Verfahren) – beschrieben werden. Die Ausdruckskraft von Metadaten variiert nicht nur sozusagen in der horizontalen, sondern auch in der vertikalen Ebene: Beschreibbar sind neben Entitäten auch Zusammenstellungen von Entitäten (Bsp. Stadt, Datenbank, Menschengruppe) oder einzelne Komponenten von Entitäten⁶⁷ [Gilliland(2008):2], man spricht hier von der Granularität der Metadaten⁶⁸.

Gerade wegen der wachsenden analogen und vor allem digitalen Vielfalt von Bibliotheksmaterialien wird der „Universalbegriff“ Metadaten nun vermehrt in Informationseinrichtungen verwendet und scheint den Begriff der bibliografischen Beschreibung bzw. der Katalogisierung langsam abzulösen. Dies geschieht nicht unumstritten und die Diskussion, ob „Metadaten“ nicht doch nur für digitale Objekte bzw. Informationsobjekte⁶⁹, oder nur für Objekte im Internet angemessen sind, ist noch nicht abgeschlossen. [Intner/Lazinger/Weihs(2006):4] Vorsichtig bleibt auch die Deutsche Nationalbibliothek (DNB) auf ihren Internetseiten:

Metadaten sind (strukturierte) Daten, die eine Ressource, eine Entität, ein Objekt oder andere Daten beschreiben. [...] Im informationswissenschaftlichen und bibliothekarischen Kontext versteht man hierunter Daten, die der Beschreibung von elektronischen Ressourcen dienen. Der Trend geht jedoch dahin, den Begriff Metadaten auch für Daten und Kataloge in Datenbanken zu verwenden, die nicht-elektronische Ressourcen beschreiben. [DNBmd(2010)]

Neben dem Entitäts- und Objektbegriff taucht hier der Ressourcen-Begriff auf, der in dieser Verwendung im deutschen Sprachgebrauch eigentlich unüblich ist und wahrscheinlich in Analogie zum Englischen entstand. Der Begriff *resource* stammt zwar aus der WWW-Kon-

67 Ausschlaggebend für den Begriff der Entität ist die *Wahrnehmung* als eine Einheit.

68 Vgl. die Definition von *granularity* in der „Introduction to metadata“ des Getty Research Institute: „The level of detail at which an information object or resource is viewed or described.“ [Baca-Glossary(2008)]

69 Vgl. die Definition von *information object* in der „Introduction to metadata“ des Getty Research Institute: „A digital item or group of items referred to as a unit, regardless of type or format, that a computer can address or manipulate as a single discrete object.“ [Baca-Glossary(2008)]

Aber auch der übergeordnete Begriff *Objekt* setzt sich in der analogen Bibliothekswelt durch, seit Öffentliche Bibliotheken ihren Nutzern audiovisuelle Medien, Brett- und Computerspiele und vieles mehr anbieten.

zeption des W3C (vgl. Uniform Resource Locator, URL), vergleichbar mit dem Metadatenbegriff hat aber auch dieser inzwischen eine Entwicklung aus dem internetspezifischen Raum (i.S. von web resource) hin zu einem allgemeinen Konzept gemacht:

A resource can be anything that has identity. Familiar examples include an electronic document, an image, a service (e.g., 'today's weather report for Los Angeles'), and a collection of other resources. Not all resources are network 'retrievable'; e.g., human beings, corporations, and bound books in a library can also be considered resources. [Berners-Lee et al(1998)]

Zudem lehnt sich die DNB damit an die englischsprachige – und internationale – Katalogisierungsterminologie an, in der bibliotheksrelevante Materialien neuerdings unter dem Begriff der *bibliographic resources* zusammengefasst werden, wie z.B. deren Definition im Statement of International Cataloguing Principles (ICP) zeigt:

An entity within the realm of library and similar collections consisting of the products of intellectual or artistic endeavour [ICP(2009):10]

Der Schritt von der *bibliographic resource* zur allgemeinen Ressource in der Definition der DNB ist im Zusammenhang mit ihrem umfassenden Metadatenbegriff nachvollziehbar. Auch in internationalen Kontext gibt es diese Verknappung auf den allgemeinen Begriff resource. So hat die ISBD Review Group der IFLA entschieden, in der aktuellen Fassung der ISBD diesen Begriff anstatt item oder publication zu verwenden. [ISBD(2007):ix] Eine resource ist hier

...[a]n entity, tangible or intangible, that comprises intellectual and/or artistic content and is conceived, produced and/or issued as a unit, forming the *basis of a single bibliographic description*. Resources include text, music, still and moving images, graphics, maps, sound recordings and video recordings, electronic data or programs, including those issued serially. [ISBD(2007):Appendix E: Glossary, Hervorhebung K.W.]

Entsprechend diesem auf Bibliotheksmaterialien fokussierten Ressourcen-Begriff behält man den Begriff der *bibliographic description*⁷⁰ bei und spricht nicht etwa von *resource description* – verständlicherweise, stellte man doch damit den Titel des gesamten Regelwerks in Frage. Der Begriff metadata wird ebenso vermieden, ist aber auch nicht notwendig: bibliografische Beschreibung beinhaltet in seiner begrifflichen Ambiguität ja sowohl die Tätigkeit wie das Ergebnis. Im Falle des Metadaten-Begriffs ist demgegenüber resource description als erklärender Vorgang notwendig, wie die Definition der National Information Standards Organization (NISO) zeigt:

In the library environment, [the term, K.W.] metadata is commonly used for any formal schema of resource description, applying to any type of object, digital or non-digital,... [NISO(2004)]

Und sie stellt fest:

Traditional library cataloguing *is a form* of metadata; MARC21 and the rule sets used with it, such as AACR2, are metadata standards. [NISO(2004), Hervorhebung K.W.]

70 „A set of bibliographic data recording and identifying a resource.“ [ISBD(2007):Appendix E: Glossary].

Zusammenfassend lässt sich der entstehende Konsens über die Definition von Metadaten als „Beschreibung von Ressourcen“ (DNB) bzw. resource description (NISO) festhalten. Dabei werden Bibliografische Beschreibung bzw. Katalogisierung als eine spezielle Form von Metadaten(-generierung), und die bibliografischen Regelwerke und Datenformate als spezifische Metadatenstandards angesehen.

3.3 Metadaten als Bibliografisches System

Die meisten Metadaten-Definitionen gehen wie die DNB-Definition von strukturierten Daten (oder Informationen) aus, während die NISO in der oben genannten Definition genauer von einem formalen Schema spricht. Bei einem solchen *Metadatenchema* handelt es sich um eine strukturierte Zusammenstellung von festgelegten Metadatenelementen, also von Merkmalskategorien, die bei der Ressourcen-Beschreibung mit den objektspezifischen Werten kombiniert werden. Genauer erklärt die DCMI auf ihrer Internetseite:

Schemas are machine-processable specifications which define the structure and syntax of metadata specifications in a formal schema language. (DCMI)⁷¹

Während die DCMI (mit Bezug auf die Maschinenlesbarkeit) zwischen Metadatenchema und Metadatenspezifikation unterscheidet, werden die Begriffe, neuerdings auch „Metadaten-sprache“, zumeist synonym gebraucht [Greenberg(2003)]. In Deutschland ist wohl der Begriff Metadatenformat am gebräuchlichsten.

Daneben wird häufig der Begriff *Metadatenstandard* bzw. metadata standard benutzt, der aber schlicht ausdrückt, dass dieser durch eine nationale oder internationale Standardisierungsorganisation anerkannt wurde oder zumindest als Quasi-Standard aufgrund seiner Verbreitung gelten darf. Er ist damit ein rein formaler Begriff, der noch nichts über die Art und den Inhalt des Datenstandards aussagt. Eine sinnvolle Einteilung von Metadatenstandards für die bibliografische Sphäre bietet Anne J. Gilliland in ihrem einführenden Aufsatz. [Gilliland(2008):3] Sie unterscheidet⁷²

- data *structure* standards (metadata element sets, schemas) – also Metadatenschemata (Definition eines Metadatensatzes bzw. der Metadatenelemente, die in diesem Metadaten-set zusammengefasst sind)

71 DCMI, siehe <http://dublincore.org/schemas/>. Allgemeiner betrachtet, definiert Tim Berners-Lee: „any document which describes, in formal way, a language or parameters of a language [sic!]“. [Berners-Lee(1999)]

72 Die Typologie basiert auf dem Aufsatz „CDWA Lite for Cataloging Cultural Objects (CCO) : a new XML Schema for the cultural heritage community“ von Karim Boughida. Hervorhebungen von Anne J. Gilliland, deutsche Bezeichnungen und Beispiele z.T. von mir, K.W.

wie die *Elementsets* von MARC und MAB2⁷³, das Dublin Core Metadata Element Set (DCMES)⁷⁴, die CDWA⁷⁵;

- data *value* standards (controlled vocabularies, thesauri, controlled lists) – also Datenwertstandards (welche die möglichen Attributwerte für die Metadatenelemente festlegen)

wie nationale Normdateien, die Dewey-Dezimalklassifikation (DDC)⁷⁶, die Medical Subject Headings (MeSH), der ISO-Standard 3166 für Länderangaben⁷⁷, das DCMI Type Vocabulary⁷⁸;

- data *content* standards (cataloging rules and codes) – also Erschließungsstandards (welche die Ableitung der Metadaten vom Objekt reglementieren)

wie RAK, AACR2, ISBD, CCO⁷⁹, DACS⁸⁰, AMIM⁸¹;

- data *format/technical interchange* standards (metadata standards expressed in machine readable form) – also Kodierungsstandards

wie MARC21, und vermehrt XML-basierte Kodierungsstandards wie MABxml⁸², die EAD DTD⁸³, die Qualified DC XML Schemas⁸⁴ oder das VRA Core 4.0 XML Schema⁸⁵.

Leider hat sich auch im englischsprachigen Raum bisher keine einheitliche Terminologie herausgebildet. So stimmt Eleta Exline z.B. mit dieser Einteilung überein, verwendet jedoch die Begriffe *schema*, *content values*, *content rules*, *syntax* (in dieser Reihenfolge). [Exline(2008)]

Wie an den Beispielen oben teilweise schon ersichtlich wird, sind diese Standards aufeinander angewiesen, um eine vollständige und funktionale Ressourcen-Beschreibung zu realisieren. So schreiben Lois Mai Chan und Marcia Lei Zeng:

For each element defined, a metadata standard usually provides content rules for how content should be included (for example, how to identify the main title), representation rules for content (for

73 MARC21 und MAB2 bieten unterschiedliche Datenformate für bibliografische Beschreibungen, Personennamen, Körperschaftsnamen, Schlagwörter und Lokaldaten, die freilich auf einer strukturellen Grundlage beruhen. Vgl. <http://www.loc.gov/marc/marcdocz.html> und <http://www.d-nb.de/standardisierung/formate/mab.htm>

74 Dublin Core Metadata Metadata Element Set (DCMES), siehe <http://dublincore.org/documents/dces/>

75 Categories for the Description of Works of Art (CDWA), siehe http://www.getty.edu/research/conducting_research/standards/cdwa/

76 Dewey-Dezimalklassifikation (DDC), siehe <http://www.ddc-deutsch.de/>

77 ISO 3166. Codes for the representation of names of countries and their subdivisions, siehe http://www.iso.org/iso/country_codes.htm

78 DCMI Type Vocabulary siehe <http://dublincore.org/documents/dcmi-type-vocabulary/>

79 Cataloging Cultural Objects (CCO), siehe <http://www.vraweb.org/ccoweb/cco/index.html>

80 Describing Archives: A Content Standard (DACS), siehe <http://www.archivists.org/catalog/pubDetail.asp?objectID=1279>

81 Archival Moving Image Materials: a cataloging manual (AMIM), siehe <http://www.loc.gov/cds/catman.html#amima>

82 MABxml, siehe <http://www.d-nb.de/standardisierung/formate/mabxml.htm>

83 Encoded Archival Description Document Type Definition (EAD DTD), siehe <http://www.loc.gov/ead/>

84 Qualified Dublin Core XML Schemas, siehe <http://dublincore.org/schemas/xmls/>

85 VRA Core 4.0 der Visual Resources Association, siehe <http://www.vraweb.org/projects/vracore4/index.html>

example, capitalization rules or standards for representing time), and allowable content values (for example, whether values must be taken from a specified controlled vocabulary or can be author-supplied, derived from text, or added by metadata creators working without a controlled term list.) [Chan/Zeng(2006):2.1]

Metadaten schemata definieren nur die Metadatenelemente, die einen dem Schema konformen Metadatenatz (metadata set) bilden. Sie können Aussagen treffen, welche Elemente unbedingt nötig sind (mandatory) und welche optional, darüber wie oft ein spezifisches Element benutzt werden und welche Werte es annehmen darf sowie zu deren Abfolge und Beziehungen zueinander. [Chan/Zeng(2006):4.3] Erschließungsstandards benötigen diese aber grundlegend zu ihrer Umsetzung: So spiegeln sich die bibliografischen Konzepte in den Metadatenelementen, in Indikatoren und Qualifikatoren wieder. Datenwertstandards liefern hierfür format- und/oder inhaltsbezogene Attributwerte aus spezifischen Kontexten, die referenzierbar und normiert sind. Für Implementierung, Austausch, Anpassung an sich wandelnde technische Standards und gegenseitige Abbildung von Metadaten schemata sind schließlich verschiedene Kodierungsstandards vonnöten. Jenn Riley merkt sowohl die Komplementarität wie auch die Abgrenzungsschwierigkeiten an, die zwischen diesen Kategorien bestehen:

Metadata element sets suffer from a similar lack of clarity of definition. While these generally provide what most think of as 'fields' for information, some also include prescribed or recommended methods for constructing the values for those fields, a function often discussed as the realm of content standards or encoding schemes (which includes controlled vocabularies). The distinction between these concepts is useful to make, but it is just as important to realize that as a practical consideration, a given standard may cross the boundaries between them in places. [Riley(2008):124]

Der populäre, übergreifende Begriff Metadatenstandard hat also insofern seine Berechtigung, da damit oft Konglomerate dieser unterschiedlichen Datenstandards gemeint sind, die funktionell aufeinander angewiesen sind. Populäre Beispiele sind etwa die Verbindung von AACR2, MARC-Feldstruktur und MARC21-Kodierung oder die Erschließung mittels Dublin-Core-Elementen, DCMI Metadata Terms und XML-Kodierung.

Metadatensysteme in diesem Sinn weisen in ihrem Zusammenwirken eine strukturelle Ähnlichkeit zur traditionellen Bibliografischen Kontrolle auf; sie können also als *Bibliografisches System* betrachtet werden, und zwar mit den folgenden Komponenten: 1. Metadaten schemata – der Definition von Metadatenätzen aus einzelnen Elementen mit Attributen; 2. Datenwertstandards – die formale und/oder inhaltliche Attributwerte für das Schema bereithalten; 3. Erschließungsstandards – die in umfassender Weise die Ableitung von Metadaten von dem Bezugsobjekt reglementieren; 4. Kodierungsstandards – welche die Speicherung und Darstellung des Metadaten schemas festlegen. Erschließungsstandards nehmen dabei die komplexe Sonderrolle ein, die vormals Katalogisierungsregelwerke inne hatten⁸⁶: sie enthalten schon für sich semantische, syntaktische und pragmatische Elemente. [Bianchini/Guerri-

⁸⁶ Oft handelt es sich ja (noch) um eben diese traditionellen Katalogisierungsregelwerke.

ni(2009):114] Inhaltliche Datenwertstandards wie KOS und Normdaten begründen nach wie vor die semantische Ausdruckskraft von Beschreibungen, formale Datenwertstandards spielen dagegen gerade bezüglich allgemeiner Interpretierbarkeit und Interoperabilität eine entscheidende Rolle. Als Kern der elektronischen Neustrukturierung des Bibliografischen Systems kann aber die formale Darstellung der bibliografischen und Normdaten mittels Metadaten schemata gelten. Diese syntaktische Formalisierung wird ständig weiterentwickelt im Hinblick auf die Interoperabilität der Schemata und die Darstellung in allgemeineren Kodierungsstandards, z.B. XML-basierte Formate.

Die hier entworfene Struktur von Bibliografischen Metadaten systemen kann schematisch folgendermaßen dargestellt werden:

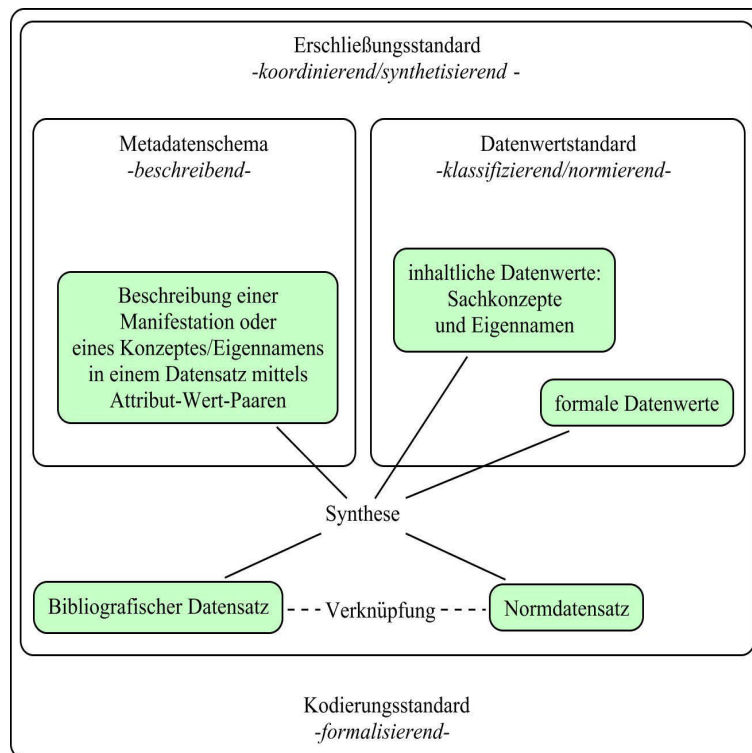


Abbildung 7: Bibliografisches Metadaten system

Bemerkenswert im Vergleich zur Klassischen Bibliografischen Kontrolle in Kapitel 2.5 ist hierbei, dass Normenkontrolle i.e.S. und Kontrollierte Vokabulare stärker in die systematische bibliografische Betrachtung einbezogen werden. Durch die eigene Definition von diesbezüglichen Datenwertstandards sind Normdaten bzw. genormte Konzepte nicht mehr nur Anhängsel einer Titelaufnahme. Aufgrund von Katalogdatenbanken entstanden zwangsläufig eigene Normdatensätze, die, wenn auch nicht inhaltlich, so doch wenigstens strukturell den bibliografischen Datensätzen ebenbürtig sind. Aus heutiger Perspektive besteht darin auch

eine Weiterentwicklung von der dokumentenzentrierten Sicht auf das Bibliografische Universum hin zu einer, die bibliografisch bedeutsame Entitäten außer dem zu erschließenden Objekt kennt. Neben der eigentlichen „Revolution“ der technischen Strukturierung in Attribut-Wert-Schemata erhält die Bibliografische Kontrolle also auch auf der konzeptionellen Ebene einen klaren Zuwachs an Systematik.

Nichtsdestotrotz bleibt ein Hauptproblem dieser Bibliografischen Kontrolle via Metadaten-systeme die ungeheure Vielfalt und damit Inkompatibilität der unterschiedlichen Standards und der dahinter stehenden Datenmodelle. Dies gilt umso mehr, wenn sich Bibliotheken verwandten Bereiche wie Gedächtnisorganisationen, Wissenschaft, Kultur, Medien mit ihren spezifischen Daten öffnen wollen. Für eine solche angestrebte Interoperabilität eignen sich in besonderer Weise system- und softwareunabhängige Datenformate, die in ihrer Elementstruktur beliebig erweiterbar sind. Eine herausragende Rolle spielt in diesem Zusammenhang die *eXtensible Markup Language* (XML).

3.4 XML-kodierte Metadaten

Seit Mitte der 90er Jahre entwickelte sich zur Datenbank-Realisierung von Metadaten die grundsätzliche Alternative der dokumentinternen Auszeichnung von einzelnen Datenelementen und ihren Attributen. Allgemeiner formuliert traten neben die stark formalisierten Kodierungsstandards, solche, die textuell für den Menschen erschließbar sind. [McCathieNevile/Méndez(2007):29] Markup- oder Auszeichnungssprachen bieten den Vorteil, dass sie maschinenlesbar und systemunabhängig einzelne Datenelemente in Dokumenten verarbeitbar machen. Während die HyperText Markup Language (HTML⁸⁷), die eine Grundlage des WWW bildet, auf die Verarbeitung zur Anzeige bzw. dem Layout von Dokumenten abzielt, wurde mit der eXtensible Markup Language (XML⁸⁸) eine Auszeichnungssprache geschaffen, die auf die (ansichtsunabhängige) strukturierte Beschreibung von Dokumenten abhebt:

Both HTML and XML are markup languages; they allow us to publish content and provide information about what role the content plays. However, there are certain differences. [...] XML was designed to describe data and to focus on specifying the nature of the enclosed data. HTML is about displaying information, while XML is about describing information. [Chowdhury/Chowdhury(2007):163]

XML basiert auf einer sehr einfachen Struktur, neben Prolog und optionalem Epilog bilden den Hauptteil eines Dokuments Elemente, die aus Daten mit der dazugehörigen Auszeichnung (Elementnamen oder tags) bestehen, und zwar in der Form <Elementname>Daten</Elementname>. [Antoniou/van Harmelen(2008):26] Die Elemente können auch über Attribute verfü-

87 HTML, aktuelle Spezifikation der Version 4.0 siehe <http://www.w3.org/TR/REC-html40>

88 XML, aktuelle Spezifikation der Version 1.0 siehe <http://www.w3.org/TR/xml/>

gen und als Subelemente ineinander verschachtelt werden – sie sind in einer Baumstruktur angeordnet, gehen also auf ein gemeinsames Wurzelement zurück, das die beschriebene Ressource im Ganzen benennt. [Legg(2007):416] Subelemente und Attribute haben letztlich dieselbe Funktion, die Aufzählung und Beschreibung von einzelnen Merkmalen des Elements, das ihnen direkt übergeordnet ist bzw. dessen Attribut sie sind.

Die Besonderheit von XML besteht darin, dass es die Grundlage für alle möglichen Anwendungen bietet und nicht nur für die Darstellung in einem Browser wie HTML:

XML is a framework that allows users to produce application-specific codes, with markup, so that the tags become meaningful in terms of data and content, thus making the resultant XML documents suitable for machine-processing. [Chowdhury/Chowdhury(2007):163]

Absolut offen ist XML nicht nur gegenüber Verarbeitungsweisen, sondern auch in Bezug auf seine Struktur und die Bedeutung (Semantik) seiner Datenelemente, deshalb wird XML auch als Metasprache bezeichnet:

...it is up to the user to define a vocabulary suitable for the application. Therefore, XML is a metalanguage for markup: it does not have a fixed set of tags but allows users to define tags on their own. [Antoniou/van Harmelen(2008):27]

So müssen jeweils verwendete Elementtypen und Attribute in der document type declaration im Datei-Prolog definiert werden, dies kann aber auch durch den Verweis auf externe Strukturdefinitionen wie Document Type Definitions⁸⁹ (DTD) oder XML-Schema-Dateien geschehen. Diese Schemasprachen definieren eine Klasse von XML-Dokumenten durch Festlegungen der Markup-Elemente, die in diesen XML-Dokumenten verwendet werden dürfen. Das Ergebnis sind zahlreiche XML-Applikationen bzw. XML-Sprachen für spezifische Anwendungsbereiche wie Mathematik, Bioinformatik, Humanressourcen- und Investmentmanagement [Antoniou/van Harmelen(2008):28], aber auch allgemein gebräuchliche Datenformate wie Atom⁹⁰, SVG⁹¹, XHTML, die durch ihre grundlegende XML-Syntax aber grundsätzlich interoperabel bleiben.

Während für DTDs eine eigene Sprache verwendet werden muss, sind XML Schemas ebenfalls in XML abgefasst. Das neuere XML-Schema-Format ist auch das Ausdrucksstärkere, es kann z.B. im Gegensatz zu DTDs festlegen, welche Datentypen für Elemente und Attribute verwendet werden dürfen sowie Reihenfolge und Anzahl von Kindelementen:

XML schemas provide uniqueness constraints and references, which denote specific attributes of the elements that make them unique and relate them to others. They support various data types and name spaces. An XML name space is a collection of names, identified by a URI (uniform resource identifier), which are used in XML documents as element types and attribute names. [Chowdhury/Chowdhury(2007):166]

⁸⁹ DTDs können auch dokumentintern definiert werden, dies wird aber nicht empfohlen.

⁹⁰ Atom Syndication Format, siehe <http://www.atompub.org/rfc4287.html>

⁹¹ ScalableVector Graphics, siehe <http://www.w3.org/TR/SVG11/>

Die Nutzbarkeit von XML-Namensräumen⁹² ist ein weiterer klarer Vorteil vom XML Schema gegenüber DTD. Diese bieten die universelle Möglichkeit, einmal definierte Elemente in eigenen XML-Dokumenten wieder zu verwenden, indem man auf einem Namensraum verweist, der die Namen dieser Elementtypen und Attribute enthält. Üblicherweise sind dies die Namensräume der verschiedenen XML Schemas bzw. XML-Sprachen. So spielen sie für die Modularität von XML-Elementen und allgemein der Konvergenz von XML-Dokumenten eine wichtige Rolle.

Grigoris Antoniou und Frank van Harmelen fassen die Vorteile von XML folgendermaßen zusammen:

XML is the de facto standard for representation of structured information on the Web and supports machine processing of information. XML supports the exchange of structured information across different applications through markup, structure, and transformations. [Antoniou/van Harmelen(2008):59]

Wegen seiner Anwendungs- und Systemunabhängigkeit, der Flexibilität zur umfangreichen Weiterentwicklung und zur Wiederverwendung von Elementen sowie der einfachen – auch „menschenslesbaren“ – Struktur wird XML vermehrt als Grundlage von Metadaten-Systemen gebraucht. So liegen auch bibliografische und Normdaten-Schemata in XML-kodierter Form vor, z.B. MABxml⁹³, XmetaDiss⁹⁴/XmetaDissPlus⁹⁵, MARCXML (Bibliographic and Authorities), MODS und MADS⁹⁶. Dies bildet die Grundlage für die Verwendung von bibliografischen und Normdaten in universellen XML-basierten Umgebungen, z.B. zum Austausch von Daten über das Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH⁹⁷). XML-Stylesheet-Sprachen ermöglichen die Präsentation eines XML-Dokuments, z.B. als unterschiedlich formatierte HTML- oder Textdokumente sowie die Transformation in andere XML-Schemata. [Antoniou/van Harmelen(2008):53-54] Die von der LoC entwickelten MARCXML Stylesheets bieten so Konvertierungen in andere XML-basierte Formate, etwa ONIX⁹⁸ und Dublin-Core-Varianten.

Die besondere Eignung der Struktur-Auszeichnungssprache XML für Metadaten wird besonders klar, wenn man sich die grundlegenden Anforderungen an diese vor Augen führt:

It is becoming generally accepted in the information community that interoperability is one of the most important principles in metadata implementation. Other basic metadata principles include simplicity, modularity, reusability, and extensibility. [Chan/Zeng(2006)]

92 Namespaces in XML, siehe <http://www.w3.org/TR/REC-xml-names/>

93 MABxml, siehe <http://www.d-nb.de/standardisierung/formate/mabxml.htm>

94 XmetaDiss, siehe <http://www.d-nb.de/standards/xmetadiss/xmetadiss.htm>

95 XmetaDissPlus, siehe <http://metadaten-twr.org/2009/11/09/xmetadissplus-fur-hochschulschriften-und-publicationen/#more-231>

96 MARCXML und Konvertierungsmöglichkeiten, siehe <http://www.loc.gov/standards/marcxml/> MODS, siehe <http://www.loc.gov/standards/mods/> und MADS, siehe <http://www.loc.gov/standards/mads/>

97 OAI-PMH, siehe <http://www.openarchives.org/OAI/openarchivesprotocol.html>

98 ONline Information eXchange (ONIX), siehe <http://www.editeur.org/8/ONIX/>

Alle diese Kriterien unterstützt XML mit seinen Derivaten in bestechender Weise. Zu bedenken bleibt aber, dass XML für sich nur Grundlage der Austauschbarkeit auf der Kodierungsebene legt. Ohne eine entsprechende Kompatibilität auf der Ebene der Metadatenschemata und Datenwertstandards stellt dies jedoch nur eine oberflächliche Interoperabilität von Metadaten sicher. Allgemeiner gesprochen, unterliegen XML und seine Derivate folgender Bedingung:

Collaboration and exchange are [only, K.W.] supported if there is an underlying shared understanding of the vocabulary. XML is well-suited for close collaboration, where domain-or community-based vocabularies are used. [Antoniou/van Harmelen(2008):61]

Einheitliche oder zumindest kompatible syntaktische und semantische Datenstrukturen auf Schema-/Datenwert-Ebene sind somit für die allgemeine Interoperabilität von bibliografischen Daten auch mit XML-Kodierung ein entscheidendes Thema – darauf wird zurückzukommen sein.

3.5 Dublin Core Metadata Element Set (DCMES)

Das Dublin Core Metadata Element Set (DCMES), oft auch kurz Dublin Core (DC) genannt, stellt einen Ansatz dar, disziplinübergreifend eine grundlegende Ressourcen-Beschreibung zu bieten, also ein Metadatenschema für Dokumente im WWW in allgemeinsten Form. Einfach anzuwenden ist Dublin Core schon aus dem Grund, dass alle Elemente in einem Metadatensatz optional sind und mehrfach verwendet werden können, wobei auch die Reihenfolge der Elemente variabel ist. Weitere grundlegende Prinzipien sind die Möglichkeit der Modifikation der einzelnen Elemente und die Erweiterungsmöglichkeit durch Elemente von anderen Schemata an sich. [Lagoze/Lynch/Daniel(1996):4-5] In der genuinen XML-Kodierung baut Dublin Core hier auf den Vorteilen von XML auf: Neben den Namensräumen des DCMES und der DC Terms kann beliebig auf weitere XML-Namensräume mit anderen Metadatenelementen verwiesen werden.

Aus bibliografischer Sichtweise wurde Dublin Core lange als zu grob und wegen der fehlenden Erschließungsregeln und der laxen inhaltlichen Definition der Datenelemente gescholten. So sprach Michael Gorman noch im Jahr 2000 diesbezüglich von einem „unstandardized, uncontrolled, ersatz cataloguing“ und von Metadaten allgemein als „a fancy name for an inferior form of cataloguing“. [Howarth(2005):52, Notes] Dem gegenüber wurde aus Sicht der praktischen Metadatenentwicklung argumentiert, dass Metadaten, und als herausragendes Beispiel Dublin Core, eben nicht in der Tradition der Katalogisierung stehen und weder deren Vorgehensweise noch Zielsetzung teilen. Wie Gradmann bereits 1998 versucht hat, klar zu stellen, weisen die Ergebnisse von computergestützter Katalogisierung und Metadatenproduktion mittels DC in ihren Ergebnissen zwar semantische Ähnlichkeiten auf, aber

...the original focus of the DC initiative was '*to facilitate resource discovery in a networked environment*' (Lagoze 1997) and thus not primarily resource **description**'. [Gradmann(1998):4, Hervorhebungen Gradmann]

Genauer beschreibt Gilliland die Absicht, die hinter Dublin Core steht bzw. die gewollte einfache Struktur von Dublin Core, die seinen besonderen Nutzen ausmacht:

...the Dublin Core Metadata Element Set (DCMES) identifies a relative small, generic set of metadata elements that can be used by any community, expert or non-expert, to describe and search across a wide variety of information resources on the World Wide Web. Such metadata standards are necessary to ensure that different kinds of descriptive metadata are able to interoperate with one other and with metadata from nonbibliographic systems [...]. [Gilliland(2008):8]

Mittlerweile hat die bibliothekarische Welt verstanden, dass mit Dublin Core keine Alternative zur Katalogisierung forciert war und schätzt⁹⁹ es als nützliches Minimaldatenset dort, wo keine sonstigen bibliografischen Metadaten erhoben werden, oder wo diese immer noch inkompatibel sind – so stellt DC heute z.B. das Minimalformat für das OAI-PMH dar, das ein allgemein anerkanntes Austauschformat von Digitalen Bibliotheken ist. Ein anderes, aktuelles Beispiel für die bibliothekarische Nutzung von Dublin Core stellt das Browser-Plugin Connexion des OCLC dar, das die automatische Extraktion von DC-Metadaten (Simple DC oder Qualified DC) aus verschiedenen Online-Ressourcen (neben HTML-, PDF-, DOC- auch MP3-Dateien) bietet. Da eine Abbildung von Dublin-Core-Elementen auf MARC-Felder vorliegt, kann dies als Grundlage zur automatischen Erstellung von MARC-Datensätzen genutzt werden. [Gonzalez(2008):140]

Angesichts einer überbordenden Medien- und Formatvielfalt erweist sich das DC-Schema auch abseits der bibliografischen Sphäre als ein zentraler Schlüssel zur minimalen Erschließung von Webdokumenten. Aus diesem Grund ist Dublin Core mittlerweile als Standard von der ANSI/NISO, ISO und Internet Engineering Task Force (IETF)¹⁰⁰ anerkannt und eine offizielle deutsche Übersetzung wurde schließlich 2007 vom Kompetenzzentrum für Interoperable Metadaten (KIM) erstellt¹⁰¹. Es wird zudem von sieben Staaten als Metadatenstandard für Regierungsdokumente verwendet sowie von einigen Organen der Vereinten Nationen, wie WHO, UNEP, FAO¹⁰² sowie der European Environmental Agency. Aus Gründen seiner „All-

99 Es sei angemerkt, dass der oben zitierte Michael Gorman schon zwei Jahre später eine qualitative Weiterentwicklung des Metadatenkonzeptes für möglich hielt, die dieses ununterscheidbar von „richtiger“ Katalogisierung macht, vgl. [Liu(2007):20].

100 NISO Z39.85-2007, siehe <http://www.niso.org/standards/z39-85-2007/> und ISO 15836:2009 siehe <http://www.iso.org/iso/search.htm?qt=15836&searchSubmit=Search&sort=rel&type=simple&published=on> sowie IETF RFC 5013, siehe <http://www.ietf.org/rfc/rfc5013.txt>

101 Dublin-Core-Metadaten-Elemente-Set, http://kim-forum.org/material/pdf/uebersetzung_dcmes_20070822.pdf

102 World Health Organization (WHO), United Nations Environmental Program (UNEP), Food and Agriculture Organization (FAO)

gemeinverständlichkeit“ wird das DCMES häufig in umfangreichere Metadatenysteme integriert, wie etwa in PRISM¹⁰³ und ETD-MS¹⁰⁴. [Harris/Weibel(2004)]

Des Weiteren hat die Dublin Core Metadata Initiative (DCMI) auf Kritik reagiert und schon ein Jahr nach der Veröffentlichung von DC 1995 mit seinen 15 (anfangs 13) grundlegenden Elementen betonte man, dass je nach Kontext unterschiedliche Metadaten schemata für ein Objekt geeignet sein können. Mit dem Warwick-Framework entwarf man ein Container-Format, in dem verschiedene Metadaten sets von DC und MARC für ein Objekt zusammengefasst werden können. [Lagoze/Lynch/Daniel(1996):9-12] In den folgenden Jahren bis 2000 wurde das nun Simple Dublin Core genannte Schema fortlaufend verfeinert. Qualified Dublin Core verfügt über *Element Refinements* zur Spezifizierung von Datenelementen (also Qualifikatoren) und die zusätzlichen Datenelemente Audience, Provenance and RightsHolder. Zudem werden *Vocabulary Encoding Schemes* (z.B. das DCMI Type Vocabulary, die Dewey Decimal Classification und Library of Congress Subject Headings) und *Syntax Encoding Schemes* (z.B. DCMI Period Encoding Scheme, Sprach- und Ländercodes von ISO und IETF, W3C Date and Time Formats Specification) als Datenwertstandards festgelegt.

Dublin Core stellt mittlerweile eher eine Familie von Metadaten schemata dar, wenn man die Application Profiles (DC-APs) betrachtet, die seit dem Jahr 2000 für unterschiedlichste Objekte oder Einsatzgebiete von Dublin Core Communities entwickelt werden. Für den bibliografischen Bereich ist hier vor allem das DC Lib¹⁰⁵, also das Application Profile für Bibliotheken zu nennen. Der grundsätzliche Gedanke der Modularität und Flexibilität, wie er in den Grundprinzipien, dem Warwick-Framework und dem Qualified DC zum Ausdruck kommt, wird hier noch einmal erweitert mit der Aufforderung an Fachcommunities, ihre spezifischen Vokabulare einzubinden und eigene Erschließungsregeln aufzustellen. Mit diesem Anwachsen des Dublin-Core-Vokabulars (vgl. DCMI Metadata Terms¹⁰⁶) und vor allem durch die W3C-Entwicklung des Resource Description Frameworks (RDF) seit 1997 wurde schließlich eine allgemeine und integrative Modellierung des „Dublin-Core-Universums“ notwendig, das sogenannte *DC Abstract Model*, auf das in Kapitel 5.1 eingegangen wird.

103 Publishing Requirements for Industry Standard Metadata (PRISM), ein umfangreiches, XML-basiertes Metadatenvokabular für Zeitschrifteninhalte, siehe <http://www.prismstandard.org>

104 Electronic Thesis and Dissertations Metadata Standard (ETD-MS), siehe <http://www.ndltd.org/standards/metadata/etd-ms-v1.00-rev2.html>

105 DC-Library Application Profile, siehe <http://dublincore.org/documents/library-application-profile/>

106 „...an up-to-date specification of all metadata terms maintained by the Dublin Core Metadata Initiative, including properties, vocabulary encoding schemes, syntax encoding schemes, and classes“, siehe <http://dublincore.org/documents/dcmi-terms/>

3.6 Kritik: Bibliografische Metadatensysteme

Charles McCathie und Eva Méndez fassen in anschaulicher Weise zusammen, welches Umdenken die Neuorientierung der Bibliografischen Kontrolle hin zur komplexen Ressourcenbeschreibung durch Metadatensysteme erfordert:

Information professionals, librarians, and specialised cataloguers are learning to adapt their skills in processing of books to *technological processing of „resources“*. *Traditional cataloguing* blends the main elements: tangible documents, processes (bibliographic description, content analysis—mainly indexing and abstracting), and products, with are fruits of this activity (typically bibliographic records collected as a catalogue). *Resource description* implies: documents like information objects, the creation-automatic or manual-of information about the attributes of those objects (metadata), making metadata records and metadata repositories. [McCathieNevile/Méndez(2007):24, Hervorhebungen K.W.]

Man kann wohl durchaus behaupten, dass diese Neuausrichtung der Bibliografischen Kontrolle im Allgemeinen gelungen ist, betrachtet man die Verbreitung von OPACs, die permanente Weiterentwicklung von bibliografischen Datenformaten und Datenbanksystemen, das Bewusstsein für Probleme der elektronischen Langzeitarchivierung sowie das aktuelle Bestreben der freien Verfügbarmachung von Daten zum Download.

Und doch hat die Entwicklung der Bibliografischen Kontrolle zu Metadatensystemen auch einige Nachteile gebracht. Die Bibliothekskataloge an sich (Band- und Kartenkatalog, Katalogdatenbank und Online Public Access Catalog, OPAC) unterlagen in kurzer Zeit einem enormen strukturellen und funktionellen Wandel. [Bianchini/Guerrini(2009):107-108] Zwar revolutionierte die computergestützte Verarbeitung Verfügbarkeit, Austauschbarkeit, Retrieval und Manipulation von bibliografischen Daten, jedoch wurde dies mit einem Verlust an bibliografischer Struktur erkauft. [Svenonius(2001):62-64] Während im Bandkatalog verschiedene Ausgaben, Übersetzungen und Exemplare eines Werkes in hierarchischer Darstellung beieinander zu finden waren (Kollokationsprinzip) und z.T. sogar Werk-Werk-Beziehungen (z.B. ausgewählte Kommentare) über Verweisungen dargestellt wurden, begann sich diese Struktur in Kartenkatalogen bereits zu lösen. Mit der Schaffung von unabhängigen Titel- und Normdatensätzen verschwand sie vollends. Elaine Svenonius zieht eine harte, aber im Kern bedenkenswerte Bilanz der Katalog-Modernisierung, wenn sie schreibt:

The tiered structure, so neatly displayed in book catalogs and more or less preserved by strict filing rules in card catalogs, has been lost almost altogether by the inevitably move to computer filing. [...] In any case, technological advance has brought with it a steady deterioration in the integrity bibliographic structures, [...] and with it, a undermining of bibliographic objectives. [Svenonius(2001):64]

Insbesondere betrifft dies die bis dato teilweise dargestellten bibliografischen Beziehungen zwischen Werken bzw. Exemplaren. Barbara B. Tillett hat bereits 1987 durch Untersuchungen von Katalogisierungsregelwerken sieben solcher Beziehungstypen identifiziert, die später in die Ausarbeitung der FRBR-Modell-Familie einfließen. Zusammengefasst unterscheidet sie [Tillett(2001):30-31]:

- primary relationships, die der konzeptionellen Unterscheidung von work, expression, manifestation, item inhärent sind;
- close content relationships, darunter equivalence, derivative, descriptive relationship;
- whole-part and part-to-part relationships, unter letzterer accompanying und sequential relationship;
- shared characteristic relationships „between different works¹⁰⁷ that share any element in common (other than an element described by one of the other types of relationships)“ [Tillett(2001):27, Hervorhebung Tillett]

Während Beziehungen der dritten Gruppe mittlerweile in Online-Katalogen nachvollziehbar dargestellt werden, und die letzte Gruppe von Koinzidenzen von Merkmalen (letztendlich also nicht-normierte Sucheinstiege) ganz dem Metadatenprinzip entspricht, sucht man Darstellungen der Beziehungen der ersten und zweiten Gruppe bisher meist vergebens. Dabei sind gerade sie es, die einen besonders engen inhaltlichen Zusammenhang zwischen Werken/Exemplaren widerspiegeln, insofern also für eine inhaltliche Suche von enormem Interesse sind.

Durch die Dekonstruktion der klassischen Titelaufnahme in bibliografische und Normdatenelemente wurden diese Elemente zwar eindeutiger darstellbar, auf der anderen Seite wurden jedoch bestimmte Arten von bibliografischen Beziehungen noch weniger sichtbar.

Ein zweites zentrales Problem der Metadatenysteme besteht in ihrer mangelnden Interoperabilität und Flexibilität. Durch die Festlegung eines Metadatenschemas sind Umfang und Struktur der Metadaten im Nachhinein nur noch schwer manipulierbar, so können etwa nur mit großem technischen Aufwand weitere bibliografische Elemente hinzugenommen werden. Zwar sind einzelne bibliografische Elemente in Metadatenystemen eben sichtbar, sie sind aber auch notwendig an diese gebunden:

...metadata is defined as a particular record structure, and the data elements are neither extendible nor reuseable outside of that structure. [Coyle(2010b):26]

Das grundsätzliche Problem, dass unterschiedliche Schemata und Datenformate die Austauschbarkeit behindern, wird dadurch noch verstärkt. Wie gezeigt, ermöglichen XML und seine Derivate zwar eine einheitliche und erweiterbare Kodierung von Metadatenstandards. Unter dieser Kodierungsebene bestehen jedoch die unterschiedlichen Metadaten schemata und Datenwertstandards weiter. Daher sind sie auf aufwändige, intellektuelle Verfahren zur Herstel-

¹⁰⁷ Barbara B. Tillett hebt aus unerklärten Gründen den Begriff work hervor, spricht danach und an anderer Stelle aber allgemein von bibliographic entities und bringt auch entsprechende Beispiele.

lung einer syntaktischen und semantischen Interoperabilität angewiesen, die Chan/Zeng in ihrem Aufsatz überblicksartig vorstellen:

At the schema level, interoperability actions usually take place before operational level metadata records are created. The actions focus on the elements (independent of individual applications). Methods used to achieve interoperability at this stage mainly include: derivation, application profiles, crosswalks, switching-across, framework, and registry. [Chan/Zeng(2006): Abschnitt 4]

Eine letzte große Hürde stellt die „Gatefunktion“ des Onlinekatalogs und die dynamische Generierung von Suchergebnissen dar. Bibliografische und Normdaten sind nicht wirklich als „im World Wide Web angekommen“ zu bezeichnen, wenn der Zugang zu ihnen nur über ein Datenbank-Interface möglich ist. So elaboriert bibliografische Metadatensysteme auch sein mögen, ihre Funktionalität steht und fällt heute mit ihrem unmittelbaren Vorhandensein im WWW, wie Karen Coyle eindringlich schreibt:

In spite of this legacy of pre-computer practice, the question that we must face, and that we must face sooner rather than later, is how we can best transform our data so that it can become part of the dominant information environment that is the Web. This is a radical change in the context for library metadata, yet it is a logical extension of the design for sharing that has been a principle of library cataloging. [Coyle(2010a):11]

Stefan Gradmann hat bereits 2005 auf dieses zentrale Manko der Online-Kataloge hingewiesen – und er dachte gleichzeitig zu Ende, was die Veröffentlichung der heutigen Bibliotheksdaten bedeuten würde:

...bibliographic information originated by libraries still largely remains buried within the 'hidden Web' – and that, as long as different layers of information remain blended in bibliographic records, the non-librarian world probably is better off without these thousands of identical bibliographic records pointing simply to different items or manifestations and thus 'polluting' search engine results with massive amounts of redundant information. [Gradmann(2005):64]

Ein Überfluss an – zum großen Teil – redundanten Daten würde die präzise Suche und automatische Verarbeitung ungemein erschweren. Die eigentliche Problematik besteht darin, dass diese Redundanz nicht ohne Weiteres zu bereinigen ist. Zum einen sind da die verschiedenen Datenformate; zum anderen wird, selbst wenn die gleichen Formate vorliegen, das Herausfiltern von redundanten Datensätzen bereits bei minimalen Unterschieden in der Erschließung zum Problem. Zudem stellt sich die Frage, ob der normale Webnutzer überhaupt von den teilweise sehr komplexen bibliografischen Metadatenformaten profitieren würde, zumindest sind sie aber ohne ausführliche Dokumentation und eine spezifische Darstellungsstruktur schwer nutzbar.

Bibliografische Metadatensysteme weisen also einerseits immanente Probleme auf, wie die Zersplitterung in eine Vielzahl von inkompatiblen und nicht flexiblen Standards und den Verlust von bibliografischen Strukturen gegenüber Band- und Kartenkatalogen. Eine besondere Herausforderung stellt aber die mangelnde Zugänglichkeit und Interoperabilität der so

strukturierten Metadaten im WWW dar. Man kann diesem Problem wohl eine gewisse Katalysatorfunktion in der Reformierung des Bibliografischen Kontrolle zusprechen, da sich gerade Bibliothekare der herausragenden Bedeutung der Informationsstruktur Internet bewusst sind.

Abseits der bibliothekarischen Diskussion wurde der Metadaten-Begriff von den Entwicklern des WWW schon früh als ein weitergehendes Konzept betrachtet. So hat Tim Berners-Lee bereits 1997 postuliert, dass der damalige (und auch noch derzeitige) Metadaten-Begriff nur der Anfang einer Entwicklung hin zu allgemein maschinenverständlichen Metadaten sein kann:

Metadata was called 'Metadata' because it started life, and is currently still chiefly, information about web resources, so data about data. In the future, when the metadata languages and engines are more developed, it should also form a strong basis for a web of machine understandable information about anything: about the people, things, concepts and ideas. We keep this fact in our minds in the design, even though the first step is to make a system for information about information. [Berners-Lee(1997)]

Und hier liegt eben hier der Schlüssel nicht nur für eine Weiterentwicklung des Metadaten-Konzept an sich, sondern auch für bibliografische Metadatensysteme im Besonderen. Unter dem Begriff Semantic Web werden die aktuellen Entwicklungen von Web-Technologien zusammengefasst, die diese Vision von Metadaten als maschinenverständliche Informationen wahr machen soll.

4 Semantische Technologien im World Wide Web

4.1 Vom WWW zum Semantic Web

Tim Berners-Lees Entwurf [Berners-Lee(1989)] eines auf dem Hypertext-Modell¹⁰⁸ beruhenden „universal linked information system“ für das CERN bildete die Grundidee des heutigen World Wide Webs. Das Ziel Berners-Lees war es, die Zersplitterung von Informationen in diskreten Datenbanken und Informationssystemen mittels übergreifenden Verbindungen zu überwinden:

There are few products which take Ted Nelson's idea of a wide "docuverse" literally by allowing links between nodes in different databases. In order to do this, some standardisation would be necessary. [Berners-Lee(1989)]

Diese Standardisierung soll durch Frontends erfolgen, welche die allgemeine Verarbeitung und Übertragbarkeit der Informationen ermöglichen:

The only way in which sufficient flexibility can be incorporated is to separate the information storage software from the information display software, with a well defined interface between them. [Berners-Lee(1989)]

Dieses Grundprinzip einer Darstellung in einem allgemeinen, verbindlichen Datenformat findet sich noch heute im HTML/HTTP-basierten¹⁰⁹ WWW. Zum Erfolg des WWW trug wesentlich bei, dass nicht nur seine Standards „offen“ sind, also für Jedermann frei verwendbar, sondern auch das sich entwickelnde Netz selbst, also das sogenannte AAA-Prinzip (Anyone can say Anything about Any topic):

It was this freedom of expression on the document Web that allowed it to take off as a bottom-up, grassroots phenomenon. [Allemang/Hendler(2008):7-8]

Die mittlerweile „Web 1.0“ genannte klassische WWW-Struktur ist dabei auf statische HTML-Dokumente ausgerichtet, deren Syntax – wie im Vergleich mit XML beschrieben – auf die Darstellung und kaum auf inhaltliche Auszeichnung ausgerichtet ist. Zudem sind diese Webseiten recht unflexibel und werden nur von einem kleineren Teil der Webnutzer produziert, da dies eine Erstellung, Bearbeitung und Versionierung per Hand oder zumindest die Bedienung eines HTML-Editors verlangt. Auch aus diesem Grund spricht man von dieser Generation des WWW als „Read-Only-Web“ oder „Document Web“.

Als „Web 2.0“ oder „Read-Write-Web“ werden dagegen jüngeren Web-Applikationen bezeichnet, die eine schnelle und unkomplizierte Herstellung von Webinhalten ohne Code-Kenntnisse für Jeden erlauben. Diese Flexibilisierung verändert auch den Charakter von Webdokumenten. Sie sind nicht nur leichter veränderbar und versionierbar, sondern bieten

108 Der Begriff des Hypertextes geht auf Theodor Holm Nelson zurück. In seinem Aufsatz „A file structure for the complex, the changing and the indeterminate“ von 1965 werden die Begriffe Hypertext und Hypermedia eingeführt. In den Büchern „Computer Lib / Dream Machines“ und „Literary Machines“ führt er die Idee seines Xanadu genannten weltweiten Hypertext-Netzwerkes weiter aus.

109 HyperText Markup Language (HTML), HyperText Transfer Protocol (HTTP)

die Möglichkeit, einfach andere Informationen einzubeziehen und miteinander zu kombinieren. Während man das Web 1.0 als eine dokumentenzentrierte Struktur ansehen kann, ist das Web 2.0 durch eine stärkere Nutzerzentrierung geprägt: Webinhalte werden in ihrer Form leichter manipulierbar, kleinteiliger und kurzlebiger (kurz: „liquider“), in den Vordergrund rückt die aktuelle Verarbeitung von einzelnen Informationen durch Web-Nutzer.

Auch wegen dieser begrüßenswerten „Demokratisierung“ des Internets verschärfen sich die grundsätzlichen Probleme des Information-Retrieval über Webdokumente bzw. ihrer Erschließung. Einerseits ist die rapide wachsende Größe des WWW an sich ein Problem, andererseits erscheint die text- bzw. wortbasierte Suche aufgrund der sprachlichen und kulturellen Heterogenität nur wenig erfolgversprechend hinsichtlich Precision und Recall. [Chowdhury/Chowdhury(2007):199-200] Eher sporadisch werden Metadaten als Erschließungsmittel benutzt, sie beschränken sich jedoch oft auf einfache formale Merkmale. Selbst wenn eine inhaltliche Erschließung erfolgt, wird dies nur durch grobe Schlagwörter und oft mit unkontrolliertem Vokabular realisiert.

Einen verbindlicheren Ansatz zur Beschreibung von Webdokumenten hat zumindest die Dublin Core Metadata Initiative in Form von allgemein anwendbaren Metadatenschemata und populären Datenwertstandards entwickelt. Die DC-Metadaten werden zumeist durch interoperable, selbstbeschreibende XML-Sprachen in die Webdokumente eingebettet. Ihre inhaltliche Tiefe wurde zwar durch das Qualified-DC-Elemente-Set und die DCMI Metadata Terms entscheidend erweitert, ohne eine breite Absicht und das Vermögen der Web-Produzenten zur intellektuellen Metadatengenerierung bleibt jedoch das durchdachtste Metadaten-system wirkungslos.

Prinzipiell lässt sich also feststellen, dass die webspezifischen Erschließungstechniken und die allgemeine Dynamik des WWW in den letzten Jahren eine ungleichzeitige Entwicklung genommen haben. Während die Metadaten-Community zu elaborierteren Beschreibungssystemen und stärkerer Standardisierung neigt, wächst das WWW schneller und weniger reglementiert denn je. So ergibt sich langsam die Einsicht, dass die Erschließung durch Metadaten (als Ergänzung zur textbasierten Suche) neue Methoden entwickeln muss, die der schieren Masse und Diversität von Webdokumenten entsprechen.

4.1.1 „Semantische“ Informationen

Die Idee des Semantic Webs versucht diese Probleme der Erschließung mit den Mitteln des Netzes selbst zu lösen. XML-basierte Struktur- und Metadatenauszeichnungen sind insofern grundlegend, da eine weltweite Erschließung nur durch die Dokumente selbst gesichert werden kann. Wie aber werden Metadaten eindeutig und trotz der Größe des WWW funktionell?

Auch hier ist die Vernetzung die zentrale Antwort. *Eindeutigkeit* oder Identität von Begriffen lässt sich mit im Web hinterlegten Definitionen sichern, als Namespace-Prinzip ist dies z.B. für XML-Sprachen und das Dublin-Core-Vokabular schon länger im Gebrauch. Egal ob es um formale oder inhaltliche Konzepte geht, ein über das Netz *referenzierbarer* Nachweis (reference) ist die einfachste Möglichkeit, diese eindeutig zu identifizieren und so allgemein nutzbar zu machen.

Das erste grundlegende Technologie des Semantic Webs ist somit die Verwendung von sogenannten Uniform Resource Identifiers (URIs)¹¹⁰, also einheitlichen Identifikatoren von Ressourcen:

...the syntax of the URI makes it possible to 'dereference' it – that is, to use all the information in the URI [...] to locate a file (or a location in a file) on the Web. [Allemang/Hendler(2008):40]

Neben der Klärung der Konzepte für den menschlichen Nutzer macht eine solche Referenzierung auch für automatische Systeme (sogenannte Agenten) „Sinn“. Sie verstehen zwar nicht die Definition oder Erläuterung, die dort gegeben wird, wohl aber, dass alles, was auf dieses Konzept referenziert, „das Selbe“ sein muss. Dies ist gemeint, wenn man von maschinenverarbeitbarer, also automatisiert herstellbarer, Semantik¹¹¹ spricht. Dass *Maschinenverarbeitbarkeit* von Informationen von entscheidender Bedeutung für die weitere Entwicklung des WWW ist, erläutert Catherine Legg anschaulich:

The web has largely followed human rather than machine formats, resulting in what is essentially a large, hyperlinked book. The Semantic Web aims to bridge this gap between human and machine readability. Given the enormous and ever-increasing number of Web pages [...], this has the potential to open unimaginable quantities of information to any applications able to traffic in machine-readable data and, thereby, to any human able to make use of such applications. [Legg(2007):410]

Damit ist die zweite grundlegende Technologie des Semantic Webs benannt, eine maschinenverarbeitbare Form von Metadaten. Spätestens hier mag man einwenden, dass klassische Metadaten, egal ob in originären Formaten oder in XML kodiert, doch schon „maschinenverarbeitbar“ wären. Aber, wie es McCathie/Méndez es prägnant fassen,

The Semantic Web is build on machine-understandable data, not just machine-readable data (MARC). [McCathieNevile/Méndez(2007):28]

Die Unterscheidung zwischen den Konzepten der maschinenverarbeitbaren (z.T. auch „maschinenverständlich“ genannt) und maschinenlesbaren Metadaten mag auf den ersten Blick kleinlich erscheinen, sie ist jedoch für die Funktion des Semantic Webs entscheidend. Die Maschinenverarbeitbarkeit rekuriert auf die beschriebene Selbstreferenzialität des Semantic

110 URI, siehe <http://www.ietf.org/rfc/rfc3986.txt> Mittlerweile gibt es auch den allgemeineren Standard für Internationalized Resource Identifiers (IRI, siehe <http://www.ietf.org/rfc/rfc3987.txt>), bei denen der US-ASCII-Satz auf Unicode/ISO 10646 ausgedehnt wurde.

111 „Semantik als Teil der Sprachwissenschaft beschäftigt sich damit, unter welchen Bedingungen eine sprachliche Struktur eine Bedeutung hat.“ [Strauch(2004):Semantik, S. 110]

Webs durch URIs, welche die Eindeutigkeit und damit Bedeutung der Metadatenkonzepte (also von Elementen und Werten) universal sichert:

A critical part of the design of the whole system is the *way* that the semantics of metadata or indeed of data are defined. [...] In the future we would like to move toward a state in which *any* metadata or eventually any form of machine readable data *carries a reference* to the specification of the semantics of all the assertions made within it. [Berners-Lee(1997), Hervorhebungen K.W.]

Maschinenverarbeitbarkeit meint hier eben nicht nur eine allgemein verbindliche Auszeichnungssyntax, sondern setzt eine eindeutige Identifizierung bzw. Referenzierung aller Konzepte voraus – und damit eine Semantik, die Maschinen genuin „verstehen“.

Der Begriff „Semantic Web“ meint im Kern eben die Anreicherung des WWW mit Informationen, die durch technische Systeme automatisch auswertbar sind, wie Tim Berners-Lee bereits 1999¹¹² schrieb:

The first step is putting data on the Web in a form that machines can naturally understand, or converting it to that form. That creates what I call a Semantic Web – a web of data that can be processed directly or indirectly by machines. [Berners-Lee/Fischetti(2000):191]

Dieser Idee folgte eine breite theoretische und praktische Entwicklungsarbeit; semantische, d.h. maschinenverarbeitbare Informationen werden aber nach wie vor als das zentrale Prinzip des Semantic Webs angesehen. Dies wird in der ungleich exakteren Definition deutlich, die Stefan Gradmann in seiner Antrittsrede 2008 formuliert hat:

Ist doch das Semantic Web in seiner gegenwärtigen Ausprägung vor allem der Versuch, Informationsinstanzen im Netz und diese zum World Wide Web verbindende Graphen in einer Weise modellierbar zu machen und mit formalisiert-differenzierten Attributen zu versehen, die letztlich diese Informationsinstanzen durch Maschinen nach Regeln prozessierbar machen. Und diese Regeln müssen in logikbasierten Sprachen formulierbar sein, um für Maschinen ausführbar zu sein. [Gradmann(2009):14]

Die Sprache von semantischen *Informationen* beruht dabei auch auf dem grundlegenden Unterschied zu bisherigen Metadaten, *was* Objekt der Beschreibung ist. Der Zweck konventioneller Metadaten ist per definitionem die Ressourcenbeschreibung, wobei Ressource hier als ein konsistentes Aggregat von Informationen, z.B. ein WWW-Dokument, verstanden wird. Damit stellt diese Ressourcenbeschreibung zum großen Teil eine formale Erschließung dar sowie zu einem kleineren Teil eine „flache“, da auf der Gesamtschau des Aggregats beruhende, sachliche Erschließung.

Sinn einer semantischen Auszeichnung, wie sie das Semantic Web ermöglichen soll, ist dagegen die „tiefe“ Annotation der einzelnen Bestandteile dieses Aggregats, also aller in ihr auftauchenden Konzepte, Zusammenhänge und Aussagen. [Coyle(2010b):12] Dementsprechend wird der Begriff Ressource im Semantic Web eben auf diese Sinnbestandteile bezogen. Die zu beschreibenden bzw. referenzierenden Ressourcen sind nicht mehr die Informationsaggregate, sondern die *atomaren* Daten und Konzepte *in* den Aggregaten.

112 Jahr der Erstausgabe

Zwar bauen semantische Metadaten, wie im Folgenden ausgeführt, auf bekannten WWW- und XML-Strukturen auf, das Verständnis für ihren grundsätzlich neuen Charakter – eindeutig und referenzierbar, maschinenverarbeitbar, atomar – ist jedoch zwingend für das Verständnis der Funktionsweise des Semantic Webs insgesamt.

4.1.2 Ein universelles Beschreibungsmodell

Den Begriff semantische *Metadaten* zu benutzen, ist nicht nur wegen der beschriebenen Unterschiede zu klassischen Metadaten heikel. Denn das Semantic Web kennt eigentlich kein Metadaten-Modell, d.h. es wird kein Metadatenschema oder Metadatensystem definiert, sondern es wird ein Metamodell, ein allgemeinstes Beschreibungsmodell, zugrunde gelegt. Tim Berners-Lee hat diese Rückführung des Metadaten-Konzeptes auf Beschreibungsaussagen bereits 1997 beschrieben:

Metadata consists of assertions about data, and such assertions typically, when represented in computer systems, take the form of a name or type of assertion and a set of parameters, just as in the natural language a sentence takes the form of a verb and a subject, an object and various clauses. [Berners-Lee(1997)]

Das vom W3C entwickelte Datenmodell Resource Description Framework (RDF)¹¹³ (1999, 2004) definiert als elementare Struktur Aussagen (statements) in Form von Subjekt, Prädikat und Objekt. Der Name RDF verweist – wie die klassische Metadaten-Definition – auf die Beschreibung von Ressourcen. Anfangs nur explizit als Begriff für Webdokumente verwendet [Berners-Lee(1997)], werden hier alle denkbaren Konzepte als *Ressource* aufgefasst und mittels statements beschreibbar gemacht. Eine *resource* muss im RDF-Modell nur eine Bedingung erfüllen: Sie muss eindeutig referenziert sein, also durch einen URI-Rückverweis eindeutig identifiziert werden. Dies gilt verpflichtend für Subjekte und Prädikate, z.T. auch für die Objekte in RDF-Aussagen.

Für die Arbeit mit Metadaten-Systemen stellt dies eine äußerst nützliche Konstruktion dar, da so in einfacher Weise Metadatenelemente mittels URI-Referenzierung eindeutig definiert werden können. Zugleich bietet die angedeutete RDF-Syntax eine sehr einfache Möglichkeit, metadatenspezifische Attribut-Wert-Paare zu formulieren. RDF kann folglich als Metastruktur für klassische Metadatenstandards verwendet werden und bietet damit einen vielversprechenden Ansatz zur Unterstützung ihrer Interoperabilität. [Haynes(2004):41]

Trotzdem sollte RDF aber nicht nur als übergreifende Metadaten-Beschreibungssyntax begriffen werden; seine Struktur wurde dagegen viel allgemeiner für die Beschreibung von jedweden Konzepten bzw. Daten entworfen. Damit zielt RDF auf ein erstes Ziel der Semantic-

¹¹³ Siehe die allgemeine Übersichtsseite zu RDF des W3C, <http://www.w3.org/RDF/>

Web-Vision ab, eine WWW-inhärente Datenschicht zu schaffen, mit der die einzelnen Konzepte in den Dokumenten, anstatt derer, sinnvoll verknüpft werden können:

The main idea of the Semantic Web is to support a distributed Web at the level of data rather than at the level of presentation. Instead of having one webpage point to another, one data item can point to another, using global references called Uniform Resource Identifiers (URIs). The Web infrastructure provides a data model whereby information about a single entity can be distributed over the web. [...] The single, coherent data model for the application is not held inside one application but rather is part of the Web infrastructure. [Allemang/Hendler(2008):7]

Man kann also RDF als das genuine Datenmodell des Semantic Webs bezeichnen, es ist in seiner konkreten Anwendung im WWW jedoch weit mehr – auf der ersten Stufe eines „Web of Data“¹¹⁴ bildet RDF eben die Grundstruktur des Semantic Webs. Durch URI-referenzierte Konzepte und ihre Verknüpfung mittels RDF-Aussagen wird ein einfachstes Geflecht von sinnhaften Verbindungen zwischen Konzepten bzw. Daten möglich. Allerdings ist diese reine Aussagenstruktur im Hinblick auf differenziertere Beziehungen zwischen Konzepten sehr eingeschränkt.

4.1.3 Formale Wissensrepräsentation

Durch die beschriebene maschinenverarbeitbare Semantik können Konzepte eindeutig identifiziert und mittels RDF in sinnhafte Zusammenhänge gebracht werden. Es ist damit jedoch nicht systematisch darstellbar, in welchen komplexeren Beziehungen Konzepte zueinander stehen. Grundsätzlich ist die Definition eines Begriffs neben beschreibenden Elementen (Merkmalsnennung) durch die Einordnung in ein System von Begriffen möglich. Es ist also sehr sinnvoll, die Bedeutung eines RDF-Konzeptes auch durch seine Beziehungen zu anderen Konzepten zu erklären. Wie beschrieben kann dies in universaler Form aber nur durch eine verbindliche Auszeichnung und Selbstreferenzialität des WWW realisiert werden. Folglich sind in Erweiterung zum Resource Description Framework Modelle und Formalisierungen zur Darstellung von Begriffsbeziehungen nötig.

Wie im Kapitel 2.4 Kontrollierte Vokabulare (KOS) erläutert wurde, hat man im Bereich der Inhaltlichen Erschließung umfangreiche Erfahrungen mit der systematischen Darstellung von Begriffsbeziehungen. Mittels Klassenbildung, hierarchischen und weiteren Relationen von Konzepten werden dort Wissensdomänen geordnet. Die Repräsentation von Wissen umfasst diese Methode der Wissensordnung und weitere Verfahren, um explizites Wissen durch vertretende Surrogate wiederzugeben. [Stock/Stock(2008):37-40]

Zur Wissensrepräsentation bzw. zur umfassenden Modellierung der komplexen Relationen zwischen Konzepten (zumeist einer Wissensdomäne) werden *Ontologien* verwendet. Sie als

114 Auf den Begriff Web of Data in Abgrenzung zum Semantic Web wird im Kapitel 4.3 Aktuelle Anwendung von RDF(S) – Linked Data näher eingegangen.

die elaborierteste Form von knowledge organisation systems zu bezeichnen, trifft nur zum Teil zu. Ontologien zeichnen sich über Kontrollierte Vokabulare hinausgehend durch die *formale* Darstellung der Begriffe und Beziehungen (der sogenannten ontologischen Primitive) sowie logische Komponenten aus:

Im Gegensatz zur Terminologie sind die Begriffe in einer Ontologie formal durch ihre Merkmale, durch Beziehungen zu anderen Begriffen, sowie durch Axiome näher charakterisiert, während allein die Festlegung einer Menge von Begriffen und ihrer Bezeichner schon eine Terminologie ausmacht. [Strauch(2004):91,Ontologie]

„Formale“ Darstellungen beruhen auf grundlegenden Prinzipien, deren Bedeutungen unabhängig von Interpretation und Verhandelbarkeit sind, also objektiv gültig sind:

Such an objective form, and the rules that govern how it works, is called a formalism. [Allemang/Hendler(2008):20]

Um ontologische Primitive und logische Komponenten wie Axiome und Regeln zu formalisieren, sind eigene Ontologiesprachen nötig. [Stock/Stock(2008):256] Diese Ontologiesprachen können in eine maschinenlesbare Syntax gefasst werden und damit automatisierte, logische Ableitungen (sogenanntes inferencing oder automated reasoning) ermöglichen. Wolfgang G. Stock und Mechthild Stock zählen die formale Darstellung in Ontologiesprachen folglich zu den Definitionsmerkmalen von Ontologien:

Unter einer Ontologie (i.e.S.) verstehen wir eine Begriffsordnung, die in einer standardisierten Sprache vorliegt, automatisches Schlussfolgern gestattet, stets über Allgemein- und Individualbegriffe verfügt sowie neben der Hierarchierelation mit weiteren spezifischen Relationen arbeitet. [Stock/Stock(2008):269]

Um konkrete Ontologien entwerfen bzw. für das Semantic Web nutzbar machen zu können, wurden vom W3C die Resource Vocabulary Description Language (RDF Schema, RDFS)¹¹⁵ und die Web Ontology Language (OWL)¹¹⁶ entwickelt. RDFS bietet hierbei die Möglichkeit, die grundlegenden Prinzipien von Begriffssystemen formal wiederzugeben: Klassenbildung und hierarchische Relationen, und dies nicht für Konzepte, sondern auch für Prädikate, also Beziehungen oder Eigenschaften von Konzepten. Darüber hinaus kann deren Gültigkeitsbereich bezüglich Subjekten und Objekten festgelegt werden. Der obligatorische Einsatz zusammen mit RDF wird oft mit „RDF(S)“ oder „RDF/S“ umschrieben:

RDF/S can be used to make statements defining and describing in a formal (machine-understandable) way, application-specific and existing vocabularies (for example, DCMES, Dublin Core Metadata Element Set) or for creating new vocabularies or new elements of an existing one. [McCathieNevile/Méndez(2007):38]

OWL baut auf RDFS auf und erlaubt darüber hinausgehend ungleich präzisere Definitionen von ontologischen Primitiven sowie die Einbeziehung von Axiomen und grundlegenden logischen Funktionen:

¹¹⁵ RDF Vocabulary Description Language, siehe <http://www.w3.org/TR/rdf-schema/>

¹¹⁶ Mittlerweile in der Version 2 vorliegend, siehe <http://www.w3.org/TR/owl2-overview/>

OWL builds upon RDF and RDF Schema; XML-based RDF syntax is used; instances are defined using RDF descriptions; and most RDFS modeling primitives are used.

Formal semantics and reasoning support is provided through the mapping of OWL on logics. Predicate logic and description logics have been used for this purpose. [Antoniou/van Harmelen(2008):152]

Damit wurden konzeptionell-technische Komponenten des Semantic Webs vorgestellt, die dessen Grundgerüst bilden und schon gegenwärtig angewendet werden. Zusammengefasst sind dies:

1. Eindeutigkeit und Referenzierbarkeit: Metadaten, bzw. allgemeiner, jegliche Konzepte, können durch einen rückverfolgbaren Nachweis im WWW universell eindeutig definiert werden. URIs, insbesondere HTTP URIs, erfüllen diese Forderung mit einer einfachen und bewährten Struktur.
2. Universelles Beschreibungsmodell: Auf der Basis der XML-Metasprache und XML-Namensräumen sind inhaltliche Beschreibungen, oder allgemeiner, Aussagen über jegliche Konzepte, mittels dem RDF-Modell möglich. RDF verfügt, ähnlich wie XML, über eine Syntax, die universalen Einsatz und Erweiterbarkeit ermöglicht.
3. Formale Wissensrepräsentation: Ontologien erlauben die formale Modellierung eines Wissensgebietes (formale Wissensorganisation) und Ontologiesprachen dessen explizite Darstellung. Neben Konzepten, Instanzen und frei definierbaren Beziehungen können Wertedefinitionen und logische Postulate bezüglich Konzepten und Beziehungen mit diesen ausgedrückt werden.

In einer Perspektive, die auf das visionäre Fernziel des Semantic Webs abhebt, treten natürlich komplexere Komponenten hinzu, welche die logische Verarbeitung der semantischen Informationen unterstützen. Formale Sprachen bieten umfassende Möglichkeiten, logische Zusammenhänge und ontologisches Wissen darzustellen. Inferenzmaschinen (reasoner) können durch logische Operationen aus diesem deklarativen Wissen deduktiv neues Wissen ableiten. Schließlich sollen zukünftig „intelligente“ Software-Agenten darauf aufbauen und eigenständig komplexe Aufträge bearbeiten können, etwa Informationen zusammentragen und aufbereiten. [Antoniou/van Harmelen(2008):13-16] Die formale Darstellbarkeit von Wissen und solche automatischen Systeme sind dabei notwendig aufeinander bezogen:

Declarative knowledge is modeled by means of ontologies while problem solving methods specify generic reasoning mechanisms. Both types of components can be viewed as complementary entities that can be used to configure new knowledge-based systems from existing reusable components. [Gómez-Pérez/Fernández-López/Corcho(2004):1]

Dies soll nur ein Verweis darauf sein, wie umfangreich und ambitioniert das „Projekt“ Semantic Web eigentlich ist. Ebenso soll mit dem sogenannten Semantic Web Layer Cake bzw. Semantic Web Stack des W3C der weitaus komplexere technische Aufbau des Semantic Webs zumindest angedeutet werden. Dessen aktuelle Version stammt aus dem Jahr 2007:

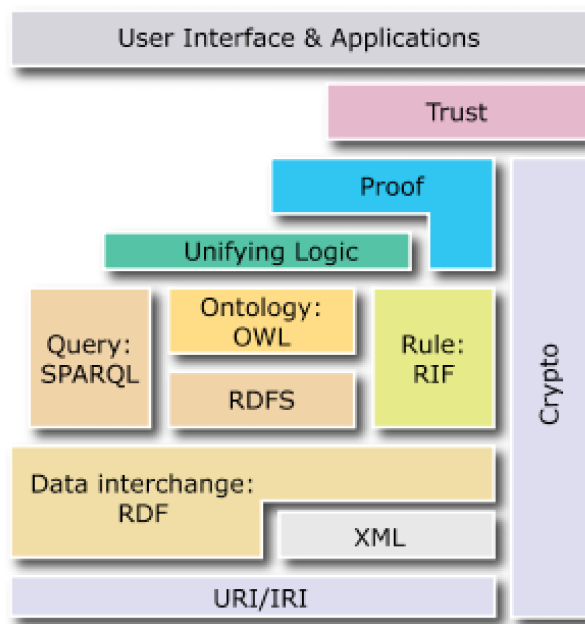


Abbildung 8: W3C - Semantic Web Stack
 (<http://www.w3.org/2007/03/layerCake-small.png>)

Ohne auf die einzelnen Komponenten tiefer eingehen zu können, wird hier klar, dass die gegenwärtige Bereitstellung von semantischen Informationen als Web of Data nur die primitivste Stufe der umfassenden Semantic-Web-Vision ist. Für die Diskussion der weiteren Entwicklung von bibliografischen Metadatensystemen sind jedoch genau die genannten grundlegenden Technologien von Interesse. Auf diese wird dementsprechend in den folgenden Abschnitten genauer eingegangen.

4.2 Resource Description Framework (RDF)

Um maschinenlesbare Metadaten zur Beschreibung von Webdokumenten einfach in das bestehende WWW zu integrieren, hat das W3C 1997 damit begonnen, ein eigenes Beschreibungsmodell für diese zu entwickeln. Seit 1999 bzw. 2004 liegt mit RDF ein einfach strukturiertes Datenmodell vor, das aufgrund seiner Allgemeinheit nicht nur Webdokumente, sondern prinzipiell jede Art von Ressource, ob real oder abstrakt, beschreiben kann.

Das RDF-Modell besteht aus Datentripeln (statements) von Subjekt, Prädikat und Objekt. Die grundlegende Struktur dafür ist ein gerichteter Graph:

A set of such triples is called an RDF graph [...]. This can be illustrated by a node and directed-arc diagram, in which each triple is represented as a node-arc-node link (hence the term "graph"). [...] The direction of the arc is significant: it always points toward the object. [RDFconcepts(2004)]

Wie in der sprachlichen Grammatik wird in einem Tripel eine Aussage über das Subjekt getroffen. Subjekte sind im RDF-Modell Ressourcen (*resources*), d.h. sie können prinzipiell alles sein, was als Konzept beschrieben werden kann. Prädikate werden im RDF-Modell auch als *properties* bezeichnet, sie sind die Beziehung oder Eigenschaft, durch die das Subjekt beschrieben wird. Dadurch, dass sie ein eigenständiges Konzept darstellen („lesen“ oder „Alter“ in den folgenden Beispielen), sind sie ebenfalls Ressourcen, über die Aussagen gemacht werden können. [Antoniou/van Harmelen(2008):67-68] Das Objekt im RDF-Modell kann schließlich ebenfalls eine Ressource sein oder ein Literal¹¹⁷, also eine schlichte Zeichenfolge, die kein Konzept repräsentiert. Dazu zwei kleine Beispiele: 1. Peter liest ein Buch. 2. Peter ist 10 Jahre alt. In der ersten Aussage ist „Peter“ das Subjekt, „lesen“ das Prädikat (Beziehung), „Buch“ das Objekt, das ein eigenes beschreibbares Konzept, also ebenfalls eine Ressource, darstellt. Im zweiten Satz ist „Peter“ ebenfalls Subjekt, „Alter“ das Prädikat (Eigenschaft), „10“ das Objekt, welches hier aber ein literaler Wert und kein eigenes beschreibbares Konzept ist. Des Weiteren werden Literale in *plain literals*, also nicht reglementierte Zeichenketten, oder *typed literals* unterschieden. Für diese typisierten Zeichenketten wird der zulässige Datentyp bzw. Wertebereich festgelegt und durch einen sogenannten datatype URI referenziert.

Die folgende Veranschaulichung des RDF-Modells stammt von der Dublin Core Metadata Initiative. Sie zeigt den gerichteten RDF-Graph mit Subjekt und Prädikat als eindeutig referenzierte Ressourcen und das Objekt, welches hier in Form des typisierten Literals „43“ im vorgegebenen Datentyp Integer vorliegt.

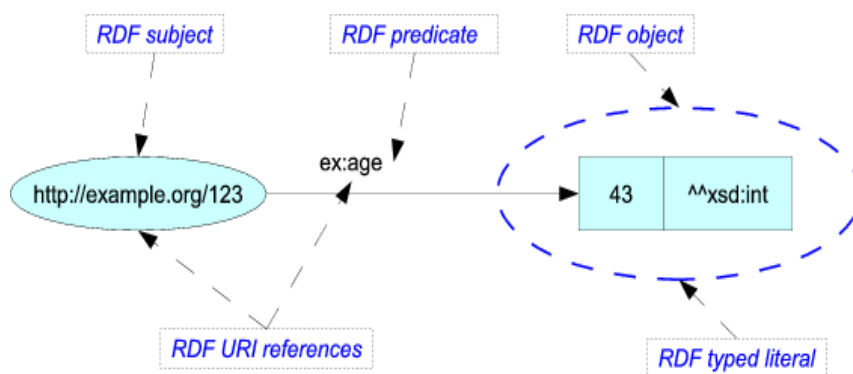


Abbildung 9: DCMI - Beschreibungsmodell RDF [DCinRDF(2008)]

Die Struktur von RDF scheint auf dem ersten Blick dem Objekt-Attribut-Wert-Schema des objektorientierten oder Entity-Relationship-Modells zu entsprechen. [McCathieNevile/Mén-

¹¹⁷„Literals are used to identify values such as numbers and dates by means of a lexical representation. [...] A literal may be the object of an RDF statement, but not the subject or the predicate.“ [RDFconcepts(2004)]

dez(2007):31] Neben der eher ungewöhnlichen Entscheidung, dass properties gleichzeitig auch resources sind, gibt es jedoch einen weiteren entscheidenden Unterschied, wie Tim Berners-Lee ausführt:

This architecture, [...] differs from the 'Entity-relationship' (ER) model and many similar models like it, including most object-oriented programming systems. In an ER model, typically every object is typed and the type of an object defines the attributes can have [sic], and therefore the assertions which are being made about it. [...] The scope of the attribute name is the entity type, just as in OOP the scope of a method name is an object type (or interface). [Berners-Lee(1997)]

Durch die notwendigerweise festgelegte Attributierung von Konzepten in objektorientierten und ER-Modellen werden die Primitive, und somit auch Beschreibungen, von vorn herein definitorisch eingeschränkt. Da RDF auf elementarer Ebene auf Typisierungen verzichtet und anstatt Attribut-Wert-Tupeln Daten-Tripel verwendet, können hier Beschreibungen von Konzepten bzw. Ressourcen gänzlich unabhängig voneinander gemacht werden. Im Ausgleich dazu müssen alle Ressourcen als notwendige Bedingung eindeutig identifiziert sein, also über einen Uniform Resource Identifier (URI) verfügen. Nur so können Aussagen, die dieselben Ressourcen betreffen, zusammengeführt werden. [Allemang/Hendler(2008):37-40]

Dies kann man als Grundgedanke des RDF-Modells bezeichnen: unabhängige Aussagen aus unterschiedlichen Quellen über dieselbe Ressource werden in ihrer Gesamtheit zur Beschreibung der Ressource verwendet. [McCarthyNeville/Méndez(2007):32]

Die konkrete Darstellung von RDF-Aussagen kann unterschiedlich erfolgen: die N-Tripel-Notation (N-Triples) ist die einfachste, sie besteht schlicht aus den URIs (bzw. gegebenenfalls dem Literal) von Subjekt, Prädikat und Objekt, die hintereinander geschrieben werden, mit einem Punkt als Abschluss des statements. [Allemang/Hendler(2008):51-52] Ein kleines Beispiel:

```
<http://dbpedia.org/resources/Goethe's_Faust>  
<http://purl.org/dc/elements/1.1/creator>  
<http://dbpedia.org/resource/Johann_Wolfgang_von_Goethe>.
```

Diese Aussage über „Goethes Faust“ (Subjekt) beinhaltet das Prädikat „creator“, wie es im Dublincore-Schema vorgesehen ist, und das Objekt „Johann Wolfgang von Goethe“. Wie zusätzlich an diesem Beispiel deutlich wird, ist die DBpedia¹¹⁸ eine zentrale Quelle für URIs von Ressourcen, die nicht direkt durch das WWW referenzierbar sind. Wenn man dem obigen property-URI `<http://purl.org/dc/elements/1.1/creator>` folgt, erhält man folgende XML-basierte Definition:

```
<rdf:Property rdf:about="http://purl.org/dc/elements/1.1/creator">  
[...]
```

¹¹⁸ DBpedia, siehe <http://dbpedia.org/> DBpedia nutzt die Einträge und das Kategorienschema der englischsprachigen Wikipedia, um daraus URIs für Ressourcen und z.T. deren Beschreibung bereitzustellen. Ausführlich zur Funktionsweise von DBpedia siehe [Auer/Lehmann/Bizer(2009):265-273].


```
<rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-  
ns#Property"/>  
[...]  
</rdf:Property>
```

Die Ressource (`rdf:about`) „creator“ wird hier mittels dem Prädikat `rdf:type` als eine RDF-property definiert. `rdf:type` ist die grundlegende property in RDF(S), mit der für Ressourcen die Zugehörigkeit zu einer Klasse von Ressourcen erklärt wird, hier eben der ganz allgemeinen Klasse `rdf:Property`.¹¹⁹

Daneben gibt es die von N-Triples abgeleitete, übersichtlichere Notation³ RDF, kurz N3, die auf die auch in XML genutzten qualified names¹²⁰ zurückgreift. In der Präambel des Dokuments werden die Präfixe, die Domänen vertreten, definiert. Das obige Beispiel in N3, zuerst die Präambel, danach das statement:

```
@prefix dbpedia:  
<http://www.dbpedia.org/resources/>  
@prefix dc:  
<http://purl.org/dc/elements/1.1/>
```

`dbpedia:Goethe's_Faust dc:creator dbpedia:Johann_Wolfgang_von_Goethe`.

Wie schon ersichtlich wurde, ist die wichtigste Darstellung für den Gebrauch im WWW (schon wegen der Auflösung von URIs) die Serialisierung von RDF(S) in XML¹²¹, dazu stehen die entsprechenden XML-Namensräume¹²² zur Verfügung. Antoniou und van Harmelen weisen darauf hin, dass im Gegensatz zu XML die Namensräume jedoch nicht nur zur Unterscheidung von Konzepten (Eindeutigkeit) dienen, sondern darüber hinaus, um bereits definierte Ressourcen zur Nachnutzung verfügbar zu machen. [Antoniou/van Harmelen(2008):75] Dieses Prinzip wurde durch die DBpedia- und Dublin-Core-Namensräume oben bereits angedeutet.

Das Wurzelement eines RDF/XML-Dokuments ist `rdf:RDF`. Ein oder mehrere statements über ein Subjekt werden durch das XML-Element `rdf:Description` eingeleitet, das Subjekt wird dabei mit `rdf:about` bzw. `rdf:ID` gekennzeichnet. Prädikat und Objekt folgen dann (jeweils) in Form eines Elements, oder in verkürzter Form als Attribut mit Attributwert. Wie in XML üblich, können Elemente auch verschachtelt werden. [Antoniou/van Harmelen(2008):77-79]

¹¹⁹ Es sei daran erinnert, dass im RDF-Modell auch properties Ressourcen sind.

¹²⁰ Qualified names, oder kurz qnames, siehe [Allemang/Hendler(2008):40-41] bzw. <http://www.w3.org/TR/REC-xml-names/#ns-qualnames>

¹²¹ RDF/XML Syntax Specification (Revised), siehe <http://www.w3.org/TR/rdf-syntax-grammar/>

¹²² Der bereits oben als URI-Referenz angegebene allgemeine RDF Syntax Namespace, siehe <http://www.w3.org/1999/02/22-rdf-syntax-ns> sowie <http://www.w3c.org/2000/01/rdf-schema>

Um RDF-Daten direkt in HTML-Dokumente einbetten zu können, gibt es aktuell zwei Möglichkeiten. Zum einen können durch GRDDL¹²³, eine W3C Empfehlung, Datentripel mittels XSLT-Stylesheets¹²⁴ aus Mikroformaten gewonnen werden. Mikroformate¹²⁵ werden kleine (X)HTML-kompatible Spezifikationen genannt, um einfache Daten zu Personen, Terminen, Orten u.a. im Code vorzuhalten, die nicht zur Anzeige, sondern zur (minimalen) inhaltlichen Auszeichnung bestimmt sind. [Allemang/Hendler(2008):63-64]

Die neuere und vielversprechendere Methode, das WWW einfach semantisch „anzureichern“, ist die W3C-Empfehlung RDF in Attributen, kurz RDFa¹²⁶. Im Moment ist RDFa nur für die Spezifikation XHTML 1.1 gültig, da XHTML im Gegensatz zu HTML erweiterbar ist. RDFa ermöglicht es in einfacher Weise, die XHTML-Elemente mit RDF-Attributen zu versehen, also neben der Darstellungsauszeichnung eine maschinenverständliche Semantik zu implementieren. Der zentrale Vorteil gegenüber Mikroformaten/GRDDL besteht in der Universalität des RDF-Modells, man erhält damit sozusagen ein „generisches Mikroformat“ [Hausenblas(2009):154]. Die Arbeit an XHTML wurde allerdings zugunsten des in Entwicklung befindlichen HTML5 eingestellt. Es ist also davon auszugehen, dass HTML5 in zentraler Weise die RDFa-Einbindung weiterführen wird, um ein „semantisches HTML“ und im weiteren Sinn das Konzept des Semantic Webs zu verwirklichen:

Während Microformats einen sehr einfachen, zielorientierten Ansatz darstellen, Metadaten in HTML einzubetten, stellt RDFa eine vollwertige Semantic Web Technologie dar, da es die Einbettung beliebiger RDF Graphen in HTML erlaubt. Dieser generische Ansatz von RDFa, in Verbindung mit der Verwendung von URIs zur Identifikation von Entitäten im Web macht RDFa zum Standard für semantisches HTML. [...] Semantisches HTML ermöglicht die Repräsentation von Metadaten und Schemata in HTML-Seiten, sodass nicht nur ein Mensch, sondern auch Software sinnvoll eine Webseite verarbeiten können. [Hausenblas(2009):147-148]

Zur Generierung, Speicherung und Verarbeitung von RDF-Daten existieren mittlerweile einige Software-Werkzeuge. Vergleichbar mit relationalen Datenbanksystemen gibt es sogenannte RDF stores (triple stores), die mit dem RDF-Datenmodell arbeiten, z.B. Sesame¹²⁷, das Jena Framework¹²⁸ oder die endnutzer-orientierte Piggy Bank¹²⁹. Um die gespeicherten Daten auswerten zu können, sind die triple stores über RDF-spezifische Abfragesprachen, in erster Linie SPARQL¹³⁰, ansprechbar. [Allemang/Hendler(2008):64-66]

123 GRDDL, Gleaning Resource Descriptions from Dialects of Languages, siehe <http://www.w3.org/TR/grddl/>

124 XSLT (Extensible Stylesheet Language Transformation) ist ein Teil der XSL, die zur Transformation von XML-Dokumenten in andere XML-Sprachen gebraucht wird, siehe <http://www.w3.org/TR/xslt>

125 Microformats, siehe <http://microformats.org/>

126 RDFa, siehe <http://www.w3.org/TR/rdfa-syntax>

127 Sesame, siehe <http://www.openrdf.org/>

128 Jena Framework, siehe <http://jena.sourceforge.net/>

129 Piggy Bank, siehe http://simile.mit.edu/wiki/Piggy_Bank

130 SPARQL Query Language for RDF, siehe <http://www.w3.org/TR/rdf-sparql-query/>

Parser/Serializer¹³¹ bilden die Schnittstelle zwischen anderen Anwendungen und triple stores, indem sie die RDF-Tripel aus der serialisierten RDF/XML-Form auslesen (parsing) bzw. entgegengesetzt RDF-Graphen in eine serialisierte Syntax überführen können.

Zur Umwandlung von Daten, die in relationalen Datenbanken oder Tabellenkalkulationen vorliegen, gibt es spezielle Konverter¹³² während Programme für das Auslesen von strukturierten Informationen aus Webseiten (also aus HTML) als (screen) scraper¹³³ bezeichnet werden.

4.3 Aktuelle Anwendung von RDF(S) – Linked Data

In der derzeitigen Diskussion ist ein bescheidenerer Umgang mit dem Semantic-Web-Begriff, wie er am Anfang von Berners-Lee, Hendler und Lassila 2001 visioniert wurde, zu bemerken. Häufiger werden heute die Begriffe Web of Data oder Linked Data verwendet, die nicht so sehr den Nimbus des „verstehenden“ oder „wissenden“ Netzes haben – das ja leider auch einen Anklang an die Künstliche-Intelligenz-Forschung hat, die wiederum weit hinter ihren Versprechungen zurückgeblieben ist. Der Begriff Web of Data hat eher das technische Zwischenziel im Blick, die bedeutungsvolle Vernetzung von Daten, während „Semantic Web“ eher das visionäre Gesamtbild inklusive der potentiellen Nutzung durch „intelligente“ Software-Agenten symbolisiert. Linked Data wird demgegenüber als die heute schon machbare, praktikable und sinnvolle Anwendung der semantischen Technologien verstanden. Auer, Lehmann und Bizer subsumieren unter Linked Data folglich die

...Menge von Best Practices zur Veröffentlichung und Verknüpfung von strukturierten Daten im Web. Grundannahme von Linked Data ist, dass der Wert und die Nützlichkeit von Daten steigen, je stärker sie mit Daten aus anderen Datenquellen verknüpft sind.
[Auer/Lehmann/Bizer(2009):261]

Auch wenn ein allgemein gültiges, semantisches HTML als Umsetzung des Semantic-Web-Konzeptes noch auf sich warten lässt, sind die technischen Grundlagen dafür zum großen Teil bereits vorhanden. Wie Antoniou und van Harmelen betonen, hängt dessen Realisierung im Moment vor allem davon ab, ob diese auf breiter Basis von Wissenschaft und normalen Internetnutzern angenommen und umgesetzt werden:

At present, the greatest needs are in the areas of integration, standardization, development of tools, and adoption of users. [Antoniou/van Harmelen(2008):8]

Auch aus dieser Perspektive ist Linked Data also eine wichtige Voraussetzung zur weiteren Entwicklung des Semantic Webs:

131 Als Beispiel sei die Raptor RDF Parser Library genannt, <http://librdf.org/raptor/>

132 Sammlungen von Convertern finden sich z.B. im Wiki des ESW, siehe <http://esw.w3.org/topic/Converter-ToRdf> und beim SIMILE-Projekt, siehe <http://simile.mit.edu/wiki/RDFizers>

133 Der scraper Solvent funktioniert z.B. als Firefox-Plugin, siehe <http://simile.mit.edu/wiki/Solvent>

Therefore, while the Semantic Web, or Web of Data, is the goal or the end result of this process, Linked Data provides the means to reach that goal. [Bizer/Heath/Berners-Lee(2009)]

In dem durchaus an die gesamte WWW-Öffentlichkeit gerichteten Report „Linked Data – the story so far“ geben Bizer, Heath und Berners-Lee eine genauere Definition des Begriffs:

Technically, Linked Data refers to data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets. [Bizer/Heath/Berners-Lee(2009):2]

Die Grundprinzipien, die für Linked-Data-Anwendungen von Tim Berners-Lee bereits 2006 benannt wurden¹³⁴, sind eben die des Semantic Webs, wobei auf komplexe Konstrukte aber zunächst verzichtet wird. Konkret umfasst Linked Data die Identifizierung aller Konzepte im WWW mit URIs und die Dereferenzierbarkeit dieser URIs über das Web (also HTTP-URIs), die Verwendung von RDF(S) und gegebenenfalls OWL zur Veröffentlichung von strukturierten Daten, sowie die Verknüpfung von Daten in verschiedenen Datenquellen durch möglichst umfassende RDF-Verweise. [Auer/Lehmann/Bizer(2009):261]

Es handelt sich also um die Forcierung einer breiten semantischen Annotation von Dokumenten und Konzepten, die überhaupt erst die Grundlage für komplexere Anwendungen bilden. Seinen Ursprung hat Linked Data in der W3C SWEO-Community als Projekt Linking Open Data¹³⁵, heute ist es mittlerweile eine allgemeine Bewegung verschiedener Produzenten im WWW. Das Linking-Open-Data-Projekt definiert ihr Ziel folgendermaßen:

The goal of the W3C SWEO Linking Open Data community project is to extend the Web with a *data commons* by publishing various *open data sets* as RDF on the Web and by setting RDF links between data items from different data sources. (SWEO Interest Group, Hervorhebungen K.W.)¹³⁶

Hier wird klar, dass neben dem Semantic-Web-Konzept ein zweites fundamentales Prinzip zugrunde gelegt wird – das der Freiheit der veröffentlichten Daten von urheberrechtlichen Beschränkungen. Im angloamerikanischen Copyright kann der Urheber von Daten(-banken) diese der Public Domain widmen¹³⁷, also auf seine Urheberschaft genuin verzichten. Eine andere Möglichkeit, Daten zu „offenen“ Daten zu machen, ist es, unentgeltlich ein einfaches Nutzungsrecht für jedermann einzuräumen, dies ist auch nach deutschem Urheberrecht mög-

134 Tim Berners-Lee, Linked data – design issues, siehe <http://www.w3.org/DesignIssues/LinkedData.html>

135 Semantic Web Education and Outreach (SWEO) Interest Group, siehe <http://www.w3.org/2001/sw/sweo/> bzw. W3C SWEO Community Project Linking Open Data, siehe <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

136 Vgl. die erwähnte Projekt-Homepage, <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

137 Siehe z.B. die Public Domain Dedication and Licence (PDDL), <http://www.opendatacommons.org/licenses/pddl/> sowie die Creative-Commons-Zero-Lizenz (CC0), <http://creativecommons.org/about/cc0>. Einige Bibliotheken haben sich in Zusammenarbeit mit dem Hochschulbibliothekszentrum Nordrhein-Westfalen (hbz) bereits zur Datenpublikation unter der CC0-Lizenz entschlossen, siehe http://www.hbz-nrw.de/projekte/linked_open_data/index_html

lich. Ziel ist es in jedem Fall, Daten zur freien und uneingeschränkten¹³⁸ Nachnutzung zur Verfügung zu stellen.

Die Linked-Data-Bewegung hat bereits eine beachtenswerte inhaltliche Breite erreicht, sozusagen einen

...global data space connecting data from diverse domains such as people, companies, books, scientific publications, films, music, television and radio programmes, genes, proteins, drugs and clinical trials, online communities, statistical and scientific data, and reviews. [Bizer/Heath/Berners-Lee(2009):2]

Dies lässt sich am besten anhand der sogenannten Linked Data Cloud¹³⁹ ablesen, welche die bisher verknüpften Datenquellen als Diagramm darstellt. Wie bereits gezeigt, ist der Rückgriff auf referenzierbare Vokabulare für die semantische Auszeichnung unabdingbar. Auch dafür existiert ein kleines Diagramm, das die gegenseitigen Bezüge der Ontologien/Vokabularien darstellt¹⁴⁰. Einige allgemein bekannte Vokabulare sind hier etwa FOAF, SIOC oder GeoNames¹⁴¹, insbesondere für den bibliothekarischen Bereich sind die Vokabulare von Dublin Core, SKOS, Bibliographic Ontology und FRBR¹⁴² von Interesse.

Über die Publikation von Datensets und Vokabulare hinaus sind auf der technischen Ebene zudem Daten-Browser und spezifische Suchmaschinen entwickelt worden.¹⁴³ Vor dem Hintergrund der quantitativen und vor allem qualitativen Entwicklung die das WWW bzw. das Web 2.0 genommen haben, steht die wahre Entfaltung von Linked Data aber eindeutig erst noch bevor. Bizer, Heath und Berners-Lee resümieren daher optimistisch:

...Linked Data realizes the vision of evolving the Web into a global data commons, allowing applications to operate on top of an unbounded set of data sources, via standardised access mechanisms. If the research challenges highlighted above can be adequately addressed, we expect that Linked Data will enable a significant evolutionary step in leading the Web to its full potential. [Bizer/Heath/Berners-Lee(2009):20]

138 Die Diskussion um eben doch einschränkende „share-alike“-Lizenzen bzw. dem Copyleft-Prinzip soll hier lediglich erwähnt werden. Das Hochschulbibliothekszentrum Nordrhein-Westfalen hat sich dazu entschlossen, die Katalogdaten der verschiedenen Bibliotheken

139 Linking Open Data dataset cloud oder kurz Linked Data Cloud, aktuelle Version siehe <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData#dbpedia-lod-cloud> oder auf dem Stand März 2009 z.B. in [Bizer/Heath/Berners-Lee(2009):5].

140 Linking Open Data Constellation (Stand Oktober 2008), siehe http://umbel.org/images/lod_constellation.html

141 Vgl. FOAF (Friend-of-a-Friend), <http://www.foaf-project.org/> SIOC (Semantically-Interlinked Online Communities), <http://sioc-project.org/> GeoNames, <http://www.geonames.org/ontology/>

142 Vgl. Dublin Core, <http://dublincore.org/documents/dc-rdf/> SKOS (Simple Knowledge Organization System), <http://www.w3.org/2004/02/skos/> Bibliographic Ontology, <http://bibliontology.com/> FRBR, <http://vocab.org/frbr/core>

143 Ausführlicher und mit aktuellen Beispielen, siehe [Auer/Lehmann/Bizer(2009):273-276].

4.4 Wissensrepräsentation im Semantic Web

4.4.1 Ontologien

Um die Vision des Semantic Webs zu verwirklichen, genügt es nicht, sich auf die grundlegenden Datenstandards zu verständigen. Die weitaus diffizilere Herausforderung besteht darin, sich über die Konzepte, über die Aussagen getroffen werden, eins zu werden:

The Semantic Web requires that information bearing entities on the Web be tagged with machine-processable meaning (semantic metadata) in a standard way. The standardization will enable the exchange, use, and reuse of information. [...] A fundamental component to this activity is the development, registration, and sharing of metadata schemas and ontologies. [Greenberg/Méndez(2007):2]

Wohlgemerkt geht es nicht (oder zumindest nur im geringem Ausmaß) darum, wahre Aussagen zu postulieren. Das Semantic Web soll sich ja per se aus elementaren RDF-Aussagen – stetig wachsend und vernetzend – selbst bilden. Aufgabe von Ontologien in der Semantic-Web-Vision ist es, die zumindest von einem Großteil der Nutzer geteilten Wissenskonzepte bereitzustellen, die in Form von RDF-Ressourcen für Aussagen verwendet werden können:

The Semantic Web can be viewed as the next generation of the World Wide Web architecture, an essential part of which is the content producers' ability to the *share the ontological information that defines the formal semantics* of the Web content. Through shared ontological commitments reasoning agents are able to reach *partial shared understanding* and thus interoperate. [Lassila/McGuinness(2001), Hervorhebungen K.W.]

Bereits erwähnt wurde die „ontologische Annahme“ in der Informationsorganisation. Allgemein befasst sich die philosophische Disziplin der Ontologie sehr grob gesagt eben mit der Natur alles Existenten. Genauer schreibt Barry Smith:

Ontology as a branch of philosophy is the science of what is, of the kinds and structures of objects, properties, events, processes, and relations in every area of reality. [Smith(2004):155]

Es geht also im umfassenden Sinn um die Charakterisierung und Kategorisierung aller Phänomene und Dinge, um die Erforschung der Strukturen der Welt.¹⁴⁴

Dies mag an das Ziel von knowledge organisation systems (KOS) erinnern. Zwar beschäftigen sich diese weniger mit der philosophischen Diskussion um die Dinge, Phänomene und ihre Strukturen, aber sie versuchen eben diese als Konzepte und Strukturen modellhaft abzubilden, sie anzuordnen – zumindest soweit, wie es dem aktuellen und überwiegend anerkannten Wissenstand entspricht und mit dem Zweck der sachlichen Beschreibung von Ressourcen.

Ontologien (notabene: Plural) werden daran anknüpfend auch die mächtigsten Formen von KOS genannt. Der Begriff, wie er in der Informationswissenschaft, aber auch der Informatik

144 Einen kurzen Abriss der Geschichte der philosophischen Ontologie bieten [Legg(2007):422-425] und [Gómez-Pérez/Fernández-López/Corcho(2004):3-5].

verwendet wird¹⁴⁵, ist somit ein anderer: es geht um eine praktische Darstellung von Wissensdomänen an Hand ihrer ontologischen Bestandteile (Primitive):

In the context of computer and information sciences, an ontology defines a set of representational primitives with which to *model a domain of knowledge or discourse*. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members). The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application. [Gruber(2009), Hervorhebung K.W.]

Tom Gruber¹⁴⁶ unterscheidet in seiner Definition in der Encyclopedia of Database Systems also Klassen (Konzepte), Attribute (Eigenschaften) und Beziehungen zwischen Klassen als die Primitive, mittels der hier Wissen dargestellt bzw. organisiert wird. Klassenbildung und Relationierung sind dabei von älteren KOS bekannt. Attributierung von Konzepten/Klassen sowie eine genauere Definition von Relationen sind dagegen in den neuen Thesauri-Standards BS8723 bzw. ISO 25964 vorgesehen. Hier nähern sich die Normen für strukturierte Vokabulare bzw. für Thesauri schon dem Ontologie-Konzept an. Die freie Definition von Relationen (im Unterschied zur Darstellung von einigen wenigen Relationsarten) ist ein Merkmal von Ontologien, dass sie Vorläufern, den Semantischen Netzen¹⁴⁷, verdanken. Diese zeichnet eine umfassende assoziative Relationierung von Konzepten aus, sie verzichten aber auf Klassenbildung (und wie Ontologien auch, auf die Terminologie-Problematik).

Darüber hinaus können sogenannte heavyweight ontologies, die auf Frames und Prädikatenlogik oder Beschreibungslogik beruhen, neben Klassen, Beziehungen und Instanzen über Funktionen und Axiome verfügen. Mittels formaler Axiome werden grundsätzliche, wahre Aussagen postuliert, die sich nicht anders formal darstellen lassen. [Gómez-Pérez/Fernández-López/Corcho(2004):11-14]

Was Ontologien aber prinzipiell auszeichnet, ist die Ausrichtung auf eine logische Verarbeitbarkeit (reasoning, inferencing) ihrer Primitive. Dies bedeutet, dass eine Ontologie idealerweise so umfassend und formalisiert beschrieben wird, dass logisch geschlussfolgert werden kann, also neue Attribute, Beziehungen usw. daraus abgeleitet werden können. Dagobert Soergel definiert Ontologien demzufolge:

In AI-related contexts, an ontology is a classification with a rich set of semantic relationships that support reasoning. [Soergel(2009):27]

145 Eine Auswahl von informationswissenschaftlichen Definitionen geben [Gómez-Pérez/Fernández-López/Corcho(2004):6-8].

146 Auf Tom Gruber geht ebenfalls eine der ersten und meistzitiertesten Definitionen von Ontologien im Sinne der Wissensorganisation zurück: „An ontology is an explicit specification of a conceptualization.“ (1993), zitiert nach [Stock/Stock(2008):13-14].

147 Das Konzept der Semantischen Netze (semantic networks, 1975) geht auf William A. Woods zurück, vgl. [Lassila/McGuinness(2001)].

Auch Catherine Legg sieht in der logischen Verarbeitbarkeit die entscheidende Schwelle, aber welcher man in in der Reihe von KOS mit steigender Komplexität¹⁴⁸ von einer Ontologie sprechen sollte. Gleichzeitig stellt sie fest:

...inferencing capability is inseparable from machine-understandable meaning... [Legg(2007):443]

Auch Klassifikationen und Thesauri verfügen über semantische Relationen – allerdings sind diese Bedeutungen nur für Menschen verständlich und auswertbar. Die formale, also maschinenverarbeitbare Darstellung von Beziehungen ist dagegen der erste Schritt, der ein automatisches Reasoning möglich macht:

...Semantic Web applications will have to draw on some kind of shared, structured machine-readable *conceptual scheme*. Thus there has been a convergence between the Semantic Web research community and an older tradition with roots in classical Artificial Intelligence (AI) research ... whose goal is to develop a *formal ontology*. A formal ontology is a machine-readable *theory* of the most fundamental concepts or 'categories' required in order to understand information pertaining to any knowledge domain. [Legg(2007):407, Hervorhebungen Legg]

Zunächst müssen also Klassen sowie ihre Eigenschaften und Relationen (properties)¹⁴⁹ in einer formalen Sprache dargestellt werden. Am Anfang steht dabei die Formalisierung der einfachsten Relation, der generischen Beziehung zwischen Klassen und Unterklassen, die eine grundlegende automatische Verarbeitung des KOS ermöglicht. Nach Ora Lassila und Deborah McGuinness ist damit die einfachste Stufe eines ontologischen Systems erreicht. [Lassila/McGuinness(2001)]

Dann können durch die Formalisierung der Eigenschaften von Beziehungen (properties), also Transitivität, Symmetrie usw., weitere Beziehungen abgeleitet werden. So kann z.B. durch die Formalisierung der Transitivität die automatische Vererbung von Merkmalen erfolgen. [Legg(2007):428] Zudem können Relationen bezüglich ihres „Einsatzbereiches“ eingeschränkt werden: domain und range bezeichnen die Gruppe von Konzepten und Werten, auf welche die Relation anwendbar ist. In der Darstellung als Subjekt-Prädikat-Objekt-Modell bezeichnet domain eben die möglichen Subjekte, range die möglichen Objekte des Prädikats. Außerdem können Instanzen (individuals) in die Ontologie aufgenommen werden.

Bestehende Ontologien werden oft nach ihren Einsatzgebieten unterschieden: auf ein Wissensgebiet begrenzte Ontologien werden als domain ontologies, für eine bestimmte Aufgabe oder Anwendung geschaffene Ontologien task ontologies genannt. Top-level und upper ontologies sind solche, die nur aus wenigen, aber allgemeinsten Konzepten bestehen, mittels derer die Interoperabilität von Domain-Ontologien hergestellt werden soll. [Gómez-Pérez/Fernández-López/Corcho(2004):32-34]

148 Vgl. die bekannte Grafik „An Ontology Spectrum“ von Ora Lassila und Deborah L. McGuinness, [Lassila/McGuinness(2001)].

149 Eigenschaften und Relationen werden in RDF ja in gleicher Weise durch properties dargestellt – je nachdem in Verbindung mit einer weiteren Ressource oder einem Literal (Wert) als Objekt.

Für die Modellierung von Ontologien bzw. das automatische Schließen gibt es mittlerweile einige Software-Werkzeuge. Zur Erstellung von konkreten Ontologien gibt es mittlerweile Ontologie-Editoren¹⁵⁰, die Eingaben auf einer grafischen Oberfläche in Ontologiesprachen überführen. Integriert in Editoren, oder auch separat gibt es Werkzeuge zur Bearbeitung, z.B. der Zusammenführung (merging, matching) von Ontologien. Unter dem Stichwort ontology engineering¹⁵¹ werden schließlich alle Methoden, Vorgänge bzw. Softwaretools zusammengefasst, die den gesamten „Lebenszyklus“ von Ontologien unterstützen. Ontologie-Suites¹⁵² vereinen dieses ontology engineering, sie unterstützen zumeist RDFS, OWL und z.T. weitere formale Sprachen und beinhalten reasoning bzw. inferencing machines, also automatisierte logische Verarbeitungsroutinen. [Synak/Dabrowski/Kruk(2009):43-44] Ebenfalls wichtig sind ontology registries bzw. „libraries“¹⁵³, um Informationen über bestehende Ontologien und ihre Beschreibung nachzuweisen und zur Nachnutzung bereitzustellen.

Tom Gruber fasst in der Encyclopedia of Database Systems die Funktionen von Ontologien im Rahmen des Semantic Webs wie folgt zusammen:

Ontologies are part of the W3C standards stack for the Semantic Web, in which they are used to specify standard conceptual vocabularies in which to exchange data among systems, provide services for answering queries, publish reusable knowledge bases, and offer services to facilitate interoperability across multiple, heterogeneous systems and databases. [Gruber(2009)]

Der funktionelle Kern von Ontologien besteht also in der möglichst genauen Formalisierung der logischen Eigenschaften bzw. Beziehungen der Ontologie-Primitive. Dies erlaubt nicht nur die Darstellung komplexer Sachverhalte wie Wissensdomänen. Durch die Formulierung von deduktiven Ableitungsregeln (reasoning oder inferencing rules) und gezielten Abfragen (querying) kann darüber hinaus ontologie-implizites Wissen automatisiert explizit gemacht werden. [Synak/Dabrowski/Kruk(2009):43]

4.4.2 RDF Schema und OWL

Zur Beschreibung von Ontologien, ihrer Primitive und logischen Eigenschaften sind eigene ausdrucksstarke formale Sprachen nötig, die Ontologiesprachen:

Similarly, while an ontology must be formulated in *some* representation language, it is intended to be a semantic level specification – that is, it is independent of data modeling strategy or implementation. [...] Ontologies are typically formulated in languages which are closer in expressive power to logical formalisms such as the predicate calculus. This allows the ontology

150 Am stärksten verbreitet ist aktuell die Open-Source-Suite protégé, siehe <http://protege.stanford.edu/> weitere Beispiele sind KAON, <http://kaon.semanticweb.org/> sowie KAON2, <http://kaon2.semanticweb.org/> und SWOOP für OWL-Ontologien, <http://code.google.com/p/swoop/>

151 Siehe dazu ausführlicher [Antoniou/van Harmelen(2008):225-244].

152 Zum Beispiel das open-source-basierte NeOn Toolkit, <http://neon-toolkit.org>

153 Eine Auswahl geben [Antoniou/van Harmelen(2008):231].

designer to be able to state semantic constraints without forcing a particular encoding strategy. [Gruber(2009), Hervorhebung Gruber]

Das W3C empfiehlt für die Einbindung in das WWW grundsätzlich zwei formale Sprachen für Ontologien: die einfachere RDF Vocabulary Description Language (RDF Schema, RDFS) und die komplexeren Varianten der Web Ontology Language (OWL).

RDF Schema baut, wie man erahnen kann, auf dem allgemeinen Beschreibungsmodell RDF auf und dient zur Beschreibung bzw. Modellierung eigener Vokabulare als Grundlage für RDF-Beschreibungen. Sie ermöglicht durch die allgemeine, formale Umsetzung der Prinzipien Klassenbildung und Vererbung die Darstellung von grundlegenden Strukturen der Wissensrepräsentation. RDF-resources können in Klassen zusammengefasst (`rdfs:Class`) und diese Klassen hierarchisch in Beziehung gesetzt werden (`rdfs:subClassOf`); RDF-properties können ebenfalls hierarchisiert werden (`rdfs:subPropertyOf`). Dean Allemang und Jim Hendler fassen die Funktionalität von RDFS mittels ihrer wichtigsten Primitive folgendermaßen zusammen:

RDFS is the schema language for RDF; it describes constructs for types of objects (`Classes`), relating types to another (`subClasses`), properties that describe objects (`Properties`), and relationships between them (`subProperty`). The Class system in RDFS includes a simple and elegant notion of inheritance, based on set inclusion [...]. [Allemang/Hendler(2008):121]

Ein Vorteil von RDFS, der bereits prinzipiell von XML/XML Schema bekannt ist, besteht darin, dass RDFS bzw. die damit dargestellten Vokabulare selbst in RDF formuliert werden. Allerdings geht RDFS in ihrer Bedeutung über andere Schemasprachen weit hinaus. Sie macht eben nicht strukturelle Vorgaben für bestimmte (schema-konforme) Objekte, sondern stellt eine allgemeine Syntax bereit, um RDF-Konzepten über die Einordnung in Begriffssysteme eine bestimmte Semantik zuzuweisen:

The key idea of the schema in RDF is that it should help provide some sense of meaning to the data. [...] This idea is not completely unfamiliar for schema languages. In the case of XML Schema, the validation process provides more than just a yes/no answer, but it also provides type information for the parsed data. RDFS builds on and formalizes this idea by providing detailed axioms that express exactly what inference can be drawn from particular data patterns. [Allemang/Hendler(2008):92, Hervorhebung Allemang/Hendler]

Durch die Formalisierung von Klassenbildung und Vererbung ergibt sich die Möglichkeit einer automatisierten Inferenz. RDFS erlaubt darüber hinaus die Definition von Gültigkeitsbereichen (in Form von Klassen) bezüglich der Subjekte (`rdfs:domain`) und der Objekte (`rdfs:range`) von properties. Dies stellt eine einfach erscheinende, aber in ihrer Konsequenz weitreichende Methode dar, wenn man sich den funktionalen Unterschied von `domain` und `range` gegenüber deren landläufiger Verwendung klarmacht:

Because of the inference-based semantics of RDFS (and OWL), domains and ranges are not used to validate information (as is the case, for example, in OO modeling and XML) but instead are used to determine new information based on old information. We have just seen how this unique aspect of

`rdfs:domain` and `rdfs:range` support [sic] particular uses of filtering and classifying information. [Allemang/Hendler(2008):115-116]

Über diese Gültigkeitseinschränkungen von Beziehungen bzw. Eigenschaften können schließlich auch Ressourcen bzw. Klassen indirekt beschrieben werden. Durch die formale Darstellung sind hier ebenfalls logische Ableitungen möglich.

Das W3C sieht im Design von RDF(S) eine Reihe von Vorteilen:

- having a simple data model
- having formal semantics and provable inference
- using an extensible URI-based vocabulary
- using an XML-based syntax
- supporting use of XML schema datatypes
- allowing anyone to make statements about any resource [RDFconcepts(2004)]

Kurz gesagt, ist RDF(S) einfach in seiner grundlegenden Struktur und Syntax, dafür aber mächtig in seiner universellen Anwend- und Erweiterbarkeit und der formalen Verarbeitbarkeit. RDF(S) ist heute damit der Quasi-Standard für semantic-web-taugliche Vokabulare bzw. „leichtgewichtige“ Ontologien¹⁵⁴ und bildet die Grundlage für komplexere Ontologiesprachen.

Die Web Ontology Language (OWL) bietet auf RDFS aufbauend viel umfangreichere Möglichkeiten. Genauer gesagt existieren drei Dialekte von OWL, die sich in ihrer Expressivität und (umgekehrt proportional dazu) Entscheidbarkeit¹⁵⁵ stark unterscheiden. OWL Full ist die „vollständige“ OWL-Sprache, die aufgrund ihrer weitreichenden Möglichkeiten jedoch nicht entscheidbar sein muss. OWL DL (descriptions logics) verfügt prinzipiell über dieselben Konstrukte, aber über Restriktionen, die eine Entscheidbarkeit auf der Grundlage von Beschreibungslogiken sicherstellt. OWL Lite ist ein Subset von OWL DL, das weitere Einschränkungen aufweist und auf eine leichte Implementierbarkeit ausgerichtet ist.¹⁵⁶ Diese Unterscheidungen sind jedoch weniger wichtig vor dem Hintergrund der variablen Auswahl von einzelnen Konstrukten und Restriktionen für praktische Anwendungen. Zudem ist OWL weiter in Entwicklung, so dass hier nur beispielhafte Möglichkeiten von OWL genannt werden sollen.

¹⁵⁴ Im Unterschied zu Ontologien im engeren Sinne (heavyweight ontologies), die über logische Komponenten verfügen, werden unter lightweight ontologies thesauri- und schlagwortsystem-ähnliche Vokabulare verstanden, die in RDFS formuliert vorliegen, vgl. [Stock/Stock(2008):256]. Sie stellen eine wichtige Komponente des Linked-Data-Konzeptes dar.

¹⁵⁵ Expressivität (expressivity): „The ability of a modeling language to describe certain aspects of the world.“ [Allemang/Hendler(2008):29], Entscheidbarkeit (decidability): „Formally, a system is decidable if there exists an effective method such that for every formula in the system the method is capable of deciding whether the formula is valid (is a theorem) in the system or not. If not, then the system is undecidable.“ [Allemang/Hendler(2008):295]

¹⁵⁶ Zu den Unterschieden zwischen OWL Full, OWL DL und OWL Lite siehe [Antoniou/van Harmelen(2008):117-119] und [Allemang/Hendler(2008):297-299].

So können in OWL etwa Klassen durch Aufzählung von Instanzen oder durch die Festlegung von property-Werten (restrictions¹⁵⁷) definiert werden. Klassen können als äquivalent, konjunkt, disjunkt, Schnittmenge oder komplementär zueinander definiert werden und daher Ableitungen über Klassenbeziehungen bzw. über die Struktur der Wissensrepräsentation selbst gemacht werden. [Allemang/Hendler(2008):183-196, 214-217, 226-229, 238-244]

Properties können invers oder äquivalent zueinander, als symmetrisch und transitiv, sowie generell als object property oder datatype property definiert werden; ihre Eigenschaften¹⁵⁸, etwa die Kardinalität und (inverse) Funktionalität, ebenfalls durch restrictions bestimmt werden. [Allemang/Hendler(2008):124-132, 142-143, 149-152, 155, 222-226] Gerade diese genauere Definition von Eigenschaften bzw. Beziehungen spielt eine große Rolle in der Annäherung von formalem Wissen an das menschliche Verstehen und Schlussfolgern, wie Rebecca Green schreibt:

The intuitive understanding that humans bring to relationships is not shared by computational devices. At the same time, the expression and manipulation of relationships is perhaps our best hope for infusing higher quality into our retrieval systems. [...] Conversely if we learn how to reason on these relationships, the higher quality may also come in the identification of relevant sources that would otherwise not have been retrieved at all, for example, if we can retrieve sources that are relevant by analogy. [Green(2001):14]

Der sinnvolle Einsatz von OWL setzt also eine genaue Analyse und daraus abgeleitete Modellierung der Wissensdomäne voraus. Nur dann können bei der Implementierung von weitreichenden logischen Konstrukten und Regeln neue, wahre Zusammenhänge und Fakten abgeleitet bzw. abgefragt werden. Allemang und Hendler betonen dabei die Vorteile, die in der Anwendung von logischen Verfahren auf die Wissensrepräsentation liegen:

Using logic as the foundation of a modeling language makes perfect sense; we can draw upon decades, or even centuries, of development work in logical formalism. The properties of various logic structures are well understood. Logic provides a framework for defining all of the inferences that our modeling language will need. [Allemang/Hendler(2008):294-295]

Abseits der logischen Formalisierung von Wissensrepräsentation erfüllen einige OWL-Tags noch administrative Aufgaben, etwa `owl:Ontology` als (nicht formal gültige) Kennzeichnung von „RDF-Datensammlungen“ als eine Ontologie, `owl:imports` bezüglich deren Zusammenführung sowie weitere zur Versionsbeschreibung der Ontologie. [Allemang/Hendler(2008):255-257]

4.4.3 Simple Knowledge Organization System (SKOS)

Das Simple Knowledge Organization System (SKOS)¹⁵⁹ ist ein vom W3C empfohlener Standard zur Darstellung von knowledge organisation systems (KOS) für Semantic-Web-Anwen-

157, „A Restriction is a class that is defined by a description of its members in terms of existing properties and classes.“ [Allemang/Hendler(2008):183]

158 Zur grundlegenden Betrachtung von Eigenschaften von Beziehungen siehe [Green(2001)].

159 SKOS, siehe <http://www.w3.org/2004/02/skos/>

dungen. Das SKOS-Vokabular¹⁶⁰ (ehemals SKOS Core, SKOS Extensions und SKOS Mapping) stellt dabei alle Konstrukte zur Beschreibung von Begriffssystemen bereit, die optionale SKOS eXtension for Labels¹⁶¹ (SKOS-XL) bietet zusätzliche Konzepte zur Definition, Beschreibung und Relationierung von Benennungen. SKOS basiert auf RDF(S)- und OWL-Konzepten und ist die zentrale Methode, um Kontrollierte Vokabulare bzw. ihre Komponenten im WWW maschinenverarbeitbar darzustellen und eindeutig zu referenzieren. In der W3C-Referenz wird die Funktionalität von SKOS wie folgt umrissen:

The SKOS data model provides a standard, low-cost migration path for porting existing knowledge organization systems to the Semantic Web. SKOS also provides a lightweight, intuitive language for developing and sharing new knowledge organization systems. It may be used on its own, or in combination with formal knowledge representation languages such as the Web Ontology language (OWL). [SKOSref(2009):Abstract]

Aufgrund der jahrzehntelangen Entwicklungsarbeit, die an den existierenden KOS geleistet wurde, sowie ihrer inhaltlichen Breite kann mit SKOS nun einfach und ohne großen Aufwand die Semantic-Web-Forderung nach referenzierbarer, geteilter Wissensrepräsentation erfüllt werden. Dies gilt natürlich nur eingeschränkt, da wie beschrieben, Ontologien als genuine Semantic-Web-Technologie über Begriffssysteme mit ihren logischen Struktur weit hinaus gehen. Gleichwohl wird bezüglich einigen KOS nicht umsonst von „lightweight ontologies“ gesprochen: so verfügen z.B. Thesauri bereits über elaborierte Relationsarten sowie explizites Zugangsvokabular und Klassifikationen und Schlagwortssysteme können die breitesten Wissensorganisationssysteme überhaupt bieten. SKOS „begnügt“ sich mit der Darstellung der Strukturen von Kontrollierten Vokabularen, also der Beziehungen zwischen Begriffen, und zwischen Begriffen und Benennungen, sowie administrative Komponenten in diesen:

Using SKOS, concepts can be identified using URIs, labeled with lexical strings in one or more natural languages, assigned notations (lexical codes), documented with various types of note, linked to other concepts and organized into informal hierarchies and association networks, aggregated into concept schemes, grouped into labeled and/or ordered collections, and mapped to concepts in other schemes. [SKOSref(2009):Synopsis]

Dies aber stellt bereits einen wichtigen Beitrag zur Realisierung von „semantischen Metadaten“ dar: die Einbeziehung von referenzierbaren, normierten Wissenskonzepten bzw. ihren Benennungen wird so auf breiter, erprobter Grundlage möglich.¹⁶²

Prinzipiell sind darüber hinausgehend Konzeptionen von Ontologien auf der Basis von Begriffssystemen möglich. Da dies jedoch umfangreiche intellektuelle Arbeit erfordert, werden anstatt der *Aussagen über die Welt* wie sie eine echte Ontologie macht, gerne mittels SKOS schlicht *Aussagen über ein Modell der Welt* dargestellt. Mit SKOS formulierte Aussagen be-

¹⁶⁰ SKOS Vocabulary Namespace Document, <http://www.w3.org/TR/skos-reference/skos.html>

¹⁶¹ SKOS-XL Namespace Document, <http://www.w3.org/TR/skos-reference/skos-xl.html>

¹⁶² Insbesondere für die Linked-Data-Bewegung, die auf höhere logische Konstrukte vorerst verzichtet und eine breite Grundlage in Form von verknüpften RDF-Daten schaffen will, spielt dies eine große Rolle.

ziehen sich auf Konzepte in Begriffssystemen und nicht auf die „realen“ Konzepte an sich. Diese Unterscheidung mag auf den ersten Blick nicht besonders erscheinen, sie ist aber unbedingt notwendig, da die Konstrukte in Begriffssystemen eben nicht logisch definiert sind, also über keine formale Explikation der Konzepte, Relationen und gegebenenfalls Axiomen verfügen. [SKOSref(2009):1.3] Dies wäre auch nur in Ausnahmefällen überhaupt möglich: KOS wurden entgegen Ontologien nicht geschaffen, um wahre Strukturen der Welt abzubilden, sondern ein kontrolliertes Vokabular über das Wissen der Welt zusammenzufassen. Logische Kontingenz spielt in ihrem Entwurf eine untergeordnete Rolle. So erfüllen z.B. im seltensten Fall alle hierarchischen Relationen in einem KOS die Definition für logische Subsumtion, folglich können die KOS-Begriffe auch nicht in einer Klassenhierarchie von RDFS bzw. OWL ausgedrückt werden, ohne dass falsche Inferenzen zu erwarten sind. [Miles/Pérez-Agüera(2007):76]

Das zentrale SKOS-Konzept zur Darstellung von Begriffen `skos:Concept` wird daher nicht als Klasse, sondern schlicht als RDF-resource mit properties wie `skos:narrower` (untergeordneter Begriff), `skos:inScheme` (Begriffssystem, dem der Begriff angehört), `skos:exactMatch` (synonymer Begriff in anderem Begriffssystem) oder `skos:prefLabel` (Vorzugsbenennung) definiert. SKOS verfügt aber in sich über Ansätze von formaler Logik [SKOSref(2009):1.5], etwa durch die Hierarchisierung von Benennungs- und Dokumentations-properties (labels und notes) [Allemang/Hendler(2008):162-163], wie z.B.:

```
skos:scopeNote rdfs:subPropertyOf skos:note . [SKOSref(2009):S17]
```

Zudem werden die Relations-properties (also solche, die Beziehungen zwischen Begriffen darstellen) `skos:broaderTransitive`, `skos:narrowerTransitive` und `skos:related`, `skos:exactMatch` durch inferenzfähige OWL-properties definiert:

```
skos:broaderTransitive
  rdf:type owl:TransitiveProperty ; [SKOSref(2009):S24]
  owl:inverseOf skos:narrowerTransitive . [SKOSref(2009):S26]
skos:related
  rdf:type owl:SymmetricProperty . [SKOSref(2009):S23]
skos:exactMatch
  rdf:type owl:TransitiveProperty . [SKOSref(2009):S45]
```

SKOS bietet neben seiner relativ einfachen Struktur weitere Vorteile. Da es auf RDF(S) und OWL basiert, ist es generell erweiterbar und es können anwendungsbezogene SKOS-Spezifikationen erstellt werden. Andererseits gerät dadurch die Grenze zwischen der schlicht maschinenverarbeitbaren Darstellung von Begriffssystemen und einer Implementierung von formalen logischen Konstrukten prinzipiell durchlässig – je nach Anwendungsszenario sind semi-formale Modellierungen möglich. [Miles/Pérez-Agüera(2007):77-78] Im Hinblick auf Web-2.0-Schlagwortssysteme (folksonomies) könnte SKOS somit eine wichtige Brücke zum Semantic Web bilden:

Looking to the future, SKOS occupies a position between the exploitation and analysis of unstructured information, the informal and socially-mediated organization of information on a large scale, and the formal representation of knowledge. [SKOSref(2009):1.1]

Und mehr noch bietet SKOS für Bibliotheken und andere Gedächtnisinstitutionen eine zentrale Möglichkeit an der Schaffung des Semantic Web mitzuwirken, indem sie ihre jahrzehntelange Arbeit und Erfahrung mit Kontrollierten Vokabularen einbringen können:

What important is that we are experts in developing information standards, and, most importantly we have the most sophisticated skills and experience in knowledge representation. In sum, if librarians transfer their skills to the semantic knitting required for a Semantic Web, we can help build a better Web. [Greenberg/Méndez(2007):7]

5 Bibliografische Kontrolle mit dem Semantic Web?

In den ersten beiden Kapiteln wurden die Struktur der klassischen Bibliografischen Kontrolle, neue Modellierungen des Bibliografischen Universums und die Funktionsweise von bibliografischen Metadaten-Systemen vorgestellt. Die Probleme, die sich durch ihre Struktur und ihre mangelnde Ausrichtung auf die zentrale Informationsstruktur WWW ergeben, wurden benannt. Danach folgte ein Überblick über die Konzeptionen und Technologien, die unter dem Begriff Semantic Web zusammengefasst werden. Ausgehend von diesen Ergebnissen soll nun untersucht werden, wie Semantic-Web-Methoden im Bereich der Bibliografischen Kontrolle angewendet werden können und welche Auswirkungen dies auf die Zukunft von bibliothekarischen Daten haben könnte.

5.1 Metadaten und RDF(S) – das DCMI Abstract Model

Das Abstract Model (DCAM¹⁶³, 2005 bzw. 2008 als deutsche Übersetzung [DCAMde(2008)]) der Dublin Core Metadata Initiative (DCMI) wurde mit dem Ziel entwickelt, die mittlerweile entstandene Vielfalt an Dublin-Core-Schemata und -Vokabularen in ein übergreifendes Strukturmodell zu fassen. Dabei richtete man sich bewusst an der W3C-Empfehlung RDF und seiner Vocabulary Language, RDFS als zukünftiges allgemeines Beschreibungsmo-
dell im WWW aus. Die Zielsetzung der DCMI bestand explizit in der Abstimmung des auch Dublin Core zugrunde liegenden klassischen Metadatenparadigmas mit dem RDF-Modell:

The Abstract Model was designed to bridge the modern paradigm of the unbounded, linked data graph and the more familiar paradigm of the validatable metadata record, providing a foundation for the development of application-independent syntax specifications and constraint languages. [DCMIspec(2010)]

Das DCAM will damit eine „abstract syntax for 'description sets' (metadata records)“ [DCMIspec(2010)] bieten – auffällig ist hier die Revision der Metadaten-Terminologie zugunsten von *Beschreibungssätzen*, dies findet seine Vollendung in der Einführung des neuen Begriffs Description Set Profile. Das DCAM besteht in sich aus drei interagierenden Informationsmodellen: dem Resource Model (Ressourcenmodell), Description Set Model (Beschreibungssatzmodell), und dem Vocabulary Model (Vokabular-Modell).

Der Kern des DCAM bildet das Ressourcenmodell (Resource Model), dem zufolge eine Ressource mittels eines oder mehrerer Eigenschafts-Wert-Paare (property-value pairs) beschrieben wird. Die Beschaffenheit des Werts wird dabei entweder als alphanumerische

¹⁶³ DCMI Abstract Model, siehe <http://dublincore.org/documents/abstract-model/>

Bezeichnung (literal value) oder als „physische, digitale oder begriffliche Entität“ (non-literal value), also als eine reale Entität oder ein Konzept, definiert:

Jeder Wert (value) ist eine Ressource (resource) - eine physische, digitale oder begriffliche Entität [i.e. ein nicht-literaler Wert (non-literal value), K.W.] oder ein Literal (literal), das mit einer Eigenschaft (property) assoziiert wird, wenn ein Paar aus einer Eigenschaft und einem Wert (property-value pair) benutzt wird, um eine Ressource (resource) zu beschreiben. [...] Ein nach RDF definiertes Literal (literal) ist eine Entität, die, um eine Ressource (resource) zu bezeichnen eine Unicode-Zeichenkette sowie optional die Sprache und/oder den Datentyp verwendet. [DCAMde(2008):5]

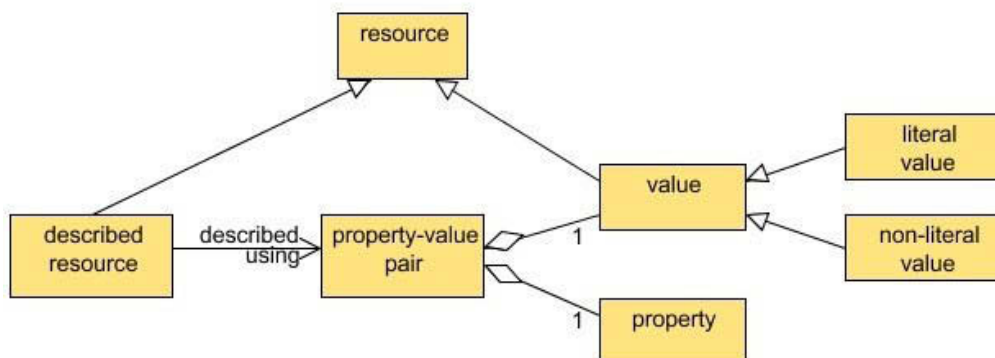


Abbildung 10: DCAM - Ressourcenmodell [DCAMde(2008):5]

Mag der Begriff Eigenschaft-Wert-Paar auch auf den ersten Blick an das klassische Attribut-Wert-Schema erinnern, das (implizit) bibliografischen Beschreibungen und (explizit) klassischen Metadatenschemata zugrunde liegt, wurde hier jedoch eine entscheidende Neuorientierung vorgenommen. Offenkundig wird der traditionelle Begriff des Metadatenelements bzw. Attributs aufgegeben zugunsten des Begriffs *property*, der wie folgt definiert wird:

Ein besonderer Aspekt, eine Eigenart, ein Attribut oder eine Beziehung, die verwendet werden, um Ressourcen (resources) zu beschreiben. [DCAMde(2008):14, Terminologie]

Diese Übernahme des RDF-Konzept *property*, wie von weiteren Begriffen¹⁶⁴, an zentraler Stelle bekräftigt zum einen die enge Bindung des DCAM an die RDF-Struktur. Zum anderen wird mit der Absage an „Metadatenelemente“ implizit der Begriff des Metadatenschemas in Frage gestellt.

An dessen Stelle tritt nun eine „Sammlung“ von Beschreibungsaussagen, die zu einer Ressource gehören – Vorgaben, wie eine solche Sammlung auszusehen hat, können dann davon unabhängig in sogenannten Description Set Profiles gemacht werden. Einerseits wird hier also die klassische Metadaten-Struktur an die des RDF-Modells angepasst, andererseits bleibt es auf einer höheren Ebene möglich, schema-ähnliche Definitionen zu treffen.

¹⁶⁴ Explizit werden diesbezüglich folgende Begriffe des DCAM genannt: resource, class, syntax encoding scheme, has domain, has range, sub-property of, sub-class of, plain value string, typed value string, vgl. [DCAMde(2008):11].

Das auf dem Ressourcenmodell aufsetzende Beschreibungssatzmodell (Description Set Model) legt dementsprechend fest, dass ein oder mehrere Aussagen (statements) über dieselbe Ressource zusammen die *description* (Beschreibung) einer Ressource bilden. Eine solche *description* kann auch über den URI der beschriebenen Ressource (described resource URI) realisiert werden. Eine oder mehrere Beschreibungen (descriptions) bilden in ihrer Gesamtheit ein Beschreibungssatz (*description set*), der in einem Metadatensatz (record) der Ressource realisiert wird. [DCAMde(2008):6-7,10] Dadurch, dass ein verwendeter nicht-literaler Wert seinerseits als resource über eine Beschreibung verfügen kann, kann ein Beschreibungssatz schließlich aus descriptions von zusätzlichen Ressourcen neben der eigentlich beschriebenen Ressource bestehen,

...z. B. eine separate Beschreibung die Metadaten zu einer Person beinhalten, die der Verfasser der beschriebenen Ressource ist. [DCAMde(2008):9]

In der deutschen Übersetzung des DCAM wird genauer erläutert:

Ein Beschreibungssatz (description set) kann beispielsweise sowohl die Beschreibung (description) eines Bildes als auch die des Künstlers enthalten. Außerdem kommt es häufig vor, dass ein Beschreibungssatz (description set) auch die Beschreibung (description) des Beschreibungssatzes (description set) selbst enthält (zuweilen auch als "administrative Metadaten" oder "Meta-Metadaten" bezeichnet). [DCAMde(2008):10]

Das Beschreibungssatzmodell definiert ein *statement* als die Realisierung eines Eigenschafts-Wert-Paars, die property wird dabei (idealerweise) durch einen property URI identifiziert, der Wert durch einen Platzhalter (value surrogate). Ein statement wird im DCAM folgendermaßen definiert:

Die Instantiierung eines Paares aus Eigenschaft und Wert (property-value pair), die aus dem URI der Eigenschaft (property URI) (dem URI, der die Eigenschaft identifiziert) und einem Platzhalter (value surrogate) besteht. [DCAMde(2008):14, Terminologie]

Dabei wird zwischen literal value surrogate und non-literal value surrogate unterschieden. Während ein *literaler Platzhalter* schlicht durch eine Zeichenkette (value string) besetzt wird, wird ein *nicht-literaler Platzhalter* durch den URI des *semantischen Codierungsschemas* (vocabulary encoding scheme URI), aus dem der Wert stammt, und dem URI des nicht-literalen Wertes (value URI) besetzt. Liegt kein URI für das Konzept vor, kann eine Zeichenkette diesen ersetzen. Das *non-literal value surrogate* soll folglich Werte aufnehmen, die aus einem Semantischen Codierungsschema, einem inhaltlichen Datenwertstandard, stammen: dazu werden z.B. das KOS sowie das spezifische Konzept benannt und referenziert oder zumindest dessen Benennung angegeben. [DCAMde(2008):6-7]

Zudem definiert das Beschreibungssatzmodell einen *value string* entweder als plain value string (einfache Zeichenkette) oder typed value string (typisierte Zeichenkette). Eine einfache Zeichenkette sollte mit einer ISO-normierten Sprachangabe und/oder Datentypbezeichnung ergänzt werden; eine typisierte Zeichenkette wird durch ein *Syntaktisches Codierungs-*

schema (syntax encoding scheme), also einem formalen Datenwertstandard, definiert, auf das per URI (syntax encoding scheme URI) verwiesen wird. Explizit wird hier auf XML-kodierte value strings eingegangen:

XML-Inhalte in einer Zeichenkette (value string) werden kenntlich gemacht, indem eine typisierte Zeichenkette (typed value string) zusammen mit der URI eines syntaktischen Codierungsschemas (syntax encoding scheme URI) aus <http://www.w3.org/1999/02/22-rdf-syntax-ns#XMLLiteral> verwendet wird. [DCAMde(2008):10]

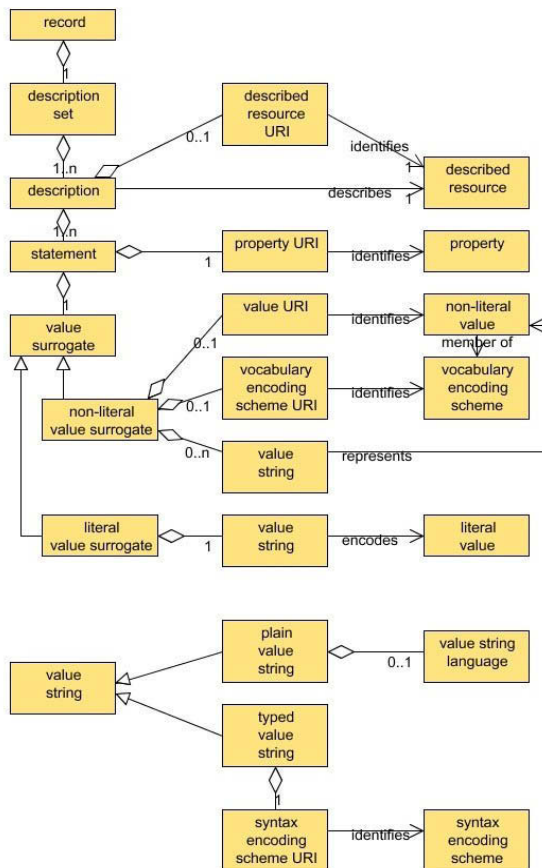


Abbildung 11: DCAM - Beschreibungssatzmodell
[DCAMde(2008):9]

Zentrales Anliegen des DCAM ist also die Referenzierung (und damit eindeutige Identifizierung) nicht nur der beschriebenen Ressource und der Eigenschaften (properties), sondern eben auch der Werte. Im Falle von nicht-literalen Werten geschieht dies durch die URIs des Konzepts und des Semantischen Codierungsschemas, aus dem es stammt; im Falle von literalen Werten, sofern möglich, durch den URI des zugrunde liegenden Syntaktische Codierungsschemas.

Auch hier ist die Ausrichtung am RDF-Modell offensichtlich. Einen Hinweis gibt schon die Wahl des Begriffes *statement*. Beschreibt er in RDF das grundlegende Tripel von Subjekt-Prädikat-Objekt (bzw. *resource-property-value*), kann man im DCAM das *statement*-Paar von *property URI* und *value surrogate* leicht um die spezifische *resource* als Subjekt ergänzen – per *definitionem* bezieht sich eine Beschreibung, und somit auch ein *statement*, auf immer genau eine Ressource. Mit der angestrebten eindeutigen Referenzierung beider Bestandteile der *statements* wird ebenfalls ein Grundprinzip von RDF angewendet.

Die Einbeziehung z.B. der DCMI Metadata Terms und verschiedener externer Vokabulare in das DCAM wird schließlich mit dem Vokabularmodell (*Vocabulary Model*) sichergestellt. Mit dem Konzept *term* wird quasi eine Superklasse aller DCAM-Konzepte geschaffen, die zur Ressourcen-Beschreibung eingesetzt werden. Diese Terme können anwendungsspezifisch in einem *vocabulary* festgelegt werden. Solch ein Vokabular im Sinne des DCAM ist ein Satz von Termen,

...der sich aus Eigenschaften (*properties*) bzw. Elementen (*elements*), Klassen (*classes*), Semantischen Codierungsschemas (*vocabulary encoding schemes*) und/oder syntaktischen Codierungsschemas (*syntax encoding schemes*) zusammensetzt. [DCAMde(2008):9]

Vocabulary ist hier also nicht mit Kontrollierten Vokabularen (KOS) zu verwechseln, die nur einen (optionalen) Teil dieser Sets von Benennungen bilden. Folglich kann ein Term also eine Eigenschaft¹⁶⁵, eine Klasse, ein semantisches Codierungsschema oder ein syntaktisches Codierungsschema sein. Dabei können durch *domains* oder *ranges* von *properties* bestimmte *Klassen* gebildet werden, die *values* des entsprechenden Eigenschaft-Wert-Paares sind dann jeweils Objekte dieser Klassen. Ressourcen können Instanzen von (im Modell nicht weiter erläuterten) Klassen sein und/oder einem Semantischen Codierungsschema angehören – entsprechend des Postulats im Ressourcenmodell, dass jeder *value* wiederum als eine *resource* begriffen werden kann. Ein Syntaktisches Codierungsschema wird als Klasse von Literalen definiert. Zudem ist grundsätzlich die Bildung von *sub-classes* und *sub-properties* ist vorgesehen. [DCAMde(2008):8-9]

¹⁶⁵Die ältere Bezeichnung für Eigenschaften (*properties*) sind die in der oberen Aufzählung genannten Elemente (*elements*).

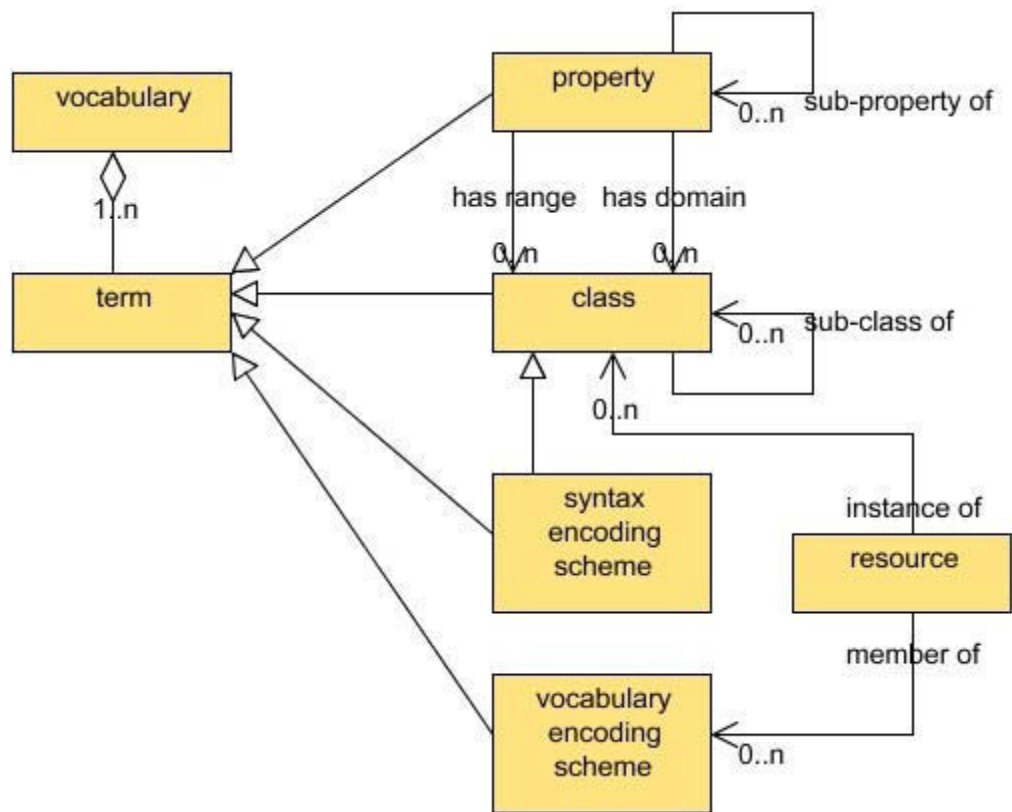


Abbildung 12: DCAM - Vokabularmodell [DCAMde(2008):9]

Durch das Vokabularmodell werden die grundlegenden Mechanismen der Klassenbildung und Vererbung in das DCAM eingeführt. Die zu beschreibende Ressourcen, sowie Eigenschaften und Werte können hierarchisiert werden. Komplexere Strukturen von Metadatenelementen werden damit durch RDF Schema (RDFS) darstellbar.

Das Dublin Core Abstract Model wurde geschaffen, um das „DC-Metadaten-Universum“ umfassend konzeptionell zu definieren. Dies erfolgte klar mit Blick auf das RDF(S)-Datenmodell und seine zukünftige Rolle bei der Etablierung des Semantic Webs. Ziel ist es, ein möglichst allgemeines Modell für Metadaten-Anwendungen anzubieten, das einerseits Kompatibilität zwischen diesen ermöglicht, und andererseits deren strukturelle Konformität zu RDF(S)-Darstellungen sicherstellt. Das DCAM könnte somit als eine Art strukturelle lingua franca für Metadatensysteme betrachtet werden, die WWW-kompatible, semantische Umsetzung dieser garantiert.

5.2 Eine neue Grundstruktur der Bibliografische Kontrolle?

Wie gezeigt wurde, ist die Konzeption von Meta-Informationen für das WWW Grundprinzip des Semantic Webs: Eindeutigkeit und Referenzierbarkeit, Maschinenverarbeitbarkeit sowie der primäre Bezug auf atomare Informationsentitäten wurden als Merkmale benannt und deren Umsetzung im Resource Description Framework (RDF) dargestellt. Die eindeutige Referenzierung der RDF-Konzepte (vorzugsweise durch HTTP-URIs) begründet hierbei einerseits Maschinenverarbeitbarkeit, andererseits die universelle Einbeziehung aller denkbaren Konzepte, gegebenenfalls durch Stellvertreter im WWW. Über diese Eigenschaften hinaus unterscheidet sich RDF in seiner Struktur grundlegend von Bibliografischen Beschreibungen und Metadatenschemata. RDF definiert keine Beschreibungsschemata mit spezifischen Elementen bzw. Attributen und Werten, sondern nur eine allgemeine Beschreibungsstruktur in Form von Aussagen. [McCathieNevile/Méndez(2007):32-33] Zudem können mit RDF Schema (RDFS) Ressourcen und Eigenschaften/Beziehungen mittels Klassen hierarchisiert werden und Geltungsbereiche für Eigenschaften/Beziehungen definiert werden; in der Ontologiesprache OWL sind schließlich umfangreiche logische Konstrukte möglich.

Vom W3C werden RDF und die darauf basierenden „Sprachen“ als zentrale Standards für Meta-Informationen im WWW (aktuell noch in der XHTML-, prospektiv in der HTML5-Recommendation) propagiert. Wenn Bibliotheksdaten also den Schritt zu einer wirklichen Internetpräsenz vollziehen sollen, ist eine Ausrichtung auf die zukünftige „Metadatenstruktur“ des WWW mehr als sinnvoll. Karen Coyle begründet dies nachdrücklich:

The change that libraries will need to make in response must include the transformation of the library's public catalog from a stand-alone database of bibliographic records to a highly hyperlinked data set that can interact with information resources on the World Wide Web. [...] For libraries, this means yet another evolutionary step in the library of our catalog: from metadata to metaDATA. [Coyle(2010a):5-6, Hervorhebung Coyle]

Auf den Netzwerkeffekt, durch den bibliografischen Daten in großem Maße profitieren könnten, weist Martin Malmsten hin:

The promise of the Semantic Web and Linked Data (Berners-Lee 2006) is that it could make data connected, simply by making it available. This, it seems, could be the perfect way for libraries to expose all of their data. [Malmsten(2008):146]

McCathieNevile/Méndez heben die schon vorhandene Verbreitung von RDF/Linked Data in unterschiedlichsten Zusammenhängen hervor, und sprechen von seinem „potential to be the backbone of the Semantic Web“ und in Anlehnung an die Universelle Bibliografische Kontrolle von einem Mittel zur – in Anführungszeichen – „Universal Web Control“. [McCathieNevile/Méndez(2007):4] Dies mag auf den ersten Blick etwas übertrieben erscheinen, sie begründen diese Analogie aber nachvollziehbar:

The Semantic Web community shared with the 'school of bibliographic control' the desire for interoperability and 'universal control' of Web resources. In that community, such interoperability

and the control it gives are based on formal languages and metadata schemas, both machine and human understandable data. [McCathieNevile/Méndez(2007):29]

Wie diese grundsätzliche Konvergenz von Zielen praktisch nutzbar wird, indem man bibliografische bzw. Metadatenstandards Semantic-Web-kompatibel gestalten kann, erläutert Gordon Dunsire anschaulich:

This requires expression of elements of the various standards using Resource description framework (RDF). More specifically, elements relating to metadata structure, such as tags, fields, and attributes, need to be expressed as classes and properties in RDF schema (RDFS), while elements relating to metadata content, such as codes and controlled vocabularies, need to be expressed in Simple knowledge organization system (SKOS). Complicated semantic relationships in vocabularies can be expressed in Web ontology language (OWL). [Dunsire(2009):6]

Die allgemeine Aussagen-Struktur von RDF und die Wissenrepräsentationsmethoden von RDFS, SKOS und Ontologiesprachen erlauben zusammen weitreichende Beschreibungsmöglichkeiten. Mit Letzteren können Elemente aus Metadaten schemata genauso wie Konzepte aus Datenwertstandards, also z.B. Normdatenlisten und Begriffssystemen, erfasst werden. Reformulierungen der verschiedenen Metadatenstandards in RDF(S)/OWL/SKOS könnten als eine langersehnte „Metasprache“ der Bibliografischen Kontrolle fungieren, wie etwa das Grundlagenpapier Understanding Metadata der NISO andeutet:

[RDF] is a data model for the description of resources on the web that provides a mechanism for integrating multiple metadata schemes. [...] Multiple namespaces can be defined, allowing elements from different schemes be combined in a single resource description. Multiple descriptions, created at different times for different purposes, can also be linked to each other. [NISO(2004):11]

Eben darin besteht der grundsätzliche Vorteil von Metadatenformulierungen in RDF. Eine einfache Zusammenführung von Datenelementen und Datenwerten aus unterschiedlichen (nicht nur bibliothekarischen) Quellen wird damit möglich:

That's what real RDF does in the wild: enabling us to put metadata from multiple vocabularies (and even from multiple sources) in the same record. Just as any collection of information in RDF can be merged, it can also be decomposed into the three-part statements we mentioned. [McCathieNevile/Méndez(2007):34]

In der vorgestellten Systematik von Metadatenstandards wurde zwischen Erschließungsstandards, Kodierungsstandards sowie Metadaten schemata und Datenwertstandards unterschieden. Während die ersten Beiden auf konkrete Einsätze von Metadaten bezogen sind, bilden Letztere sozusagen die strukturelle und inhaltliche Basis für verschiedene Anwendungsszenarien. Diese sind es daher auch, auf welche die aktuellen Reformulierungen mittels Semantic-Web-Methoden in erster Linie abzielen.

5.3 Bibliografische Metadaten schemata in RDF(S)

Mittlerweile liegen einige traditionelle bibliografische Metadaten schemata in RDF(S) vor. Vorreiter ist hier ebenfalls die Dublin Core Metadata Initiative (DCMI), die nach den ersten Spezifikationen von DC Simple und Qualified in RDF(S) (2002) diese durch die Repräsentation

tion ihres weitaus umfassenderen Dublin Core Abstract Models in RDF [Nilsson et al(2008)] ersetzt haben. Die frühen Darstellungen von Dublin Core in RDF und die ebenfalls vorhandenen Mappings von DC zu MARC und MAB bieten zudem das Potential, um existierende bibliografischen Daten über die DC-Transformation in RDF darstellbar zu machen. Die LoC bietet hierfür auch ein eigenes Stylesheet an, das MARCXML in RDF/XML-kodiertes DC Simple überführt. [Svensson(2007):4]

2008 haben Rob Styles, Danny Ayers und Nadeem Shabir ein Verfahren zur direkten Darstellung von MARC-21-Daten in RDF [Styles/Ayers/Shabir(2008)] vorgestellt. Die MARC-21-Elemente, die in einer komplizierten Struktur aus Unterfeldern, Identifikatoren usw. vorliegen, werden zunächst in leicht lesbare RDF-Konzepte (also resources und properties) umgewandelt¹⁶⁶, um die MARC-Attribute-Wert-Paare damit als RDF-Aussagen reformulieren zu können. Hier ein Auszug aus dieser vorläufigen Beschreibung in RDF/N3-Syntax:

```
@base <http://example.com/a_marc_record> .
@prefix marc21: <http://example.com/marc21#> .
[]
  marc21:controlNumber
    "9cbbe7fc3a7346d99c281979d45b679c";
#Following data comes from fixed positions in the
Leader
  marc21:recordStatus "New";
  marc21:recordType "Language material";
  marc21:bibliographicLevel "Monograph/item";
  marc21:encodingLevel "Full";
#Following data comes from fixed positions in 008
  marc21:recordCreated
    "1999-08-31"^^xsd:dateTime;
  marc21:publicationStatus "Published";
  marc21:placeOfPublication "England";
  marc21:language "English";... [Styles/Ayers/Shabir(2008):4]
```

Der zweite Schritt besteht dann in der Umwandlung der als literale Werte vorliegenden Objekte in RDF-Ressourcen, die per definitionem über einen URI identifiziert/referenziert werden müssen. Alle Zeichenkette werden dazu normalisiert und einem Namespace zugeordnet, so dass HTTP-URIs anstatt der literal values verwendet werden können. Styles/Ayers/Shabir verweisen an dieser Stelle auf das FOAF-Vokabular zur Beschreibung von Personen, auf SKOS für Begriffe aus Kontrollierten Vokabularen sowie auf andere bibliografische Vokabulare in RDF (MarcOnt, Bibliographic Ontology, RDF Book Mashup¹⁶⁷), die zur Einbeziehung in die MARC-Beschreibung geeignet wären.

¹⁶⁶ Dabei merken sie als potentielle Alternative das MARC-Derivat MODS an, dass sie als „very close being just“ a „ontology based on the full Library of Congress MARC21“ [Styles/Ayers/Shabir(2008):5] bezeichnen. Tatsächlich liegt seit 2006 eine Formulierung von MODS in RDF (2006) von Stefano Mazzocchi inklusive eines spezifischen RDFizers vor, siehe http://simile.mit.edu/wiki/MARC/MODS_RDFizer

¹⁶⁷ Bei RDF Book Mashup von Chris Bizer und Tobias Gauß handelt es sich eigentlich nicht um ein Vokabular, sondern um ein Wrapper-Tool, das APIs von Amazon und Google Base abfragt und die Suchergebnisse in RDF umwandelt, siehe <http://www4.wiwiss.fu-berlin.de/bizer/bookmashup/> RDF Book Mashup nutzt dafür Konzepte aus Dublin Core, SKOS, FOAF und GoodRelations, ein Beispiel: <http://www4.wiwiss.fu-berlin.de/bookmashup/doc/books/3631456794>

Außerdem wird von ihnen eine Routine vorgeschlagen, um die entsprechenden FRBR-Entitäten Werk, Manifestation und Expression (inklusive URIs) aus den RDF-Daten abzuleiten. Die Autoren beziehen sich dabei auf Matching-Kriterien, die im Rahmen einer breiteren Analyse¹⁶⁸ von Tom Delsey an der LoC entwickelt wurden, um aus MARC-Datensätzen Informationen zu Gruppe-1-Entitäten zu extrahieren. Seit 2005 liegen dank Ian Davis und Richard Newman zudem die wichtigsten FRBR-Konzepte als RDF-Vokabular vor [Davis/Newman/D'Arcus(2009)], sodass aufgrund dieser beiden Vorarbeiten Styles/Ayers/Shabir einfache manifestations-, expressions- und werkbezogene statements in ihre Beispieldaten formulieren konnten. [Styles/Ayers/Shabir(2008):Appendix C]

Etwa zeitgleich hat die schwedische Nationalbibliothek an einer Überführung ihrer Verbundkatalog-Daten in eine RDF-Darstellung gearbeitet, Martin Malmsten hat die Ergebnisse 2008 in dem Aufsatz Making a Library Catalogue Part of the Semantic Web vorgestellt. [Malmsten(2008)] Hier wurden erstmals, ähnlich wie bei Styles/Ayers/Shabir angedacht, alle in MARC kodierten bibliografischen und Normdatensätze sowie (generierte) FRBR-Werke mit URIs ausgestattet und als RDF-Aussagen reformuliert. Die RDF-Daten stehen in einem triple store zu Verfügung, der über SPARQL abgefragt werden kann. Um „Schnittstellen“ zum weiteren Semantic Web zu schaffen, werden die Vokabulare Dublin Core, FOAF, SKOS, DBpedia und Annotea¹⁶⁹ einbezogen. [Malmsten(2008):147-149] Die weiteren semantic-web-spezifischen Entwicklungen um den LIBRIS-Onlinekatalog¹⁷⁰ werden unter <http://blog.libris.kb.se/semweb/> dokumentiert. Martin Malmsten fasst zusammen:

A goal when creating the new version of the LIBRIS web interface was to make the information presented to a normal user transparently available to machines/web robots as well. It was also obvious that information not intrinsic to the record itself, such as user annotations and connections to other records could be made available this way. [Malmsten(2008):146]

Neben Dublin Core und MARC als den beiden zentralen Metadatenschemata wurden bereits bibliografische RDF-Vokabulare entwickelt, die einen synthetischen Ansatz verfolgen und damit genuin der RDF-Philosophie folgen. Oben wurden bereits die MarcOnt Ontology sowie die Bibliographic Ontology erwähnt.

¹⁶⁸Functional analysis of MARC 21 Bibliographic and Holdings formats (2002), siehe <http://www.loc.gov/marc/marc-functional-analysis/functional-analysis.html> In ähnlicher Weise werden beim Online Computer Library Center (OCLC) seit 2003 Verfahren zur Gruppierung von Datensätzen bezüglich FRBR-Werken erforscht. [Hickey/O'Neill(2005)]

¹⁶⁹Annotea Schema bzw. Annotea Bookmark Schema sind kleine RDF-Vokabulare zur Beschreibung von Annotationen und Bookmarks (Favoriten), die ja üblicherweise von Internetnutzern selbst erstellt werden, siehe <http://www.w3.org/2001/Annotea/>

¹⁷⁰LIBRIS-OPAC, englischsprachiges Interface, siehe <http://libris.kb.se/?language=en>

Die MarcOnt Ontology¹⁷¹ wurde als integrative bibliografische Ontologie am Digital Enterprise Research Institute (DERI) Galway im Zusammenhang mit der Digital-Library-Software JeromeDL¹⁷² entwickelt. In der Version 1 vereinigte MarcOnt alle Konzepte aus BibTeX, Dublin Core und zum Teil MARC 21 in einer OWL-Ontologie. [Dabrowski/Synak/Kruk(2009)] Mittlerweile liegt die Version 2.0¹⁷³ vor, die gegenüber dem Vorgänger vereinfacht wurde (BibTeX und DC werden weiter voll unterstützt), und die FOAF-Konzepte `foaf:Person`, `foaf:Organization`, `foaf:Group` und `foaf:Agent` sowie das zentrale SKOS-Konzept `skos:Concept` (z.B. als `range` für `marcont:hasKeyword`, `marcont:hasTopic`) einbezieht. [Dabrowski/Synak/Kruk(2009):118-119] Zudem wurde für die Beschreibung von unkontrollierten Schlagwörtern (Tags) die S3B Tagging Ontology¹⁷⁴ eingeführt. [Dabrowski/Synak/Kruk(2009):114-115]

Die Bibliographic Ontology¹⁷⁵ (BIBO) wird seit 2005 von Frédéric Giasson und Bruce D'Arcus als übergreifende bibliografische OWL-Ontologie entwickelt (aktuelle Version: 2009/11/04), die eher am Endnutzer orientiert ist, indem sie auf Kompatibilität zu Textverarbeitungen und variable Expressivität und Zitationsstile Wert legt. [Dabrowski/Synak/Kruk(2009):119-120] „Bibliontology“ nutzt Konzepte von Dublin Core inklusive DC Terms, Element Refinements und Encoding Schemes, sowie zahlreiche andere Vokabulare, etwa FOAF, Event und Timeline Ontology, Creative Commons Schema¹⁷⁶ wie man der aktuellen Spezifikation¹⁷⁷ entnehmen kann. Giasson/D'Arcus erläutern auf der Homepage sehr anschaulich den Grundgedanken von Bibliographic Ontology – wie von Linked Data im Allgemeinen:

Instead of covering all topics within the Bibliographic Ontology [sic] itself, we describe the basic topics and build into a larger framework - RDF - that allows us to take advantage of work elsewhere on more specific description vocabularies.

RDF provides the Bibliographic Ontology with a way to mix together different descriptive vocabularies in a consistent way. Vocabularies can be created by different communities and groups as appropriate and mixed together as required, without needing any centralized agreement on how terms from different vocabularies can be written down in XML or N3. [...] In summary then, RDF is self-documenting in ways which enable the creation and combination of vocabularies in a devolved manner. [Giasson/D'Arcus(2009)]

171 MarcOnt, siehe <http://www.marcont.org/>

172 DERI, siehe <http://www.deri.org/> und JeromeDL, <http://www.jeromedl.org/>

173 MarcOnt Ontology Specification 2.0, http://www.marcont.org/ontology/index.html#term_marcont-2.0.owl

174 S3B Tagging Ontology, siehe <http://s3b.corrib.org/tagging/spec/>

175 Bibliographic Ontology, siehe <http://www.bibliontology.com/>

176 Event Ontology, siehe <http://motools.sourceforge.net/event/event.html> Timeline Ontology, <http://motools.sourceforge.net/timeline/timeline.html> Creative Commons Schema, <http://creativecommons.org/ns>

177 Bibliographic Ontology Specification, Revision: 1.3 2009, siehe <http://bibotools.googlecode.com/svn/bibo-ontology/tags/1.3/bibo.xml.owl>

5.3.1 FRBR als bibliografischen Ontologie?

Die neuen und in Diskussion befindlichen Konzeptmodelle FRBR, FRAD und FRSAD wurden im ersten Kapitel vorgestellt. Es ist klar, dass auch die ersten beiden Konzeptionen, die als veröffentlichtes Dokument vorliegen, nicht als endgültig betrachtet werden können – schon allein aus dem Grund, da das Zusammenspiel der drei Modelle, insbesondere auch durch den aktuellen Entwurf des FRSAD, nicht geklärt ist. Zudem wurden aus praxisnahen Sichtweisen verschiedene kritische Anmerkungen und Verbesserungsvorschläge eingebracht. [Coyle(2010b):24-25] Gleichwohl ist mit der FRBR-Familie¹⁷⁸ der Grundstein für eine analytisch begründete Neuausrichtung des Bibliografischen Kontrolle gelegt worden. Deren Teilbereiche wurden einer Funktionsanalyse unterworfen und Elemente und Attribute, die in bibliografischen Beschreibungen, Normdaten und Metadatenschemata verwendet werden, in Form eines allgemeinen Entitäten-Relationen-Modells redefiniert. Es ist in gewissem Sinn erstaunlich, aber ein nicht zu leugnender Fakt, dass ein übergreifendes Modell des Bibliografischen Universums – trotz einer breiten und elaborierten Vielfalt von bibliografischen Schemata und Formaten – bisher fehlte. Den Unterschied zwischen so einer expliziten Modellierung und impliziten Modellen, wie sie in diversen Schemata und Formaten enthalten sind, formuliert Jenn Riley anschaulich:

Both abstract and conceptual models can be either explicit or implicit. Most elements sets in the library community fall into the implicit category – there is *no formal explicitly defined conceptual model* underlying MARC or MODS, for example. Metadata elements sets in this implicit category are generally defined by their encodings, rather than specifying both a conceptual model and one or more bindings for that model into an encoding syntax. [Riley(2008):125, Hervorhebung K.W.]

Eine allgemeine Abstraktion von bibliografischen Entitäten, ihrer Attribute und Beziehungen stellt die genuine Leistung der FRBR-Rahmenwerks dar. Ein übergreifendes Konzeptmodell ist von entscheidender Bedeutung für eine wirkliche Metadaten-Interoperabilität: Nicht nur auf der Kodierungsebene wie XML, sondern durch die verbindliche Klärung der Bedeutung von bibliografischen Konzepten. Diese gibt FRBR eben nicht durch die Erstellung und Dokumentation eines weiteren Metadatenschemas, sondern durch eine abstrakte Modellierung des Bibliografischen Universums – die durch Neutralität gegenüber Datenmodellen und spezifischen Kodierungen Grundlage für verschiedene Realisationen bilden kann. [Maxwell(2008):1] Besonders offenbar wird dieser Fortschritt in der Systematik und Darstellbarkeit der bisher verborgenen Beziehungen zwischen den bibliografischen Entitäten [Coyle(2010b):23], die das Navigieren zwischen diesen als eine stark erweiterte Funktionalität des Katalogs ermöglicht:

¹⁷⁸ Da es sich im Folgenden um allgemeine Überlegungen handelt, wird hier „FRBR“ – wie oftmals – stellvertretend für die gesamte Familie von Konzeptmodellen benutzt.

The entity-relationship model has great potential both to streamline data creation operations such as cataloging and, from a user's perspective, to allow much better search functionality and greater clarity in information display. [Maxwell(2008):4]

Die konsequente Anwendung des FRBR-Modell auf Onlinekataloge¹⁷⁹ wird das klassische Katalog-Verständnis als Ansammlung von (in geringem Ausmaß verknüpften) Titelaufnahmen bzw. Datensätzen notwendig auflösen. Klassische Bibliografische Kontrolle und Metadaten-systeme beschränkten sich auf wenige Entitäten, zu beschreibende Manifestationen und zu normierende Eigennamen und Begriffe. Dafür bieten Attribut-Wert-Schemata eine sinnvolle Struktur. Das FRBR-Modell will dagegen das Gesamtbild der bibliografischen Beziehungen abbilden und nutzbar machen. Dies lässt sich nur durch eine komplexe Netzstruktur von Entitäten, die über Beziehungen und gemeinsame Attribute verbunden sind, realisieren. Es wurde bereits gezeigt, dass bibliografische Metadaten als Dekonstruktion von Titelaufnahmen betrachtet werden können, indem einzelne bibliografische Elemente in Form eines Attribut-Wert-Schemas ansprechbar wurden. Es wurde ebenso festgestellt, dass die Elemente aber notwendig an diese Schemata gebunden sind, die zudem wenig flexibel und interoperabel sind, da es ihnen an einer übergreifenden und expliziten Bedeutungsstruktur fehlt. FRBR bietet nun zum einen ein Modell, dass die bibliografische Sphäre mit einer solchen gemeinsamen Konzeption unterlegt. Zum anderen werden die wenigen Beschreibungen von Entitäten mittels Attribut-Wert-Schemata prospektiv durch die Wiedergabe des umfassenden Geflechts aller Entitäten, Attribute und Beziehungen abgelöst. Prinzipiell könnte man also von einer zweiten Dekonstruktion, eben der der Metadaten-schemata, sprechen. Bibliografische Elemente, die vorher als Attribute in einem Metadaten-schemata „gefangen“ waren, werden im ER-Modell unabhängig – eben zur Entität, die selbst über Attribute und Beziehungen verfügen kann. Bisher vernachlässigte bibliografische Entitäten wie Expressionen, Familien oder Themen können somit separat beschrieben werden. Ein weiterer zentraler Vorteil von FRBR gegenüber Metadaten-schemata besteht in der grundsätzlichen Erweiterbarkeit der Struktur, d.h. auch der spezifischen Anpassung. Praktisch bedeutet dies: Welche Konzepte wie in einem Anwendungsszenario beschrieben werden, ist in einer netzartigen Struktur sehr viel freier definierbar, grundsätzlich handelt es sich in jedem Fall stets um einen Ausschnitt aus dieser Struktur, somit bleiben alle spezifischen Beschreibungen aus unterschiedlichen Szenarien kompatibel.

FRBR kann folglich in Abgrenzung zu Metadaten-schemata als bibliografisches Vokabular bzw. leichtgewichtige Ontologie verwendet werden. Als Basis für die zukünftige Darstellungsform von bibliografischen Daten und Beziehungen wird zur Zeit die Umsetzung der FRBR-Konzeption in Katalogsystemen erprobt. Die Eingangsbeispiele von

¹⁷⁹ Diese Umgestaltung von Onlinekatalogen wird oft mit „FRBRizing“ oder deutsch „FRBRisierung“ umschrieben.

Styles/Ayers/Shabir und Malmsten haben dies schon gezeigt. Dort werden Entitäten der Gruppe 1 zusätzlich zu MARC-RDF-Realisationen verwendet, nachdem sie aus bestehenden Titeldatensätzen abgeleitet wurden.

Das Online Computer Library Center (OCLC) beschäftigt sich dagegen schon länger, unabhängig von Semantic-Web-Technologien, mit der versuchsweisen Anwendung der FRBR¹⁸⁰, seit 2002 wird an dem Langzeitziel der „FRBRisierung“ des WorldCat gearbeitet. [Hickey/O'Neill(2005)] Daraus haben sich einige Unterprojekte entwickelt, etwa der FRBR Work-Set Algorithm¹⁸¹, der durch den Abgleich von Titel- und Verfasserinformationen (aus Titel- und Personendatensätzen) die verschiedenen Manifestationen eines Werkes zusammenführt und in einem Werk-Datensatz bzw. einer Werk-Seite („work level record“) aggregiert. Diese Werk-Seite existiert bislang nur in HTML (bzw. die Manifestations-Informationen in einfachem XML), verfügt aber über automatisch generierte Grafiken etwa zu Publikationsgeschichte und Notationen. Der FictionFinder¹⁸² baut auf dem Work-Set Algorithm auf und stellt sozusagen eine schlanke Variante dieser Werk-Seiten für 2,9 Millionen Werke der Belletristik zur Verfügung. Es leuchtet natürlich ein, dass gerade bei einer Vielzahl von Auflagen und Medienformen, wie es im Bereich der Schönen Literatur häufiger der Fall ist, die Anwendung von FRBR-Primärbeziehungen ein vielversprechendes Versuchsfeld ist.

Weder der FRBR-Report noch die nachfolgenden Studien¹⁸³ geben eine konkrete Empfehlung zur Umsetzung in der Bibliografischen Kontrolle, mit den gleichzeitigen Entwicklungen der maschinenverarbeitbaren Web-Beschreibungsstandards des W3C ergeben sich aber äußerst günstige Überschneidungen. Das FRBR/FRAD/FRSAD-Universum kann, obwohl nicht geplant, mit dem RDF-Beschreibungsmodell und dem Ontologie-Konzept aufgrund struktureller Ähnlichkeit gut abgebildet werden. [Coyle(2010b) I:7] Eben dies wurde bereits 2005 von Stefan Gradmann vorgeschlagen:

The proposal is to rethink the technical platforms for librarian metadata implementation in terms of Semantic Web technology and to do so using FRBR as a kind of pivot concept. In that sense, the proposal is not to view FRBR as a kind of ontology to be expressed in RDF, but rather to consider it a kind of specific meta-ontology in the field of librarian information objects, which would have to be expressed using RDFschema (or OWL) as a consequence and which, in turn, would be a suitable basis for catalogue implementation using RDF. [Gradmann(2005):70]

Ian Davis und Richard Newman haben im selben Jahr mit der konkreten Definition von FRBR in RDF(S)/OWL/SKOS begonnen. [Davis/Newman/D'Arcus(2009)] Diese maschinen-

180OCLCs FRBR-Entwicklungsprojekte sind auf folgender Übersichtsseite zusammengefasst: <http://www.oclc.org/research/activities/past/orprojects/frbr/default.htm>

181OCLC FRBR Work-Set Algorithm, siehe Beispiele von „work pages“ <http://frbr.oclc.org/research/pages/> und die Projektseite <http://www.oclc.org/research/activities/past/orprojects/frbralgorithm/algorithm.htm>

182OCLC FictionFinder, siehe <http://fictionfinder.oclc.org/> bzw. die Projektseite <http://www.oclc.org/research/activities/fictionfinder/default.htm>

183 In der FRSAD-Studie wird jedoch u.a. auf die Kompatibilität mit SKOS hingewiesen.

verarbeitbare Repräsentation beschränkt sich auf „Kernkonzepte“ von FRBR, also die Entitäten aller drei FRBR-Gruppen sowie deren Beziehungen untereinander, nicht jedoch die Attribute, die ja auch im FRBR-Report nicht mit dem Anspruch einer Vollständigkeit aufgeführt sind.

Zusammenfassend lässt sich also eine gewisse Bandbreite von bibliografischen Metadaten-schemata ausmachen, die in RDF(S) vorliegen oder deren Formulierung aktuell betrieben wird. Zum einen sind dies klassische Schemata wie Dublin Core und MARC21, zum anderen synthetische Vokabulare wie MarcOnt und Bibliographic Ontology. Dazu kommt das Bestreben, das FRBR-ER-Modell als bibliografische Ontologie zu gebrauchen und dieses komplexe Beziehungsgeflecht in RDF, d.h. auch als Linked Data, ausweisen zu können.

Die andere, netzartige Struktur, die FRBR gegenüber dem klassischen Metadaten-schema besitzt, prädestiniert es dazu, bibliografischen Daten zur Kompatibilität mit dem entstehenden Semantic Web zu verhelfen. Und umgekehrt stellen Semantic-Web-Technologien die Mittel bereit, um FRBR darstellbar und nutzbar zu machen. In der radikalen Umsetzung bedeutet dies eine Abkehr vom Attribut-Wert-basierten Schema hin zu atomaren Beschreibungsaussagen. Diese können in variablen Aggregationen oder „Ansichten“ zusammengeführt, aber ebenso explorativ genutzt werden, und sie führen über den bibliografischen Bereich hinaus, indem sie über Beziehungen zu anderen Vokabularen verfügen:

Not only we gain this navigation capability within the library system, but the entities, relationships, and various attributes, *freed from the confines of the bibliographic record*, could begin to interact with data in other networks, including the mother of all networks, the World Wide Web. [Coyle(2010b):24, Hervorhebung K.W.]

5.4 Datenwertstandards in RDF(S) und SKOS

Neben der Umwandlung von bibliografischen Metadaten-schemata in RDF(S) besteht die zweite zentrale Herausforderung für eine WWW-orientierte Bibliografische Kontrolle darin, Datenwertstandards, sowohl inhaltliche wie formale, ebenfalls in Semantic-Web-Standards zu überführen. Hier sei angemerkt, dass in der oben beschriebenen Bibliographic Ontology bereits einige Datenwertstandards einbezogen sind, etwa die von der Dublin Core Metadata Initiative anerkannten Encoding Schemes. In der DC-Terminologie wird ebenfalls zwischen Vocabulary Encoding Schemes (also sachliche Standards) und Syntax Encoding Schemes (formale Standards) unterschieden. Der Bereich der formalen Datenwertstandards stellt eher eine allgemeine Normierungsaufgabe dar, die durch ISO, W3C oder IETF wahrgenommen wird. Sachliche Datenwertstandards sind dagegen klassische Mittel und Erzeugnisse von Gedächtnisinstitutionen, die inhaltliche Erschließung betreiben. Wie dargestellt, lassen sich hier traditionell der Bereich der Normdaten und der der Kontrollierten Vokabulare (KOS) unter-

scheiden. In diesem Abschnitt sollen Beispiele für die WWW-orientierte Darstellung dieser inhaltlichen Datenwertstandards im Bibliotheksbereich vorgestellt werden.

Der Nutzen, der sich durch die maschinenverarbeitbare Bereitstellung von Normdaten und KOS ergibt, ist offensichtlich. Deren Nachnutzbarkeit und Interoperabilität ist nicht nur für Bibliotheken und verwandten Einrichtungen von großer Bedeutung, sondern stellt auch einen wichtigen Service für viele Internet-Nutzer dar. Für die Allgemeinheit sind Datenwertstandards vermutlich sogar attraktiver als bibliografische Beschreibungsdaten, wie Harper/Tillett schreiben:

Incorporating elements used in library resource descriptions into the sets of resource properties that are used in to enable the Semantic Web is a large step, and enables library metadata to interoperate with Dublin Core and other RDF encoded metadata. Authority records, library thesauri, and library controlled vocabularies, if converted into formats that support Semantic Web technologies, have an *even greater potential* for revolutionizing the way users – and machines – interact with information on the Internet. [Harper/Tillett(2007):53, Hervorhebung K.W.]

Wie Ontologien für das Semantic Web allgemein, stellen Vokabulare bzw. leichtgewichtige Ontologien für die bibliografischen RDF-Beschreibungen bedeutungsvolle Konzepte bereit. Dabei verfügen sie im Ansatz über eine maschinenverarbeitbare Semantik, die es erlaubt, logische Ableitungen herzustellen. Neben der eindeutigen Referenzierung von RDF-Ressourcen ist hier der W3C-Standard zur Darstellung von Begriffssystemen in RDF(S), SKOS, von zentraler Bedeutung. Bei dessen Erläuterung wurde auch auf die Möglichkeit hingewiesen, OWL-Konzepte einzubeziehen und so semiformal zu modellieren.

Die Schwierigkeiten bei der Umsetzung von bibliografischen Metadatenschemata in RDF(S) sind vor allem auf die Existenz mehrerer konkurrierender Vokabulare nebeneinander und die fehlende bzw. nur versuchsweise Spezifikation des zukünftigen Rahmenwerks FRBR zurückzuführen. Im Bereich der inhaltlichen Datenwertstandards erscheinen diese Hindernisse zumindest teilweise kleiner, so stellt die Formulierung in SKOS für Begriffssysteme den unumstrittenen Weg dar. Personennormdaten sind die ersten Normdaten i.e.S., die mittels RDF(S) dargestellt werden. Dies geschieht bislang vorrangig mit bekannten, einfachen Vokabularen, jedoch werden auch hier Probleme mit komplexeren Strukturen offenbar.

5.4.1 Normdaten im Semantic Web

Normdaten, insbesondere Personennormdaten, werden von großen Bibliotheken schon vermehrt als Linked Data zur Verfügung gestellt. Hervorzuheben sind hierbei die von der Linking-Open-Data-Community eingeführten personenbezogenen Vokabulare FOAF und SIOC, an die Umsetzung von bibliothekarischen Konzepten zur Beschreibung von Personen (sowie Körperschaften) tastet man sich dagegen gerade erst heran.

Ein zentrales Desiderat der Normenkontrolle bestand und besteht in einer Verknüpfung der nationalen Normdaten. Dieses Ziel verfolgt das Projekt der Virtual International Authority

File¹⁸⁴ (VIAF), das gemeinsam von der LoC, der DNB und der Bibliothèque nationale de France (BnF) voran getrieben und vom Online Computer Library Center (OCLC) gehostet wird. In einer Anlaufphase seit 2002 wurden die nationalen Personennamensätze der LCNAF und der PND jeweils mit Daten zur Individualisierung angereichert, über Protokolle der Open Archives Initiative¹⁸⁵ (OAI) zusammengeführt und aufeinander bezogen (Matching). Schließlich wurde ein recherchierbarer OAI-Server aufgebaut, der sich per Metadaten-Harvesting in den nationalen Normdatenbanken eigenständig aktualisiert. [Tillett(2004):38] Aktuell beteiligen sich 15 Institutionen, vor allem nationalbibliografische Zentren, mit 18 Personennamen-Normdateien an der VIAF. Ein Ausbau in Bezug auf geografische Normdaten ist geplant. Die VIAF entspricht der Vision einer gleichrangigen Vernetzung von nationalen Normdaten, wie sie die MLAR Working Group in ihrer Studie [MLARreport(1998)] formuliert hat. Dem Nutzer wird es so möglich, mit seinen individuellen bzw. kulturellen Präferenzen (Namensformen, Sprache, Schrift) international nach Personen (und damit auch ihren Werken) zu recherchieren. [Hickey(2009)]

Zudem wird im Zusammenhang mit der VIAF für dieses Jahr die Einführung eines International Standard Name Identifier¹⁸⁶ (ISNI) für natürliche und Rechtspersonen sowie fiktionale Charaktere avisiert, der zwingend über einen URI identifiziert sein soll. British Library, BnF und OCLC sind neben anderen internationalen Vereinigungen Gründungsmitglieder der ISNI Registration Authority, die Einbeziehung des ISNI in Personennormdateien, insbesondere der VIAF, steht damit bereits fest. Community-übergreifend werden die zahlreichen Namensformen in unterschiedlichen Sprachen und Kontexten zusammengeführt und ein einfacher, universeller Nachweis von Personen möglich. Insbesondere schließt dies eine eindeutige Referenzierbarkeit von (bibliografisch bedeutsamen) Personen als Ressourcen im Sinne des Semantic Webs ein.

Die VIAF bietet heute schon ihre Daten neben einer normalen Browser-Darstellung auch als RDF-Tripel bzw. Linked Data an. Wie im folgenden auszugsweisen Beispiel¹⁸⁷ deutlich wird, werden die unterschiedlichen Repräsentationen von Personen durch die property `skos:exactMatch` verbunden und das FOAF- und DCMI-Terms-Vokabular benutzt:

```
<rdf:RDF xml:base="http://viaf.org/">
<skos:Concept rdf:about="viaf/73981446.rwo">
  <skos:inScheme rdf:resource="viaf-scheme/#concept"/>
  <skos:inScheme rdf:resource="viaf-scheme/#personalNames"/>
  <skos:altLabel>Goethe, Christiane  $\frac{4}{5}$  von  $\frac{7}{3}$  1765-1816</skos:altLabel>
[...]
```

184 VIAF, siehe <http://viaf.org/> und <http://www.oclc.org/research/activities/viaf/>

185 Open Archives Initiative, siehe <http://www.openarchives.org/>

186 ISNI, siehe <http://www.isni.org/>

187 RDF/XML-Darstellung des VIAF-Datensatzes zu Christiane Vulpius, <http://viaf.org/viaf/73981446.rdf>


```
<skos:exactMatch rdf:resource="http://viaf.org/processed/NKC|
jn20000701921"/>
<skos:exactMatch rdf:resource="http://id.loc.gov/authorities/n81041761#con-
cept"/>
[...]
<skos:exactMatch rdf:resource="http://libris.kb.se/auth/207600"/>
<skos:exactMatch rdf:resource="http://d-nb.info/gnd/118628011"/>
[...]
<skos:exactMatch rdf:resource="viaf/73981446.m21"/>
<skos:exactMatch rdf:resource="viaf/73981446.unimarc"/>
<skos:exactMatch rdf:resource="viaf/73981446.viaf"/>
<skos:exactMatch rdf:resource="http://en.wikipedia.org/wiki/Christiane_Vul-
pius"/>
<skos:exactMatch rdf:resource="http://dbpedia.org/page/Christiane_Vulpius"/>
<skos:changeNote xml:lang="en">Transcribed by agency: OCoLC</skos:changeNo-
te>
<skos:changeNote xml:lang="en">Modified by agency: OCoLC</skos:changeNote>
<dcterms:identitifer>73981446</dcterms:identitifer>
<dcterms:created
rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2009-03-
03T12:03:20+00:00</dcterms:created>
<dcterms:modified
rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2009-09-
21T11:33:17+00:00</dcterms:modified>
<dcterms:type xml:lang="en">person</dcterms:type>
<foaf:page rdf:resource="http://www.worldcat.org/wcidentities/lccn-n81-
041761"/>
<foaf:page rdf:resource="http://en.wikipedia.org/wiki/Christiane_Vulpius"/>
<foaf:page rdf:resource="http://dbpedia.org/page/Christiane_Vulpius"/>
<foaf:name>Goethe, Christiane  $\frac{4}{5}$  von  $\frac{7}{3}$  1765-1816</foaf:name>
<foaf:name>Vulpius, Christiane, 1765-1816</foaf:name>
</skos:Concept>
</rdf:RDF>
```

Mit der breiten Beschreibung von Personen (z.B. Einbeziehung des Wikipedia-Artikels) und der Herstellung der internationalen maschinenverarbeitbaren Verknüpfung wird die VIAF mit Sicherheit eine zentrale Quelle für bibliografische Personennormdaten im WWW werden. Dass diese über den bibliothekarischen Bereich hinaus wirkmächtig sein kann, hat Barbara Tillett schon 2004 angemerkt:

This stage is in future, but reminds us of the opportunities that libraries have now to contribute to the infrastructure of the future Internet environment. We can envision a shared international authority file being an integral part of a future 'Semantic Web'. [Tillett(2004):38]

Recht überraschend hat die Deutschen Nationalbibliothek im April dieses Jahres ebenfalls die testweise Verfügbarmachung von Personen- und Sachnormdaten in Semantic-Web-Formaten bekannt gegeben¹⁸⁸. Sie beziehen sich dabei explizit auf die Ziele und beabsichtigte Unterstützung der Linked-Data-Community:

Die Deutsche Nationalbibliothek (DNB) möchte mit ihren größtenteils intellektuell gepflegten Daten einen entscheidenden Beitrag zur Stabilität und Zuverlässigkeit der 'Linked-Data-Cloud' leisten. Mit ihren qualitativ hochwertigen Daten beabsichtigt die DNB, eine der tragenden Säulen des semantischen Netzes zu werden. [Hannemann(2010)]

In der von Jan Hannemann erstellten Dokumentation, die über das Wiki der DNB zugänglich ist [Hannemann(2010)], wird die geplante technische Infrastruktur des DNB Linked Data

¹⁸⁸ Vgl. die Email von Vera Binz im Namen der DNB auf der InetBib-Mailingliste vom 16.04.2010, <http://www.ub.uni-dortmund.de/listen/inetbib/msg41983.html>

Service als ein über SPARQL ansprechbaren triple store vorgestellt. Vorerst sind die RDF/XML-Daten über eine eingeschränkte Instanz des Onlinekatalogs unter <http://devel.d-nb.de/linked/> recherchierbar oder in Notation3 per FTP-Download zugänglich.

Hannemann dokumentiert zudem die allgemeinen Erwägungen in der DNB, welche Ontologien sinnvoll in die Normdaten-Repräsentation einzubeziehen wären. Dabei werden wenige bestehende Ontologien verworfen (darunter die Bibliographic Ontology aufgrund Nichtkonformität mit dem FRBR/FRAD-Modell), mehrere spezifischere Ontologien dagegen als potentielle punktuelle Ergänzungen gewürdigt, etwa GeoNames, die Music Ontology, das Relationship Vocabulary und eventuell auch die MarcOnt-Ontologie. Außerdem wird eine eingehendere Prüfung bzw. Verwendung der DCMI Metadata Terms avisiert. Das FOAF-Vokabular, das für Personenbeschreibung prädestiniert ist, weist bei genauerer Betrachtung Probleme in der Konformität zu bibliografischen Personendatenmodellen auf: verfügt es einerseits über eine zu begrüßende höhere Granularität von Namenselementen (`foaf:familyName`, `foaf:givenName`, `foaf:firstName` und `foaf:lastName`), fehlt andererseits die Unterscheidung von präferierten und alternativen Namensformen sowie die eindeutige Zuordnung von Namensbestandteilen zueinander. [Hannemann(2010)]

Als Grundlage für ihre Normdaten-Repräsentation verwendet die DNB vorerst die Spezifikation FRBR Entities for RDA¹⁸⁹ aus den RDA Element Sets bei der NSDL Registry¹⁹⁰, da die offizielle FRBR-Definition durch die FRBR Review Group noch aussteht. Die DNB betont den zukünftig zentralen Stellenwert, den „sowohl die Ontologie als auch das Regelwerk RDA (Resource Description and Access)“ [Hannemann(2010)] im Bibliothekswesen haben wird.

Demzufolge besteht das synthetische Vokabular¹⁹¹, in dem die Personen-Normdaten gefasst werden, hauptsächlich aus RDA-Konzepten für die FRBR/RDA-Gruppe-2-Entitäten. Diese liegen ebenso bei den RDA Elements Sets der NSDL Registry bereits vor: die RDA Group 2 Elements (Präfix: `rdaGr2`) sowie die RDA Relationships for Persons, Corporate Bodies, Families (`rdaRelGr2`). Angereichert wird dies durch properties aus FOAF (`foaf:page`) und dem Relationship Vocabulary zur Definition von familiären Beziehungen (Präfix: `relationship`). Da die Regeln für die Namensansetzungen in RDA noch nicht festgelegt sind, wurde provisorisch ein eigener Namensraum (`gnd`) mit Konzepten geschaffen, um weitere vorhandene Daten aus der Personennamendatei (PND) darstellen zu können, etwa die properties `gnd:preferredNameForThePerson` bzw. `gnd:variantNameForThePerson`. Es wird

189 FRBR Entities for RDA von Diane Hillmann und Gordon Dunsire, <http://metadataregistry.org/schema/show/id/14.html>

190 RDA Vocabularies in der NSDL Registry, siehe <http://metadataregistry.org/rdabrowse.htm>

191 Das gesamte Mapping der PICA-Datenfelder in RDF-Konzepte dieser Vokabulare ist einsehbar, <https://wiki.d-nb.de/download/attachments/34963694/Tabelle1.pdf?version=1&modificationDate=1269939679000>

darauf hingewiesen, dass diese GND-Konzepte vorläufige Definitionen sind, die Änderungen durchlaufen können. Auch hier soll auszugsweise die RDF/XML-Repräsentation des Datensatzes zu Christiane von Goethe¹⁹² dargestellt werden:

```
<rdf:RDF>
<rdf:Description rdf:about="http://d-nb.info/gnd/118628011">
<gnd:variantNameForThePerson>Vulpius, Johanna Christiana Sophia</gnd:variantNameForThePerson>
[...]
<gnd:preferredNameForThePerson rdf:parseType="Resource">
<gnd:prefixName>von</gnd:prefixName>
<gnd:usedRules>RAK-WB</gnd:usedRules>
<gnd:surname>Goethe</gnd:surname>
<gnd:foreName>Christiane</gnd:foreName>
</gnd:preferredNameForThePerson>
<foaf:page>http://de.wikipedia.org/wiki/Christiane_von_Goethe</foaf:page>
<gnd:variantNameForThePerson>Goethe, Johanna Christiane von</gnd:variantNameForThePerson>
[...]
<gnd:variantNameForThePerson rdf:parseType="Resource">
<gnd:prefixName>von</gnd:prefixName>
<gnd:usedRules>RAK-WB</gnd:usedRules>
<gnd:surname>Goethe</gnd:surname>
<gnd:foreName>Johanna Christiane</gnd:foreName>
</gnd:variantNameForThePerson>
<rdaGr2:biographicalInformation xml:lang="de">Ehefrau J. W. von Goethes</rdaGr2:biographicalInformation>
<relationship:spouseOf rdf:resource="http://d-nb.info/gnd/118540238"/>
<gnd:preferredNameForThePerson>Goethe, Christiane von</gnd:preferredNameForThePerson>
<gnd:publicationOfPerson xml:lang="de">Goethe, Johann W. von: Briefwechsel</gnd:publicationOfPerson>
<gnd:publicationOfPerson xml:lang="de">Goethe, Johanna C. von: Briefe an Nicolaus Mayer. - 1887</gnd:publicationOfPerson>
<rdaGr2:placeOfBirth xml:lang="de">Weimar</rdaGr2:placeOfBirth>
<owl:sameAs>http://dbpedia.org/resource/Christiane_Vulpius</owl:sameAs>
<rdaGr2:dateOfBirth>1765</rdaGr2:dateOfBirth>
<rdaGr2:placeOfDeath xml:lang="de">Weimar</rdaGr2:placeOfDeath>
<gnd:countryCode>XA-DE</gnd:countryCode>
<rdaGr2:dateOfDeath>1816</rdaGr2:dateOfDeath>
<gnd:variantNameForThePerson>Vulpius, Johanna Christiana Sophia</gnd:variantNameForThePerson>
<rdaGr2:identifierForThePerson>(DE-588c) 4064083-8</rdaGr2:identifierForThePerson>
<relationship:parentOf rdf:resource="http://d-nb.info/gnd/11854022X"/>
[...]
<owl:sameAs>http://viaf.org/viaf/73981446</owl:sameAs>
<relationship:siblingOf rdf:resource="http://d-nb.info/gnd/118770489"/>
<rdaGr2:identifierForThePerson>(DE-588) 118628011</rdaGr2:identifierForThePerson>
<gnd:variantNameForThePerson>Vulpius, Johanna C.</gnd:variantNameForThePerson>
</rdf:Description>
</rdf:RDF>
```

Gemäß dem FRAD-Modell werden hier durch die Elemente `gnd:preferredNameForThePerson` und `gnd:variantNameForThePerson` mehrere Namensentitäten für eine Person geschaffen, um gegebenenfalls für diese Namensbestandteile und Ansetzungsregeln darstellen zu können. Modellhaft wird dies in der Dokumentation der DNB folgendermaßen ausgedrückt:

¹⁹² DNB-Normdatensatz Christiane von Goethe, siehe <http://devel.d-nb.de/linked/gnd/118628011/about>

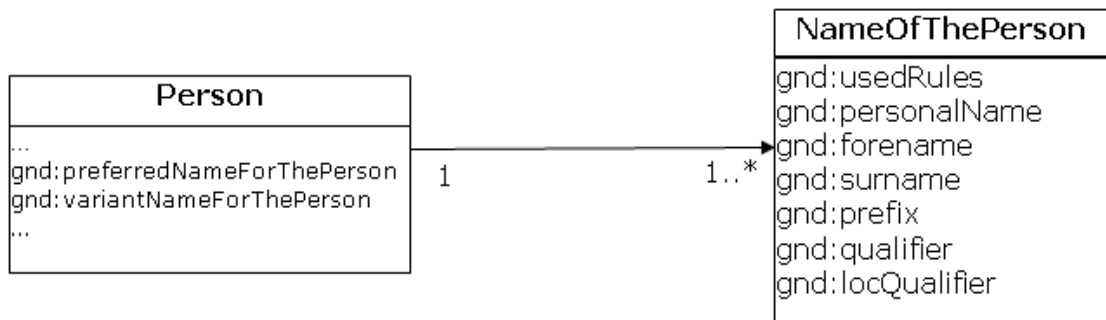


Abbildung 13: DNB - Namensentitäten im Linked-Data-Service-Prototyp [Hannemann(2010)]

Wie bei der VIAF werden Verknüpfungen zum Wikipedia-Artikel und zur DBpedia-Ressource gemacht, ebenso natürlich die Verknüpfung zum VIAF-Datensatz. Mit der Darstellung von familiären Beziehungen durch das Relationship Vocabulary und die Einbeziehung von PND-spezifischen Daten, z.B. durch das Element `gnd:publicationOfPerson` und die granulare Darstellung von Namensentitäten einer Person, geht die DNB-Realisierung von Personennormdaten als Linked Data ein Stück über die VIAF hinaus.

Letztlich zeigt der direkte Vergleich der RDF/XML-Daten auch, dass je nach Schwerpunktsetzung relativ unterschiedliche Repräsentationen der gleichen Ressource möglich sind. Dies unterstreicht jedoch nur den Grundansatz des Semantic Webs:

A different motivation for modeling in the Semantic Web is to form an integrated picture of some sort of domain by *federating information from multiple sources*. [...] The model provides a recipe for adding new information to incomplete information so it can be federated with other sources. [Allemang/Hendler(2008):296, Hervorhebung K.W.]

Abschließend sei erwähnt, dass die Library of Congress für dieses Jahr eine Veröffentlichung ihrer Name Authority Files (LCNAF), also Personen- und Körperschaftsnormdaten, angekündigt hat.

5.4.2 Kontrollierte Vokabulare (KOS) im Semantic Web

Das Simple Knowledge Organization System (SKOS) stellt die unangefochtene Methode zur Darstellung von Kontrollierten Vokabularen (KOS) für das Semantic Web dar. Es ist folglich besonders für Gedächtnisinstitutionen eine zentrale Aufgabe, die von ihnen kuratierten Begriffssysteme als RDF/SKOS-Spezifikationen auszuweisen und so zur maschinenverarbeitbaren Nachnutzung im WWW bereitzustellen. Trotz einer recht einfachen Struktur bietet SKOS weitreichende Möglichkeiten, die in der W3C-Referenz folgendermaßen umschrieben werden:

...making the accumulated experience and wisdom of knowledge organization in the library and information sciences accessible, applicable and transferable to the technological context of the Semantic Web, in a way that is complementary to existing Semantic Web technology (and in particular formal systems of knowledge representation such as OWL),... [SKOSref(2009):1.1]

Mittlerweile ist SKOS in der bibliothekarischen Welt als zukünftiger Standard anerkannt und es liegen bereits einige Kontrollierte Vokabulare in dieser WWW-tauglichen Form vor. Nachfolgend sollen zwei Schlagwortsysteme und eine Klassifikation beispielhaft vorgestellt werden.¹⁹³

Ein international wichtiges Schlagwortsystem neben der SWD, die Library of Congress Subject Headings (LCSH), werden von der LoC unter <http://id.loc.gov/authorities/search/> in RDF/XML oder Notation3 zur Recherche und als Download angeboten. Die Umwandlung der im MARC21-Authority-Format kodierten LCSH in SKOS haben Ed Summers, Antoine Isaac, Clay Redding und Dan Krech vorgestellt.¹⁹⁴ [Summers et al(2008)] Ed Summers hat die LCSH in SKOS dann zunächst unter der privaten Domain <http://lsh.info/> veröffentlicht. Im Mai 2009 hat die LoC schließlich ihren zentralen Service ID.LOC.GOV gestartet und die Testinstanz ersetzt. Die LoC plant, dort weitere Vokabulare wie MARC-Datenwertstandards (Relator, Geographic Areas, Language Codes) und den Thesaurus of Graphic Materials (TGM) zu publizieren.

Die analytische Abbildung der MARC-Normdaten-Felder auf das SKOS-Vokabular wird von Summers et al. als unkompliziert beschrieben und durch die Library of Congress Control Number (LCCN) jedes Normdatensatzes ist die Generierung eines persistenten HTTP-URIs für die beschriebenen Konzepte einfach.

Neben der umfassenden Nutzung des SKOS-Vokabulars beziehen Summers et al. die Dublin-Core-Konzepte `dcterms:created`, `dcterms:source`, `dcterms:modified` und `dcterms:lcc` ein, letztere property weist das Objekt als eine Notation des in den DCMI Metadata Terms registrierten Encoding Schemes Library of Congress Classification (LCC) aus. [Summers et al(2008):26-28] Als Beispiel soll hier der LCSH-Datensatz für das Sachschlagwort Madeira wine in der RDF/XML-Serialisierung aufgeführt werden:

```
<rdf:RDF>
<rdf:Description rdf:about="http://id.loc.gov/authorities/sh85146975#concept">
<skos:prefLabel xml:lang="en">Wine and wine making</skos:prefLabel>
</rdf:Description>
<rdf:Description rdf:about="http://id.loc.gov/authorities/sh85050983#concept">
<skos:prefLabel xml:lang="en">Fortified wines</skos:prefLabel>
</rdf:Description>
<rdf:Description rdf:about="http://id.loc.gov/authorities/sh85079519#concept">
<skos:inScheme rdf:resource="http://id.loc.gov/authorities#topicalTerms"/>
```

¹⁹³ Weitere Beispiele finden sich etwa im Forschungsbereich Terminology Services von OCLC, siehe <http://tspi.lot.oclc.org/resources/index.html>. Ebenfalls erwähnenswert sind der in 22 Sprachen vorliegende Thesaurus der Europäischen Union, Eurovoc, siehe http://idi.fundacionctic.org/classifications_endpoint/eurovoc und der an der Deutschen Zentralbibliothek für Wirtschaftswissenschaften gepflegte Standard-Thesaurus Wirtschaft, http://zbw.eu/stw/versions/latest/tech_background/about.de.html

¹⁹⁴ Vorarbeit hat hierfür bereits 2006 Corey A. Harper geleistet, indem er, ebenfalls mit Bezug auf die LCSH, ein XSLT-Stylesheet zur Umwandlung von Normdaten in MARCXML in RDF/SKOS entworfen hat, vgl. [Harper(2006)].

```
<skos:inScheme rdf:resource="http://id.loc.gov/authorities#conceptScheme"/>
<dcterms:created
rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">1986-02-
11T00:00:00-04:00</dcterms:created>
<skos:prefLabel xml:lang="en">Madeira wine</skos:prefLabel>
<owl:sameAs rdf:resource="info:lc/authorities/sh85079519"/>
<dcterms:modified
rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">1986-02-
11T00:00:00-04:00</dcterms:modified>
<skos:broader rdf:resource="http://id.loc.gov/authorities/sh85146975#con-
cept"/>
<skos:broader rdf:resource="http://id.loc.gov/authorities/sh85050983#con-
cept"/>
<rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
</rdf:Description>
</rdf:RDF>
```

Summers et al. machen allerdings auch auf Probleme bei der Umsetzung der LCSH in RDF/SKOS aufmerksam. So sind präkoordinierte Schlagwortketten sowie die Unterscheidung in verschiedene Schlagworttypen – in den LCSH werden geografische, Genre-/Form-, Zeit- und Sachschlagwörter unterschieden – in der SKOS-Konzeption nicht vorgesehen. Eine Verfeinerung von SKOS, die mehrgliedrige, z.B. facetiierte Konzepte darstellbar macht, wäre hier für viele Schlagwortsysteme ein Desiderat. Die Darstellung von Schlagworttypen ließe sich etwa anwendungsbezogen mit der Definition von Subklassen von `skos:Concept` realisieren. [Summers et al(2008):32] Auf die Verknüpfungsmöglichkeiten mit bereits vorhandenen Vokabularen wie DBpedia und GeoNames wird in dem Aufsatz bereits hingewiesen; jetzt existiert bereits ein teilweises, manuelles Mapping zu SKOS-Konzepten des Répertoire d'autorité-matière encyclopédique et alphabétique unifié (RAMEAU), dem französischen Schlagwortsystem, das mittlerweile ebenfalls als Linked Data vorliegt¹⁹⁵. Insgesamt werden die Möglichkeiten von Semantic-Web-Technologien zur Bereitstellung und automatischen Verarbeitung von KOS im WWW aber hervorgehoben:

The experiment highlighted the areas where SKOS and semantic web technologies excel: the identification and interlinking of resources; the reuse and mix-ability of vocabularies like SKOS and Dublin Core; the ability to extend existing vocabularies where generalized vocabularies are lacking. [Summers et al(2008):33]

Als eine zweite RDF(S)-Realisierung eines Schlagwortsystems soll hier ebenfalls die Testversion des Linked Data Service der DNB erwähnt werden. Diese umfasst eben nicht nur Personennormdaten, sondern bildet auch den Einstieg in die WWW-orientierte Darstellung der Schlagwortnormdatei (SWD) [Hannemann(2010)], die mittels den SKOS-properties `prefLabel`, `altLabel`, `definition`, `scopeNote`, `broader`, `related` modelliert. Außerdem finden Konzepte aus den DCMI Metadata Terms (`dcterms:identifier`, `dcterms:language`) und dem erwähnten eigenen Vokabular (`gnd:countryCode`)

¹⁹⁵ RAMEAU subject headings as SKOS linked data, siehe <http://www.cs.vu.nl/STITCH/rameau/index.html>

Verwendung, wie in den Beispieldaten des Sachschlagwortes Madeira <Getränk> ersichtlich wird:

```
<rdf:RDF>
<rdf:Description rdf:about="http://d-nb.info/gnd/7560734-7">
<skos:definition xml:lang="de">Likörwein von der gleichnam. Insel</skos:de-
finition>
<skos:altLabel xml:lang="de">Madeirawein</skos:altLabel>
<skos:broader xml:lang="de">Likörwein</skos:broader>
<dcterms:identifizier>(DE-588c) 7560734-7</dcterms:identifizier>
<skos:prefLabel xml:lang="de">Madeira <Getränk></skos:prefLabel>
<gnd:countryCode>XA-PT</gnd:countryCode>
</rdf:Description>
</rdf:RDF>
```

Nicht erfasst wurden bisher geografische, Personen-, Körperschafts-, Werk- sowie mehrgliedrige Sachschlagwörter. Des Weiteren sind, wie oben ersichtlich, Oberbegriffe bisher nur als literale Werte einbezogen, nicht als referenzierte Ressourcen; die Repräsentation der SW-D-Gruppen sowie der Klassifikationsnotationen der Dewey Decimal Classification (DDC) fehlen ebenso. Zur Darstellung von mehrgliedrigen Schlagwörtern wird in der Dokumentation aber bereits eine Lösungsstrategie vorgestellt: Man will sich dort an der SKOS-Formulierung des Thesaurus Sozialwissenschaften (TheSoz)¹⁹⁶ von Benjamin Zapilko und York Sure [Zapilko/Sure(2009)] orientieren.

Eine Einbeziehung der DDC-Notationen erscheint hier leicht machbar, da OCLC, Rechteinhaber an der Dewey Decimal Classification (DDC), zumindest die dreistelligen Notationen für die nicht-kommerzielle Nutzung und Weitergabe frei gegeben hat und diese ebenfalls in Semantic-Web-Formaten zur Verfügung stellt.¹⁹⁷ Unter <http://dewey.info/> stehen die Beschreibungen der Dewey-Klassen in zehn Sprachen, sowohl für die normale Browser-Ansicht in XHTML+RDFa, als auch rein in RDF/XML bereit, ebenso ist eine Abfrage in SPARQL möglich. Zu beachten ist aber, dass OCLC die Daten unter der Creative-Commons-Lizenz BY-NC-ND 3.0¹⁹⁸ verbreitet, d.h. eine Bearbeitung oder Veränderung nicht möglich ist. Die Lizenz ist in der RDF-Repräsentation eingebettet, wie in diesem Beispiel¹⁹⁹ zu erkennen ist:

```
<rdf:RDF>
<rdf:Description rdf:about="http://dewey.info/class/415/2009/03/about.de">
<rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
<xhv:license rdf:resource="http://creativecommons.org/licenses/by-nc-nd/3.0/" />
<cc:attributionName>OCLC Online Computer Library Center, Inc.</cc:attributionName>
<cc:attributionURL rdf:resource="http://www.oclc.org/dewey/" />
<dct:isVersionOf rdf:resource="http://dewey.info/class/415/" />
<dct:language rdf:datatype="http://purl.org/dc/terms/RFC4646">de</dct:language>
<skos:notation rdf:datatype="http://dewey.info/schema-terms/Notation">415</skos:notation>
<skos:inScheme rdf:resource="http://dewey.info/scheme/2009/03/about.de" />
```

196 TheSoz, siehe <http://www.gesis.org/dienstleistungen/tools-standards/thesaurus-sozialwissenschaften/>

197 Siehe hierzu: Dewey Summaries as Linked Data, <http://www.oclc.org/dewey/webservices/default.htm>

198 CC BY-NC-ND 3.0, siehe <http://creativecommons.org/licenses/by-nc-nd/3.0/>

199 RDF/XML-Darstellung der deutschsprachigen DDC-Klasse 415 Grammatik, <http://dewey.info/class/415/2009/03/about.rdf>

```
<skos:prefLabel xml:lang="de">Grammatik</skos:prefLabel>
<skos:broader rdf:resource="http://dewey.info/class/41/2009/03/about.de"/>
<cc:morePermissions
rdf:resource="http://www.oclc.org/dewey/about/licensing/" />
</rdf:Description>
</rdf:RDF>
```

Die durchdachte Syntax der URIs für die DDC-Klassen [Panzer(2008)] ermöglicht es, sämtliche Sprachversionen eines Konzeptes zusammen, und andererseits die einzelnen Versionen von Konzepten zu referenzieren. Dies ist von enormer Bedeutung für eine ständig in Weiterentwicklung begriffene Klassifikation, die zudem eine internationale Gültigkeit anstrebt.

Anhand von Beispielen wurde gezeigt, wie bibliografische Metadaten schemata und Datenwertstandards mittels Semantic-Web-Technologien in eine Darstellungsform überführt werden können, welche die enthaltenen Daten in allgemeingültiger Form im WWW zugänglich macht. Dabei erlangen die bibliografischen Daten und Normdaten Eigenschaften, die charakteristisch für semantische Informationen sind. Ihre komplexe Struktur wird auf einfache Ausgabengefüge heruntergebrochen und die verwendeten Konzepte durch Referenzierung eindeutig definiert. Im weitesten Sinne bedeutet dies, dass bibliothekarische Metadaten in eine vollkommen andere Datenstruktur einfließen, in der weder herkömmliche Datenbanken noch Datensätze eine Rolle spielen. Der Vorreiter dieser Struktur, die ein unendlicher Graph von verknüpften Daten aus höchst verschiedenen Kontexten sein könnte, wird heute als Linked Data Cloud bezeichnet.

Die sich daraus ergebende Frage lautet, wie dieses neue Metadaten-Paradigma die bibliothekarische Arbeit jenseits der einfachen Verfügbarmachung im Internet verändern wird. Denn soviel scheint klar: „Linked Bibliographic/Authority Data“ kann die Reichweite und inhaltliche Tiefe von Katalogen und Bibliografien zwar ungeheuer erweitern, aber keinen lokalen status quo ersetzen. [Coyle(2010a):29] Und um verlässliche und konsistente Beschreibungen zu erreichen, sind nach wie vor Erschließungsstandards nötig, vor dem Hintergrund von „befreiten Daten“ vielleicht mehr denn je.

5.5 Ein neuer Erschließungsstandard: Resource Description and Access (RDA)

Eine tiefgründige Behandlung des neuen Erschließungsstandards bzw. Rahmenwerks der Bibliografischen Kontrolle, das zudem noch am Anfang steht, kann in dieser Arbeit nicht geleistet werden. Der Vollständigkeit halber soll abschließend aber ein kleiner Ausblick auf dieses gegeben werden.

Mit der Funktionsanalyse von bibliografischen Daten und Normdaten, welche die FRBR-, FRAD- und FRSAD-Studien geleistet haben und vor allem mit ihrer Neuorientierung auf ein

Entitäten-Relationen-Modell, stehen nicht nur die herkömmlichen Datenstrukturen zur Disposition, sondern auch die traditionellen Katalogisierungsregelwerke. Aus diesem Grund und mit dem Ziel, einen international gültigen bibliothekarischen Erschließungsstandards für alle Medien und Materialien zu schaffen, wurde die anstehende Revision des angloamerikanischen Regelwerks AACR2 durch eine grundlegende Neukonzeption – das Regelwerk Resource Description and Access (RDA) – ersetzt. Geleitet wird die Erarbeitung der RDA vom Joint Steering Committee for Development of RDA²⁰⁰, der deutsch-österreichisch-schweizerische Standardisierungsausschuss²⁰¹ für das Bibliothekswesen ist daran aktiv beteiligt. Die Arbeitsstelle für Standardisierung der DNB hat mit den Expertengruppen des Standardisierungsausschusses die verschiedenen Teilentwürfe kommentiert, ebenso den letzten Gesamtentwurf [RDAdraft(2008)], der November 2008 veröffentlicht wurde. Das RDA Toolkit²⁰², das die Implementierung erleichtern soll, wurde am 23.06.2010 offiziell publiziert.

Die Zielsetzung von RDA wird im ersten Satz der Introduction des Gesamtentwurfs wie folgt zusammengefasst:

RDA provides a set of guidelines and instructions on formulating data to support resource discovery. [RDAdraft(2008):1]

Dabei sollen die RDA „all types of content and media“ abdecken. Die Daten, die gemäß den RDA formuliert werden sollen, werden zum einen als „data [...] to describe a resource“ bzw. descriptive data und zum anderen als „data [...] to describe an entity associated with a resource (a person, family, corporate body, concept, etc.)“ bzw. authority point control data [RDAdraft(2008):1-2 bzw. RDAscope(2009):2] benannt.

Hier sei an die sich durchsetzende Definition von bibliografischen Metadaten als „resource description“ erinnert. Ziel der Beschreibung der RDA sind aber nicht nur die Ressourcen im Sinne von Dokumenten und Objekten, sondern auch die (realen oder abstrakten) Entitäten, die mit diesen in Verbindung stehen. Letzteres entspricht, wenn auch bisher nicht so formuliert, eben der Aufgabe der Normenkontrolle von Eigennamen und Sachkonzepten. Dies ist insofern nicht weiter verwunderlich, da der Teilbereich Katalogisierung der Bibliografischen Kontrolle traditionell die anderen Teilbereiche Bibliografischer Beschreibung, Normenkontrolle und Kontrollierte Vokabulare zusammenführend reglementiert hat. Bemerkenswert ist jedoch, dass diese abstraktere Terminologie keinen Selbstzweck darstellt, sondern auf die Neustrukturierung auf der Basis der Konzeptmodelle FRBR und FRAD verweist. Der entscheidende Unterschied zu bisherigen Katalogisierungsstandards ist, dass RDA kaum mehr als solcher bezeichnet werden kann: Ziel von RDA ist weniger ein Katalog – also die struktu-

200 Joint Steering Committee for Development of RDA, siehe <http://www.rda-jsc.org/>

201 Standardisierungsausschuss, siehe http://www.d-nb.de/standardisierung/afs/afs_org.htm#standardisierungsausschuss

202 RDA Toolkit, siehe <http://www.rdatoolkit.org/>

rierte Anordnung von konsistenten, formal und inhaltlich deskriptiven Surrogaten – denn die Darstellung von bibliografischen Sachverhalten gemäß einem Entitäten-Relationen-Modell.

In der deutschen Übersetzung des RDA-Prospectus' wird dies folgendermaßen ausgedrückt:

Ein zweites Schlüsselement in der Gestaltung der RDA ist die Einführung einer klaren Trennung zwischen der Erfassung der Daten einerseits und der Präsentation der Daten andererseits. Der Schwerpunkt der RDA wird auf der Bereitstellung von Richtlinien und Regeln zur Datenerfassung für die Wiedergabe der Merkmale und Beziehungen liegen, die mit den in den FRBR und FRAD definierten Entitäten in Verbindung stehen. [RD Apro(2009):3]

Wird in den FRBR/FRAD-Modellen ein bibliografisches Netz von Entitäten, Relationen und Attribute der Entitäten gezeichnet, werden in der Überführung in den Erschließungsstandard RDA Relationen ebenfalls als Attribute der Entitäten gefasst. Dies ist notwendig, da die komplexe ER-Struktur in einzelne Beschreibungen aufgelöst werden muss, um sie linear darstellen zu können. So werden Attribut/Relation-Wert-Paare (property-value pairs) zur Beschreibung der Entitäten gebildet, wobei die properties als RDA-Elemente (elements) bezeichnet werden. Dabei können Elemente in einer generischen Hierarchie angeordnet werden (element sub-types) als auch partitiv untergliedert werden (sub-elements). [RDAScope(2009):2-3] In der weiteren Systematisierung der RDA-Elemente folgt das Scope-and-Structure-Papier der RDA zum einen den allgemeinen Attribut-Typen des <indec> Metadata Frameworks²⁰³ (label, quantity, quality, type, role) und legt zum anderen das DCMI Abstract Model (DCAM), genauer, das Beschreibungssatzmodell, zugrunde. Dort können Werte für die Attribute (hier: Elemente) vordefiniert werden, indem literale oder nicht-literale Platzhalter und gegebenenfalls semantische oder syntaktische Codierungsschemata festgelegt werden.²⁰⁴ Die so vordefinierten Elemente werden in den RDA zur Beschreibung der einzelnen FRBR/FRAD-Entitäten festgelegt, dabei existieren besondere Regelungen und Ausnahmen für spezifische Arten von Inhalten und Medien bzw. für spezifische Entitäten. [RDAScope(2009):4-6] Folgendes wird deutlich: RDA unterscheidet sich zu bisherigen Erschließungsstandards auch darin, dass es nicht nur Regeln und Instruktionen zur Gewinnung von Daten vorgibt, sondern ein explizites Modell der bibliografischen Entitäten, ihrer Attribute und Beziehungen beinhaltet. Darüber hinaus definiert RDA einige interne und externe Wertstandards in Form von Listen und Vokabularen. Auf der Grundlage der FRBR/FRAD-Modelle, mit der umfassenden Definition von Klassen und Typen von Attributen und der Einbeziehung von Datenwertstandards kann RDA auch als ein umfassendes bibliografisches ontologisches System verstanden werden.

203 Godfrey Rust; Mark Bide: The <indec> metadata framework : principles, model and data dictionary, siehe http://www.doi.org/topics/indec/indec_framework_2000.pdf

204 Siehe ausführlicher dazu [Hillmann et al(2010)].

So scheint denn auch einerseits die Ausrichtung auf das Resource Description Framework (RDF) recht offensichtlich, wenn man die Darstellung der Attribute und Relationen der Entitäten in einer einfachen Beschreibungsstruktur sowie die an das DCAM angelehnte Struktur der RDA-Elemente betrachtet. Allerdings bleiben die RDA-Papiere diesbezüglich recht unverbindlich und betonen dagegen die Neutralität gegenüber Umsetzungen in konkrete Metadatenschemata und sogar Kodierungen generell:

RDA does not specify a record syntax for the encoding or presentation of descriptive data or access point control data. Property/value statements formulated according to the guidelines and instructions in RDA are treated as discrete statements that can be stored or presented in a variety of record syntaxes. [RDAscope(2009):6, Hervorhebungen original]

Das Scope-and-Structure-Papier kündigt des Weiteren verschiedene unterstützte Darstellungsformen an:

Mappings of RDA elements to a select number of encoding and presentation syntaxes (e.g., MARC 21, ISBD) are provided in RDA appendices.

Planning is underway to develop an RDA element vocabulary and value vocabularies that would support the encoding of RDA data in RDF-compliant XML. [RDAscope(2009):6-7]

Dass hierbei besonders auf die Kompatibilität zu den „alten“ Standards abgehoben wird, ist aus Sicht der Übertragung der bestehenden Katalogdaten natürlich äußerst wichtig. Jedoch vermisst man in Anbetracht der eher beiläufigen Anmerkung zur RDF-Repräsentation eine konsequente Zukunftsvision für die praktischen Umsetzung der RDA.

Tatsächlich werden in den Appendices D und E des RDA-Gesamtentwurfs bisher nur ISBD-, AACR2- und MARC21-Abbildungen gegeben, selbst ein einfaches Dublin-Core-Mapping fehlt. Insofern verwundert es nicht, dass die DNB in ihrem Kommentar zum Full Draft eine nachdrücklichere Ausrichtung auf zukünftige WWW-Standards fordert. In den General comments on RDA as a whole mahnt sie einen expliziten Punkt Compliance to the Semantic Web an:

At present there are several attempts to create ontologies for bibliographic data but all of them come from outside the library community. At the conceptual level, RDA is a step in the right direction but without a connection to the Semantic Web it will be irrelevant outside the library world. A representation of the core RDA concepts in RDF is urgently needed. At least, the JSC should define official URIs for the key concepts of RDA. [DNBrda(2009):2]

Anscheinend hat es das Joint Steering Committee for Development of RDA (JSC) vorgezogen, sich diesem Problem zuerst praktisch zu nähern bzw. erst versuchsweise die RDA in RDF(S) zu definieren und mit der offiziellen Empfehlung des Standards abzuwarten. Auch in der nichtbibliothekarischen Metadaten- und Semantic-Web-Community erkannte man den dringenden Bedarf, RDA – als im Idealfall international verbindliche bibliografische Ontologie – in maschinenverarbeitbarer Form bereitzustellen. Aus diesem Grund wurde schon im April/Mai 2007 eine Kooperation zwischen dem JSC und der Dublin Core Metadata Initiative (DCMI) ins Leben gerufen, die eine öffentlich zugängliche RDF(S)-Repräsentation der

RDA erarbeiten sollte. Die DCMI/RDA Task Group²⁰⁵ hat unter Führung von Diane Hillmann und Gordon Dunsire schließlich RDA-Elemente und die in RDA benutzten Datenwertstandards in der National Science Digital Library (NSDL) Metadata Registry unter <http://metadatabrowser.org/rdbrowse.htm> veröffentlicht. [Hillmann et al(2010)] Die NSDL Metadata Registry soll eine zentrale nationale und internationale Plattform für die Entwicklung und den Nachweis von semantischen Vokabularen werden, ähnlich der DCMI Registry. Metadatenschemata, Datenwertstandards and Application Profiles sollen „menschenslesbar“ und maschinenverarbeitbar vorgehalten sowie automatisierte Mappings zwischen diesen vorgenommen werden; die Vokabulare werden unter der CC-BY-NC-SA-Lizenz²⁰⁶ veröffentlicht.

Die RDA-Repräsentation in RDF(S) gliedert sich in die *RDA Element Sets* zur Darstellung der RDA-Elemente (also vergleichbar mit Metadatenschemata) und die *RDA Vocabularies* zur Darstellung der in RDA benutzten Datenwertstandards (value vocabularies) in RDF(S) und SKOS. [Coyle(2010b):30] Grundsätzlich werden FRBR/RDA-Entitäten als RDFS-Klassen und RDA-Elemente (also FRBR/FRAD-Attribute und -Relationen) als RDF-properties definiert, wie es im Scope-and-Structure-Papier bereits theoretisch angekündigt wurde.

Die RDA Element Sets umfassen dabei die FRBR Entities for RDA²⁰⁷, also die vorläufige FRBR/FRAD-Entitäten-Repräsentation der DCMI/RDA Task Group, die durch die offizielle Registrierung der FRBR Review Group noch abgelöst werden soll. Weitere RDA Element Sets bilden die RDA-Elemente, die jeweils den drei FRBR-Gruppen als Attribute zugeordnet sind, die RDA-Elemente, die sich aus den Relationen zwischen den Entitäten ergeben sowie die RDA Roles, die spezifische properties und sub-properties für Gruppe-2-Entitäten bereitstellen. Da laut RDA einige Attribute auf verschiedene Entitäten angewendet werden können, müssen im Sinne der eindeutigen Referenzierung von Konzepten (hier: properties) jeweils einzelne auf die Entitäten bezogene RDA-Elemente geschaffen werden sowie eine super-property zu diesen, die das allgemeine Konzept repräsentiert. So existiert etwa die property `titleProper` als allgemeines Konzept sowie die sub-property `titleProperManifestation` als konkrete property zur Beschreibung einer Ressource, die der Klasse `Manifestation` angehört. [Coyle(2010b):27-29]

Die Definition von RDA in RDF(S) der DCMI/RDA Task Group wird voraussichtlich eine bedeutende Rolle in der internationalen Praxis der Bibliografischen Kontrolle spielen. Auch aus diesem Grund beteiligt sich die Arbeitsstelle für Standardisierung der DNB bereits durch die Übersetzung der RDA Vocabularies RDA Media Type, RDA Content Type und RDA

205 DCMI/RDA Task Group, siehe <http://dublincore.org/dcmirdataskgroup/>

206 Creative Commons BY-NC-SA, siehe <http://creativecommons.org/licenses/by-nc-sa/3.0/>

207 Diese wurden bereits bei der Vorstellung des Linked Data Services der DNB oben erwähnt.

Carrier Type²⁰⁸. Der entscheidende Vorteil einer RDF-Kodierung der Metadaten- und Datenwertstandards, die RDA implizit vorgibt, besteht eben darin, dass es die Metadaten- und Vokabularstrukturen in Form von einzelnen referenzierten Aussagen verfügbar macht. Auch wenn also spezifische RDA Element Sets und RDA Vocabularies als Ganzes formuliert werden, sind die dortigen Konzepte aber immer auch separat einsetzbar, flexibel kombinierbar und erweiterbar. Dies bedeutet, dass Metadatensets umfassend skalierbar und anpassbar werden [Coyle(2010b):31-32] und die Definition von Application Profiles, d.h. von anwendungsbezogenen Metadatensets, in freier Form möglich wird:

In an RDF-compliant application profile, elements and vocabularies can be taken from any suitable defined set, and many Semantic Web applications work with a mix of elements from numerous sources. [...] The key, however, is that all of these customized profiles would be compatible with each other because they all would make use of the same defined and registered metadata properties. [Coyle(2010b):34-35]

5.6 Fazit: Bibliografische Kontrolle mit dem Semantic Web?

RDF(S) und passende, leichtgewichtige Ontologien könnten einen effektiven Weg bieten, die strukturelle und inhaltliche Komplexität von bibliografischen Metadaten einheitlich darzustellen. Die äußerst einfache Grundstruktur garantiert nicht nur die Interoperabilität der einzelnen Komponenten des Bibliografischen Systems; dieses wäre zudem hochgradig anpassungsfähig und durch allgemein referenzierte Konzepte universell verarbeitbar bzw. „verständlich“. Diese Vision eines einheitlichen und zugleich WWW-tauglichen Metadatenmodells hat Stefan Gradmann bereits 2005 aufgezeigt:

Thus [...] such an integrated approach would also create grounds for an integrated, WWW-transparent global model for librarian metadata successfully transgressing the divide which separates formal and subject metadata. [Gradmann(2005):72]

In der allgemeinen Nutzbarmachung von bibliografischen Daten für die Internetöffentlichkeit liegt ein zentrales Argument für „semantischen Metadaten“. Charles McCathieNevile und Eva Méndez sehen hierin aber nicht nur eine bedeutende Forderung an die Bibliotheken, sondern gleichzeitig die Chance, bibliothekarische Kernkompetenzen für das Semantic Web bzw. für die Erschließung des Webs auszuspielen:

Likewise, librarians need the Semantic Web to make their metadata interoperate [...] in the entirety of the Web, redefining the traditional strength and skills of library information organization for the Semantic Web and/or Universal Access era. [McCathieNevile/Méndez(2007):30]

Wenn so aufbereitete Bibliotheksdaten als Linked Data im Internet publiziert werden, werden neue Arten ihres Gebrauchs und ihrer Interaktion möglich. Der entscheidende Vorteil gegenüber den heute verfügbaren Daten im Netz besteht darin, dass sie zunächst formal und

²⁰⁸Übersetzungen von RDA Vocabularies in der NSDL Metadata Registry, siehe <http://metadataregistry.org/user/show/id/103.html>

maschinenverarbeitbar vorliegen. Auf die Chancen, die solche allgemein verarbeitbaren Metadaten generell bieten, wies Gradmann ebenfalls hin:

Inference-based functional models could then be built on this technical basis, generating completely new services for metadata retrieval and also simplifying and automating much of routine cataloguing and indexing work. [Gradmann(2005):72]

Darauf aufbauend sind im WWW verschiedenste Nutzungen und Weiterverarbeitungen möglich, intelligente Routinen zur Aufbereitung für bibliothekarische Laien genauso wie Verschmelzungen mit anderen Datenbeständen. Dadurch würden auch die bibliografischen Daten selbst auf eine breitere informationelle Basis gestellt:

[T]he Semantic Web may well be the library's best ally in its effort to redefine itself in the new Web environment. [...] The network of protocols and standards emerging from the World Wide Web Consortium could dramatically widen the library community's access to other experts in different disciplines, making their specialized knowledge electronically available to supplement and enhance library records. [Campbell(2007):19]

Beim ambitionierten Projekt einer virtuellen Sammlung von (freien) kulturellen Gütern Europas, Europeana, hat man anscheinend entschieden, diesen Weg weiter zu verfolgen. So wird im ersten Europeana White Paper die Einbeziehung des Linked-Data-Paradigmas eindeutig benannt:

Europeana should thus be seen as a big aggregation of digital representations of cultural artefacts together with rich contextualisation data and embedded in a Linked Open Data architecture... [Gradmann(2010):9]

Darüber hinaus wird klar gestellt, dass das Ziel der bibliografischen Daten der Europeana eben nicht nur ein Katalog sein soll, sondern dass sie Teil des World Wide Web werden sollen, auch wenn dies ein neues „Eigenleben“ der Daten und damit einen Kontrollverlust bedeutet:

Our aim is to integrate the data layer providing semantic context for Europeana object representations as seamlessly in the LOD architecture as possible. This implies giving up some autonomy: the very idea of 'control' becomes obsolete to some extent that way – but the gain in functionality and rich context will be considerable and – above all – this step makes Europeana part of a much larger community and in a way simply an integrated part of the WWW, the biggest interoperability framework the world has ever seen. [Gradmann(2010):12]

In diesem Kapitel sollten jedoch nicht nur die generellen Möglichkeiten einer Bibliografischer Kontrolle auf der Basis von semantischen Technologien angedeutet werden, sondern auch aktuelle Ansätze zur konkreten Umsetzung dieser Idee vorgestellt werden. Es existieren dazu bereits einige engagierte Unternehmungen aus verschiedenen professionellen Hintergründen heraus, wie der Metadaten-Community, der Semantic-Web-Community und vermehrt auch der Bibliotheken. Während die Dublin Core Metadata Initiative ihr Metadaten-Universum mit dem Dublin Core Abstract Models (DCAM) an das RDF(S)-Paradigma angepasst hat, differenzieren sich Linked-Data-Angebote und Ontologien bzw. Vokabulare weiter aus. Einige dieser Entwicklungen wurden direkt von bibliothekarischen Initiativen über-

nommen oder werden mitgenutzt, etwa die essentiellen Vokabulare zur Beschreibung von Personen und von Begriffssystemen, FOAF und SKOS, sowie die DBpedia als zentrale Referenz für nicht direkt adressierbare Konzepte.

Die systematische RDF(S)-Umsetzung von Metadatenystemen aus einer genuin bibliothekarischen Perspektive steckt dagegen noch in den Kinderschuhen. Die Forderung nach elaborierten Ontologien zur Darstellung von bibliografischen Daten und Normdaten stellt hierbei den entscheidenden Punkt dar. Dass die außerhalb der bibliothekarischen Community entwickelten hybriden Vokabulare MarcOnt und Bibliographic Ontology nur als Zusatz fungieren können, ist relativ klar. Die unabhängig von der Semantic-Web-Diskussion entwickelten neuen Konzeptmodelle beruhen auf einer Funktions- und Entitätenanalyse des bibliografischen Feldes und bieten sich daher für die Strukturierung einer bibliografischen Ontologie an. Der neue bibliothekarische Erschließungsstandard Resource Description and Access (RDA) basiert eben auf den FRBR/FRAD-Entwürfen und legt der zukünftigen Katalogisierung somit ein explizites Modell von bibliografischen Entitäten, ihren Attribute und Beziehungen zugrunde.

Mit der Formulierung der RDA Element Sets und RDA Vocabularies in RDF(S) wird nicht nur dieses Entitäten-Relationen-Modell des Bibliografischen Universums als ontologisches System anwendbar, sondern auch Metadaten schemata und Datenwertstandards in einer gemeinsamen, maschinenverarbeitbaren Datenstruktur zusammenführt. Alle Metadatenelemente und Datenwerte, die im Katalogisierungsregelwerk RDA vorgesehen sind, werden in einer einheitlichen Syntax einzeln adressierbar und kombinierbar bereitgestellt. Dies birgt das Potential von individuellen Metadatensets bzw. Application Profiles in Form von kumulierten Beschreibungsaussagen, wie es etwa das DCAM entwirft. Eine konsequente RDF(S)-basierte Bibliografische Kontrolle würde somit auch starr definierte bibliografische Datensätze und Normdatensätze generell in Frage stellen. Karen Coyle schreibt:

Where RDF differs most from the present system of bibliographic records is in allowing bibliographic descriptions to interact, extend, and influence each other and to interact at a statement level with other data from library and nonlibrary sources. [Coyle(2010b):35-36]

Diesen Möglichkeiten trägt der RDA-Entwurf nicht ausreichend Rechnung. Auch wenn seine Formatneutralität hervorgehoben wird, ist davon auszugehen, dass zur Umsetzung der RDA das RDF-Datenmodell aufgrund seiner Universalität zukünftig bevorzugt wird. Das Zusammenspiel von semantischen Methoden und RDA könnte folgendermaßen schematisch dargestellt werden:

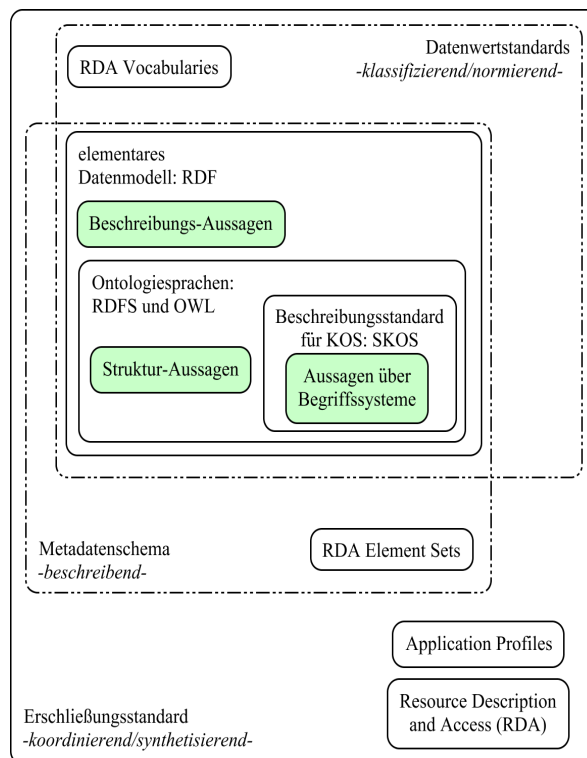


Abbildung 14: Bibliografische Kontrolle im Semantic Web

Ausgehend von der Skizze von Bibliografischen Metadatenystemen in Kapitel 3.3 lässt sich hier erkennen, dass die Bereiche von Metadaten schemata und Datenwertstandards in ihrer bisherigen Abgrenzung verblassen. Beide Funktionen werden durch Beschreibungsaussagen auf der Basis von RDF(S) bzw. Ontologien und SKOS realisiert; alle Konzepte, ob Metadaten elemente oder Datenwerte, sind referenziert und damit maschinenverarbeitbar. Letztlich verliert mit dieser syntaktischen Gleichheit die generelle Unterscheidung zwischen diesen Teilbereichen der Bibliografischen Kontrolle an Bedeutung. Mit der freien Definition von Application Profiles, also der Festlegung von zu verwendenden Metadaten schema und Datenwertstandards bzw. einzelnen Elementen und Werten, sind ganz unterschiedliche „Bibliografische Beschreibungen“ möglich. Dies ist besonders interessant vor dem Hintergrund, da nun alle bibliografische Entitäten beschrieben werden können und sollen. Spätestens hier wird das ursprüngliche Verständnis von Metadaten schemata und Datenwertstandards obsolet, wenn z.B. ein Sachkonzept im Zentrum der bibliografischen Arbeit steht.

Metadaten systeme dekonstruierten bereits die Titelaufnahmen der klassischen Bibliografischen Kontrolle, machten einzelne Datenelemente ansprechbar und brachten eine systematischere Sichtweise auf das Bibliografische Universum. Durch eine zweite Dekonstruktion der komplexen und oft inkompatiblen Metadatenstandards und ihre Rückführung auf ein gemein-

sames, universelles Aussagenmodell könnte eine weitere Entwicklungsstufe genommen werden, hin zu strukturell einfacheren, aber ungleich wirkmächtigeren bibliografischen Metadaten.

6 Schlussbemerkung

In dieser Arbeit wurde in einem ersten Komplex versucht, die grundsätzlichen Entwicklungslinien der Bibliografischen Kontrolle nachzuvollziehen, sowie deren aktuelle Herausforderungen und mögliche Lösungsansätze darzustellen.

In einer historisch-systematischen Erläuterung wurde das System der „klassischen“ Bibliografischen Kontrolle vorgestellt, in dessen Zentrum die Bibliografische Beschreibung bzw. das zu beschreibende Objekt stand. Alle Resultate der bibliografischen Teilbereiche wurden in einer Titelaufnahme vereinigt. Eine gesonderte, eingehende Betrachtung der dazu eingesetzten, z.T. genormten Begriffe setzte sich nur langsam durch. Mit dem Siegeszug der computergestützten Katalogisierung und dem weltweiten intellektuellen und bibliothekarischen Austausch wurde eine strukturiertere Abbildung der bibliografischen Entitäten und Beziehungen, kurz des Bibliografischen Universums, nötig.

Technisch wurde diese Forderungen durch Bibliografische Metadatensysteme zum Teil gelöst. Mit der Dekonstruktion von Titelaufnahmen wurden die einzelnen bibliografischen Elemente, nun Metadatenelemente und Datenwerte, ansprechbar und strukturiert darstellbar. Allerdings gingen damit nicht nur bisher aufgezeichnete, wertvolle bibliografische Beziehungen verloren, sondern wurde die benötigte Interoperabilität der Metadatenstandards durch die Vielzahl an Formaten und Kodierungen zu einem Problem. Überdies sind Katalogdaten nun zumeist in einer Datenbank „gefangen“, die nur über eine spezifische Suchmaske erreichbar sind.

Unterdessen näherte sich die bibliothekarische Community auf einer konzeptionellen Ebene der Zukunftsfrage der Bibliografischen Kontrolle an. Mit dem Entwurf von drei Entitäten-Relationen-Modellen versuchte man sich über Struktur und Funktionen des Bibliografischen Universums klar zu werden und damit die Grundlage für eine erneuerte bibliothekarische Erschließung zu schaffen. Mit der Einführung eines Mehr-Ebenen-Modells von Werk, Expression, Manifestation und Exemplar können Merkmale und Beziehungen nun nicht nur exakter zugeordnet werden, sondern auch Zusammenhänge, die Medienarten und Kunstformen überspringen, einbezogen werden. Allerdings bleiben einige Fragen des Verhältnisses von FRBR, FRAD und FRSAD zueinander genauso ungeklärt wie deren adäquate Umsetzung in der praktischen Katalogisierung.

Mit dem Entwurf eines neuen Erschließungsstandards mit intendierter internationaler Gültigkeit sollen die Vorstellungen der ersten beiden Konzeptmodelle implementiert werden. Dabei wird insbesondere die medienübergreifende Anwendungsmöglichkeit und die Trennung von Form und Inhalt der Katalogisierung hervorgehoben. Jedoch wird, wie auch z.T. in der neuen Erklärung zu den internationalen Katalogisierungsprinzipien, die Kompatibilität zu den bisherigen Standards der ISBD und den MARC21-Formaten festgeschrieben.

Beide Dokumente zielen darauf ab, die klassischerweise vernachlässigten Entitäten neben dem bibliografischen Objekt systematischer und umfangreicher zu beschreiben, also die Normenkontrolle i.e.S. in ähnlicher Weise zur Bibliografischen Beschreibung auszugestalten. Außen vor bleiben vorerst in beiden Fällen jedoch Entitäten der sachlichen Normenkontrolle, also Begriffe aus Kontrollierten Vokabularen.

Bibliografische Kontrolle sieht sich heute also mit zwei fundamentalen Fragestellungen konfrontiert. Auf der einen Seite fordern neue Konzeptmodelle und darauf aufbauende Regularien eine genauere und systematischere Abbildung des Bibliografischen Universums, insbesondere auch der adäquaten Einbeziehung von sämtlichen bedeutsamen Entitäten. Auf der anderen Seite brachten Metadatensysteme, die zu einer ersten klareren Strukturierung der Bibliografischen Kontrolle führten, nicht zu unterschätzende Nachteile mit sich, allen voran die mangelnde Sichtbarkeit und Allgemeinverständlichkeit bzw. Interoperabilität.

Als zweiter Komplex in dieser Arbeit wurden Theorien und Methoden, die unter dem Begriff Semantic Web zusammengefasst werden, umfassend erläutert. Die Einbeziehung von semantischen Informationen, und damit einer Möglichkeit der „Selbsterschließung“ des World Wide Webs, erscheint als ein vielversprechender Ausweg aus der Wachstums- und Erschließungsproblematik. Die maßgeblichen Techniken hierfür sind eine primitive „Semantik“ von Konzepten, d.h. ihre eindeutige Identifizierung durch WWW-Referenzierung, die Nutzung von formalen Sprachen und logischen Konstrukten, sowie eine allgemeines, einfaches Datenmodell in Form von elementaren Aussagen. Auch wenn die Vision des Semantic Webs weit über diese Grundlagen hinaus geht, werden mit der begonnenen Umsetzung in Form eines „Web of Data“ bereits WWW-interne Verknüpfungen von Konzepten bzw. Daten möglich. Dies stellt die genuine Leistung zur Erschließung des Internets dar.

Hier stellt sich die Frage, ob solche semantischen Informationen, die das bisherige Verständnis von Metadaten in Frage stellen, also auch für die Bibliografische Kontrolle von Interesse sein können. Schon durch den Desiderat, bibliothekarische Daten endlich „richtig“ in das Internet zu stellen, aber auch durch die erwähnten Probleme der Metadatensysteme, wurden diese Ideen dankbar von einzelne Bibliotheken und Informationswissenschaftlern aufgenommen. Unterstützung erhält die bibliothekarische Öffentlichkeit durch die Metadaten-Community, die ihre Standards an das RDF(S)-Paradigma anpasst und die Semantic-Web-Entwickler, die mit dem SKOS ein wichtiges Element für die Darstellbarkeit bibliografischer Daten geschaffen haben. In der öffentlichen Nutzbarmachung von Begriffssystemen liegt eine Dienstleistung, für die Gedächtnisorganisationen geradezu prädestiniert sind – diese reicht über Formulierung in SKOS hinaus bis zur Entwicklung von Ontologien für Wissensdomä-

nen. Zudem wird an der Entwicklung von spezifischen bibliografischen Ontologien gearbeitet, z.T. in Bezug auf allgemeinere Semantic-Web- und Metadaten-Vokabulare, aber auch in direktem Bezug auf die bibliothekarischen Konzeptmodelle FRBR und FRAD. Somit wird in allen wichtigen Teilbereichen von Bibliografischer Kontrolle bzw. von Metadaten-Systemen die Reformulierung durch das RDF(S)-Modells bereits erprobt. Dabei ist klar geworden, dass etwa das SKOS-Vokabular noch nicht vollends in der Lage ist, die Komplexität von Schlagwortsystemen und Thesauri wiederzugeben. Ebenso wurde festgestellt, dass personenbezogene Vokabulare zwar eine gute Ergänzung zu Personennormdaten darstellen, auf jeden Fall aber ein eigenes bibliothekarisches Vokabular nötig ist. Die Umwandlung von Körperschaftsnormdaten und die vollständige Transformation von OPAC-„Titelaufnahmen“ zu Beschreibungen von FRBR-Entitäten stellt eine weitere Herausforderung dar.

Insgesamt hat man den Vorteil erkannt, dass nicht nur bibliothekarische Begriffssysteme, sondern auch die Neukonzeptionierung des Bibliografischen Universums erstaunlich gut mit dem RDF(S)/Ontologien-Paradigma des Semantic Webs harmonieren. Die netzartige Struktur aus Entitäten mit Attributen und Beziehungen lässt sich leicht in die Aussagenform von Subjekt, Prädikat und Objekt überführen. Umgekehrt betrachtet stellen Semantic-Web-Technologien adäquate Mittel bereit, um das komplexe bibliografische Geflecht linear darstellbar und navigierbar zu machen – und dies in einer allgemeinverständlichen und WWW-tauglichen Form.

Mit der Definition von RDA-Struktur hat diese Entwicklung seinen vorläufigen Höhepunkt erreicht. RDA Element Sets, also Metadatenschemata und RDA Vocabularies, also Datenwertstandards, bieten nun eindeutig referenzierte bibliografische Konzepte im WWW an, die zur Erstellung von Application Profiles frei kombiniert werden können. Damit wurde ein adaptives und erweiterbares ontologisches System zur umfassenden Beschreibung von jeglichen Ressourcen geschaffen. Aufgrund der RDF(S)-Syntax ist dies kompatibel zur wachsenden Zahl von außerbibliothekarischen Schemata und Standards und kann durch die zukünftige W3C Recommendation HTML5 auch leicht in Websites eingebunden werden.

Mit der eindeutigen Ansprechbarkeit von einzelnen Konzepten und der Rückführung auf elementare Beschreibungsaussagen werden Metadatenschemata und Datenwertstandards aufgebrochen und verlieren ihre Bedeutung als geschlossene, verbindliche Blöcke. Bibliografische Metadaten auf der Basis von RDF(S) und adäquaten Ontologien wären in ihrer strukturellen und inhaltlichen Interoperabilität und Adaptivität, in ihrer Reichweite und Wirkmächtigkeit den heutigen Metadaten-Systemen weit überlegen. Gleichzeitig wäre ein solcher Richtungswechsel durchaus geeignet, die begonnene strukturelle Neuordnung des Bibliografischen Universums sinnvoll zu unterstützen. Allerdings bedarf es dazu auch den konsequenten Mut,

traditionelle Konzepte und Methoden der Bibliografischen Kontrolle zu überdenken. Der Gewinn, den Bibliotheken und andere Gedächtnisinstitutionen dadurch erwarten dürfen, wird die Mühen wahrscheinlich aufwiegen. Charles McCathie und Eva Méndez, dem Mann des W3C und der wissenschaftliche Bibliothekarin, soll das Schlusswort vorbehalten bleiben:

The universal library is still a utopia despite the Web. Notwithstanding, the Semantic Web and the technologies used in its construction let us dream again of universal access to knowledge, which in a more modern and technological way we would call something like: 'interoperable access to distributed digital knowledge'. [McCathieNevile/Méndez(2007):24]

7 Abbildungsverzeichnis

FRBR - Entitäten der Gruppe 1 und Primärbeziehungen [FRBRde(2009):13].....	19
FRBR - Entitäten der Gruppe 2 und "Verantwortlichkeits"-Beziehungen [FRBRde(2009):14].....	19
FRBR - Entitäten der Gruppe 3 und "Themen"-Beziehungen [FRBRde(2009):15].....	19
FRAD - Konzeptmodell für Normdaten [FRADdraft(2007):7].....	35
FRSAD - Konzeptmodell [FRSADdraft(2009):25].....	48
Klassische Bibliografische Kontrolle.....	50
Bibliografisches Metadatensystem.....	59
W3C - Semantic Web Stack (http://www.w3.org/2007/03/layerCake-small.png).....	78
DCMI - Beschreibungsmodell RDF [DCinRDF(2008)].....	79
DCAM - Ressourcenmodell [DCAMde(2008):5]	97
DCAM - Beschreibungssatzmodell [DCAMde(2008):9].....	99
DCAM - Vokabularmodell [DCAMde(2008):9].....	101
DNB - Namensentitäten im Linked-Data-Service-Prototyp [Hannemann(2010)].....	116
Bibliografische Kontrolle im Semantic Web.....	128

8 Literaturverzeichnis

- [Allemang/Hendler(2008)] *Semantic web for the working ontologist : modeling in RDF, RDFS and OWL* / Allemang, Jean; Hendler, James A. Burlington, MA : Morgan Kaufmann.
- [ANSI/NISO Z39.19(2005)] ANSI/NISO Z39.19 - guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda, MD : NISO. URL: <http://www.niso.org/kst/reports/standards/>.
- [Antoniou/van Harmelen(2008)] *A semantic web primer* / Antoniou, Grigoris; van Harmelen, Frank. 2nd edition. Cambridge, MA; London, England : MIT Press.
- [Auer/Lehmann/Bizer(2009)] Semantische Mashups auf Basis Vernetzter Daten / Auer, Sören; Lehmann, Jens; Bizer, Christian. In: *Social Semantic Web : Web 2.0 - was nun?* / Blumauer, Andreas; Pellegrini, Tassilo (Hg.). Berlin, Heidelberg : Springer. S. 259-286.
- [Baca-Glossary(2008)] Glossary. In: *Introduction to metadata* / Baca, Murtha (Hg.). Los Angeles, CA : Getty Publications. URL: http://www.getty.edu/research/conducting_research/standards/intrometadata/glossary.pdf.
- [Berners-Lee(1989)] Information management : a proposal / Berners-Lee, Tim. URL: <http://www.w3.org/History/1989/proposal.html>.
- [Berners-Lee(1997)] Metadata architecture (last edit date: 2000/09/21) / Berners-Lee, Tim. URL: <http://www.w3.org/DesignIssues/Metadata.html>.
- [Berners-Lee et al(1998)] Uniform Resource Identifiers (URI) : generic syntax / Berners-Lee, T.; Fielding, R.; Irvine, U.C.; Masinter, L. URL: <http://www.rfc-editor.org/rfc/rfc2396.txt>.
- [Berners-Lee/Fischetti(2000)] *Weaving the web : the past, present and future of the World Wide Web by its inventor* / Berners-Lee, Tim; Fischetti, Mark. London : Texere.
- [Berners-Lee/Hendler/Lassila(2001)] The semantic web / Berners-Lee, Tim; Hendler, James; Lassila, Ora. In: *Scientific American*. New York, NY : Scientific American. Vol. 284, No. 5. S. 34-43. URL (Kopie): <http://www.jeckle.de/files/tblSW.pdf>.
- [Bianchini/Guerrini(2009)] From bibliographic models to cataloging rules : remarks on FRBR, ICP, ISBD, and RDA and the relationships between them / Bianchini, Carlo; Guerrini, Mauro. In: *Cataloging & classification quarterly*. Philadelphia, PA : Taylor & Francis Group. Vol. 47, No. 2. S. 105-124.
- [Bizer/Heath/Berners-Lee(2009)] Linked Data - the story so far / Bizer, Christian; Heath, Tom; Berners-Lee, Tim. In: *International journal on semantic web and information systems*. Hershey, PA : IGI Global. Vol. 5, Iss. 3. S. 1-22. URL (Preprint): <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>.
- [Buizza(2004)] Bibliographic control an authority control from Paris Principles to the present / Buizza, Pino. In: *Cataloging & classification quarterly*. Binghamton, NY : Harworth Information Press. Vol. 38, No. 3/4. S. 117-133.
- [Bush(1945)] As we may think / Bush, Vannevar. In: *The Atlantic monthly*. Boston, MA : Atlantic Monthly. Vol. 176, No. 1. S. 101-108. URL: <http://www.theatlantic.com/past/docs/unbound/flas-hbks/computer/bushf.htm>.
- [Campbell(2007)] The birth of the new web : a Foucauldian reading of the semantic web / Campbell, D. Grant. In: *Cataloging & classification quarterly = Knitting the semantic web* / Greenberg, Jane; Méndez, Eva (Hg.). Binghamton, NY : Harworth Information Press. Vol. 43, No. 3/4. S. 9-20.
- [Chan/Zeng(2006)] Metadata interoperability and standardization : a study of methodology part I; achieving interoperability at the schema level / Chan, Lois Mai; Zeng, Marcia Lei. In: *D-Lib magazine*. Reston, VA : Corporation for National Research Initiatives. Vol. 12, No. 6. URL: doi:10.1045/june2006-chan.

- [Chowdhury/Chowdhury(2007)] *Organizing information : from the shelf to the web* / Chowdhury, G. G.; Chowdhury, Sudatta. London : Facet.
- [Cordeiro(2003)] From library authority control to network authoritative metadata sources / Cordeiro, Maria Ines. In: *Subject retrieval in a networked environment : proceedings on the IFLA satellite meeting held in Dublin, OH, 14-16 August 2001* / McIlwaine, I. C. (Hg.). München : Saur. S. 131-139. URL: <http://dlist.sir.arizona.edu/archive/00000118/01/authoritycontrol2001.doc>.
- [Coyle(2009)] FRASAD. In: *Coyle's InFormation* / Coyle, Karen. URL: <http://kcoyle.blogspot.com/2009/08/frasad.html>.
- [Coyle(2010a)] *Library technology reports = Understanding the semantic web : bibliographic data and metadata* / Coyle, Karen. Chicago, IL : American Library Association. Vol. 46, Iss. 1.
- [Coyle(2010b)] *Library technology reports = RDA vocabularies for a twenty-first-century data environment* / Coyle, Karen. Chicago, IL : American Library Association. Vol. 46, Iss. 2.
- [Dabrowski/Synak/Kruk(2009)] Bibliographic ontology / Dabrowski, Maciej; Synak, Macin; Kruk, Sebastian Ryszard. In: *Semantic digital libraries* / Kruk, Sebastian Ryszard; McDaniel, Bill (Hg.). Berlin, Heidelberg : Springer. S. 103-122.
- [Davis/Newman/D'Arcus(2009)] Expression of core FRBR concepts in RDF (first issued: 2005-07-15) / Davis, Ian; Newman, Richard, D'Arcus, Bruce. URL: <http://vocab.org/frbr/core.html>.
- [DCAMde(2008)] Übersetzung des Dublin Core Metadata Initiative Abstract Model (DCAM) / Baker, Thomas; Rühle, Stefanie. Göttingen : Kompetenzzentrum Interoperable Metadaten. URL: http://www.kim-forum.org/material/pdf/uebersetzung_dcam_20081111.pdf.
- [DCinRDF(2008)] Expressing Dublin Core metadata using the Resource Description Framework (RDF) / Nilsson, Mikael; Powell, Andy; Johnston, Pete; Ambjörn, Naeve. URL: <http://dublincore.org/documents/2008/01/14/dc-rdf/>.
- [DCMIspec(2010)] DCMI Specifications [Homepage] / Dublin Core Metadata Initiative. URL: <http://dublincore.org/specifications/>.
- [Dextre Clarke(2001)] Thesaural relationships / Dextre Clarke, Stella G. In: *Relationships in the organization of knowledge* / Bean, Carol A.; Green, Rebecca (Hg.). Dordrecht : Kluwer. S. 37-52.
- [DIN 1463-1(1987)] DIN 1463-1:1987-11 : Erstellung und Weiterentwicklung von Thesauri ; Einsprachige Thesauri / Deutsches Institut für Normung. Berlin : Beuth.
- [DNBfrad(2007)] Comments on Functional Requirements for Authority Data (FRAD) (2nd world-wide review) / Hengel, Christel; Henze, Gudrun; Pfeifer, Barbara. Leipzig, Frankfurt am Main, Berlin : Deutsche Nationalbibliothek. URL: http://www.d-nb.de/standardisierung/pdf/comments_frad_dnb_1.pdf.
- [DNBmd(2010)] Metadaten [Homepage] / Deutsche Nationalbibliothek. URL: <http://www.d-nb.de/standardisierung/metadaten/metadaten.htm>.
- [DNBrda(2009)] Comments on „RDA – Resource Description and Access“ : constituency review of November 2008 full draft / Deutsche Nationalbibliothek, Arbeitsstelle für Standardisierung. URL: http://www.d-nb.de/standardisierung/pdf/comments_rda_full_draft.pdf.
- [Dunsire(2009)] UNIMARC, RDA and the semantic web. In: *World Library and Information Congress : 75th IFLA General Conference and Assembly, 23-27 August 2009, Milan, Italy ; proceedings*. URL: www.ifla.org/files/hq/papers/ifla75/135-dunsire-en.pdf.
- [Exline(2008)] Metadata made simple / Exline, Eleta. URL: www.nhlibrarians.org/presentations/metadata.ppt.
- [Foulonneau/Riley(2008)] *Metadata for digital resources* / Foulonneau, Muriel; Riley, Jenn. Oxford, England : Chandos.
- [FRADdraft(2007)] Functional requirements for authority data : a conceptual model ; draft 2007-04-01 / IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). URL: <http://www.ifla.org/VII/d4/Franar-ConceptualModel-2ndReview.pdf>.

- [FRBR(1998)] Functional requirements for bibliographic records : final report / IFLA Study Group on the Functional Requirements for Bibliographic Records. München : Saur.
- [FRBRde(2009)] Funktionale Anforderungen an bibliografische Datensätze : Abschlussbericht der IFLA Study Group on the Functional Requirements for Bibliographic Records ; Stand: Februar 2009 / Deutsche Nationalbibliothek. URN: <urn:nbn:de:101-2009022600>.
- [FRSADdraft(2009)] Functional requirements for subject authority data (FRSAD) : a conceptual model ; 2nd draft 2009-06-10 / IFLA Working Group on Functional Requirements for Subject Authority Records (FRSAR). URL: <http://nkos.slis.kent.edu/FRSAR/report090623.pdf>.
- [GARR(2001)] Guidelines for authority records and references / IFLA Working Group on GARE Revision. München : Saur. URL: <http://www.ifla.org/VII/s13/garr/garr.pdf>.
- [Geyer-Hayden(2009)] Wissensmodellierung im Semantic Web / Geyer-Hayden, Barbara. In: *Social Semantic Web : Web 2.0 - Was nun?* / Blumauer, Andreas; Pellegrini, Tassilo (Hg.). Berlin, Heidelberg : Springer. S. 127-146. URI: <doi:10.1007/978-3-540-72216-8>.
- [Giasson/D'Arcus(2009)] Bibliographic Ontology Specification : specification document - 4 November 2009 / Giasson, Frédéric; D'Arcus, Bruce. URL: <http://bibliontology.com/specification>.
- [Gilliland(2008)] Setting the stage / Gilliland, Anna J. In: *Introduction to metadata* / Baca, Murtha (Hg.). Los Angeles, CA : Getty Publications. S. 1-19. URL: http://www.getty.edu/research/conducting_research/standards/intrometadata/setting.pdf.
- [Gómez-Pérez/Fernández-López/Corcho(2004)] *Ontological engineering : with examples from the areas of knowledge management, e-commerce and the semantic web* / Gómez-Pérez, Asunción; Fernández-López, Mariano; Corcho, Oscar. London; Berlin; Heidelberg : Springer.
- [Gonzalez(2008)] OCLC's work with the Dublin Core Metadata Element Set / Gonzalez, Linda M. In: *Journal of library metadata*. Binghamton, NY : Routledge. Vol. 8, No. 2. S. 139-145. URI: <doi:10.1080/10911360802087333>.
- [Gradmann(1998)] Cataloguing vs. metadata : old wine in new bottles? / Gradmann, Stefan. In: *Proceedings of the 64th IFLA General Conference, August 16 - August 21 1998, Amsterdam, Netherlands*. URL: <http://archive.ifla.org/IV/ifla64/007-126e.htm>.
- [Gradmann(2005)] rdfs:frbr : towards an implementation model for library catalogs using semantic web technology / Gradmann, Stefan. In: *Cataloging & classification quarterly = Functional Requirements for Bibliographic Records (FRBR) : hype or cure-all?* / Le Boeuf, Patrick (Hg.). Binghamton, NY : Haworth Information Press. Vol. 39, No. 3/4. S. 63-75.
- [Gradmann(2009)] Signal, Information, Zeichen : zu den Bedingungen des Verstehens in semantischen Netzen ; Antrittsvorlesung 28. Oktober 2008 / Gradmann, Stefan. URL: <http://nbn-resolving.de/urn:nbn:de:kobv:11-10096931>.
- [Gradmann(2010)] Knowledge = information in context : on the importance of semantic contextualisation in Europeana ; Europeana white paper 1 / Gradmann, Stefan. URL: <http://www.scribd.com/doc/32110457/Europeana-White-Paper-1>.
- [Green(2001)] Relationships in the organization of knowledge : an overview / Green, Rebecca. In: *Relationships in the organization of knowledge* / Bean, Carol A.; Green, Rebecca (Hg.). Dordrecht : Kluwer. S. 3-18.
- [Greenberg(2003)] Metadata and the World Wide Web / Greenberg, Jane. In: *Encyclopedia of library and information science* / Drake, Miriam A. (Hg.). New York, NY : Dekker. S. 1876-1888.
- [Greenberg/Méndez(2007)] Introduction : toward a more library-like web via semantic knitting / Greenberg, Jane; Méndez, Eva. In: *Cataloguing & classification quarterly = Knitting the semantic web* / Greenberg, Jane; Méndez, Eva (Hg.). Binghamton, NY : Haworth Information Press. Vol. 43, No. 3/4. S. 1-8.
- [Gruber(2009)] Ontology / Gruber, Tom. In: *Encyclopedia of database systems* / Liu, Ling; Özsu, M. Tamer (Hg.). New York, NY : Springer. URL (Preprint): <http://tomgruber.org/writing/ontology-definition-2007.htm>.

- [GSARE(1993)] Guidelines for subject authority and reference entries / Working Group on "Guidelines für Subject Authority Files" of the Section on Classification and Indexing of the IFLA Division of Bibliographic Control. München : Saur.
- [Hannemann(2010)] Dokumentation des Linked Data Service Prototyps der DNB / Hannemann, Jan. URL: <https://wiki.d-nb.de/display/LDS/Dokumentation+des+Linked+Data+Service+Prototyps+der+DNB>.
- [Harper(2006)] Encoding Library of Congress Subject Headings in SKOS : authority control for the semantic web / Harper, Corey A.. In: *Metadata for knowledge and learning : International Conference on Dublin Core and Metadata Applications, DC-2006 in Manzanillo, Colima, Mexico, 3-6 October 2006* / Dekkers, Makx; Fera, Lourdes (Hg.). Colima, Mexico : Universidad de Colima : Dublin Core Metadata Initiative. URL: <http://dcpapers.dublincore.org/ojs/pubs/article/view/842/838>.
- [Harper/Tillett(2007)] Library of Congress controlled vocabularies and their application to the semantic web / Harper, Corey A.; Tillett, Barbara B.. In: *Cataloging & classification quarterly = Knitting the semantic web* / Greenberg, Jane; Méndez, Eva (Hg.). Binghamton, NY : Harworth Information Press. Vol. 43, No. 3/4. S. 47-68.
- [Harris/Weibel(2004)] Metadata : a worldwide library effort / Harris, P.; Weibel, S. L. In: *ISO focus*. Genève : International Organization for Standardization. Vol. April. S. 31-32. URL: <http://www.oclc.org/research/staff/weibel/isofocus.pdf>.
- [Hausenblas(2009)] Anreicherung von Webinhalten mit Semantik : Microformats und RDFa / Hausenblas, Michael. In: *Social Semantic Web : Web 2.0 - Was nun?* / Blumauer, Andreas; Pellegrini, Tasilo (Hg.). Berlin, Heidelberg : Springer. S. 147-158. URI: <doi:10.1007/978-3-540-72216-8>.
- [Haynes(2004)] *Metadata for information management and retrieval* / Haynes, David. London : Facet.
- [Hellige(2007)] Library of the Future-Visionen in den 1950-60er Jahren : vom zentralen Wissenspeicher zum verteilten Wissensnetz / Hellige, Hans Dieter. In: *INFORMATIK 2007 : Informatik trifft Logistik ; Beiträge der 37. Jahrestagung der Gesellschaft für Informatik e.V. 24.-27. September 2007 in Bremen. Band 2* / Koschke, Rainer; Herzog, Otthein; Rödiger, Karl-Heinz; Ronthaler, Marc (Hg.). Bonn : Gesellschaft für Informatik. S. 491-496.
- [Hickey(2009)] The Virtual International Authority File : expanding the concept of universal bibliographic control / Hickey, Thomas. In: *NextSpace : the OCLC newsletter*. Dublin, OH : Online Computer Library Center. No. 13. S. 18-19. URL: <http://www.oclc.org/nextspace/013/research.htm>.
- [Hickey/O'Neill(2005)] FRBRizing OCLC's WorldCat / Hickey, Thomas B.; O'Neill, Edward T.. In: *Cataloging & classification quarterly = Functional Requirements for Bibliographic Records (FRBR) : hype or cure-all?* / Le Boeuf, Patrick (Hg.). Binghamton, NY : Haworth Information Press. Vol. 39, No. 3/4. S. 239-251.
- [Hillmann et al(2010)] RDA vocabularies : process, outcome, use / Hillmann, Diane; Coyle, Karen; Phipps, Jon; Dunsire, Gordon. In: *D-Lib Magazine*. Reston, VA : Corporation for National Research Initiatives. Vol. 16, No. 1/2. URL: <http://www.dlib.org/dlib/january10/hillmann/01hillmann.html>.
- [Howarth(2005)] Metadata and bibliographic control : soul-mates or two solitudes? / Howarth, Lynne C. In: *Cataloging & classification quarterly*. Binghamton, NY : Haworth Information Press. Vol. 40, No. 3/4. S. 37-56.
- [ICP(2009)] Statement of international cataloguing principles (ICP) / IFLA Cataloguing Section and IFLA Meetings of Experts on an International Cataloguing Code. URL: www.ifla.org/VII/s13/icc/imeicc-statement_of_principles-2008.pdf.
- [ICPde(2009)] Erklärung zu den internationalen Katalogisierungsprinzipien / Deutsche Nationalbibliothek, Arbeitsstelle für Standardisierung. URL: http://www.ifla.org/files/cataloguing/icp/icp_2009-de.pdf.
- [Intner/Lazinger/Weihs(2006)] *Metadata and its impact on libraries* / Intner, Sheila S.; Lazinger, Susan S.; Weihs, Jean. Westport, CT : Libraries Unlimited.

- [ISBD(2007)] International standard bibliographic description (ISBD) / ISBD Review Group. München : Saur. URL: http://www.ifla.org/files/cataloguing/isbd/isbd-cons_2007-en.pdf.
- [Kuhlen(2004)] *Informationsethik : Umgang mit Wissen und Information in elektronischen Räumen* / Kuhlen, Rainer. Konstanz : U. V. K. Verlagsgesellschaft.
- [Lagoze/Lynch/Daniel(1996)] The Warwick framework : A container architecture for aggregating sets of metadata = Cornell University computer science technical report ; TR96-1593 / Lagoze, Carl; Lynch, Clifford A.; Daniel, Ron Jr. URL: <http://hdl.handle.net/1813/7248>.
- [Lassila/McGuinness(2001)] The Role of Frame-Based Representation on the Semantic Web = KSL Tech Report ; KSL 01-02 / Lassila, Ora; McGuinness, Deborah L. URL: ftp://ftp.ksl.stanford.edu/pub/KSL_Reports/KSL-01-02.html.
- [Legg(2007)] Ontologies on the Semantic Web / Legg, Catherine. In: *Annual review of information science and technology*. San Francisco, CA : Wiley Interscience. Vol. 41, Iss. 1. S. 407-451.
- [Lindenthal(2005)] Grundlagen der inhaltlichen Erschließung von Texten und Objekten : Begriff, Bezeichnungen, und warum man Ordnung hält / Lindenthal, Jutta. URL: <http://www.jlindenthal.de/mb/Terminologie.pdf>.
- [Lindenthal/Scheven(2009)] ISO 25964 – Standard der Information und Dokumentation : „Thesauri and interoperability with other vocabularies“ ; [Präsentation, gehalten auf der ISKO-Tagung 2009] / Lindenthal, Jutta; Scheven, Esther. URL: http://isko.gesis.org/d-a-ch/bonn2009/Lindenthal_Scheven-Isko-bonn-2009-ISO25964.pdf.
- [Liu(2007)] *Metadata and its applications in the digital library : approaches and practices* / Liu, Jia. Westport, CT : Libraries Unlimited.
- [Malmsten(2008)] Making a library catalogue part of the Semantic Web / Malmsten, Martin. In: *Metadata for semantic and social applications : proceedings of the International Conference on Dublin Core and Metadata Applications, Berlin, 22 - 26 September 2008* / Greenberg, Jane; Klas, Wolfgang (Hg.). Göttingen : Universitätsverlag. S. 146-150. URL: <http://dcpapers.dublincore.org/ojs/pubs/article/view/927/923>.
- [Maxwell(2008)] *FRBR : a guide for the perplexed* / Maxwell, Robert L. Chicago : American Library Association.
- [McCathieNevile/Méndez(2007)] Library cards for the 21st century / McCathieNevile, Charles; Méndez, Eva. In: *Cataloging & classification quarterly = Knitting the semantic web* / Greenberg, Jane; Méndez, Eva (Hg.). Binghamton, NY : Harworth Information Press. Vol. 43, No. 3/4. S. 21-45.
- [Miksa(2008)] You need my metadata: demonstrating the value of library cataloging / Miksa, Shawne D. In: *Journal of library metadata*. Binghamton, NY : Routledge. Vol. 8, No. 1. S. 23-36. URI: <doi:10.1300/J517v08n01_03>.
- [Miles/Pérez-Agüera(2007)] SKOS : Simple Knowledge Organisation for the Web / Miles, Alistair; Pérez-Agüera, José R.. In: *Cataloging & classification quarterly = Knitting the semantic web* / Greenberg, Jane; Méndez, Eva (Hg.). Binghamton, NY : Haworth Information Press. Vol. 43, No. 3/4. S. 69-83.
- [MLARreport(1998)] Mandatory data elements for internationally shared resource authority records / UBCIM Working Group on Minimal Level Authority Records and ISADN. The Hague : International Federation of Library Associations and Institutions. URL: <http://archive.ifla.org/VI/3/p1996-2/mlar.htm>.
- [Nilsson et al(2008)] Expressing Dublin Core metadata using the Resource Description Framework (RDF) / Nilsson, Mikael; Powell, Andy; Johnston, Pete; Ambjörn, Naeve. URL: <http://dublincore.org/documents/dc-rdf/>.
- [NISO(2004)] Understanding metadata / National Information Standards Organization. Bethesda, MD : NISO. URL: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>.
- [On the record(2008)] On the record : report / The Library of Congress Working Group on the Future of Bibliographic Control. URL: <http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>.

- [Panyr(2006)] Thesauri, Semantische Netze, Frames, Topic Maps, Taxonomien, Ontologien - begriffliche Verwirrung oder konzeptionelle Vielfalt? / Panyr, Jiri. In: *Information und Sprache : Beiträge zu Informationswissenschaft, Computerlinguistik, Bibliothekswesen und verwandten Fächern ; Festschrift für Harald H. Zimmermann* / Harms, Ilse; Luckhardt, Heinz-Dirk; Giessen, Hans W. (Hg.). München : Saur. S. 139-151.
- [Panzer(2008)] Cool URIs for the DDC : towards webscale accessibility of a large classification system / Panzer, Michael. In: *Metadata for semantic and social applications : proceedings of the international conference on Dublin Core and Metadata Applications, Berlin, 22 - 26 September 2008, DC 2008: Berlin, German* / Greenberg, Jane; Klas, Wolfgang (Hg.). Göttingen : Universitätsverlag. S. 183-190. URL: <http://dcpapers.dublincore.org/ojs/pubs/article/viewFile/932/928>.
- [Payer(2009)] Grundlagen der Formalerschließung : Skript; Kapitel 2: Bibliographische Beschreibung / Payer, Margarete. URL: <http://www.payer.de/grundlagenfe/fegscr02.htm>.
- [PP(1961)] Statement of Principles / The International Conference on Cataloguing Principles. Paris. URL: http://www.d-nb.de/standardisierung/pdf/paris_principles_1961.pdf.
- [RDAdraft(2008)] RDA full draft / Joint Steering Committee for Development of RDA. URL: http://www.rdatoolkit.org/constituencyreviewfiles/constituency_review.zip.
- [RD Apro(2009)] RDA Prospectus / Deutsche Nationalbibliothek, Arbeitsstelle für Standardisierung. URL: http://www.d-nb.de/standardisierung/pdf/prospectus_dt_09.pdf.
- [RDAscope(2009)] Resource Description and Access (RDA) : scope and structure / Joint Steering Committee for Development of RDA. URL: <http://www.rda-jsc.org/docs/5rda-scoperev4.pdf>.
- [RDFconcepts(2004)] Resource Description Framework (RDF) : concepts and abstract syntax ; W3C Recommendation 10 February 2004 / Klyne, Graham; Carroll, Jeremy J. URL: <http://www.w3.org/TR/rdf-concepts/>.
- [Riley(2008)] Moving from a locally-developed data model to a standard conceptual model / Riley, Jenn. In: *Culture and identity in knowledge organization : proceedings of the Tenth International ISKO Conference, 5-8 August 2008, Montréal, Canada* / Arsenault, Clément (Hg.). Würzburg : Ergon. S. 124-130.
- [Riley(2009)] Thoughts on FRISAD. In: *Inquiring Librarian* / Riley, Jenn. URL: <http://inquiringlibrarian.blogspot.com/2009/07/thoughts-on-frisad.html>.
- [SKOSref(2009)] SKOS Simple Knowledge Organization System reference : W3C Recommendation 18 August 2009 / Miles, Alistair; Bechhofer, Sean. URL: <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>.
- [Smiraglia(2005)] Introducing metadata / Smiraglia, Richard P.. In: *Cataloging & classification quarterly*. Binghamton, NY : Haworth Information Press. Vol. 40, No. 3/4. S. 1-15.
- [Smith(2004)] Ontology / Smith, Barry. In: *Blackwell guide to the philosophy of computing and information* / Floridi, Luciano (Hg.). Oxford : Blackwell. S. 155-166. URI: <doi:10.1111/b.9780631229193.2003.00013.x>.
- [Soergel(2009)] Digital libraries and knowledge organization / Soergel, Dagobert. In: *Semantic digital libraries* / Kruk, Sebastian Ryszard; McDaniel, Bill (Hg.). Berlin, Heidelberg : Springer. S. 9-39.
- [Stock/Stock(2008)] *Wissensrepräsentation : Informationen auswerten und bereitstellen* / Stock, Wolfgang G.; Stock, Mechthild. München : Oldenbourg.
- [Strauch(2004)] *Grundlagen der praktischen Information und Dokumentation : begründet durch Klaus Laisiepen, Ernst Lutterbeck und Karl-Heinrich Meyer-Uhlenried. Band 2: Glossar : zusammengestellt und redigiert von Dietmar Strauch* / Kuhlen, Rainer; Seeger, Thomas; Strauch, Dietmar (Hg.). 5., völlig neu gefasste Ausgabe. München : Saur.
- [Styles/Ayers/Shabir(2008)] Semantic MARC, MARC21 and the semantic web / Styles, Rob; Ayers, Danny; Shabir, Nadeem. In: *Proceedings of the Linked Data on the Web Workshop, Beijing, China, April 22 2008* / Bizer, Christian; Heath, Tom; Idehen, Kingsley; Berners-Lee, Tim (Hg.). URL: <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-369/paper02.pdf>.

- [Summers et al(2008)] LCSH, SKOS and Linked Data / Summer, Ed; Isaac, Antoine; Redding, Clay; Krech, Dan. In: *Metadata for semantic and social applications : International Conference on Dublin Core and Metadata Applications ; 22 - 26 September 2008, Berlin* / Greenberg, Jane; Klas, Wolfgang (Hg.). Göttingen : Universitätsverlag. S. 25-33. URL: <http://edoc.hu-berlin.de/conferences/dc-2008/summers-ed-25/PDF/summers.pdf>.
- [Svenonius(2001)] *The intellectual foundation of information organization* / Svenonius, Elaine. Cambridge, MA : MIT Press.
- [Svensson(2007)] National libraries and the semantic web : requirements and applications / Svensson, Lars G. In: *Proceedings of the International Conference on Semantic Web & Digital Libraries : ICDS 2007, Bangalore, India, February 21-23 2007* / Prasad, ARD; Madalli, Devika P. (Hg.). Bangalore, India : Indian Statistical Institute. S. 101-108. URL: http://drtc.isibang.ac.in:8080/bitstream/1849/326/1/010_p36_svensson_formatted.pdf.
- [Synak/Dabrowski/Kruk(2009)] Semantic web and ontologies / Synak, Marcin; Dabrowski, Maciej; Kruk, Sebastian Ryszard. In: *Semantic digital libraries* / Kruk, Sebastian Ryszard; McDaniel, Bill (Hg.). Berlin, Heidelberg : Springer. S. 41-54.
- [Tillett(2001)] Bibliographic relationships / Tillett, Barbara B. In: *Relationships in the organization of knowledge* / Bean, Carol A.; Green, Rebecca (Hg.). Dordrecht : Kluwer. S. 19-35.
- [Tillett(2004)] Authority control : state of the art and new perspectives / Tillett, Barbara B.. In: *Cataloging & classification quarterly*. Binghamton, NY : Harworth Information Press. Vol. 38, No. 3/4. S. 23-41.
- [Warnke(2008)] Hypertext : oder die Befreiung des Geistes durch die Maschine ; Antrittsvorlesung an der Leuphana Universität Lüneburg, 5. November 2008 / Warnke, Martin. URL: <http://weblab.uni-lueneburg.de/kulturinformatik/warnke/antritt/Hypertext.html>.
- [Zapilko/Sure(2009)] Converting the TheSoz to SKOS / Zapilko, Benjamin; Sure, York. Bonn : GESIS Leibniz-Institut für Sozialwissenschaften. URL: http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2009/TechnicalReport_09_07.pdf.
- [Zeng/Zumer(2009)] Introducing FRISAD and mapping it with SKOS and other models : (based on the work of the FRISAR Working Group) / Zeng, Marcia Lei; Zumer, Maja. In: *Proceedings of the World Library and Information Congress : 75th IFLA General Conference and Assembly, 23-27 August 2009, Milan, Italy*. URL: www.ifla.org/files/hq/papers/ifla75/200-zeng-en.pdf.

Alle URLs wurden am 30.06.2010 letztmalig auf ihre Gültigkeit geprüft.