

Carsten Klee

Vokabulare für bibliographische Daten

Zwischen Dublin Core und bibliothekarischem Anspruch

Bibliotheken auf dem Weg ins Web der Daten

Bibliographische Daten werden mehr denn je gebraucht. Sie werden außerhalb der Bibliothekswelt sowohl in kleinen als auch in großen Kontexten¹ nachgenutzt und sind unverzichtbares Instrument für die Wissenschaft². Im Web der Daten³ werden qualitativ hochwertige bibliographische Daten dringend als normierendes Instrument benötigt. Ähnlich wie Titelaufnahmen Normdateien referenzieren, können beliebige Ressourcen bibliographische Daten referenzieren, wenn diese als Linked Data⁴ zur Verfügung stehen. Nun zeigen die aktuellen Entwicklungen, dass viele nationale wie ausländische Bibliotheken, ob nun einzeln oder über ihren Bibliotheksverbund, sich dieser Aufgabe annehmen und ihre Daten in RDF publizieren.

Die Auswahl von Vokabularen, die Bibliotheken zur Verfügung haben, um ihre Daten zu beschreiben, ist dabei nicht einfach zu überblicken, und viele Vokabulare entsprechen nicht unbedingt den Ansprüchen, die klassische Katalogisierung an ein Datenformat stellt. Dabei wird aber schnell übersehen, dass ganz andere Ansprüche für die bibliographischen Daten im Kontext des World Wide Web bestehen, als im Kontext eines Bibliothekskataloges.

Dieser Beitrag versucht aus Sicht von Bibliotheken, die ihre Daten als Linked Data publizieren möchten, einen Überblick über die derzeit verfügbaren Vokabulare zur Beschreibung von bibliographischen Daten zu geben und diese zu vergleichen, ohne jedoch zu sehr ins Detail zu gehen. Dabei wird ein besonderes Augenmerk auf den Paradigmenwechsel gelegt, der die Ablösung bibliothekarischer Austauschformate durch einfache, aber weit verbreitete und damit viel genutzte Vokabulare bedeutet.

1 Zum Beispiel in der persönlichen Literaturverwaltung (kleiner Kontext) oder in der Wikipedia als Quellenangabe für die Artikel (großer Kontext).

2 Siehe Schulze.

3 Zu dem Begriff „Web der Daten“ siehe Berners-Lee.

4 Der Begriff „Linked Data“ symbolisiert hier nur die Mindestanforderung. Das zusätzlich offene Daten die Möglichkeiten der Nachnutzung vervielfältigen, steht außer Frage.



Was ist ein Vokabular?

Der Ausdruck Vokabular ist eigentlich ungenau und eher ein Sammelbegriff für Sets von Termen⁵ unterschiedlicher Nutzungsbestimmung. Die englischen Begriffe ‚Ontology‘, ‚Vocabulary‘, ‚Element Set‘, ‚Encoding Scheme‘ etc. beschreiben besser die Bestimmung der jeweiligen Termsammlung. Ontologien (engl. Ontologies)⁶ sind einfach ausgedrückt Container für diese Konzepte zu einem bestimmten Gebiet. In ihnen gliedern sich die Konzepte in Element Sets, das sind Attribute oder auch Eigenschaften (engl. Properties) der Ressource, und Vocabularies, das sind die Werte, die bestimmte Eigenschaften annehmen können. Eine weitergehende Erläuterung und Unterscheidung der einzelnen Begriffe kann an dieser Stelle jedoch nicht geleistet werden. Zur besseren Verständlichkeit soll der Begriff ‚Vokabular‘ hier stellvertretend für alle diese Termsammlungen genutzt werden. Im Abschnitt *Aufbau von Vokabularen* wird jedoch noch einmal genauer darauf eingegangen, welche Ebenen Vokabulare haben können und wie diese zum Einsatz kommen.

Für das Verständnis von Vokabularen ist es zunächst ausreichend zu wissen, dass Vokabulare helfen, Dinge⁷ zu beschreiben, indem sie Konzepte definieren, um diese später mit den Ressourcen zu verknüpfen.

Im Web der Daten kann jeder sein eigenes Vokabular erfinden und benutzen. Dazu braucht man nichts weiter zu tun, als einem URI unter seiner eigenen Domain eine bestimmte Bedeutung zu geben. Damit kreiert man seinen eigenen Namensraum (engl. Namespace). Dass jeder sein eigenes Vokabular benutzen kann, ist gerade für Bibliotheken mit ihren wenigen aber alles bestimmenden Regelwerken und Normdateien sehr gewöhnungsbedürftig. Man muss allerdings verstehen, dass sich das Web mit vielfältigen Aspekten der gesamten Welt befasst, während Bibliotheken lediglich die in ihr existierenden Publikationen, also nur einen kleinen Ausschnitt, beschreiben. Möglich gemacht werden soll dies durch die sehr einfache Erweiterbarkeit des Webs der Daten durch jeden, der über eine Domain verfügt:

```
@prefix dbp: <http://dbpedia.org/resource/> .
@prefix myvocab: <http://mywebsite.de/vocab/> .
dbp:The_Name_of_the_Rose myvocab:12345 dbp:Umberto_Eco .
```

⁵ Wird eine Vokabel aus einem Vokabular mit einem URI identifiziert, spricht man allgemein von einem Term.

⁶ Gemeint sind hier Semantic Web Ontologien, die zumeist in OWL (Ontology Working Language) verfasst sind.

⁷ Wenn ein Ding durch einen Resource Identifier (URI) identifiziert wird, wird es Ressource genannt.

In dem obigen Beispiel wird kaum jemand anderes außer mir wissen, welche Bedeutung `http://mywebsite.de/vovab/12345` hat, denn unter dem URI findet man keine Webseite, die eine Beschreibung der Bedeutung geben könnte, die ich dem URI zugewiesen habe. Für Linked Data ist das jedoch völlig ausreichend (wenn auch nicht empfohlen). In diesem Fall wäre es vielleicht sinnvoll, wenn ich den Term so benenne, dass ein Dritter zumindest erraten kann, welche Bedeutung er hat:

```
dbp:The_Name_of_the_Rose myvocab:creator dbp:Umberto_Eco .
```

Noch besser wäre es, wenn ich unter dem URI auch ein Dokument hätte, das eine Beschreibung enthält, was dieser Term zu bedeuten hat. Wenn dann noch die Beschreibung des Terms in einer maschineninterpretierbaren Form vorläge, wäre mein Vokabular fast perfekt. Nur, mein Vokabular hat einen kleinen Schönheitsfehler. Den von mir erfundenen Term gibt es schon in einem anderen Vokabular, welches viel bekannter ist als meines und viel mehr genutzt wird. Darum habe ich zwei Optionen. Entweder ich nutze meinen Term weiterhin und erweitere die Beschreibung meines Terms durch eine Beziehung zu dem Term des bekannteren Vokabulars (Vocabulary Alignment):

```
@prefix myvocab: <http://mywebsite.de/vocab/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix dct: <http://purl.org/dc/terms/> .
myvocab:creator owl:equivalentProperty dct:creator .
```

... oder ich nutze gleich das andere, bekanntere Vokabular. Denn der Sinn der Vokabulare ist es ja, die Daten für andere nachnutzbar zu machen. Und das geschieht am besten dadurch, dass möglichst bekannte und viel verbreitete Vokabulare genutzt werden.

Aber egal, ob man sich für das bekanntere oder sein eigenes Vokabular entscheidet, für die Bewertung jedes Vokabulars gelte folgendes 5-Sterne-Schema⁸:

- * Veröffentliche dein Vokabular im Web unter einem stabilen URI und unter einer offenen Lizenz.
- ** Stelle eine von Menschen lesbare Dokumentation und minimale Metadaten, wie Autor, Herausgeber, Erstellungsdatum, Änderungsdatum, Versionsnummer zur Verfügung.

⁸ Siehe Vatant.

- *** Stelle Beschreibungen, wenn möglich in verschiedenen Sprachen, zur Verfügung, um dein Vokabular in mehrsprachigen Kontexten nutzbar zu machen.
- **** Mache dein Vokabular unter Verwendung von W3C-Standards (RDF, SPARQL) und Content Negotiation über sein Namespace-URI verfügbar, und zwar in maschinenlesbarer wie auch in einer für Menschen verständlichen Form.
- ***** Linke zu anderen Vokabularen, indem du Elemente nachnutzt, anstatt sie neu zu erfinden.

Aufbau von Vokabularen

Um Vokabulare nachnutzen zu können, muss man ihren Aufbau verstehen, der glücklicherweise zumeist sehr einfach gehalten ist. Aus Anwendersicht gibt es bei Vokabularen bis zu vier Ebenen, welche die im Abschnitt *Was ist ein Vokabular?* erwähnten Konzepte enthalten:

- Classes⁹
- Object Properties
- Data Properties
- Datatypes

Wenn ein Ding¹⁰ (engl. Thing) einer Klasse¹¹ (engl. Class) zugeordnet wird, dann wird das Ding zum Individuum (engl. Individual) und ist gleichzeitig eine Instanz (engl. Instance) dieser Klasse. Wenn mehrere Individuen einer Klasse zugeordnet werden, dann teilen sie sich ihre jeweiligen Klasseneigenschaften.¹²

```
@prefix dbp: <http://dbpedia.org/resource/> .
@prefix bibo: <http://purl.org/ontology/bibo/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
dbp:The_Name_of_the_Rose rdf:type bibo:Book .
```

⁹ Genauer gesagt gibt es noch Super- und Subclasses.

¹⁰ Ding oder Thing (owl:Thing) ist die Klasse aller Klassen. Alle Individuen (alles was ist) sind Mitglieder dieser Klasse.

¹¹ Klassen kann man auch im Allgemeinen daran erkennen, dass sie zumeist groß geschrieben werden.

¹² Dinge können zu Instanzen mehrerer Klassen werden. D.h. man kann sie mehreren Klassen zuordnen, wenn diese Klassen nicht disjunkt sind.

In diesem Beispiel wird das Ding beschrieben durch die Ressource `http://dbpedia.org/resource/The_Name_of_the_Rose`, der Klasse `http://purl.org/ontology/bibo/Book` zugeordnet und ist nun also eine Instanz dieser Klasse.

Eigenschaften, auch Attribute genannt, werden unterteilt in Object Properties, das sind Eigenschaften, die Individuen verbinden:

```
@prefix dbp: <http://dbpedia.org/resource/> .
@prefix dct: <http://purl.org/dc/terms/> .
dbp:The_Name_of_the_Rose dct:creator dbp:Umberto_Eco .
```

... und Data Properties, welche Individuen mit Literalen verbinden:

```
dbp:The_Name_of_the_Rose dct:title "Der Name der Rose"@de .
```

Zum Schluss gibt es noch Datatypes, welche Literale „typisieren“, also genauere Auskunft darüber geben, um welche Art von Literal¹³ es sich handelt.

```
@prefix dbp: <http://dbpedia.org/resource/> .
@prefix ex: <http://example.org/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
dbp:The_Name_of_the_Rose ex:temporal "1327"^^xsd:gYear .
```

Aus einer guten Dokumentation eines Vokabulars kann abgeleitet werden, wie eine Eigenschaft (engl. Property) verwendet werden darf. So gibt ‚`rdfs:domain`‘ (Definitionsbereich) Auskunft darüber, welche Klasse das Subjekt eines Tripels instantiiert, in dem die jeweilige Eigenschaft angewendet wird. ‚`rdfs:range`‘ (Wertebereich) sagt aus, welcher Klasse das Objekt eines entsprechenden Tripels angehören sollte:¹⁴

```
@prefix dct: <http://purl.org/dc/terms/> .
@prefix dbp: <http://dbpedia.org/resource/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix ex: <http://example.org/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
#Ein dct:creator muss immer eine Instanz der Klasse dct:Agent
#oder einer Subclass sein.
dct:creator rdfs:range dct:Agent .
```

¹³ Zum Begriff Literale siehe Carroll.

¹⁴ Definitionsbereich und Wertebereich lassen Inferenzregeln zu: Hat ein Individuum eine bestimmte Eigenschaft, kann abgeleitet werden, welcher Klasse es angehören muss.

```
#Ein dct:title muss immer ein Literal sein
dct:title rdfs:range rdfs:Literal .
```

Eigenschaften, die einen Datatype als Range haben, werden auch Datatype Properties genannt:

```
ex:temporal rdfs:range xsd:dateTime ;
#Alle Individuen, die die Eigenschaft ex:temporal haben
#sind Instanzen der Klasse dbp:Work oder einer Subclass
    rdfs:domain dbp:Work .
```

Ist kein `rdfs:domain` angegeben, gehört die Eigenschaft automatisch der Klasse `owl:Thing`.

Bei der Modellierung von RDF-Daten sollte nach Möglichkeit darauf geachtet werden, dass die Reglementierung der Verknüpfung von Individuen und Eigenschaften nicht missachtet wird. Zu Problemen würde es allerdings erst kommen, wenn so genannte Reasoner¹⁵ zum Einsatz kommen.

Vokabulare für bibliographische Daten

Nachdem nun die Grundlagen von Vokabularen erläutert wurden, werden nun die einzelnen Vokabulare untersucht, die für die Beschreibung von bibliographischen Daten nützlich sein könnten. Dies kann allerdings nur oberflächlich und in Auswahl geschehen. Dennoch sollte diese Liste als guter Ausgangspunkt dienen können, um den Kern von bibliographischen Daten abzudecken.

Dublin Core Element Set & Metadata Terms

Element Set

Namespace-URI: <http://purl.org/dc/elements/1.1/>

Spezifikation: <http://dublincore.org/documents/dces/>

¹⁵ Siehe Semantic Reasoner.

Metadata Terms

Namespace-URI: <http://purl.org/dc/terms/>

Spezifikation: <http://dublincore.org/documents/dcmi-terms/>

Zusammenfassung: Die Dublin Core Vokabulare enthalten die grundlegendsten Terme zur Beschreibung von Dokumenten, wie Titel (title), Verfasser (creator), Erscheinungsvermerk (publisher) usw. Die Dublin-Core-Vokabulare sind die am bekanntesten und mit dem FOAF-Vokabular die am meist genutzten¹⁶ (siehe Abbildung 1). Die meisten Vokabulare, die Dokumente beschreiben wollen, erweitern entweder diese Vokabulare oder verweisen auf sie durch Vocabulary Alignment. Der Unterschied der beiden Vokabulare liegt vor allem darin, dass die Metadata Terms genauer definiert und teilweise verfeinert worden sind.¹⁷

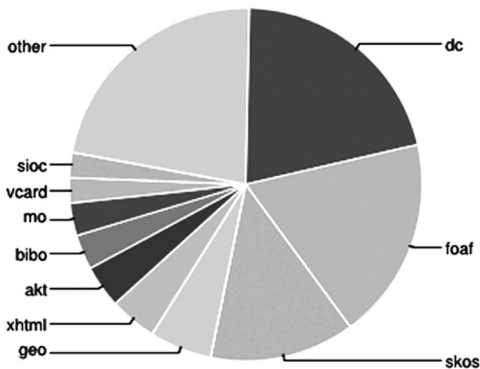


Abbildung 1: Verteilung der am häufigsten genutzten Vokabulare¹⁸

Friend of a Friend (FOAF)

Namespace-URI: <http://xmlns.com/foaf/0.1/>

Spezifikation: <http://xmlns.com/foaf/spec/>

Zusammenfassung: Ähnlich wie die Dublin-Core-Vokabulare stellt FOAF grundlegende Terme zur Verfügung. FOAF konzentriert sich dabei um Beziehungen zwischen Agenten (Personen, Gruppen, Organisationen) und Dokumenten. FOAF

¹⁶ Siehe Bizer et al.

¹⁷ Für einen guten Einblick in die unterschiedlichen Anwendungen der Dublin Core Vokabulare siehe das Dokument http://wiki.dublincore.org/index.php/User_Guide/Publishing_Metadata.

¹⁸ Abbildung übernommen aus Bizer et al.

ist neben Dublin Core einer der am meisten genutzten Vokabulare¹⁹ (siehe Abbildung 1).

Publishing Requirements for Industry Standard Metadata (PRISM)

Namespace-URI: <http://prismstandard.org/namespaces/basic/2.0/>

Spezifikation: http://www.prismstandard.org/specifications/2.0/PRISM_prism_namespace_2.0.pdf

Zusammenfassung: PRISM ist ein Vokabular, das ursprünglich Verlagen dazu diente, ihre Metadaten zu vereinheitlichen. Das PRISM-Vokabular reicht allerdings auch in die Welt der Bibliotheken und der im weitesten Sinne bibliographischen Daten hinein. Ontologien wie die Bibliographic Ontology und FRBR-aligned Bibliographic Ontology haben Terme dieses Vokabulars übernommen. PRISM ist daher als ein bekanntes und weit verbreitetes Vokabular anzusehen. Es ist aber leider ein schlechtes Beispiel, was das 5-Sterne-Schema angeht, da es in fast allen Punkten Schwachstellen aufweist.

Bibliographic Ontology (BIBO)

Namespace-URI: <http://purl.org/ontology/bibo/>

Spezifikation: <http://bibliontology.com/specification>

Zusammenfassung: Die Bibliographic Ontology ist das meist genutzte Vokabular, das speziell für die Beschreibung bibliographischer Daten entwickelt wurde (siehe Abbildung 1). BIBO versucht die grundlegendsten Klassen (Buch, mehrbändiges Werk, fortlaufendes Sammelwerk, Zeitschrift etc.), sowie die gebräuchlichen Eigenschaften (wie Titel, Autor, Verlag etc.) als auch tiefer gehende Beschreibungsmöglichkeiten (ISBN, ISSN, Seitenzahl, Herausgeber etc.) zusammenzufassen. Dabei macht die Ontologie vorbildlichen Gebrauch von anderen Vokabularen wie Dublin Core, FOAF, PRISM oder der Event-Ontologie, indem es die Klassen und Eigenschaften nachnutzt. BIBO ist der ideale Ausgangspunkt zur Beschreibung von bibliographischen Daten und wird daher in den meisten Linked-Library-Data-Projekten verwendet.

¹⁹ Siehe Bizer.

Schema.org

Namespace-URI: <http://schema.org/>

Spezifikation: <http://schema.org/docs/schemas.html>

Zusammenfassung: Schema.org ist ein Vokabular, das eine ganze Reihe von Schemata zu unterschiedlichen Themen umfasst. Schemata sind Sets von Klassen und deren Eigenschaften zu bestimmten Themen. Von besonderem Interesse dürfte in diesem Kontext das Schema CreativeWork mit den Klassen Article, Book, Map, MusicRecording, Photograph, WebPage etc. sein. Schema.org wurde speziell zur breiten Anwendung durch Webmaster in Zusammenhang mit HTML5, Microdata und RDFa entwickelt und wird zum Beispiel von OCLC im WorldCat²⁰ genutzt. Dennoch ist Schema.org allein nicht ausreichend, um umfassend bibliographische Daten zu beschreiben. Das Fehlen von Klassen wie Periodical oder Newspaper gleicht OCLC durch den Einsatz eines ergänzenden Vokabulars²¹ aus.

Expression of Core FRBR Concepts and Expression of Extended FRBR Concepts in RDF

FRBR Core

Namespace-URI: <http://purl.org/vocab/frbr/core#>

Spezifikation: <http://vocab.org/frbr/core.html>

FRBR Extended

Namespace-URI: <http://purl.org/vocab/frbr/extended#>

Spezifikation: <http://vocab.org/frbr/extended.html>

Zusammenfassung: Die beiden FRBR Vokabulare bilden die Repräsentation der Functional Requirements for Bibliographic Records. Da nach der Publikation der FRBR im Jahre 1998, 2005 immer noch ein durch Bibliothekare erstelltes Vokabular auf sich warten ließ, überführte Ian Davis die FRBR-Begriffe – allerdings nur sehr oberflächlich – nach RDF. Die SPAR-Ontologie Essential FRBR in OWL2

²⁰ Siehe OCLC.

²¹ Das ergänzende Vokabular ist unter <http://purl.org/library/> zu finden, hat aber noch Entwurfsstatus. Für die weitere Entwicklung hat OCLC/Richard Wallis eine W3C Community Group gegründet, siehe <http://www.w3.org/community/schemabibex/>.

DL und Teile des RDA-Vokabulars sind ebenfalls – teilweise abweichende – RDF-Umsetzungen des FRBR-Modells.

Semantic Publishing and Referencing Ontologies (SPAR)

Übersicht: <http://sempublishing.svn.sourceforge.net/viewvc/sempublishing/SPAR/index.html>

Zusammenfassung: Die Semantic Publishing and Referencing (SPAR) Ontologies bilden eine Reihe von sich ergänzenden Ontologie-Modulen für die Erstellung von maschinenlesbaren RDF-Metadaten für alle Aspekte des Publizierens, Referenzierens und Zitierens.²²

FaBiO, the FRBR-aligned Bibliographic Ontology

Namespace-URI und Spezifikation: <http://purl.org/spar/fabio/>

Zusammenfassung: FRBR-aligned Bibliographic Ontology ist ähnlich wie die Bibliographic Ontology ein Vokabular, das speziell in Hinblick auf die Beschreibung von bibliographischen Daten entwickelt wurde. Dabei ist FaBiO anders als BIBO²³ in das FRBR-Modell eingegliedert. Die Klassen und Eigenschaften von FaBiO können somit immer einer WEMI-Ebene zugeordnet werden. FaBiO ist weniger verbreitet als BIBO, und es gibt bis jetzt kaum Beispiele für den Einsatz der Ontologie.

Essential FRBR in OWL2 DL

Namespace-URI und Spezifikation: <http://purl.org/spar/frbr/>

Zusammenfassung: Ähnlich wie die FRBR-Umsetzung in RDF von Ian Davis, spiegelt die Essential FRBR in OWL2 DL Ontologie die Basiskonzepte der Functional Requirements for Bibliographic Records wieder. Durch die Beschreibung der Ontologie in OWL2 DL²⁴ können die FRBR-Konzepte und ihre Beziehungen aller-

22 Neben den beiden hier erwähnten Vokabularen gibt es außerdem folgende SPAR-Ontologien: the Publishing Roles Ontology (PRO), the Publishing Status Ontology (PSO), the Publishing Workflow Ontology (PWO), the Citation Typing Ontology (CiTO), the Bibliographic Reference Ontology (BiRO), the Citation Counting and Context Characterization Ontology (C4O), the Document Components Ontology (DoCO).

23 Für einen tiefer gehenden Vergleich der beiden Ontologien FaBiO und BIBO siehe Shotton.

24 Siehe dafür folgende Dokumente: <http://www.w3.org/TR/owl2-direct-semantics/> und <http://www.w3.org/TR/2009/REC-owl2-overview-20091027/>.

dings sehr viel umfangreicher abgebildet werden, als es das schon beschriebene FRBR-Vokabular kann.

The RDA (Resource Description and Access) Vocabularies

Übersicht²⁵: <http://rdvocab.info/>

Zusammenfassung: Entwickelt von der DCMI/RDA Task Group bilden die RDA Vocabularies die RDF-Repräsentation der RDA-Elemente und ihrer Eigenschaften. Die RDA Vocabularies sind unterteilt in 6 Element Sets (Vokabulare mit Properties) und ca. 70 Vocabularies (Vokabulare mit Klassen). Da RDA auf FRBR aufbaut, sind auch die Eigenschaften und Klassen den WEMI-Ebenen²⁶ zugeordnet. Allerdings wurden die Vokabulare so modelliert, dass parallel übergeordnete Terme genutzt werden können, die außerhalb des FRBR-Modells stehen (generalized properties)²⁷. Große Teile des Vokabulars haben noch den Status eines Diskussionsvorschlags und sind daher noch nicht abgeschlossen. Da die RDA-Vokabulare insgesamt sehr umfangreich sind, RDA als Regelwerk erst noch eingeführt werden muss und sich die Bibliothekswelt noch nicht auf RDF als Datenformat festgelegt hat, ist die Rezeption der Vokabulare noch verhalten.

CIDOC CRM / FRBRoo

Übersicht: http://www.cidoc-crm.org/official_release_cidoc.html

Zusammenfassung: Bibliographische Daten im FRBR-Kontext sind, gerade wenn es um fortlaufende Sammelwerke geht, aufgrund ihrer Komplexität nicht ohne Probleme im Web der Daten abzubilden. FRBRoo erweitert das CIDOC Conceptual Reference Model (CRM) und „erlaubt es, aufgrund der netzartigen Struktur des CRM, bibliographische Informationen in Linked-Data-Kontexte zu übertragen“. Genauer beschrieben wird das in dem Beitrag *FRBR, Serials und CIDOC CRM - Modellierung von fortlaufenden Sammelwerken unter Verwendung von FRBRoo* von Hans-Georg Becker.

²⁵ Übersicht der Vokabulare ISBD Elements, ISBD Content Form, ISBD Content Qualification of Dimensionality, ISBD Content Qualification of Motion, ISBD Content Qualification of Sensory Specification, ISBD Content Qualification of Type, ISBD Media Type.

²⁶ WEMI: Work, Expression, Manifestation, Item

²⁷ Einen guten Einblick in die Entwicklung der RDA Vocabularies bietet Hillmann.

ISBD Elements und Vocabularies

Übersicht: <http://iflstandards.info/ns/isbd/>

Zusammenfassung: Die ISBD Vokabulare sind die RDF-Repräsentation der Elemente des IFLA-Regelwerkes International Standard Bibliographic Description (ISBD) und enthalten damit alle wichtigen Angaben einer Titelaufnahme. Ohne das Regelwerk genau zu kennen, ist es schwer abzuschätzen, in wie weit die ISBD Vokabulare dem Regelwerk entsprechen. Allerdings dürfte festzuhalten sein, dass die ISBD Vokabulare, vielleicht aufgrund ihrer Regelwerksbasis mit internationalem Anspruch oder aufgrund ihrer einfachen, flachen und übersichtlichen Struktur und gebräuchlichen Konzepte, weite Verbreitung in bibliothekarischen Linked Library Data Projekten finden. Oft werden ISBD Terme als Ergänzung zur Bibliographic Ontology (BIBO) genutzt.

MARC21 Vocabularies

Übersicht: <http://marc21rdf.info/elements/>

Zusammenfassung: Die MARC21 Vokabulare sind im Gegensatz zu den ISBD Vokabularen sehr viel umfangreicher. Sie bestehen aus neun Element Sets (MARC-Kategorien OXX-8XX) und über 100 Vocabularies. Kurz nach Veröffentlichung des Vokabulars gab es aus der Community starke Kritik an der Eins-zu-eins-Umsetzung der MARC-Kategorien.²⁸ Die RDF-Repräsentation habe keine großen Vorteile gegenüber einem MARC-Datensatz. Die MARC21-Vokabulare scheinen daher noch keine weite Verbreitung gefunden zu haben. Dennoch dürften Teile der Vokabulare ebenso wie die ISBD-Vokabulare eine gute Ergänzung zu anderen Vokabularen darstellen.

MODS Ontology

Namespace-URI und Spezifikation: <http://simile.mit.edu/2006/01/ontologies/mods3#>

Zusammenfassung: Das MODS (Metadata Object Description Schema) wird von Bibliotheken hauptsächlich im Bereich der Beschreibung von digitalen Objekten und deren Langzeitarchivierung eingesetzt. Daher ist das Vokabular am ehesten

²⁸ Siehe <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind1109&L=dc-rda&D=0&P=1704> und nachfolgende Diskussion.

für eine Konvertierung von MODS-XML nach MODS-RDF geeignet. Dafür geeignete Konvertierungstools sind nicht schwer zu finden.

Document Availability Information Ontology (DAIA)

Namespace-URI: <http://purl.org/ontology/daia/>

Spezifikation: <http://uri.gbv.de/ontology/daia/>

Zusammenfassung: Mit DAIA kann die Verfügbarkeit von Dokumenten in Bibliotheken und anderen Informationseinrichtungen beschrieben werden. Die von Jakob Voß entwickelte Ontologie wurde zwar nicht für die klassische bibliographische Formalbeschreibung entwickelt, ist aber in Teilen, wenn es etwa um die Beziehung zwischen Institution und Dokument geht, auch für den vorliegenden Kontext relevant.

LOC Authorities and Vocabularies

Übersicht: <http://id.loc.gov/>

Zusammenfassung: Authorities and Vocabularies ist der Sammelbegriff für die Vokabulare der Library of Congress, die eine Vielzahl von gebräuchlichen Begriffen und unter anderem die gesamten Library of Congress Subject Headings (LCSH) als Ressourcen zur Verfügung stellen.

Subject Headings (LCSH)

Spezifikation: <http://id.loc.gov/authorities/names.html>

Namespace-URI: <http://id.loc.gov/authorities/subjects/>

Zusammenfassung: Die Schlagwortnormdatei der Library of Congress

Name Authority File (NAF)

Spezifikation: <http://id.loc.gov/authorities/names.html>

Namespace-URI: <http://id.loc.gov/authorities/names/>

Zusammenfassung: Namen von Personen, Organisationen, Ereignissen, Orten und Titeln

Children's Subject Headings (LCSHAC)

Spezifikation: <http://id.loc.gov/authorities/childrensSubjects.html>

Namespace-URI: <http://id.loc.gov/authorities/childrensSubjects/>

Zusammenfassung: Erweiterung der LCSH auf dem Gebiet Kinder und Jugend

Genre/Form Terms for Library and Archival Materials (LCGFT)

Spezifikation: <http://id.loc.gov/authorities/genreForms.html>

Namespace-URI: <http://id.loc.gov/authorities/genreForms/>

Zusammenfassung: Anders als Schlagwörter beschreiben diese Terme eindeutig das Genre des Werks.

Thesaurus for Graphic Materials

Spezifikation: <http://id.loc.gov/vocabulary/graphicMaterials.html>

Namespace-URI: <http://id.loc.gov/vocabulary/graphicMaterials/>

Zusammenfassung: Kann genutzt werden, um Inhalt aber auch Form und Genre von graphischen Materialien zu beschreiben.

MARC Code List for Relators

Spezifikation: <http://id.loc.gov/vocabulary/relators.html>

Namespace-URI: <http://id.loc.gov/vocabulary/relators/>

Zusammenfassung: Die MARC Code List for Relators beschreiben die Beziehung zwischen einer bibliographischen Ressource und einer Person oder Organisation. Die Relators wurden für den Gebrauch in MARC 21 codierten bibliographischen Daten entwickelt. Die meisten Relatorterme sind Refinements des Dublin Core Terms contributor (an entity responsible for making contributions to the content of the resource). Das Vokabular ist ein ideales Instrument, wenn es um die Verknüpfung von bibliographischen Ressourcen und Normdaten geht.²⁹

²⁹ Ein gutes Beispiel für den Einsatz von MARC Relator Roles findet sich unter <https://wiki1.hbz-nrw.de/x/vYU7>.

MARC Countries

Spezifikation: <http://id.loc.gov/vocabulary/countries.html>

Namespace-URI: <http://id.loc.gov/vocabulary/countries/>

Zusammenfassung: Liste von Ländercodes, zu unterscheiden von ISO 3166.

Geographic Areas

Spezifikation: <http://id.loc.gov/vocabulary/geographicAreas.html>

Namespace-URI: <http://id.loc.gov/vocabulary/geographicAreas/>

Zusammenfassung: Terme für geographische Regionen oder Areale. Zum Beispiel „Europa“ oder „Deep Space“.

MARC List for Languages

<http://id.loc.gov/vocabulary/languages>

ISO639 Languages

Verschiedene Versionen des ISO-Standards 639

<http://id.loc.gov/vocabulary/iso639-1>

<http://id.loc.gov/vocabulary/iso639-2>

<http://id.loc.gov/vocabulary/iso639-5>

BIBFRAME-Vokabular

Namespace-URI und Spezifikation: <http://bibframe.org/vocab/>

Zusammenfassung: Die „Bibliographic Framework Transition Initiative“ arbeitet an der Ablösung von MARC 21 als Austauschformat und Z39.50 als Protokoll, und entwickelt die entsprechenden Nachfolgetechnologien auf Basis von Linked-Data-Standards. Das Bibframe-Vokabular spielt dabei eine zentrale Rolle und wurde im November 2012 in einem ersten Entwurf publiziert. Es soll nicht nur existierende MARC-Daten in RDF abbilden können, sondern auch nach RDA katalogisierte Daten, wie auch bibliographische Daten von nicht-bibliothekarischen Institutionen. Aus diesem Grund ist das Vokabular weniger detailliert

als etwa RDA und FRBR. So wird beispielsweise das aus FRBR bekannte WEMI-Modell nicht übernommen, sondern nur zwei entsprechende Klassen – „Work“ und „Instance“ – geprägt. Damit verschiedene Akteure leicht dezentral (lokale) Informationen zu einer bibliographischen Ressource ergänzen können und dabei Provenienzangaben vorliegen, wurde die Klasse „Annotation“ eingeführt.

Das Vokabular ist noch in der Entwicklung und die Details werden auf der Bibframe-Mailingliste diskutiert.³⁰

Anwendung von Vokabularen

Die Auswahl an Vokabularen zur Beschreibung von bibliographischen Daten ist beträchtlich. Die hier vorgestellten Vokabulare bilden mithin nur einen Ausschnitt von den Vokabularen, die sinnvoll zum Einsatz gebracht werden können.

Die Entscheidung für ein Vokabular aber muss erst gar nicht getroffen werden. RDF bietet durch seine Erweiterbarkeit immer die Möglichkeit, mehrere Vokabulare zu nutzen. Nehmen Sie sich daher das Beste aus allen Welten.

Das Hochschulbibliothekszentrum hbz nutzt für seinen Linked Data Service lobid.org³¹ bei Fertigstellung dieses Dokuments dreizehn Vokabulare. Eher muss eine Auswahl über die konkurrierenden Terme der unterschiedlichen Vokabulare³² getroffen werden.³³

Bei der Vermischung von Vokabularen ist allerdings darauf zu achten, dass nicht jede Eigenschaft jeder Klasse zugeordnet werden kann:

```
@prefix mods: <http://simile.mit.edu/2006/01/ontologies/mods3#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix dbp: <http://dbpedia.org/resource/> .
@prefix bibo: <http://purl.org/ontology/bibo/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
mods:extent rdfs:domain mods:Description .
dbp:The_Name_of_the_Rose rdf:type bibo:Book ;
#Das ist nicht richtig
                                mods:extent „632 pages“ .
dct:extent rdfs:range dct:SizeOrDuration .
```

30 <http://listserv.loc.gov/listarch/bibframe.html>

31 <http://lobid.org>

32 <http://prefix.cc/dc,dct,bibo,daia,frbr,isbd,gn,lv,marcrel,rdf,owl,geo,foaf>

33 Wie unterschiedlich die Anwendung der einzelnen Terme der Vokabulare sein kann, zeigt die Arbeit der „Titeldaten Gruppe“ innerhalb der DINI-AG KIM. Siehe <https://wiki.d-nb.de/display/DINIAGKIM/Titeldaten+Gruppe> .


```
#Das ist umständlicher, aber richtig
dbp:The_Name_of_the_Rose dct:extent[ rdf:value „632 pages“ ].
```

In den obigen Beispielen wird einmal mit der `bibo:Book` Instanz `dbp:The_Name_of_the_Rose` die Eigenschaft `mods:extent` benutzt. Da `mods:extent` aber ein Definitionsbereich (domain) von `mods:Description` hat, und die Klasse `bibo:Book` in keinem Verhältnis zu `mods:Description` steht, darf die Eigenschaft `mods:extent` nicht genutzt werden. Sie ist keine Eigenschaft von `bibo:Book`, sondern nur von `mods:Description`. Die Eigenschaft `dct:extent` darf allerdings mit der Instanz `dbp:The_Name_of_the_Rose` genutzt werden. Für `dct:extent` ist kein Definitionsbereich angegeben und wird damit automatisch `owl:Thing`, der Klasse aller Klassen, zugeordnet. Allerdings hat `dct:extent` ein Wertebereich (range) von `dct:SizeOrDuration`, ist also eine Object Property, und darf somit nicht mit Literalen verwendet werden. Die Modellierung ist daher an dieser Stelle etwas umständlicher.

Soweit auf solche Modellierungsprobleme geachtet wird, steht es einem frei, Vokabulare nach eigenem Geschmack zu mischen. Nicht selten aber wird es vorkommen, dass der benötigte Term in keinem der Vokabulare vorhanden ist. Bevor die Beschreibung der Daten durch den Einsatz von ähnlichen oder sehr allgemeinen Termen zu ungenau wird, sollte man sich dafür entscheiden, selber einen Namensraum zu besetzen und eigene Terme zu nutzen. Diese können dann durch Vocabulary Alignment mit Termen anderer Vokabulare in Relation gesetzt werden. Ein guter Einstieg, um Fehler zu vermeiden und um eine möglichst große Nachnutzung seiner Daten zu erreichen, ist das Online-Buch „Linked Data Patterns“ von Leigh Dodds und Ian Davis.³⁴

Zusammenfassung

An Vokabularen zur Beschreibung bibliographischer Daten im Web der Daten gibt es eine große Auswahl, die hier nur überblicksweise und in Ausschnitten beleuchtet werden konnte. Anders als die bibliothekarischen Austauschformate MARC oder MAB, sind die in Linked-Library-Data-Projekten eingesetzten Vokabulare einfach und überschaubar. Große umfangreiche Vokabulare werden eher weniger genutzt. Alles in allem entspricht dies den Erfahrungen, die im Zusammenhang mit dem Semantic Web und Linked Data gemacht wurden: Kleine übersichtliche und gut dokumentierte Vokabulare setzten sich durch. Je mehr

³⁴ Siehe Dodds.

Vokabulare es zudem geben wird, desto wichtiger wird es sein, dass diese gut dokumentiert sind und dem 5-Sterne-Schema entsprechen.

Gerade wenn es um offene Daten geht (Linked Open Data), wird die Verwendung von weit verbreiteten und gut dokumentierten Vokabularen umso wichtiger, da dadurch die Interoperabilität für Dritte, die Ihre Daten nachnutzen, steigt. Und Interoperabilität ist wichtig, denn wenn man seine Daten freigibt, dann möchte man auch, dass sie nachgenutzt werden. Aber auch für geschlossene Datenräume ist Interoperabilität ein entscheidender Vorteil für das Datenmanagement.

Aber egal, ob nun offene Daten oder nicht, bei der Verwendung von Vokabularen, genau wie bei der Erstellung von Metadaten, ist immer auf seine Stakeholder oder die Community zu achten. In der Bibliothekswelt bedeutet das, dass zum einen die Welt ihre Daten nachnutzen möchte. In der Kommunikation mit dieser wird es nötig sein, die weit verbreiteten Vokabulare zu nutzen, um die Nachnutzung in möglichst vielen Kontexten zu ermöglichen. Es bedeutet aber auch, dass sich auf Vokabulare verständigt werden muss, um Daten unter Bibliotheken auszutauschen.

Die klassischen Austauschformate sind veraltet und werden bald ersetzt werden müssen. Gerade mit der bevorstehenden Einführung von RDA wird es nötig sein, andere Wege zu finden, um die Daten sinnvoll zu beschreiben. An Vokabularen für bibliographische Daten mangelt es jedenfalls nicht und die vielen existierenden Library-Linked-Data-Projekte zeigen, dass auf diesem Gebiet eine Menge passiert.

Quellen

- Berners-Lee, Tim: Linked Data. URL: <http://www.w3.org/DesignIssues/LinkedData.html> (zuletzt geprüft 20.04.2012)
- Bizer, Chris; Jentzsch, Anja; Cyganiak, Richard: State of the LOD Cloud. Version 0.3, 09/19/2011. URL: http://www4.wiwiwiss.fu-berlin.de/lodcloud/state/2011-09_index.html (zuletzt geprüft 10.10.2012)
- Carroll, J. J.; Klyne, G.: Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation. 10 February 2004. URL: <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/#section-Literals> (zuletzt geprüft 24.04.2012)
- Dodds, Leigh; Davis, Ian: Linked Data Patterns : A pattern catalogue for modelling, publishing, and consuming Linked Data. 19. August 2011. URL: <http://patterns.dataincubator.org> (zuletzt geprüft 03.05.2012)
- Hillmann, Diane; Coyle, Karen; Phipps, Jon u.a.: RDA Vocabularies: Process, Outcome, Use. In: D-Lib Magazine, 16.1/2 (2010). URL: <http://www.dlib.org/dlib/january10/hillmann/01hillmann.html> (zuletzt geprüft 24.04.2012)

- Ideenlehre. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 14. September 2012, 06:51 UTC. URL: <http://de.wikipedia.org/w/index.php?title=Ideenlehre&oldid=108049170> (zuletzt geprüft 06.10.2012)
- OCLC Online Computer Library Center, Inc: OCLC adds Linked Data to WorldCat.org. 20. Juni 2012. URL: <http://www.oclc.org/news/releases/2012/201238.htm> (zuletzt geprüft 10.10.2012)
- Schulze, Carsten M.: Mikroformate für bibliographische Daten : Vergleich verschiedener Konzepte zur semantischen Annotation. Fachhochschule Potsdam, FB Informationswissenschaften, Diplomarbeit, 2008. Abschnitt Bibliographische Daten im WWW. Seite 3 ff.
- Semantic reasoner. In Wikipedia, The Free Encyclopedia. Bearbeitungsstand: 7. Oktober 2012, 17:42. URL: http://en.wikipedia.org/w/ihttp://blog.hubjects.com/2012/02/is-your-linked-data-vocabulary-5-star_9588.html [Index.php?title=Semantic_reasoner&oldid=513866223](http://blog.hubjects.com/2012/02/is-your-linked-data-vocabulary-5-star_9588.html) (zuletzt geprüft 07.10.2012)
- Shotton, David: Comparison of BIBO and FaBiO. In: Open Citations and Semantic Publishing, vom 29.07.2011. URL: <http://opencitations.wordpress.com/2011/06/29/comparison-of-bibo-and-fabio/> (zuletzt geprüft 24.04.2012)
- Vatant, Bernard: Is your linked data vocabulary 5-star?. URL: http://blog.hubjects.com/2012/02/is-your-linked-data-vocabulary-5-star_9588.html (zuletzt geprüft 21.04.2012)