



**Kontrollierte Vokabulare der Library of Congress, der Virtual International Authority File und ihre Anwendung im Semantic Web**

**Barbara B. Tillett und Corey Harper**

**Meeting: 147 - IFLA-CDNL Alliance for Bibliographic Standards (ICABS)**

**Simultaneous Interpretation: yes**

*WORLD LIBRARY AND INFORMATION CONGRESS: 73RD IFLA GENERAL CONFERENCE AND COUNCIL*  
19-23 August 2007, Durban, South Africa  
<http://www.ifla.org/iv/ifla73/index.htm>

**Abstract:**

*Dieser Vortrag basiert auf einem Artikel, den die Autorin zusammen mit Corey Harper für die Zeitschrift *Cataloging & Classification Quarterly* geschrieben hat („Library of Congress controlled vocabularies and their application to the Semantic Web”, by Corey A. Harper and Barbara B. Tillett, v.43, no. 3 / 4, 2006). Er befasst sich mit den kontrollierten Vokabularen der Library of Congress (Library of Congress Subject Headings, Library of Congress Classification und Library of Congress/NACO Authority Files) und beschreibt das VIAF-Projekt (Virtual International Authority File). Diese kontrollierten Vokabulare können hoffentlich als Bausteine des Semantic Web dazu dienen, internationale Normdaten aus verlässlichen Quellen zum Nutzen der Benutzer in aller Welt zu verlinken.*

Der 2001 in der Zeitschrift *Scientific American* erschienene Artikel von Tim Berners-Lee zum Semantic Web<sup>1</sup> beschreibt erstmals die Idee eines verlinkten, universellen Datennetzes. Er erwähnte dort, dass voneinander unabhängige Gruppen an ähnlichen Konzepten arbeiten und eine gemeinsame Sprache benötigen, um besser zusammenarbeiten zu können. **Bibliotheken und die Entwickler des Semantic Web haben gemeinsame Ziele: die Benennung von Begriffen, die Benennung von Entitäten und das Zusammenführen verschiedener Formen dieser Benennungen.** Bibliothekarische Werkzeuge werden seit Jahrzehnten entwickelt und sind wertvolle Quellen für verknüpfte Daten. Wir müssen sie jetzt nur zu neuen Werkzeugen weiterentwickeln, welche die Infrastruktur des Semantic Web unterstützen können. Wir gehen von einem breiten Einverständnis aus, das uns bei der Entwicklung von Werkzeugen für die Zukunft hilfreich sein kann. Den Bibliotheken kommt hierbei eine wichtige Rolle zu.

Ein erster großer Schritt wird es sein, unsere kontrollierten Vokabulare in Semantic-Web-Standards umzuwandeln, wie zum Beispiel die Web Ontology Language (OWL), so dass sie für neue Anwendungen zur Verfügung stehen können. Darüber hinaus gibt es die Semantic-

Web-Technologie Simple Knowledge Organisation System Core (SKOS) für die Codierung der Thesaurus-Inhalte. So stellen beide, SKOS und OWL, für unsere Normdaten wichtige Wege dar, die Normdateninformationen zu übersetzen. Ebenso hilfreich für das Semantic Web ist es, unsere bibliografischen Daten in RDF (Resource Description Framework) zu übersetzen. Ein erster Schritt in diese Richtung erfolgt gerade mit den Dublin Core und RDA (Resource Description and Access) Communities.<sup>2</sup>

Die hohe Akzeptanz und Wertschätzung der Beständigkeit und Vertrauenswürdigkeit von kontrolliertem bibliothekarischem Vokabular hilft uns außerordentlich, Dinge voran zu bringen.

Ein Großteil der bisherigen Arbeit, insbesondere für digitale Bibliotheken, besteht darin, Bibliotheksbestände im Netz bekannt zu machen, durch Digitalisierung der Inhalte ausgewählter Sammlungen. Bibliothekarische Normdaten haben ein ständig wachsendes Potenzial, Nutzern dabei zu helfen, im Web zu finden, was sie brauchen, indem sie Suchwege ermöglichen, die sie mit relevanten Informationen verbinden. Der Begriff Web 2.0 wird oft benutzt, um benutzerinteraktive Systeme im Internet zu beschreiben: Anwendungen, welche die Nutzer befähigen, Dinge die sie finden, zu personalisieren, eigene Kommentare, eigene „Folksonomien“ oder Tags hinzuzufügen – wir würden das Schlagwörter oder Sucheinstiege nennen –, um Dokumente, die sie in ihrer eigenen Umgebung nutzen wollen, leichter auffindbar und benutzbar zu machen.

Hier einige Beispiele für kontrollierte Vokabulare, die von der Bibliotheks-Community geschaffen wurden: die Dewey Decimal Classification (DDC) und die Library of Congress Classification (LCC). Ebenso die Schlagwortverzeichnisse – die Library of Congress Subject Headings (LCSH), zwei Thesauri für grafisches Material: Subject Terms (TGM I) und Genre & Physical Characteristic Terms (TGM II), Guidelines on Subject Access to Individual Works of Fiction, Drama, etc. (GSAFD) und der Ethnographic Thesaurus. Derzeit gibt es keine Suchinstrumente, welche die syndetischen Strukturen (Cross-Konkordanzen) dieser kontrollierten Vokabulare nutzen können.

Kontrollierte Vokabulare können die Werkzeuge und Dienste des Semantic Web und des Web 2.0 wesentlich effektiver nutzbar machen, indem sie Strukturen und Verknüpfungen für Inhalte und für Namen von Personen und Dingen usw. zur Verfügung stellen.

Darüber hinaus besteht durch die Technologie des Semantic Web auch ein Gewinn für die kontrollierten Vokabulare. Die Darstellung der syndetischen Strukturen solcher Vokabulare mittels gebräuchlicher Frameworks wie OWL oder SKOS macht die Interoperabilität zwischen den einzelnen Vokabulare verständlicher.

Vokabulare in dieser Weise darzustellen und zu modellieren, erleichtert die Bestimmung und Auswertung von Begriffsrelationen und Äquivalenzen in unterschiedlichen kontrollierten Vokabularen. Das könnte zum Beispiel einige der Clustering-Möglichkeiten verbessern, die in gerade in der Entwicklung befindlichen Katalogsystemen der nächsten Generation auftauchen.

OCLC hat versuchsweise Standardsoftware, wie Microsoft Word, bei der Anwendung von kontrollierten Vokabularen eingesetzt. Im OCLC Terminologies Project werden Word-Nutzer befähigt, kontrollierte Vokabulare zu konsultieren und Schlagwörter zu finden, die sie in ihren Text einbinden können, ohne Word verlassen zu müssen.<sup>3</sup> Wir können uns vorstellen, dass Katalogisierer über ein ähnliches Werkzeug verfügen, welches ihnen einen schnellen Zugriff auf passende Begriffe in einem kontrollierten Vokabular innerhalb eines effizienten, Zeit und Katalogisierungskosten einsparenden Systems ermöglicht. Das Terminologies Project legt

bereits die Grundlagen für diese Entwicklung durch die Integration der Microsoft Office Werkzeuge in die Katalogisierungs-Software OCLC Connexion.

Ebenso können wir unsere bereits existierenden Normdateien für Personen in einer webfreundlicheren Struktur zugänglich machen, so dass sie für Web-Anwendungen genutzt werden können. Es besteht die Hoffnung, dass der Virtual International Authority File sich letztlich in eine solche Ressource entwickelt.

Das Projekt Virtual International Authority File (VIAF) ist ein lange gehegter Traum. Ein entsprechendes Modell wird in Zusammenarbeit mit der Deutschen Nationalbibliothek, der Library of Congress, der Bibliothèque Nationale de France und OCLC ausprobiert. Die ersten Ergebnisse werden getestet, aber wir müssen noch die Benutzer-Oberfläche und Möglichkeiten entwickeln, um über das Web besser auf die Daten zugreifen zu können. Ich möchte Sie an dieser Stelle auf verschiedene IFLA-Veröffentlichungen und andere Berichte über das VIAF-Projekt hinweisen und sehe davon ab, sie hier zu wiederholen. Die Grundidee des VIAF ist ein System mit internationaler Beteiligung, das die wichtigsten Normdateien der Welt in einem frei zugänglichen System miteinander verbindet, welches die Normdatensätze für dieselben Entitäten zusammenführt, und alle Normdatensätze für die Verwendung in vielen Anwendungen verfügbar macht. Diese gemeinsame Datei würde dazu beitragen, Katalogisierungskosten in den einzelnen Bibliotheken zu reduzieren, mithilfe eines Tools zur Anzeige in einer vom Benutzer gewünschten Sprache oder Schrift für Personennamen (und später auch für Körperschaften und geografische Namen). Ein einzelnes System wie dieses würde weitaus nützlicher sein, als alle Normdateien weltweit getrennt benutzen zu müssen, und könnte von Web-Systemen benutzt werden, um bei der Recherche die Präzision wahren zu lassen, die zurzeit bei Suchmaschinen wie Google leider fehlt.

Diese Sicht auf Normdatensätze bedeutet, dass alle Namensvarianten, die im Normdatensatz zusammengetragen sind, und zusätzlich die verlinkten Normdatensätze suchbar und zur Anzeige durch den Endnutzer verfügbar sein können. In der Vergangenheit haben Bibliotheken nur eine einzige Namensform als normierte Anzeigeform gewählt, und ein solcher Standardeintrag wird auch weiterhin benötigt werden, wenn ein Nutzer keine bestimmte Sprache oder Schrift auswählt.

Wir hoffen, das VIAF-Projekt auch auf nicht-lateinische Schriften ausdehnen zu können.

Wenn wir VIAF außerdem auf die Semantic-Web-Communities ausdehnen, kann die Stärke unseres Normdatensystems in den verschiedensten Arten von Anwendungen benutzt werden. Mehrere Wege öffnen sich zu den Ressourcen von oder über eine gesuchte Person oder zu anderen interessanten Informationsquellen, die mit ihr verbunden sind – ihr Blog, ein Wikipedia-Eintrag, Bestände von oder über diese Person in der Bibliothek vor Ort, biografische Informationen in Lexika oder Enzyklopädien, Zeitschriftenartikel, die von der oder über die gesuchte Person geschrieben wurden, ihre Artikel, die über Internet-Handel bezogen werden können, und vieles andere mehr. Diese Querverbindungen tragen auch dazu bei, die Glaubwürdigkeit und Authentizität dessen, was man im Netz findet, zu bewerten oder den Zusammenhang mit den persönlichen Interessen der gesuchten Person herzustellen, die wiederum Rückschlüsse auf das zulassen, was sie schreibt.

Wie können wir nun ausgehend von unseren heutigen Normdaten dorthin kommen? Zunächst müssen unsere Normdatensätze in einem maschinenlesbaren Format vorliegen, besser noch in Formaten, wie OWL oder SKOS, die für das Semantic-Web gemacht wurden. Ein weiterer Schritt ist es, jeden Begriff mit einem URI zu versehen oder stattdessen die XML-Version der

Normdatensätze in SKOS abzubilden. Eine weitere Alternative ist, das Format „on the fly“ aus dem Originalformat zu konvertieren. Um die Daten in jeder einzelnen Anwendung benutzen zu können, kann der VIAF zum Beispiel in MARCXML gehalten werden, woraus die Daten je nach Bedarf in RDF oder SKOS konvertiert werden können. Ebenso könnte er Systeme anreichern, die Folksonomien verwenden.

So können Informationen aus Normdateien in einer Reihe von anderen Anwendungen sehr kurzfristig weiterverwendet werden. Die eindeutige Benennung von Normdaten mit URIs ermöglicht es, diese URIs in anderen Umgebungen weiter zu verwenden, so dass andere Beschreibungen von Personen mit diesen Normdatensätzen verbunden werden können, die ihrerseits wieder mit den Werken der Person verbunden sind. Neu aufkommende Standards für die Codierung von Beziehungen zwischen Personen, wie das Projekt Friend of a Friend (FOAF), können für diese Informationen von großem Nutzen sein.<sup>4</sup> FOAF liefert Mechanismen, um Beziehungen sowohl zwischen verschiedenen Personen als auch zwischen Individuen und den verschiedenen Dingen, die sie machen, zu dokumentieren. Die Möglichkeit, URIs von VIAF-Datensätzen in FOAF-Beschreibungen einzubringen, hilft der Bibliotheks-Community, ihre Rolle als Anbieter zuverlässiger und glaubwürdiger Dateninhalte auszuweiten. So würde es einfacher, einen Teilnehmer in einem Blog oder den Autor eines Wikipedia-Eintrags als eine angesehene Autorität zu einem bestimmten Thema auszumachen.

Wie wir in unserem Artikel aus dem Jahr 2007 angemerkt haben, „ist es ein Teil der Vision des Semantic Web, durch die Entwicklung von Werkzeugen, die den Benutzern helfen, die Suche zu verfeinern und weiterzuentwickeln, Hilfestellung bei der Ressourcenermittlung zu geben sowie Unterstützung bei der Navigation in den Suchergebnissen zu bieten.“<sup>5</sup> Die von Bibliotheken entwickelten kontrollierten Vokabulare können ein leistungsstarkes und wunderbares Hilfsmittel sein, um die Resultate für die Endnutzer im Web zu verbessern.

Ein wichtiges Treffen, das vom 30. April bis 1. Mai 2007 in London veranstaltet wurde, brachte Vertreter von SKOS, Semantic Web, Dublin Core, IEEE/LOM und RDA (Resource Description and Access) auf Einladung der RDA-Herausgeber zusammen. Das Treffen hatte zum Ziel, verschiedene Datenmodelle (einschließlich FRBR und FRAD) zu prüfen und die Nützlichkeit der RDA für einige der Metadaten-Communities zu untersuchen. Die Ergebnisse dieses Treffens wurden unverzüglich veröffentlicht, und es wurden Empfehlungen für die Suche nach Finanzierungsmöglichkeiten zur Entwicklung eines RDA Application Profile gegeben, welches folgende Anforderungen beinhaltet:

- Entwicklung eines RDA Element Vocabulary
- Entwicklung eines RDA DC (Dublin Core) Application Profile basierend auf FRBR und FRAD und
- Entwurf eines RDA Value Vocabulary für RDF/RDFS/SKOS

Wie in der Mitteilung steht, „wird der Nutzen unserer Arbeit darin bestehen, dass:

- die Bibliotheks-Community einen Metadatenstandard erhält, der mit der Web-Architektur kompatibel ist und vollständige Interoperabilität mit anderen Semantic-Web-Initiativen bietet,
- die DCMI-Community ein Bibliotheksanwendungsprofil erhält, das auf DCAM und FRBR basiert (was ein viel beachtetes Beispiel sein wird, dem andere folgen können),
- die Semantic-Web-Community einen umfangreichen Pool wohldurchdachter Metadaten-Begriffe erhält, welche sie weiter verwenden kann und

- eine weitere Akzeptanz von RDA erreicht wird.“

Das Einbringen von Erfahrung, Standards und kontrollierten Vokabularen aus dem Bibliotheksbereich in das Semantic Web wird zum Vorteil für die Bibliotheks- und Metadaten-Communities sein. Die Nutzer werden die Hauptnutznießer sein.

Übersetzung: Renate Behrens-Neumann, Deutsche Nationalbibliothek

---

<sup>1</sup> Berners-Lee, T., Hendler, J., & Lasilla, O. "The Semantic Web" [Elektronische Version]. *Scientific American*, 284, no.5, 2001, p. 34-43. <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>

<sup>2</sup> Siehe die Mitteilung über die Ergebnisse der Sitzung vom 30. April 30 – 1. Mai 2007 in London "Data Model Meeting" unter <http://www.bl.uk/services/bibliographic/meeting.html>

<sup>3</sup> Vizine-Goetz, Diane. "Terminology services: making knowledge organization schemes more accessible to people and computers." *OCLC Newsletter*, no. 266, 2004. Zugänglich unter: <http://www.oclc.org/news/publications/newsletters/oclc/2004/266/>

<sup>4</sup> Siehe die Homepage des Projekts Friend of a Friend. <http://www.foaf-project.org/>

<sup>5</sup> Harper, Corey A. & Tillett, Barbara B., "Library of Congress controlled vocabularies and their application to the Semantic Web," *Cataloging & Classification Quarterly*, v.43, no. 3/4, 2006.