

# Bibliotheken als Akteure im Semantic Web

Veranstaltung in der Bibliothek der  
Westsächsischen Hochschule Zwickau

von **SILVIA CZERWINSKI**

**D**as Semantic Web wird schon seit längerem im bibliothekarischen Umfeld thematisiert. So finden sich seit über fünf Jahren sowohl Vorträge auf den Bibliothekstagen und -kongressen wie auch Artikel in den einschlägigen Fachzeitschriften. Nach einer zunächst eher theoretisch geprägten Debatte gibt es inzwischen auch ganz konkrete Entwicklungen: So war die vom Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (HBZ) organisierte erste Konferenz zu „Semantic Web in Bibliotheken“ (<http://www.swib09.de>) Ende 2009 in kürzester Zeit ausgebucht. Das HBZ ist außerdem die erste deutsche Einrichtung, die in Kooperation mit Kölner Bibliotheken Katalogdaten als Open Data veröffentlichte. Aus dem Südwestdeutschen Bibliotheksverbund planen die Universitätsbibliotheken Tübingen und Konstanz, in Kürze nachzuziehen. Zudem fand im Mai diesen Jahres im Rahmen des „Leipziger Semantic

Web Tag“ an der Universität Leipzig die Gründung des deutschen Zweiges der Open Knowledge Foundation statt (<http://okfn.de>).

Die Hochschulbibliothek Zwickau nahm die Aktualität des Themas zum Anlass, einen der Hauptakteure des Diskurses um das Semantic Web in Bibliotheken zu einem Vortrag einzuladen. Jakob Voß ist Entwickler in der Verbundzentrale des GBV in Göttingen und forscht seit seiner Zeit bei Wikimedia zu Metadaten und Semantic Web. Wikimedia setzt sich konsequent für freie Daten ein. So ist DBPedia einer der wichtigsten Knoten im Semantic Web. Außerdem ist es gelungen, in Kooperation mit der DNB die PND als Open Data zur Verfügung zu stellen.

Das Interesse an dem Vortrag war groß. Obwohl der Weg nach Zwickau für viele sehr weit war, kamen ca. 20 Teilnehmer persönlich. Per Online-Live-

stream schalteten sich nochmals ca. 35 Zuschauer zu. Außerdem ließ ein Professor des Studiengangs Informatik der Westsächsischen Hochschule den Stream in seiner Vorlesung mitlaufen, wodurch weitere 45 Studierende erreicht wurden. Die Folien des Vortrags lassen sich online bei Slideshare nachlesen (<http://www.slideshare.net/nichtich/semantic-web-und-linked-open-data-eine-bibliothekarische-geschichte>). Zudem wird in Kürze ein Videomitschnitt veröffentlicht (<http://biblog.fh-zwickau.de/2010/03/29/semantic-web-und-linked-data>).

Der Schwerpunkt des Vortrags lag nicht nur in einer Einführung in Semantic Web und Open Linked Data. Vielmehr zeigte Voß, dass Bibliotheken mit ihrem Wissen und ihren Kompetenzen in Datenstrukturierung und Datenspeicherung wichtige Akteure im Bereich des Semantic Web sind bzw. sein können. Zunächst ging er darauf ein, was diese Daten sind, die im Web als Open Data miteinander verlinkt werden, um dann zum eigentlichen Semantic Web überzuleiten.

#### Open Linked Data

Damit die innere Struktur des Semantic Webs begreifbar wird, muss vorab die grundsätzliche Bedeutung von Daten erläutert werden. Voß definiert Daten als „Strukturen aus unterscheidbaren Zeichen, die aufgrund von Vereinbarungen Informationen darstellen“. Daten sind somit als Träger von Informationen anzusehen, die jedoch erst verstanden werden können, wenn ihnen eine intersubjektive bzw. maschinenlesbare Konnotation zugrunde liegt. Über diese Bedeutung muss ein Konsens erzielt worden sein, d.h. sie muss in irgendeiner Form standardisiert sein. Während dies im allgemeinen Sprachgebrauch relativ unkompliziert ist, braucht es für die Datenverarbeitung mittels Computer klare Zuordnungen. Einzelne Daten – bspw. die Titeldaten eines Buches – werden als Metadaten strukturiert und zu Datensätzen zusammen gefasst. Im bibliothekarischen Bereich begegnen wir diesen Datenkonglomeraten vorrangig in Form von Titeldatensätzen. Für das Semantic Web ist die Möglichkeit der Zerschneidung dieser Sätze in möglichst kleine Teile (sogenannte Tripel) von höchster Wichtigkeit. Je kleiner und differenzierter diese Teile sind, desto mannigfaltiger lassen sie sich mit anderen Tripeln rekombinieren. Aus der bibliothekarischen Erschließung ist dieses Prinzip u.a. von Ranganathans Facettenklassifikation bekannt.

Bibliotheken sind nicht nur spezialisiert auf das Erfassen und Strukturieren von Daten, sie sind auch seit jeher verantwortlich für deren Speicherung. Durch die Speicherung von Daten werden sie einfacher kopierbar und rekombinierbar. Im Laufe der Geschichte von Bibliotheken bzw. des Bibliothekskatalogs wurden die Methoden zu Speicherung von strukturierten Daten immer weiter verfeinert. Von der Bibliographie in Buchform über den Bandkatalog und den Zettelkatalog hin zu relationalen

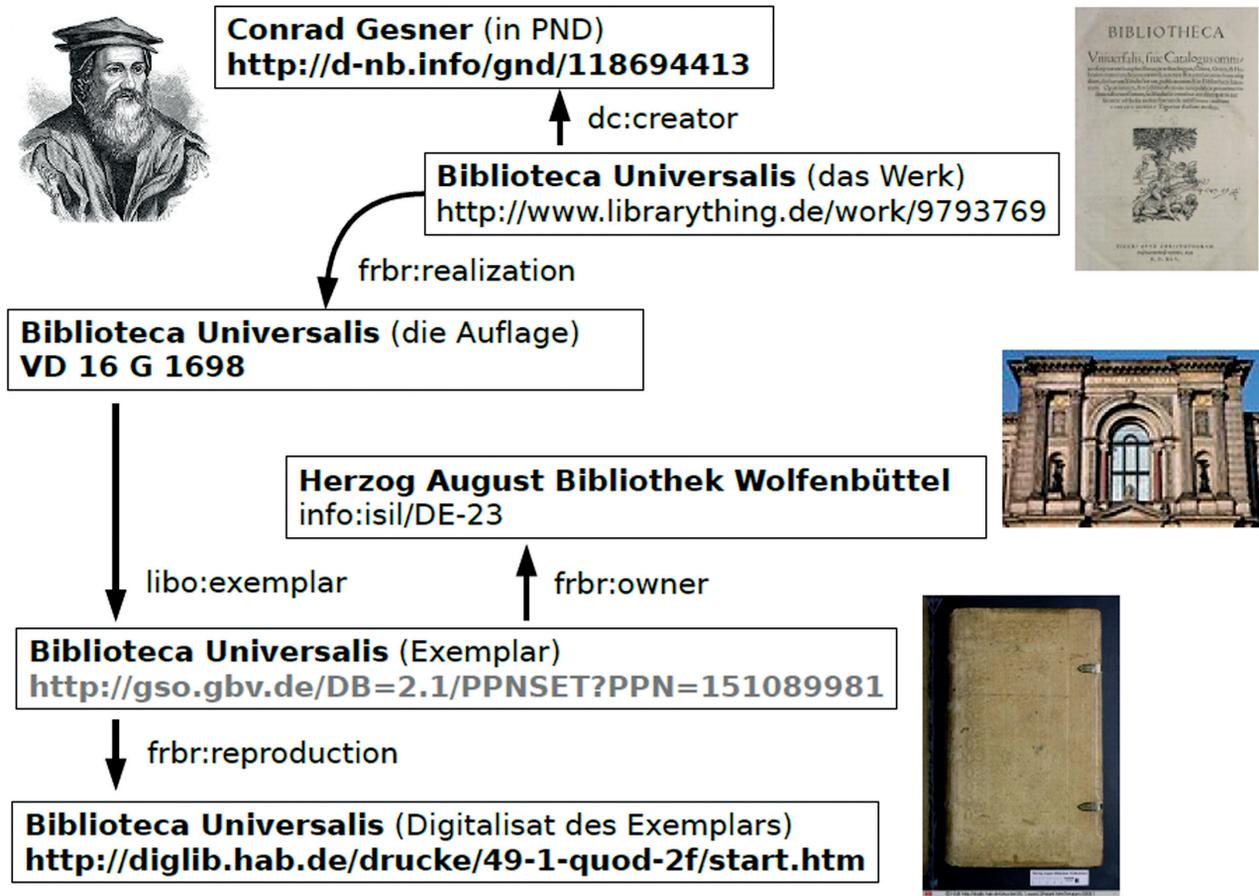
Datenbanken erweiterten sich sowohl die schon erwähnte Kopierbarkeit und Rekombinierbarkeit als auch die Sucheinstiege. Mit dem Einsatz von Computern kann heute über Random Access ein direkter Zugriff auf jedweden Punkt einer strukturierten Datenmenge erfolgen. Wichtig ist an dieser Stelle einerseits die Zergliederung in kleinste Dateneinheiten und gleichzeitig deren Zusammenführung in eine strukturierte Menge.

Für die standardisierte Struktur von Daten kann die Karteikarte als anschauliches Beispiel dienen. Über die hier vereinbarten Standards lassen sich Daten relativ problemlos miteinander austauschen. Für den einfachen Austausch der Daten gilt selbstverständlich, dass es nicht zu viele inkompatible Standards zu einem Bereich geben darf. Dies ist leider im BID-Bereich eher selten der Fall.

Direkte Verweise zwischen Daten wurden bereits mit der Karteikarte realisiert. Für das Web ist jedoch die Idee der bidirektionalen Verweisung von Daten maßgeblich. Ted Nelson entwickelte hierfür schon 1964 das Prinzip der Hyperlinks, welches von Tim Berners-Lee zum World Wide Web weitergeführt wurde. Der Aufbau des WWWs besteht zum einen in der einheitlichen Adressierung von Daten mittels URLs und zum anderen in deren einheitlichem Transport über das Protokoll HTTP.

Obwohl diese Standards mittlerweile einige Jahrzehnte alt sind, werden sie nach Voß von Bibliotheken dennoch bis heute nicht konsequent eingehalten bzw. teilweise völlig missverstanden. In der Mailingliste NGC4Lib fasst Tim Spalding, der Gründer von LibraryThing, die Situation in den meisten Bibliotheken wie folgt zusammen: „Before libraries join the ‚Semantic Web‘ [...] can they join the ‚Web Web‘? For example, it would be nice if I could link to a book in a library catalog, and if library catalogs were in Google.“ Bis auf wenige Ausnahmen (bspw. viele der auf Open Source beruhenden Kataloge wie der UB der TU Hamburg-Harburg, der Kölner KUG sowie der Schwedischen Nationalbibliothek) stellen heutige Bibliothekskataloge de facto „Datensilos“ dar. Das bedeutet, dass sie als Gatekeeper oder Portale agieren, hinter denen die eigentlichen Daten versteckt sind. Zudem sind die einzelnen Seiten des Katalogs lediglich innerhalb von Sessions abzurufen, wodurch keine stabilen URLs existieren. Ein einmal gefundener Titeldatensatz hat somit das nächste Mal eine andere Adresse und ist für den Nutzer als Link unbrauchbar. Für den Nutzer ist ein stabiler Link auf einen Titel von Interesse, damit er u.a. diesen in einer Mail oder per Chat an einen Kommilitonen weiter geben oder den Titel in seinem Literaturverwaltungsprogramm speichern kann. Voß kommt demnach zu dem Urteil, dass Bibliothekskataloge zwar im Web, aber nicht Teil des Webs sind.

Nur durch die Generierung von sogenannten Permalinks kann eine stabile Referenzierbarkeit im Web statt finden, so dass die Daten nicht mehr im Deep Web verschwinden, sondern als Teil des Webs



bestehen. Das Prinzip der stabilen Referenzierbarkeit liegt der Idee von „Linked Data“ zugrunde. Wenn die Daten über einen Permalink verfügen, haben sie einen eindeutigen Identifikator und können darüber mit anderen Daten verbunden werden. Voß betonte, dass für die Unveränderlichkeit der Permalinks – bspw. bei Umzug der Domain oder dem Wechsel auf andere Systeme – nicht die Technik, sondern die Menschen verantwortlich sind. Tim Berners-Lee, der neben dem Web auch für das Semantic Web als Erfinder gelten kann, sagte dazu bereits 1998: „What makes a cool URI? A cool URI is one which does not change. What sorts of URI change? URIs don't change: people change them.“

Im Semantic Web werden aus Linked Data – so Jakob Voß – „zweckdienliche Informationen“. Diesen Informationen liegt wiederum der Standard RDF zugrunde. Hierüber lassen sich Tripel von Objekten bilden z.B. „Romeo und Julia“, „wurde geschrieben von“, „Shakespeare“. Jedes Objekt in diesem Tripel ist eindeutig identifizierbar und referenzierbar. Dies bedeutet – und damit schließt Voß den Kreis – dass die eigentlichen Daten als kleinste Teile im Semantic Web in standardisierter Form vielfältig miteinander verknüpft, kopiert und rekombiniert werden können.

Bibliotheken verfügen über eine enorme Menge an Daten, worüber sie ein mächtiger Teil des (Seman-

tic) Web sein könnten. Angesichts der derzeitigen personellen und finanziellen Lage von Bibliotheken stellte Voß jedoch heraus, dass die Formatierung der Daten in RDF zunächst im Hintergrund stehe. Wichtiger sei es, die Daten überhaupt erst einmal zu publizieren und zwar als Open Data (<http://opendata-network.org>). Denn nur über die freie Verfügbarkeit der Daten können diese von anderen weiter verarbeitet und genutzt werden. Wie dies konkret aussehen kann, zeigten zuletzt wie oben erwähnt das HBZ oder die ungarische Nationalbibliothek, die ihre Katalogdaten als Open Data veröffentlichten. Einen aktuellen Überblick zu Linked Open Data Projekten liefert das World Wide Web Consortium (W3C) unter <http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>.

Im Anschluss an den Vortrag entwickelte sich eine lebhafte Diskussion, in der u. a. die Frage des Copyrights gestellt wurde. Voß erläuterte, wie dies bei Open Linked Data durch freie Lizenzen und Verfahren wie Creative Commons Zero geregelt ist. Zusammenfassend zeigte die Veranstaltung an der Westsächsischen Hochschule Zwickau anschaulich, dass Semantic Web und Linked Data Zukunftsfelder sind, auf die Bibliotheken unmöglich verzichten können.

