

SWIB13 – DIE FÜNFTE »SEMANTIC WEB IN LIBRARIES CONFERENCE« BEFASST SICH MIT NÜTZLICHKEIT, NUTZBARKEIT UND NUTZUNG DES SEMANTIC WEB FÜR BIBLIOTHEKEN

Zum fünften Mal richteten das Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz) und die Deutsche Zentralbibliothek für Wirtschaftswissenschaften – Leibniz-Informationszentrum Wirtschaft (ZBW) die internationale Fachtagung zum Thema »Semantic Web in Bibliotheken« (SWIB) aus. Die SWIB13 fand vom 25. bis 27. November 2013 in Hamburg statt. Es nahmen 180 Menschen aus 23 Ländern (Europa, Amerika, Asien, Afrika) sowie eine Katze (Finnland) teil. Zusätzlich wurde die SWIB13 per Live-Video-Stream angeboten und international verfolgt.

Ziel der SWIB ist, substantiell über Entwicklungen im Umfeld von »Linked Open Data« (LOD) in der Bibliothekswelt zu informieren, den Austausch von Erfahrungen, Werkzeugen und Ideen zu fördern und zum Querdenken zu ermutigen. Die SWIB bringt Neulinge und »alte Hasen« zusammen, kombiniert die Bereiche Anwendung, IT-Entwicklung und Forschung, verbindet Bibliotheks-, Museums- und Archivmitarbeiter und -mitarbeiterinnen. Entsprechend bieten die Workshops am ersten Tag eine Einführung in LOD, einen Austausch für Doktoranden sowie weitere themenspezifische Tutorien. Am zweiten und dritten Tag umfasst das Programm Keynotes und Vorträge sowie spontane Lightning Talks. Die Gemeinschaft wird durch ausgedehnte Pausen, gute Verpflegung vor Ort und ein gemeinsames Abendessen gefördert. Kommuniziert wird auch via Twitter (Hashtag #swib13). Die Konferenzsprache ist Englisch. Programme, Videomitschnitte und Folien sind über die Website verfügbar (<http://swib.org/swib13/programme.php>).

Im Folgenden fokussiert dieser Bericht, inwiefern Werkzeuge, Datenquellen und Beschreibungsformen des semantischen Web für Bibliothekare, Webentwicklerinnen und die Internetöffentlichkeit operativ nützlich und nutzbar sind – und genutzt werden, auch ohne »Semantic Web«-Expertise.

Das Semantic Web sollte die Arbeit von Bibliothekarinnen und Bibliothekaren erleichtern
Gleich zu Beginn der Hauptkonferenz rückte **Dorothea Salo** (Universität Wisconsin-Madison) mit einer stimulierenden Keynote die täglichen Aufgaben von Bibliothekarinnen und Bibliothekaren in den Mittelpunkt. Die Technologien des Semantic Web müssten fokussiert bibliothekarische Arbeit unterstützen – Ontologien allein seien irrelevant. Auch käme es nicht darauf

an, was man dank IT-Fachwissen mit den Technologien bewerkstelligen könne, sondern darauf, was damit für Bibliothekarinnen und Bibliothekare in ihrem Arbeitsalltag möglich sei. Ferner plädierte sie dafür, LOD-Strategien nicht nur auf große Einrichtungen, sondern insbesondere auch auf kleine, mit wenig Personal und IT ausgestattete Bibliotheken auszurichten.

Eine typische bibliothekarische Aufgabe will die Universitätsbibliothek Mannheim mit ihrem Mashup für den Medienerwerb unterstützen. Als Entscheidungshilfe für Fachreferenten ermittelt das Tool aus LOD-Quellen zu bestimmten DDC-Stellen (Dewey Decimal Classification) Publikationen und stellt dann für die einzelnen Publikationen relevante Angaben zusammen wie z. B. RVK-Notation (Regensburger Verbundklassifikation), DDC-Notation, Schlagwörter oder Bibliotheken mit der Publikation im Bestand. Startpunkt ist dabei z. B. die deutsche oder britische Nationalbibliografie. Der Prototyp »OpenCat« der französischen Nationalbibliothek (BnF) erlaubt lokalen Bibliotheken, mit Technologien des Semantic Web und ohne Anpassung ihres Bibliothekssystems einen Katalog im Web zu veröffentlichen, der ihre lokalen Daten (Signatur, Verfügbarkeit, Schlagwörter) mit BnF-Daten (Werk-bezogene Daten, Links zu Digitalisaten) und externen LOD-Quellen (Snippets, Links zu Konferenzseiten, Biografien, virtuellen Ausstellungen, Bildern) ergänzt.

Die Pflege von Normdaten ist für die bibliothekarische Arbeit fundamental. Auch das vom Global Forum on Agricultural Research (GFAR) geleitete Projekt AgriVIVO produziert und pflegt Normdaten über Personen, Organisationen und Projekte in der agrarwissenschaftlichen Forschung. Viele dieser Akteure seien in VIAF (Virtual International Authority File) noch nicht

LOD-Strategien sollten auch kleine Bibliotheken einbeziehen

Normdaten via LOD-Cloud anreichern und nachnutzen



Abb.: Internationales Publikum im Bürgerhaus Wilhelmsburg in Hamburg

Foto: ZBW

enthalten, wie **Valeria Pesce** (GFAR) konstatierte. Das Sheet Music Consortium ermittelt aus Notendruckern Verleger – die, so **Stephen Davison**, traditionell nicht in Normdateien enthalten seien – und veröffentlicht sie als LOD, ergänzt aber auch VIAF. Institutionelle Repositorien seien historisch gesehen »Datensilos«, meinte **Vitali Peil** (Universitätsbibliothek Bielefeld). Die Universitätsbibliothek Bielefeld publiziere nun als Teil ihres Repositoriums »PUB«-Normdaten von Personen, Organisationen, Abteilungen, Projekten, Auszeichnungen, später kämen Verknüpfungen zur GND (Gemeinsame Normdatei), zu ORCID (Open Researcher and Contributor ID) etc. hinzu. Im Rahmen des langfristigen Projekts »Germania Sacra« werden aus Büchern Normdaten zu Personen, Klöstern, Orten und Zeiträumen extrahiert und die GND um weitere Einträge, Angaben und Verknüpfungen ergänzt, berichtete ein Lightning Talk. Sollte die schwedische Nationalbibliothek, fragte umgekehrt **Martin Malmsten** (schwedische Nationalbibliothek), selber deutsche Autoren erfassen? Lieber fokussiere sie bei der Erschließung den eigenen Kern und konsumiere ansonsten verfügbare LOD.

Der Wirkungsgrad bibliothekarischer Arbeit verstärkt sich, wenn Medien über Thesaurus-Mappings Schlagwörter aus weiteren Klassifikationssystemen zugeordnet werden. An der Hochschule der Medien Stuttgart werden für automatische Mappings zwischen Klassifikationssystemen zunächst Exemplare, Ausgaben und Auflagen zu Werken gruppiert, um einen höheren Anteil an Mehrfachindexierungen als Ausgangspunkt zu erhalten. Im Rahmen der Erkennung von z. B. Personen, Orten oder Themen in textuellen Inhalten von Videos bildet das Hasso Plattner Institut GND-Schlagwörter und DBpedia mehrsprachig aufeinander ab, um die Stärken der beiden Datenquellen zu kombinieren.

Die komfortable Nutzbarkeit von LOD-Quellen lässt noch Wünsche offen

Der tatsächlichen Nachnutzung verfügbarer LOD-Quellen stünden noch Mängel vor allem bei ihrer Dokumentation entgegen, berichtete **Jose Manuel Barreiro Cruz** (Universität Valencia). So zeigte sich bei der Betrachtung von Kandidaten für LOD-Quellen zur Anreicherung mittelalterlicher Manuskriptdigitalisate, dass viele Datensets nicht registriert und kaum mit Metadaten beschrieben seien sowie keine Lizenz ausgewiesen. Auch wären die Datensets nur spärlich untereinander verknüpft.

Bezüglich des technischen Zugriffs auf LOD-Quellen befand **Fabian Steeg** (hbz), man solle nicht Semantic-Web-Experte sein müssen, um LOD zu nutzen. Da-

her stelle das hbz Titel- und Normdaten als LOD nicht nur über die im semantischen Web verbreiteten Wege SPARQL-Schnittstelle (SPARQL Protocol and RDF Query Language) und einen Vollabzug der Daten (RDF-Dump) bereit, sondern auch über eine Webentwicklern geläufigere Web-API. Trotz der grundlegenden Verteiltheit von Linked Data, so **Simeon Warner** (Cornell Universitätsbibliothek), seien lokale Kopien oft nötig. Das Sitemaps-basierte Framework »ResourceSync« böte hierzu Möglichkeiten, von einer Datenquelle ausgewählte Ressourcen zu beziehen und bei Änderungen synchronisiert zu halten.

Verschiedentlich gäben zwar lokale Systeme einen Überblick über Änderungen in Thesauri, einen Standardweg gebe es aber noch nicht. Dem soll das Projekt »skos history« abhelfen, das **Joachim Neubert** (ZBW) in einem Lightning Talk vorstellte. Bei der Publikation der Inhalte in ihrem institutionellen Repositorium PUB in Form von LOD intendiert die Universitätsbibliothek Bielefeld, die Veränderungen an den Metadaten anzuzeigen. Mit URIs (Uniform Resource Identifier) referenzierbare Versionen von LOD-Metadaten werden für »edoweb«, den Archivserver des Landesbibliothekszentrums Rheinland-Pfalz und ebenfalls Thema eines Lightning Talks, angestrebt.

Linked Open Data kann auch für Webnutzerinnen und -nutzer Gewinn bringen

Nach **Richard Wallis** (OCLC) exponierten Bibliotheken ihre Bestände nicht gut im Web. Dass sich die Sichtbarkeit von Bibliotheken immens erhöht, wenn die Bestände nicht im sogenannten »Deep Web« des Katalogs verborgen bleiben, sondern ihre Auffindbarkeit durch Suchmaschinen erleichtert wird, hatte die BnF bereits bei der SWIB12 gezeigt. Ihr Prototyp »OpenCat« erlaubt nun bei einer Suche in einem Verbundkatalog die Lokalisierung der gefundenen Bibliotheksressourcen auf einer Landkarte. Um Nutzerinnen und Nutzer detaillierter über Bestände informieren zu können, wird innerhalb der DINI-AG KIM eine (Mikro-)Ontologie für Bestandsinformationen modelliert. Wie Informationen aus Europeana von Suchmaschinen präsentiert, aber auch wie sie in sozialen Netzwerken geteilt werden, wird seitens Europeana beeinflusst. Zum Einsatz kommen dabei Technologien wie schema.org, Facebook Open Graph und Twitter Cards.

Die Veröffentlichung von LOD in einer für Nutzer verständlichen, ansprechenden Weise ist noch eine Herausforderung. Das Content Management System »Drupal«, Inhalt eines Workshops, bietet jedoch die Anreicherung von Webseiten mit der Technologie »RDF in Attributes« (RDFa). RDFa-Angaben gäbe Drupal als HTML-Attribute wieder, die Webseiten sind dann also

gleichermaßen Teil des »Web of Data«, für Suchmaschinen findbar und für Menschen lesbar.

Linked Open Data als Beschreibungsform

entfaltet auch intern mehr Potential

Formate seien oft im Weg, bei Formaten beschäftige man sich mehr mit Formalien, statt Dinge zu beschreiben, und während das Katalogisierungsformat MARC (Machine-Readable Cataloging) Zeichenketten benenne, referenziere RDF mit URIs Entitäten, argumentierte Martin Malmsten in seiner Keynote. Für ihr neues Bibliothekssystem, das sie selbst von Grund auf neu entwickeln, verwende die schwedische Nationalbibliothek daher LOD als primäre Technologie und »halte Monster und Formate draußen«. Ähnlich attestierte Salo, die den Verdross kleinerer Bibliotheken über die aufwändige Katalogisierung kennt, MARC eine zu hohe Komplexität, und identifizierte hier Potential für LOD als Basis alternativer Beschreibungsformen. Bei AgriVIVO, einem Suchportal zur Förderung der wissenschaftlichen Kollaboration in der Landwirtschaft, nimmt LOD ebenfalls eine zentrale Rolle ein. AgriVIVO's Basis VIVO sei vollständig Ontologie-getrieben – sofort nach dem Design bzw. einer eventuellen Anpassung der Ontologie und ohne weitere Programmierung ließe sich VIVO entsprechend der Ontologie mit Daten befüllen, und es stünde ein Nutzerinterface zur Verfügung, so Pesce. Auch edoweb nutzt RDF als internes Datenmodell, und die NTNU Universitätsbibliothek Trondheim befasst sich mit der Katalogisierung in RDF, eine Konversion entfällt somit, wie zwei Lightning Talks offenbarten. Die nationale Gestaltung und Verwendung von Ontologien in Bibliotheken, Museen, Archiven, Regierungsinstituten etc. fördert und vereinheitlicht die finnische Nationalbibliothek im Projekt »ONKI«. Aus LOD als Beschreibungsform ergeben sich dann auch weitere Vorteile, wie z. B. die Erkennung von Fehlern in der Erschließung bei der schwedischen Nationalbibliothek. Einige Datenkonsumenten gehen inzwischen dazu über, LOD-Quellen nicht zu kopieren, sondern zu verknüpfen und auf Änderungen zu achten. Diesen Weg gehen z. B. die schwedische Nationalbibliothek und die Universität Valencia, deren Bookviewer die aktuell betrachteten Medien mit Angaben aus LOD-Quellen ergänzt.

Die SWIB war auch 2013 wieder informativ und kommunikativ, anregend und vergnüglich, und die Teilnehmerinnen und Teilnehmer zeigten sich insgesamt sehr zufrieden. Die Bemühungen, die SWIB als Plattform für den (internationalen) Austausch der bibliothekarischen LOD-Community aufzubauen, scheinen Erfolge zu zeitigen. So nutzten mehrere Vorhaben Ergebnisse anderer Körperschaften, die bereits früher

auf der SWIB präsentiert worden waren. Auch bezeichnete Salo in einem Konferenzbericht im »Library Journal« die SWIB als »the premiere conference for semantic web technologies in libraries«.

Die SWIB14 wird, in freundlicher Kooperation mit der Bibliothek der Friedrich-Ebert-Stiftung, vom 1. bis 3. Dezember 2014 in Bonn stattfinden.

SWIB14 Anfang Dezember in Bonn

DIE VERFASSER

Dr. Timo Borst, Leiter der Abteilung Innovative Informationssysteme und Publikationstechnologien, ZBW – Leibniz-Informationszentrum Wirtschaft (Kiel/Hamburg), Tel.: 0431 – 8814-312, E-Mail: t.borst@zbw.eu

Aenne Löhden, Abteilung Innovative Informationssysteme und Publikationstechnologien, ZBW – Leibniz-Informationszentrum Wirtschaft (Kiel/Hamburg), Tel.: 040 – 42843-453, E-Mail: a.loehden@zbw.eu

Adrian Pohl, Programmbereich Linked Open Data, Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz), Tel.: 0221 – 40075-235, E-Mail: pohl@hbz-nrw.de

»Linked Data or die« (Martin Malmsten)

Archiv **DIGITALISIERUNG IM ARCHIV – NEUE WEGE DER BEREITSTELLUNG DES ARCHIVGUTS. 18. ARCHIVWISSENS- SCHAFTLICHES KOLLOQUIUM AM 26. UND 27. NOVEMBER 2013 IN MARBURG**

Mit über 160 Teilnehmern war das von der Archivschule Marburg ausgerichtete 18. Archivwissenschaftliche Kolloquium am 26. und 27. November 2013 in Marburg gut besucht. In diesem Jahr stand es unter dem Motto »Digitalisierung im Archiv – Neue Wege der Bereitstellung des Archivguts« und fand vor dem Hintergrund des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projekts »Produktivpiloten Digitalisierung von archivalischen Quellen« statt, in dem aktuell eine Road Map und Workflows für eine flächendeckende Digitalisierung und Bereitstellung deutschen Archivguts entwickelt wird. In dem Kolloquium nun wurde ein Blick über den Produktivpiloten

»Produktivpiloten Digitalisierung von archivalischen Quellen«