



How do Libraries Find their Way onto the Semantic Web?

*Conference Report from the SWIB09
(Semantic Web in Libraries '09) in Cologne*

Dr Timo Borst, Birgit Fingerle, Joachim Neubert

ZBW — German National Library of Economics, Leibniz Information
Centre for Economics, Düsternbrooker Weg 120, 24105 Kiel, Germany
t.borst@zbw.de

Anette Seiler

Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen (hbz),
seiler@hbz-nrw.de

The conference programme, abstracts and lectures can be found online at
<http://www.swib09.de> (predominantly in German).

Abstract

The First Semantic Web/Linked Data Conference of German libraries took place in Cologne on 24 and 25 November 2009 and revealed much interest in advancing from a web of documents to a web of data. The contributions covered advanced applications at the Library of Congress and the Swedish National library as well as advocacy for linked data approaches and the use of semantic web tools in the library and cultural heritage domain. The German National Library announced the publication of their authority files as linked open data, with first prototypes to become available in mid-2010, and this constituted a highlight of the conference. Discussions about ontologies to be used, application examples and a summary of practical experiences completed the event.

Key Words: semantic web; linked data; authority files

The event ‘Semantic Web in Libraries — SWIB09’ was conceived as a ‘conference for innovative librarians’ and took place for the first time on 24 and 25 November 2009 in Cologne. Aiming to familiarise libraries with the semantic web, the first day offered an introduction and several practical examples illustrating the potential of semantic web applications. The second day revolved around practical examples demonstrating the challenges and possible solutions for libraries. The conference was hosted by the University Library Centre of North Rhine-Westphalia (hbz) and the German National Library of Economics (ZBW) and attracted international speakers and participants. For capacity reasons, the number of participants was limited to one hundred on the first and to sixty on the second day. A significantly higher number of registrations showed the great need for further exchanges on this topic.

What is the Semantic Web?

Jakob Voss, software developer at the Head Office of the Common Library Network (VZG/GBV), opened the conference with a brief presentation on the semantic web. With reference to conventional bibliographic catalogue entries, he outlined the change from a web of linked documents to a web of linked data that characterises the semantic web. Various institutions in the academic environment already publish their data as linked data. Voss pointed out that currently libraries do not use all the opportunities offered by the web, hiding the contents of their databases in the ‘deep web’. A lack of stable URLs often hindered linking to an entry in a library catalogue by traditional means.

Linked data, on the other hand, offer new opportunities for the division of labour, said Voss. It will be possible for libraries to link bibliographic data of a publication provided by its publisher to other data, such as the name authority file (PND) or the exact location in the library. In this way data published by other persons or libraries about a particular publication can be re-used. The library is able to present selected information to its users on a website customised for their needs. The benefits of such scenarios of use for linked data are obvious. Because the web itself is the database, the need for specialist software and data conversions will diminish. The laborious collection of data in proprietary databases and duplicating efforts in similar institutions will become redundant, because the relevant information will be

gathered from various sources. The library will publish only those data itself that are not yet available elsewhere.

Why Should Libraries Engage in the Semantic Web?

According to Stefan Gradmann, Professor of Knowledge Management at the Humboldt University in Berlin (HU Berlin), libraries should engage in the semantic web in order to remain visible at all in the future knowledge society. In addition, libraries could develop attractive and innovative services for their users, such as improved search functionalities, by building on semantic web technologies. Libraries in particular could also make specific contributions to the semantic web. In this context Gradmann reminded the audience that Tim Berners-Lee, the inventor of the web and a pioneer of the semantic web, had named 'Catalogs on the web' a killer application of the semantic web as far back as 2000. In contrast to many other database providers, libraries possess reliable metadata and context data (such as name or corporate body authority files, classifications and thesauri). These are highly attractive for linking with other data. For the transfer onto the semantic web a suitable standard is already available in SKOS which was developed specifically for classifications and thesauri. For other aspects insufficiently considered until now, such as how to deal with the change of meaning or the long-term availability of information and knowledge on the semantic web, libraries could contribute their experience.

In agreement with Gradmann, Dr. Bernhard Haslhofer, research assistant at the Department of Distributed and Multimedia Systems of Wien University, saw great potential in the incorporation of classical mechanisms of information organisation by libraries (identifiers, metadata, controlled vocabularies) as 'open linked data' into the semantic web. 'Linked data' is an attempt to continue these mechanisms on the web. Libraries might even be able to contribute a building-block to a free and community-driven alternative to commercial applications such as Google books. A necessary prerequisite for this would be that libraries no longer keep their data in closed data silos, but provide them freely on the internet under an appropriate licence.

In his lecture Jürgen Kett (German National Library, DNB) asserted that universities, libraries, museums and archives are integral parts of the semantic web, as they give it the necessary reliability and stability.

What is Happening in Other Libraries?

Several contributions showed that linked data now provide a suitable technology for turning the semantic web into a reality for libraries. Ed Summers, software developer at the Library of Congress (LoC) and responsible for the LoC Subject Headings, which represent the largest linked data stocks in libraries so far, referred to the many linkage options and applications of linked data. Headings from different vocabularies could be correlated to provide the basis for better search options. Press clippings from a project such as 'Chronicling America: Historic American Newspapers' could be linked to pictures on 'Flickr' and geographic coordinates, thus illustrating and locating the event.

Fig. 1: Ed Summers of the Library of Congress.



Anders Söderbäck from the Swedish National Library showed how a classical library catalogue could be transferred to linked open data and subsequently to a modern web application and demonstrated what part semantic web technologies could play in this process. Felix Ostrowski, IT developer at the hbz, delineated a model of deep integration of ontologies directly into the programming language, using repository software as his example for 'linked applications'. André Hagenbruch of the University Library Bochum presented a project for a linked data-based library portal intended to integrate not only bibliographic data, but also administrative information such as personnel data or opening hours.

Jürgen Kett offered interesting insights into the linked data strategy of the German National Library (DNB), which includes the publication of the name authority file (PND) for re-use free of charge. In another lecture, Kett and his colleague Dr. Lars G. Svensson provided details of their current linked data project. In the first phase, URI and SPARQL access (SPARQL Protocol and RDF Query) will be allowed to the SWD (subject headings authority file) and PND (name authority file) datasets including their links to international authority files. A beta version is scheduled to go online mid-2010. Dr. Timo Borst, head of IT development at the German National Library of Economics (ZBW) picked up on this and demonstrated the benefit of the DNB's planned linked data service by means of the integration of person authority files (PND) into the ZBW's document server EconStor. An experimental SPARQL endpoint of the DNB served as technical access point for querying PND data as linked data. An autosuggest web service for PND — built at this endpoint — enabled a lightweight enhancement of the document server application and allowed thus for a more controlled indexing of authors' data. In an analogous move, integrating the standard thesaurus for economics (STW) as linked data into EconStor was intended to support controlled indexing with subject headings and improved search functionalities through search term extension.

What Can Libraries Learn from Other Domains?

Karin Teichmann, Head of the Graphic Collection at the German Book and Writing Museum of the DNB (German National Library), presented the 'CIDOC Conceptual Reference Model' as an ontology developed in the cultural heritage domain. It is an ISO standard and allows the merging of differently structured and distributed information by means of an overarching meta-model, thereby solving integration problems; solving such problems is a frequent requirement in libraries, too.

Professor Gradmann and Marlies Olensky, researcher at the Humboldt University in Berlin, presented the semantic data layer of *Europeana* which uses the W3C standard SKOS and linked data to support user requests, browsing and results presentation.

Anja Jentsch, researcher at the Freie Universität (Free University) Berlin, gave a report on the extensive linked data activities within DBpedia, a semantic web version of Wikipedia. Apart from the infrastructure for linked data

she demonstrated the integration of DBPedia data into browser, search and portal interfaces, calling it an important node for linked data because of the great number of data and links to other linked data sets.

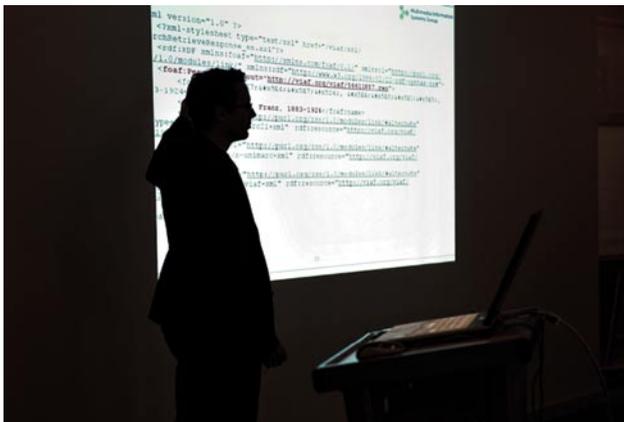
Elena Semanova, researcher and independent ontology expert, offered insight into another pillar of the semantic web: ontologies, their connection with natural speech and their possible uses.

What Could be the First Steps Towards the Semantic Web?

Anders Söderbäck said that it could be very useful for linked data projects to assume a different perspective and regard the library's catalogue as a network. A fundamental step in the direction of the semantic web would then be the legally safeguarded publication of their data as linked data. Joachim Neubert, IT developer at the ZBW (German National Library of Economics), summarised practical experiences from the linked data projects of the ZBW and gave precise tips on how to start on the semantic web. He recommended picking a manageable field and dataset. Licensing and attribution problems should be addressed early. Data storage and maintenance as well as the library's work routines should remain unaltered as far as possible.

As Jakob Voss described in his opening lecture, the use of stable HTTP URIs (URLs) for catalogue entries and authority files is a basic rule that libraries

Fig. 2:



should adhere to when publishing linked data. According to the rules for linked data as laid down by Tim Berners-Lee, phrasing an offer of relevant information to these URIs in the semantic web standards RDF (Resource Description Framework) or SPARQL should be the next step. For this, ontologies could be used describing what kinds of objects there are, their attributes and the relations between them. The last step would be linkage to external URIs in order to open up new resources.

When publishing linked data, libraries should apply open standards already available and help to develop these further, instead of relying on in-house developments. Several contributors stressed the importance of this rule. Gradmann mentioned that *Europeana* switched from a planned in-house development to the re-use of already existing standards. For ontologies, several quasi-standards are available in SKOS, Dublin Core, FRBR, FOAF (Friend of a friend, an ontology for the description of persons and social networks) or OAI-ORE (Open Archives Initiative — Object Reuse and Exchange). The 'Bibliographic Ontology' (bibo) presented by Jakob Voss in another lecture could develop into another standard. Presently available as version 1.3, this ontology is based on existing vocabularies such as Dublin Core or FOAF and is in continuous development through a community of currently around 150 people. It can be used to describe types of documents, compilations and events, as well as the publication status. Since at present there is no other ontology for bibliographic data in RDF as extensive, well documented and openly discussed, there was no alternative to 'bibo' for transferring bibliographic data onto the semantic web, according to Voss. In this context he asked a provocative question: will 'bibo' supersede bibliographic data formats? Svensson contended in his lecture that 'bibo' concentrated too much on citation management and therefore put RDA (Resource Description and Access) at the centre of the DNB's planning. His conclusion was that there is no optimal ontology for library data.

Neubert cited a number of practical tips, tools and tutorials for the creation of linked data. He made recommendations on how to guarantee the persistence of URIs and how to design them practically. He also reported his positive experience with mailing lists, where the linked data community could be invited to give feedback on one's own applications.

Finally, libraries should take great care to publish their data with the least possible restrictions imposed by licences in order to ensure widest

re-usability. This point was made most emphatically by Patrick Dankowski, emerging Technologies Librarian at CERN, in his passionate plea for 'free data': linked data can only unfold the whole power of their network effects if the data are published under a free licence. Danowski recommended posting them as 'public domain' data.

What Next?

During the course of SWIB09 it became evident that many questions and problems remain. Some of these were summarised by Söderbäck: The present library system and the prevalent bibliographic mindset are both just as unsuited to linked data as the current legal system. But Söderbäck also entertained hopes that 'All this will change ...'.

Jürgen Kett moderated a final discussion round gathering ideas on 'What next?'. One suggestion among others was to install a communication platform within the framework of the semantic web activities of the W3C in order to continue the exchange of information and ideas on linked data projects. Many also expressed a need for a repeat conference in 2010.