Publishing Aleph data as Linked Open Data

IGeLU, Haifa 12.9.2011

Dr. Silke Schomburg
Agenda

1. Background
2. Four Steps to Publishing Linked Open Data
3. What's Ex Libris supposed to do?
1. Background
Motivation

➢ Use of international and cross-domain **web standards** leads to
  ➢ Increased **discoverability**
  ➢ **Multiple usability** of one dataset
  ➢ Maximized **interoperability & re-usability**
  ➢ **Flexible** data model
Expected ROI

Synergy effects through standardization of technology stack:

- Quantitative: saved resources
- Qualitative: improved services
A sustainable and consequent strategy of publishing and linking data on the web requires the data to be open*.

* [http://www.opendefinition.org/](http://www.opendefinition.org/)
2. Four Steps to Publishing Linked Open Data
Step 1: Find cooperating libraries

- hbz provides technology for union catalog
- Data is created cooperatively
- hbz didn`t publish all data on its own because of legal uncertainties
- Four union libraries joined the initiative
Step 2: Export Data

➔ Using Aleph Publishing module
  ➔ Concept: bibl. records are enriched with authority & holdings Data and written to Oracle table Z00P (XML structure)

➔ Problems:
  ➔ Performance issues with incremental updates
    ➔ disproportion between incoming/outcoming data updates
    ➔ workaround: irregular updates as full dumps
Raw Data Now!

http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html
Step 3: Publishing raw data

- Load to **HTTP server**
- **Describe** published datasets
- Attach an **open license**
  (for recommended open data licenses see [http://www.opendefinition.org/licenses/#Data](http://www.opendefinition.org/licenses/#Data))
- **Register** datasets at CKAN.net
- Problem with serving consistent, standardized MAB-coded data
On this wiki page you’ll find up-to-date information about available Open Data exports from the hbz union catalog. If you have any questions please contact serweb[at]hbz-nrw[dot]de or post a comment on this page. (You’ll have to login/register first.)

Since April 2011 there is a changelog. Information about older data exports from March 2010 can be found here.

There is also a download statistic site.

All data which is available for download here is published under a Creative Commons CC0 licence. The data thus is in the Public Domain which means it belongs to everybody and can be used without restrictions for any purpose.

**Recent Open Data publications**

**Linked-Open-hbz-Data**

<table>
<thead>
<tr>
<th>Name</th>
<th>Open hbz data as Linked Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Publisher</strong></td>
<td>hbz</td>
</tr>
<tr>
<td><strong>Licence</strong></td>
<td><a href="https://creativecommons.org/publicdomain/zero/1.0/">CC0</a></td>
</tr>
<tr>
<td><strong>Export Date</strong></td>
<td>2011-09-01</td>
</tr>
<tr>
<td><strong>Publication Date</strong></td>
<td>2011-09-06</td>
</tr>
<tr>
<td><strong>Format</strong></td>
<td>text/NTriples</td>
</tr>
<tr>
<td><strong>Format documentation</strong></td>
<td>See <a href="https://github.com/hbz/linked-open-hbz-data">here</a></td>
</tr>
<tr>
<td><strong>Extent</strong></td>
<td>544 MB / 158 M triples</td>
</tr>
<tr>
<td><strong>Checksum</strong></td>
<td>f9b74e5f4f456e253af793af017</td>
</tr>
<tr>
<td><strong>Notes</strong></td>
<td>This archive contains RDF triples of all records which have been published as Open Data.</td>
</tr>
</tbody>
</table>

**Duisburg-Essen University Library Catalog, Campus Essen**

| Name                      | hbz records with Dartmud University Library holdings |

Source: [https://wiki1.hbz-nrw.de/display/SEM/Recently+published+Open+Data+exports](https://wiki1.hbz-nrw.de/display/SEM/Recently+published+Open+Data+exports)
Step 4: Identifiers & data conversion

- **Choose** suitable RDF **vocabulary** for representing bibliographic data: Bibo
- **Map** and **convert raw data** to RDF
- **Mint HTTP URIs** for described resources: [http://lobid.org/resource/{internalID}](http://lobid.org/resource/{internalID})
- **Publish RDF**
  - Triple store (4store) with SPARQL interface
  - Pubby framework for HTML representation
http://semantic.ckan.net/group/?group=http://ckan.net/group/lld, as of 2011/09/06
3. What is Ex Libris supposed to do to promote Linked Open Library Data?
Provide open interfaces for legacy and future systems

- **Provide streaming interface** (event driven => real time) for all library data
- **Use open standards** (like XMPP) for interfaces
Enable Open Data

- Don‘t use ‚open‘ as marketing term but in the clear sense of open definition

- Alma in general:
  - Provide possibility to **export all data** (e.g. statistics, user data) **in open formats**

- Alma‘s community zone:
  - **Openly licensed & accessible** as a whole **by everyone**
SWIB11

“Semantic Web in Libraries" (SWIB) hosted by hbz and ZBW – Leibniz Information Centre for Economics

Hamburg, 28. - 30.11.2011

Web stream will be available

http://swib.org/
Thanks.

Questions?

With pleasure now/here or also later to semweb@hbz-nrw.de