

Information—Integration—Intelligence Solutions



Semantic web technology and the publishing world: what can they offer each other?

Bob DuCharme
2010 Semantic Web Technology Conference
June 23, 2010



Introductions

- Presentation and all its URLs:
<http://snee.com/semtech/2010>
- Me: Solutions Architect at TopQuadrant;
formerly XML, SGML guy at RIA, Moody's,
LexisNexis, Innodata Isogen
- Weblog: <http://www.snee.com/bobdc.blog>



Publishing “world”

- Book publishers
- Magazine publishers
- Print, web, CD/DVD-ROM
- Libraries
- Library and publisher service companies
- Aerospace, auto manufacturers, mobile phone companies...



The issues

- Print revenue declining
- “Semantic” becoming a marketable buzzword
- Blind men, elephant, etc.



Two use case classes

- Customer Sam Smith wants to see Italian Christmas dessert recipes



Two use case classes

- Customer Sam Smith wants to see Italian Christmas dessert recipes
- Publishing staff editor Jane Jones wants to publish a hardcover book of Italian Christmas dessert recipes



Sample extra issues for Jane

- Images with suitable resolution for print publication?
- Images and recipes that they're licensed to use in a print book?



CMS vendor says...

- Flexible metadata!
- Easier searching!





Problems with vendor proposal

- Limited metadata flexibility
- If they support any standards, they won't support overlapping ones



Image resolution?

- Japan Electronic Industries Development Association **EXIF** X Resolution, Y Resolution, Resolution Unit
- Adobe's **XMP** XResolution, YResolution, ResolutionUnit
- Library of Congress **MIX** (Metadata for Images in XML) resolutionLevels, maximumOpticalResolution, opticalResolutionUnit, xOpticalResolution, yOpticalResolution



How can semantic tech help?

- Let you store and search on whatever metadata comes with image and content resource
- Easier searching using defined relationships between standards



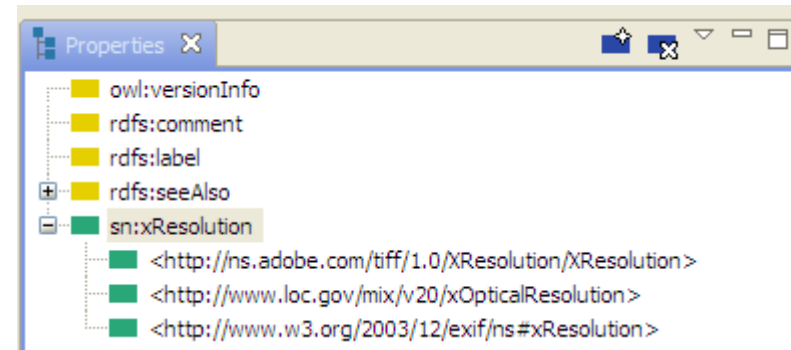
Relating same property from different standards

```
@prefix sn: <http://www.snee.com/ns/imagemetadata#> .
```

```
<http://ns.adobe.com/tiff/1.0/XResolution/XResolution>  
  rdfs:subPropertyOf sn:xResolution .
```

```
<http://www.loc.gov/mix/v20/xOpticalResolution>  
  rdfs:subPropertyOf sn:xResolution .
```

```
<http://www.w3.org/2003/12/exif/ns#xResolution>  
  rdfs:subPropertyOf sn:xResolution .
```





More overlapping standards

- PRISM embargoDate, expirationDate, rightsAgent
- ONIX for Licensing Terms, ONIX for Publications Licenses
- XACML (EXtensible Access Control Markup Language) specification of policies
- ODRL (Open Digital Rights Language)
- ccREL (The Creative Commons Rights Expression Language)
- ACAP (Automated Content Access Protocol)



What can semweb world learn from publishing world?

- Semantic web researchers and “Field of Dreams” ontologies
- Real data, real metadata
- Vocabulary development experience



“Next Generation” Cataloging and Metadata Creation Pilot



Automated capture, crosswalk and enhancement of publisher ONIX metadata

- Output in MARC and ONIX to benefit both library and publishing communities
- ONIX enriches MARC data and MARC enriches ONIX data
- Mapping between library and publisher terminologies
- OCLC pilot program with publishers, vendors and libraries



Learning from their vocabulary

WIKIPEDIA
The Free Encyclopedia

Main page

Contents

Featured content

Current events

Random article

Interaction

About Wikipedia

Community portal

Recent changes

Contact Wikipedia

Donate to Wikipedia

Help

Toolbox

Print/export

Crosswalk (metadata)

From Wikipedia, the free encyclopedia



This article is an **orphan**, as few or no other articles link to it. Please introduce links to this page from related articles; suggestions are available. *(February 2009)*

A **crosswalk** is a table that shows equivalent elements (or "fields") in more than one [database](#). It maps the elements in one [metadata](#) scheme to the equivalent elements in another scheme. For example, this is a crosswalk from [MARC](#) to [Dublin Core](#):

| MARC field | Dublin Core element |
|--|---------------------|
| 260\$c (Date of publication, distribution, etc.) | → Date.Created |
| 522 (Geographic Coverage Note) | → Coverage.Spatial |
| 300\$a (Physical Description) | → Format.Extent |

Crosswalks show people where to put the data from one scheme into a different scheme. They are often used by libraries, archives, museums, and other cultural institutions to translate data to or from [MARC](#), [Dublin Core](#), [TEI](#), and other metadata schemes. For example, say an archive has a MARC record in their catalog describing a manuscript. If the archive makes a digital copy of that manuscript and wants to display it on the web along with the information from the catalog, it will have to translate the data from the MARC catalog record into a different format such as [MODS](#) that is viewable in a webpage. Because MARC has different fields than MODS, decisions must be made about where to put the data into MODS. This type of "translating" from one format to another is often called "metadata mapping" or "field mapping," and is related to "[data mapping](#)," and "[semantic mapping](#)."

Crosswalks also have several technical capabilities. They help databases using different metadata schemes to share information. They help metadata harvesters create union catalogs. They enable search engines to search multiple databases simultaneously with a single query.

Contents [hide]

- 1 Challenges for crosswalks
 - 1.1 Difficulties in mapping
- 2 Examples
- 3 References
- 4 See also



Baselines for looking at the elephant

- **RDFa** Store “arbitrary” metadata (e.g. workflowStage = “3”)
- **SKOS** an ontology that publishers can understand and extend
- **SPARQL Query Results XML Format** semweb app output that publishers can work with easily



Conclusion

- Let's use standards to help publishers attack their difficult problems
- Let's apply our tools to their extensive data and metadata
- Learn from their long history of metadata management
- Work from the common ground with us that they understand



W3C Library Linked Data Incubator Group



| |
|--|
| W3C Activities |
| Incubator Activity |
| • Library Linked Data XG |
| Participants |
| Wiki space |
| Public XG Mailing List |
| Public LLD Community Mailing List |
| Charter |
| Search Incubator Activity |
| <input type="text"/> <input type="button" value="Go"/> |

[Incubator Activity](#) > W3C Library Linked Data Incubator Group

The mission of the Library Linked Data incubator group is to help increase global interoperability of library data on the Web, by bringing together people involved in Semantic Web activities—focusing on Linked Data—in the library community and beyond, building on existing initiatives, and identifying collaboration tracks for the future.

The group will explore how existing building blocks of librarianship, such as metadata models, metadata schemas, standards and protocols for building interoperability and library systems and networked environments, encourage libraries to bring their content, and generally re-orient their approaches to data interoperability towards the Web, also reaching to other communities. It will also envision these communities as a potential major provider of authoritative datasets (persons, topics...) for the Linked Data Web. As these evolutions raise a need for a shared standardization effort within the library community around (Semantic) Web standards, the group will refine the knowledge of this need, express requirements for standards and guidelines, and propose a way forward for the library community to contribute to further Web standardization actions.

See the [charter](#) and the group's [Wiki space](#) for more information.

- [News](#)
- [Deliverables](#)
- [Meetings](#)
- [Minutes](#)