Accessing Cultural Heritage using Semantic Web Techniques

DACH Master, 19 March 2009
Antoine Isaac
Vrije Universiteit Amsterdam, KB
aisaac@few.vu.nl

Background: me

• CATCH (NWO)
  – Continuous Access To Cultural Heritage
  – Computer science research projects applied to CH

• STITCH
  – Semantic Interoperability To access CH
  – Exchanging and integrating metadata

• W3C Semantic Web Deployment Group
  – SKOS

Topics

• A Web of data
• Smart data
• Demos

The Web for humans

• A city
• The city’s location
• Hyperlinks anchored to words

Meaning

SW problem: the Web for computers?

The Semantic Web vision: a web of (smart) data
A Web of resources

- Web-enabled Identifiers (URIs)
- Coming from different information spaces
- myVoc: http://example.org/myVocabulary/

Data in an RDF “graph”

Resource Description Framework: structured data as triple statements

- Links coming from different spaces

More than traditional metadata

- Web-based resources allow distribution/sharing/linking of:
  - documents
  - description vocabularies
  - (meta)data

Web of data: Open Linked Data

Interesting for CH?

Example CH data network

Johan Stapel, KB
Can we have that for the CH metadata?

CH case: Libris
- http://libris.kb.se/
- Swedish Library as linked data

Martin Malmsten, Dublin Core 2008

Linked descriptions of resources in Libris

External links in Libris: Library of Congress Subject Headings
Ed Summers et. al., Dublin Core 2008

Searching using multiple vocabularies

Semantic interoperability problem
Blue triangles: (collection-)specific vocabularies
Semantic alignment as a solution

Can we have more?

- Web of data
- Smart data: more "semantics"

Creating vocabularies of “building blocks” for RDF graphs

Ontologies

- Ontologies specify description vocabularies which can be shared
  - subject, Article
- Give formal definition to vocabulary elements
  - Every Article is a Document

Machine-readable definitions

CH case: Europeana Thought Lab

Allows deduction of new facts & control of existing facts by reasoning engines
Semantics for querying

- The query:
  - vra:subject
  - skos:broader
  - rma:Egypt

- The existing description:
  - rma:depicts
  - skos:broader
  - rma:Egypt
  - rma:gezicht_in_cairo

- Why is there a match? For the Europeana ontology, every rma:depicts statement implies a vra:subject statement.

More than traditional metadata

- **Flexible** reasoning: a same base can be easily added with new descriptions from different places, using different ontologies
  - dach09:shows_DACH_Student

- Requirement: semantically connect these ontologies
  - dach09:shows_DACH_Student "implies" vra:subject

- SW principle: meaning is accessible with the data, not encoded in external programs
  - SW principle: meaning is accessible with the data, not encoded in external programs

Message

- A web of (meta)data
  - Descriptions of resources
  - Easy to share and interconnect

- Smart data
  - Machine-readable definitions for the data

- Relies on open standards
  - W3C's URI, XML, RDF, OWL, SPARQL, SKOS...

- Can be crucial for CH!

Links

- **Semantic Web at W3C**
  - [http://www.w3.org/2001/sw/](http://www.w3.org/2001/sw/)

- **SKOS**
  - [http://www.w3.org/2004/02/skos/](http://www.w3.org/2004/02/skos/)

- **Cultural Heritage and Semantic Web projects**
  - eCulture, [http://e-culture.multimedian.nl/](http://e-culture.multimedian.nl/)
  - Libris, [http://libris.kb.se](http://libris.kb.se)
  - Europeana Thought Lab, [http://europeana.eu/portal/thought-lab.html](http://europeana.eu/portal/thought-lab.html)
  - CATCH, [http://www.w3.org/catch](http://www.w3.org/catch)

We can stop here!

- Questions?

- Next slides:
  - Links
  - Intro to SKOS: porting CH vocabularies on the SW
Now, how to have SW data?

• Creating born-SW data
  – Annotation of documents
• Porting existing data
• Enriching existing data
  – Information extraction from text

Example: CH Metadata

• Use of controlled documentary languages
  – Thesauri, classification systems, subject heading lists

Example: CH vocabulary (Iconclass)

<table>
<thead>
<tr>
<th>Abstract, Mon-representational Art</th>
<th>Nature</th>
</tr>
</thead>
<tbody>
<tr>
<td>earth, world, celestial body</td>
<td></td>
</tr>
<tr>
<td>257 animals</td>
<td></td>
</tr>
<tr>
<td>257 31 birds</td>
<td></td>
</tr>
<tr>
<td>257 31:1 group of birds</td>
<td></td>
</tr>
<tr>
<td>257 31:0 songbirds</td>
<td></td>
</tr>
<tr>
<td>257 31:1 predatory birds</td>
<td></td>
</tr>
<tr>
<td>257 34:0 owls</td>
<td></td>
</tr>
<tr>
<td>257 35 ornamental birds</td>
<td></td>
</tr>
<tr>
<td>257 36:0 water-birds</td>
<td></td>
</tr>
<tr>
<td>257 31:0 shore-birds and wading-birds</td>
<td></td>
</tr>
<tr>
<td>257 38:0 walker and rafter birds</td>
<td></td>
</tr>
<tr>
<td>257 39:0 other birds</td>
<td></td>
</tr>
</tbody>
</table>

Example: SKOS graph

• Observation: there are many KOS models/formats:
  – Thesauri, classification schemes, etc...
• But also common features, used by typical applications
  – Lexical information, semantic links
• SKOS (Simple Knowledge Organization System)
• W3C
• Model to represent existing vocabularies on the SW in a simple way
  – Comparable to DC, for conceptual vocabularies
Networking controlled vocabularies in SKOS

animals
cats
wildcats

animal
human
object

ex1:animals

ex2:animal

skos:exactMatch
skos:broaderMatch

ex1:cats