Accessing Cultural Heritage using Semantic Web Techniques

DACH Master, 19 March 2009

Antoine Isaac

Vrije Universiteit Amsterdam, KB
aisaac@few.vu.nl

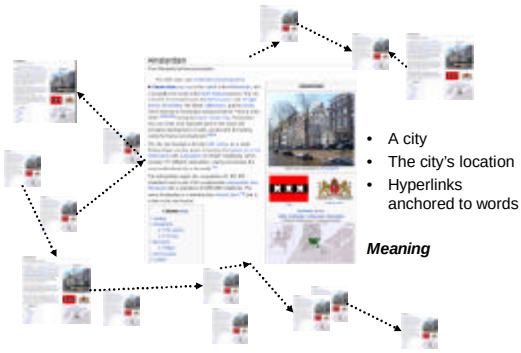
Background: me

- 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

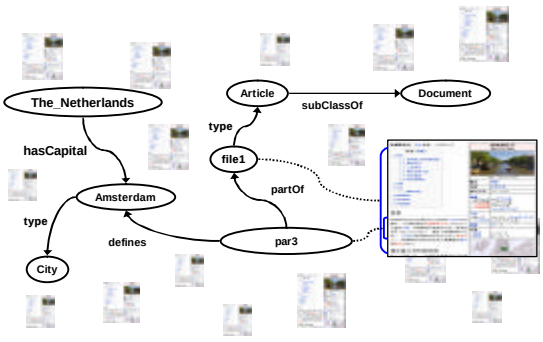
- A Web of data
- Smart data
- Demos

The Web for humans

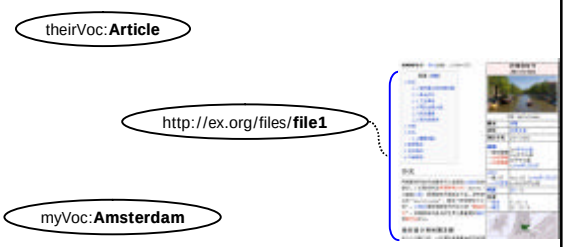


SW problem: the Web for computers?

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



A Web of resources

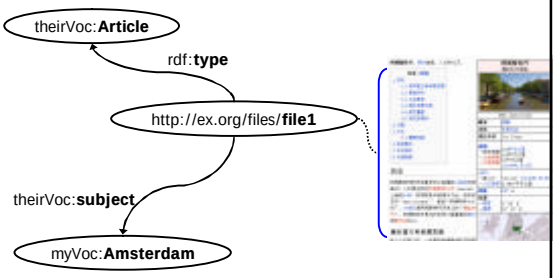


- Web-enabled Identifiers (URIs)
- Coming from different information spaces

myVoc: = http://example.org/myVocabulary/

Data in an RDF “graph”

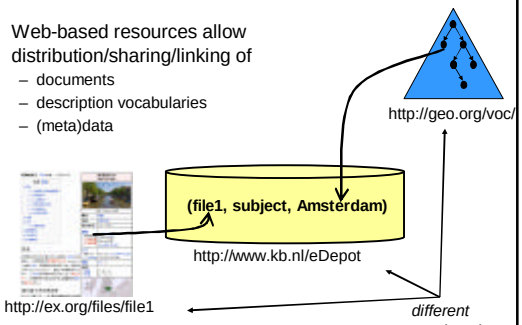
Resource Description Framework : structured data as triple statements



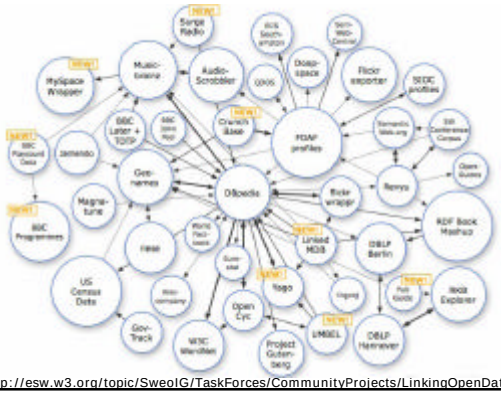
Links coming from different spaces

More than traditional metadata

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



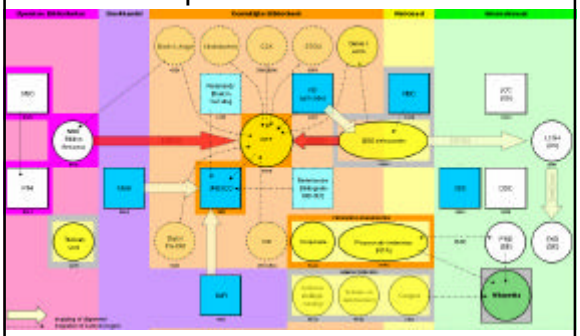
Web of data: Open Linked Data



http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

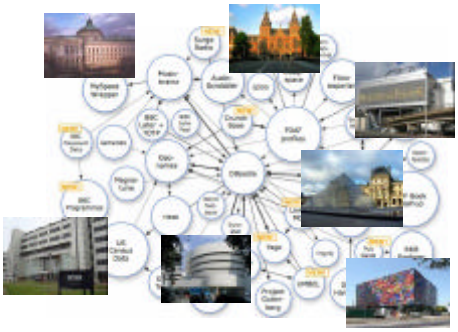
Interesting for CH?

Example CH data network



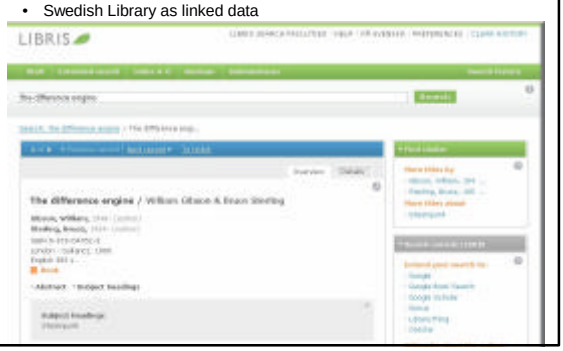
Johan Stapel, KB

Can we have that for the CH metadata?

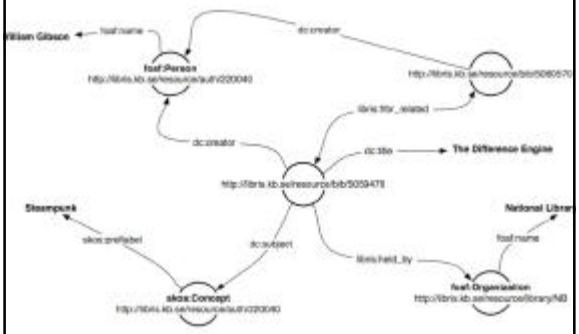


CH case: Libris

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

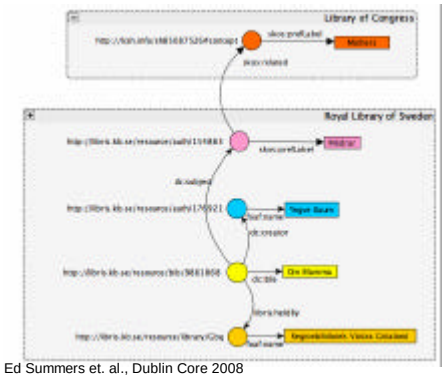


Linked descriptions of resources in Libris



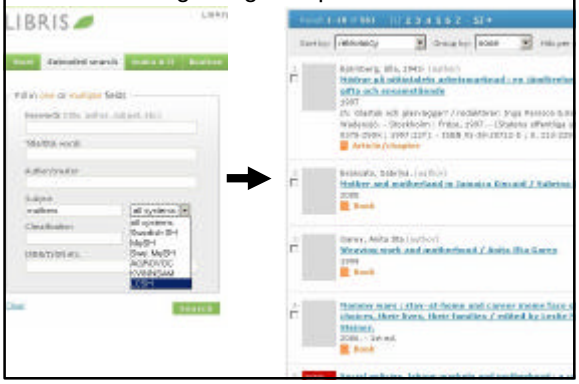
Martin Malmsten, Dublin Core 2008
http://dc2008.de/wp-content/uploads/2008/09/malmsten.pdf

External links in Libris: Library of Congress Subject Headings

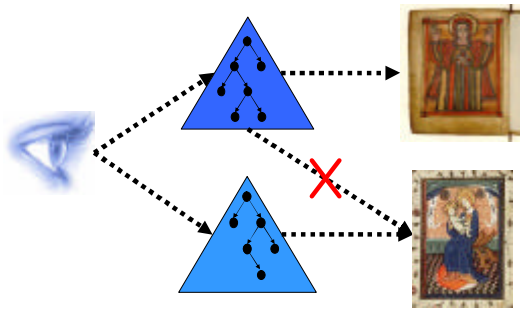


Ed Summers et. al., Dublin Core 2008
http://dc2008.de/wp-content/uploads/2008/09/summers-isaac-redding-krech.pdf

Searching using multiple vocabularies

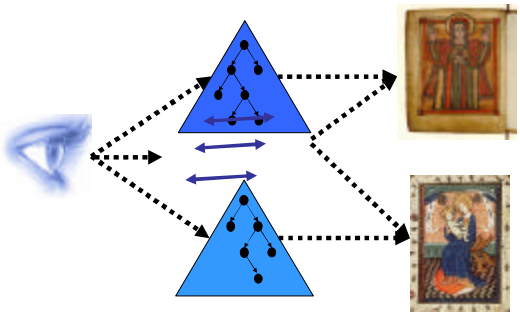


Semantic interoperability problem



Blue triangles: (collection-)specific vocabularies

Semantic alignment as a solution

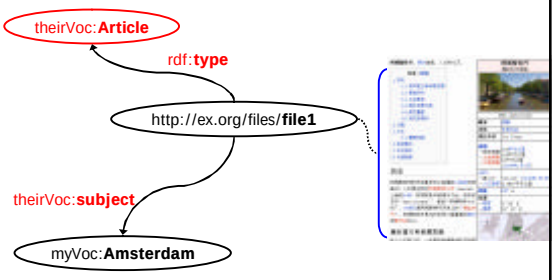


Cf. Manuscript demo: http://stitch.cs.vu.nl/BNF_KB_demo.html

Can we have more?

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

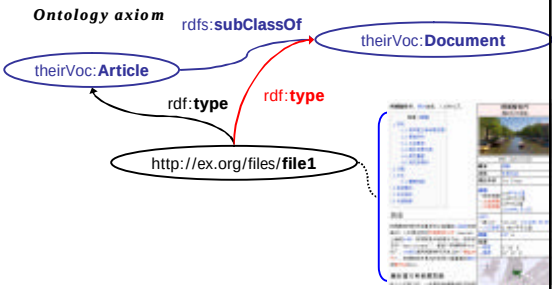
Creating vocabularies of “building blocks” for RDF graphs



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



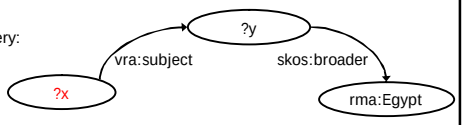
Allows deduction of new facts & control of existing facts by reasoning engines

CH case: Europeana Thought Lab

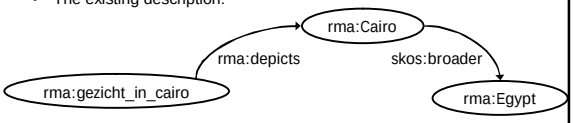
http://europeana.eu/portal/thought-lab.html

Semantics for querying

- The query:



- The existing description:



- 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

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!

We can stop here!

- Questions?
- Next slides:
  - Links
  - Intro to SKOS: porting CH vocabularies on the SW

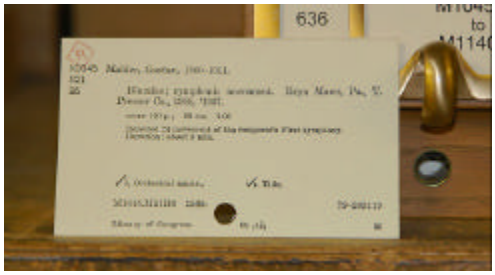
Links

- Semantic Web at W3C
  - http://www.w3.org/2001/sw/
- SKOS
  - http://www.w3.org/2004/02/skos/
- Cultural Heritage and Semantic Web projects
  - MuseumFinland, http://www.museosuomi.fi/
  - eCulture, http://e-culture.multimedian.nl/
  - Libris, http://libris.kb.se
  - Europeana Thought Lab: http://europeana.eu/portal/thought-lab.html
  - STITCH, http://stitch.cs.vu.nl/demo/
  - CATCH, http://www.nwo.nl/catch

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

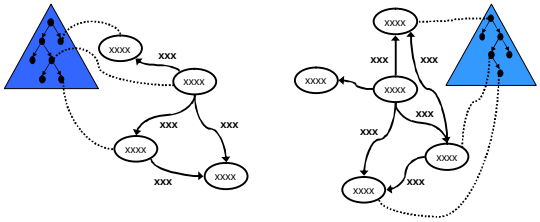


- Use of controlled **documentary languages**
  - Thesauri, classification systems, subject heading lists

Example: CH vocabulary (Iconclass)

0 Abstract, Non-representational Art
1 Religion and Magic
2 Nature
25 earth, world as celestial body
25F animals show images >25
25F3 birds show images >25
25F31 groups of birds show images < 5
25F32 song-birds show images >25
25F33 predatory birds show images >25
25F34 owls show images < 25
25F35 ornamental birds show images < 25
25F36 water-birds show images >25
25F37 shore-birds and wading-birds show images >25
25F38 walker and runner birds show images < 25
25F39 other birds show images >25

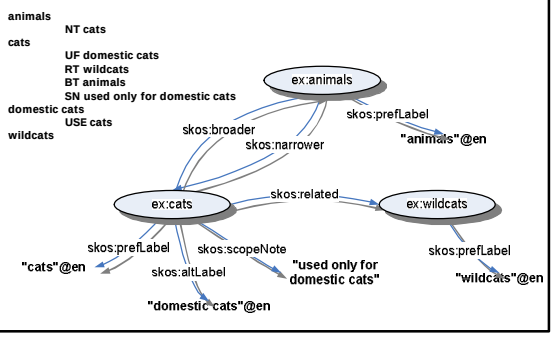
Porting controlled vocabularies to the Semantic Web



SKOS

- 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

Example: SKOS graph



Networking controlled vocabularies in SKOS