Semantic interoperability of data values, use and matching of ontologies and unstructured vocabularies

Jacco van Ossenbruggen
VU/CWI Amsterdam

Image by Thierry Noir, see http://commons.wikimedia.org/wiki/Image:Berlinermauer.jpg

Image courtesy of Paul Gramen Raven, CC-SA license, see http://www.flickr.com/photos/armchairanarchist/2869662267/

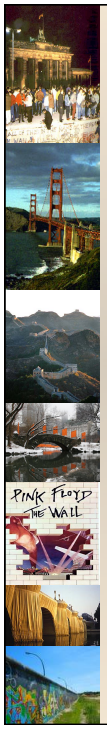
Semantic interoperability of data values, use and matching of ontologies and unstructured vocabularies

Jacco van Ossenbruggen
VU/CWI Amsterdam

4

Tearing Down Walls
&
Building Bridges

Steps towards a Culture Web

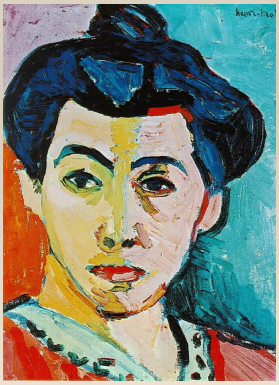


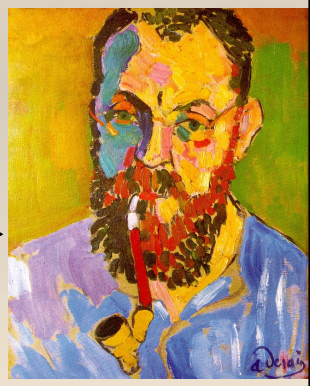
Interoperability: tearing down the walls between collections

- Musea have increasingly nice websites
- But: most of them are driven by stand-alone collection databases
- Data is isolated, both syntactically and semantically
- If users can do cross-collection search, the individual collections become more valuable!

6

The Web: "open" documents and links





URL

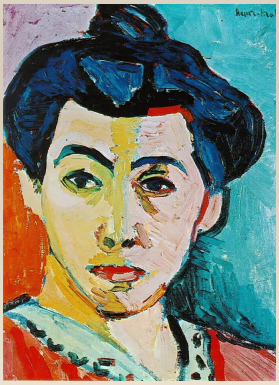
Web link

URL

7

The Semantic Web: "open" data and links

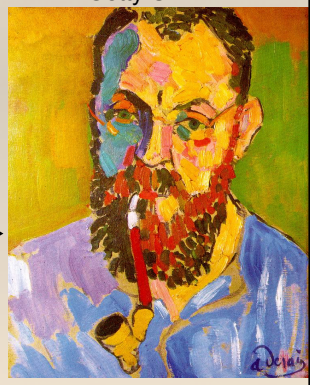
Painting
"Green Stripe (M^me^ Matisse)"
Royal Museum of Fine Arts, Copenhagen



creator

Dublin Core

Painter
"Henri Matisse"
Getty ULAN

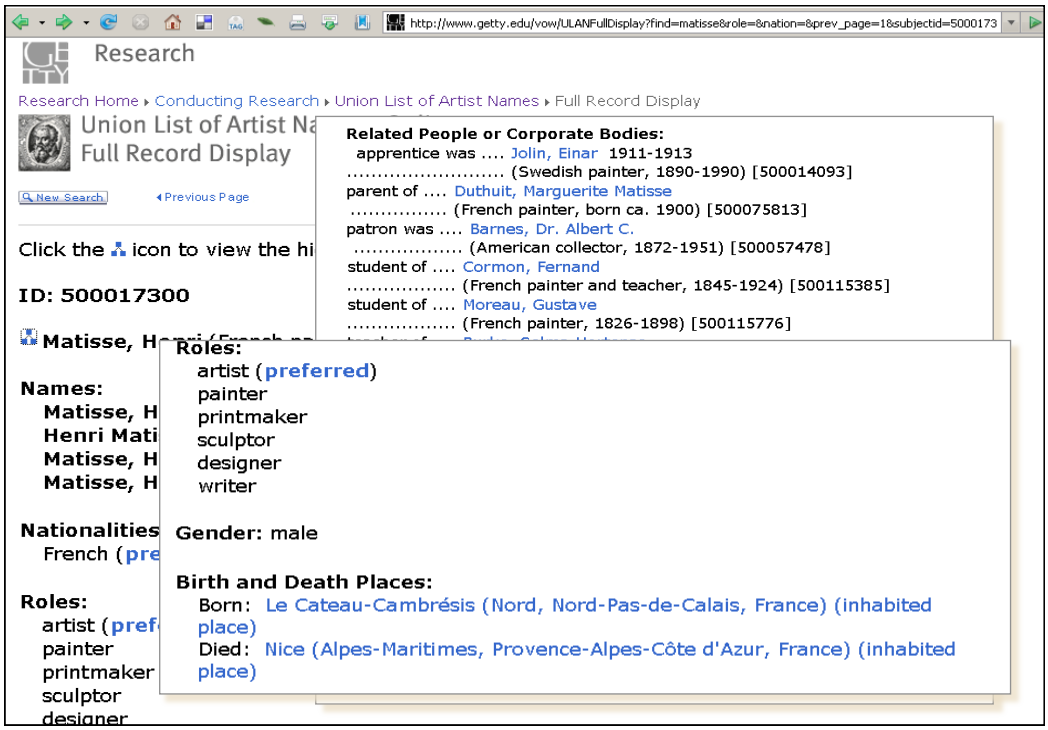


URL

Web link

URL

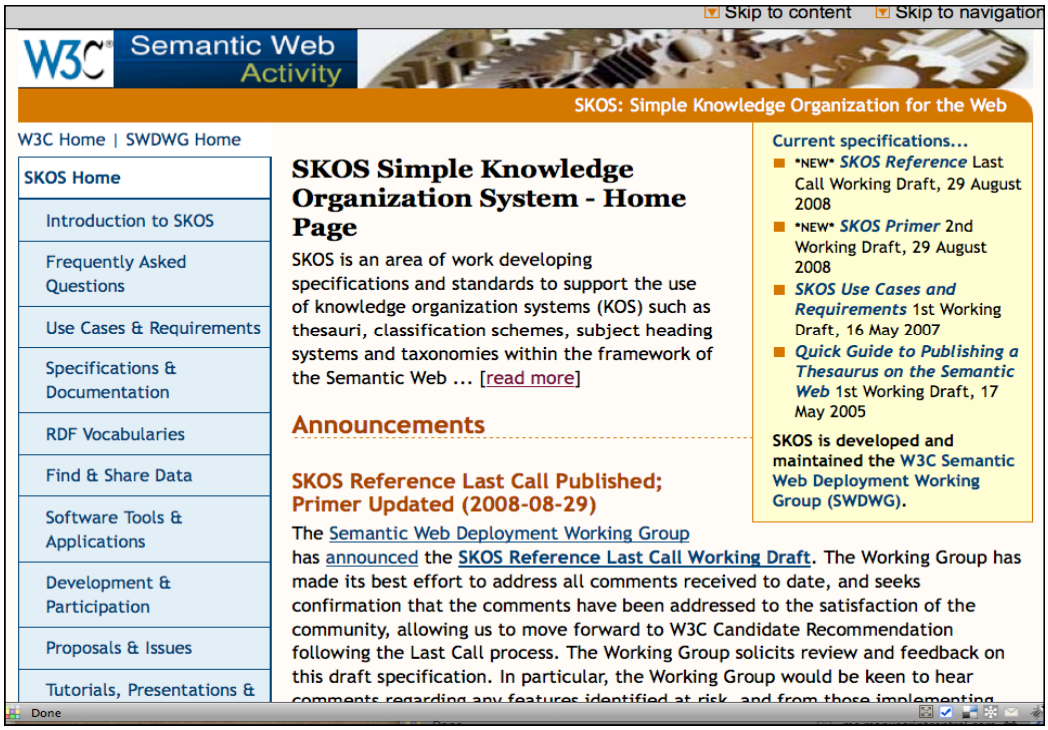
8

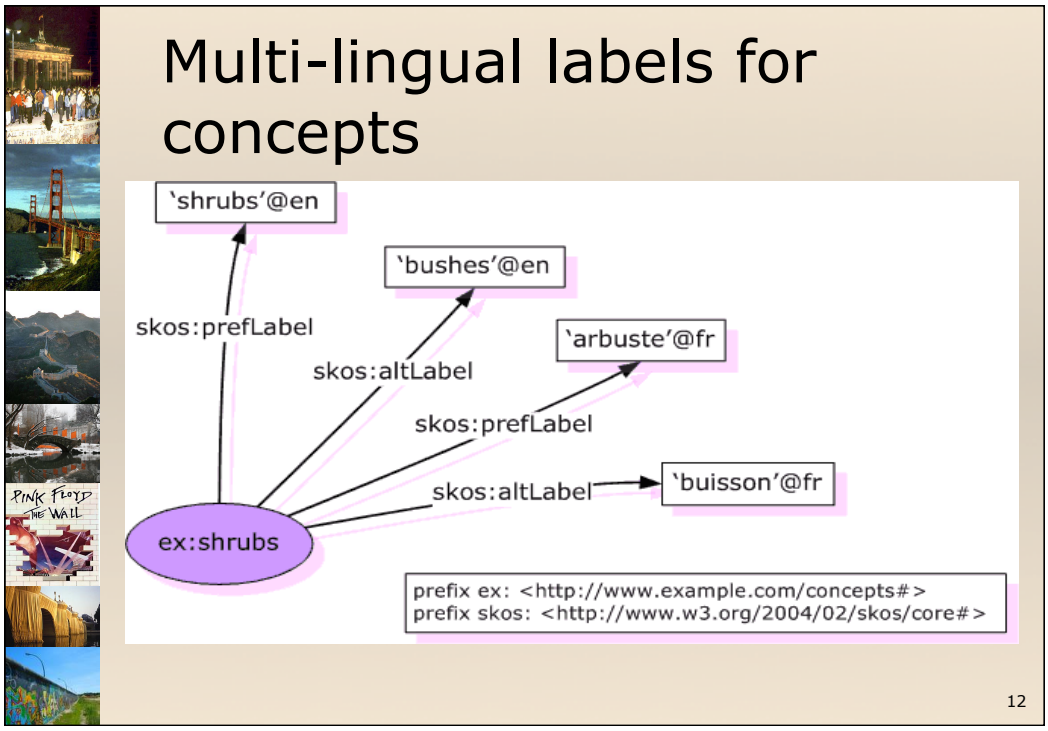


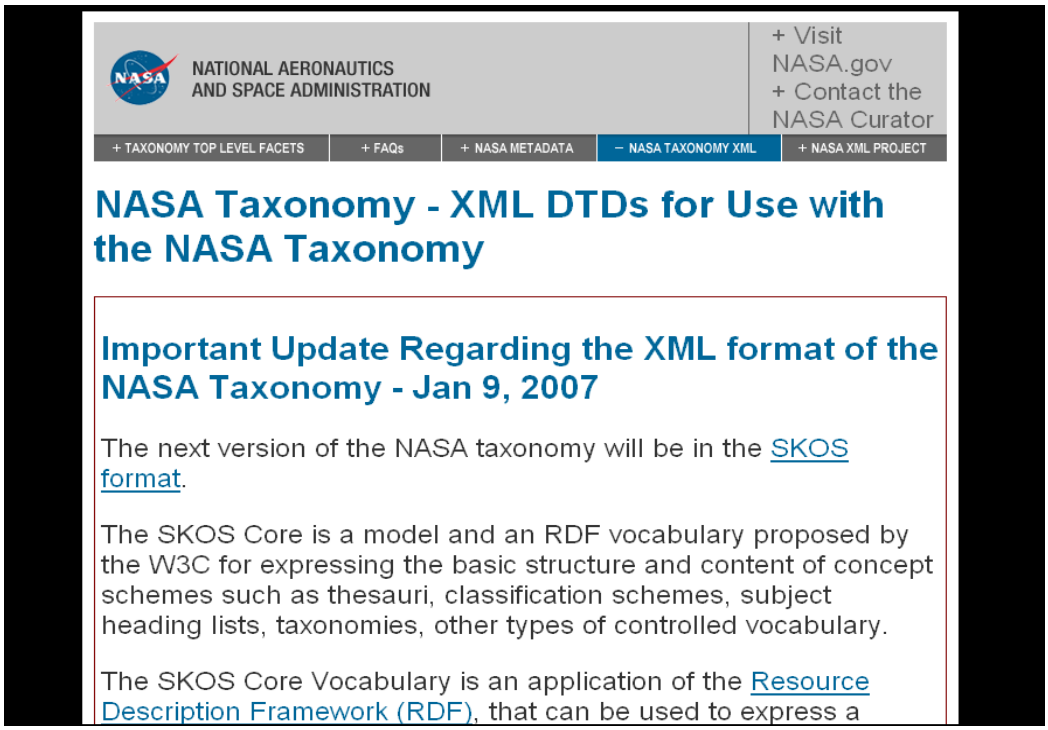
Levels of interoperability

- Syntactic interoperability
  - using data formats that you can share
  - XML family is the preferred option
- Semantic interoperability
  - How to share meaning / concepts
  - Technology for finding and representing semantic links

10

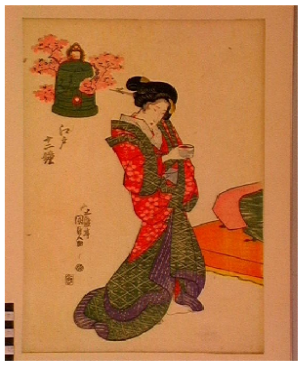

Skip to content
Skip to navigation
W3C Semantic Web Activity
SKOS: Simple Knowledge Organization for the Web
W3C Home | SWDWG Home
SKOS Home
Introduction to SKOS
Frequently Asked Questions
Use Cases & Requirements
Specifications & Documentation
RDF Vocabularies
Find & Share Data
Software Tools & Applications
Development & Participation
Proposals & Issues
Tutorials, Presentations &
SKOS Simple Knowledge Organization System - Home Page
SKOS is an area of work developing specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web ... [read more]
Announcements
SKOS Reference Last Call Published; Primer Updated (2008-08-29)
The Semantic Web Deployment Working Group has announced the SKOS Reference Last Call Working Draft. The Working Group has made its best effort to address all comments received to date, and seeks confirmation that the comments have been addressed to the satisfaction of the community, allowing us to move forward to W3C Candidate Recommendation following the Last Call process. The Working Group solicits review and feedback on this draft specification. In particular, the Working Group would be keen to hear
Current specifications...
NEW SKOS Reference Last Call Working Draft, 29 August 2008
NEW SKOS Primer 2nd Working Draft, 29 August 2008
SKOS Use Cases and Requirements 1st Working Draft, 16 May 2007
Quick Guide to Publishing a Thesaurus on the Semantic Web 1st Working Draft, 17 May 2005
SKOS is developed and maintained the W3C Semantic Web Deployment Working Group (SWDWG).
Done


Multi-lingual labels for concepts
'shrubs'@en
'bushes'@en
'arbuste'@fr
'buisson'@fr
skos:prefLabel
skos:altLabel
skos:prefLabel
skos:altLabel
ex:shrubs
prefix ex: <http://www.example.com/concepts#>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
12



Principle 1: semantic annotation

- Description of web objects with "concepts" from a shared vocabulary



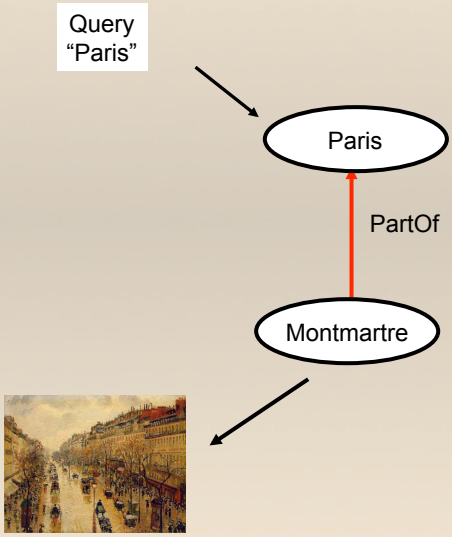
Description:

recordnumber	23727;
timestamp	2001-06-22;
type	cultural;original;
collector	Johannes Frederik van Ov Zeldzaamheden; Utagawa
series	360;
Culture	Japans;
Date	1800-1829;
Description	1883 JAPAN aankoop;
Identifier in Current Repository	360-4564;
Creation Site	Japan;
Current Repository	RMV;
Material	papier;
Measurements.Format	oban, 25.5 cm x 37.5 cm;
Style/Period.Period	Edo;
Title	Edo junisho;
Type	prenten;
type	Work;

14

Principle 2: semantic search

- Search for objects which are linked via concepts (semantic link)
- Use the type of semantic link to provide meaningful presentation of the search results



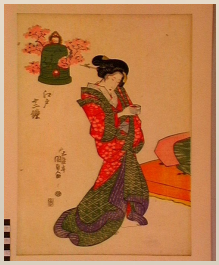
15

Term disambiguation is key issue in semantic search

- Post-query
  - Sort search results based on different meanings of the search term
  - Mimics Google-type search
- Pre-query
  - Ask user to disambiguate by displaying list of possible meanings
  - Interface is more complex, but more search functionality can be offered

16

Principle 3: vocabulary alignment



AAT style/period
Edo (Japanese period)
Tokugawa

AAT is Getty's Art & Architecture Thesaurus

SVCN period
Edo

SVCN is local in-house ethnology thesaurus

17

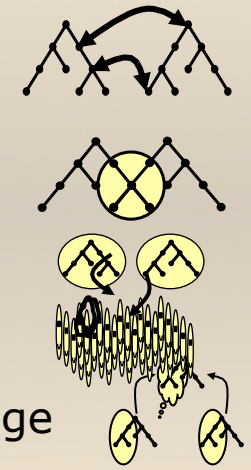
The myth of a unified vocabulary

- In large virtual collections there are always multiple vocabularies
  - In multiple languages
- Every vocabulary has its own perspective
  - You can't just merge them
- But you can use vocabularies jointly by defining a limited set of links
  - "Vocabulary alignment"
- It is surprising what you can do with just a few links

18

Common approaches to ontology matching

- Linguistics & structure
- Shared vocabulary
- Instance-based matching
- Shared background knowledge



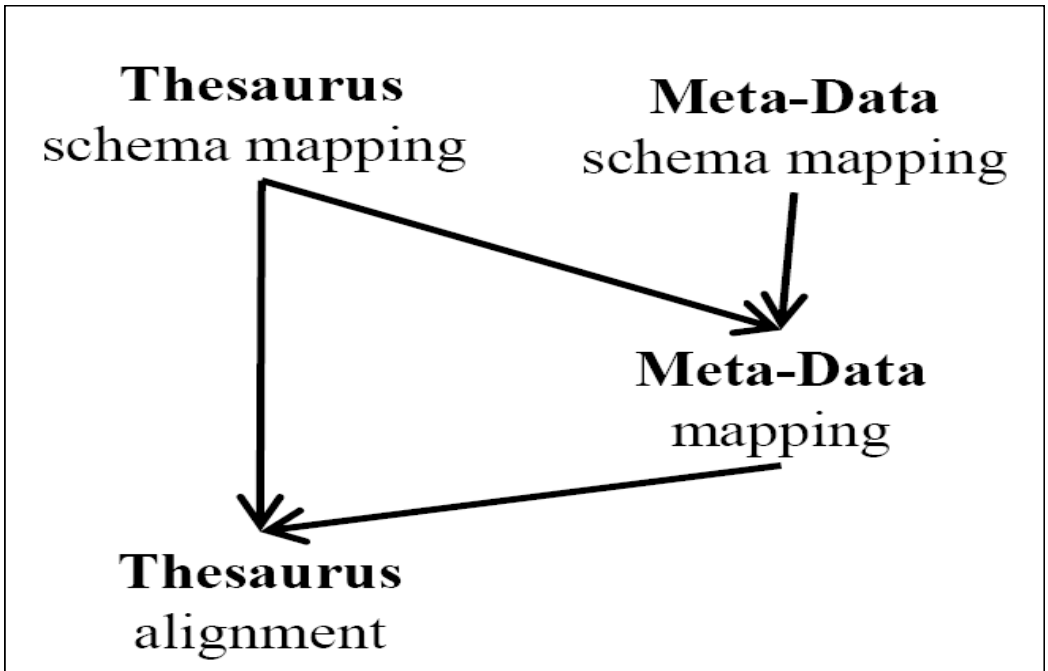
Original slide by Frank van Harmelen

19

Learning alignments

- Learning relations between art styles in AAT and artists in ULAN through NLP of art historic texts
  - "Who are Impressionist painters?"

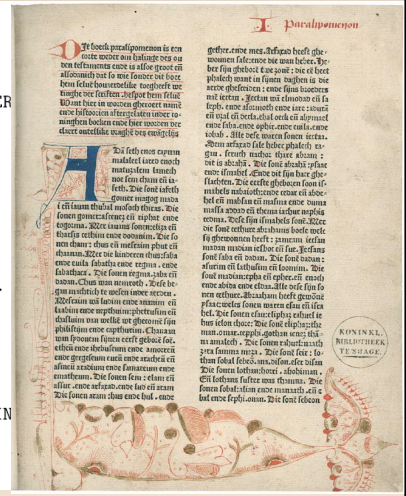
Artist Name	*IS*	*In GS*
edgar degas	0.0699	1
edouard manet	0.0548	1
pierre-auguste renoir	0.0539	1
morisot, berthe	0.0393	1
gogh, vincent van	0.0337	0
cassatt, mary	0.0318	1
cezanne, paul	0.0302	1



21

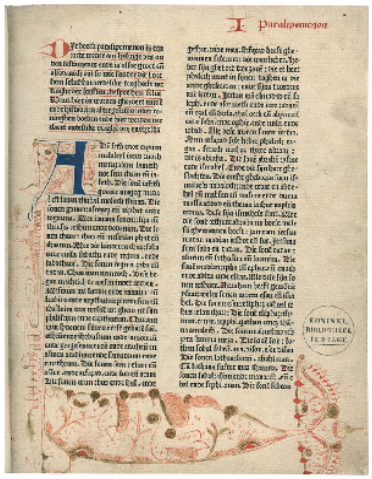
Example textual annotation

```
<inm:Record>
    <inm:NUMMER>6</inm:NUMMER>
    <inm:TITEL>Delftse Bijbel...</inm:TITEL>
    <inm:TITEL_EN>Delft Bible...</inm:TITEL_EN>
    <inm:MAKER>Yemantszoon, Mauricius : d</inm:MAKER
    <inm:OBJECT>tekstbladzijde</inm:OBJECT>
    <inm:TECHNIEK>boekdruk</inm:TECHNIEK>
    <inm:DATERING>10 jan. 1477</inm:DATERING>
    <inm:CLASSIFICATIE>D</inm:CLASSIFICATIE>
    <inm:ORIGINEEL>Bijbel. Oude
        Testament...</inm:ORIGINEEL>
    </inm:REPRODUCTIE>
    <inm:TWNAAM/>
    <inm:TWOND>typografische vormgeving</inm:TWOND>
    <inm:TWOND>bijbels</inm:TWOND>
    <inm:TWGEO>Delft</inm:TWGEO>
    <inm:OMSCHRIJVING>Eerste bijbel die in het
        Nederlands verscheen...</inm:OMSCHRIJVING>
    <inm:OMSCHRIJVING_EN>The first Bible to
    appear in the Dutch language...</inm:OMSCHRIJVIN
    <inm:AFMETINGEN>27 x 20 cm</inm:AFMETINGEN>
    ...
</inm:Record>
```



22

Resulting semantic annotation (rendered as HTML with RDFa)



Description:

classificatie	Geschiedenis van de boekdrukkunst;
drukker	Meer, Jacob Jacobszoon van der; Yermantszoon, Mauricius;
origineel	Bijbel. Oude Testament. - Delft: Jacob Jacobszoon van der Meer en Mauricius Yermantszoon, 10 jan. 1477, dl. 2, p. 1;
Date	10 jan. 1477;
Description	The first Bible to appear in the Dutch language, known as the Delft Bible. It consists of the Old Testament only and is an anonymous adaptation of the - again anonymous - History Bible of 1360. It is an example of an incunabulum where the hand-written book still served as an example for lay-out and design. Contrary to many other incunabula, the place of origin, the names of the printers and even the day of its completion are mentioned in the colophon.;
Measurements.Dimensions	27 x 20 cm;
rights.copyright	Den Haag Koninklijke Bibliotheek;
Source	Bibliopolis;
Subject	bibles; incunabula; initials; ornamental borders; rubrications; typographical design;
subject.geographicPlace	Delft;
Technique	letterpress printing;
Title	Delft Bible, printed in Delft by Jacob Jacobszoon van der Meer and Mauricius Yermantszoon, 1477;
Type	tekstbladzijde;
type	Work;

Used as value to describe other resources:

BBB_169E56_1477_P1.JPG;	relation.depicts

Perspectives

- Openness seems contagious
  - sometimes cannot be open in only a part of your workflow
- Social barriers have to be overcome!
  - "open door" policy
  - Involvement of general public => issues of "quality"

Caveats for museum software

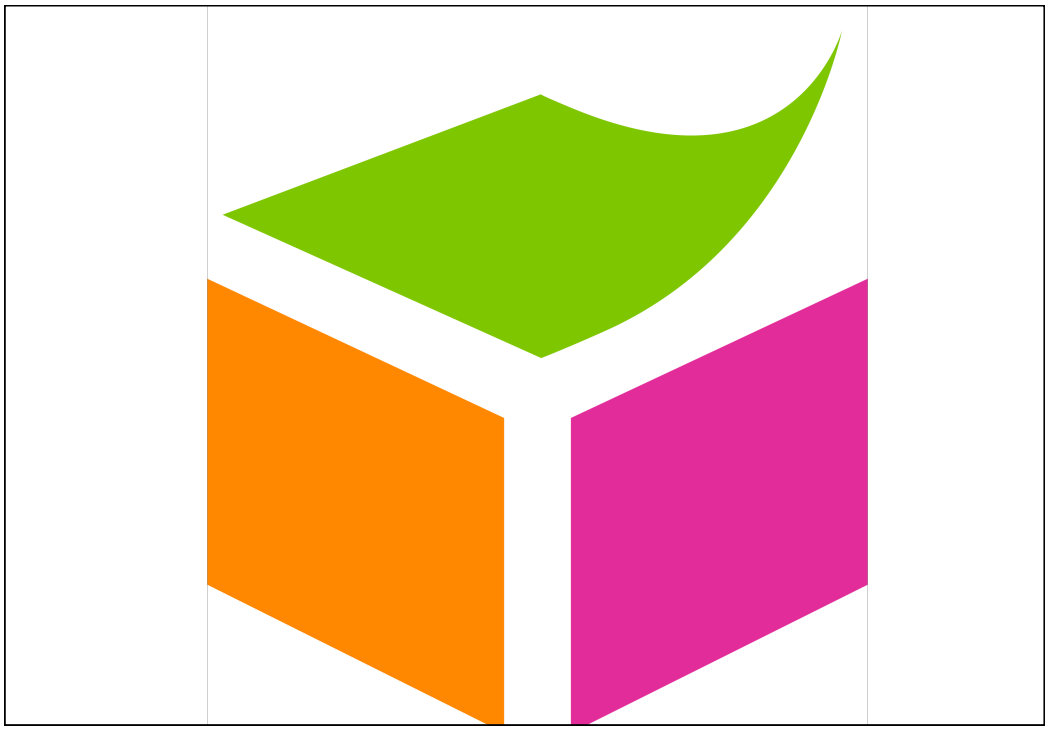
- Be wary of Flash
  - Accessibility
- Make sure you can connect others and other can connect to you
  - "Don't buy software which does not support standard open API's"
- Export facilities to common formats (XML, RDF, DC, SKOS, …)

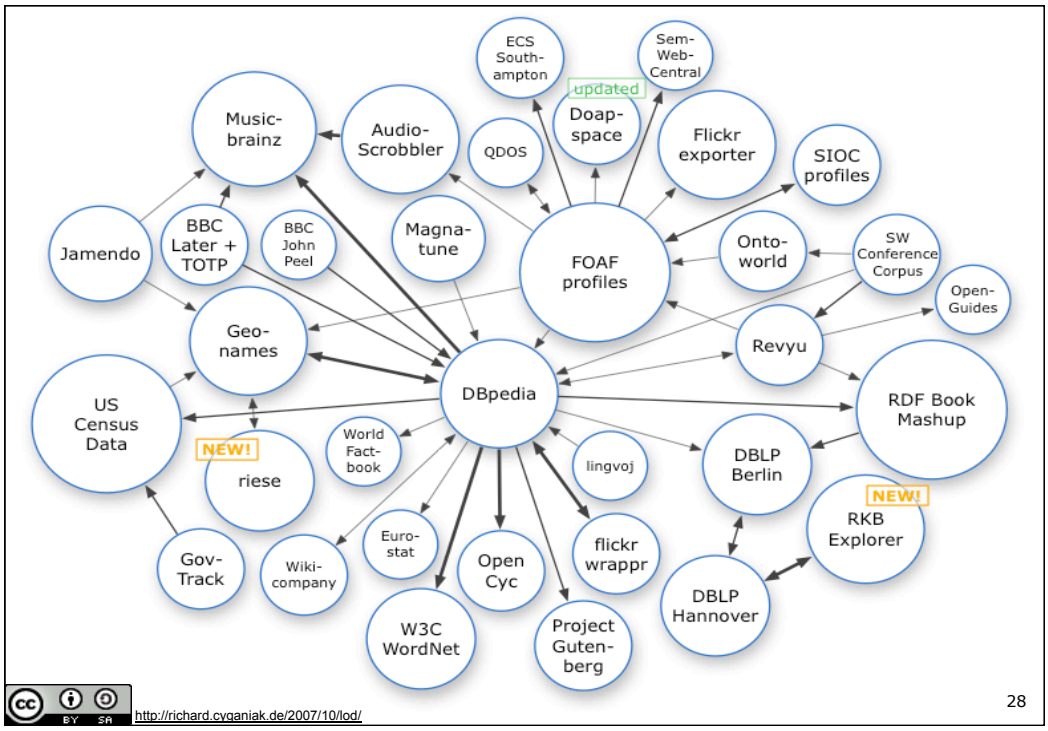
25

MoveMyData.org

Because if you can't move it it's not really yours

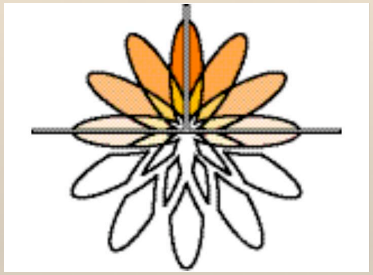
http://movemydata.org/




ECS South-ampton
Sem-Web-Central
updated
Doap-space
Flickr exporter
SIOC profiles
Music-brainz
Audio-Scrobbler
QDOS
Magna-tune
FOAF profiles
Onto-world
SW Conference Corpus
Open-Guides
Jamendo
BBC Later + TOTP
BBC John Peel
Geo-names
Revyu
DBpedia
RDF Book Mashup
US Census Data
NEW!
riese
World Fact-book
lingvoj
DBLP Berlin
NEW!
RKB Explorer
Euro-stat
flickr wrappr
Gov-Track
Wiki-company
Open Cyc
DBLP Hannover
W3C WordNet
Project Guten-berg
http://richard.cyganiak.de/2007/10/lod/
28

http://e-culture.multimedian.nl

- Part of the Dutch knowledge-economy project **MultimediaN**
- Partners: VU, CWI, UvA, DEN, ICN
- People:
  Alia Amin, Lora Aroyo, Mark van Assem, Victor de Boer, Lynda Hardman, Michiel Hildebrand, Laura Hollink, Marco de Niet, Borys Omelayenko, Marie-France van Orsouw, Jacco van Ossenbruggen, Guus Schreiber Jos Taekema, Annemiek Teesing, Anna Tordai, Jan Wielemaker, Bob Wielinga
- Artchive.com, RKD, Rijksmuseum Amsterdam, Dutch ethnology musea (Amsterdam, Leiden), National Library (Bibliopolis)



29