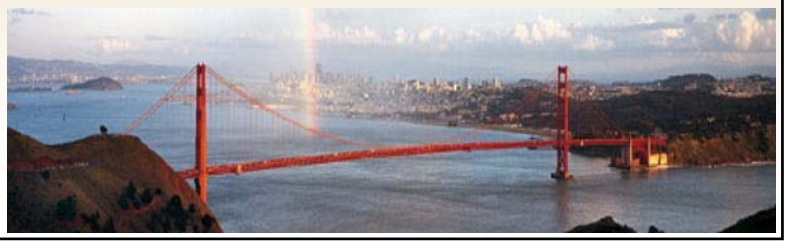
Tearing down walls
and
Building bridges

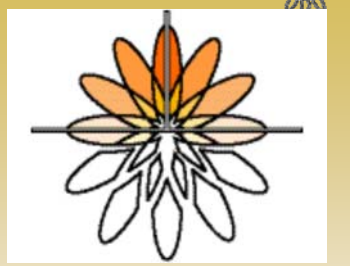
Principles and pragmatics of a Semantic Culture Web



Acknowledgements

- Part of large Dutch knowledge-economy project **MultimediaN**
- Partners: VU, CWI, UvA, DEN,ICN
- People:

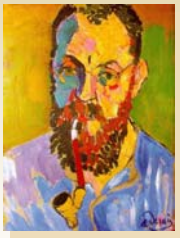
  Alia Amin, Lora Aroyo, Mark van Assem, Victor de Boer, Lynda Hardman, Michiel Hildebrand, Laura Hollink, Marco de Niet, Borys Omelayenko, Marie-France van Orsouw, Jacco van Ossenbruggen, Guus Schreiber Jos Taekema, Annemiek Teesing, Anna Tordai, Jan Wielemaker, Bob Wielinga
- Artchive.com, Rijksmuseum Amsterdam, Dutch ethnology musea (Amsterdam, Leiden), National Library (Bibliopolis)



The Web: resources and links

URL

Web link



URL

The Semantic Web: typed resources and links

Painting
"Woman with hat
SFMOMA

URL

Dublin Core

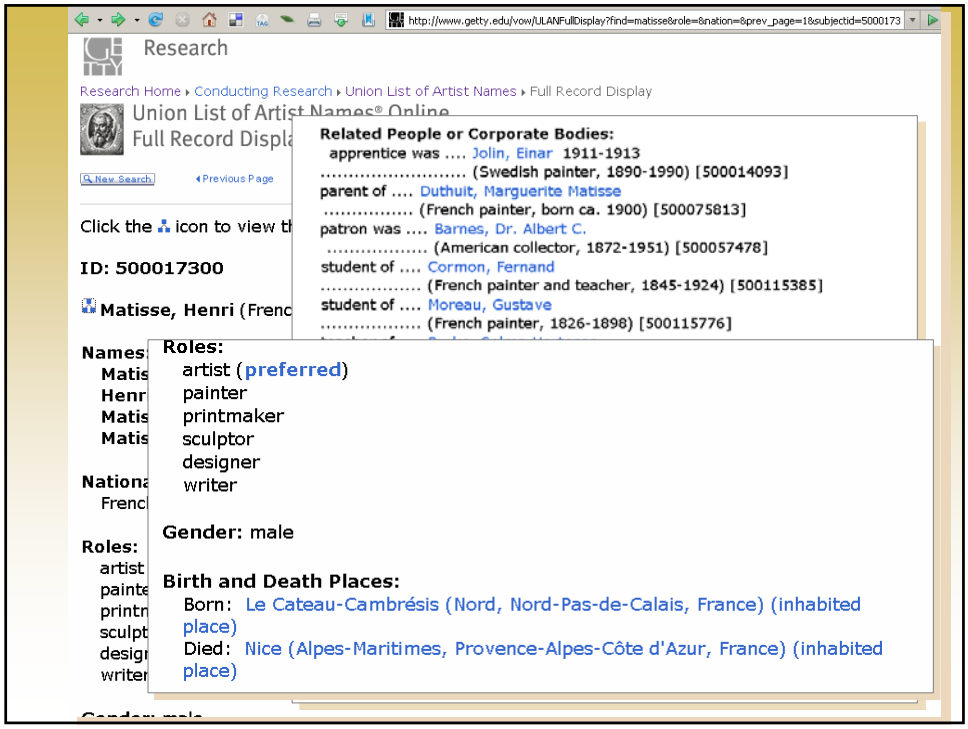
creator

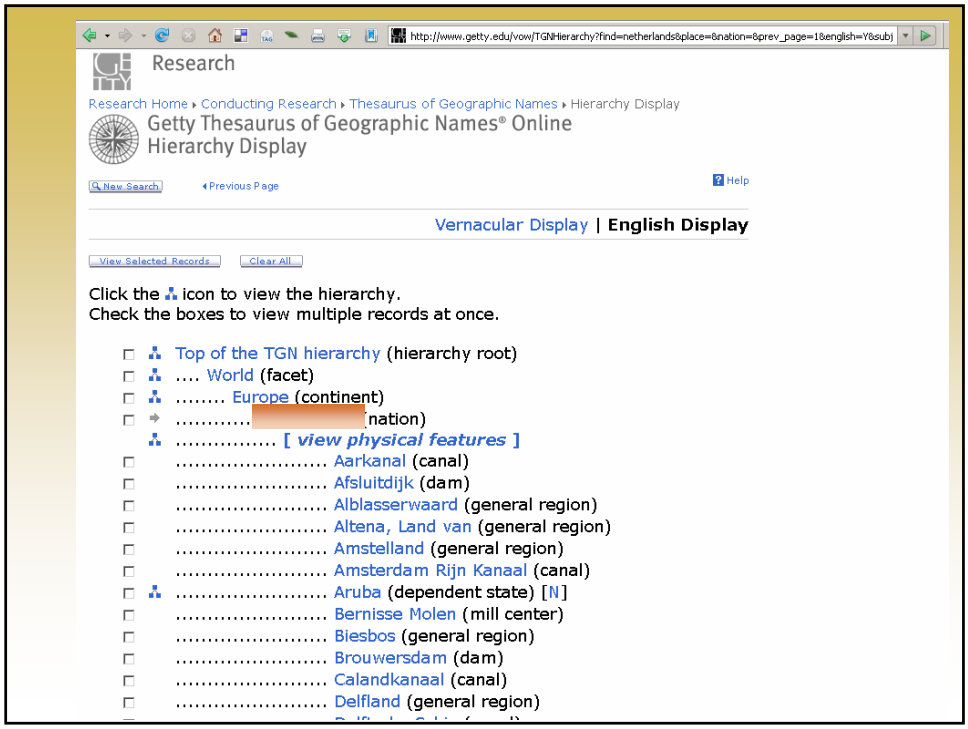
Web link

ULAN

Henri Matisse

URL


http://www.getty.edu/vow/ULANFullDisplay?find=matisse&role=&nation=&prev_page=1&subjectid=5000173
Research
Research Home ▸ Conducting Research ▸ Union List of Artist Names ▸ Full Record Display
Union List of Artist Names® Online
Full Record Displa
New Search
Previous Page
Click the icon to view th
ID: 500017300
Matisse, Henri (Frenc
Names
Matis
Henr
Matis
Matis
Nationa
Frenc
Roles:
artist
painte
printn
sculpt
desigr
writer
Related People or Corporate Bodies:
apprentice was .... Jolin, Einar 1911-1913
..... (Swedish painter, 1890-1990) [500014093]
parent of .... Duthuit, Marguerite Matisse
..... (French painter, born ca. 1900) [500075813]
patron was .... Barnes, Dr. Albert C.
..... (American collector, 1872-1951) [500057478]
student of .... Cormon, Fernand
..... (French painter and teacher, 1845-1924) [500115385]
student of .... Moreau, Gustave
..... (French painter, 1826-1898) [500115776]
Roles:
artist (preferred)
painter
printmaker
sculptor
designer
writer
Gender: male
Birth and Death Places:
Born: Le Cateau-Cambrésis (Nord, Nord-Pas-de-Calais, France) (inhabited place)
Died: Nice (Alpes-Maritimes, Provence-Alpes-Côte d'Azur, France) (inhabited place)


http://www.getty.edu/vow/TGNHierarchy?find=netherlands&place=&nation=&prev_page=1&english=Y&subj
Research
Research Home ▸ Conducting Research ▸ Thesaurus of Geographic Names ▸ Hierarchy Display
Getty Thesaurus of Geographic Names® Online
Hierarchy Display
New Search
Previous Page
Help
Vernacular Display | English Display
View Selected Records
Clear All
Click the icon to view the hierarchy.
Check the boxes to view multiple records at once.
Top of the TGN hierarchy (hierarchy root)
.... World (facet)
..... Europe (continent)
..... nation)
..... [view physical features]
..... Aarkanal (canal)
..... Afsluitdijk (dam)
..... Alblasserwaard (general region)
..... Altena, Land van (general region)
..... Amstelland (general region)
..... Amsterdam Rijn Kanaal (canal)
..... Aruba (dependent state) [N]
..... Bernisse Molen (mill center)
..... Biesbos (general region)
..... Brouwersdam (dam)
..... Calandkanaal (canal)
..... Delfland (general region)

Principle 1: semantic annotation

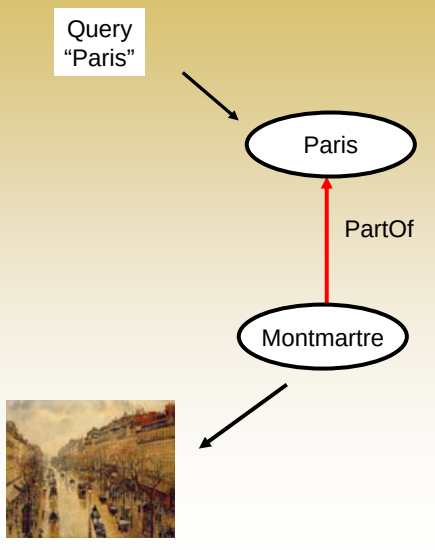
- Description of web objects with “concepts” from a shared vocabulary

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;

Principle 2: semantic search

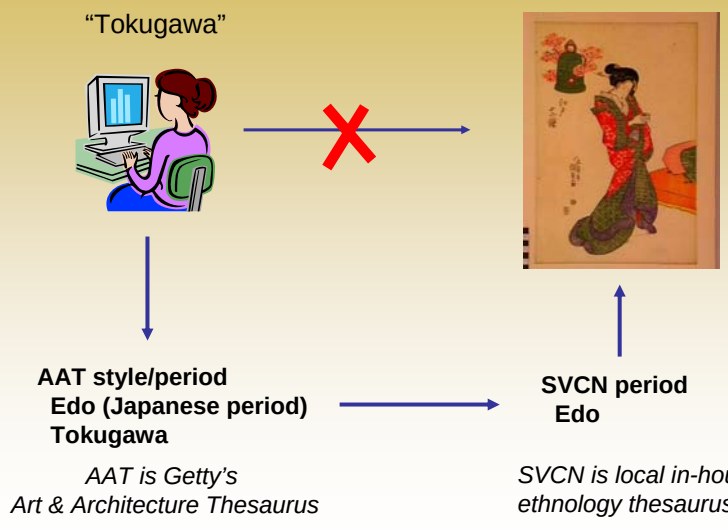
- Search for objects which are linked via concepts (semantic link)
- Use the type of semantic link to provide meaningful presentation of the search results



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

Principle 3: vocabulary alignment



A link between two thesauri

Edo

Description:

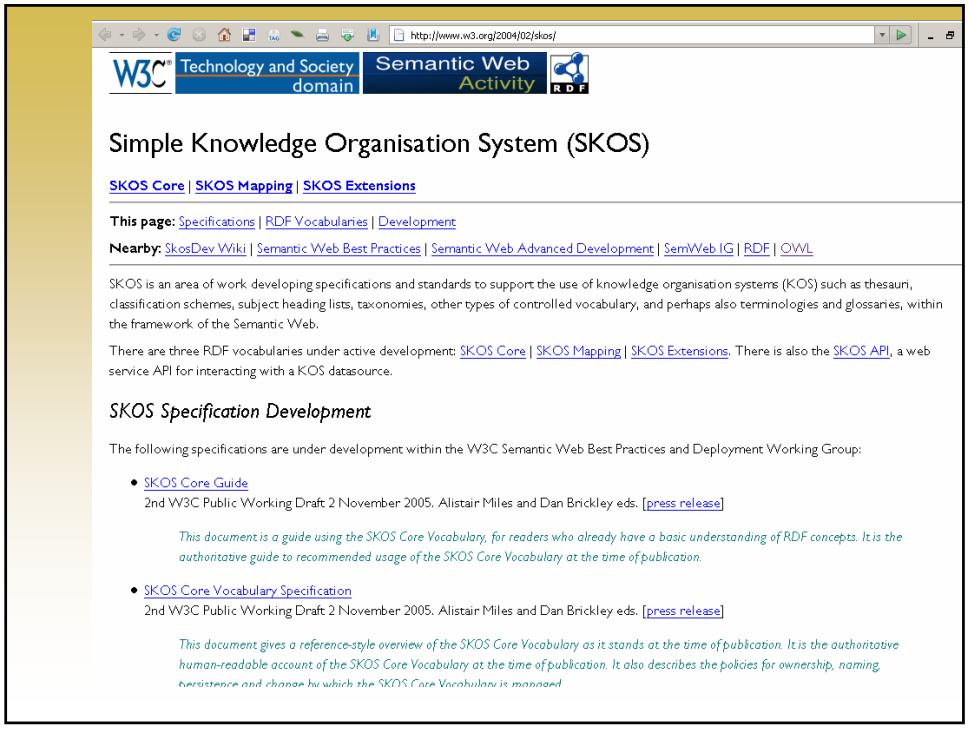
labelPreferred	Edo;Edo;
labelPreferred.Singular	Edo;
preferred parent	Japanse vroege moderne perioden;
source	RMV; RMV-WEB;
type	SVCN Concept;
label	OVM.AAB.AAA.AAB.AAE.AAB.AAB.AAA.AAE.AAA;
sameAs	Edo (Japanese period);

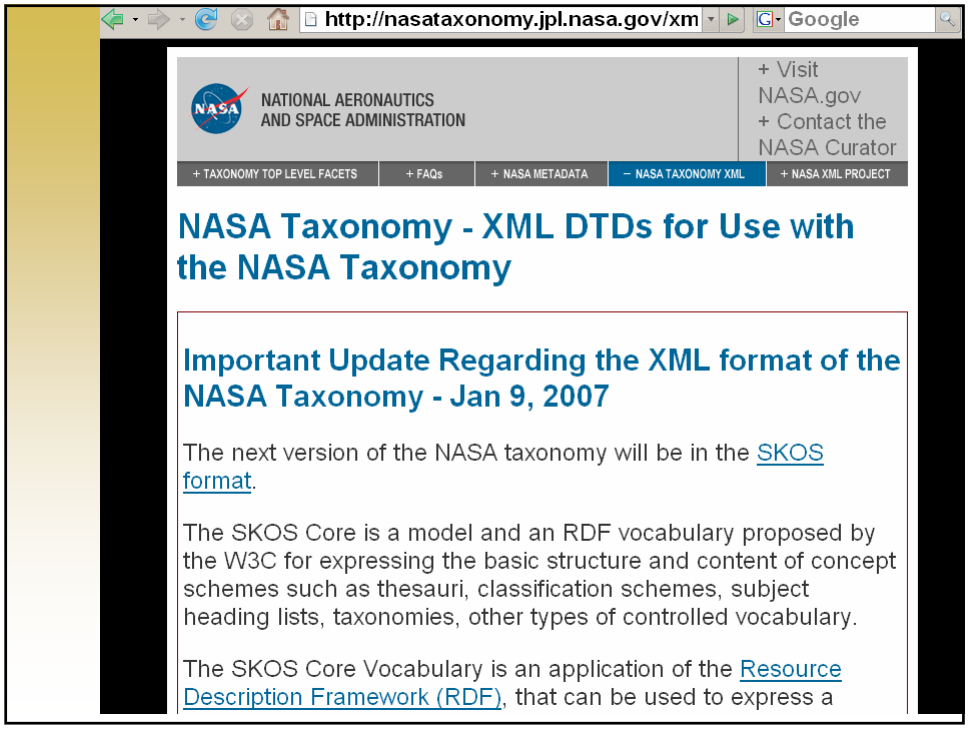
Used as value to describe other resources:




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


http://www.w3.org/2004/02/skos/
W3C
Technology and Society domain
Semantic Web Activity
RDF
Simple Knowledge Organisation System (SKOS)
SKOS Core | SKOS Mapping | SKOS Extensions
This page: Specifications | RDF Vocabularies | Development
Nearby: SkosDev Wiki | Semantic Web Best Practices | Semantic Web Advanced Development | SemWeb IG | RDF | OWL
SKOS is an area of work developing specifications and standards to support the use of knowledge organisation systems (KOS) such as thesauri, classification schemes, subject heading lists, taxonomies, other types of controlled vocabulary, and perhaps also terminologies and glossaries, within the framework of the Semantic Web.
There are three RDF vocabularies under active development: SKOS Core | SKOS Mapping | SKOS Extensions. There is also the SKOS API, a web service API for interacting with a KOS datasource.
SKOS Specification Development
The following specifications are under development within the W3C Semantic Web Best Practices and Deployment Working Group:
SKOS Core Guide
2nd W3C Public Working Draft 2 November 2005. Alistair Miles and Dan Brickley eds. [press release]
This document is a guide using the SKOS Core Vocabulary, for readers who already have a basic understanding of RDF concepts. It is the authoritative guide to recommended usage of the SKOS Core Vocabulary at the time of publication.
SKOS Core Vocabulary Specification
2nd W3C Public Working Draft 2 November 2005. Alistair Miles and Dan Brickley eds. [press release]
This document gives a reference-style overview of the SKOS Core Vocabulary as it stands at the time of publication. It is the authoritative human-readable account of the SKOS Core Vocabulary at the time of publication. It also describes the policies for ownership, naming, persistence and change by which the SKOS Core Vocabulary is managed

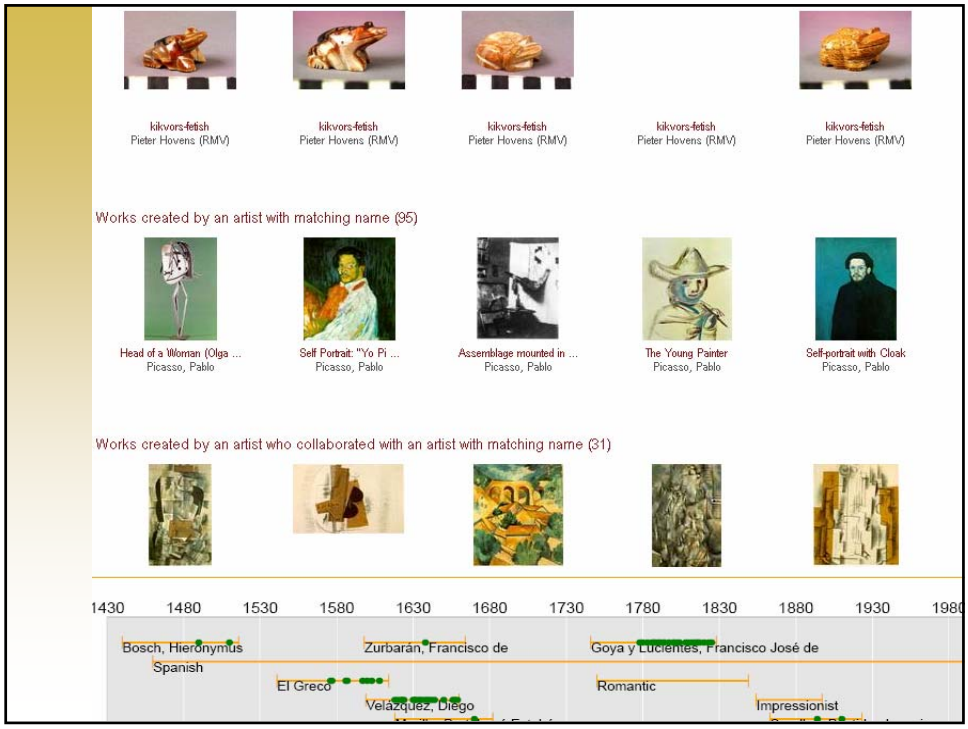

http://nasataxonomy.jpl.nasa.gov/xm
Google
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
+ Visit NASA.gov
+ Contact the NASA Curator
+ TAXONOMY TOP LEVEL FACETS
+ FAQs
+ NASA METADATA
– NASA TAXONOMY XML
+ NASA XML PROJECT
NASA Taxonomy - XML DTDs for Use with the NASA Taxonomy
Important Update Regarding the XML format of the NASA Taxonomy - Jan 9, 2007
The next version of the NASA taxonomy will be in the SKOS format.
The SKOS Core is a model and an RDF vocabulary proposed by the W3C for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, other types of controlled vocabulary.
The SKOS Core Vocabulary is an application of the Resource Description Framework (RDF), that can be used to express a

Cross collection search

http://e-culture.multimedian.nl

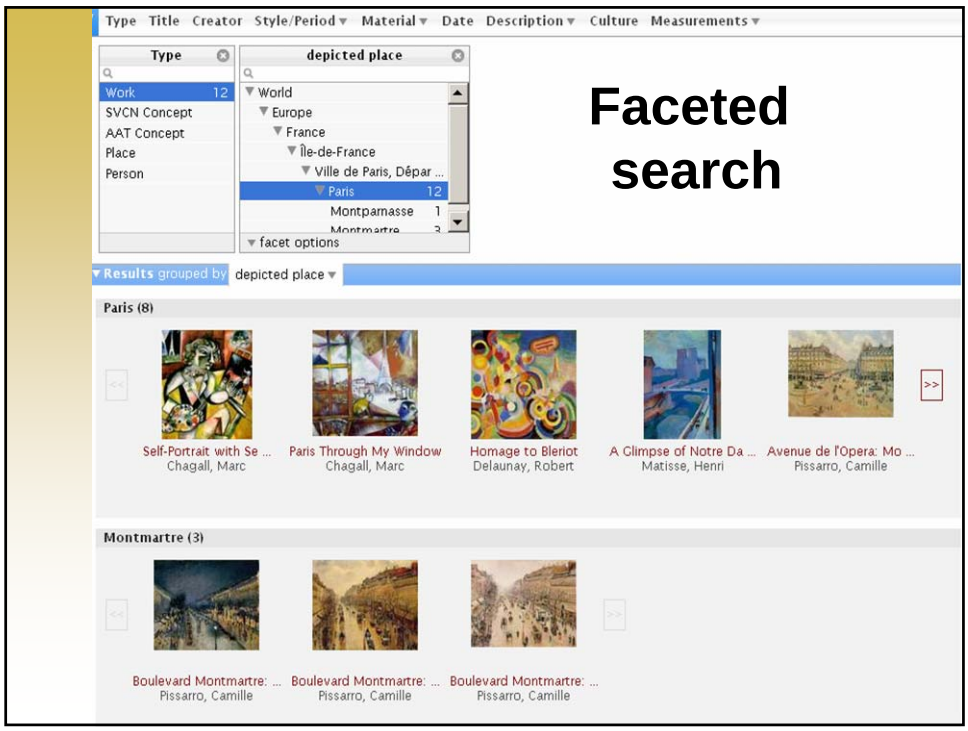
Term disambiguation is key issue in semantic search

- 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



Keyword search with semantic clustering

1. Btree of literals plus Porter stem and metaphone index
2. Find resources with matching labels
   - Default resources are “Work”s
3. Find related resources by one-way graph traversal
   - **owl:inverseOf** is used
   - Threshold used for constraining search
4. Cluster results (group instances)

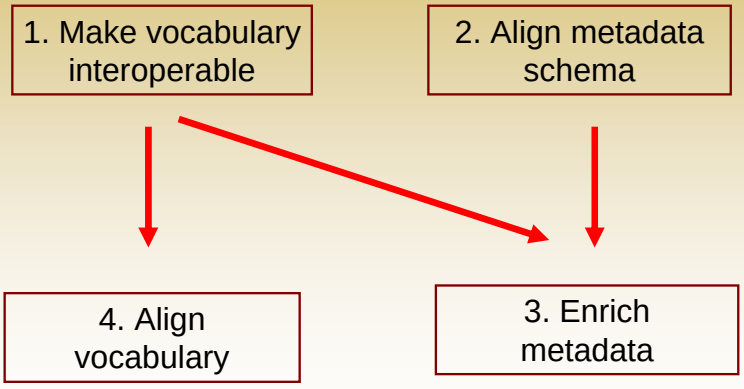


Faceted search

What do you need to do to make your collection part of a Semantic Culture Web?

Four activities

From metadata to semantic metadata



Activity 1: making vocabulary interoperabie

- Make your vocabularies available in the Web standard RDF
- Many organizations are already do this
- W3C provides the SKOS template to make this almost straightforward
- Effort required: at most a few days

Activity 2: aligning the metadata schema

- Specify your collection metadata scheme as a specialization of Dublin Core
- With RDF/OWL this is easy/trivial!
- Cf. DC Application Profiles

Aligning VRA with Dublin Core

- VRA is specialization of Dublin Core for visual resources
- VRA properties “material.medium” and “material.support” are specializations of Dublin Core property “format”

```
vra:material.medium rdfs:subPropertyOf
  dc:fotmat .
vra:material.medium rdfs:subPropertyOf
  dc:format .
```

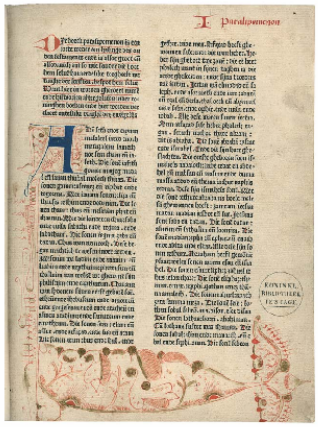
Activity 3: enriching the metadata

- Extracting additional concepts from an annotation
  - Matching the string “Paris” to a vocabulary term
- Information-extraction techniques exists (and continue to be developed)
- Effort required can be up to a few weeks
  - The more concepts, the better, but no need to be perfect!

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:OMSCHRIJVING_EN>
    <inm:AFMETINGEN>27 x 20 cm</inm:AFMETINGEN>
    ...
</inm:Record>
```

Resulting semantic annotation (rendered as HTML with RDFa)



Description:

classificatie	Geschiedenis van de boekdrukkunst;
drukker	Meer, Jacob Jacobszoon van der; Yemantszoon, Mauricius;
origineel	Bijbel. Oude Testament. - Delft: Jacob Jacobszoon van der Meer en Mauricius Yemantszoon, 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 Yemantszoon, 1477;
Type	tekstbladzijde;
type	Work;

Used as value to describe other resources:

BBB_169E56_1477_P1.JPG;	relation.depicts

Activity 4: aligning the vocabulary

- Find semantic links between vocabulary links
  - Derain (ULAN) related-to Fauve (AAT))
- Automatic techniques exists, but performance varies
- Often combination of automatic and manual alignment
- Effort strongly dependent on vocabularies
  - But “a little semantic goes a long way” (Hendler)

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

Extracting additional knowledge from scope notes

Neo-Expressionist

Description:

vp descriptiveNote	Refers to the style of art, predominantly of painting, inspired by German Expressionism that gained popularity in Italy, Germany, and America in the late 1970s and early 1980s. The style is characterized by large, figurative works, crudely and rapidly painted, often with objects imbedded in their surfaces, such as broken plates or straw.;
vp id	300022189;
vp labelNonPreferred	Energism;
vp labelPreferred	Neo-Expressionist;
vp preferred parent	aat:post-1945 fine arts styles and movements;

Perspectives

- Basic Semantic Web technology is ready for deployment
- Web 2.0 facilities fit well:
  - Involving community experts in annotation
  - Personalization, myArt
- Social barriers have to be overcome!
  - “open door” policy
  - Involvement of general public => issues of “quality”

Planned: annotation interfaces

- Brings in issues w.r.t. trust and access control
  - Quality issue
- Involving users in the annotation process
  - Stimulating people to contribute
- User profiling

Planned: large-scale deployment

- Moving to 100-150M triples, 30+ collections
- Scalability is key
  - Load balancing, 32+Gb RAM
  - Search algorithms
- Methodology for including new collections
- Continuous user and usability studies