Publishing vocabularies on the Semantic Web using SKOS - with examples from Cultural Heritage

SemAst 2009

Antoine Isaac

Vrije Universiteit Amsterdam, National Library of the Netherlands
aisaac@few.vu.nl
Preamble: acknowledgements

• Participants of the W3C Semantic Web Deployment Working Group

http://www.w3.org/2006/07/SWD/
Talk overview

- Problem statement
- SKOS: a guided tour
- Some more details
- And/or demos
Knowledge Organization Systems? (KOS)

- Domain-specific vocabularies
  - Medicine: UMLS, SNOMED, MESH, Galen
  - Art history: AAT, ULAN
  - Geography: TGN
  - Food: Agrovoc
  - Libraries: LCSH, DDC, UDC

- Generic vocabularies
  - Lexical vocabularies: WordNet
  - Country codes, ...
Example: CH Metadata

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

**Browse by subject**
using the Iconclass classification system

<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategory</th>
<th>Show images</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Abstract, Non-representational Art</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Religion and Magic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Nature</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>earth, world as celestial body</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25F</td>
<td>animals</td>
<td>show images &gt;25</td>
<td>1</td>
</tr>
<tr>
<td>25F3</td>
<td>birds</td>
<td>show images &gt;25</td>
<td>1</td>
</tr>
<tr>
<td>25F31</td>
<td>groups of birds</td>
<td>show images &lt; 5</td>
<td>1</td>
</tr>
<tr>
<td>25F32</td>
<td>song-birds</td>
<td>show images &gt;25</td>
<td>1</td>
</tr>
<tr>
<td>25F33</td>
<td>predatory birds</td>
<td>show images &gt;25</td>
<td>1</td>
</tr>
<tr>
<td>25F34</td>
<td>owls</td>
<td>show images &lt; 25</td>
<td>1</td>
</tr>
<tr>
<td>25F35</td>
<td>ornamental birds</td>
<td>show images &lt; 25</td>
<td>1</td>
</tr>
<tr>
<td>25F36</td>
<td>water-birds</td>
<td>show images &gt;25</td>
<td>1</td>
</tr>
<tr>
<td>25F37</td>
<td>shore-birds and wading-birds</td>
<td>show images &gt;25</td>
<td>1</td>
</tr>
<tr>
<td>25F38</td>
<td>walker and runner birds</td>
<td>show images &lt; 25</td>
<td>1</td>
</tr>
<tr>
<td>25F39</td>
<td>other birds</td>
<td>show images &gt;25</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Human Being, Man in General</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Society, Civilization, Culture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Abstract Ideas and Concepts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>History</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Bible</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Literature</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Classical Mythology and Ancient History</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Show only notations used in the manuscript database
Show full Iconclass hierarchy

**Search by keyword**

Search
Why are we interested in that now?

• Museums, libraries, archives

• Crucial technological evolutions
  – Digitization
  – Web

• Trend: (web) access to CH objects from different collections
Europeana.eu case

Providing access to 2 million objects from 50 institutions
CH metadata and SW

• To implement this, metadata is key

• Problems
  – how to access seamlessly different collections?
  – how to allow other applications to exploit (knowledge about) CH objects?
  – can solutions be flexible?

• *The Semantic Web can be useful for this*
Can we have that for the CH metadata?

http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData
SKOS Focus:
how to port KOSs to the SW?
SKOS

• Simple Knowledge Organization System
SKOS

• Observation: there are many KOS models/formats:
• But also common features, used by typical classes of applications

• SKOS is a model to represent KOSs on the SW in a *simple* way
  – Ontology for *concepts*
Not owl:Class?

• Ontologizing KOSs is possible, but:
  – they are large
  – and have loose semantics
    • Car wheel BroaderTerm Car

• Existing information can be ported as such
  – loose semantics can be useful for many applications!
    • Search, annotation
SKOS

- Concepts and Concept Schemes
- Lexical properties
- Semantic relations
- Notes
Thesaurus example

animals
  NT (narrower term) cats

cats
  UF (used for) domestic cats
  RT (related term) wildcats
  BT (broader term) animals
  SN (scope note) used only for domestic cats

domestic cats
  USE cats

wildcats

ISO 2788 model
SKOS concepts and labels

- SKOS is concept-oriented
  - Concepts are first-order resources
  - Labels are RDF literals
(Multilingual) labels

ex:animals
  skos:prefLabel "animals"@en
  skos:prefLabel "animaux"@fr

skos:Concept
  rdf:type
Semantic relations

cats

RT (related term) wildcats
BT (broader term) animals

ex:cats — skos:related — ex:wildcats

ex:animals — skos:broaden — ex:cats
Documenting concepts

```
skos:note
|-- skos:definition
|-- skos:scopeNote
|   |-- skos:example
|   |-- skos:historyNote
|   |-- skos:editorialNote
|   |-- skos:changeNote
```

ex:cats

"used only for domestic cats"
Example: SKOS graph

animals

NT cats

cats

UF domestic cats
RT wildcats
BT animals
SN used only for domestic cats
domestic cats
USE cats

wildcats

ex:animals

skos:prefLabel
"animals"@en

ex:cats

skos:broader
ex:animals

skos:narrower
ex:wildcats

skos:related
ex:animals

"cats"@en
skos:prefLabel
"used only for domestic cats"

"domestic cats"@en
skos:altLabel

"wildcats"@en
skos:prefLabel
Some other features

• Collections of concepts
  – skos:Collection, skos:OrderedCollection, skos:member

• Explicit representation of Concept Schemes
  – skos:ConceptScheme, skos:inScheme, skos:hasTopConcept

• Notations
  – skos:notation

• Mapping relations
  – coming now…
CH case: Libris

- http://libris.kb.se/
- Swedish Library as linked data
Linked descriptions of resources in Libris

Martin Malmsten, Dublin Core 2008
External links in Libris:
Library of Congress Subject Headings

Ed Summers et. al., Dublin Core 2008
Searching using multiple vocabularies
Semantic interoperability problem
Using semantic alignment
Networking controlled vocabularies in SKOS

animals
cats
wildcats

animal
human
object

ex1:animals

skos:exactMatch

ex1:cats

skos:broadMatch

ex2:animal
Abstract

As the astronomical information processed within the Virtual Observatory becomes more complex, there is an increasing need for a more flexible representation of quantities, concepts, and processes not confined to things easily placed in a FITS image, or expressed in a catalogue or a table. This document describes a format for vocabularies based on the W3C’s Resource Description Framework (RDF) and Simple Knowledge Organization System (SKOS).
Mass adoption!

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.

http://nasataxononomy.jpl.nasa.gov/
WHAT IS A TAXONOMY?

A taxonomy is a standards-based classification scheme used to organize electronic content.

HOW ARE TAXONOMIES USED?

Taxonomies can improve many aspects of on line information management.

They can be used to:

- make search more robust
- facilitate data interoperability
- allow business analysis across content in disparate repositories
- support records management for long term archiving
- enable role-based content delivery for portals
- kick off work flows for process based information transactions

HOW IS THE NASA TAXONOMY USED?

The NASA taxonomy provides first steps towards the unification of the NASA information space by documenting a high level set of terms that can be used for mapping together varying data structures. Reconciliation of terms and topics is essential to understanding NASA discoveries in a larger context. Find out more in our Resources section.

- Current version of the NASA Taxonomy
- Taxonomy Facets (or branches)
- Terms, Definitions, Synonyms, Relationships

- For Developers
- Current NASA Core Metadata Specification
- DTD Files
- SKOS Files
- Editorial Style Guide

- Maintenance
Questions?

- Tricky bits
- Demos
About some SKOS modeling choices

- Model constructs
- Formal semantics

- How much interoperability does porting to SKOS really allow?
  - Are there different ways to convert similar things?
  - Different interpretations of SKOS constructs?
  - Things impossible to convert?
Preamble: W3C standardization process

- Input: draft specification
- Collect use cases & derive requirements
- Create issues list: requirements that cannot be handled by the draft spec
- Propose resolutions for issues
- Get consensus on new spec
- Find two independent implementations for each feature in the spec
- Continuously: ask for public feedback/comments

Guus Schreiber
SKOS Use Cases and Requirements

W3C Working Draft 16 May 2007

This version:
http://www.w3.org/TR/2007/WD-skos-ucr-20070516/

Latest version:
http://www.w3.org/TR/skos-ucr

Previous version:
This is the first public Working Draft

Editors:
Antoine Isaac, Vrije Universiteit Amsterdam, aisaac@few.vu.nl
Jon Phipps, Cornell University, jphipps@madcreek.com
Daniel Rubin, Stanford Medical Informatics, dlrubin@stanford.edu
Example use case and requirement

• 2.3 Use Case #3 — Semantic search service across mapped multilingual thesauri in the agriculture domain

“This application coming from the AIMS project […] includes some more specific links […] String-to-String relationships …”

<table>
<thead>
<tr>
<th>acronym</th>
<th>Food and Agriculture Organization</th>
<th>FAO</th>
</tr>
</thead>
<tbody>
<tr>
<td>spelling_variant</td>
<td>organisation</td>
<td>organization</td>
</tr>
<tr>
<td>translation</td>
<td>vache</td>
<td>cow</td>
</tr>
</tbody>
</table>

“Requires: […] R-RelationshipsBetweenLabels”
Example issue: relationships between lexical labels

“R-RelationshipsBetweenLabels
  Representation of links between labels associated to concepts

• In previous SKOS spec, labels are literals
• Literals cannot be subject of an RDF property
Example issue: relationships between lexical labels

In the end we did it, but as an extension

prefix ex: <http://www.example.com/concepts#>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
Semantics of broader

- Is **skos:broader** "transitive"?
- It can be wrong, sometimes!
- **skos:broader** is not transitive in general
Semantics of broader

**skos:broader** has a super-property **skos:broaderTransitive** with semantics of “has ancestors”

1. every **broader** implies a **broaderTransitive**
2. **broaderTransitive** is transitive!
Semantics of broader

• skos:broader is not (a sub-property of) rdfs:subClassOf!

• Problem: people having KOSs with transitive hierarchies may just assert skos:broaderTransitive statements
Other semantics of relations

• **broad**er and **narrow**er are inverse of each other

• **related** is symmetric
Semantics of SKOS

• This tells what should be explicit or not in a SKOS conversion, and what can (shall) be inferred from it
  – Important for building applications using SKOS

• Beware: this sometimes requires reasoning!
Is that damn thing useful?

• At least it's there!
  – A proposed standard to represent KOS on the SW

• It allows to publish KOSs
  – Simple, with minimal commitment
  – For most KOS features, conversion is smooth

• It allows to develop applications with re-usable & interoperable components
  – It can also be extended
Conclusion

Despite some issues, SKOS contributes to enhance interoperability of KOSs
Thanks!

- Reminder: comments highly welcome on SKOS
  SKOS Reference: http://www.w3.org/TR/skos-reference
  SKOS Primer: http://www.w3.org/TR/skos-primer

- Some pointers:
  SKOS: www.w3.org/2004/02/skos/
  Europeana: www.europeana.eu
  Libris: libris.kb.se
  eCulture (semantic search): eculture.cs.vu.nl
  STITCH (vocabulary alignment and repository):
    stitch.cs.vu.nl/demo.html
SKOS Demos: searching, browsing and repositories

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