Publishing Vocabularies on the Web

Guus Schreiber
Antoine Isaac
Vrije Universiteit Amsterdam
Acknowledgements

- Alistair Miles, Dan Brickley, Mark van Assem, Jan Wielemaker, Bob Wielinga
- Participants of the W3C Semantic Web Best Practices and the Semantic Web Deployment Working Groups
Overview

- Issues in conversion to RDF/OWL
  - Example: Union List of Artist Names (ULAN)
  - Example: WordNet 2.0
- Work within the W3C Semantic Web Deployment Working Group
  - SKOS model for thesauri
  - Recipes for Web access to published vocabularies
  - RDFa: embedding RDF metadata in HTML
Thesauri / vocabularies

- Controlled vocabularies
  - *Thesauri, classification schemes, taxonomies, subject heading lists, authority lists…*

- Large bodies of knowledge that represent consensus in particular domains

- Often lots of implicit semantics available

- Semantic Web Challenge showed that thesauri are important resources for SW applications

- Representation is typically relational database and/or XML
Example thesauri

- 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, FrameNet
  - Currencies, country codes, …
ISO standard for representing thesauri

- **Term**
  - Preferred term (USE)
  - Non-preferred term (USED FOR)

- **Hierarchical relation between terms**
  - Broader/narrower term (BT/NT)
    - Generic
    - Partitive

- **Association between terms (RT)**
Typical conversion process

- Two steps
  - Step 1: “As is” conversion
    - Keep original names/constructs
    - Make implicit semantics explicit (not trivial!)
    - Decisions on whether to keep all information
  - Step 2: adding semantics
    - Separate file(s)
    - Interpretation of thesauri features, e.g. hyponym relation as `rdfs:subClassOf`
    - May require (lots of) additional research
Example thesaurus: ULAN

- 300,000 “Subject” records (artists and art institutions)
  - with biographical information (place/time birth/death)
  - and relations to other artists (student-of, …)
- Large XML file with all data
- Basic representation:
  - association links between subjects
  - preferred/non-preferred terms relations between subjects and terms
Click the icon to view the hierarchy.

**ID: 500000351**

**Koninck, Philips de** (Dutch painter and draftsman, 1619-1688)

**Note:** History and portrait painter who is today most well-known for his naturalistic panoramic bird’s-eye view landscapes.

**Birth and Death Places:**
- **Born:** Amsterdam (North Holland, Netherlands) (inhabited place)
- **Died:** Amsterdam (North Holland, Netherlands) (inhabited place)

**Related People or Corporate Bodies:**
- related to (familial) .... Koninck, Salomon (Dutch painter, printmaker, and draftsman, 1609-1656) [500027532]
- sibling of .... Koninck, Jacob, the elder (Dutch painter and engraver, ca. 1616-1708) [500024292]
- student of .... Rembrandt van Rijn (Dutch painter, draftsman, and printmaker, 1606-1669) [500011051]
<Associative_Relationships>
  <Associative_Relationship>
    <Historic_Flag>NA</Historic_Flag>
    <Relationship_Type>
      1102/student of
    </Relationship_Type>
    <Related_Subject_ID>
      <VP_Subject_ID>500011051</VP_Subject_ID>
    </Related_Subject_ID>
  </Associative_Relationship>
</Associative_Relationships>
Conversion issues

- XML and RDF/OWL are inherently different
  - XML = thesaurus document structure
  - RDF = thesaurus document content
- Redundant/meaningless information in XML file
  
  `<Associative_Relationships>
  <Historic_Flag>NA</Historic_Flag>
  </Associative_Relationships>`
- How to represent “student of”?
  - Subproperty of `Associative_Relationship` is probably preferred
  - Needs to be derived from the data; not part of schema
XML fragment of ULAN: terms

<Non-Preferred_Term>
  <Term_Text>Koning, Philips Aertsz. de</Term_Text>
  <Term_ID>1500207734</Term_ID>
  <Display_Order>34</Display_Order>
  <Vernacular>Vernacular</Vernacular>
</Non-Preferred_Term>
Conversion issues

- Do we include all information in the conversion?
  - Display order
- Should each term have a URI?
- Making language explicit
  - “vernacular” means the string is written in the original language
  - Multi-linguality is an important issue for thesauri
RDF/OWL Representation of WordNet

W3C Working Draft 19 June 2006

This version:
http://www.w3.org/TR/2006/WD-wordnet-rdf-20060619/

Latest version:
http://www.w3.org/TR/wordnet-rdf/

Previous version:
This is the first published version

Editors:
Mark van Assem, Vrije Universiteit Amsterdam
Aldo Gangemi, ISTC-CNR, Rome
a depression forming the ground under a body of water; "he searched for treasure on the ocean bed"
WordNet: internal representation

<table>
<thead>
<tr>
<th>SynsetID</th>
<th>Order</th>
<th>LexForm</th>
<th>Type</th>
<th>SenseNum</th>
</tr>
</thead>
<tbody>
<tr>
<td>s(108644031,1)</td>
<td>1</td>
<td>'bed'</td>
<td>n,3</td>
<td>2</td>
</tr>
<tr>
<td>s(108644031,2)</td>
<td>2</td>
<td>'bottom'</td>
<td>n,5</td>
<td>1</td>
</tr>
<tr>
<td>s(102719813,1)</td>
<td>1</td>
<td>'bed'</td>
<td>n,1</td>
<td>51</td>
</tr>
</tbody>
</table>

g(108644031,'(a depression forming the ground under a body of water; "he searched for treasure on the ocean bed")').
g(102719813,'(a piece of furniture that provides a place to sleep; "he sat on the edge of the bed"; "the room had only a bed and chair")').
WordNet URIs

- What URIs should be chosen?
  - SynSet, WordSense, Word

- URI name:
  - ID? => difficult for human interpretation
  - Human-readable concatenation

```plaintext
wn:synset-bank-noun-2
synset denoted by second sense of “bank”
```

```plaintext
wn:wordsense-bank-noun-1
wn:word-bank
```
Implicit WordNet semantics

“The ent operator specifies that the second synset is an entailment of first synset. This relation only holds for verbs.”

- Example: [breathe, inhale] entails [sneeze, exhale]
- Semantics (OWL statements):
  - Transitive property
  - Inverse property: entailedBy
  - Value restrictions for VerbSynset (subclass of Synset)
Data access

- `<rdf:RDF>`
  - `<rdf:Description about="http://www.w3.org/2006/03/wn/wn20/instances/synset-bank-noun-2">`
    - `<wn20schema:synsetld>108639924</wn20schema:synsetld>`
    - `<rdfs:label>bank</rdfs:label>`
    - `<rdf:type rdf:resource="http://www.w3.org/2006/03/wn/wn20/schema/NounSynset"/>`
    - `<wn20schema:gloss>`
      (sloping land (especially the slope beside a body of water); "they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents")
    - `<wn20schema:gloss>`
  - `<wn20schema:hyponymOf rdf:resource="http://www.w3.org/2006/03/wn/wn20/instances/synset-slope-noun-1"/>`
   `</rdf:Description>`
- `</rdf:RDF>`

- Query for WordNet URI returns “concept-bounded description”
Overview

- Issues in conversion to RDF/OWL
  - Example: Union List of Artist Names (ULAN)
  - Example: WordNet 2.0

- Work within the W3C Semantic Web Deployment Working Group
  - SKOS model for thesauri
  - Recipes for Web access to published vocabularies
  - RDFa: embedding RDF metadata in HTML
W3C Semantic Web Deployment Working Group

Making vocabularies/thesauri/ontologies available on the Web

http://www.w3.org/2006/07/SWD/
SWD goals

- Schema for interoperable RDF/OWL representation of vocabularies
  - SKOS
- Publication guidelines
  - URI management, representation of versions
- Embedding RDF in (X)HTML pages
  - RDFa
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
Multi-lingual labels for concepts

ex:shrubs

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

- skos:note
  - skos:definition
    - skos:scopeNote
    - skos:example
    - skos:historyNote
  - skos:editorialNote
  - skos:changeNote
Semantic relation: broader and narrower

prefix ex: <http://www.example.com/concepts#>  
prefix skos: <http://www.w3.org/2004/02/skos/core#>
Semantic relations: related

prefix ex: <http://www.example.com/concepts#>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
Collections: role-type trees

milk
  <$milk by source animal$>
  .. buffalo milk
  .. cow milk
  .. goat milk
  .. sheep milk
Adding semantics

- Adding OWL statements
  - `skos:related` rdf:type `owl:SymmetricProperty`
  - `skos:broader` owl:inverseOf `skos:narrower`

- Inference rules
  - Collection membership rule
    
    $(?s \ skos:narrower \ ?c) \ (\ ?c \ skos:member \ ?t)$

  $\rightarrow \ (?s \ skos:narrower \ ?t)$

- Interpreting thesaurus relations such as `broader` as `subClassOf` can be useful but is often imprecise
SKOS semantics: concepts are not the real things
Indexing a resource with a SKOS concept


prefix ex: <http://www.example.com/concepts#>
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix foaf: <http://xmlns.com/foaf/0.1/>
Semantic alignment links

- Learning relations between thesauri is an important form of additional semantics
  - Example: AAT contains styles; ULAN contains artists, but there is no link
  - Availability of this kind of alignment knowledge is extremely useful
  - Cf. demo

Warning: unstable part of SKOS!
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 amended spec
- Find two independent implementations for each feature in the spec
- Continuously: ask for public feedback/comments

(YES, YOU!)
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”
“R-RelationshipsBetweenLabels

Representation of links between labels associated to concepts

The SKOS model shall provide means to represent relationships between the terms associated with concepts. Typical examples are [...]”

- In current SKOS spec labels are represented as literals
- This is a problem because literals have no URI, so cannot be subject of an RDF property
- Possible resolutions:
  - Labels/terms as instances of a new class
  - Relaxing constraints on label property
Example issue: relationships between lexical labels

\[ 
\text{ex:shrubs} \quad \text{skosext:translation} \quad ? 
\]
SWD goals

- Schema for interoperable RDF/OWL representation of vocabularies
  - SKOS
- Publication guidelines
  - URI management, representation of versions
- Embedding RDF in (X)HTML pages
  - RDFa
Recipes for vocabulary URIs

- Simplified rule:
  - Use “hash” variant for vocabularies that are relatively small and require frequent access
    http://www.w3.org/2004/02/skos/core#Concept
  
  - Use “slash” variant for large vocabularies, where you do not want always the whole vocabulary to be retrieved
    http://www.w3.org/[[...]/instances/synset-bank-noun2
Data access

- `<rdf:RDF>`
  - `<rdf:Description about="http://www.w3.org/2006/03/wn/wn20/instances/synset-bank-noun-2">`
    - `<wn20:schema:synsetId>108639924</wn20:schema:synsetId>`
    - `<rdfs:label>bank</rdfs:label>`
    - `<rdfs:type rdf:resource="http://www.w3.org/2006/03/wn/wn20/schema/NounSynset"/>`
  - `<wn20:schema:gloss>`
    - (sloping land (especially the slope beside a body of water); "they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents")
  - `<wn20:schema:hyponymOf rdf:resource="http://www.w3.org/2006/03/wn/wn20/instances/synset-slope-noun-1"/>`
- `</rdf:Description>`
- `</rdf:RDF>`

- Query for WordNet URI returns “concept-bounded description”
Recipes for serving RDF

- Persistent URIs and version-specific content
  
  **HTTP 303 redirection**
  - Client asking http://example.org/voc#myClass
  - Client redirected to
    
    http://example.org/voc-files/voc-version3.rdf#myClass

- For more information and other recipes, see:
  
  http://www.w3.org/TR/swbp-vocab-pub/
SWD goals

- Schema for interoperable RDF/OWL representation of vocabularies
  - SKOS

- Publication guidelines
  - URI management, representation of versions

- Embedding RDF in (X)HTML pages
  - RDFa
A RDFa sample

Regular HTML

<h1>Photo Album #12345: Vacation in the South of France</h1>
<h2>created by Mark Birbeck</h2>

HTML with RDFa

<h1 property="dc:title">Vacation in the South of France</h1>
<h2>created by <span property="dc:creator">Mark Birbeck</span></h2>

Resulting RDF statements

<> dc:title "Vacation in the South of France"^^XMLLiteral .
<> dc:creator "Mark Birbeck"^^XMLLiteral .
Linking to other resources

Regular HTML

This document is licensed under a
<a href="http://creativecommons.org/licenses/by-nc/2.5/">
  Creative Commons Non-Commercial License
</a>.

HTML with embedded RDF

This document is licensed under a
<a rel="cc:license"
  href="http://creativecommons.org/licenses/by-nc/2.5/">
  Creative Commons Non-Commercial License
</a>.
Statements about other resources:
photo example

<li><img src="/user/markb/photo/23456" />,
   <span about="/user/markb/photo/23456" property="dc:title">
   Sunset in Nice
   </span></li>

<li><img src="/user/markb/photo/34567" />,
   <span about="/user/markb/photo/34567" property="dc:title">
   W3C Meeting in Mandelieu
   </span></li>
RDFa demo

- Having time, feeling lucky and online?
- Slides
More information

Semantic Web Deployment Working Group

The mission of this Working Group is to provide guidance in the form of W3C Technical Reports on issues of practical RDF development and deployment practices in the areas of publishing vocabularies, OWL usage, and integrating RDF with HTML documents.

Also On This Page — Charter and History • Working Group Resources

Mailing list and wiki

- Mailing list archives: public-swd-wg
- Wiki

Related Working Drafts:

- Best Practice Recipes for Publishing RDF Vocabularies
  W3C Working Draft 14 March 2006, Miles, Baker, Swick (eds.)
- RDF/A Primer 1.0: Embedding RDF in XHTML
  W3C Working Draft 10 March 2006, Adida, Birbeck (eds.)
- SKOS Core Vocabulary Specification
  W3C Working Draft 2 November 2005, Miles, Brickley (eds.)
- SKOS Core Guide
  W3C Working Draft 2 November 2005, Miles, Brickley (eds.)
- Metainformation Module and Metainformation Attributes Module of XHTML 2.0
  W3C Working Draft 26 July 2006

Current and Upcoming Events

- W3C teleconferences on Tuesdays at 1500 UTC
- 4-10 November 2007: Advisory Committee and Technical Plenary combined, Cambridge, MA (USA). See announcement.
- more W3C Events

Nearby

- Deliverables and Schedule
- Mailing list archives: public-swd-wg
- Wiki
- issues tracker
- Notes on scribing
Thanks

- Reminder: we ask for feedback!
  - Questions and comments highly welcome

- aisaac at few.vu.nl
- schreiber at cs.vu.nl

- Continue for demo?
SKOS Demo: browsing and alignment

- Feeling lucky and online?

Back
**Demo: SKOS, browsing and alignment**

### BNF KB Integrated Search (single view Iconclass,mandraNewNONE mapping)

Images from the *Illuminated Manuscripts Collection of the National Library of the Netherlands* combined with the *Bibliothèque nationale de France Mandragore Collection*

version 0.20 (serql=sesame)

**MAPPING:** mandraBASE mandraNewNONE

**VIEWS:** COMBINED VIEW SINGLE VIEW IC SINGLE VIEW IC(Eng) SINGLE VIEW IC(Ger) SINGLE VIEW MANDRA

**Collection**

<table>
<thead>
<tr>
<th>Collection</th>
<th>IconClass(Eng)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bibliothèque nationale de France Mandragore Collection (2170)</td>
<td>LOCAL (2310)</td>
</tr>
<tr>
<td></td>
<td>Abstract Ideas and Concepts (85)</td>
</tr>
<tr>
<td></td>
<td>Bible (1819)</td>
</tr>
<tr>
<td></td>
<td>Classical Mythology and Ancient History (393)</td>
</tr>
<tr>
<td></td>
<td>History (200)</td>
</tr>
<tr>
<td>Illuminated Manuscripts Collection of the National Library of the Netherlands</td>
<td>Human Being, Man in General (894)</td>
</tr>
<tr>
<td></td>
<td>Literature (56)</td>
</tr>
<tr>
<td></td>
<td>Nature (954)</td>
</tr>
<tr>
<td></td>
<td>Religion and Magic (1679)</td>
</tr>
<tr>
<td></td>
<td>Society, Civilization, Culture (1952)</td>
</tr>
</tbody>
</table>

**Subject vocabulary, collection 1**
Demo: SKOS, browsing and alignment

Hierarchical path from root to selected subject

Possible specialization for selected subject

version 0.20 (serql-sesame)

MAPPING: mandraBASE mandraNewNONE
Refine your search further within these categories:

<table>
<thead>
<tr>
<th>Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illuminated Manuscripts</td>
</tr>
<tr>
<td>Collection of the National Library of the Netherlands (3)</td>
</tr>
</tbody>
</table>

IconClass(Eng): Nature > earth, world as celestial body > animals > amphibians

Found 3 objects

Possible specialization for selected subject
Demo: SKOS, browsing and alignment

Semantic alignment of subjects activated

Found 11 objects

Document from Collection 2

version 0.20 (serql=sesame)

MAPPING: mandraBASE mandraNewNONE

Refine your search further within these categories:

<table>
<thead>
<tr>
<th>Collection</th>
<th>Illuminated Manuscripts Collection of the National Library of the Netherlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bibliothèque nationale de France Mandragore Collection</td>
<td>(6)</td>
</tr>
<tr>
<td>IconClass(Eng): all &gt; Nature &gt; earth, world as celestial body &gt; animals &gt; amphibians</td>
<td></td>
</tr>
<tr>
<td>LOCAL</td>
<td>tailless amphibians</td>
</tr>
</tbody>
</table>

These terms define your current search. Click the X to remove a term.

IconClass(Eng): Nature > earth, world as celestial body > animals > amphibians
**Demo: SKOS, browsing and alignment**

<table>
<thead>
<tr>
<th>Title:</th>
<th>Subject from voc2 aligned to voc1:amphibians</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fable : le cobra et les grenouilles</td>
<td></td>
</tr>
<tr>
<td>Picture:</td>
<td><img src="http://visualiseur.bnf.fr/Visualiseur?Destination=Mandrag" alt="Image of a book page with a cobra illustration" /></td>
</tr>
<tr>
<td>MANDRAGORE:</td>
<td>plante</td>
</tr>
<tr>
<td>MANDRAGORE:</td>
<td>naja</td>
</tr>
<tr>
<td>MANDRAGORE:</td>
<td>grenouille</td>
</tr>
<tr>
<td>MANDRAGORE:</td>
<td>fable</td>
</tr>
<tr>
<td>MANDRAGORE:</td>
<td>arbre</td>
</tr>
<tr>
<td>DATE:</td>
<td>13e siècle</td>
</tr>
</tbody>
</table>
RDFa demo: a page with RDFa

Academic

- research publications,
- invited presentation slides, or
- teaching experience.

If you’re looking for more "official" information, you can also find my

- CV (January 2007 version).
- bio (September 2006 version).

Various Projects

- StopBadware Working Group Member
- Creative Commons Technology Advisory Board member, and representative to the W3C.
  Chair of the RDF-in-XHTML Task Force.
- Berkman Center for Internet and Society affiliate
Academic

- research publications,
- invited presentation slides, or
- teaching experience.

If you're looking for more "official" information, you can also find my

- CV (January 2007 version).
- bio (September 2006 version).

Various Projects

- StopBadware Working Group Member
- Creative Commons Technology Advisory Board member, and
  representative to the W3C.
  Chair of the RDF-in-XHTML Task Force.
RDFa demo: displaying triples

Academic

- **research publications**, invited presentation slides, or teaching experience.

If you’re looking for more "official" information, you can also find my

- **CV** (January 2007 version).
- **bio** (September 2006 version).

RDFa Triples

<foaf:currentProject> <http://creativecommons.org>  
- **StopBadware** Working Group Member  
- **Creative Commons** Technology Advisory Board member, and representative to the W3C.  
  Chair of the **RDF-in-XHTML** Task Force.