How to deal with evolving vocabulary when it comes to linking ...?

Gerard Coen, Ronald Siebes, Andrea Scharnhorst

@dansknaw, @gerardcoen, ORCID: 0000-0001-9915-9721

October 1st, GO INTER IN KoM, Cologne
Topics

- Linked Open Data & Knowledge Organization Systems
- Versioning Linked Data
- The DANS KOS Observatory
  - Linked Data & Community Maturity
  - Sustainability of KOS
- Future perspectives

Gerard Coen
Project Leader

Data Archiving & Networked Services (DANS)

- KOS Observatory Project
- FAIRsFAIR
- EOSC Synergy
Linked Open Data - Ten Things

Learning
Understand and practice the Semantic Web and LOD basics.

Exploring
Inventory of your data.

Defining
Define the URI (Uniform Resource Identifier) naming strategy.

Resolving
Consider resolvability when a person or machine visits the URI.

Transforming
Generate the URIs for the selected concepts and relations according to the URI naming strategy.

Mapping
Map your Linked Data from your newly defined namespace to similar concepts and relations within the LOD.

Enriching
Enrich your data with information from the LOD.

Exposing
Define how people can get access to your LD: a data-dump, a SPARQL endpoint or a Web API.

Promoting
Publish and disseminate the value of your data via visualisations and workflows.

Sustaining
Ensure sustainability of your data.

- What is conceptual? Written in the text.
- What is technical? Where to go to find the tools, or resources.

https://librarycarpentry.org/Top-10-FAIR/2019/09/05/linked-open-data/
KOS Registries

Metadata Standard

Heritage Data
Linked Data Vocabularies for Cultural Heritage

EU Vocabularies

VEST / AgroPortal
MAP OF STANDARDS

BioPortal

Finto

Library of Congress

DANS

Data Archiving and Networked Services
What are KOS?

KOS are characterized by different structures and functions, and are also used as Knowledge Representation Resources (KRR). Examples include: ontologies, controlled vocabularies, subject headings, thesauri, classification schemes etc.

- Knowledge organization systems are used to organize materials for the purpose of retrieval and to manage a collection (Hodge, 2000)

- A semantic roadmap for navigation and orientation through relationships and concepts (Zeng, 2008)

- To find information and make sense of it (Soergel, 2009)

- A knowledge organization system is a scheme that models a structure (i.e. elements and mutual relationships) of an organized set of knowledge (Bratkova & Kucerova, 2014)
Versioning

Version information should be available at the **KOS level** (e.g. version release information) but also at the **Data level** in the case that concepts/relationships are modified over time.

Provenance information can also document what has influenced a resource to be modified, when, by whom, and how changes have been made.

**Use cases for extended change information:**

- People maintaining a derived/adapted subset of a KOS
- People maintaining mappings to other vocabularies, and applications supporting them (using terms from an external ontology)
- Automatic or semi-automatic indexing applications which make use of the KOS and/or its mappings
  - Search applications which make use of the KOS and/or its mappings
- Data Stewards wanting to learn about new and deprecated concepts
- Data Stewards (and supporting applications) re-indexing large sets of documents

Lookup (Ideal Scenario)

User/Machine follows a URI with a version tag for an older version.

Lookup service returns the requested version

http://ourxiv.com/resource/v1.1/astromy/pluto

Other Deprecated Version

Provenance

Lookup service also returns versions where changes were made to the concept

http://ourxiv.com/resource/v1.2/astromy/pluto

Requested Version

Provenance

http://ourxiv.com/resource/v1.3/astromy/pluto

Default Version

Provenance

Lookup service also returns the default version - incase the requestor is not aware of the update.
## The DANS KOS Observatory

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Schema Name/Title</th>
<th>Phenomena included</th>
<th>Maintenance organization</th>
<th>Place: Publisher, date.</th>
<th>Earlier versions (editions)</th>
<th>History of versioning</th>
<th>Version Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOS.HSS.35</td>
<td>MDA archaeological objects thesaurus</td>
<td></td>
<td>Institut für Museumsforschung (GETTY Art History Information)</td>
<td>n/a</td>
<td>Version 0.9 B</td>
<td><em>This beta release is a draft for your review and evaluation before its final release in February 1996—p. [3] of Quick reference card.</em>*</td>
<td></td>
</tr>
<tr>
<td>KOS.HSS.35.1</td>
<td>Archaeological Objects</td>
<td></td>
<td>Frau Prof. Hagedorn-Saupe</td>
<td>n/a</td>
<td>Version 1.0</td>
<td>Accompanies special issue of Visual resources, v. 11, no. 3-4, 1996.</td>
<td></td>
</tr>
<tr>
<td>KOS.HSS.36</td>
<td>Archaeological Sciences</td>
<td></td>
<td>El Centro de Documentación (GETTY Art History Information)</td>
<td>n/a</td>
<td>Version 0.9 B</td>
<td><strong>Revised March 2014 by Patricia Harping,</strong> earlier versions issued in print.</td>
<td></td>
</tr>
<tr>
<td>KOS.HSS.38</td>
<td>Cargo Thesaurus</td>
<td></td>
<td>Instituto centrale per il catalogo dei musei</td>
<td>Online Version</td>
<td>Revised 2016</td>
<td>Murtha Baca, and Patricia Harping, Editors.</td>
<td></td>
</tr>
<tr>
<td>KOS.HSS.39</td>
<td>Components Thesaurus</td>
<td></td>
<td>Nomenclature Committee, th</td>
<td></td>
<td>Version 1.0</td>
<td>The current version 4.0, that was released in 2007, is expressed as an XML schema in order to support the interoperability and exchange of VRA Core records. The unrestricted schema imposes no requirements on the values entered into any of the elements, sub-elements, or attributes, and may be useful for those who want to exchange legacy data.</td>
<td></td>
</tr>
<tr>
<td>KOS.HSS.40</td>
<td>Defence of Britain Thesaurus</td>
<td></td>
<td>Nomenclature Committee, th</td>
<td></td>
<td>Version 4.0</td>
<td>Core Schemas, Unrestricted Version</td>
<td></td>
</tr>
<tr>
<td>KOS.HSS.41</td>
<td>Event Thesaurus</td>
<td></td>
<td>Nomenclature Committee, th</td>
<td></td>
<td>Version 3.0</td>
<td>The current version 4.0, that was released in 2007, is expressed as an XML schema in order to support the interoperability and exchange of VRA Core records. The unrestricted schema imposes no requirements on the values entered into any of the elements, sub-elements, or attributes, and may be useful for those who want to exchange legacy data.</td>
<td></td>
</tr>
<tr>
<td>KOS.HSS.42</td>
<td>Evidence Thesaurus</td>
<td></td>
<td>Nomenclature Task Force</td>
<td></td>
<td>Version 1.0</td>
<td><a href="http://old.cidoc-crm.org/docs/guide/guide.htm">http://old.cidoc-crm.org/docs/guide/guide.htm</a></td>
<td></td>
</tr>
<tr>
<td>KOS.HSS.43</td>
<td>Farmstead Thesaurus</td>
<td></td>
<td>Getty Art History Information</td>
<td></td>
<td>Version 6.2.1</td>
<td>Editorial Status: Open-In progress</td>
<td></td>
</tr>
<tr>
<td>KOS.HSS.44</td>
<td>First World War Thesaurus</td>
<td></td>
<td>Getty Art History Information</td>
<td></td>
<td>Version 6.2.2</td>
<td>A TDWG task group was created to revise the Darwin Core, and a ratified metadata standard (version 1.0) was adopted. The Darwin Core was originally developed by the Natural History Metadata Group in 2000. The Darwin Core is a metadata standard that can be used to describe biological research data.</td>
<td></td>
</tr>
<tr>
<td>KOS.HSS.45</td>
<td>Heritage Crime Thesaurus</td>
<td></td>
<td>Getty Art History Information</td>
<td></td>
<td>Version 1.0</td>
<td>WP3 working group &quot;Identifying standards and developing recommendations&quot;.</td>
<td></td>
</tr>
<tr>
<td>KOS.HSS.46</td>
<td>Historic Aircraft Thesaurus</td>
<td></td>
<td>Getty Art History Information</td>
<td></td>
<td>Version 1.0</td>
<td>WP3 working group &quot;Identifying standards and developing recommendations&quot;.</td>
<td></td>
</tr>
<tr>
<td>KOS.HSS.47</td>
<td>Historic Characterisation Thesaurus</td>
<td></td>
<td>Getty Art History Information</td>
<td></td>
<td>Version 1.0</td>
<td>WP3 working group &quot;Identifying standards and developing recommendations&quot;.</td>
<td></td>
</tr>
<tr>
<td>KOS.HSS.48</td>
<td>Manner of Loss List</td>
<td></td>
<td>Getty Art History Information</td>
<td></td>
<td>Version 1.0</td>
<td>WP3 working group &quot;Identifying standards and developing recommendations&quot;.</td>
<td></td>
</tr>
<tr>
<td>KOS.HSS.49</td>
<td>Maritime Craft Thesaurus</td>
<td></td>
<td>Getty Art History Information</td>
<td></td>
<td>Version 1.0</td>
<td>WP3 working group &quot;Identifying standards and developing recommendations&quot;.</td>
<td></td>
</tr>
<tr>
<td>KOS.HSS.49.1</td>
<td>Maritime Fixtures and Fittings</td>
<td></td>
<td>Getty Art History Information</td>
<td></td>
<td>Version 1.0</td>
<td>WP3 working group &quot;Identifying standards and developing recommendations&quot;.</td>
<td></td>
</tr>
</tbody>
</table>

**International Coding**

**Classification (ICC)**

<table>
<thead>
<tr>
<th>Relevant ICC code for the KOS.</th>
</tr>
</thead>
</table>

**NARCIS Classification**

<table>
<thead>
<tr>
<th>Relevant NARCIS class for the KOS.</th>
</tr>
</thead>
</table>

**KOS Types Vocabulary**

| The appropriate Dublin Core NKOS group ‘KOS Types Vocabulary’ label for the KOS. |
Some communities have many stairs to climb!

A rough illustration of the semantic gradient

Weaker semantics

Glossaries

Thesauri

Controlled vocabularies

Data models

Taxonomies

Ontologies

Stronger semantics

Modified from Blumauer and Pellegrini (2006)

Source: https://5stardata.info/en/
Endangered KOSs: 404 Errors & other issues

Reference Rot = (Link Rot + Content Drift)
Semantic Change
Subject Ontogeny

Sustainability

LICENSING
Maintenance
Persistence
Stability

Versioning
FAIRness
Ethics
On "FAIR Vocabularies"

At all levels: The (meta)data should facilitate the growth of the science and knowledge graphs and allow for the possibility of reproducibility/replication of experiments, understanding of data and access to resources over time, for credit to be attributed.

Registry level: A common catalog of KOSs is needed. Or a registry harmonising the existing data (Maybe on EOSC Portal)

KOS level: Ontology Engineers, Researchers, and Data Stewards need descriptive, structural, and administrative metadata to be able to record/store/retrieve KOS in registries and repositories. The metadata should be rich in terms of describing content with a sufficient amount of minimum/core metadata to allow for inclusion in a common catalog.

Data level: There must be KOS capable of describing/annotating a vast variety of data covering all of science. Interoperability should be of a high enough quality to allow researchers from different communities to reuse data.
The future of the KOSo

- How can we continue to build a resource (KOSo) which requires a very high amount of curation/data stewardship?
- How can we encourage existing registries (especially those which are more advanced) to begin to consider versioning and archiving?
- How can we encourage ontology engineers to consider the FAIRness of their resources?
- How can we stimulate the creation of LD infrastructure in less mature communities?
Any questions?

gerard.coen@dans.knaw.nl   @gerardcoen
https://dans.knaw.nl/   @dansknaw

DANS is an institute of KNAW en NWO