Interoperability Approaches in KOS Vocabulary Development

Observable Changes In the LOD-enabled mashup culture

https://rdcu.be/PgZW
https://doi.org/10.1007/s00799-018-0241-2
Interoperability Approaches in KOS Vocabulary Development

1 Derivation:
- Derived vocabularies
- Microthesaurus

2 Expansion:
- Leaf nodes
- Satellite vocabularies
- Open umbrella structure

3 Integration/Combination:
- Metathesaurus
- Heterogeneous meta-vocabulary

4 Interoperation/Shared/Harmonization:
- Shared/bridge scheme
- Reference ontologies
- Virtual harmonization through linking

Are the newly generated vocabularies took similar ways used prior to the 21st century?


Interoperability Approaches in KOS Vocabulary Development

Observable Changes
In the LOD-enabled mashup culture

• The new functions and differences observed in current approaches are the results of applying semantic technologies.
  • Each *thing* is named with an URI + a domain name prefix (maintaining the original semantics and linguistic decisions while being reusable).
  • Machine-understandable, machine-processable data.
  • Benefit from semantic technologies and the available open tools.

• Communities grow quickly, spread widely, involving many contributors;
• The number of projects and vocabs increased dramatically;
• Reuse and connect, not in silos;
• No hero ‘master’;
• Vocabularies and Vocabulary Services are together.

Are they challenge us to think further regarding the KOS development?
Observable Changes
In the LOD-enabled mashup culture

Let’s observe some real KOS products

1. A unified vocab for a domain
2. A shared concept scheme from only the most used entries of multiple vocabs
3. A heterogeneous meta-vocabulary
4. Connecting KOS concepts to real things
5. Shared, unconventional mashups
1. A unified scheme for a domain

TOP. 700+ plant characteristics: plant traits and environmental associations.

- Source: http://www.top-thesaurus.org/annotationInfo?viz=1&trait=Frost%20tolerance

• Source: http://www.top-thesaurus.org/annotationInfo?viz=1&trait=Frost%20tolerance
2. A shared concept scheme

1. Take [only] the 10,000 most-used concepts from each, by 2014/10.

2. Automatically map them to each other, by 2015/03.


4. GACS Beta. 2016/05.

Source: Compiled based on Baker, Thomas et al. 2016.
GACS (Global Agricultural Concept Scheme) for Agrisemantics*

- GACS is seen as the first step towards improving the coherence and interoperability of agricultural data.

- GACS as a hub linking user-oriented thesauri with semantically more precise domain ontologies.

- Domain ontologies link to datasets about food and agriculture, in order to make that data more interoperable and reusable.

*Agrisemantics is an emerging community network of semantic assets relevant to agriculture and food security.

GACS (Global Agricultural Concept Scheme)

GACS concept

Broader, narrower, and related concepts

Alternative labels

Translated in up to 25 languages

Mapped to source thesauri

3. A *heterogeneous* meta-vocabulary

Encompasses:
(1) the different conceptions of a taxon,
(2) the temporal order of the changes, and
(3) the references to scientific publications whose results justify these changes.

Allows multilingual, multi-opinions … in a unified view.
4. Enriching the KOS-at-hand and connecting to real things

**skos:Concept ➔ Real-World Things**

- **People are People and Places are Places**
  - In order to describe something accurately they need to be labeled as those specific types of Things
- **foaf:focus** allows FAST Controlled Vocabulary terms (skos:Concept) to be connected to URIs that identify real-world entities

Source: extracted screenshots (2018-08-21)
From http://fast.oclc.org/searchfast/

http://experimental.worldcat.org/fast/35588/
John F. Kennedy’s entry in FAST is enriched with other sources.

- The **DBpedia** identifiers allow FAST terms to include detailed information that is usually excluded in authority records.

- The enrichment allows FAST terms to take advantage of all of the various string values included in **VIAF** without having to manually include the values in the RDF triples for the specific term.
• The GeoNames data is used to power MapFAST, which is a Google Maps mash-up.

(cont.) 4. Enriching the KOS-at-hand and connecting to real things

Swissbib, a provider for bibliographic data in Switzerland
http://linked.swissbib.ch

- 30,773 links to DBpedia
- 20,714 links to VIAF
- high precision values
- generated in reasonable expenditures of time

Source: Figure 8. Data flow diagram of the interlinking procedure in the Weissbib project
Source: Bensmann et al. (2017), p.8
5. Shared, unconventional mashups

Wikipedia

Authority

Authority

Authority

Image captured 2018-08-21, five more name authority than one year ago. 

https://en.wikipedia.org/wiki/Frank_Lloyd_Wright
Wikidata – as an Authority base/ Knowledge Base

- Defined and controlled
- Mapped to other Vocab IDs
- Linked to examples
b. (cont.) Wikidata – as an Authority base/ Knowledge Base

“cotton”
(cont.) 5. Shared, unconventional mashups

The portion in the Wikimedia (linked with Wikidata) entry for category “Dresses” with subcategories organized by facets and hierarchies.

https://commons.wikimedia.org/wiki/Category:Dresses
2017-07-12
No hero ‘master’?

https://commons.wikimedia.org/wiki/Category:Dresses
2018-08-21 Note category notation changes
**Tool: Mix’n’Match**

This tool lists entries of some external databases (over 1000 catalogs), and allows users to match them against Wikidata items.

[https://tools.wmflabs.org/mix-n-match/#/](https://tools.wmflabs.org/mix-n-match/#/)

---

<table>
<thead>
<tr>
<th>Catalog groups</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
</tr>
<tr>
<td>Wolfram Language entity</td>
</tr>
<tr>
<td>Archives</td>
</tr>
<tr>
<td>Art</td>
</tr>
<tr>
<td>Biography</td>
</tr>
<tr>
<td>Biology</td>
</tr>
<tr>
<td>Books</td>
</tr>
<tr>
<td>Encyclopedia</td>
</tr>
<tr>
<td>Entertainment</td>
</tr>
<tr>
<td>Food</td>
</tr>
<tr>
<td>General</td>
</tr>
<tr>
<td>Geography</td>
</tr>
</tbody>
</table>

**Search catalogs**

*Start typing here*

**Latest catalogs**

- Württembergische Kirchengeschichta person identifier for a person in the Evangelical church in Württemberg
- INE municipality code identifier for Spanish municipalities, by the Spanish Statistical Institute
- BNP works Works at the BNP
- BNP authors Authors at BNP
- Library of Congress Genre/Form Terms ID for the Library of Congress genre and forms
- PRELIB person identifier for a person on the Projet de recherche en littérature website

**Catalogs by property class**

<table>
<thead>
<tr>
<th>Group</th>
<th>Catalogs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authority control for people</td>
<td>426</td>
</tr>
<tr>
<td>Catalogs without Wikidata property</td>
<td>255</td>
</tr>
<tr>
<td>Identifier</td>
<td>153</td>
</tr>
<tr>
<td>Authority control</td>
<td>102</td>
</tr>
<tr>
<td>Software</td>
<td>93</td>
</tr>
<tr>
<td>Organisations</td>
<td>65</td>
</tr>
<tr>
<td>Films</td>
<td>54</td>
</tr>
<tr>
<td>Authority control for artists</td>
<td>53</td>
</tr>
<tr>
<td>Film industry</td>
<td>53</td>
</tr>
<tr>
<td>Authority control for works</td>
<td>49</td>
</tr>
<tr>
<td>Encyclopedias</td>
<td>48</td>
</tr>
<tr>
<td>Authority control for places</td>
<td>45</td>
</tr>
<tr>
<td>Sports hall of fame</td>
<td>44</td>
</tr>
<tr>
<td>Authority control for writers</td>
<td>42</td>
</tr>
<tr>
<td>Politics</td>
<td>35</td>
</tr>
<tr>
<td>Medicine</td>
<td>33</td>
</tr>
<tr>
<td>Video games</td>
<td>30</td>
</tr>
<tr>
<td>Taxa</td>
<td>29</td>
</tr>
<tr>
<td>Artworks</td>
<td>27</td>
</tr>
<tr>
<td>Music</td>
<td>25</td>
</tr>
<tr>
<td>Biology</td>
<td>24</td>
</tr>
</tbody>
</table>
# Authority control

<table>
<thead>
<tr>
<th>Group</th>
<th>Catalogs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authority control for people</td>
<td>126</td>
</tr>
<tr>
<td>Catalogs without Wikidata property</td>
<td>255</td>
</tr>
<tr>
<td>Identifier</td>
<td>153</td>
</tr>
<tr>
<td>Authority control</td>
<td>102</td>
</tr>
<tr>
<td>Software</td>
<td>93</td>
</tr>
<tr>
<td>Organisations</td>
<td>65</td>
</tr>
<tr>
<td>Films</td>
<td>54</td>
</tr>
<tr>
<td>Authority control for artists</td>
<td>53</td>
</tr>
<tr>
<td>Film industry</td>
<td>53</td>
</tr>
<tr>
<td>Authority control for works</td>
<td>49</td>
</tr>
<tr>
<td>Encyclopedias</td>
<td>48</td>
</tr>
<tr>
<td>Authority control for places</td>
<td>45</td>
</tr>
<tr>
<td>Sports hall of fame</td>
<td>44</td>
</tr>
<tr>
<td>Authority control for writers</td>
<td>42</td>
</tr>
<tr>
<td>Politics</td>
<td>35</td>
</tr>
<tr>
<td>Medicine</td>
<td>33</td>
</tr>
<tr>
<td>Video games</td>
<td>30</td>
</tr>
<tr>
<td>Taxa</td>
<td>29</td>
</tr>
<tr>
<td>Artworks</td>
<td>27</td>
</tr>
<tr>
<td>Music</td>
<td>25</td>
</tr>
<tr>
<td>Biology</td>
<td>24</td>
</tr>
</tbody>
</table>

- **HMBID**: identifier of a person or topic in the History of Modern Biomedicine database.
- **ISO 4217**: code de monnaitre ISO 4217 (inclut la bitcoin).
- **GND Fictive Places**: GND entries about fictive places.
- **INAO Label pages**: Food quality standards for French AOC.
- **EuFashion Thesaurus**: Vocabulary of European Fashion.
- **FOIH styles and cultures**: styles and cultures in Flemish organization for Immovable Heritage this.
- **PM20 special subjects**: Sondermappe Sacharchiv Pressemappe 20. Jahrhundert.
- **ISO 15624 numeric code**: numeric code for a writing system in ISO 15624.
- **Caijifora**: item number for taxa in Califora.
- **MeSH Humanities**: Medical Subject Headings for Humanities.
- **FOIH periods**: periods in Flemish organization for Immovable Heritage thesauri.
- **MeSH Publication Characteristics**: Medical Subject Headings for Publication Characteristics.
- **DoS Events**: Dictionary of Sydney / events.
- **FFW**: Contains articles about software, file formats, file types, file extension.
- **MeSH Named Groups**: Medical Subject Headings for Named Groups.
- **MeSH Psychiatry and Psychology**: Medical Subject Headings for Psychiatry and Psychology.
- **DoS Natural Features**: Dictionary of Sydney / natural features.
- **GeoNames**: Identifier for feature classes in GeoNames.
- **NSW Flora ID**: Identifier for a plant taxon, in the NSW Flora Online.
- **MeSH Disciplines and Occupations**: Medical Subject Headings for Disciplines and Occupations.
- **FOIH taxa**: species in Flemish organization for Immovable Heritage thesauri.

[https://tools.wmflabs.org/mix-n-match/#/group/ig_authority_control](https://tools.wmflabs.org/mix-n-match/#/group/ig_authority_control) 2018-08-22
“rhyta”

Match results on Mix’n’match, 2018-06-16

“Dunhuang cave”
Interoperability Approaches in KOS Vocabulary Development

Observable Changes In the LOD-enabled mashup culture

<table>
<thead>
<tr>
<th>No significant change:</th>
</tr>
</thead>
<tbody>
<tr>
<td>From conceptual and structural points of view, the newly generated vocabularies,</td>
</tr>
<tr>
<td>derived from the existing ones, took similar approaches used prior to the 21st century.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observable changes:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The new functions and differences observed in current approaches are the results of applying semantic technologies.</td>
</tr>
<tr>
<td>• Each thing is named with an URI + a domain name prefix (maintaining the original semantics and linguistic decisions while being reusable).</td>
</tr>
<tr>
<td>• Machine-understandable, machine-processable data.</td>
</tr>
<tr>
<td>• Benefit from semantic technologies and the available open tools.</td>
</tr>
<tr>
<td>• Communities grow quickly, spread widely, involving many contributors;</td>
</tr>
<tr>
<td>• The number of projects and vocabs increased dramatically;</td>
</tr>
<tr>
<td>• Reuse and connect, not in silos;</td>
</tr>
<tr>
<td>• No hero ‘master’;</td>
</tr>
<tr>
<td>• Vocabularies and Vocabulary Services are together.</td>
</tr>
</tbody>
</table>

Are they challenge us to think further regarding the KOS development?

How & Who to measure?

• Quality
• Sustainability
• Applicability
• Usability/Re-use-ability
• Being “authority”?
• Functional “requirements”?

As a community, we need to think about these.
References


O’Neill, Ed, and Jeff Mixter 2013. (1) The case for faceting (2) FAST Linked Data mechanics. In 76th Annual Meeting of the American Society for Information Science and Technology (ASIS&T), Montreal, Canada, Nov. 2-6, 2013.

