Knowledge Organization Systems (KOS) in the Semantic Web: Is the Whole Greater Than the Sum of Its Parts?
-- A Multi-Dimensional Review

Marcia Lei Zeng
School of Information, Kent State University, Ohio, USA
&
Philipp Mayr
GESIS - Leibniz Institute for the Social Sciences, Cologne, Germany

1. Background

1.1 About this study

Main sources of the study:
• Sessions of KOS at international conferences
• Research-based journal publications
• Theses and dissertations
• Professional conferences and summits
• NKOS workshops [http://nkos.slis.kent.edu](http://nkos.slis.kent.edu)
• NKOS Bibliography

Other sources where cases were exposed to the authors:
• LOV on Google+
• Getty Vocab Google Group
• Getty Share
• Social media sources: tweets, blogs, Facebook groups
• Ontolog-Forum
• LODLAM challenges and sessions
• GitHub entries such as OpenSKOS, NatLibFi/Skosmos, JSKOS, etc.

Zeng, ISKO-UK 17/09/12
1.2 KOS trending (1/3)

In the BARTOC registry
(thesaurus, ontology, classification)
KOS registered:
-(2016-05): 1836
-(2017-08-25): 2,753
RDF: 297

In the Datahub

LOD KOS registered
-(2016-03): 1251 found
-(2017-08): 1662 found

with tags “thesaurus”, “classification”, “taxonomy”, “ontology”, “terminology”
(Note: some are tagged with multiple categories. Some have multiple editions.)

http://bartoc.org/

https://datahub.io/
1.2 KOS trending (2/3)

**Growth of the KOS for the information and knowledge organization tasks**

**Classification, Subject headings**
- LIS professionals
- I&A database producers

**+ Thesaurus**
- LIS professionals
- I&A database producers

**++ Ontology, knowledge base**
- LIS professionals
- I&A database producers
- Anyone
Conventional KOS have always been quick adapters of new technologies in their publishing venues and applications.

The release of a LOD KOS product represents a turning point for the producer or provider of a vocabulary; but what are the results?
1.3. Explanation of the term “LOD KOS”

Covering:
- “value vocabularies”
- “light-weight ontologies”, “application ontologies”
Pride and Prejudice

Jane Austen (1775-1817)

-resource

"value vocabularies"

Zeng, ISKO-UK 17/09/12
1.4. Features of a LOD KOS (1/5)

• A LOD KOS must follow the principles of Linked Data (Berners-Lee 2006) and must be openly available.

1. Use **URIs** as names for things
2. Use **HTTP** URIs so that people can look up those names
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs so that they can discover more things

[http://www.w3.org/DesignIssues/LinkedData.html](http://www.w3.org/DesignIssues/LinkedData.html)
1.4. Features of a LOD KOS (2/5)

a) A KOS = a **concept scheme**, comprising a set of **concepts** where each *concept* must be named by a **URI** or **IRI**.

- Using a unique identifier to represent an entity or resource is one of the basic solutions to providing machine-processable, disambiguated data.

URI (Uniform Resource Identifier)

IRI (Internationalized Resource Identifier)

http://dbpedia.org/page/Himalayas

[Image of a mountain]

http://zh.wikipedia.org/wiki/喜马拉雅山脉

http://my.wikipedia.org/wiki/ſімғу_ſімғуңғылар

http://ja.dbpedia.org/page/ヒマラヤ山脈
1.4. Features of a LOD KOS (3/5)

a) A KOS = a concept scheme, comprising a set of concepts
   -- where each concept must be named by a URI or IRI;
   -- and HTTP URIs should be used when released as LOD.

Example of a thesaurus entry in resource-property-value triples, viewing from a Web interface.

Note that each concept is represented by an http URI.

Source: AGROVOC concept http://aims.fao.org/aos/agrovoc/c_6599.html
1.4. Features of a LOD KOS (4/5)

b) Data of a LOD KOS are expressed as RDF triples and may be encoded using any concrete RDF syntax.

Example of a thesaurus entry encoded using RDF/XML.
Source: LCSH concept http://id.loc.gov/authorities/subjects/sh2007006251
c). The end-product of a LOD KOS may be available as a RDF data-dump or accessed through a SPARQL endpoint.
1.5 LOD KOS Vocabulary Services (1/5)
-- Repositories and portals hosting KOS full products

1) Individual vocabulary provider

2) Individual institute as the provider of all vocabularies produced in the institution
   • Library of Congress Linked Data Services – Authorities and Vocabularies [http://id.loc.gov/](http://id.loc.gov/)

3) Country-based provider

4) Domain-oriented portal
   • E.g., BioPortal [www.bioportal.bioontology.org](http://www.bioportal.bioontology.org) , and many others
BioPortal Statistics

- Ontologies: 585
- Classes: 8,130,555
- Resources Indexed: 48
- Indexed Records: 39,537,360
- Direct Annotations: 95,468,433,792
- Direct Plus Expanded Annotations: 144,789,582,932

2017-07-25 data
Usage data about each vocabulary

http://bioportal.bioontology.org/ontologies/MESH
(cont.) LOD KOS Vocabulary Services (4/5)

5) Middleware that provides tools for end-users to use/reuse published vocabularies
   Skosprovider http://skosprovider.readthedocs.io

6) Upper ontology facilitated multiple vocabularies’ concept- and entity-mapping and reuse
   E.g., Linked Open Ontology cloud KOKO https://finto.fi/koko/en/
   E.g., Upper Mapping and Binding Exchange Layer (UMBEL) http://umbel.org/
(cont.) LOD KOS Vocabulary Services (5/5)

-- **Registries** offer information *about* vocabularies

1) Registry of KOS
   - E.g., BARTOC (Basel Register of Thesauri, Ontologies & Classifications) [https://bartoc.org/](https://bartoc.org/)

2) Registry of vocabularies (“property vocabularies” and “value vocabularies”) that are published with Semantic Web languages
   - E.g., LOV (Linked Open Vocabularies) [http://lov.okfn.org/dataset/lov](http://lov.okfn.org/dataset/lov) : 600+ registered, some are value vocabularies.

3) Registry of LOD products, including KOS
   - Searching for various KOS types resulted with over 1000+, after verification.
2. Preliminary Findings

Examined functional changes of KOS after they were released as LOD.

Findings are organized around personas that representing typical users of LOD KOS.

- Some specific cases are used as examples.
- Attention is on summarizing the **general issues** and **milestones** identified by this study.
- Best practices and experimental approaches are presented together with the possible challenges and hurtles.
2.1. For LOD Dataset Producers, LOD KOS vocabularies enable their data to become 4-star and 5-star Linked Open Data

Tim Berners-Lee [https://www.w3.org/DesignIssues/LinkedData.html](https://www.w3.org/DesignIssues/LinkedData.html)

Zeng, ISKO-UK 17/09/12
The options and actions related with KOS in the LOD dataset production

- **No controlled values**
  - Need to populate controlled vocabs in a dataset

- **Controlled, but local**
  - Need to map to standard or popular KOS

- **Controlled, but not on LOD**
  - Need to use LOD vocabs (with URIs)
DP-I.

**Situation:** Dealing with semi-structured and unstructured data that have no controlled values for the named entities and topics.

**Purpose:** create LOD datasets from scratch.

### Milestones:
1. Identify the entities.
2. Put the entities into structured data.
3. Clean up the newly structured data, with local control.
4. **Encode the entities with standardized KOS vocabularies (as strings).**
5. Obtain URIs for names of entities provided by the LOD KOS datasets.
6. Use http URIs for names of any entities.

### Why try to control?

**Possible issues:**
- Place names change through time (e.g., “Saint Petersburg,” “Leningrad”);
- Alternate names exist (e.g., “New York City,” “NYC,” and multilingual labels);
- Same name is associated with multiple locations (“St. Petersburg” in the world);
- Unidentifiable places;
- Unnamed places;
- Cartographic versus geographical placement;
- Feature typing /categorizing results are incorrect or inconsistent.

### Examples:
- Digitized materials, textual or non-textual, in silos
- Archival finding aids
- Oral history transcripts
- Merged local files
- ... ...
**Case: Linked Jazz**  [http://linkedjazz.org/](http://linkedjazz.org/)

Focuses on digital archives of jazz history to expose relationships between musicians and reveal their community’s network.

**Milestones:**
1. Identify the entities.
2. Put the entities into structured data.
3. Clean up the newly structured data, with local control.
4. **Encode the entities with standardized KOS vocabularies (as strings).**
5. **Obtain URIs for names of entities provided by the LOD KOS datasets.**
6. Use http URIs for names of any entities.

Step 1, get the names from the transcripts; map to name authorities (VIAF, LC name authority, and Dbpedia) or establish new authority entries.
Step 2. Find the names in the documents based on the established name authority file.

Step 3. Describe the relationships using a relationship ontology the team developed.

http://linkedjazz.org/
**Steps:**
- Represent the thesaurus in a **machine-understandable way** for automating machine-assisted indexing processes;
- Facilitate its integration into retrieval tools;
- Ensure the consistency of indexing even though the thesaurus evolves;
- Facilitate the process of updating and maintaining the thesaurus;
- Express all the concepts already represented in the thesaurus; and
- Use standards and models related to thesauri and controlled vocabularies for interoperability purposes (Sibille-de Grimoüard 2014).

**Case: Archives of France (Sibille-de Grimoüard 2014)**

One of the key tasks involves the conversion of existing KOS into LOD before applying them as standard value vocabs in all datasets to be integrated.

https://francearchives.fr/article/37828
**DP- 3.**

Having datasets that have been using value vocabularies in structured data, turning them into 4 star and 5-star LOD

**Milestones:**
1. Identify the entities.
2. Put the entities into structured data.
3. Clean up the newly structured data, with local control.
4. **Encode the entities with standardized KOS vocabularies (as strings).**
5. Obtain URIs for names of entities provided by the LOD KOS datasets.
6. Use http URIs for names of any entities.

Image source: Zeng, 2012
Possible issues for each of the datasets currently at the 3-star level and on the way to the 4-star and 5-star

• If it has used local controlled vocabularies
  ➢ the terms used or the form representing the concepts and named entities may be different from standardized controlled vocabularies.

• If it has used pre-LOD vocabulary
  ➢ there might be no URIs/IRIs yet. How to obtain the URIs/IRIs to replace the strings of a named entity or concept?

• If a decision of mapping is made
  ➢ which vocabulary and how many vocabularies will be involved, since in a subject domain and a community there could be more than one standard vocabulary?

• If it needs to map the local controlled lists to a standardized LOD KOS,
  ➢ human resources and quality control are most critical and could be challenging.

• For a dataset formed through aggregation
  ➢ in addition to the above issues, synonyms and acronyms occur in the data provided by different sources. Heavy disambiguation and semantic conflict controls are needed.
The 5-star datasets in the LOD Cloud indicated the essential role of LOD KOS Vocabularies

Source: Annotated by the author on the LOD CLOUD 2014-08-30 image
Zeng, ISKO-UK 17/09/12
2. Preliminary Findings

**DP Summary**

**For LOD Dataset Producers**

LOD KOS vocabularies
- are the source of http URIs/IRIs for named entities and concepts used in data-transformation;
- empower the owners of data to
  - convert and publish their data under the LOD principles, with high quality and trustworthy linkages in RDF triples;
  - transforming anyone’s database into LOD Datasets, reaching 4- and 5- stars;
  - create machine-understandable and machine-processable data for any users, machine or human.

The creation of a KOS vocabulary involves tremendous intellectual efforts and human resources, thus the openly available, well-established, and constantly-maintained vocabularies are invaluable engines for the LOD datasets.
2. Preliminary Findings

2.2 For vocabulary producers (VPs) who are involved in the development and enrichment of KOS, LOD approaches lead to unconventional processes and results.
VP-1. Creating new value vocabularies for a particular project’s products by extracting the components from a comprehensive KOS vocabulary

• Example: creating a microthesaurus
The units (facets, hierarchies) were recommended to be used.
Steps:
1. Go to Getty Vocab LOD SPARQL Endpoint: http://vocab.getty.edu/sparql
2. Choose ‘Queries’.
3. Choose “Descendants of a Given Parent” from the template, click. → Now, the template's text will show on the right.
4. Click ‘SPARQL” to get the query text up.

VP-1. Creating new value vocabularies for a particular project's products by extracting the components from a comprehensive KOS vocabulary.
5. Use this ID in the query, send the query.
6. Get the dataset, download.
VP-2. Creating a unified scheme for a domain based on multiple existing KOS vocabularies

Case: TOP

Case: *Global Agricultural Concept Scheme (GACS)*
The definition uses multiple sources of the concepts, coded with the URI from the original namespaces.

**Source:** [http://www.top-thesaurus.org/annotationInfo?viz=1&trait=Frost%20tolerance](http://www.top-thesaurus.org/annotationInfo?viz=1&trait=Frost%20tolerance)
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.
Case: **Global Agricultural Concept Scheme (GACS)**

- **GACS concept**
- Broader, narrower, and related concepts
- Alternative labels
- Translated in up to 25 languages
- Mapped to source thesauri

VP-3. Creating a heterogeneous meta-vocabulary

**Case: TaxMeOn**

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.

Source: http://onki.fi/onkiskos/kerambycids/
VP-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.

Case: FAST

Case: Swissbib

Case: FAST

Source: extracted screenshots (2017-07-12)
From http://fast.oclc.org/searchfast/
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.
Case: **FAST**

- The GeoNames data is used to power MapFAST, which is a Google Maps mash-up.
(cont.) VP-4. Enriching the KOS-at-hand and connecting to real things

Case: **Swissbib**

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
VP-5. Enhancing semantic consistency of data through shared, unconventional mashup KO activities

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
Another example:
Europeana Fashion Vocabulary

Vocabulary Development End-products

No significant change:

• Conceptual and structural: similar approaches that was developed prior to the 21st century.
  • making microthesauri and satellite vocabularies,
  • creating a super structure,
  • direct mapping or employing a switching system,
  • crowd-sourcing, post-control, etc.

Observable changes:

• 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.

Zeng, ISKO-UK 17/09/12
2. Preliminary Findings

2.3 For Researcher (RS) end-users, LOD KOS products can become knowledge bases and provide semantic-rich discoveries.

Examples:
- Universal Protein Resource (UniProt)
- Getty Vocabularies LOD
- Cadastre and Land Administration Thesaurus (CaLAThe)
RS-1. Using well-developed KOS products, high quality and relevant knowledge bases are now available for researchers to use easily.

- LOD KOS can be used for
  - obtaining special graphs or datasets for very complicated questions, and
  - revealing unknown relationships.

Could a LOD KOS dataset be considered
- as a knowledge base?
- as the foundation of a network analysis?
- as the building blocks of a framework for research in humanities and science?
Examples

1. Select all taxa from the UniProt taxonomy: (show)
2. Select all bacterial taxa, and their scientific name, from the UniProt taxonomy: (show)
3. Select all E-Coli K12 (including strains) UniProt entries and their amino acid sequence: (show)
4. Select the UniProt entry with the mnemonic 'A4_HUMAN': (show)
5. Select a mapping of UniProt to PDB entries using the UniProt cross-references to the PDB database: (show)
6. Select all cross-references to external databases of the category '3D structure databases' of UniProt entries that are classified with the keyword '3Fe-4S': (show)
7. Select all UniProt entries, and their recommended protein name, that have a preferred gene name that contains the text 'DNA': (show)
8. Select the preferred gene name and disease annotation of all human UniProt entries that are known to be involved in a disease: (show)
9. Select all human UniProt entries with a sequence variant that leads to a 'loss of function': (show)
10. Select all human UniProt entries with a sequence variant that leads to a tyrosine to phenylalanine substitution: (show)
11. Select all UniProt entries with annotated transmembrane regions and the regions' begin and end coordinates on the canonical sequence: (show)
12. Select all UniProt entries that were integrated on the 30th of November 2010: (show)
13. Was any UniProt entry integrated on the 9th of January 2013? (show)
14. Construct new triples of the type 'HumanProtein' from all human UniProt entries: (show)
15. Select all triples that relate to the EMBL CDS entry AA089367.1: (show)
16. Select all triples that relate to the taxon that describes Homo sapiens: (show)
17. Select the average number of cross-references to the PDB database of UniProt entries that have at least one cross-reference to the PDB database: (show)
18. Select the number of UniProt entries for each of the EC (Enzyme Commission) second level categories: (show)

http://sparql.uniprot.org/
One can obtain special RDF graphs or datasets for very complicated questions, and revealing unknown relationships.

### Getty Vocabularies: LOD

#### 4.1 TGN-Specific Queries

- **4.1** Places by Type
- **4.2** Places with English or GVP Label
- **4.3** Places by Direct and Hierarchical Type
- **4.4** Breakdown of Sovereign States by Type
- **4.5** Inhabited Places That Were Sovereign States
- **4.6** Places by Type and Parent Place
- **4.7** Places by Type, with placeTypePreferred
- **4.8** Places by Triple FTS
- **4.9** Places by FTS Parents
- **4.10** Capitals by Association
- **4.11** Members of the European Union
- **4.12** Members of the United Nations
- **4.13** Geo Chart with sgvizler
- **4.14** Column Chart with sgvizler
- **4.15** Countries and Capitals By Type and Containment
- **4.16** Places by Coordinate Bounding Box
- **4.17** Places Within Bounding Box
- **4.18** Places by Type Within Bounding Box
- **4.19** Places Outside Bounding Box (Overseas Possessions)
- **4.20** Places Nearby Each Other

[http://vocab.getty.edu/queries#Top-level_Subjects](http://vocab.getty.edu/queries#Top-level_Subjects)
Steps: [http://vocab.getty.edu/](http://vocab.getty.edu/) => Queries

1. go to 4.18,
2. click on that SPARQL sign for 4.18,
3. submit.

Demo: Look for **castles around The Netherlands**
(within the boundary of 50.787185 3.389722 53.542265 7.169019)
Demo: Look for castles around The Netherlands

(4) Download the datasets in a selected format.

Additional:
(5) Click on any castle’s ID, (6) open the single data record for this concept. Download the dataset as you wish.
(7) You may click on the Website to see its normal html view.
Example 2.
Following these geographic places located on the Silk Road, using the **geo-coordinates** TGN provided, get all of them in one dataset.
Demo: Look for caves on or around the ancient Silk Road

“caves” within bounding box (24.75083 28.95778 43.80722 108.92861)
“caves” within bounding box (24.75083 28.95778 43.80722 108.92861)

- Got a dataset of over 200 caves spread in various countries, all done within a few minutes.
- Each URI also brings the full data for each cave and other related information.
- The dataset is available for downloading with various formats.
RS-2. Name authorities offer foundational structured data for network analyses.
**Steps:**
1. go to 5.2;
2. click on that SPARQL sign for 5.2;
3. replace the ID of the person you want to find. **Note there are two IDs to replace.**
4. Submit.

**Query:** Find associative relationships of **ulan:500020307 Wright, Frank Lloyd** (American architect, 1867-1959); showing **relationship type, associated persons, each person’s preferred name, preferred display biography**, and other notes.
<table>
<thead>
<tr>
<th>rel</th>
<th>x</th>
<th>name</th>
<th>bio</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>gvp:ulan1000_related_to</td>
<td>ulan:5000077136</td>
<td>Sullivan, Francis</td>
<td>Canadian architect and draftsman, 1882-1929</td>
<td></td>
</tr>
<tr>
<td>gvp:ulan1101_teacher_of</td>
<td>ulan:500125693</td>
<td>Lustig, Alvin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gvp:ulan1106_apprentice_of</td>
<td>ulan:500035235</td>
<td>Ayala Valva, Francisco d’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gvp:ulan1106_apprentice_of</td>
<td>ulan:500255770</td>
<td>Beharka, Roi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gvp:ulan1106_apprentice_of</td>
<td>ulan:500249945</td>
<td>Besinger, Cyril</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gvp:ulan1106_apprentice_of</td>
<td>ulan:500236661</td>
<td>Drake, Blaine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gvp:ulan1106_apprentice_of</td>
<td>ulan:500236682</td>
<td>Drake, Hulda</td>
<td>Biefry</td>
<td></td>
</tr>
<tr>
<td>gvp:ulan1106_apprentice_of</td>
<td>ulan:500585665</td>
<td>Karflik, Vladimir</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gvp:ulan1106_apprentice_of</td>
<td>ulan:500001446</td>
<td>Tafel, Edgar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gvp:ulan1202_patron_of</td>
<td>ulan:500071769</td>
<td>Hanna, Jean Shuman</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gvp:ulan1217_employee_of</td>
<td>ulan:500001455</td>
<td>Sullivan, Louis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gvp:ulan1218_employee_of</td>
<td>ulan:500003109</td>
<td>Griffin, Marie McNairy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gvp:ulan1218_employee_of</td>
<td>ulan:500001158</td>
<td>Griffin, Walter Burley</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gvp:ulan1218_employee_of</td>
<td>ulan:500202006</td>
<td>Guerrero, Patricia</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
More findings: Searching on the Web using such an URI, e.g., “ulan:500020307”, the results will bring:

- this artist’s Wikipedia pages in all languages,
- the DBpedia (rich structured data, modeled by DBpedia Ontology),
- the museums hosting the artists’ works (such as MoMA),
- the libraries that have books about this artist,
- and more.

Results: Websites for this ‘thing’ **ulan: 500020307**

Frank Lloyd Wright | MoMA
https://www.moma.org/artists/6459
American, 1867–1959. A pragmatist, technical innovator, and independent thinker, Frank Lloyd Wright designed cities and buildings and their interior furnishings ...

Frank Lloyd Wright: prairie houses - Catalog - UW-Madison Libraries
https://search-ld.library.wisc.edu/catalog/9910026403802121
Publication Details Cite/Export. Creator photographs by Alan Weintraub ; text by Alan Hess ; with contributions by Kathryn Smith ; Format Books ; Contributors.

Frank Lloyd Wright - DBpedia Wikidata
wikidata.dbpedia.org/page/Q5604

Frank Lloyd Wright - TheInfoList.com
www.theinfolist.com/pg/SummaryGetPub?findGo=Frank%20Lloyd%20Wright ...

Frank Lloyd Wright - Wikivisually
wikivisually.com/ang-cs/wiki/Frank_Lloyd_Wright ...

Frank Lloyd Wright - Project Gutenberg Self-Publishing - eBooks ...
www.gutenberg.org/articles/frank_lloyd_wright ...

Frank Lloyd Wright - LiveBinder
www.livebinders.com/playby/?id=747 ...

My-Proxy - index.php
http://vocabulary.eduta.gov/ulan/500020307 ...
http://www.munzing.de/search/go/document.jsp?id= ...

Zeng, ISKO-UK 17/09/12
RS-3. User-friendly displays of KOS provide visually enriched understanding.

The graphic overview of the group “Activity” of the Cadastre and Land Administration Thesaurus (CaLAThe).

Source: captured from http://cadastralvocabulary.org/
RS Summary --
For Researcher (RS) end-users, LOD KOS products can become knowledge bases and provide semantic-rich discoveries.

What are found:
- There are great and endless potential of LOD KOS.
- The semantic rich structure and high quality controlled vocabulary now can be used integratedly and innovatively, beyond simply existing as controlled vocabularies or standardized name authorities.
- LOD KOS datasets should be considered as:
  - knowledge bases,
  - the foundation of network analyses,
  - the building blocks of a framework for research in humanities and science.

What are needed in action:
- extended functionality of the KOS beyond being the value vocabularies:
- The final LOD KOS deliverables need appropriate practices for the
  - implementation,
  - extension,
  - access, and
  - use of W3C standards
- The barrier resides in the communication about the KOS rather than in their structures, formats, or contents.
- Need to realize the limitations of normal web-based search and browsing.
- Need to take the full advantages of machine-processable data that are much more powerful and useful than the previous machine-readable status.
Conclusions (1/2)

• Although it is possible to use each available component of KOS independently, the real power lies in the skillful coordination of all.
  • The Semantic Web standards such as SKOS, OWL, RDFS, and SPARQL have paved the way for the conventional KOS to become LOD datasets.
  • There have been tremendous and continuous needs for KOS of all kinds, across domains, and worldwide.

• When the two sides embrace and KOS join the mainstream in the 21st century, the opportunities for using the semantic-rich LOD KOS is much greater than ever before due to the fact:
  • LOD KOS data are *machine-understandable*, *-processable*, and *–actionable* (instead of just being *machine-readable*);
  • the Semantic Web connects *things* instead of *strings*.

“The whole is greater than the sum of its parts”
- Aristotle*

Conclusions (2/2)

• Call for more needed collaborations between the knowledge organization communities and the semantic technology communities.

• Encourage working with real end-user researchers who will
  • bring their domain expertise, express their information needs, and help us to understand the information-seeking behaviors;
  • lay out the questions that KOS knowledge bases can aim to answer; and
  • help the growth of the KOS user communities with a variety of new objectives.

“The whole is greater than the sum of its parts”
- Aristotle*

THANK YOU!

Knowledge Organization Systems (KOS) in the Semantic Web: Is the Whole Greater Than the Sum of Its Parts?
-- A Multi-Dimensional Review

Marcia Lei Zeng
School of Information, Kent State University,
Ohio, USA
&
Philipp Mayr
GESIS - Leibniz Institute for the Social Sciences, Cologne, Germany

References (1/3)


References (2/3)


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.


Zeng, ISKO-UK 17/09/12
References (3/3)


