

Panel: Knowledge Organization Systems (KOS) as Linked Data Services

Joseph A Busch

Taxonomy Strategies
Washington, D.C.
jbusch@
taxonomystrategies.com

Vivian Bliss

Taxonomy Strategies
Washington, D.C.

Marjorie Hlava

Access Innovations
Albuquerque, NM
mhlava@accessinn.com

Imma Subirats

Food and Agriculture
Organization of the UN
Imma.Subirats@fao.org

Marcia Lei Zeng*

Kent State University
Kent, OH
mzeng@kent.edu

Maja Žumer

University of Ljubljana
Ljubljana, Slovenia
Maja.Zumer@ff.uni-lj.si

ABSTRACT

Knowledge organization systems (KOS) are the key elements for interlinking the values of Linked Data datasets. Existing KOS products (including well-known thesauri and classification systems) need to be transformed into the constructs of Linked Data services, much more beyond what a traditional thesaurus or classification were/are developed, but how? What can these Linked Data KOS bring to various information services? The panelists will share their experiences and lessons learned. The presentation starts from the processes that have been developed to build, maintain, and publish a newly derived taxonomy using W3C SKOS standard, to make it extensible and scalable to support new information services in scholarly publishing workflow. The next presentation will share experiences and lessons learned from the implementation of a platform built for publishing and distribution of multiple KOS and for automatic indexing using these KOS. The third case is a Linked Data bibliographic service, a RDF-aware system and a mash up application which relies on the use of a multilingual thesaurus that has been published and maintained as a Linked Data dataset, and has been mapped to more than ten KOS. For the datasets that have not used standardized KOS,

the panel will present methods that can be applied when generating or deriving project-focused micro-thesauri from the Linked Data KOS and give an overview of the use cases of Linked Data KOS in information services.

Keywords

Knowledge Organization Systems, KOS, Linked Data, SKOS, Thesaurus publishing and distribution platform Mashup, RDF Triples, information services.

THE PANELISTS AND PRESENTATIONS

All panelists will give presentations based on the real-world and daily-use experiences, from the processes to the tools, and from results to lessons learned. The presenters have been very active in the areas of networked knowledge organization systems for many years and more recently on the frontline of Linked Data services. A majority of the panelists are from industries and services, not academic. The short biographies and presentation descriptions are listed below, arranged in presentation order.

1. Changing from a Hierarchical Classification to a Linked Data Standard: American Physical Society (APS) Taxonomy Case Study

Joseph A Busch, Taxonomy Strategies

Vivian Bliss, Taxonomy Strategies

Since 1975, American Physical Society (APS) *Physical Review* article submission, editorial assignments and journal tables of contents have been based on the *Physics and Astronomy Classification Scheme* (PACS) developed by the American Institute of Physics (AIP), a 5 to 6-level hierarchical classification system. In 2013 APS decided to

develop a new taxonomy. AIP is no longer maintaining the classification scheme, and APS wanted a single scheme for all types of content (journal articles, conference papers and website content), and to develop new capabilities for topic-based online services such as faceted navigation, targeted alerts, personalized subscriptions, etc. The new taxonomy includes several discrete sets of categories (facets) allowing content to be categorized along distinct dimensions rather than trying to use a single concept as with PACS. APS also wanted to base the new taxonomy on Linked Data standards, in particular the SKOS knowledge organization standard, and is using the PoolParty thesaurus management tool to build, maintain and publish the new faceted taxonomy. This presentation will discuss the process that has been developed to build, maintain and ultimately publish the new *APS Taxonomy* so that is, as far as possible, backward compatible with the legacy content categorized using PACS, and is extensible and scalable to support new information services.

Short biography: Mr. Busch is the Founder and Principal Consultant of Taxonomy Strategies. Taxonomy Strategies guides global companies, government agencies, international organizations and not-for-profits such as Nike, Oracle, the U.S. Environmental Protection Agency, the International Monetary Fund and Harvard Business Publishing in developing metadata frameworks and taxonomy strategies. Before founding Taxonomy Strategies, Joseph Busch held management positions at Interwoven, Metacode Technologies, the Getty Information Institute and PriceWaterhouse. He is a Past President of the ASIS&T, and a member of the Executive Committee and Chair-elect of the Advisory Board of the Dublin Core Metadata Initiative.

Vivian Bliss, Taxonomy Strategies

Short biography: Vivian Bliss is a Taxonomy Strategies associate who focuses on creating and applying taxonomies and metadata to improve content management and the user experience on the World Wide Web and corporate intranets. Ms. Bliss has been instrumental in pioneering information architecture and involved in developing several taxonomy management tools and processes. She was a member of the Microsoft Windows Vista User Experience Team and Natural User Interface Platform Team. In addition to earning a Master's in Library and Information Science from the University of Illinois at Urbana-Champaign, Ms. Bliss also earned a J.D. at the University of Kansas.

2. How about Linked data and triples - making them a reality

Marjorie Hlava, Access Innovations

Presentation description: There is much discussion about the Linked Data, triples, RDF, TURTLE and other evolving protocols and standards. What about implementation?

Who is doing it? Where is this happening in the world outside of academia? Is it working? Is it scalable? Does text really benefit from this treatment? American Institute of Physics (AIP) took the bold step of moving from a traditional classification system to a full Z39.19 compliant thesaurus and then made it available as a basis for sharing throughout the physics, astronomy, and mathematics communities. Several subsidiary thesauri have been created including those for astronomy, acoustics, physics and others. It is now used to support search and editorial production for the AIPP (AIP Publishing). Mapping to the previous PACS (*Physics and Astronomy Classification Scheme*) codes (two editions) was completed. The legacy content conversion with automatic indexing at 95+% accuracy levels of over 900,000 articles is complete. A triple store of over one million triples has been implemented for search. Semantic fingerprinting for every article, institution and author is updated daily. Using a combination of MarkLogic, Data Harmony and Publishing technologies a cutting edge publishing and distribution platform is now in daily use. This paper describes that journey and lessons learned.

Short biography: Marjorie M.K. Hlava is President, Chairman, and founder of Access Innovations, Inc. She is past president of NFAIS and past president of the ASIS&T and the 1996 recipient of ASIST's prestigious Watson Davis Award. She has published more than two hundred articles and several books on information science topics. She also holds three patents. Her research areas include furthering the productivity of content creation and the governance layer for information access through automated indexing, thesaurus development, taxonomy creation, natural language processing, machine translations, and machine aided indexing.

3. Using KOS as the Connectors of Linked Datasets

Imma Subirats, Food and Agriculture Organization of the United Nations

Marcia Lei Zeng, Kent State University

Presentation description: This presentation provides evidence, examples, and use cases of using KOS as the backbones or connectors of linked datasets. The case is AGRIS (International System for Agricultural Science and Technology), a database with more than 7.7 million structured bibliographical records provided by the Food and Agriculture Organization (FAO) of the United Nations. In addition to providing common bibliographic services, AGRIS runs a mash-up Web application that links to related resources (bibliographic, statistical, factual, biographical, and geographic data, and country profile) from several other major resources on the Web. At the back-end, all metadata records are expressed in RDF triples. The Web application of AGRIS uses a KOS, the *AGROVOC* thesaurus, as its backbone. The published

multilingual thesaurus is completely encoded with SKOS. One of the two major approaches to interlinking is using AGROVOC formal alignments to other thesauri. AGROVOC is mapped to more than ten important KOS. The mapping results are integrated into the SKOS-XL encoded thesaurus entries for both human and machine's direct consumption.

While new applications of Linked Data KOS are enlightening, questions remain about datasets that have not used standardized KOS. The presentation provides a summary of methods that can be applied when generating or deriving project-oriented micro-thesauri and lists from the Linked Open Data KOS such as the faceted *Art and Architecture Thesaurus*. The presentation will also give an overview of the use cases of Linked Data KOS in information services.

Short biography: Imma Subirats has been working as knowledge and information management officer at the Food and Agriculture Organization of the United Nations (FAO) since 2006. She works on the advise of standards, tools and good practices for the management and exchange of data to academic, research, private and governmental institutions worldwide. She is also actively promoting open access and open data in the agricultural research context, notably through the CIARD network, assuring the quality of repository content through implementing metadata standards, thesauri, and other forms of authority control. In recent years, she has been working on the facilitation of the AIMS community and portal, space for accessing and discussing information management standards, tools and methodologies with the objective to connect information specialists worldwide to support the implementation of structured and linked information and knowledge.

Short biography: Marcia Zeng is professor of Kent State University. She has chaired and served on standards committees and working groups for IFLA, SLA, ASIST, NISO, and ISO. She has been a Working Group member of

the ISO 25964 Thesauri and Interoperability with Other Vocabularies. She recently completed her role as an invited expert for the W3C Linked Library Data Incubator Group and for the Asset Description Metadata Schema (ADMS) of the European Union. She currently serves as a member of the Executive Board of the International Society for Knowledge Organization (ISKO) and the Dublin Core Metadata Initiative (DCMI) Advisory Board. She is also Director-at-large of ASIS&T.

MODERATOR:

Maja Žumer, University of Ljubljana

(Moderator)

Short biography: Maja Žumer is Professor of Information Science at the University of Ljubljana (Slovenia). Her research interests include design and evaluation of information retrieval systems, end-user interfaces, and conceptual modeling. She has been involved in IFLA working groups, NISO committees, and several EU projects. She has received several international and national research grants. She also served as a member of IFLA FRBR Review Group, chair of the IFLA Working Group for Guidelines for National Bibliographies in the Digital Age and co-chair of IFLA Working Group on the Functional Requirements for Subject Authority Records (FRSAR). Currently she serves as chair of IFLA Classification and Indexing Standing Committee.

REFERENCES

AGRIS 2.0. FAO of the UN. Updated 2014/01/24.

<http://aims.fao.org/openagris>

Getty Vocabularies as Linked Open Data. 2014/02/19.

<http://www.getty.edu/research/tools/vocabularies/lod/>

* To whom all correspondence should be sent.