Developing a Dublin Core Application Profile for the Knowledge Organization Systems (KOS) Resources

by Marcia Lei Zeng and Gail Hodge

Knowledge Organization Innovation: Design and Frameworks

EDITOR'S SUMMARY

Since 2000, members of the Networked Knowledge Organization Systems/Services (NKOS) group have been working on specifications for metadata to describe the variety of knowledge organization systems (KOS), including classification schemes, thesauri and taxonomies. At the 2010 Dublin Core Metadata Initiative (DCMI) International Conference, a task group was formed to create a Dublin Core Application Profile (DCAP) for KOS to facilitate the use of KOS in networked settings. The DCAP details metadata used for a KOS application, addressing how the KOS is to function, the items and interrelationships described, the metadata terms used and their rules for use and the machine syntax used for encoding data. Building on past work, the task group is making progress toward its numerous goals, including production of a KOS Type Vocabulary, integrating the KOS taxonomy and KOS typology into a single workable product and ultimately testing the KOS DCAP and Typology on a variety of actual knowledge organization systems.

KEYWORDS

knowledge organization systems

metadata

Dublin Core

Marcia Lei Zeng is a professor at the School of Library and Information Science, Kent State University. She can be reached by email at mzeng<at>kent.edu.

Gail Hodge is a senior information scientist with Information International Associates, Inc. (IIa). She can be reached by email at ghodge<at>iiaweb.com.

The term *knowledge organization system* is intended to encompass all types of schemes for organizing information and promoting knowledge management. The acronym, KOS, has been used both as a singular (to represent a kind of resource that is different from other resource types) and a plural (when referring to KOS instances). Different families of KOS, including thesauri, classification schemes, subject heading systems and taxonomies, are widely recognized and applied in both modern and traditional information systems.

Various types of KOS have been increasingly embraced as (World Wide Web) services to facilitate resource discovery and retrieval. Different agents, services and applications need to communicate with KOS data in the form of transfer, exchange, transformation, mediation, migration and integration. It has been recognized that information about a KOS, including its data model, type, protocol, status, responsible body, available format, affectivity (at what time, location and/or use is the content applicable or valid) and other descriptive data, is integral to terminology registries, service registries, vocabulary users (machine or human) and retrieval systems. At a minimum level, metadata for KOS resources will describe specific characteristics of a KOS, facilitate the discovery of KOS resources, assist in the evaluation of such resources for a particular application or use, and facilitate sharing, reuse and collaboration. However, no widely accepted protocol for describing KOS resources currently exists.

In late 2009, members of the Networked Knowledge Organization Systems/ Services (NKOS) group proposed to develop a Dublin Core Application Profile (DCAP) for the KOS resources. The usage board of DCMI (Dublin

Core Metadata Initiative) formally endorsed the effort in 2010. A task group named DCMI/NKOS was formed in early 2010 with invited experts from various countries who have been active in the NKOS group [1, 2]. NKOS (http://nkos.slis.kent.edu/) is a community of more than 300 practitioners from more than 10 countries who are interested in the use of KOS in networked environments. The group's significant efforts have led to 18 successful full-day NKOS workshops held in both the United States and Europe since 1998.

In addition, the NKOS members have pioneered a number of specifications, including the following:

- *NKOS Registry - Draft Thesaurus Description Attributes* (1996-1998)
- *NKOS Registry Reference Document for Data Elements* (1998 – 2001)
- *Vocabulary Markup Language (ML): Metacode strawman DTD* (2000)
- *Taxonomy of Knowledge Organization Systems* (2000 and on)

The goals outlined by the DCMI/NKOS task group incorporate most of these products completed by NKOS members over the preceding years. The DCMI/NKOS task group is developing a DCAP (discussed below), which will be tested by professionals and researchers. Related to this DCAP is a *KOS Type Vocabulary* (also to be discussed). The task group first met at the 2010 DCMI International Conference (October 20-22, 2010, in Pittsburgh, Pennsylvania), where the group held a half-day workshop and a number of informal meetings.

The KOS Application Profile Components

According to the *Guidelines for Dublin Core Application Profile* [3], a DCAP is a document (or set of documents) that specifies and describes the metadata used in a particular application. To accomplish these goals, a profile does the following:

- Describes what a community wants to accomplish with its application (functional requirements)
- Characterizes the types of things described by the metadata and their relationships (domain model)
- Enumerates the metadata terms to be used and the rules for their use (Description Set Profile and Usage Guidelines)
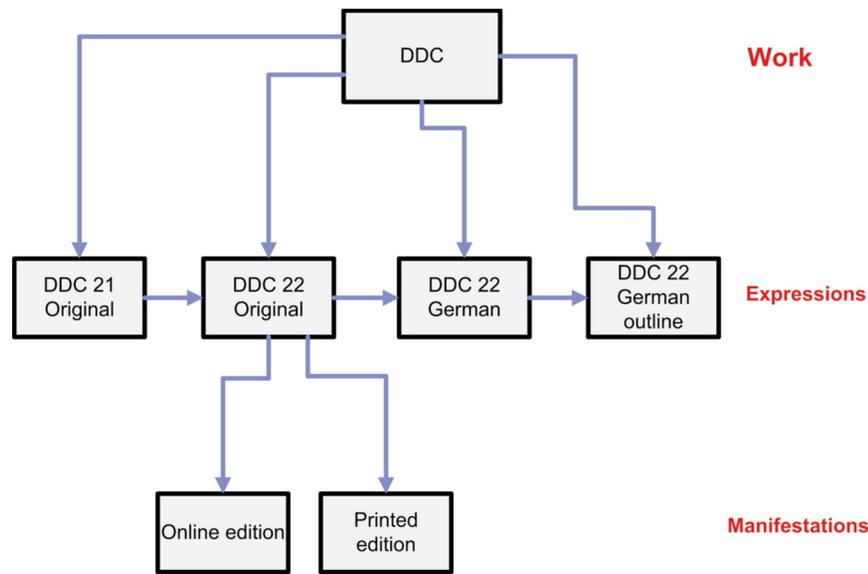
- Defines the machine syntax that will be used to encode the data (Syntax Guidelines and Data Formats)

The functional requirements. The DCAP requires a profile to be built on what a community wants to accomplish with its application. The task group has considered that the AP will be developed based on a) the characteristics of individual KOS instances, as well as b) the interactions between/among KOS instances, such as versioning, modularization, use and reuse, cross-referencing and multiple-representation. Initial usages of this AP would include terminology registries, service registries, vocabulary users (machine or human) and retrieval systems.

The NKOS group had, actually, initiated discussions regarding use case for KOS registries back in 1996. A draft set of thesaurus attributes was developed by Gail Hodge and revised in 1998 [4] based on the *Controlled Vocabulary Registry* developed by Linda Hill and Interconnect Technologies in 1996. The 1998 document specified functions from seven perspectives: product information, scope and usage, characteristic of descriptors, size of set of descriptors, labels for relationships, other product information, terms and conditions, and vendor/provider information. At the 2010 DCMI/NKOS workshop, Marjorie Hlava presented a strawman based on comprehensive research of existing registries [5]. Although the registries exhibit very different properties, Hlava was still able to identify and categorize many common, required properties found in five major groups: general information, scope and usage, vocabulary characteristics, terms and conditions, and provider. Many of these properties are also appropriate for other purposes such as vocabulary use (by both machines and humans) and retrieval systems.

The domain model. The domain model of an application profile aims to characterize the types of things described by the metadata and their relationships. The discussions of the task group on the domain model paid special attention to the needs to be emphasized in the functional requirements that are important to KOS cases, such as edition, translations, translations with modifications, adaptations and subsets. KOS instances may contain complicated relationships and present challenges for description, especially when translations are done for different versions of an original

FIGURE 1. DDC as an example interpreted with the FRBR model, with the original DDC as the work.

to encode the data (Syntax Guidelines and Data Formats) [8]. Previously, the NKOS group has worked on the metadata element sets, especially the *NKOS Registry Reference Document for Data Elements* developed by Diane Vizine-Goetz in 1998-2001 and updated in 2001 [9]. The data elements in the 2001 version are consistent with the Dublin Core Metadata Element Set (DCMES), with additional elements for details describing the characteristics of the basic components of a KOS scheme, such as entity type, entity value, relationships, information given (in terms of each concept in a scheme) and arrangement. Because the 2001 update was developed before the DCMI Metadata Terms (DCTERMS) was released, some qualifiers were used as elements in the 2001 version Vizine-Goetz developed. There is a possibility

FIGURE 2. A generalized DCMI Application Profile model based on FRBR model.
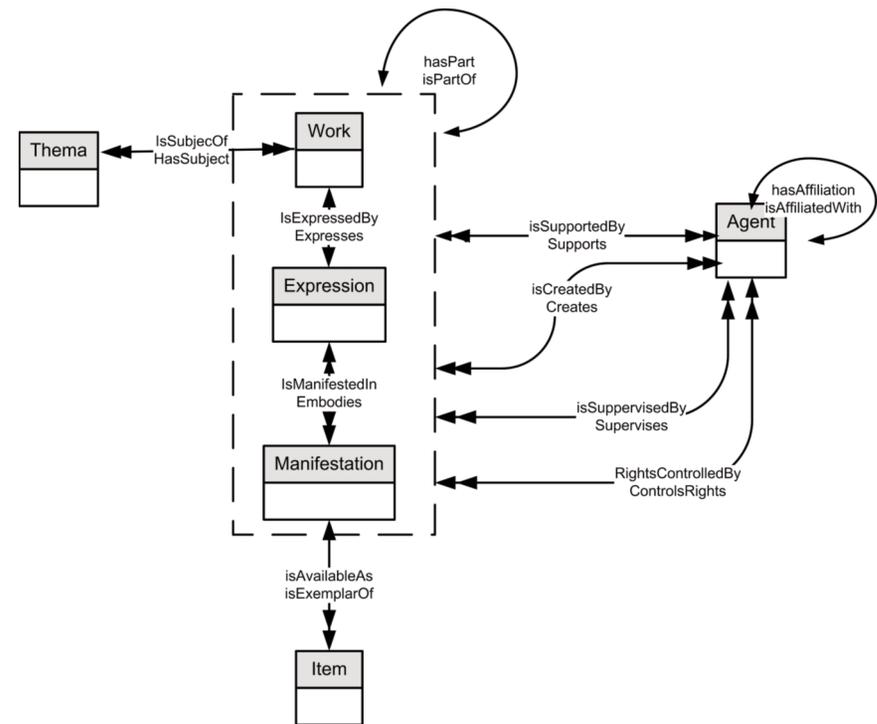

KOS instance. Prior to the KOS AP effort, there were other DCAPs (such as the *Scholarly Works Application Profile*, SWAP) [6] that used FRBR as the basic model. The DCMI/NKOS task group decided to follow SWAP's approach and use FRBR as the base for the domain model. Using the *Dewey Decimal Classification* (DDC) as an example, we can show alternative treatments of *work* and *manifestation*. For example, the original DDC can be considered as the work (Figure 1), or an individual version of DDC (for example, DDC 21) can be considered as the *work*.

On this topic, Maja Zumer presented the proposal of using a generalized FRBR AP model [7]. The General DCAP Domain Model declares relationships at the most general level. Given this approach the KOS AP will need to add, revise or eliminate specific relationships. (See Figure 2.)

Metadata terms. DCAP requirements include enumerating the metadata terms to be used and the rules for their use (Description Set Profile and Usage Guidelines) as well as defining the machine syntax that will be used

that the DCMI/NKOS team will revisit these elements, map them with the identified functional requirements and domain model (for example, what elements are for *work*, *expression*, *manifestation* or *item*), and re-map with DCTERMS, which will finally lead to the development of the description set profile.

The KOS Types Vocabulary

The DCMI/NKOS task group's charter includes producing a *KOS Type Vocabulary*. It will include various types of KOS, defined based on characteristics such as structure and complexity, the relationships between concepts and historical function.

Proposals for a taxonomy of KOS types (or taxonomy of taxonomies) were developed earlier by the NKOS group members Gail Hodge and Marjorie Hlava, respectively, and revised by Linda Hill and Marcia Zeng. The *Taxonomy of Knowledge Organization Sources/Systems* [10] adopted by the NKOS group based on Gail Hodge's monograph *Systems of Knowledge Organization for Digital Libraries* [11] categorizes KOS types into the three basic groups: term lists (authority files, glossaries, dictionaries and gazetteers), categorization and classification (subject headings, classification systems, taxonomies and categorization systems) and relationship groups (thesauri, semantic networks and ontologies). As more types were included (pick lists, directories, synonym rings) and consideration was paid to those that not only present the conceptual structures but also provide facts (name authority files and digital gazetteers), the taxonomy of KOS types was revised by Linda Hill and Marcia Zeng to include metadata-like models in addition to the term list group. The taxonomy and a visualized display can be found in the NKOS website [12]. All KOS types are distinguished through the two variables: the degree of control introduced (from natural language to controlled language) and the strength of their semantic structure (from weakly structured to strongly structured), corresponding to the major functions of KOS.

Doug Tudhope presented a KOS typology in 2006 in which he also began to consider KOS purposes and contexts of use in the typology [13]. For each typical type of KOS (classification, thesaurus) the typology framework would indicate the attributes such as entities, relationships (internal), typical application to objects in domain of interest and relationship applying concepts to objects in domain. The KOS typology has been extended to become a KOS spectra in 2010 [14]. It is a tentative typology of KOS that is based on the KOS and their context such as intrinsic/extrinsic characteristics and KOS types or instances such as essential/accidental) [15]. The typology and the previous KOS taxonomy complement each other. Together they provide options for typing/categorizing a KOS instance in a registry (or other use cases) based on the characteristics or structure/functions. Some of these characteristics are also important properties in describing a KOS instance.

Conclusion and Next Steps

The DCMI/NKOS task group's overriding goal is to establish an application profile for the KOS resources. This effort is building on and incorporating work completed by NKOS members over the last decade. Major progress was achieved during the DC-2010 conference, yet a great volume of work needs to be done before an application profile can be fully developed. At the current stage, the use case study has been limited to registries. Other use cases need to be collected. The proposed general domain model also needs to be tailored to meet the functional requirements for the KOS resources description. The metadata terms will be further revised based on the updated related standards. Finally, the KOS taxonomy and KOS typology need to be harmonized into a functional product. Both the element sets and KOS taxonomy will need to be tested by users once the DCAP draft is released. The DCMI/NKOS task group looks forward to collaborating with professionals and related groups. ■

Resources Mentioned in the Article

[1]   *DCMI/NKOS Task Group official webpage:* http://dublincore.org/groups/nkos.

[2]   *DCMI/NKOS Task Group working wiki:* www.metadataetc.org/wiki/dcmi-nkos/doku.php.

[3]   Coyle, K., & Baker. T. (2009). *Guidelines for Dublin Core Application Profiles.* Dublin Core Metadata Initiative (DCMI). Retrieved March 6, 2011, from http://dublincore.org/documents/profile-guidelines/.

[4]   *NKOS Registry – Draft Set of Thesaurus Attributes.* (Last modified July 30, 1998). Retrieved March 6, 2011, from http://nkos.slis.kent.edu/Thesaurus_Registry.html.

[5]   Hlava, M. K. (October 21, 2010). NKOS Task Group – KOS functional requirements. Presented at the *DC2010 DCMI-NKOS Task Group Meeting, Pittsburgh, PA.*

[6]   *Scholarly Works Application Profile* (SWAP). (Last modified 2009). SWAP Working Group, JISC Digital Repositories Programme. Coordinated by Julie Allinson and Andy Powell. www.ukoln.ac.uk/repositories/digirep/index/Scholarly_Works_Application_Profile.

[7]   Zumer, M., Zeng, M.L., & Salaba, A. (2010). FRBR: A generalized approach to Dublin Core Application Profiles. In *DC-10: International Conference on Dublin Core and Metadata Applications* (pp.21-30). DCMI. Retrieved March 6, 2011, from http://dcpapers.dublincore.org/ojs/pubs/article/viewArticle/1024.

[8]   Coyle, K., & Baker, T. (2009). *Guidelines for Dublin Core Application Profiles.* DCMI. http://dublincore.org/documents/profile-guidelines/.

[9   *NKOS Registry – Reference document for data elements.* (June 21, 2001). (3rd version). Retrieved March 6, 2011, from http://nkos.slis.kent.edu/registry3.htm and http://staff.oclc.org/~vizine/NKOS/Thesaurus_Registry_version3_rev.htm.

[10] *Taxonomy of Knowledge Organization Sources/Systems* [web page]: http://nkos.slis.kent.edu/KOS_taxonomy.htm.

[11] Hodge, G. (2000). *Systems of knowledge organization for digital libraries: Beyond traditional authority files.* Washington, DC: Council on Library and Information Resources. Retrieved March 6, 2011, from www.clir.org/pubs/reports/pub91/contents.html.

[12] The figure is extracted from Zeng, M. L. (2008). Knowledge organization systems. *Knowledge Organization, 35* (2-3), 160-182.

[13] Tudhope, D. A. (2006, September 21). Tentative typology of KOS: Towards a KOS of KOS [PowerPoint slides]? Presented at the *5th European Networked Knowledge Organization Systems (NKOS), 10th ECDL Conference, Alicante, Spain. September 21, 2006.* Retrieved March 6, 2011, from www.ukoln.ac.uk/nkos/nkos2006/presentations/tudhope.ppt.

[14] Rocha Souza, R., Tudhope, D., & Barcellos Almeida, M. (2010). The KOS spectra: A tentative typology of knowledge organization systems. In *Paragdigms and conceptual systems in knowledge organization: Proceedings of the 11th ISKO International Conference, Rome, 2010* (pp.122-129). Wurzburg: Ergon Verlag. Retrieved March 6, 2011, from www.ergonverlag.de/en/start.htm?information-_library_sciences_advances_in_knowledge_organization_2.htm.

[15] DCMI-NKOS Task Group. (2010). *Kos_typology_2010.* Retrieved March 6, 2011, from www.metadataetc.org/wiki/dcmi-nkos/doku.php?id=kos_typology_2010. Extract from [13].