

A Domain Model for Describing and Accessing KOS Resources—Report of the Development of a KOS Dublin Core Application Profile

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

Marcia Lei Zeng
Kent State University,
USA
mzeng@kent.edu

Marjorie M. K. Hlava
Access Innovations, Inc.
USA
mhlava@accessinn.com

Abstract

Due to the dynamic and complex characteristics of knowledge organization systems (KOS), the need for a multi-layered model to present the complex relationships among KOS resources is widely recognized. This paper reports on a continuous effort to establish such a model. Prior to this paper, a general Dublin Core Application Profile (DCAP) model was proposed using the FRBR (Functional Requirements for Bibliographic Records) conceptual model as the foundation. The current study applies the general DCAP model to KOS. It introduces a domain model that is suitable for describing and accessing KOS resources and incorporates the core attributes that the NKOS (Networked Knowledge Organization Systems) group has surveyed and summarized. The domain model includes the entities, relationships between entities, and the core attributes of the entities in the context of user tasks. The results reflect access-related properties that are complementary to the previous research by NKOS group members that focused on description-related metadata properties commonly found in KOS registries.

Keywords: application profile; domain model; KOS properties; vocabulary registries

1. Introduction

Knowledge organization systems (KOS) encompass a wide range of types of structured vocabularies that represent concepts within certain knowledge domains. They are mainly used for resource and information organization, information retrieval, and knowledge representation. Classification systems, categorization systems, taxonomies, and thesauri are the most well known examples of KOS. The problem addressed by this paper is the description of, and access to, a KOS as a whole (i.e., a “concept scheme” as referred to by the SKOS Simple Knowledge Organization System Reference (SKOS, 2009)). For the concepts, labels, and relationships among concepts included in a KOS, another conceptual model has been established and is described in the Functional Requirements for Subject Authority Data (FRSAD) report (FRSAD, 2011).

1.1. The Need for a Multi-layered Modeling Approach for KOS Application Profile

In developing a Dublin Core application profile (DCAP) for describing KOS resources, the DCMI-NKOS Task Group (2010-) has been following the requirements set by the Guidelines for Dublin Core Application Profiles (Coyle and Baker, 2009) which requires the establishment of a domain model that characterizes the types of things described by the metadata and their relationships. While it seems obvious on the surface that any KOS resource (such as a thesaurus or a classification scheme) may be described using DC terms, a closer look would reveal their unique characteristics and necessity for extension.

The need for developing a DCAP for the KOS resources is a consequence of the dynamic nature of KOS. Unlike many of the published works, almost all KOS resources need to be continuously developed. A KOS scheme or system would lose its value and credibility if not constantly updated. In addition to micro-level updates, new versions with a significant amount of changes may be regularly released. This paper will use the example of the *ASIS&T Thesaurus*. Its

third edition has had extensive extensions of the coverage, content, and organization, and has different authors from the first and second editions.

Like the *ASIS&T Thesaurus*, the later versions of a KOS scheme are usually based on previous ones, fully or partially, while the authorship may change. More significantly, the KOS works are usually not developed or used as stand-alone resources. Reuse, mapping, re-alignments, and derivation are common use cases. It is important to know the relationships among the different KOS works to enable implementation and interoperability.

Translation of a KOS is far more complex than translation of other types of works. Translations of a KOS can be symmetrical, locally tailored, or selective. Adding to the complexity of the relationships within KOS work families are the formats, in which they are available: the content of a thesaurus or a classification scheme is usually stored in databases, with available outputs in multiple formats.

Furthermore, the translation, extraction, and reuse could be at different levels or limited to a subset of the original work. Using FRBR (Functional Requirements for Bibliographic Records, 1998) terminology, they can be: an expression of a work in a different language, a translation of an English version (an expression), a selected translation of a French version which is a translation itself (an expression based on another expression), or an extraction of a classification work that is partially released as RDF triples for Linked Open Data purpose (a specific manifestation).

All of these dynamic and complex characteristics call for a multi-layered model to present the complex relationships among KOS resources. As the first step, the general model for a DCMI Application Profiles (AP), based on FRBR, was proposed in 2010 (Figure 1) (Žumer, Zeng, and Salaba, 2010). It resulted from a study of existing domain models for APs.

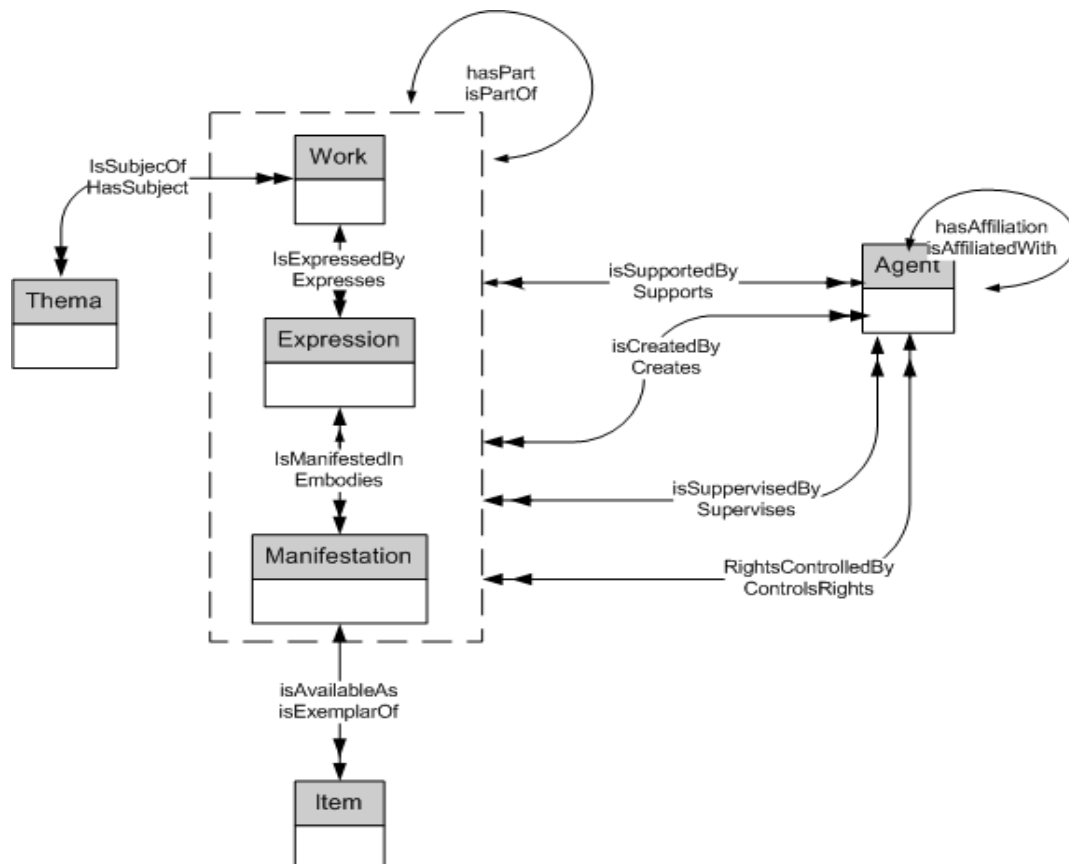


FIG. 1. General domain model

In the current study we have applied this general AP model to the specific domain of KOS and incorporated the core attributes that the NKOS group have surveyed and summarized during the last decade.

2. Modeling KOS with FRBR

KOS resources as intellectual products can be modeled using FRBR. Without getting into a discussion of whether a new version is a new work or a new expression when significant changes of title, content, and authorship occur, we can illustrate this model using the ASIS&T Thesaurus as an example.

2.1. The Entities and Relationships

The ASIS&T Thesaurus was first published in 1994. In addition to the new versions (created by different authors), multiple translations (translated by different translators) have been published or used internally. The thesaurus has been released for different needs in the online and Web environment with various formats. When modeling them according to the FRBR entities (work, expression, and manifestation), we get:

- the thesaurus as a whole is a work (the identification starts with “W”);
- different versions (such as Version 1994 in English, Version 2005 in English, Version 2012 in French) are different expressions of this work (the identification starts with “E”); and
- the printed edition of the 2010 English version and the SKOS Linked Data representation of the same version, which are examples of manifestations (the identification starts with “M”).

A subset of expressions and manifestations of this work is illustrated in Table 1. Attributes and relationships are examples only.

TABLE 1: A subset of entities for the *ASIS&T Thesaurus*

Entity	ID	Relationships (examples)	Attributes (examples)
ASIS&T Thesaurus	W1		Type of KOS: thesaurus Subject: information science; information technology; librarianship
Original <i>ASIS Thesaurus of Information Science and Librarianship</i>	E1	realises W1 createdBy J. Milstead	Language: English CreationDate:1994
2 nd edition, 1996	E2	realises W1 basedOn E1 creatdBy J. Milstead	Language: English CreationDate:1996
3 rd edition, 2005 <i>ASIS&T Thesaurus of Information Science, Technology, and Librarianship</i>	E3	realises W1 basedOn E2 createdBy A. Redmond-Neal and M. Hlava publishedBy: Information Today	Language: English CreationDate: 2005
French version of 3 rd edition, ASIS&T Thesaurus in French	E4	realises W1 translationOf E3 createdBy Mabrouka el Hachani	Language: French CreationDate: 2012
Arabic version of 3 rd edition	E5	realises W1 translationOf E3 createdBy inprogress	Language: Arabic Expected CreationDate: 2012

Spanish version of 3 rd edition	E6	realises W1 translationOf E3 createdBy inprogress	Language: Spanish Expected Creation Date: 2012
German version of 3 rd edition	E7	realises W1 translationOf E3 createdBy inprogress	Language: German Expected Creation Date: 2012
3 rd edition, database	M1	embodies E3 publishedBy Access Innovations, Inc. on behalf of ASIS&T; distributed by Information Today Inc.	Format: CD ROM using Data Harmony Thesaurus Master
3 rd edition, database	M2	embodies E3 publishedBy Access Innovations, Inc. on behalf of ASIS&T	Format: Standards-compliant XML file for query from other applications
3 rd edition, printed	M3	embodies E3 publishedBy Access Innovations, Inc. on behalf of ASIS&T	Number of pages: 344 Format: PDF file
3 rd edition, SKOS Linked Data	M4	embodies E3 publishedBy Access Innovations, Inc. on behalf of ASIS&T	Number of triples: ... Format: XML file
3 rd edition, Web reproduction	M5	embodies E3 publishedBy Access Innovations, Inc. on behalf of ASIS&T	Number of entries: ... Format: HTML webpages

2.2. User Tasks

To develop the domain model, we identified three major types of users for KOS registries (also known as “terminology registries”): (a) KOS developers, (b) information retrieval system developers, and (c) end-users (including all other users). The AP must meet their specific needs of both description and access. In analyzing their general use tasks, we consider that:

- The owner(s)/creator(s) of a KOS would want to publish, share, and allow reuse and mapping of their work. They register and publish their systems and thus expose the KOS product(s) to interested parties.

The role of a producer and a user may switch throughout the whole process. From a user’s (rather than a developer’s) point of view, the use cases can be summarized as the following:

- Other KOS developers may be interested in an existing KOS for reuse or as examples of good practice. They may create derivative works based on an existing KOS.
- Information retrieval system (IRS) developers may want to reuse, implement, and evaluate a KOS, as well as to apply a KOS to a collection to support searching and/or navigation.
- End users and researchers may be involved in terminology-related research and exploration within a subject domain. They may want to evaluate, align, or compare KOS resources.

No matter what the end goals are, all users will need to find, identify, select, obtain, and explore KOS resources. Therefore, the appropriate metadata and its presentation need to support these tasks. These needs and contexts can be directly interpreted using the FRBR user tasks (with FRSAD extension):

- using the data to find a KOS that corresponds to the user's stated search criteria (e.g., in the context of a search for all KOS on a given subject, or a search for a KOS issued under a particular title);
- using the data retrieved to identify a KOS (e.g., to confirm that the KOS described in a record corresponds to the document sought by the user, or to distinguish between two KOS products or two editions that have the same title);

- using the data to select a KOS that is appropriate to the user's needs (e.g., to select a KOS in a particular language, or to choose a release of a KOS that is compatible with the hardware and operating system available to the user);
- using the data in order to acquire or obtain access to the KOS described (e.g., to place a purchase order or to access online an electronic KOS product stored on a remote computer);
- using the data to explore the different KOS that are available in a registry (e.g., get acquainted with the subject coverage of a KOS or discover available KOS in a specific domain).

2.3. Core Relationships Between Entities

To support these tasks, attributes and relationships need to be defined. Proposed core relationships and attributes are listed in Table 2.

TABLE 2: Core relationships between entities

Entity	Core relationships	Extensions	Examples
Work	Based on (a work)		“Répertoire de vedettes-matière (RVM)” is based on the “Library of Congress Subject Headings”
	Part of (a work)		“Abridged Dewey Decimal Classification” is part of the full “Dewey Decimal Classification”
Expression	Realizes (a work)		Original “ASIS Thesaurus of Information Science and Librarianship” realises “ASIS&T Thesaurus”
	Part of (an expression)	Outline of	Outline of “Dewey Decimal Classification”
		Excerpt of	Class 780, Music, of “Dewey Decimal Classification”
	Based on (an expression)	Translation of	“ASIS&T Thesaurus of Information Science, Technology, and Librarianship 3 rd edition” in a different language
		Abridgement of	Top levels only of “ACM Computing Classification System (CCS)” used for web navigation tree
		Extension of	Class Q, Medicine, of “Library of Congress Classification” expanded for specialized use
Manifestation	Embodies (an expression)	Version of	Thesaurus B’s 5-year review and revision
		Other relationship (with an expression)	Thesaurus B mapped with Thesaurus C, resulting in a table of mapping
	Part of (a manifestation)	Mapped with	LCSH2011-04-26 expressed in SKOS embodies LCSH modified 2011-04-26
Manifestation	Embodies (an expression)		Thesaurus A implemented in a retrieval system
			Volume II of “Library of Congress Subject Headings, 30th edition (2007)”

Relationships are essential for the explore task, but may also support the identify task.

2.4. Mapping the NKOS Core Attributes to the User Tasks

The core attributes have been proposed by the NKOS group members based on the analysis of existing KOS repositories during the past several years. They were discussed at the DCMI-NKOS group workshops and other informal meetings during 2010 and 2011. This section attempts to assign these core attributes to the FRBR entities (work, expression and manifestation) and to align them with the user tasks. This is the foundation of an application profile. An integrated presentation of all these attributes in relation to the user tasks they support can be found at the wiki site of the DCMI-NKOS Task Group.¹

TABLE 3: *Work* attributes
(The core attributes are listed first, followed by extensions at the bottom.)

Core Attributes of <i>Work</i>	Supporting User Tasks:				
	Find	Identify	Select	Obtain	Explore
Title	x	x			
Identifier	x				
Description		x	x		
Type of KOS	x	x	x		
Creator	x	x			
Rights	x		x	x	
Subject	x	x	x		
Relationships to other entities		x			x
Audience	x	x	x		
Supporting documentation			x		

TABLE 4: *Expression* attributes and relationships

Core Attributes of <i>Expression</i>	Supporting User Tasks:				
	Find	Identify	Select	Obtain	Explore
Title	x	x			
Identifier	x				
Contact				x	
Description		x	x		
Creator	x	x			
Language	x	x	x		
Size		x	x		
Date updated	x		x		
Rights	x		x	x	
Relationships to other entities		x			x
Used by			x		
Frequency of update			x		
Audience	x	x	x		
Supporting documentation			x		

¹ http://wiki.dublincore.org/index.php/Core_Elements

TABLE 5: *Manifestation* attributes and relationships

Core Attributes of <i>Expression</i>	Supporting User Tasks:				
	Find	Identify	Select	Obtain	Explore
Title	x	x			
Identifier	x				
Contact				x	
Description		x	x		
Creator	x	x			
Publisher	x	x	x		
Format	x	x	x		
Rights	x		x	x	
Relationships to other entities		x			x
Used by			x		
Services offered			x		
Supporting documentation			x		

In this proposal, some attributes (e.g., title, identifier, creator) can be found at different levels. In other words, all three entity types (work, expression, and manifestation) can have these attributes. It should be noticed that, in each particular case, the attribute values would be different. Let's refer to Table 1 examples:

- The thesaurus as a whole is a work. Its title is "ASIS&T Thesaurus".
- Different versions (such as Version 1994 in English, Version 2005 in English, Version 2012 in French) are different expressions of this work. Each expression has its own title.
- The printed edition of the 2010 English version and the SKOS Linked Data representation of the same version are examples of manifestations, with different titles.

3. Conclusion

The multi-layered model presented in this paper reflects the requirements for describing and accessing KOS resources that have unique, dynamic, and complex characteristics. Researchers in the DCMI-NKOS Task Group have previously done comprehensive research based on the description-related metadata elements commonly found in KOS registries (Golub and Douglas, 2008). This paper takes a complementary approach, starting from a consideration of the user tasks and putting the attributes that are important to work, expression, and manifestation in the context of user tasks. Using the FRBR family models and terminology, the domain model included in the proposed DCAP brings both description and access aspects together. The next step will be to align the NKOS core attributes with the DCMI terms. The Task Group will also provide templates for describing KOS resources in registries under the same model. A test will be conducted to use the DCAP for KOS resources in describing a wide range of variant KOS instances, such as the multilingual or mixed-language schemes, the derived expressions of an original work, the aggregated products of multiple KOS works, the database-driven systems (in which the KOS management tends to be on the individual concept and label level, rather than on the whole system level), and various releases of a born-digital concept scheme.

References

- ASIS&T Thesaurus of Information Science, Technology, and Librarianship. (2005). Third Edition. Edited by Alice Redmond-Neal and Marjorie M. K. Hlava. Information Today, Inc.
- Coyle, Karen and Thomas Baker. (2009). Guidelines for Dublin Core Application Profiles. <http://dublincore.org/documents/profile-guidelines/>
- DCMI/NKOS Task Group official webpage (2010-): <http://dublincore.org/groups/nkos/>;
- DCMI/NKOS Task Group Working wiki (2010-): http://wiki.dublincore.org/index.php/DCMI_NKOS_Task_Group
- Golub, Koraljka and Douglas Tudhope. (2008). JISC Terminology Registry Scoping Study (TRSS) report. <http://www.jisc.ac.uk/media/documents/programmes/sharedservices/trss-report-final.pdf>
- Functional Requirements for Subject Authority Data, A Conceptual Model (FRSAD). (2011). IFLA Working Group on Functional Requirements for Subject Authority Records (FRSAR). Eds. Zeng, Marcia L, Maja Zumer, and Athena Salaba. Berlin/Munich: De Gruyter Saur.
- Functional Requirements for Bibliographic Records - Final Report. (1998). IFLA Study Group on the Functional Requirements for Bibliographic Records (FRBR). Munich: K.G. Saur.
- SKOS Simple Knowledge Organization System Reference. (2009). W3C Recommendation, 18 August 2009. Eds. Miles, Alistair and Sean Bechhofer. <http://www.w3.org/TR/skos-reference/>
- Žumer, Maja, Marcia Lei Zeng, and Athena Salaba. (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. <http://dcpapers.dublincore.org/ojs/pubs/article/viewArticle/1024>.