Establishing correspondences between ISO-25964 and SKOS constructs

Johan De Smedt
TenForce
Belgium

Antoine Isaac
Europeana & Vrije Universiteit Amsterdam
The Netherlands

Marcia Lei Zeng
Kent State University
USA

Stella Dextre Clarke
Project Leader, ISO NP 25964
UK

This presentation reports an initiative to establish correspondences between the data model expressed in the newly published ISO 25964-1, *Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval,* and the constructs of the W3C *SKOS Simple Knowledge Organization System.* ISO 25964-1 replaces ISO 2788 and ISO 5964, the standards that influenced SKOS' design from the beginning. Since an Appendix of the *SKOS Primer* (prepared by Antoine Isaac and Ed Summers) presents a correspondence table between SKOS constructs and the withdrawn standards ISO 2788 and 5964, this is now due for updating. Responding to the call by the ISO 25964 Working Group chair Stella Dextre Clarke, an ad-hoc task force led by Johan De Smedt has therefore been developing a new correspondence table between ISO 25964 and the SKOS/SKOS-XL Models. Based on this work, the ad-hoc task force and the editor of *SKOS Primer,* Antoine Isaac, are also preparing a proposal for a SKOS/XL extension that would capture the semantics defined by ISO 25964-1.

For years, thesauri have been used mostly in contexts where humans controlled or mediated the search process. In the standards and guidelines prior to ISO 25964, the distinction between concepts (on the one hand) and the terms representing concepts (on the other) was made in a rather fuzzy way. It has been realized that a data model would help to enforce consistency in thesaurus construction and mapping, and to enable greater interoperability between thesauri and other vocabularies. Compliance with a data model would also enhance performance at all stages from design of the thesaurus through development, management, and exchange. With the demand for (and under the pressure of) machine understandable and processable thesaurus data in the 21st century, a data model of thesaurus emerged in the newer standard developments, first through the British Standard BS 8723, *Structured vocabularies for information retrieval* Part 5 (also known as DD 8723-5). The model was further refined in the ISO 26964-1 and released in 2011. Meanwhile, in the course of the last decade, the general SKOS data model has also been developed, to address the need for exchanging data in a wide range of knowledge organization systems.

The ISO 25964 data model and the SKOS data model are largely compatible, particularly when the SKOS-XL extension for labels is taken into account. Concepts are the central modeling primitive of both ISO 25964 and SKOS. This has been especially emphasized in ISO 25964 and expressed through its data model. Although the tags "BT", "NT" and "RT" are retained (because

they are widely used in many existing thesauri), the standard clarifies that the relationships these tags indicate are between concepts, not terms. The data model of ISO 25964 is built on six basic classes: *Thesaurus* (mapped to *skos:ConceptScheme*), *ThesaurusArray* (a subclass of *skos:Collection), ThesaurusConcept (*mapped to *skos:Concept), ThesaurusTerm (*mapped to *skos-xl:Label), Note* (mapped to the property *skos:note*,) and *VersionHistory* (proposed to be mapped to a data set version history). In order to enable successful thesaurus exchange in a wide range of situations, the ISO standard provides for sophisticated as well as simple thesauri. Therefore, although a rather simple model would be enough to meet the basic recommendations required by the standard, the ISO 25964 data model has much greater complexity so as to accommodate a number of optional features. Some of them (such as relationships between terms) are not covered in SKOS. A SKOS extension may be able to handle those extra features as well as some more specific types of semantic relationships. This presentation hopes to give an update of the status of such an effort and engage colleagues of the NKOS community in further discussions.

References

DD8723-5, Data Model for BS 8723. In: *Structured vocabularies for information retrieval. Part 5. Exchange formats and protocols for interoperability*. London: British Standards Institution. Available at: http://schemas.bs8723.org/Model.aspx

ISO 25964-1:2011, *Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval*. Geneva: International Organization for Standards, August 8, 2011.

ISO 25964-1 Schema and Data Model. http://www.niso.org/schemas/iso25964/#schema

W3C Recommendation, *SKOS Simple Knowledge Organization System Reference.* W3C Recommendation, August 18, 2009. Alistair Miles and Sean Bechhofer, eds. Available at: http://www.w3.org/TR/skos-reference/

W3C Recommendation. *SKOS eXtension for Labels (SKOS-XL)*. In: SKOS Simple Knowledge Organization System Reference, Appendix B. W3C, August 18, 2009. Available at: http://www.w3.org/TR/2009/REC-skos-reference-20090818/#xl

Correspondences between ISO-2788/5964 and SKOS constructs. In: SKOS Simple Knowledge Organization System Primer. W3C Working Group Note 18 August 2009. Antoine Isaac, A. and Ed Summers, eds. Available at: http://www.w3.org/TR/skos-primer/#seccorrespondencesISO