Integrating Knowledge Organization Systems Registries with Metadata Registries

Hugo Manguinhas[1], José Borbinha[1],

[1] INESC-ID – Instituto de Engenharia de Sistemas e Computadores,
Apartado 13069, 1000-029 Lisboa, Portugal
{hugo.manguinhas, jlb}@ist.utl.pt

Abstract. Metadata Registries (MDR) are information systems defined to manage and publish metadata related to information and data models. An inherent part of the data models registered within an MDR is the vocabularies used to assign values to data elements. These vocabularies relate to concepts that can be managed by Knowledge Organization Systems (KOS). This paper discusses how an MDR can import and take advantage of the information managed within a KOS.

Keywords: MDR, ISO11179, KOS.

1 Introduction

An **information model** is an abstract description of how information is represented in a specific context (a business area, an organization or a community), consisting therefore in the definition of the relevant terms, the relations between them, and the vocabularies or rules to assign values to them. A **data model** (schema) is the representation of an information model for a specific syntax (XML Schema, SQL-DDL, etc.) and/or storage method (file, database, etc.), which defines the data elements and arrangements between them.

The information needed to describe information and data models has been traditionally called metadata (in opposition to the term "data", to refer to the data objects representing the instances of the information entities). Associated to that resulted the concept of **Metadata Registry** (MDR), defined as an information system to manage and publish metadata. A MDR is thus a central location (e.g. repository) where metadata is stored and maintained in a controlled environment, which usually happens inside a specific organization, but it also can be the case of a multi-organizational scenario.

An inherent part of the data models registered within an MDR are the vocabularies used to assign values to data elements. In fact, in an MDR these vocabularies are separated into two different spaces: a "conceptual space", which defines the terms (concepts); and a corresponding "data or code space", which defines the values used to encode those concepts. A typical example of a vocabulary is for example the ISO

639 standard for languages codes[1] (which can be considered as the conceptual space) which relate to two code spaces (the 2 letter and 3 letter codes). As an example of its use, is the 2 letter code space used as range for the "xml:lang" attribute[2] of XML formats. Another similar example is the ISO 3166 standard for country codes[3] used in MARC formats to encode the country of the publication.

In a broader view, we argue that a conceptual space managed within a MDR is similar to the information defined within a thesaurus, taxonomy, classification scheme or ontology managed by a Knowledge Organization System (KOS). Considering this assumption, it would be beneficial to define open and flexible ways to integrate the information managed within KOS into an MDR.

2 Integration of a KOS registry with an ISO/IEC 11179 compliant MDR

The ISO/IEC 11179[4] is the standard that defines the concepts behind an MDR, addressing "the semantics of data", "the representation of data", and "the registration of the descriptions of that data". Considering that information models in a specific context can be designed reusing other existing information models (especially in the case of defining profiles) which already may be managed by other MDR systems, the sharing of information among registries is becoming a key issue. The ISO 20944 - Metadata Registries Interoperability and Bindings (MDR-IB) family of standards addresses that problem.

As mentioned before, the information managed within a KOS also can relevant for reuse in a MDR, thus raising a new interoperability challenge.

The development of KOS registries is currently still at its beginning, nevertheless it has been gathering a lot of interest within the community to make that information sharable. Concerning interoperability among metadata and KOS registries, we believe such could be achieved through the following approaches:

- In a first approach, if we can realize that a KOS service was implemented according to the ISO 11179 series of standards (as proposed by the XMDR project [1]), than that integration would be immediately possible through the service interfaces defined by the ISO 20944 - MDR-IB.
- A second approach would be to build a wrapper for the KOS registry, making it MDR-IB compliant. This would be the easiest approach to implement.
- A third approach would be to import to the KOS information present in the KOS registry within the MDR, through a service's interfaces. This would be eventually harder, if no standard interface for KOS registries is defined.

As discussed in the introduction, we argue that the information managed by KOS relate to the information managed in a MDR, more precisely to its conceptual space. In ISO 11179, the conceptual space is represented by the Data Element Concept and

[1] http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=22109
[2] http://www.w3.org/XML/1998/namespace
[3] http://www.iso.org/iso/english_country_names_and_code_elements
[4] http://metadata-standards.org/

Conceptual regions (see Figure 1), while the data space is represented by the Value Domain and Data Element regions. Note that the conceptual space can also be defined using the Classification Domain only if it were to be used as a classification scheme within the MDR. Another kind of information that is of interest to this issue is the relationships between the terms defined within KOS. This information is commonly represented in an ISO 11179 through the "ConceptRelationship" defined as relationships between "Concept" classes. A potential usage of this information is to clarify other relationships that may be defined between data elements that can related to those concepts.

A particular requirement for KOS information to be integrated into an MDR is to be uniquely identified within the KOS registry, which must assure that it does not change in time. This is essential to guarantee an effective integration between them.

Fig. 1. Class Diagram representing the subset of the ISO 11179 metamodel for the Conceptual Domain region.

1 Conclusion

This paper discusses the possibility of MDR systems to take advantage of the information manage in KOS registries through the integration of the two registries. It concludes that this integration can be achieved and with minimum effort.

References

1. Bargmeyer, B.: eXtended Metadata Registries (XMDR). Presentation at the 7th NKOS Workshop at JCDL 2005. http://nkos.slis.kent.edu/2005workshop/Bargmeyer.ppt