

Achieving interoperability for metadata registries using comparative object modeling

Yu Rang Park^a, Ju Han Kim^{a, b}

^aSeoul National University Biomedical Informatics (SNUBI),

^bDivision of Biomedical Informatics, Seoul National University College of Medicine Seoul 110-799, Korea

Abstract

Achieving data interoperability between organizations relies upon agreed meaning and representation (metadata) of data. For managing and registering metadata, many organizations have built metadata registries (MDRs) in various domains based on international standard for MDR framework, ISO/IEC 11179. Following this trend, two public MDRs in biomedical domain have been created, United States Health Information Knowledgebase (USHIK) and cancer Data Standards Registry and Repository (caDSR), from U.S. Department of Health & Human Services and National Cancer Institute (NCI), respectively. Most MDRs are implemented with indiscriminate extending for satisfying organization-specific needs and solving semantic and structural limitation of ISO/IEC 11179. As a result it is difficult to address interoperability among multiple MDRs. In this paper, we propose an integrated metadata object model for achieving interoperability among multiple MDRs. To evaluate this model, we developed an XML Schema Definition (XSD)-based metadata exchange format. We created an XSD-based metadata exporter, supporting both the integrated metadata object model and organization-specific MDR formats.

Keywords:

Metadata registry, ISO/IEC 11179, Interoperability, Object model

Introduction

Metadata provide data with context and are used to facilitate the understanding, usage and management of data both by human and computers. Meta Data Registry (MDR) is a system for managing these metadata contents. For managing and describing metadata, ISO/IEC JTC 1/SC 32 established a standard for MDR, named ISO/IEC 11179 [1], on which many MDRs in various domains have been built and introduced. In biomedical domains, there are two representative MDRs, United States Health Information Knowledgebase (USHIK) from U.S. Department of Health & Human Services [2] and cancer Data Standards Registry and Repository (caDSR) from National Cancer Institute (NCI) [3]. For evaluating metadata functionality, we also developed an MDR, named Clinico-Histopathological Metadata Registry (CHMR), for describing

and managing data elements related to clinical trial and research data [4].

Recently, several studies have shown semantic and structural limitations of ISO/IEC 11179 [5-7]. They are categorized into three problems. The first problem is the absence of semantic or syntactic linkage of shared concepts between components [6]. The second and third ones are structural limitations. ISO/IEC 11179 doesn't support a structure for metadata extension [5] and usage model [7]. To solve these problems and to satisfy organization-specific needs, most MDRs are built with indiscriminate extending. As a result, metadata inconsistency problem makes it difficult to support interoperability among MDRs. In other words, it results structural and semantic inconsistencies between MDRs. In this paper, we are focusing only on structural inconsistency between MDRs. To address the structural inconsistency problem, we propose an integrated metadata object model and metadata exchange format. Also we implement a metadata exporting system.

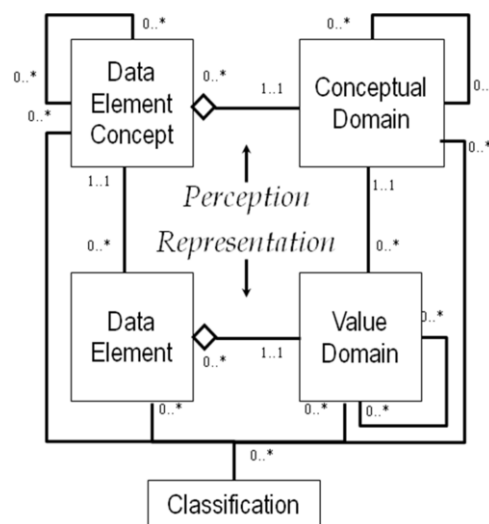


Figure 1- High-level meta-model of ISO/IEC 11179

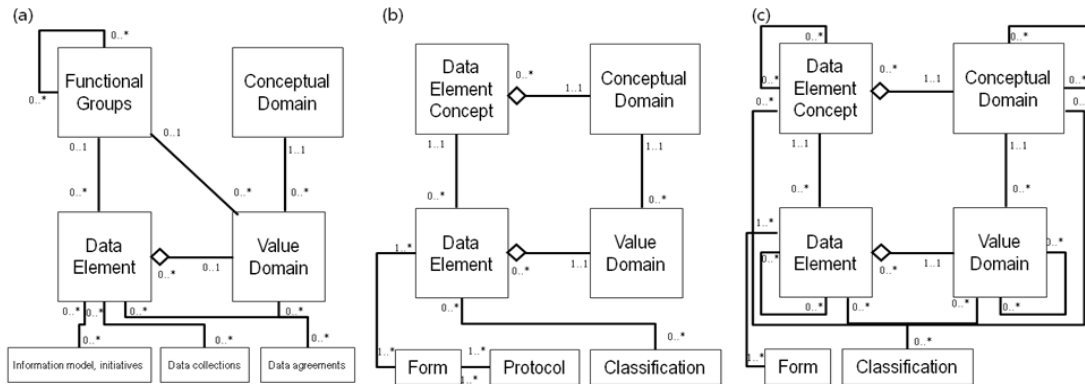


Figure 2- Reconstructed scheme for Metadata Registries (a) USHIK, (b) caDSR and (c) CHMR

Background: ISO/IEC 11179

ISO/IEC 11179 is an international standard for describing and registering metadata. Figure 1 shows the most important five components of ISO/IEC 11179, Data Element, Data Element Concept, Value Domain, Conceptual Domain and Classification. The Data Element is foundational concepts in ISO/IEC 11179 model and an associated component between Data Element Concepts and Value Domain. Data Element Concept and Value domain is a set of attributes describing conceptual meaning and representational characteristics of data, respectively. Conceptual domain and Value Domain can be viewed as logical and physical code sets, respectively. Conceptual Domains support Data Element Concepts and Value Domains support Data Elements. Classification provides classification scheme to associate the components of ISO/IEC 11179.

Materials and Methods

We select two public MDRs (USHIK and caDSR) with our own MDR called CHMR. USHIK and caDSR do not provide direct access to database schema or structure of metadata. For reconstructing the scheme of MDRs, we analyzed the XML files exported from caDSR and the metadata content structure of USHIK. Finally, we designed an integrated metadata object model (IMOM) and XML Schema Definition (XSD) to solve the structural inconsistency among MDRs. We downloaded 14,331, 39,572 and 13,835 data elements from USHIK, caDSR and CHMR, respectively, and stored them in an IMOM-enabled database, having 67, 738 data elements in total. We also implemented a system exporting metadata as XML files in organization-specific and IMOM formats regardless of the source of the metadata.

Results

Structural differences among MDRs

We reconstructed the scheme of MDRs (Figure 2) and compared the differences between MDRs. Relatively large number of modifications from the ISO/IEC definitions were found in USHIK and caDSR. Two of the most common modifications are as follows. First, most self associations are eliminated in USHIK and caDSR (Figures 2(a) and (b)). Second, newly defined component, named “Form,” which is a set of data elements, is added in caDSR and CHMR (Figure 2(c)). Organization-specific modifications are as follows. USHIK renames Data Element Concept as “Functional Groups” and changes required associations to optional ones between Data Element and Functional Groups and Data Element and Value Domain. New associations were added between Functional Groups and Value domain (Figure 2(a)). caDSR added “Protocol” that represents a collection of “Form” (Figure 2(b)). CHMR added a self association in Data Element to describe the relationship between Data Elements (Figure 2(c)).

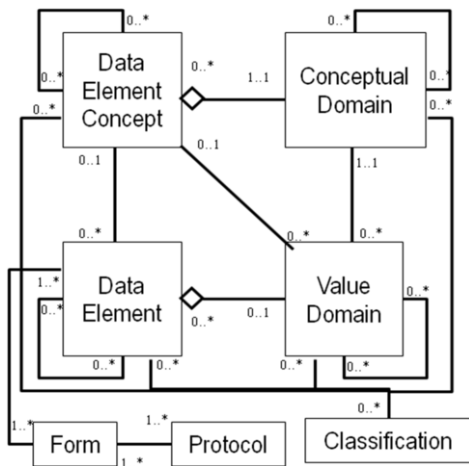


Figure 3- Integrated model

Integrated metadata object model and metadata exchange format

For covering all of the modifications in the three MDRs, we designed IMOM illustrated in Figure 3. The full object model represented in UML is available at <http://bmesh.snubi.org/MetadataExchange/>. There are two basic concepts of IMOM design. First, “do not eliminate organization specific extensions.” Second, “basic attributes of ISO/IEC 11179 should be maintained.” As a result, IMOM has 64 classes including five new classes related to the “Form” and “Protocol” components. Cardinalities of several associations were modified and four new associations were added.

To support metadata sharing among MDRs, we designed two types of XSD-based metadata exchange formats, one based on the full IMOM covering the entire attributes including organization-specific ones and the other covering the basic attributes only. The basic attributes are minimum specification for describing metadata defined by ISO/IEC 11179. Figure 4 shows a part of IMOM-based XSD file for metadata exchange. Both types of XSD files are available at <http://bmesh.snubi.org/MetadataExchange/>

Metadata Exporter

To evaluate IMOM and metadata exchange formats, we developed a metadata exporter (available at <http://bmesh.snubi.org/MetadataExchange/>), supporting five XSD formats including the three organization-specific MDR formats and the two XSD formats based on IMOM and basic attributes.

Conclusion and Discussion

Recently, many organizations and research groups pay attention for extending and improving metadata contents [8-10]. Most efforts have been made for only one MDR or a domain. Achieving interoperability is important not only for an MDR but also for between MDRs. However, until now, research trends on metadata managing have been focused on just one MDR. The problem related to achieving interoperability among multiple MDRs should be researched.

In this paper, we proposed an integrated metadata object model and exchange format to solve the problem related to the structural inconsistencies among MDRs. We also implemented a metadata exporting system to support exchanging in multiple MDRs. To the best of our knowledge, none of the previous works have considered the interoperability among multiple MDRs. Our effort provides a basic framework to solve the structural inconsistency problems for achieving interoperability among multiple MDRs. In the next step of our research, we will focus on semantic inconsistency problems in multiple MDRs.

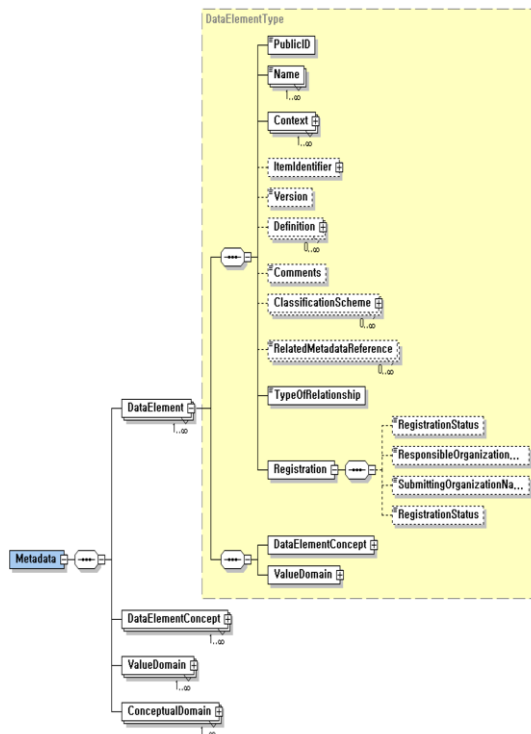


Figure 4- The fragment of integrated model XML Schema Definition file

Acknowledgments

This study was supported by grant from the Korea Health 21 R&D Project, Ministry of Health, Welfare and Family Affairs, Republic of Korea (0405-BC02-0604-0004). Educational training of Y.R.P was supported by a grant from the Korean Pharmacogenomics Research Network (A030001), Ministry of Health & Welfare, Korea

References

- [1] ISO/IEC JTC 1/SC 32. ISO/IEC 11179: Information technology-metadata registries-part 1~6. ISO/IEC JTC 1, 2003.
- [2] United States Health Information Knowledgebase. Available at <http://ushik.ahrq.gov/> Accessed October 12, 2009.
- [3] Komatsoulis GA, Warzel DB, Hartel FW, et al. caCORE version 3: Implementation of a model driven, service-oriented architecture for semantic interoperability. J Biomed Inform. 2008 Feb; 41(1)106-23.
- [4] Park YR, Kim JH. Metadata Registry and management system based on ISO 11179 for cancer clinical trial information system. AMIA Annu Symp Proc. 2006:1056.

- [5] Nadkarni PM, Brandt CA. The common data elements for cancer research: remarks on functional and structure. *Methods Inf Med.* 2006;45(6):594-601.
- [6] Davies J, Harris S, Crichton C, et al. Metadata standards for semantic interoperability in electronic government. In *Proceedings of the 2nd International Conference on Theory and Practice of Electronic Governance 2008*. Cairo, Egypt ACM New York, NY, USA.
- [7] Solbrig, HR. Metadata and the reintegration of clinical information: ISO 11179. *MD Comput.* 2000 17(3): 25-8.
- [8] Kunz I, Lin MC, Frey L. Metadata mapping and reuse in caBIG. *BMC Bioinformatics.* 2009 Feb 5; 10 suppl2:S4.
- [9] Kwon JH, Song CY, Moon CJ, et al. Bridging real world semantics to model world semantics for taxonomy based knowledge representation system. *J Comput Sci & Technol.* 2005 20(3):296-308.
- [10] Lim JE, Choi OH, Na HS, et al. A methodology for semantic similarity measurement among metadata based information system. In *Proceedings of the Fourth International Conference on Software Engineering Research 2006*. Seattle, Washington, USA.

Address for correspondence

Ju Han Kim, M.D., Ph.D.
Chief, Division of Biomedical Informatics
Seoul National University College of Medicine
Seou 110799, Korea
Email: juhan@snu.ac.kr
WWW: <http://informatics.snu.ac.kr/>