MEDINFO 2013
C.U. Lehmann et al. (Eds.)

doi:10.3233/978-1-61499-289-9-618

Establishing Semantic Interoperability of Biomedical Metadata Registries using Extended Semantic Relationships

Yu Rang Park[a], Young Jo Yoon[a], Hye Hyeon Kim[a], Ju Han Kim[a, b]

[a] *Seoul National University Biomedical Informatics (SNUBI), Division of Biomedical Informatics,*
[b] *Systems Biomedical Informatics Research Center, Seoul National University, Seoul 110-799, Republic of Korea*

Abstract

Achieving semantic interoperability is critical for biomedical data sharing between individuals, organizations and systems. The ISO/IEC 11179 MetaData Registry (MDR) standard has been recognized as one of the solutions for this purpose. The standard model, however, is limited. Representing concepts consist of two or more values, for instance, are not allowed including blood pressure with systolic and diastolic values. We addressed the structural limitations of ISO/IEC 11179 by an integrated metadata object model in our previous research. In the present study, we introduce semantic extensions for the model by defining three new types of semantic relationships; dependency, composite and variable relationships. To evaluate our extensions in a real world setting, we measured the efficiency of metadata reduction by means of mapping to existing others. We extracted metadata from the College of American Pathologist Cancer Protocols and then evaluated our extensions. With no semantic loss, one third of the extracted metadata could be successfully eliminated, suggesting better strategy for implementing clinical MDRs with improved efficiency and utility.

Keywords:

Metadata registry, ISO/IEC 11179, interoperability, Semantic relationship, object model.

Introduction

Achieving semantic interoperability is fundamental and critical for sharing biomedical information between individuals, organizations, and systems [1]. Metadata is a key component of interoperability for data exchange, mapping, and interpreting across various domains. A metadata description and registration standard was established by ISO/IEC, named ISO/IEC 11179 [2]. The ISO/IEC 11179 Metadata Registry (MDR) standard provides a semantically precise structure for metadata. Many organizations have built MDRs based on this standard for managing and registering metadata. Following this trend, many healthcare organizations have increasingly adopted and implemented MDRs for managing data semantics for biomedical research, Electronic Health Record (EHR) systems, and clinical trials [3-8]. We implemented a MDR, named Clinico-Histopathological Metadata Registry (CHMR), for the data management of clinical research and trials [9].

Recently, several studies have shown semantic and structural limitations of ISO/IEC 11179 [10-13]. The limitations of ISO/IEC 11179 can be categorized as the following. As the scope and coverage of an MDR increases, most of the data elements are newly defined and not reused from existing data elements [11]. Accordingly, the number of related data elements will increase. Another limitation is the granularity of data element concepts; all concepts exist at a single level, with no means of inter-relating [11, 13]. The standard model provides no structure for semantic and syntactic relationships between related concepts and between parent-child relations or sub-components.

These limitations are mainly due to the single semantic perspective. The single semantic perspective dictates that the semantics of metadata are determined by an association of object class and properties in the data element concept. The single perspective weakens the semantic representation of metadata. Thus one suffers from finding appropriate metadata in an MDR, resulting low metadata reusability. Although the standard allows for multiple Classification Scheme and Derived Data Elements, the implication is that no addition to the semantics of the administered items is recommended.

To address the single perspective problem, additional structural and semantic extension to ISO/IEC 11179 may be needed. The structural limitations were addressed in our previous study by extending the standard model, the Integrated Metadata Object Model (IMOM) [14]. The present study targets the semantic limitations of ISO/IEC 11179.

Methods

In our previous study, in an effort to strengthen the semantic relationship between data elements, we extended the ISO/IEC 11179 model. Two classes were added, a self-association class, named data_element_relationship and a new class Data_Element_Relationship with two attributes, data_element_relationship_type and data_element_relationship_type_description.

In the present study, we define a semantic relationship that can be applied to the data_element_relationship_type attribute in the Data_Element_Relationship class. Defining semantic relationship between concepts is a well-established topic in the development of controlled vocabularies. According to the definition of ANSI/NISO 239.19 'Guidelines for the construction, format, and management of monolingual controlled vocabulary', there are three types of semantic relationships; equivalency, hierarchy, and associative [15].

The equivalency and hierarchy relationships are found in ISO/IEC 11179. Two data element that have the same data element concept would be in an equivalency relationship (synonymy). A derived data element seems to be in a hierarchy relationship (generic or instance). We subdivide the

associative relationship into three different relationships: dependent, composite, and variable. One data element can employ multiple relationships with different data elements. Detailed definitions of these associative relationships will be described in the results section.

To evaluate the semantic relationships of metadata, we used a clinical contents standard - College of American Pathologists (CAP) cancer protocol (October 2009 version) [16]. We chose six among the 55 CAP cancer protocols. For the purpose of evaluation the six protocols were all part of the head and neck category. The evaluation process consisted of the following three steps. First, we manually extracted metadata from the six CAP cancer protocols. Then the extracted metadata was integrated according to the ISO/IEC 11179. Finally, we applied the semantic relationships to the metadata and compared the number of metadata for each step for evaluation.

Results

Three types of semantic relationships

According to the ISO/IEC 11179 standard, one data element is a logical combination of a data element concept and a value domain. This structure is limited. Concepts consisting of two or more values are not allowed, for instance, 'blood pressure' should be represented by two values; systolic pressure and diastolic pressure. To overcome this and other limitations, we defined three new types of semantic relationships between metadata.

(1) **Dependency relationship**. Some clinical data elements are supposed to be activated or deactivated by the response (value) of a different data element. They are regarded as conditionally dependent. This type of link should be called a dependency relationship. Figure 1 shows an example of dependency relationship. The data element "DE: Margin types of invasive carcinoma" is activated only when the response to the data element "DE: Margins of kidney carcinoma" is "Margin(s) involved by invasive carcinoma".

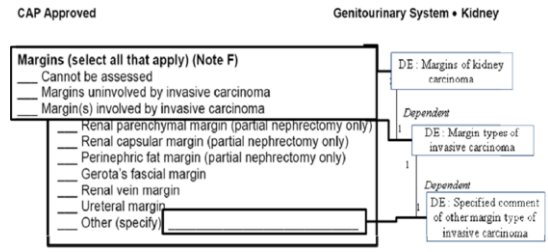

Figure 1 - Dependency relationship between data elements for DE: Margins of kidney carcinoma

(2) **Composite relationship**. Composite data elements are a set of data elements linked by a composite relationship. Figure 2 shows an example of a composite data element for the histological types of larynx carcinoma. The larynx consists of various tissue types. Due to this, various types of tumor and tissue specific histology growth may occur in the larynx. Therefore the composite data element "DE: Histology types of Larynx carcinoma" consists of the three sub data elements, "DE: Histology type's variants of Squamous Cell Carcinoma", "DE: Histology types of Neuroendocrine carcinoma", and "DE: Histology types of Minor Salivary Gland Carcinoma".

(3) **Variable relationship.** A variable data element is a data element that is linked to an appropriate value list for the corresponding variable. A variable data element represents any value data elements such as medications or laboratory tests. Figure 3 shows an example of variable data elements for medications. According to the standard model, one has to define a data element for an adverse drug event for each medication. These 'adverse drug event' data elements have no semantic or syntactic difference except for the specific medication name. One variable-enabled data element referring to a dictionary of medications in the form of 'DE: Adverse Drug Reaction of (drug) X' can replace all the 'adverse drug event' data elements. For medications, we used the RxNorm (version UMLS 2009AA) [17] standard.

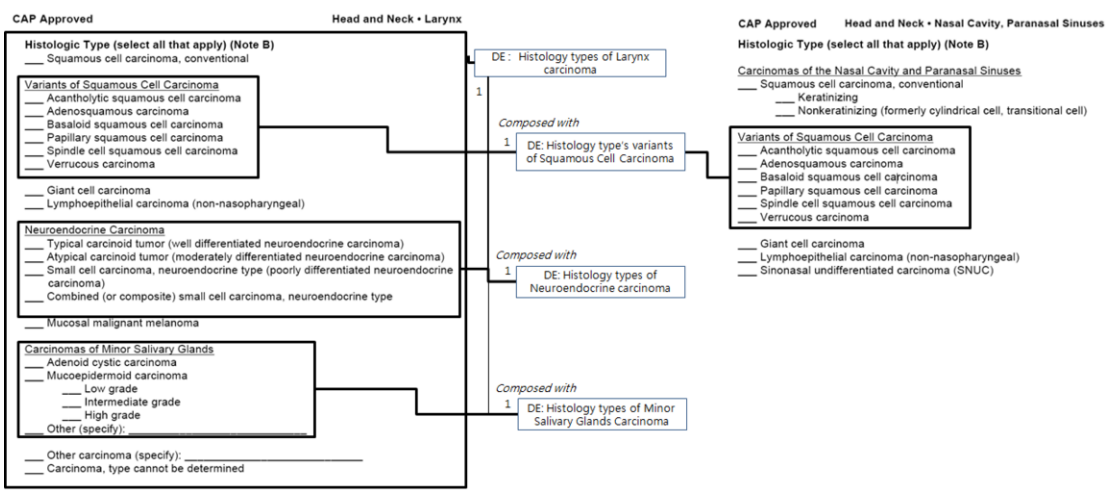

Figure 2 - Composite relationship between data elements for DE:Histologic types of Larynx carcinomas

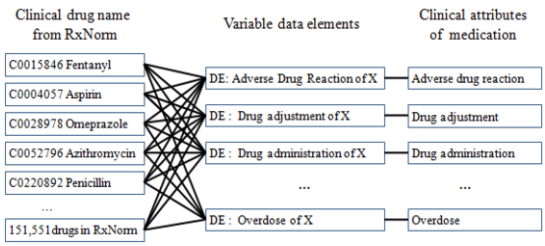

Figure 3 – Variable relationship for drugs

Evaluation

To evaluate the semantic relationships, we applied the extended relationships on the metadata from the six head and neck CAP cancer protocols. The evaluation process consisted of the following three steps; metadata extraction, metadata integration, and semantic-relationship application. We counted the number of metadata generated during each step as a measure of structural efficiency. The structural efficiency represents a MDRs ability to organize data elements. Table 1 shows the evaluation results. The first row shows the total numbers of raw metadata extracted for each cancer protocol. The second row shows the numbers of metadata that are integrated by the ISO/IEC 11179 manual. The numbers of integrated metadata were reduced by 27% compared to the numbers of extracted metadata. The third row shows the numbers of metadata after the application of the semantic relationships. The numbers of metadata were reduced by 66% compared to the numbers of the original metadata extraction.

Table 1 – Reducing the number of the data elements by introducing semantic relationships

CAP Cancer protocol	Larynx	Lip and Oral Cavity	Major Salivary Glands	Nasal Cavity and Paranasal Sinuses	Pharynx	Thyroid	Total
Metadata Extraction	79	85	59	81	89	91	484
Metadata integration	67	73	50	71	75	57	357
Application of semantic relationships	54	53	45	57	66	52	167

Discussion

The ISO/IEC 11179 MDR standard has been recognized as one of the most powerful solutions for achieving semantic interoperability in biomedical domains. Several studies, however, have demonstrated the semantic and structural limitations of the MDR standard. To solve the structural limitations, we previously proposed an object model, IMOM [14]. In this study, we addressed the semantic limitations of the MDR standard by defining extended semantic relationships.

Traditional MDRs do not define or register any associative relationships between data elements. In this study, the semantic relationships between data element can be maintained by extending the standard model (IMOM) by defining three associative relationships. The evaluation results indicated that applying semantic relationships to metadata helps to reduce the overall number of metadata without semantic loss.

Reduction of the number of data elements at the metadata integration and semantic-relationship application stages implies different meanings in terms of efficiency and utility. The former means improved reusability of data elements and the latter demonstrates improved semantic representation of related data elements. Redundant data elements were efficiently eliminated during the integration step, resulting in an increased chance for the reuse of already-defined data elements. Organizing data elements into more efficient and semantically richer structures during the semantic-relationship application step reduced the numbers of data elements to an even greater degree. Introducing these two steps to the organization of data elements may improve the efficiency and utility of MDRs.

The eXtended MetaData Registry (XMDR) project was established to propose extensions to the ISO/IEC 11179 family of metadata registry standards to promote more diverse types of metadata and enhanced capability for semantic specifications and queries [18]. The intension of the XMDR project, however, was not to solve the limitations of the standard model. Thus, the XMDR model has the same limitations as ISO/IEC 1179. Davis et al. [13] provide a modified standard model to satisfy the specific requirements of electronic governance. The modified model was simplified from the standard model by eliminating two main components; Conceptual Domain and Data Element Concepts. This model may support the implementation of electronic governance systems, but the model cannot interoperate with other standard-based MDRs.

Conclusion

In the present study, we directly addressed the intrinsic semantic and structural limitations of ISO/IEC 11179. Previously we mitigated the structural limitations by introducing an integrated object model and exchange format. In this study, we attempted to overcome the semantic limitations by defining associative relationships for MDRs. To our best knowledge, there has been no previous research that has successfully overcome the limitations of the standard model. Our present effort provides a foundation for a solution to the inconsistency problems in a single or multiple MDRs. The present study, however, is limited in proper evaluation of the variable relationship due to the unavailability of drug information in CAP cancer protocols. Future work should focus on comparisons of efficiency and expressiveness between controlled vocabularies and metadata. Comparisons of the complexity and usability between our extended model and the standard ISO/IEC 11179 model should also be addressed.

Acknowledgments

This research was supported by a Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012-0000994) with YRP's educational grant (2012R1A6A3A01011585).

References

[1] Jiang G, Solbrig HR, Iberson-Hurst D, et al. A Collaborative Framework for Representation and Harmonization of Clinical Study Data Elements Using Semantic MediaWiki. AMIA Summits Transl Sci Proc, 2010. 2010: p. 11-5.

[2] ISO/IEC JTC 1/SC 32. ISO/IEC 11179: Information technology-metadata registries-part 1~6. ISO/IEC JTC 1, 2004

[3] Stover PJ, Harlan WR, Hammond JA, et al. PhenX: a toolkit for interdiscipli-nary genetics research. Curr Opin Lipidol, 2010. 21(2): p. 136-40.

[4] Clinical Data Interchange Standards Consortium. Clinical Data Acquisition Standards Harmonization: Basic Data Collection Fields for Case Report Forms. [cited December 8, 2012]; Available from: http://www.cdisc.org/standards/cdash/downloads/CDASH_STD-1_0_2008-10-01.pdf.

[5] Buetow KH and Niederhuber J. Infrastructure for a learning health care system: CaBIG. Health Aff (Millwood), 2009. 28(3): p. 923-4; author reply 924-5.

[6] Loring DW, Lowenstein DH, Barbaro NM, et al. Common data elements in epilepsy research: development and implementation of the NINDS epilepsy CDE project. Epilepsia, 2011. 52(6): p. 1186-91.

[7] Warzel DB, Andonaydis C, McCurry B, et al. Common data element (CDE) management and deployment in clinical trials. AMIA Annu Symp Proc, 2003: p. 1048.

[8] Liu D, Wang X, Pan F, et al. Harmonization of health data at national level: a pilot study in China. Int J Med Inform, 2010. 79(6):450-8.

[9] Park YR and Kim JH. Metadata registry and management system based on ISO 11179 for Cancer Clinical Trials Information System. AMIA Annu Symp Proc, 2006: p. 1056.

[10] Solbrig HR, Metadata and the reintegration of clinical information: ISO 11179. MD Comput, 2000. 17(3): p. 25-8.

[11] Nadkarni PM and Brandt CA, The Common Data Elements for cancer research: remarks on functions and structure. Methods Inf Med, 2006. 45(6): p. 594-601.

[12] Richesson RL and Nadkarni PM, Data standards for clinical research data collection forms: current status and challenges. J Am Med Inform Assoc, 2011. 18(3): p. 341-6.

[13] 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

[14] Park YR and Kim JH. Achieving interoperability for metadata registries using comparative object modeling. Stud Health Technol Inform, 2010. 160(Pt 2): 1136-1139.

[15] NISO Standard (ANSI), Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies, in 5.3.3 Semantic Relationships 2005.

[16] Ghossein R, Update to the College of American Pathologists reporting on thyroid carcinomas. Head Neck Pathol, 2009. 3(1): p. 86-93.

[17] Bennett CC, Utilizing RxNorm to support practical computing applications: capturing medication history in live electronic health records. J Biomed Inform, 2012. 45(4): p. 634-41.

[18] eXtended MetaData Registry (XMDR) Project. [cited March 26, 2013]; Available from: http://en.wikipedia.org/wiki/XMDR.

Address for correspondence

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