Cross-walking Health Content Standards Using the ISO/IEC 11179 Metadata Registries Standard

By Denise B. Warzel, MSc, and Dianne M. Reeves, RN, MSN

Key challenges in clinical and translational biomedical research and personalized medicine include the need to integrate multidisciplinary, heterogeneous data sets. This requires both reusable data elements and content standards that unambiguously describe data within and across diverse data sources.

Models for content standards describing how to record and store data elements for later retrieval and use, and models for content standards comprised of data elements are being described differently by different healthcare and medical research communities, such as the Intermountain Healthcare/GE Healthcare (Intermountain) Clinical Data Model, and the International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) 11179 Metadata Registry Standard (ISO/IEC 11179). The use of different models for describing content standards and data standards can inhibit semantic interpretability and interoperability. This article describes and compares the ISO/IEC 11179 and the Clinical Data Model.

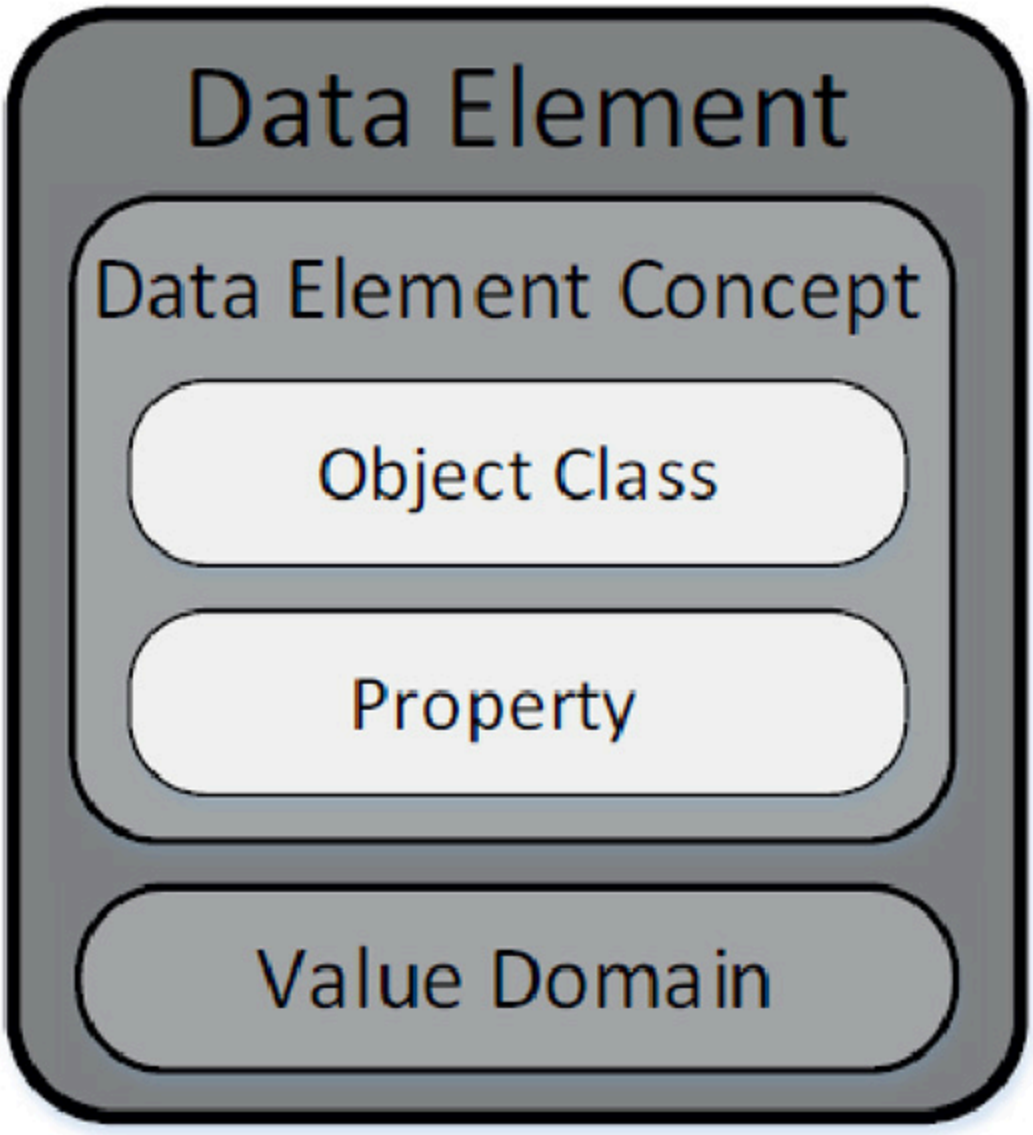
Standard Representation of Semantic Content

Standards can be established for individual data elements like height and weight, but sometimes standards need to describe a set of data elements grouped together, which is called a content standard, like the data elements needed to collect "blood pressure." The National Cancer Institute (NCI) and Intermountain recognized the need to define collections of data elements that are tightly related in order to standardize data collection and create content standards.

To enable semantic interoperability across groups establishing content standards, it is important that the model for describing the content standard is compatible. Intermountain uses the Clinical Data Model, developed in conjunction with GE Healthcare, for expressing clinical element models, which are content standards, while NCI uses ISO/IEC 11179. One key difference between these content model standards is that the Clinical Data Model includes the metadata describing the data and the actual data, whereas the ISO/IEC 11179 is only the metadata describing the data—the structure does not include the actual data.

The ISO/IEC 11179 standard states that a data element is comprised of two parts—its meaning and its representation: 1-Data Element Concept (semantics) and 2-Value Domain (representation). In turn, the Data Element Concept is comprised of two parts as well: Object Class and Property. Figure 1 below illustrates the ISO/IEC 11179 data element model.

Figure 1: Data Element Model



Source: International Organization of Standardization (ISO) and International Electrotechical Commission (IEC). "Information technology - Metadata registries. Part 1: Framework. Draft International Standard. ISO/IEC 11179-1:2014(E), Metadata Registries Framework." http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html.

At Intermountain, an effort was undertaken to explore the opportunity to use the ISO/IEC 11179 metadata registry standard to register the clinical element models standards, which could serve to facilitate the translation of expressions of semantically similar data elements between and across healthcare and research communities through the use of the data element model in Figure 1. Such a capability would enhance the opportunity to discover unforeseen links between clinical and translational research data and the potential for secondary use in precision medicine.

ISO/IEC 11179 includes the data element derivation, which is a structure or mechanism to describe a content standard that is composed of one or more data elements. The data element derivation includes a rule and method for combining the data elements. Like data elements, data element derivations can be used for consistent registration and dissemination of content standards such as those described by clinical element models.

In 2000, NCI adopted the practice of recording content standards based on the ISO/IEC 11179 model to enable language- and application-independent representation and redistribution of standards. NCI calls these structures common data elements. Common data elements are entered into NCI's ISO/IEC 11179-based metadata registry entitled the cancer Data Standards Registry and Repository (caDSR).[1] Common data elements are linked to standard terminologies, such as the NCI Thesaurus,

LOINC, and SNOMED, which provide common and unambiguous semantics and representational rules wherever they are used. For common data elements that are required within a given clinical context, such as "blood pressure," the ISO/IEC 11179 data element derivation was considered, but due to gaps in its features NCI extended ISO/IEC 11179 to capture case report form modules more fully describing the content standard. Case report form modules contain information from common data elements that describe each question text, permitted data values, relevant units of measure, default values, instructions, ordering, and dependencies between questions. Case report form modules are composed into standard case report form templates, defining the related content that is important to the correct usage of the data (i.e., attribution and provenance). Case report form standard templates are shared across thousands of NCI clinical trials ensuring that data are captured exactly the same way.

Intermountain recognized that collections of data elements—composed into case report forms in a standard registry, such as caDSR—enhance widespread access to standard-based content standards as well as the ability to identically interpret, use, and re-use such content, thus increasing semantic interoperability. While NCI case report forms are used by the NCI Clinical Trials Network and other NCI stakeholders, the case report forms model, though an extension of ISO/IEC 11179, is not a widespread healthcare IT standard.

Harmonizing NCI and Intermountain Content Standards Models

Intermountain explored the use of ISO/IEC 11179 data element derivations as opposed to clinical data models to capture the "BloodPressurePanel clinical element models" and again found the data element derivation model lacking certain features needed to support content standards.[2]

Clinical element models can be composed of one or more content standards that include features such as whether the item is mandatory or optional, cardinality, qualifiers, and modifiers making them more like information models. Qualifiers narrow the meaning of the choices or data value, but don't change its meaning. Modifiers are elements that must be used in conjunction with the data because they significantly modify its meaning.[3] The BloodPressurePanel clinical element models contain three smaller clinical element models: systolic blood pressure, diastolic blood pressure, and mean arterial pressure, each of which includes contextual, attribution, and provenance elements such as the qualifiers for body location, body position, method or device used, and a modifier, which is the subject or patient information.

Note that a single attribute in ISO/IEC 11179, unit of measure, is described by a structure consisting of 34 attributes in a Clinical Data Model. The attribute named "unit" in the Clinical Data Model, if literally mapped to the ISO/IEC 11179 unit of measure attribute, would end up with a value of "physical quantity" instead of an actual unit of measure such as millimeters. Only after expanding the physical quantity unit into its lower level attributes do we find the unit of measure comparable to that which is required in ISO/IEC 11179. There is also a lot of information in the common data elements for the clinical element model for which there is not a placeholder in ISO/IEC 11179.

It is sufficient for the case of cross-walking between a content standard specified using Intermountain Clinical Data Model and one specified using ISO/IEC 11179, that the essential ISO/IEC information—units of measure—can be extracted from the Clinical Data Model structure, even though it is not a one-to-one mapping. Other ISO/IEC 11179 data element attributes were populated using the BloodPressurePanel clinical element modeling formation by directly matching attribute values, while others required human interpretation. A few examples are shown in Table 1 below.

Table 1: Partial Mapping of ISO/IEC 11179 Attributes to Clinical Element Model (CEM) Attributes

ISO/IEC 11179 Attribute	CEM Attribute
Unique Identifier	Key

Version	N/A
Preferred Name	CEM Name, Qualifier Name, Modifier Name
Preferred Definition	Definition, or Information
Value Domain Representation	N/A
Value Domain Enumerated Permissible Values: Value and Value Meaning • *Note: When this experiment was first conducted for an AMIA Poster in 2014, the CEM value sets did not include a Value, but a Value has been added and is visible in the CEM Browser as of the writing of this article*[4]	ValueSet Value and Value Meaning
Object Class	Type + Human interpretation of the CEM Name
Property	Key + Human interpretation of the CEM Name
Reference Document	CEM Attribution
Designations	CEM Key, CEM XML Tag, CEM Name

The ISO/IEC 11179 attributes that describe the semantics of the clinical element model were added when curating the clinical element model as a data element derivation in caDSR. The primary semantic component is the data element concept defining the conceptual domain of the common data elements, and is linked to terminology or ontologies in two parts: the Object Class, describing the thing in the real world that the data element is about—in this case “Blood Pressure”—and the Property naming the characteristic being measured or observed. These semantic components were created by reading the clinical element model information which is represented as text. The Value Domain Representation concept annotations were also a result of human interpretation of the clinical element model, in this case the representation class for the overarching BloodPressurePanel clinical element model was determined to be “Measurement.”

Though NCI does not define a blood pressure panel (a panel is a set of tests frequently collected together), as noted, the authors were able to correlate many of the clinical element model BloodPressurePanel attributes with ISO/IEC 11179 attributes, including manually discerning the appropriate Object Class and Property.[5] The Object Class was defined as the concept “Blood Pressure” linking it to the NCI Thesaurus, with concept identifier C54706. Once the authors registered these attributes in the caDSR the BloodPressurePanel became semantically linked to existing caDSR common data elements with the same Object Class, demonstrating that the ISO/IEC 11179 data element model can be used to record and discover different expressions for semantically similar elements originating in different communities.

There are eight existing caDSR data elements associated with the same Object Class (Blood Pressure, C54706) as the BloodPressurePanel clinical element model. This type of linking (mapping) between semantically similar content initially

represented by different content models paves the way for data integration needed to query for and conduct comparative analysis between research data and clinical data, the cornerstone of precision medicine. Data that has been associated with the BloodPressurePanel clinical element model, and data in a different dataset that has been collected with individual content standards, could still be linked using the information registered in an ISO/IEC 11179 registry. This could enable clinicians to discover the existence of relevant clinical trial participant data, such as test results, in order to query for and retrieve only those patients' information that match a given patient's results, helping doctors retrieve the most relevant patient data to determine optimal treatment.

NCI identified gaps in ISO/IEC 11179 Data Element and Data Element Derivations model in order to represent the BloodPressurePanel, including the need to represent the cardinality of each data element in the composition, and had previously identified the need to support mapping the data element concept components in Figure 1 to one or more ontologies, supporting multiple semantic expressions for a single item. The latter provides researchers an entry point for exploring deeper meaning of collected data across different terminologies by using the links and mappings to ontology concepts to explore knowledge that has been represented in the ontologies through biocuration.

The results of comparing the content models used by NCI and Intermountain for standard representation of clinical content standards demonstrate that Intermountain common data elements can be represented in ISO/IEC 11179 by adding the semantics using the mechanisms expressed in Figure 1, and by transforming and mapping the essential attributes needed to describe the data, providing a way to translate between two content standards. While ISO/IEC 11179 data element derivations requires some additional attributes to be fully expressive of the clinical data model content standard, these results suggest that mapping the semantics of different expressions of clinical content standards described by different healthcare content standards may be supported and normalized by using the ISO/IEC 11179 standard.

The semantic details captured in the ISO/IEC 11179 model through the data element concept, and the characteristics of the physical data values in the Value Domain can increase opportunities for mapping and transforming data that has been represented differently but is semantically the same to support a variety of secondary use scenarios.

Notes

[1] National Cancer Institute. "caDSR Wiki." May 5, 2016. https://wiki.nci.nih.gov/display/caDSR/caDSR+Wiki.

[2] IHC Health Services. "The Clinical Element Model Browser User Guide." April 2012. www.clinicalelement.com/partials/user-guide.html.

[3] Coyle, Joey et al. "Clinical Element Datatypes." Mayo Clinic. November 14, 2008. http://informatics.mayo.edu/sharp/images/f/f4/CEDatatypes20081114.pdf.

[4] OpenCEM Browser. "Blood Pressure Panel." www.opencem.org/#/20150922/Intermountain/BloodPressurePanel.

[5] Intermountain Healthcare Clinical Element Models. "CEM BloodPressurePanel Registered in caDSR TEST Context." National Cancer Institute's CDE Browser. https://cdebrowser.nci.nih.gov/CDEBrowser/search?elementDetails=9&FirstTimer=0&PageId=ElementDetailsGroup&publicId=3421600&version=1.0.

Denise B. Warzel (warzeld@mail.nih.gov) and Dianne M. Reeves (reevesd@mail.nih.gov) are program managers in the Cancer Informatics Branch at the National Cancer Institute.

Article citation:
Warzel, Denise B; Reeves, Dianne M. "Cross-walking Health Content Standards Using the ISO/IEC 11179 Metadata Registries Standard" *Journal of AHIMA* 87, no.7 (July 2016): 46-49.