e-Health – For Continuity of Care
C. Lovis et al. (Eds.)

© 2014 European Federation for Medical Informatics and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License.
doi:10.3233/978-1-61499-432-9-273

An Approach for Utilizing Clinical Statements in HL7 RIM to Evaluate Eligibility Criteria

Richard BACHE[a, b,1], Christel DANIEL[c], Julie JAMES[d], Sajjad HUSSAIN[c], Mark McGILCHRIST[e] , Brendan DELANEY[b] and Adel TAWEEL[a,b]
[a] *Dept. of Informatics, School of Natural and Mathematical Sciences, King's College London, UK*
[b] *Dept. of Primary Care and Public Health Sciences, King's College London, UK*
[c] *Assistance Publique - Hôpitaux de Paris & INSERM, U1142, Paris, France*
[d] *Centre for Health Informatics and Multiprofessional Education, University College London, UK*
[e] *Health Informatics Centre, University of Dundee, Scotland*

Abstract. The HL7 RIM (Reference Information Model) is a commonly used standard for the exchange of clinical data and can be employed for integrating the patient care and clinical research domains. Yet it is not sufficiently well specified to ensure a canonical representation of structured clinical data when used for the automated evaluation of eligibility criteria from a clinical trial protocol. We present an approach to further constrain the RIM to create a common information model to hold clinical data. In order to demonstrate our approach, we identified 132 distinct data elements from 10 rich clinical trails. We then defined a taxonomy to (i) identify the types of data elements that would need to be stored and (ii) define the types of predicate that would be used to evaluate them. This informed the definition of a pattern used to represent the data, which was shown to be sufficient for storing and evaluating the clinical statements required by the trials.

Keywords. HL7 RIM, Electronic Healthcare Records, Clinical Trials, Eligibility Criteria

Introduction

A major barrier to repurposing clinical data from Electronic Healthcare Record (EHR) systems or Clinical Data Warehouses (CDWs) to support clinical trial design and execution is that the information systems in patient care and clinical research domains represent data differently. The collective international efforts of multiple standard development organizations (such as ISO, HL7, CDISC, etc.) currently focus on defining the various standards required to achieve computable semantic interoperability and may help to bridge the gap between clinical research and patient care by enabling a scalable semantic interoperability framework [1].

Various initiatives and standards have been documented [2-5] that aim to define a shared set of semantically unambiguous and context-neutral common *data elements*.

[1] Corresponding Author

This affords dynamic mappings between data structures and semantics of varying data sources and ensures that queries are defined using centrally managed common data elements. These features codify a shared Common Information Model (CIM) and can be mapped to the individual EHR or CDW models. The Electronic Healthcare Records for Clinical Research (EHR4CR) project has developed a platform for researchers to define eligibility criteria to identify eligible patients in clinical trials using controlled access to data held securely at clinical sites [6,7]. The set of eligibility criteria is formally defined as a query that is evaluated using purely structured patient data (i.e. not free text, images etc.) drawn from hospital EHR systems. Given that the HL7 RIM (Reference Information Model) was used to construct the repository of patient data, the question we address is how this data can, hailing from multiple and diverse EHR systems, be represented and queried canonically. We describe here how we defined a pattern for achieving this end. This was then used to implement a data warehouse as a relational. Eligibility criteria were evaluated automatically on data extracted from multiple warehouse instances from heterogeneous hospital EHR systems.

1. Methods

In the domain of patient care, several large-scale efforts have been underway for over a decade with the goal of specifying both the structure and semantics of patient clinical information in a manner that enables computable semantic interoperability between diverse systems. Two major contributions to the interoperability of clinical information currently dominate internationally: ISO EN 13606, with its implementation in the openEHR Foundation [8,9], and HL7 RIM, which became an ISO standard in 2003 and its implementations such as the Clinical Document Architecture (CDA), based on the Clinical Statement Message Pattern (CSPM) [10].

The HL7 RIM is a widely used clinical information exchange standard, which was adopted by the EHR4CR project for querying heterogeneous EHRs through a standard interface. Common data elements were defined in order to specify additional constraints on the high-level EHR4CR Constrained Information Model (CIM) and to represent the fine-grained clinical information required to evaluate eligibility criteria.

In order to devise a pattern for the storage and use of those clinical statements, which were required to evaluate eligibility criteria, we selected ten rich evaluative clinical trial protocols to cover the range of clinical areas covered by the project, specifically: cancer, cardiovascular disease (especially arrhythmias), and neurology. From analysis of these, we then devised a pattern that would support the data elements and corresponding criteria that were amenable to automatic evaluation.

The set of ten clinical trials was selected by EFPIA (European Federation of Pharmaceutical Industries and Associations) partners within EHR4CR. Each of these clinical trials was running in more than one of the participating university hospitals. From the 269 free text eligibility criteria from the ten clinical trials, 99 computable eligibility criteria that were relevant in the context of feasibility were manually pre-processed by the EFPIA partners. This yielded 132 data elements and their corresponding value sets, of which there are 32, representing the fine-grained clinical information included in eligibility criteria constructs.

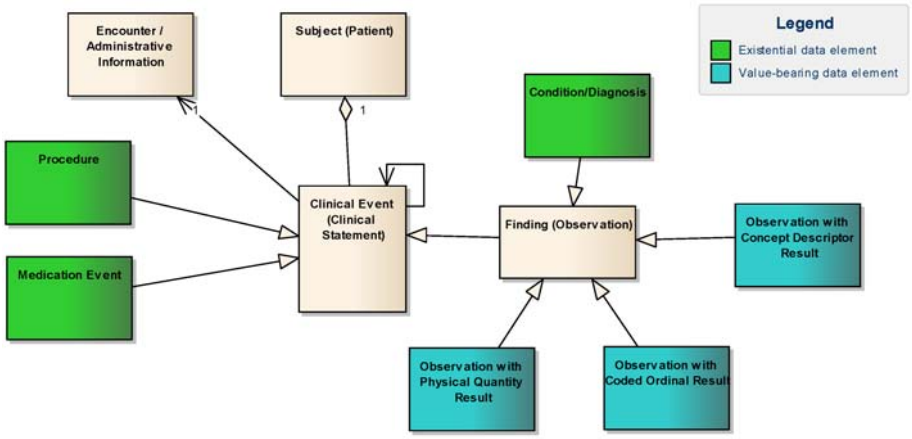

Figure 1: Object Model of Part of the EHR4CR Constrained Information Model, showing data element classification used in EHR4CR

2. Results

The 132 data elements correspond to different types of data elements, and these types were classified by the EHR4CR project: demographic statements (gender, birth date), diagnoses, clinical observations, lab test results, anatomic pathology observations, procedures and medications. Figure 1 shows a diagram of the relevant part of the CIM.

Any data element in the CIM may be further classified as being either *value bearing* or *existential*. A *value-bearing* element records the value of some measurement, score or categorisation. This value may be represented as either a number, usually with a unit of measurement, e.g. an HbA1c test, or a concept defined in some medical terminology e.g. male. An *existential* element denotes the recording of an event without any value being associated with it. *Existential* is used here in the mathematical rather than philosophical sense. In the EHR4CR CIM, procedures, medications, diagnoses and certain observations such as pregnancy and drug abuse are existential data elements; note that dosage of medications is currently ignored since it was not found in the eligibility criteria analysed. Diagnoses are considered existential because coding systems such as ICD-10 give a single code to represent a condition; there is no numeric or categorical qualifier. A key distinction between value-bearing and existential data elements is that the former form a series and the latter do not. A new observation of an HbA1c test adds to and in some senses updates a previous result of the same type. A diagnosis of a broken wrist cannot add to or update a diagnosis of lung cancer.

Date of birth and administrative gender, which are demographic statements, are represented as attributes of the HL7 *Living Subject* class. So, in keeping faithful to HL7, there are date of birth, administrative gender, deceased indicator and date of death fields. Mechanisms for recording procedure and medication statements in HL7, which are always existential, are well defined using the *Procedure* and *Substance Administration* classes respectively. For a procedure, a coded concept is used to specify the procedure described using the ISO21090 data type CD meaning a "concept descriptor". Similarly, a substance administration instance uses a coded concept to

describe the relevant medication consumed. The *Observation* class, however, caters for a broad range of clinical facts and encompasses diagnoses and clinical findings such as vital signs, laboratory tests and assessments. An observation instance is described by a pair: a 'code' field recording the nature of the observation and a 'value' field. According to HL7, the former is of type CD and the latter may be of any data type (structured or unstructured). To define our pattern effectively, a further restriction was needed.

HL7 already provides a mechanism for representing a diagnosis. Here, the 'code' attribute is set to a CD for the concept of 'diagnosis' and the value attribute is also of type CD and is used to hold the actual diagnosis in some medical terminology e.g. ICD-10. It was therefore natural to represent existential findings in the same fashion, so that the 'code' attribute is set to a CD meaning 'finding' and the actual finding, also a CD is held in the 'value'. For value-bearing observations, the 'code' field holds the CD for measurement or assessment type (e.g. fasting blood glucose) in some terminology and the 'value' field holds the result of the recorded value. Only three types of value were found in the trials analysed:

- Physical quantity (ISO21090 type PQ) where there is a numeric quantity and a unit of measurement. Vital signs and many laboratory tests fall into this category;
- Coded Ordinal (ISO21090 type CO) where there is an ordinal scale of values to which both a numeric value and a meaning (as a code) is also assigned. An example is the ECOG score of ambulatory status in cancer patients;
- Coded Value (ISO21090 type CD) where there is no numeric value or order over the range of values; the value is itself a concept e.g. HIV status, which can be 'positive' or 'negative'.

An EHR4CR Eligibility Criteria Model (ECM) was defined to support the querying for the full range of eligibility criteria deemed necessary by the project. A description of the ECM is contained elsewhere [11]. The part of the model relevant to the pattern proposed here is the form of predicates. Each criterion entered by a user contains exactly one predicate that evaluates to be *true* or *false* for a given patient. From a relevant clinical statement relating to that patient it is possible to determine whether the criterion includes or excludes him. Three types of predicate are supported:

- Existential – the predicate is deemed true if at least one instance of the data element has been recorded. This is used for existential elements.
- Numerical range – the predicate is deemed true if the numerical value for the data element lies within a specified range. This is used for physical quantities (PQ) and coded ordinals (CO).
- Finite set of non-numeric values – the predicate is deemed true if the value lies in this set and is used for coded ordinals (CO) and concept descriptors (CD).

Coded ordinals may be used in two types of predicates since either the (usually integer) numeric or coded value can be used.

The EHR4CR clinical data warehouse schema is based on the EHR4CR CIM. Tables exist for the major classes such as the Procedure, Substance Administration and Observation and their dimensions: Subject, Encounter and Participation. Some ISO 21090 data types, which would normally be given their own table, were de-normalised and placed directly in the Observation table, such as physical quantities (with their units) and coded ordinal values. With few exceptions, all coded concepts were placed within a single table for better management. The data warehouse is a core component

of a larger platform that allows the automatic evaluation of eligibility criteria. Further details are given elsewhere [7]. This implementation was found to be sufficient to store and evaluate all data elements found in any of the computable criteria from the 10 trials.

3. Discussion

Since the eligibility criteria evaluated by the EHR4CR platform require to be computed precisely and without human intervention, we record only clinical statements that may be expressed as structured data. Since any clinical statement will either be existential or value bearing (i.e. scalar, ordinal and categorical data), the pattern should support all future data elements since it is difficult to imagine how structured clinical data could be represented otherwise. Certainly, all the data elements identified in the ten clinical trial protocols analysed to date are representable in this form. The EHR4CR platform does also access legacy i2b2 [12] databases. However mapping of the CD and CO elements is problematic since this system was not designed with these data types in mind.

The proposed pattern provides a comprehensive means for expressing clinical statements used in both patient care and clinical research that uses structured data. The constrained pattern has been demonstrated as an effective way of recording clinical observations and affording reasoning about them. Potentially this could be extended beyond the evaluation of eligibility criteria, for example for storing data that is used for epidemiological research where data is subjected to statistical analysis.

References

[1] Hammond WE, Jaffe C, Kush RD. Healthcare Standards Development. The Value of Nurturing Collaboration. J. Ahima Am. Heal. Inf. Manag. Assoc. 2009 Jul; 80(7):44–50;

[2] Komatsoulis GA, Warzel DB, Hartel FW, Shanbhag K, Chilukuri R, Fragoso G, 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.

[3] Warzel DB, Andonaydis C, McCurry B, Chilukuri R, Ishmukhamedov S, Covitz P. Common Data Element (CDE) Management and Deployment in Clinical Trials. Amia Annu. Symp. Proc. 2003;1048.

[4] Jiang G, Solbrig HR, Iberson-Hurst D, Kush RD, Chute CG. A Collaborative Framework for Representation and Harmonization of Clinical Study Data Elements Using Semantic MediaWiki, Amia Summits Transl. Sci. Proc. 2010 Mar 1;2010:11–5.

[5] ISO/IEC 11179, Home Page for ISO/IEC 11179 Information Technology - Metadata registries [Internet]. 2014 Feb 9 [cited 2014 May 6]. Available from: http://metadata-standards.org/11179/

[6] EHR4CR: Electronic Health Records for Clinical Research [Homepage on the Internet] [cited 6 May 2014]. Available from: http://www.ehr4cr.eu/

[7] Chen Y, Bache R, Miles S, Cuggia M, Soto-Rey I, Taweel A, A SOA-based Platform for Automating Clinical Trial Feasibility Study, In Macedo M Editor, Proceedings of the IADIS International conference e-Health 2013, 2013: 84-94

[8] Kalra D, Singleton P, Milan J, Mackay J, Detmer D, Rector A, et al., Security and Confidentiality Approach for the Clinical E-Science Framework (CLEF). Methods Inf. Med. 2005;44(2):193–7.

[9] OpenEHR. OpenEHR [Homepage on the Internet]. 2014 [cited 6 May 2014] Available from: http://www.openehr.org/

[10] HL7. Health Level Seven International - Homepage [Homepage on the Internet]. 2014 [cited 29 Mar 2013]. Available from: http://www.hl7.org/

[11] Bache R, Taweel A Miles S, Delaney B , An Eligibility Criteria Query Language for Heterogeneous Data Warehouses, Methods Inf. Med. to appear.

[12] i2b2-Informatics for Integrating Biology & the Beside: [Homepage on the Internet] 2014 [cited 6 May 2014]. Available from http://www.i2b2.org/