Journal of Biomedical Informatics 62 (2016) 232–242

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Developing a data element repository to support EHR-driven phenotype algorithm authoring and execution

Guoqian Jiang [a,*], Richard C. Kiefer [a], Luke V. Rasmussen [b], Harold R. Solbrig [a], Huan Mo [c], Jennifer A. Pacheco [d], Jie Xu [e], Enid Montague [e,f], William K. Thompson [f], Joshua C. Denny [c,g], Christopher G. Chute [h], Jyotishman Pathak [i]

[a] Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, MN, USA
[b] Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
[c] Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, USA
[d] Center for Genetic Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
[e] Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
[f] School of Computing, DePaul University, Chicago, IL, USA
[g] Department of Medicine, Vanderbilt University, Nashville, TN, USA
[h] School of Medicine, Johns Hopkins University, Baltimore, MD, USA
[i] Division of Health Informatics, Weill Cornell Medical College, Cornell University, New York City, NY, USA

ARTICLE INFO

Article history:
Received 28 May 2016
Accepted 4 July 2016
Available online 5 July 2016

Keywords:
Quality Data Model (QDM)
HL7 Fast Healthcare Interoperability Resources (FHIR)
Metadata standards
Semantic Web technology
Phenotype algorithms

ABSTRACT

The Quality Data Model (QDM) is an information model developed by the National Quality Forum for representing electronic health record (EHR)-based electronic clinical quality measures (eCQMs). In conjunction with the HL7 Health Quality Measures Format (HQMF), QDM contains core elements that make it a promising model for representing EHR-driven phenotype algorithms for clinical research. However, the current QDM specification is available only as descriptive documents suitable for human readability and interpretation, but not for machine consumption. The objective of the present study is to develop and evaluate a data element repository (DER) for providing machine-readable QDM data element service APIs to support phenotype algorithm authoring and execution. We used the ISO/IEC 11179 metadata standard to capture the structure for each data element, and leverage Semantic Web technologies to facilitate semantic representation of these metadata. We observed there are a number of underspecified areas in the QDM, including the lack of model constraints and pre-defined value sets. We propose a harmonization with the models developed in HL7 Fast Healthcare Interoperability Resources (FHIR) and Clinical Information Modeling Initiatives (CIMI) to enhance the QDM specification and enable the extensibility and better coverage of the DER. We also compared the DER with the existing QDM implementation utilized within the Measure Authoring Tool (MAT) to demonstrate the scalability and extensibility of our DER-based approach.

Abbreviations: QDM, Quality Data Model; EHR, electronic health record; eCQMs, electronic clinical quality measures; HQMF, Health Quality Measures Format; DER, data element repository; FHIR, Fast Healthcare Interoperability Resources; CIMI, Clinical Information Modeling Initiatives; MAT, Measure Authoring Tool; eMERGE, Electronic Medical Records and Genomics; SHARP, Strategic Health Information Technology Advanced Research Project; HMORN, HMO Research Network; PCORnet, National Patient-Centered Clinical Research Network; PhEMA, phenotype execution and modeling architecture; NQF, National Quality Forum; MDR, Metadata Registry; W3C, World Wide Web Consortium; RDF, Resource Description Framework; OWL, Web Ontology Language; SPARQL, SPARQL Protocol and RDF Query Language; URI, Uniform Resource Identifier; HCLS, Semantic Web Health Care and Life Sciences; ITS, Implementable Technology Specifications; BD2K, Big Data to Knowledge; bioCADDIE, biomedical and healthCAre Data Discovery and Indexing; CEDAR, Center for Expanded Data Annotation and Retrieval; NCI, National Cancer Institute; caDSR, Cancer Data Standards Registry and Repository; API, application programming interface; MMS, meta-model schema; KNIME, Konstanz Information Miner; VSAC, Value Set Authority Center; OMG, Object Management Group; CTS2, Common Terminology Services 2; AMI, Acute Myocardial Infarction; IRA, inter-rater agreement.

* Corresponding author at: Mayo Clinic, Department of Health Sciences Research, 200 First Street, SW, Rochester, MN 55905, USA.
E-mail address: Jiang.Guoqian@mayo.edu (G. Jiang).

http://dx.doi.org/10.1016/j.jbi.2016.07.008
1532-0464/

1. Introduction

The creation of phenotype algorithms (i.e., structured selection criteria designed to produce research-quality phenotypes) and the execution of these algorithms against electronic health record (EHR) data to identify patient cohorts have become a common practice in a number of research communities, including the Electronic Medical Records and Genomics (eMERGE) Network [1–3], the Strategic Health Information Technology Advanced Research Project (SHARP) [4,5], the HMO Research Network (HMORN) [6,7] and the National Patient-Centered Clinical Research Network (PCORnet) [8]. However, there exists a limited toolbox enabling the creation of reusable and machine-executable phenotype algorithms which has hampered effective cross-institutional research collaborations [9].

To address this overarching challenge, we are actively developing a phenotype execution and modeling architecture (PhEMA) [10] (http://projectphema.org/) to enable: (1) unambiguous representation of phenotype algorithm logic and semantically rich patient data; (2) effective execution of the phenotype algorithm to generate reproducible and sharable results; and (3) a repository to share phenotypes and execution results for collaborative research. The Quality Data Model (QDM) has been chosen in the PhEMA project as an information model for representing phenotype algorithms. QDM was developed by the National Quality Forum (NQF) for representing EHR-based electronic clinical quality measures (eCQMs). In conjunction with the HL7 Health Quality Measures Format (HQMF), QDM contains core elements that make it a promising model for representing phenotype algorithms for clinical research [11,12]. However, currently the QDM specification [13] is available only as descriptive text documents, which require human interpretation and implementation for broader use and machine consumption. We believe that a standards-based, semantically annotated rendering of the QDM data elements is critical to support the development of phenotype algorithm authoring and execution applications.

The objective of this study is to develop and evaluate a data element repository (DER) that provides standard representations and machine-readable service APIs for data elements extracted from the QDM specification. The system architecture and tooling choices and their evaluations are described in the following sections.

2. Background

2.1. NQF QDM

The NQF QDM describes clinical concepts in a standardized format to enable electronic quality performance measurement in support of operationalizing the Meaningful Use Program in the United States. It consists of two modules: a data model module and a logic module [6]. The data model module is used to represent clinical entities (e.g. diagnoses, laboratory results) and includes the notions of category, datatype, attribute, and value set comprising concept codes from one or more terminologies. A QDM element encapsulates a certain category (e.g., Medication) with an associated datatype (e.g., "Medication, Administered"). Each datatype has a number of associated attributes (e.g., Dose). Fig. 1 shows the QDM element structure [13]. In QDM elements, value sets can be used to define possible codes for the QDM element's definition or the QDM elements' attributes. The logic module includes logical, comparison, temporal, and subset operators and functions. These may be combined to constrain combinations of data model entities (e.g. Diagnosis A AND (COUNT(Medication B) > 5)). As of July 2015, the latest release of QDM is version 4.1.2 [13]. Table 1

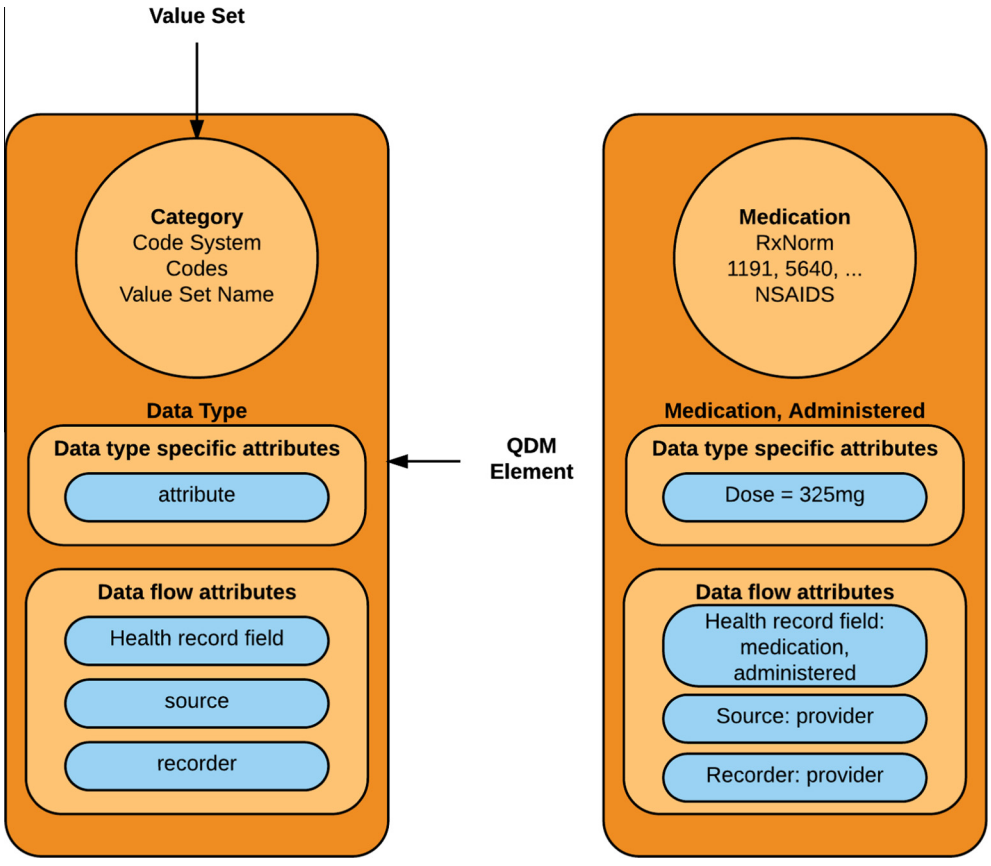

Fig. 1. Quality Data Model (QDM) element structure. (Reproduced using the source from the QDM element specification [13].)