Journal of Biomedical Informatics xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

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

A federated semantic metadata registry framework for enabling interoperability across clinical research and care domains

A. Anil Sinaci [a,b,*], Gokce B. Laleci Erturkmen [b]

[a] Department of Computer Engineering, Middle East Technical University, 06800 Ankara, Turkey
[b] SRDC Software Research & Development and Consultancy Ltd., ODTU Teknokent Silikon Blok No. 14, 06800 Ankara, Turkey

ARTICLE INFO

Article history:
Received 20 February 2013
Accepted 25 May 2013
Available online xxxx

Keywords:
Interoperability
Metadata Registry/Repository
ISO/IEC 11179
Semantic web
Linked Data
Common Data Elements

ABSTRACT

In order to enable secondary use of Electronic Health Records (EHRs) by bridging the interoperability gap between clinical care and research domains, in this paper, a unified methodology and the supporting framework is introduced which brings together the power of metadata registries (MDR) and semantic web technologies. We introduce a federated semantic metadata registry framework by extending the ISO/IEC 11179 standard, and enable integration of data element registries through Linked Open Data (LOD) principles where each Common Data Element (CDE) can be uniquely referenced, queried and processed to enable the syntactic and semantic interoperability. Each CDE and their components are maintained as LOD resources enabling semantic links with other CDEs, terminology systems and with implementation dependent content models; hence facilitating semantic search, much effective reuse and semantic interoperability across different application domains. There are several important efforts addressing the semantic interoperability in healthcare domain such as IHE DEX profile proposal, CDISC SHARE and CDISC2RDF. Our architecture complements these by providing a framework to interlink existing data element registries and repositories for multiplying their potential for semantic interoperability to a greater extent. Open source implementation of the federated semantic MDR framework presented in this paper is the core of the semantic interoperability layer of the SALUS project which enables the execution of the post marketing safety analysis studies on top of existing EHR systems.

1. Introduction

As the adoption of electronic health records (EHRs) increases, there has been a growing potential of exploiting this data both for enabling better care of patients by sharing the collected data across care organization, and also for enabling clinical research and quality assessment studies through secondary use of EHR. It is a well-accepted fact that one of the key challenges to be addressed to fulfill this great potential is enabling syntactic and semantic interoperability.

A major barrier to repurposing clinical data of EHRs for clinical research studies (clinical trial design, execution and observational studies) is that information systems in both domains – patient care and clinical research – use different data models and terminology systems. This means that data within each system is stand-alone and not interoperable. As stated by ISO [1], "One of the prerequisites for a correct and proper use and interpretation of data is that both users and owners of data have a common understanding of the meaning and descriptive characteristics of that data. To guarantee this shared view, a number of basic attributes have to be defined".

In line with this vision, many of the efforts which try to facilitate the exchange of EHRs for better care of the patient or to enable secondary use of EHRs for supporting clinical research and patient safety studies have already been developing Common Data Element (CDE) models. A few examples can be summarized as follows:

- The Health Information Technology Standards Panel (HITSP) has defined the C154: Data Dictionary Component [2] as a library of the HITSP defined data elements to facilitate the consistent use of these data elements across various HITSP selected standards. These data elements are served through PDF documents and spreadsheets. For example, HITSP C32 [3] which describes the HL7/ASTM Continuity of Care Document (CCD) [4] content for the purpose of health information exchange, marks the elements in CCD document with the corresponding HITSP C154 data elements to establish common understanding of the meaning of the CCD elements.

* Corresponding author at: SRDC Software Research & Development and Consultancy Ltd., ODTU Teknokent Silikon Blok No. 14, 06800 Ankara, Turkey. Fax: +90 312 210 1837.
E-mail addresses: anil@srdc.com.tr (A.A. Sinaci), gokce@srdc.com.tr (G.B. Laleci Erturkmen).

1532-0464/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jbi.2013.05.009

Abbreviations

Abbreviation	Description
BRIDG	*Biomedical Research Integrated Domain Group* – Develops a domain analysis model which aims to produce a shared view of the dynamic and static semantics for the domain of protocol-driven research and its associated regulatory artifacts
CDE	*Common Data Element* – The smallest meaningful data container in a given context
CDISC	*Clinical Data Interchange Standards Consortium* – A global, open, multidisciplinary, non-profit organization that is establishing standards to support the acquisition, exchange, submission and archive of clinical research data and metadata
CDISC2RDF	A development initiative in order to make the standards from CDISC available using semantic web standards and Linked Data principles
CDISC CDASH	*Clinical Data Acquisition Standards Harmonization* – Describes recommended basic standards for the collection of clinical trial data
CDISC SDTM	*Study Data Tabulation Model* – Provides a general framework for describing the organization of information collected during human and animal studies
CDISC SHARE	*CDISC Shared Health And Research Electronic Library* – A project under CDISC which aims to support computable semantic interoperability across multiple standards including, but not limited to those developed by CDISC
CRO	*Clinical/Contract Research Organization* – Provides R&D support to the pharmaceutical, biotechnology, and medical device industries in the form of research services outsourced on a contract basis
CS	*Classification Scheme* – In the meta-model of ISO/IEC 11179 standard, a Classification Scheme is a container of the classifiers for all kinds of administered items including the Common Data Elements
CSI	*Classification Scheme Item* – In the meta-model of ISO/IEC 11179 standard, a Classification Scheme Item acts as a classifier for the administered items (include the Common Data Elements) and each Classification Scheme Item belongs to a Classification Scheme
DEX	*Data Element Exchange* – A new interoperability profile which is under development by the IHE Quality, Research and Public Health (QRPH) domain
EDC	*Electronic Data Capture* – EDC systems are used for the collection of clinical data in electronic format for use mainly in human clinical trials and these systems are widely adopted pharmaceutical companies and clinical research organizations (CRO)
EHR	*Electronic Health Record*
FDA	*U.S. Food and Drug Administration*
FHIM	*Federal Health Information Model* – Managed by the Office of the National Coordinator for Health IT (ONC), FHIM seeks to develop a computationally independent model for the agencies of the Federal Health Architecture in the US
GE/IH CEM	*GE/Intermountain Healthcare Clinical Element Models*
GDSR	*Roche Global Data Standards Repository* – A metadata repository internally used in Roche in order to support clinical trials
HITSP	*Health Information Technology Standards Panel* – A partnership between the public and private sectors for the purpose of harmonizing and integrating standards that will meet clinical and business needs for sharing information among organizations and systems
HITSP C154	*HITSP Data Dictionary Component* – Defines the library of Data Elements that may be used by HITSP constructs in standards based exchanges. The Data Elements are organized into modules such as Medications, Advance Directives and Immunizations
HITSP C32	*HITSP Summary Documents Using HL7 Continuity of Care Document (CCD) Component* – Describes the document content summarizing a consumer's medical status for the purpose of information exchange. The content may include administrative (e.g., registration, demographics, insurance, etc.) and clinical (problem list, medication list, allergies, test results, etc.) information
HL7	*Health Level 7*
HL7 CDA	*HL7 Clinical Document Architecture* – A document markup standard that specifies the structure and semantics of clinical documents for the purpose of exchange between healthcare providers and patients
HL7/ASTM CCD	*HL7/ASTM Continuity of Care Document* – Defines a number of constraints on HL7 CDA standard to foster interoperability of clinical data to allow physicians to send electronic medical information to other providers without loss of meaning
HL7 RIM	*HL7 Reference Information Model* – The shared model between all HL7 domains and, as such, is the model from which all domains create their messages
I2B2	*Informatics for Integrating Biology and the Bedside* – Developing a scalable computational framework to address the bottleneck limiting the translation of genomic findings and hypotheses in model systems relevant to human health. Implements a central information model in order to manage the data interoperability
IHE	*Integrating the Healthcare Enterprise*
ISO	*International Organization for Standardization*
LOD	*Linked Open Data* – A set of methods and a philosophy for publishing data on the web so that it can be interlinked and reused across applications
MDR	*Metadata Registry/Repository* – A specialized database of metadata which describe data constructs
METeOR	*Metadata Online Registry* – Australia's repository for national metadata standards for health, housing and community services statistics and information
NEHTA	*National E-Health Transition Authority* – Established by the Australian, State and Territory governments, addresses a broad range of application domains under eHealth. NEHTA develops computable clinical content definitions known as Detailed Clinical Models
NIEM	*National Information Exchange Model* – A community-driven, government-wide, standards-based approach to exchanging information in the US
OC	*Object Class* – In the meta-model of ISO/IEC 11179 standard, an Object Class is the concept behind the Common Data Elements (CDE). A CDE is a composition of an Object Class (i.e. Patient) and a Property (i.e. Gender)
OMOP	*Observational Medical Outcomes Project* – A public-private partnership trying to identifying the most reliable methods for analyzing huge volumes of data drawn from heterogeneous sources. OMOP develops the Common Data Model in order to standardize the data format used in disparate data sources for the purposes of clinical research
OWL	*Web Ontology Language* – A set of knowledge representation languages maintained by World Wide Web Consortium (W3C) for authoring ontologies

Please cite this article in press as: Sinaci AA, Laleci Erturkmen GB. A federated semantic metadata registry framework for enabling interoperability across clinical research and care domains. J Biomed Inform (2013), http://dx.doi.org/10.1016/j.jbi.2013.05.009