View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Journal of Biomedical Informatics 46 (2013) 784–794

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 7 June 2013

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.

© 2013 Elsevier Inc. All rights reserved.

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

REST	*Representational State Transfer* – A software architecture style which was developed by W3C for designing distributed systems
RDF	*Resource Description Framework* – A set of specifications maintained by World Wide Web Consortium (W3C) for conceptual modeling of data (describing meta-models) through a variety of syntax notations and data serialization formats
SALUS	*Scalable, Standard based Interoperability Framework for Sustainable Proactive Post Market Safety Studies* – European Commission supported research project addressing the interoperability issues between clinical research and patient care domains for pharmacovigilance activities
S&I	*The Standards and Interoperability (S&I) Framework* – A collaborative community of participants from the public and private sectors who are focused on providing the tools, services and guidance to facilitate the functional exchange of health information
S&I CEDD	*The S&I Framework Clinical Element Data Dictionary* – A repository of clinically-relevant data elements and their corresponding definitions. The S&I Framework CEDD is intended to capture the unambiguously-defined data elements identified and reused by the initiatives of the S&I Framework
S&I ToC	*The S&I Framework Transitions of Care Initiative* – It aims to improve the exchange of core clinical information among providers, patients and other authorized entities electronically in support of meaningful use for improvement in the quality of care
UML	*Unified Modeling Language*
URI	*Uniform Resource Identifier* – A string of characters used to identify and refer to a resource on the Internet
XML	*Extensible Markup Language*

- The Federal Health Information Model (FHIM) [5] develops a common computationally independent model for EHRs.
- The Transitions of Care Initiative (ToC) [6] maintains the S&I Clinical Element Data Dictionary (CEDD) [7] as a repository of data elements to improve the electronic exchange of core clinical information among authorized entities in support of meaningful use and improvement in the quality of care. The Query Health [8] initiative extends this data dictionary, and establishes Query Health CEDD to enable an architecture for querying distributed EHRs in order to aggregate healthcare data for collecting quality measures and monitoring disease outbreaks.
- The Clinical Data Interchange Standards Consortium (CDISC) [9] provides common dataset definitions in (a) Study Data Tabulation Model (SDTM) [10] for enabling the submission of the result data sets of regulated clinical research studies to the FDA and in (b) Clinical Data Acquisition Standards Harmonization (CDASH) [11] for integrating SDTM data requirements into the Case Report Forms.
- The Biomedical Research Integrated Domain Group (BRIDG) [12] developed the Domain Analysis Model (DAM), which harmonizes CDISC data standards with the HL7 Reference Information Model (RIM) [13]. The BRIDG model unifies the concepts in the clinical care and research domains and creates a shared generic representation for each data element.
- Mini-Sentinel [14] is a pilot project to create an active surveillance system to monitor the safety of FDA-regulated medical products by accessing pre-existing electronic healthcare records. It proposes a Common Data Model (CDM) so that analytic applications can run on a uniform model. This model is maintained in a PDF document and partner EHR Systems are expected to translate the EHR data to this common model.

There are other similar efforts to define CDEs and accompanying data models like Observational Medical Outcomes Project (OMOP) [15] Common Data Model, GE/Intermountain Healthcare Clinical Element Models [16], National E-Health Transition Authority (NEHTA) Detailed Clinical Models [17] and I2B2 data model [18]. These are defined either as data dictionaries or through abstract data models which try to ensure interoperability within the boundaries of the associated initiatives. For instance, the query services, analysis methods or data exchange protocols envisioned by these initiatives can seamlessly run on top of the agreed common data element models which are set of core data elements. However, when it comes to achieving a broader range of interoperability, these efforts fall short: proliferation of common data element models does not help to solve the interoperability problem. Exchange of EHRs for the care of patients or secondary use of EHRs is not directly possible across these initiatives. For example, it is not directly possible to query an EHR which conforms to FHIM model through the query services provided by Query Health unless a mapping to Query Health CEDD is achieved first. When a researcher defines the data set to be collected for an observational study through CDISC SDTM variables, it does not become readily possible to extract these data sets from EHRs which can provide medical summaries of eligible patients through HITSP C32 documents. The use of different set of CDEs such as CDISC SDTM variables and HITSP Data Elements does not solve the problem of interoperability; yet it is not practical to expect all of these diverse initiatives and projects to stick to the same common model, and to use the same set of CDEs.

In this paper, we present a federated metadata registry framework where machine processable definitions of CDEs across domains can be shared, reused, and semantically interlinked with each other to address this semantic interoperability challenge.

1.1. Objective

In order to solve the interoperability problem within/between clinical research and care domains, several organizations are publishing common data element dictionaries and common models as described above. Although these efforts ensure interoperability within the selected domain for the selected use cases, interoperability across application domain boundaries is not automatically possible. These stem from the following facts:

- Common data element model development efforts are often disparate form each other. Although previous efforts are examined, most of the time, a common model is created from scratch.
- Most of the time, the specifications for these CDE sets and common models are in unstructured text files.
- Some of these efforts examine previous ones and reuse some CDEs proposed by others, and sometimes provide partial mappings to other CDE dictionaries. For example, S&I CEDD reuses elements from HITSP C154, NEHTA and FHIM; HITSP C32 provides mapping between HITSP C154 data elements to the elements of HL7 CCD. However, these are maintained in several different spreadsheets or in PDF documents. Hence, it is not possible to process or query this data.

We believe there is a need for a more coordinated approach that would allow machine processable definitions of CDEs defined by different efforts to be searched, allow CDEs to be reused and to be linked with each other and the mappings/links/relations between different CDEs in different domains can be queried to address semantic interoperability.

In this paper, we present a framework that facilitates all of these through the use of federated semantically enabled metadata registries (MDR) conforming to ISO 11179 standard [1] where CDEs maintained in different MDRs can be uniquely identified, queried and linked with each other through Linked Data principles.

2. Semantic metadata registry

The first challenge we would like to address is to maintain the definitions of CDEs in machine processable manner rather than keeping them in PDF documents or spread sheets so that it becomes possible to search and query them. For this we have selected to adopt ISO/IEC 11179 – Metadata Registries (MDR) standard.

2.1. ISO/IEC 11179

This standard addresses management of the semantics of data elements: it provides a standard metadata model for the representation of data elements and provides a methodology for the registration of the descriptions of data elements through this standard model to an MDR. The aim is to facilitate accurate common understanding of the data elements over time, space and applications. In ISO/IEC 11179 model, a data element is represented through its components, basically through a triple: Object Class, Property and Value Domain. Through the proposed model, unambiguous semantics of all these components is formally defined. In this way, management of the components of data elements and reuse of these components is also facilitated.

There are numerous adoptions of ISO/IEC 11179 registries [19–25] to address semantic interoperability, several of which are in healthcare domain. These central metadata registries are used to maintain a set of common data elements (CDEs) in the selected domain so that data sources and data requesters can agree on unambiguous semantics of the selected data elements in the chosen domain. To address interoperability at a larger scale, it should be possible to link and reuse CDE definitions across application domains which can be greatly enabled by a semantically interlinked federated MDR framework. Centralized metadata registries would not scale as it is not practical to manage CDEs within different application domains in a single registry; each set of common data elements can evolve in time, there should be a more flexible mechanism to manage and exploit linked set of CDEs across domains.

In the next section, we present the steps required to establish such a linked MDR framework. Afterwards, the design and promises of this framework are explained through an example scenario.

2.2. Proposed extensions to ISO 11179 Standard to achieve federated metadata management for semantic interoperability

A federated MDR framework should enable the following basic functionalities:

- Searching CDEs maintained by different MDRs.
- Retrieving standard specification of a selected CDE from an MDR.
- Re-using CDEs maintained in a different MDR by referencing to the respective CDE.

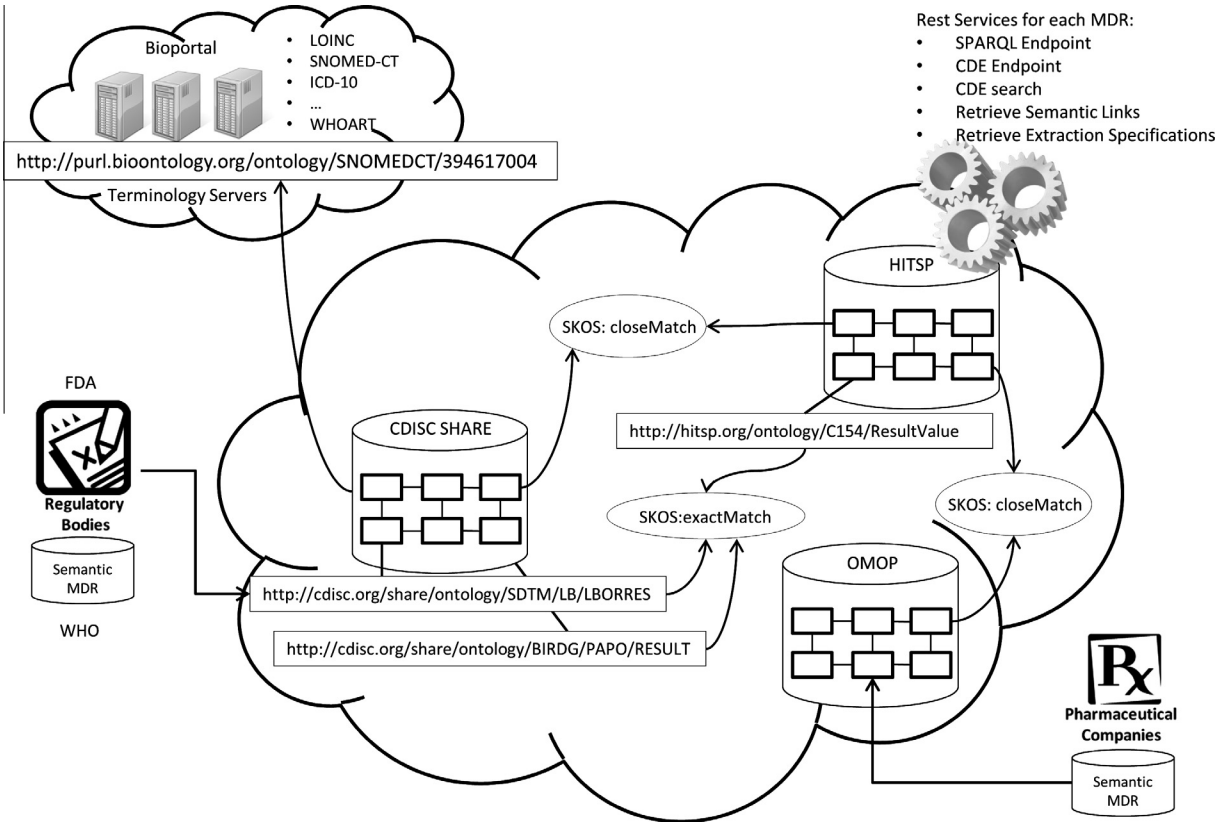

Fig. 1. Federated semantic MDR framework. Within the LOD cloud, each MDR maintains a set of CDEs together with the corresponding components and relations. The CDEs are linked to CDEs of other MDRs through KOSs and annotated with terminology systems.

To facilitate semantic interoperability more effectively across domains, a semantically linked federated MDR framework should support some additional functionality:

- It should be possible to link and semantically associate the CDEs across different MDRs in reference to well-accepted knowledge organization system (KOS) ontologies and terminology systems.
- It should be possible to easily query these semantic relationships within and across MDRs.

We have chosen to apply Linked Open Data (LOD) principles as the basis of this semantically linked federated MDR framework. Linked Data [26] is a recommended best practice for exposing, sharing and connecting pieces of data, information and knowledge on the semantic web using URIs and RDF [27]. It provides a natural way to expose the CDEs maintained in different MDRs openly in the LOD cloud and interrelate them with each other as depicted in Fig. 1.

2.2.1. MDRs in the LOD cloud

In order to integrate the MDRs within the LOD cloud, the following principles are adopted:

- Each CDE is uniquely identified by a URI.
- Each CDE is dereferenceable, that is, MDRs provide the necessary HTTP-REST services for looking up CDEs by using their unique URIs.
- Each MDR provides semantic RDF descriptions of CDEs, which are accessible through the HTTP Services provided. When a CDE is looked up through its URI, the RDF description of the CDE is returned where all context of the CDE is presented in RDF: each RDF property is interpreted as a hyperlink to other linked open MDR resources. This automatically opens up access to more data which is usually referred to as the "follow-your-nose principle". To enable this, we have created an OWL ontology from ISO/IEC 11179 meta-model [28]. We designed the ontology with OWL-Lite which is the lightest sublanguage of OWL with highest simplicity and lowest complexity. When a CDE is looked up, its RDF description in conformance to ISO/IEC 11179 meta-model is returned which includes links to other related MDR resources like the "object class", "property", "value domain", "enumerated value lists", "context" and "classification scheme items" that this CDE is related to. It should be noted that each of these resources are also maintained as uniquely identifiable LOD resources, hence, not only the CDEs but all objects within ISO/IEC meta-model are readily available through LOD principles: i.e. openly accessible with unique URIs with semantic descriptions attached.

2.2.2. Linking CDEs to terminology systems

In the Semantic MDR, it is possible to annotate CDEs with external terminology systems. Inline with the LOD approach, in our implementation, links of CDEs to terminology system codes are also referred through their unique URIs in the LOD cloud. BioPortal [29] already provides a wide range of terminology resources through the LOD principles where each terminology code is uniquely identified with a URI. In the Semantic MDR, for each terminology system a "Classification Scheme" (CS) resource is created as shown in Fig. 2. When an MDR resource is going to be related with a code from a terminology system, a "Classification Scheme Item" (CSI) resource is created under this CS resource and linked with the MDR resource. The unique URI of the terminology system code is recorded in the "value" property of CSI resource. In this way, all the CDEs across different MDRs annotated with the same terminology system code will be linked with the unique resource description created for the terminology system code, which directly provides a means to search and link CDEs across domains through LOD principles.

2.2.3. Linking CDEs to other CDEs

In our approach, it is possible to set other semantic links between the CDEs maintained in different MDRs as a part of semantic description of the CDE. For recording the semantic relationships between CDEs across MDRs, the "Classification Scheme" (CS) con-

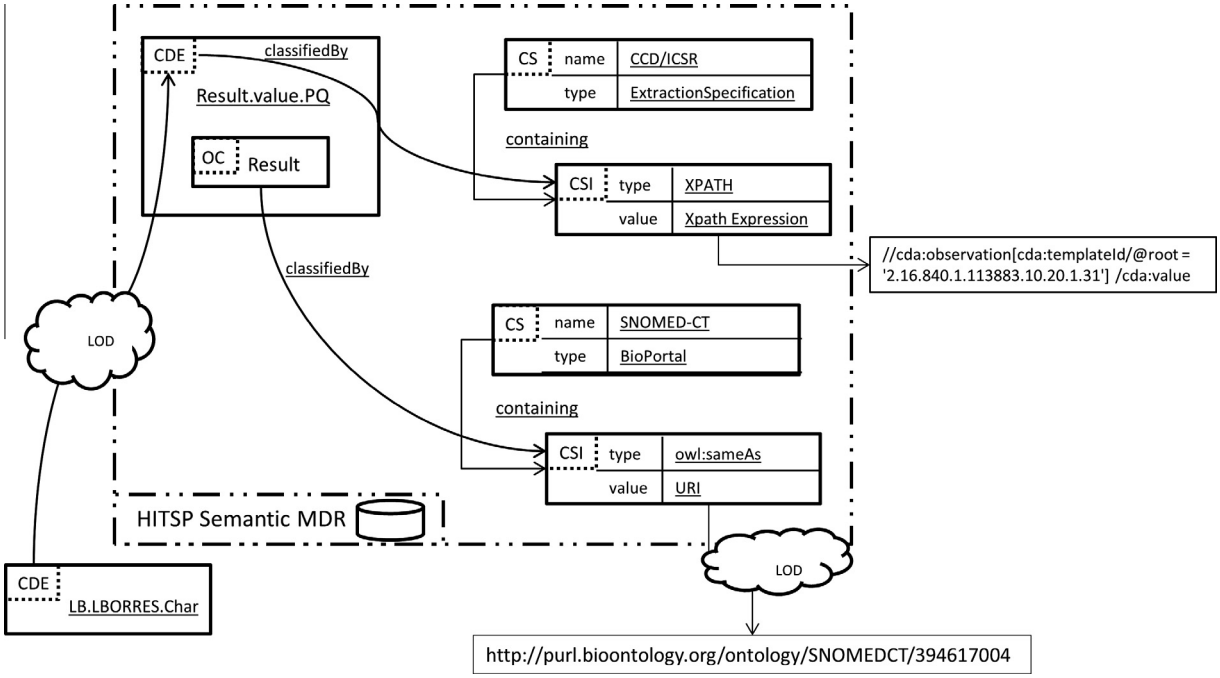

Fig. 2. Annotations and links of a CDE and its Object Class (OC) inside a Semantic MDR. The OC is annotated with a concept (term) from SNOMED-CT which is maintained under BioPortal through owl:sameAs property. The CDE has an "Extraction Specification" which is an XPATH expression pointing the exact place of the CDE in HL7 CCD models. These annotations and links are modeled through Classification Scheme and Classification Scheme Item elements of ISO/IEC 11179 model.

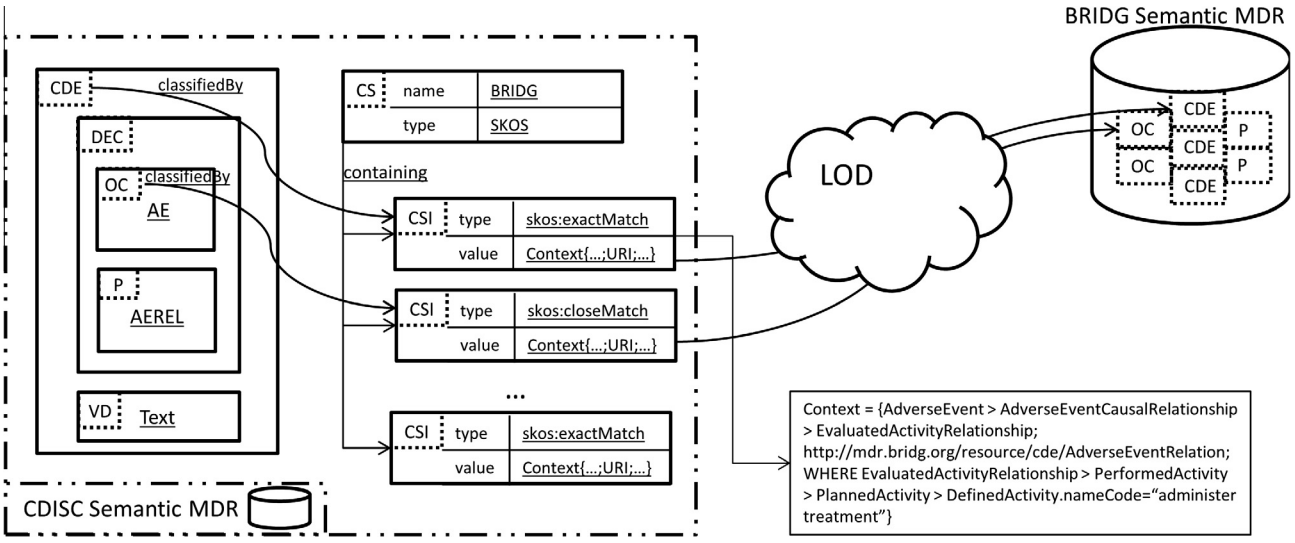

Fig. 3. The components of a CDE together with their classifications for the LOD links to other CDEs and components residing in a different Semantic MDR. The CDE, AE.AEREL.Text, has a skos:exactMatch relation with the CDE – AdverseEventRelation – in the Semantic MDR which holds the BRIDG CDEs. If needed, CDE mappings to other CDEs can be given through a Context in which some pre-conditions and rules can be specified.

structs available in ISO/IEC 11179 model are utilized. For each external MDR, a CS resource is created as presented in Fig. 3. Whenever a semantic relationship is to be created between CDEs, CSI resources are created and linked with the source CDE where the "type" property is set as the URI of the semantic relationship and "value" attribute is set as the unique URI of the target CDE. For identifying the semantic relationships, we are using upper KOS ontologies like SKOS [30]. In particular, SKOS "closeMatch" and "exactMatch" properties are exploited. By using such already existing semantic resource sets like SKOS, we ensure that CDEs are properly interlinked with each other via the other well-known LOD resources.

2.2.4. Linking CDEs to extraction specifications

One of the additional functionalities we would like to enable through a federated MDR framework is retrieving "extraction specifications" for a CDE defined in a selected domain, from a content model in a different domain. CDEs are often abstract data element definitions, which are later used to annotate the actual data elements in implementation dependent models which carry clinical content. These implementation dependent models are called "Content Models". For example, HITSP C154 data elements are used to annotate parts of CCD content models to indicate the unambiguous meaning of CCD elements. Maintaining the links between these abstract CDEs with the implementation dependent content models through an MDR architecture would facilitate retrieving machine processable extraction specifications that can be used to enable dynamic interoperability across different domains [31]. Through our federated MDR framework, it will become automatically possible to extract the SDTM annotated data sets from a medical summary conforming to HITPS C32 content model specifications (annotated with C154 data elements).

In our approach, each MDR can maintain several different content models as the implementations of the abstract CDEs served over the MDR. For the supported content models, MDR can maintain extraction specifications for each CDE available in the registry as a machine-processable specification of accessing the corresponding part of the content model. Content models can be XML documents, database schemes or RDF instances. Based on the type of the content model, different types of extraction specifications can be supported. An extraction specification is any script which can be executed on its associated content model. Current implementation of the Semantic MDR supports three types of extraction specifications:

1. XPath [32]: If the content model specification is based on XML Schema [33] and the data is serialized in XML, then it can be queried through XPath scripts. As shown in the example of Fig. 2, the information pointed by a CDE can be extracted from HL7 CCD based patient summaries when there are XPath scripts in the extraction specifications of the CDEs.
2. SPARQL [34]: If the content model specification is based on RDF and the data is residing in RDF graphs, then SPARQL scripts can be executed on the graph to retrieve the pointed information. An example is shown in Table 1 which is a part of the "ExtractionSpecification" for the CDE – "Patient.Allergy.Severity".
3. SQL [35]: If the content model is a relational model and data is residing in legacy relational databases, then SQL scripts can be executed to retrieve the associated information with the CDEs.

In the Semantic MDR, a "Classification Scheme" (CS) resource is created for each content model. The "type" of this CS is set as "ExtractionSpecification". For each extraction specification linked to the CDE resources, a "Classification Scheme Item" (CSI) resource is created where "type" property is set from the value set {XPath, SQL, SPARQL} and "value" property contains the extraction expression as presented in Fig. 2.

2.2.5. Federation through Linked Data principles

Within the generic meta-model as shown in Figs. 2 and 3, all external relations are indicated through "Classification Scheme

Table 1
SPARQL script to retrieve the severity information for the Allergy on a Patient. The target content model is SALUS Common Information Model [36] which can serve data through RDF graphs.

```
SELECT ?severity
  WHERE {
    ?pt a salus:Patient
    ?pt salus:allergy ?allergy
    ?allergy salus:severity ?severity
  }
```

Item"s (CSI) which are grouped under the "Classification Scheme"s. Therefore modeling a link to another CDE is similar to a link to a term (concept) in an external terminology system (i.e. SNOMET-CT) or to an extraction specification pointing to an implementation dependent model. That is, from the perspective of the Semantic MDR, these external resources are all metadata, but they are expected to follow the Linked Data principles and adopt well-known semantic schemes (like the SKOS), ontologies (like the ISO/IEC 11179 ontology) or the standardized serializations (like the IHE DEX profile). The beauty behind the federated semantic MDR framework is that, it does not enforce the compliance to all of the mentioned specifications. It can use and deduce as much knowledge as it can acquire from the linked resources. For example, in the current implementation we make use of the REST endpoints of BioPortal for terminology annotations. BioPortal provides the RDF serializations of the terms through well-known knowledge organization systems such as SKOS; as a result, we can automatically process a number of attributes such as labels and unique identifiers. Hence, if two different CDEs from two different Semantic MDRs are classified by the same term coming from BioPortal, a search through the federated architecture with the identifier of or a keyword belonging to that term would successfully find the two CDEs. These links to terminology systems can also be used for searching CDEs from the federated MDR framework, as a next step the extraction specifications of the discovered CDEs from the selected content models can be retrieved.

Apart from the best practices, it is a known fact that most of the existing EHR systems do not use standard terminologies or groupers. Instead, they use their local, proprietary coding schemes and vocabularies for data annotation. Making this existing legacy EHR data available for clinical research is a challenging task and there needs to be some additional effort in order to succeed the data interoperability with the existing systems. Our framework minimizes this effort because direct manipulation of the legacy data is not required. One needs to introduce the CDEs of the local coding system or content models used by the legacy systems to a local Semantic MDR and establish the appropriate mappings to the other standard based CDEs in the federated MDR framework. For example, in Fig. 2, it can be assumed that if a local CDE is used to annotate the lab results in an EHR system and that these local CDE is linked with the HITSP C154 CDE – "Result.value.PQ", then from the mappings of "Result.value.PQ" different extraction specifications can be reached. In this example the XPATH expression pointing to the exact location of the CDE in HL7/ASTM CCD model can be retrieved from the federated semantic MDR framework.

2.3. Design and implementation

The Semantic MDR provides the capabilities of a metadata registry and a metadata repository at the same time. While we utilize several services for the federated architecture of the semantic metadata registries, we also implement web based, easy-to-use graphical user interfaces for the management of CDEs including browsing, searching, editing and automatic importing in order to meet the requirements of a metadata repository. In this paper, we focus on the registry specific capabilities of the Semantic MDR for the federated and systematic communication among them.

2.3.1. Ontology of ISO/IEC 11179 Metamodel [28]

ISO/IEC 11179 provides a relational model for the structure of the MDRs through its entity-relationship diagrams. To be able to add semantic capabilities such as handling inter-links between CDEs and handling external links to other repositories, terminology systems and classification schemes, etc., the Semantic MDR has been built on top of a triple store which bases the knowledge on the ontological representation of the ISO/IEC 11179 metamodel. While building the ISO/IEC 11179 ontology, we adopt OWL [37] such that an OWL resource for each metamodel construct is created according to the mappings given in Table 2. In addition to the direct mappings of the ISO/IEC 11179 constructs, all relationships (i.e. class hierarchies, class-to-class relations) have been reflected to the ontology in order to be fully compliant with the metamodel.

Table 2
Mapping of ISO/IEC 11179 metamodel constructs to OWL constructs.

ISO/IEC 11179 metamodel construct	OWL construct
Class	owl:Class
Attribute	owl:DatatypeProperty
Composite attribute	owl:ObjectProperty
Class relationship	owl:ObjectProperty

2.3.2. MDR knowledge base

The Semantic MDR opens up several services in various layers which take root from its MDR Knowledge Base as shown in Fig. 4. The goal is to enable the federated communication through Linked Data principles. RESTful part of these services and its use in succeeding the interoperability between the clinical research and patient care domains are introduced in the following sections.

Since powerful semantic capabilities following Linked Data approach requires more sophisticated data management than the relational model, we base the data persistence on top of a Triple Store component as presented in Fig. 4. Apache Jena [38] has been adopted as the RDF framework which also has native support for OWL ontologies. Apache Jena has a built-in triple store backend, Jena TDB [39]. Our Triple Store components can selectively use either Jena TDB or Virtuoso [40] which is another high performance triple store implementation. They provide native SPARQL support, and have pros and cons over each other according to the usage context [41]. That's why the Semantic MDR provides a driver component which can automatically be integrated with both of the triple store implementations.

The Semantic MDR develops different types of importers in order to automatically populate the knowledge base with the CDEs of widely used content models. The current implementation can import OMOP CDM v4.0, SDTM v1.3, CDASH v1.1, HITSP C154 v1.0 and part of the BRIDG model using different serialization formats including SQL, RDF and comma-separated values.

2.3.3. RESTful interface

Once all of the links to external terminology systems, CDEs in external MDRs and links to content models become a part of the semantic description of a CDE, Semantic MDRs can open some simple REST services to ease the semantic query of CDEs across MDRs. A full list of proposed REST services is presented in Table 3. Through these services, it becomes possible to perform federated queries on MDRs to retrieve semantic descriptions of CDEs and process these for achieving semantic interoperability across domains. A sample application is presented in the next section.

2.4. Exploiting linked metadata registries for semantic interoperability

In our hypothetical scenario, which reflects one of the pilot application scenarios from SALUS project [36], a study data manager in a pharmaceutical company aims to design the data collection set for a new trial. The objective is to prepare a properly annotated study design document so that it can be automatically populated with patient data coming from HL7 CCD based content models through the information retrieval process of the Federated Query Service. The flow of the scenario is depicted in Fig. 5 and the steps are described in the following:

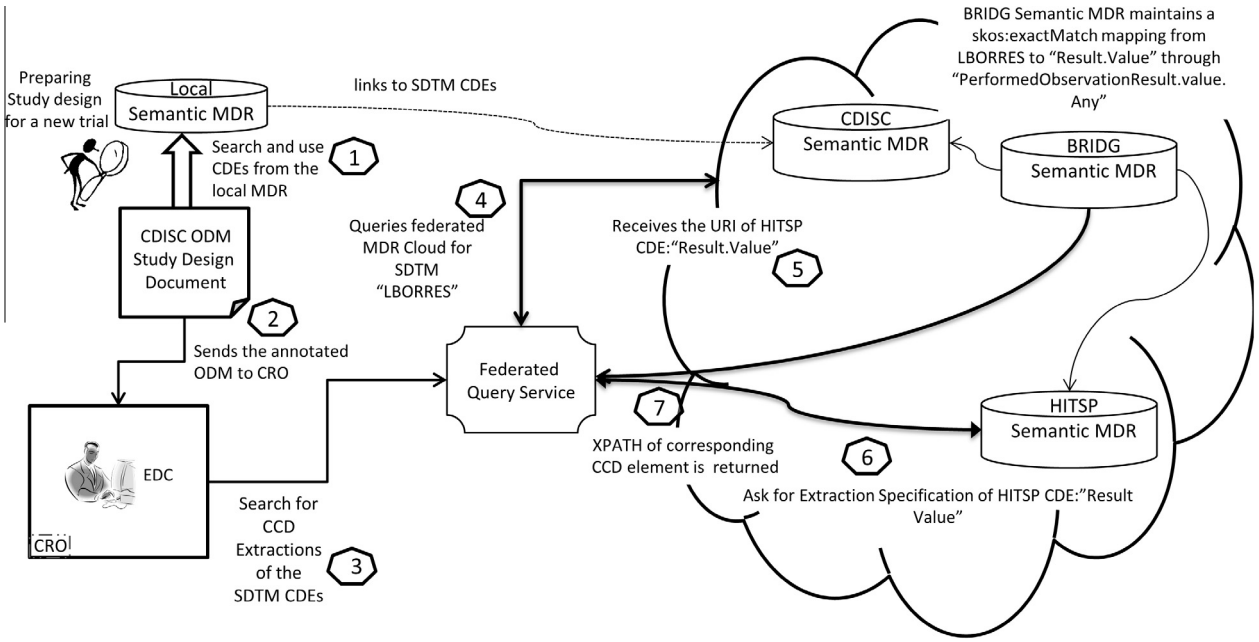

Fig. 4. High-level view of the architecture of the Semantic MDR Service Layer. At the bottom, there is a triple store serving as a backend for the MDR Knowledge Base. Above the triple store, there is a 3 layered API to perform semantic operations on this Triple Store. Semantic Data Manipulation API is a direct implementation of the ISO/IEC 11179 metamodel which reflects the operations to the underlying RDF graph. MDR API an abstraction layer which hides the complex details of the ISO/IEC 11179 metamodel and provides easy-to-use methods for the data manipulation.

Table 3
RESTful Services of a Semantic MDR.

Name	Description
SPARQL endpoint	Native SPARQL support. Functionalities of all other REST services can be provided by the SPARQL endpoint. RDF and SPARQL aware systems can build many semantic applications by consuming the SPARQL endpoints of Semantic MDRs
CDE endpoint	Given the URI, retrieve the full RDF description of the CDE from an MDR
CDE search	Parameterized search for CDEs through the allowed properties defined in ISO/IEC 11179 meta-model. For example, query CDEs by "Object Class" and/or "Classification Scheme Item". In this way it becomes possible to search CDEs annotated with a specific terminology system code
Semantic links	Retrieve all semantic links of the CDEs to the CDEs in other MDRs. Given the URI of a CDE (source), MDR returns the URIs of the other CDEs (target) interlinked with the source CDE, together with the URIs of the semantic relationships between these CDEs (e.g. skos:exactMatch). The requester can then directly lookup the full semantic description of the target CDEs, as unique URIs of CDEs will already direct the user to the MDR where it is maintained in
Extraction specification	Retrieve "extraction specifications" for a CDE in a selected domain for a supported content model. Input is the URI of the CDE and URI of the content model. Please note that the HL7 CCD content models provided by HITSP or IHE Patient Care Coordination Domain are already uniquely identifiable through OIDs

1. The study manager searches the local MDR of her organization to retrieve the data elements together with their descriptions for the selected set of variables in the data collection set. The local MDR returns a list of data element descriptions, including the unique URIs of the matching SDTM CDEs maintained by the MDR managed by CDISC.
2. The study manager prepares the study protocol as a CDISC ODM document annotated with SDTM CDEs and sends it to the Contract Research Organization (CRO).
3. The electronic data capture (EDC) system of the CRO automatically processes the study protocol, and tries to map the data items identified in the data collection set to the parts of HL7 CCD medical summary documents of study patients it collects from the participating care organizations.
4. EDC queries the federated MDR framework for extraction specifications of the SDTM CDEs from HL7 CCD format. If the CRO is using a Semantic MDR, then the federated search system is already embedded into the MDR. Otherwise, the federated query service end-point is invoked by the CRO's EDC. The service asks for extraction specifications of each SDTM CDE to the registered MDRs through the RESTful interfaces (Table 3).
5. None of the MDRs directly provide the extraction specification of the selected SDTM CDE (say LBORRES which stands for "results of a lab test") from HL7 CCD format. The query service asks for the Semantic Links of LBORRES to the registered MDRs. In our example scenario – a Semantic MDR maintaining BRIDG model data elements [12] – provides a mapping between the "LBORRES" CDE in CDISC SDTM domain to the "PerformedObservationResult.value.Any" CDE in BRIDG domain. It also maintains a mapping between "PerformedObservationResult.value.Any" CDE and the "Result.Value" CDE from HITSP domain. Hence, when the federated query service asks for Semantic Links of LBORRES, BRIDG MDR returns two URIs of
   a. "PerformedObservationResult.value.Any" from BRIDG
   b. "Result.Value" from HITSP
6. "Result.Value" CDE is served in a Semantic MDR hosted by HITSP which is linked with "PerformedObservationResult.value.Any" CDE through "SKOS:exactMatch" semantic relationship. The federated MDR search system now looks up to the HITSP MDR to retrieve extraction specification of "Result.Value" CDE in RDF format, and the extraction specification to HL7 CCD content model is available as "cda:observation[cda:templateId/

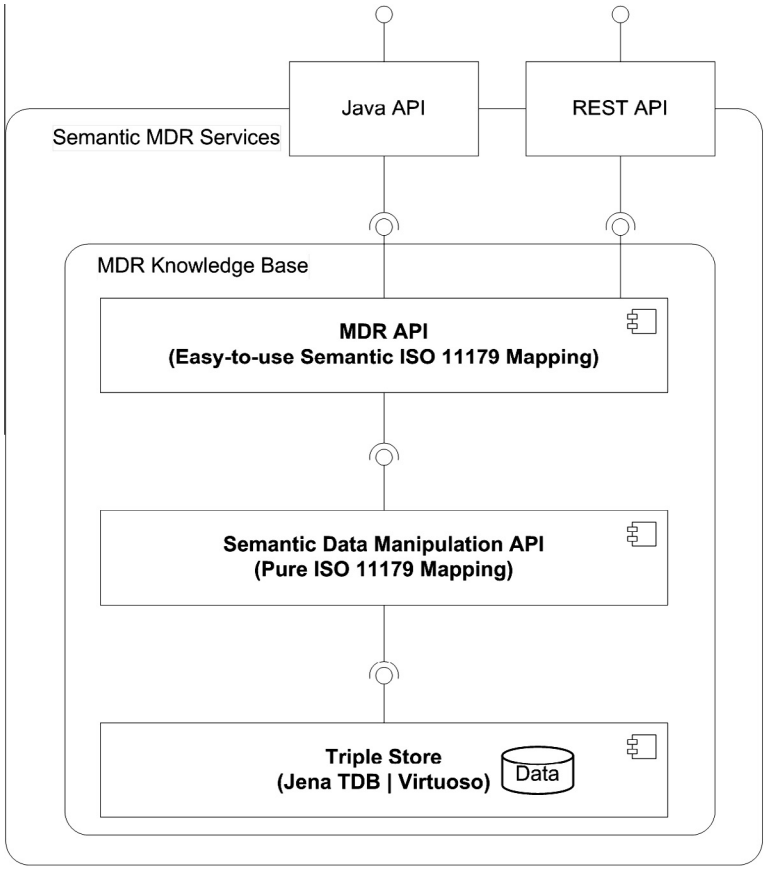

Fig. 5. Step-by-step representation of the scenario in which the federated semantic MDR framework is used for the interoperability of clinical research and clinical care domains. A properly annotated study design document can be automatically populated through the information retrieval process of HL7 CCD based content models with the help of the Federated Query Service. The service makes use of the simple REST interfaces of the Semantic MDRs which are introduced in Table 3.

@root = '2.16.840.1.113883.10.20.1.31']/cda:value" as an XPATH query.

7. In this way, the EDC is able to retrieve the required data elements in the data collection set from the HL7 CCD documents provided for each study visit by the participating organizations.

A similar flow can be achieved through retrieving the RDF descriptions of CDEs by calling the CDE endpoints, and by processing these RDF descriptions where semantic links and links to extraction specifications are already available.

As depicted in the example scenario, through the proposed federated MDR framework, it is possible to facilitate interoperability across clinical research and care domains although different standards and different CDEs are in use. Similar to this scenario, another use case can be automatic population of case safety reports to notify adverse drug events through ICSR documents [42].

3. Related work

There are several efforts trying to address the interoperability between the clinical research and patient care domains. One major approach to the problem of semantic interoperability is to build a common data model where the interoperating systems are required to interact through this well-defined data model. This can be classified as a top-down approach where a top-level knowledge model agreement is forced for the underlying data models of the interoperating parties for successful data exchange. The research behind OMOP CDM [43], FDA Mini-Sentinel [44], I2B2 [45], STRIDE [46] and EU-ADR [47] are among some of the efforts that adopt this top-down strategy to reuse existing EHR data for the clinical research purposes. In addition to these projects, Laleci et al. [48] builds a semantic common information model to exchange data between clinical research and clinical care systems. Weber et al. [49] presents a prototype of a federated query tool for clinical data repositories through a common information model.

Another major approach is to identify the CDEs of the content models of the interoperating systems and provide direct mappings between them. Apart from the strict relations within a content model, this approach attaches more importance to the elicitation of the data elements. Fadly et al. [50] presents mapping algorithms to identify semantic coherence between clinical care and clinical research data elements in order to pre-populate electronic Case Report Forms. Jiang et al. [51] presents a prototype implementation for CDISC SHARE using already available semantic tools where they try to provide an environment for CDE management. Kunz et al. [52] utilize a repository of CDEs to help developers reuse appropriate CDEs to enable interoperability of their systems. Pathak et al. [53] analyses the effects of adoption of CDEs in large-scale epidemiological and genome-wide studies on cross-study analysis.

There are also several standardization efforts addressing the problem of semantic interoperability in question. The IHE Drug Safety Content (DSC) [54] and Clinical Research Data Capture (CRD) [55] profiles are two efforts to address pre-filling of safety reports and case report forms (CRF) by retrieving the data from medical summaries expressed in HL7 CCD format. However, both of these profiles propose static XSLT mappings between a predefined medical summary template in CCD and a generic CRF form. This approach is not flexible and extensible; these XSLT mappings are only valid for the given pre-population data formats; once

these pre-population data templates are modified due to emerging requirements, new mappings are needed [56,57]. The new IHE Data Exchange (DEX) [31] profile proposal in IHE QRPH domain addresses the shortcomings of IHE CRD and DSC profiles. A metadata registry is envisioned to maintain the research and healthcare CDEs, and the exact correspondences between them. In this paper, we extend this idea by providing a semantically linked federated MDR framework to show how the DEX idea can scale in the presence of disparate CDE definition efforts by different organizations.

Our proposal tries to unify the top-down and bottom-up approaches with an analogy to the unification of old, well-established methods with newly emerging, latest technologies. Tao et al. [58] already shows the value in using OWL to represent relational meta-models, including ISO/IEC 11179. To the best of our knowledge, our work is the first attempt to apply a comprehensive set of semantic web technologies with the commonly adopted MDR standard – ISO/IEC 11179 – through the Linked Data principles.

4. Discussion

We have developed the semantically linked federated metadata registries presented in this paper within the scope of SALUS project to address the interoperability gap between clinical research and clinical care domains. The example scenario presented in Section 2.4 in fact reflects the requirements of one of the selected scenarios in our project by one of the pharmaceutical companies, Roche in particular.

There is an important initiative in addressing the interoperability between care and research domains through maintaining common data elements: CDISC SHARE [57] aims to harmonize the CDEs used in clinical research and care domains. It is envisioned that CDE definitions will be built upon BRIDG DAM where they are annotated with CDISC data sets like CDASH and SDTM, and other CDISC terminologies. We believe such a global CDE registry maintained by a standardization body, CDISC, can be in the core of our federated MDR framework. Other local MDRs of research initiatives or pharmaceutical organizations can semantically link their CDEs with the existing CDEs in SHARE as explained in the motivating scenario presented. To facilitate interoperability in larger scale, it would be very good if the CDEs provided by SHARE can be directly or indirectly interlinked with the CDEs proposed by HITSP, FHIM and ToC CIM through the federated MDR framework we propose.

In addition to this, the clinical research community has already started discussing about the benefits of using semantic web technologies, in particular Linked Data approach [23,59,60] for managing clinical research data. The Roche Global Data Standards Repository (GDSR) [23] MDR has been implemented based on semantic web technologies as a triple store. Another important effort in this direction is the CDISC2RDF initiative [61] which aims to make the standards from CDISC like CDASH and SDTM available through RDF descriptions. There is also an effort to represent HL7 Model Interchange Format in OWL. Our architecture will enable linking these semantic data set definitions with the CDEs provided by other MDRs through Linked Data principles.

Although our main focus is interoperating clinical care and research domains, we believe that the semantically linked federated metadata registry approach has the potential to be used to address interoperability challenges in different contexts: one such scenario can be addressing interoperability challenges of sharing EHRs across different systems. Our approach can be utilized to address this problem with different complexities: If we assume that the EHR systems expose EHR data with some standard based but different content models, like HL7 CCD based templates and ISO/EN 13606 EHRExtracts, then the Semantic MDR can be used to maintain the CDEs in these different formalisms (like HITSP C154 data elements) and extraction specifications to these different content models. However it is a reality that not all of the legacy EHR systems implement these standard based interfaces, which makes the problem more challenging. The Semantic MDR approach can still be used in this context, yet this time, we shall assume that local CDEs should be created and maintained by EHR vendors. It is a well-known fact that even the two different deployments of an EHR system in two different settings will result in different underlying database schemas depending on the needs of the local institutions. The Semantic MDR architecture can first be used to maintain the "extraction specifications" of the local CDEs maintained by the EHR Vendor to these different database schemas as SQL scripts. As a next step, these local CDEs can be linked with the CDEs maintained by standardization bodies, like HITSP C154 data elements. In this way, with a multi-layered MDR architecture, interoperability across systems can be addressed in a scalable manner.

5. Results and conclusion

In this paper, we introduce a federated semantic metadata registry framework where machine processable definitions of CDEs across domains can be shared, reused, and semantically interlinked with each other through Linked Data principles. We demonstrate how such a framework can be utilized to address this semantic interoperability challenge across application domains through an example scenario.

There are already several adoptions of MDRs [19–25]. Some of them are maintained in a single organization like Roche GDSR [23], some are at project level in a specific domain like caDSR [19], some are at national level for eHealth domain like METeOR MDR [21] in Australia, some are at national level but not restricted to a specific domain like NIEM US Federal metadata registry [62], and some are at global level like CDISC SHARE [57] addressing interoperability across domains but covering a selected set of data sets. On top of these, there are efforts to define core set of data elements through spreadsheets, PDF documents or UML models like HITSP C154 [2] and FHIM [5] respectively. Our approach is not a disruptive effort; instead it builds upon and complements all of these as follows: through a semantically linked federated MDR framework, we believe these efforts can be linked with each other, multiplying their potential for semantic interoperability to a greater extent.

Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under Grant Agreement No. ICT-287800, SALUS Project (Scalable, Standard based Interoperability Framework for Sustainable Proactive Post Market Safety Studies).

References

[1] ISO/IEC. ISO/IEC 11179: information technology – Metadata Registries (MDR) Parts 1–6, 2nd ed.
[2] HITSP. C 154: HITSP data dictionary. <http://www.hitsp.org/ConstructSet_Details.aspx?&PrefixAlpha=4&PrefixNumeric=154> [accessed 15.04.13].
[3] HITSP. C 32: HITSP summary documents using HL7 Continuity of Care Document (CCD) component. <http://www.hitsp.org/ConstructSet_Details.aspx?&PrefixAlpha=4&PrefixNumeric=32> [accessed 15.04.13].
[4] HL7/ASTM. Continuity of Care Document (CCD). <http://www.hl7.org/documentcenter/public_temp_DC68F8CB-1C23-BA17-0CB6B9727B87B502/pressreleases/20070212.pdf> [accessed 15.04.13].
[5] FHIMS. Federal Health Information Model. <http://www.fhims.org/content/420A62FD03B6_root.html> [accessed 15.04.13].
[6] S&I Framework. Transitions of Care Initiative (ToC). <http://wiki.siframework.org/Transitions+of+Care+(ToC)+Initiative> [accessed 15.04.13].

[7] S&I Framework. S&I Clinical Element Data Dictionary (CEDD) WG. <http://wiki.siframework.org/S%26I+Clinical+Element+Data+Dictionary+WG> [accessed 15.04.13].
[8] S&I Framework. Query health. <http://wiki.siframework.org/Query+Health> [accessed 15.04.13].
[9] Clinical Data Interchange Standards Consortium (CDISC). <http://www.cdisc.org/> [accessed 15.04.13].
[10] CDISC. Study Data Tabulation Model (SDTM). <http://www.cdisc.org/sdtm> [accessed 15.04.13].
[11] CDISC. Clinical Data Acquisition Standards Harmonization (CDASH). <http://www.cdisc.org/cdash> [accessed 15.04.13].
[12] BRIDG. The Biomedical Research Integrated Domain Group (BRIDG) model. <http://www.bridgmodel.org/> [accessed 15.04.13].
[13] HL7. HL7 Reference Information Model (RIM). <http://www.hl7.org/implement/standards/rim.cfm> [accessed 15.04.13].
[14] FDA. Sentinel initiative – mini-sentinel. <http://mini-sentinel.org/> [accessed 15.04.13].
[15] OMOP. Observational Medical Outcomes Project (OMOP). <http://omop.fnih.org/> [accessed 15.04.13].
[16] GE/Intermountain Healthcare. Clinical Element Models (CEM). <http://www.clinicalelement.com/> [accessed 15.04.13].
[17] NEHTA. Detailed clinical models. <http://www.nehta.gov.au/connecting-australia/terminology-and-information/detailed-clinical-models> [accessed 15.04.13].
[18] I2B2. I2B2 star schema. <https://www.i2b2.org/events/slides/Workshop1.pdf> [accessed 15.04.13].
[19] Komatsoulis GA et al. CaCORE version 3: implementation of a model driven, service-oriented architecture for semantic interoperability. J Biomed Inform 2008;41:106–23.
[20] UK CancerGrid. <http://www.cancergrid.org/> [accessed 15.04.13].
[21] AIHW. MeteOR: Metadata Online Registry. <http://meteor.aihw.gov.au/> [accessed 15.04.13].
[22] AHRQ. The United States health information knowledgebase. <http://ushik.ahrq.gov/> [accessed 15.04.13].
[23] Forsberg K, Malfait F. Semantic models for CDISC based standard and metadata management. Belgium: CDISC Interchange Europe Brussels; 2012. <http://kerfors.blogspot.com/2012/05/semantic-models-for-cdisc-based.html> [accessed 15.04.13].
[24] Stausberg J, et al. A national metadata repository for empirical research in Germany. In: 15th International open forum on metadata registries. Berlin (Germany); 2012 May.
[25] Rimatzki B, Haas P. Implementation of an ISO/IEC 11179 based metadata registry to foster interoperability of health Telematics applications. In: 15th International open forum on metadata registries. Berlin (Germany); 2012 May.
[26] Linked Data. <http://linkeddata.org/> [accessed 15.04.13].
[27] W3C. Resource Description Framework (RDF). <http://www.w3.org/RDF/> [accessed 15.04.13].
[28] SALUS Resources. OWL ontology for ISO/IEC 11179-3 model. <https://github.com/sinaci/semanticMDR/blob/master/core/src/main/resources/model/salus.mdr.owl> [accessed 15.04.13].
[29] Whetzel PL et al. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. Nucleic Acids Res 2011;39(Web Server issue):W541–5.
[30] W3C. SKOS: Simple knowledge organization system. <http://www.w3.org/2004/02/skos/> [accessed 15.04.13].
[31] IHE. Data Exchange (DEX) profile proposal. <ftp://ftp.ihe.net/Quality/2013_2014_YR_7/QRPH_Technical/2012-11-23_DEX-DetailedProposal_lb.docx> [accessed 15.04.13].
[32] W3C. XML Path Language (Xpath). <http://www.w3.org/TR/xpath/> [accessed 15.04.13].
[33] W3C. XML schema. <http://www.w3.org/XML/Schema> [accessed 15.04.13].
[34] W3C. SPARQL query language for RDF. <http://www.w3.org/TR/rdf-sparql-query/> [accessed 15.04.13].
[35] ISO. ISO/IEC 9075-1:2011 – Structured Query Language (SQL). <http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=53681> [accessed 15.04.13].
[36] SALUS: scalable, standard based interoperability framework for sustainable proactive post market safety studies. <http://www.salusproject.eu> [accessed 15.04.13].
[37] W3C. Web Ontology Language (OWL). <http://www.w3.org/TR/owl-features/> [accessed 15.04.13].
[38] Apache. Jena. <http://jena.apache.org/> [accessed 15.04.13].
[39] Apache. Jena TDB. <http://jena.apache.org/documentation/tdb/> [accessed 15.04.13].
[40] Open Link Software. Virtuoso universal server. <http://virtuoso.openlinksw.com/> [accessed 15.04.13].
[41] Bizer C, Schultz A. The Berlin SPARQL benchmark. Int J Semant Web Inf Syst 2009;5(2):24.
[42] FDA. ICSR: Individual Case Safety Reports. <http://www.fda.gov/ForIndustry/DataStandards/IndividualCaseSafetyReports/default.htm> [accessed 15.04.13].
[43] Reisinger SJ et al. Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases. J Am Med Inform Assoc 2010;17:652–62.
[44] Curtis LH et al. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. Pharmacoepidem Drug Saf 2012;21:23–31.
[45] Kohane IS, Churchill SE, Murphy SN. A translational engine at the national scale: informatics for integrating biology and the bedside. J Am Med Inform Assoc 2012;19:181–5.
[46] Lowe HJ et al. STRIDE – an integrated standards-based translational research informatics platform. AMIA Annu Symp Proc 2009;2009:391–5.
[47] Avillach P et al. Design and validation of an automated method to detect known adverse drug reactions in MEDLINE: a contribution from the EU-ADR project. J Am Med Inform Assoc 2013;20(3):446–52.
[48] Laleci GB, Yuksel M, Dogac A. Providing semantic interoperability between clinical care and clinical research domains. IEEE J Biomed Health Inform 2013;17(2):356–69.
[49] Weber GM et al. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. J Am Med Inform Assoc 2009;16:624–30.
[50] Fadly AE et al. Integrating clinical research with the Healthcare Enterprise: from the RE-USE project to the EHR4CR platform. J Biomed Inform 2011;44(1):S94–S102.
[51] Jiang G et al. A collaborative framework for representation and harmonization of clinical study data elements using semantic MediaWiki. AMIA Summits Transl Sci Proc 2010;2010:11–5.
[52] Kunz I, Lin MC, Frey L. Metadata mapping and reuse in caBIG. BMC Bioinform 2009;10(2):S4.
[53] Pathak J et al. Mapping clinical phenotype data elements to standardized metadata repositories and controlled terminologies: the eMERGE Network experience. J Am Med Inform Assoc 2011;18:376–86.
[54] IHE. Drug Safety Content profile (DSC). <http://www.ihe.net/Technical_Framework/upload/IHE_QRPH_TF_Supplement_Drug_Safety_Content_DSC_TI_2009-08-10.pdf> [accessed 15.04.13].
[55] IHE. Clinical Research Data Capture profile (CRD). <http://wiki.ihe.net/index.php?title=Clinical_Research_Data_Capture_-_(CRD)> [accessed 15.04.13].
[56] Bain L, Evans J, Lastic PY. Mapping EHR data to a research case report form: how a metadata repository, CDISC's SHARE, can improve the IHE profile Clinical Research Data (CRD). In: 15th International open forum on metadata registries, Berlin, Germany, 2012 May.
[57] CDISC. SHARE. <http://www.cdisc.org/cdisc-share> [accessed 15.04.13].
[58] Tao C et al. Towards semantic-web based representation and harmonization of standard meta-data models for clinical studies. AMIA Summits Transl Sci Proc 2011;2011:59–63.
[59] Anderson B, Forsberg K. Linked Data, an opportunity to mitigate complexity in pharmaceutical research and development. In: Proceedings of the 1st international workshop on linked web data management, NY, USA, 2011.
[60] Forsberg K. Linking clinical data standards, CDISC Interchange Europe Brussels, Belgium, 2012. <http://kerfors.blogspot.com/2011/05/linking-clinical-data-standards.html> [accessed 15.04.13].
[61] CDISC2RDF. <http://cdisc2rdf.com/> [accessed 15.04.13].
[62] NIEM. National Information Exchange Model. <https://www.niem.gov/> [accessed 15.04.13].