Standard-based EHR-enabled applications for clinical research and patient safety: CDISC – IHE QRPH – EHR4CR & SALUS collaboration

Christel Daniel, MD, PhD, Anil Sinaci, MSc; David Ouagne, PhD; Eric Sadou; Gunnar Declerck, PhD; Dipak Kalra MD, PhD; Jean Charlet, PhD; Kerstin Forsberg; Landen Bain; Charlie Mead; Sajjad Hussain, PhD; Gokce B. Laleci Erturkmen, PhD

1INSERM, U1142, LIMICS, F-75006, Paris, France; Sorbonne Universités, UPMC Univ Paris 06, UMR_S 1142, LIMICS, F-75006, Paris, France; 2CCS SI Patient, AP-HP, Paris, France; 3Software Research, Development and Consultancy, Ankara, Turkey; 4University College London, UK; 5AstraZeneca, Sweden; 6CDISC; 7W3C

Abstract
Integration profiles collaboratively developed by CDISC and IHE for integrating data from Electronic Health Records (EHRs) with clinical research and pharmacovigilance are limited to resolving lexical/syntactic data integration issues and do not address semantic barriers. This paper describes the collaboration between two European projects – EHR4CR and SALUS – in implementing ISO/IEC 11179-based metadata registries (MDRs) and semantically integrated cross-platform data access. A common “semantic MDR” provides a framework for bi-directional/cross-MDR mapping and federated queries are enabled using the newly-defined IHE Data Exchange (DEX) profile. In the pilot implementation, mappings for 178 EHR4CR and 199 SALUS metadata elements were persisted in the semantic MDR. The DEX profile was then used to access semantically equivalent data elements in SALUS or EHR4CR participating EHR systems. ISO/IEC 11179-based MDRs and DEX integration profile address the goal of developing pan-EU computable semantic integration of data from clinical care, clinical research, and patient safety platforms.

Keywords: Electronic Health Records, Biomedical Research, Adverse Drug Reaction Reporting Systems, Pharmacovigilance, Terminology as Topic

1 Introduction & background
Electronic Health Records (EHRs) contain a large variety of patient-centric data. A number of investigators have noted that the ability to integrate data from EHRs with that from other domains – e.g. clinical research and post-market patient safety – could provide significant value to both domain-specific and population-centric research [1,2]. Specific topics of interest include providing trial planners with a better understanding of the available cohorts [3,4,5,6] and targeted patient recruitment [7]. Others have addressed the efficiencies of “single-source data entry” at the point of clinical care [8,9]. Finally, ongoing reporting of post-market adverse drug events could be substantially improved if patient safety monitoring platforms had ongoing access to EHR data [10,11]. However, because EHRs are not designed with a primary focus on cross-patient data aggregation, data integration between EHRs and the more difficult scope of cross-domain integration, initiatives for integrating EHRs and Clinical Research or Patient Safety are often limited to non-scalable, one-off, system (or vendor)-specific efforts.

1.1 Two European projects: EHR-enabled clinical research (EHR4CR) and patient safety (SALUS)
The EHR4CR project (http://www.ehr4cr.eu/) is an IMI (Innovative Medicines Initiative) project funded by European Union’s Seventh Framework Programme and by in-kind contributions from member companies of the European Federation of Pharmaceutical Industries and Associations (EFPIA). The EHR4CR project is one of the largest public-private partnerships focused on providing adaptable and scalable solutions for reusing data from hospital EHRs for Clinical Research in various diseases. Implementations have been installed at 11 pilot sites throughout five European countries (France, Germany, Poland, Switzerland and United Kingdom). Collectively, the EHRs from the pilot sites contain data from over 7,000,000 patients.

The SALUS project (http://www.salusproject.eu/) is a STREP project funded by European Commission ICT Programme, eHealth Unit. Combining the strengths of individual case safety reports with EHR data, the SALUS project is focused on creating the necessary semantic and technical interoperability infrastructure to enable efficient and effective secondary use of EHR data in support of pro-active post-market safety studies.

Table 1 summarizes the EHR4CR and SALUS use cases as one of three high-level functional categories.

Table 1. Applicability IHE integration and content profiles in the EHR4CR and SALUS contexts
1.2 Semantic interoperability

One of the main challenges in integrating cross-domain data is semantic alignment of data collected in disparate contexts by different systems. Conceptual frameworks often base solutions on the existence of a common model of “shared semantics.” Common models must be based on the adoption and integration of multiple standards that themselves must be consistent, coherent, and cross-compatible [12,13,14]. Unfortunately, standards in clinical care, clinical trials, and patient safety monitoring have often been developed through parallel – and therefore somewhat inconsistent – efforts.

In the domain of patient care, efforts have focused on specifying both the syntax and the semantics of clinical information. The HL7 Reference Information Model (RIM) and EN 13606 standards define the semantics of patient care data and clearly demonstrate the need for “layers of semantic expressiveness” including: i) generic reference information models of concepts and relationships (e.g. CEN/ISO 13606, openEHR Reference Model, or HL7 RIM) each capable of binding terms from terminology models (e.g. SNOMED, LOINC, etc.) and associated with a data type models such as ISO 21090; and ii) more detailed models (e.g. CEN/ISO 13606 or openEHR Archetypes/Templates, or HL7 Detailed Clinical Models (DCMs), that instantiate generic reference models (e.g. HL7’s Clinical Document Architecture (CDA) meta-standard and the derived Continuity of Care Document (CCD)) [15,16]. In the domain of clinical research, the Clinical Data Information Standards Committee (CDISC) non-profit organization has developed a number of standards for study design including (Study Design Model (SDM))[17], study data collection (Operational Data Model (ODM)) [18], study data analysis (Analysis Data Model (ADaM)), and submission to the regulatory bodies (Study Data Tabulation Model (SDTM)). Historically, CDISC standards were not defined using the “semantic layers” described for clinical care. However, in 2004, CDISC, HL7, the National Cancer Institute (NCI), CDISC, HL7, and the FDA began the development of the Biomedical Research Integrated Domain Group (BRIDG) model containing the layered representations of the semantics of regulated clinical research data and CDISC standards can now be represented as BRIDG constructs [19]. In the domain of patient safety monitoring, the Individual Case Safety Report (ICSR) was developed using HL7 RIM-based constructs and is therefore defined using a layered approach [20].

In the context of interoperability between clinical care, clinical research, and patient safety monitoring systems, the term of metadata (literally "data about data") is used to distinguish “data collection structures” from “subject-level responses” that populate those structures, i.e. instance-level. Metadata should be described using well-defined metadata schema so as to represent the semantics of the instance data and will include concepts and relationships as well as bindings to terminologies, controlled vocabularies, taxonomies, etc. Metadata scheme may be expressed in a number of different programming languages e.g. HTML, XML, UML, RDF, etc. The core international standard used to define metadata is ISO/IEC 11179. This standard provides the definition of a "data element" registry, describing disembodied data elements. It is important to note that ISO/IEC 11179 covers just the definition of elements and does not dictate the persistence structures or retrieval strategies. In the healthcare domain, another ISO standard – ISO 21090 – plays a key role in the ISO/IEC 11179-based data element definitions since it provides the appropriate formal representation of the data type for Data Element Concept and of any type of the Value Domain data type. ISO 21090 especially provides a formal of the coded data types and addresses the binding with terminological systems.

Achieving broad-based, scalable and computable semantic interoperability across multiple domains requires the integration of multiple standards. Integrating the Healthcare Enterprise (IHE) (http://www.ihe.net) has emerged as the organization addressing this need. “Real-world usage scenarios” that can be instantiated using existing standards are published by IHE as Integration Profiles. Each profile defines a series of “transactions” which specify how existing standards should be applied to meet the overarching business goal. A set of integration and content profiles developed by several IHE committees - Quality, Research, and Public Health (QRPH), Patient Care Coordination
(PCC) and Information Technology Infrastructure (ITI) domains - collectively address syntactic issues of cross-vendor interoperability and are applicable in the EHR4CR and SALUS contexts [21-22].

1.3 Objective: EHR4CR-to-SALUS Semantic Interoperability

Each project develops solutions to achieve scalable, generalizable, cross-platform computable semantic interoperability. Our hypothesis is that cross-platform queries are achievable by implementing an additional “layer” of metadata defining mappings between each project’s metadata. Our goal is to develop a semantic MDR persisting mappings between project-specific metadata and to implement the newly defined IHE Data Exchange (DEX) profile for accessing the common metadata and enable achieve cross-project – i.e. EHR4CR-to-SALUS – interoperability [23].

2 Methods

Independently designed and implemented, project-specific metadata models were developed to provide robust semantic definitions of all data elements needed to address each project’s respective use cases. Cross-project semantic alignment was accomplished using a “semantic MDR”. Cross-project, federated, semantically integrated queries were defined and managed using the DEX profile’s “Retrieve Metadata” interface.

2.1 Metadata registries and semantic services

We defined the core content of the EHR4CR ISO/IEC 11179-based MDR as a set of data element definitions derived from HL7 RIM-based metadata models (such as HL7 CCD) bound to terminologies. This core content was enriched by specific data element definitions required to represent the semantic content in the scope of EHR4CR use cases. The concepts used in the definitions of the central data elements were mapped to corresponding local terms used in pilot sites [24]. We implemented semantic services used by the Query Builder of the Protocol Feasibility Study and Patient Recruitment modules of the EHR4CR platform (use case A). The Query Builder acting as a “metadata consumer” retrieves data elements from the MDR so that users can represent eligibility criteria as formal queries that are compliant with the common model.

The SALUS semantic MDR is implemented on top of a Triple Store layer that persists a relational model of ISO/IEC 11179-compliant [24] data elements and uses SKOS-based cross mappings of all metadata used by SALUS stakeholders. During the Individual Case Safety Report form population (use case B), the pre-population of the form is performed through the mappings of the data elements retrieved from the MDR.

A cross-project semantic MDR was set up in order to include mappings between SALUS and EHR4CR metadata and thereby became Sharing semantics across projects. Both EHR4CR and SALUS project maintain project-specific metadata models and the mappings of the resident data element definitions to local implementation-dependent models. To achieve cross-project interoperability, there is need for an interoperability specification to seamlessly share the definition of these data elements and the associated “extraction specifications.” The Data Exchange (DEX) profile focuses on providing uniform access to shared semantics via the “Retrieve Metadata” transaction.

Figure 1. Mappings between EHR4CR and SALUS MDRs and use of IHE DEX Retrieve Metadata transaction for cross-platform query execution
3 Results

3.1 EHR4CR semantic services

The current version of the EHR4CR MDR includes 105 data elements corresponding to the semantic content of the eligibility criteria of 13 clinical trials. EHR4CR data elements are related to demographic statements (gender, birth date), diagnosis (diagnosis types (n=25 SNOMED CT codes e.g. admitting diagnosis, principal diagnosis, etc) and associated value set consisting of n=12,318 ICD10 codes), findings (n=30 SNOMED CT codes and associated units or value sets), lab test results (n=34 LOINC codes and associated units or value sets), anatomic pathology observations (n=9 LOINC and associated units or value sets), procedure (value set of n=38 SNOMED CT codes), medication (substance administration)value set consisting of n=5,655 ATC codes). Pilot sites mapped their local data structures and terminologies to the EHR4CR data elements. Using the query builder of the EHR4CR workbench, a Study Manager combines EHR4CR data elements plus logical and temporal operators to populate query-templates designed for representing formally the eligibility criteria of the clinical trial. Once a query has been constructed, it is transformed into local specific representations based on the target systems thereby identifying patients meeting specific inclusion/exclusion criteria.

3.2 SALUS semantic services

The SALUS Semantic MDR provides a federated metadata repository where machine-processable definitions of data elements across domains are shared, re-used, and semantically interlinked to enable semantic interoperability and contains services that perform all required transforms between client systems. SALUS has identified a number of local models used in the interoperating sites, e.g. OMOP CDM, ASTM/HL7 CDD, E2B, and several other proprietary models whose semantics are relevant to the supported use cases. The Individual Case Safety Report form is based on the E2B content model, and SALUS Semantic MDR has the necessary mappings between the E2B fields and SALUS data elements. On the other hand, SALUS data elements have also the necessary mappings to the ASTM/HL7 CCD content model fields and the SALUS Common Model fields. Once MDR-resident common data elements have been mapped to corresponding elements in the local content models, all data transformations can be performed using Semantic MDR-resident reasoning tools.

3.3 Sharing Data Elements and query specifications across projects

EHR4CR and SALUS data elements were cross-mapped to support cross-project semantic alignment. Although for some data element mappings – e.g. patient gender, patient birth date, discharge diagnosis, procedure – were relatively easy to map (i.e. 1-to-1), the majority required more complex mappings secondary to the underlying differences in levels-of-abstraction that exist between the two MDRs. In particular, most of the SALUS Data Elements were defined using high-level generic content models) that have been constructed using generic terms such as “Result.” In contrast, EHR4CR data elements correspond to highly specific content models corresponding to specific results (e.g. Glucose [Moles/volume] in Serum or Plasma), vital signs (e.g. Systolic Blood Pressure) or problems (e.g. ECOG performance status). In the SALUS MDR, similar specificity is represented by constraining general elements using run-time, query-specific binding of value sets, a decision that was made in order to support automated terminology reasoning on data elements. Table 1 illustrates an example of mapping between two specific observations defined in the EHR4CR Data Elements and the corresponding generic SALUS Data Elements. As a result, only 5% of the SALUS Data elements were able to be directly mapped to EHR4CR data elements through skos:exactMatch. In contrast, 99% of the test set of EHR4CR data elements were mapped to SALUS data elements using skos:narrowMatch.

Table 1. Two examples of mapping between specific EHR4CR data elements and high level generic SALUS data elements

| Highly specialized EHR4CR HL7 v3 Construct & Data elements | ISO 11179 Model Construct | Corresponding generic SALUS HL7 v3 Construct & Data elements |
|------------------------------------------------------------|---------------------------|-------------------------------------------------------------|
| Observation.code                                          | Data Element Concept      | Result.code                                                 |
| 271649006-Systolic Blood Pressure-SNOMEDCT                | value_domain_datatype     | ANY (PQ, CD, CO)                                           |
| Observation.value                                         | value_domain_unit_of_     |                                                             |
| Physical Quantity (PQ)                                    | measurement              |                                                             |
| Unit (e.g. if data type is PQ) = mmHg                      |                           |                                                             |
Once mappings between EHR4CR and SALUS MDRs are established, interoperability is managed using the DEX profile’s “Retrieve Metadata” interface. For example, an EHR4CR request for patients with “Date of Birth > 1960” and “ECOG performance status >2” data and “Recent weight loss” to be performed SALUS pilot sites EHRs first locates the semantic links between the EHR4CR data elements and the corresponding SALUS data elements, then retrieves the mapping specifications of this data elements to local data base schemas as database queries. Qualified patients are returned to the metadata consumer in system-friendly form based on transformations derived for MDR mapping information.

4 Discussion & Conclusion

4.1 Contribution

In order to accomplish cross-domain semantic interoperability between domains/projects, there must be a single semantically unambiguous, processable, sharable, and technology-neutral metadata model, i.e. semantic metadata registry (MDR). The semantic MDR model should be based on a metadata definition standards such as ISO/IEC 11179. Similar construction of the semantic MDR and domain-/project-specific “local” MDRs enables efficient transformations/mappings of MDR elements to semantic MDR elements. Achieving computational semantic interoperability thus becomes a matter of defining a set of semantically unambiguous and context-neutral common metadata definitions, the universe of “shared semantics.” We first utilized a set of IHE profiles that collectively address the syntactic (non-semantic) issues involved in developing computational interoperability. We then identified the need for an additional profile to address the core semantic barrier: access to semantically annotated metadata. The DEX (Data Exchange) profile provides the technical specification for access to MDRs. The DEX profile enables the specification of queries over heterogeneous systems, projects, and domains. We demonstrated an application of DEX both within and between the two projects (SALUS and EHR4CR) i.e. across three domains (patient care, clinical research, and patient safety monitoring) and multiple participating pilot operational systems.

Our experience in designing MDRs for cross-platform semantic interoperability strongly suggests the importance of two specific implementation principles: i) the utility of using ISO/IEC 11179 as the meta-meta standard around which to construct both project-specific MDRs and the common semantic MDR; and ii) the advantage of using Semantic Web (SW) tools and technologies for the representation and sharing of cross-domain semantics. In particular, the SW approach is of considerable value since it eliminates the brittle binding of semantics to syntactic schema representational models such as RDBMS tables of XSD document trees, and instead places a serialization-independent “graph” representation of semantics – both concepts and their relationships – on the table as a “first-class citizen.”

4.2 Limitations and perspectives

The “devil in the details” is the construction of various mappings that are persisted at the two metadata “levels,” i.e. the local MDR for each project as well as the common, cross-project semantic MDR. A comprehensive description of the issues faced by each project in defining and implementing a scalable solution for achieving and maintaining mappings between the elements stored in their MDRs and the various local models of multiple EHRs and CDWs is out of the scope of this paper. Rather, the paper focused on the details of the mapping between EHR4CR and SALUS MDRs, i.e. on the content of the semantic MDR. Generally speaking, the mapping between SALUS and EHR4CR MDRs was eased by the fact that both projects refer to similar RIM-based layered metadata representations. However, in developing this mapping, we discovered that in the current version of the SALUS MDR, metadata elements had been defined at a higher level of abstraction then those in EHR4CR. This disparity
was addressed during the mapping using skos:narrowMatch (e.g. Glucose [Moles/volume] in Serum or Plasma as part of Result and Systolic Blood Pressure as part of Vital Signs) as the mapping predicate.

Semantic alignment of concepts must still be managed via human intervention. However, when the underlying representation is based on graphs rather than tables or document trees, harmonization efforts can occur “bottom-up” rather than “top-down” without disruption of global schemas, and are always focused on “pure” semantic alignment devoid of technology bindings. The SALUS project has demonstrated the value proposition of implementing their semantic MDR using a Semantic Web-based approach. We believe that the adoption of a similar approach to the representation of both semantic standards and their derivative products will prove to be the most effective tool to realize the benefits derivable from computable semantic interoperability.

Acknowledgements: We thank the members of EHR4CR WPG2 for their contribution to the development of the EHR4CR platform. We thank Jean Charlet who provided expertise on semantic web technologies and ontology mapping.

Funding: This work was supported by the Innovative Medicines Initiative Joint Undertaking, under grant agreement no 115189, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies’ in kind contribution (IMI JU web address www.imi.europa.eu). This work was also supported by 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).

Authorship and Contributorship: CD, AS, GL designed the study. CD, DK and CM developed the literature review method, with input from GD. CD, GL, SH drafted the original paper with AS. GL, LB led the analysis with substantial input from CM. DO, ES, GD and DK reviewed the original paper. CM, LB, KF and DK revised the paper for submission. CD, LB and CM responded to peer reviewers.

Competing interests: None.

Provenance and peer review: Not commissioned; externally peer reviewed.

5 References
1  Prokosch H-U, Ries M, Beyer A, et al. IT infrastructure components to support clinical care and translational research projects in a comprehensive cancer center. Stud Health Technol Inform 2011;169:892–6.
2  Murphy SN, Dubey A, Embi PJ, et al. Current State of Information Technologies for the Clinical Research Enterprise across Academic Medical Centers. Clinical and translational science 2012;5:281–4.
3  Dugas M, Lange M, Müller-Tidow C, et al. Routine data from hospital information systems can support patient recruitment for clinical studies. Clin Trials 2010;7:183–9.
4  Weng C, Tu SW, Sim I, et al. Formal representation of eligibility criteria: a literature review. J Biomed Inform 2010;43:451–67.
5  Murphy EC, Ferris FL 3rd, O’Donnell WR. An electronic medical records system for clinical research and the EMR EDC interface. Invest Ophthalmol Vis Sci 2007;48:4383–9.
6  Kohane IS, Churchill SE, Murphy SN. A translational engine at the national scale: informatics for integrating biology and the bedside. Journal of the American Medical Informatics Association: JAMIA 2012;19:181–5.
7  Cuggia M, Besana P, Glasspool D. Comparing semi-automatic systems for recruitment of patients to clinical trials. Int J Med Inform 2011;80:371–88.
8  Kush R, Alschuler L, Ruggeri R, et al. Implementing Single Source: the STARBRITE proof-of-concept study. J Am Med Inform Assoc 2007;14:662–73.
9  El Fadly A, Rance B, Lucas N, et al. Integrating clinical research with the Healthcare Enterprise: from the RE-USE project to the EHR4CR platform. J Biomed Inform 2011;44 Suppl 1:S94–102.
10 Bates DW, Evans RS, Murff H, et al. Detecting adverse events using information technology. J Am Med Inform Assoc 2003;10:115–28.
11 Linder JA, Haas JS, Iyer A, et al. Secondary use of electronic health record data: spontaneous triggered adverse drug event reporting. *Pharmacoepidemiol Drug Saf* 2010;19:1211–5.

12 European Commission. Semantic Interoperability for Better Health and Safer Healthcare. 2009.http://ec.europa.eu/information_society/activities/health/docs/publications/2009/2009semantic-health-report.pdf

13 Mead CN. Data interchange standards in healthcare IT—computable semantic interoperability: now possible but still difficult, do we really need a better mousetrap? *J Healthc Inf Manag* 2006;20:71–8.

14 Hammond WE, Jaffe C, Kush RD. Healthcare standards development. The value of nurturing collaboration. *JAHI M* 2009;80:44–50; quiz 51–52.

15 HL7 DCM. Detailed Clinical Models - HL7Wiki. 2013.http://wiki.hl7.org/index.php?title=Detailed_Clinical_Models (accessed 29 Mar2013).

16 Goossen WTF. Using detailed clinical models to bridge the gap between clinicians and HIT. *Stud Health Technol Inform* 2008;141:3–10.

17 CDISC. Medical Research Study Design in XML (SDM). 2013.http://www.cdisc.org/study-trial-design (accessed 29 Mar2013).

18 CDISC. Operational Data Model (ODM) - Certification & Archiving and Interchange of Metadata. 2013.http://www.cdisc.org/odm (accessed 8 Apr2012).

19 Fridsma DB, Evans J, Hastak S, et al. The BRIDG project: a technical report. *J Am Med Inform Assoc* 2008;15:130–7.

20 ICH. E2B (R2), Electronic transmission of individual case safety reports - Message specification (ICH ICSR DTD Version 2.1), Final Version 2.3, Document Revision Feb. 1, 2001. 2011.

20 IHE QRPH. IHE Clinical Research Document (CRD). 2010.http://wiki.ihe.net/index.php?title=Clinical_Research_Data_Capture(CRD)

21 IHE QRPH. IHE QRPH Drug Safety Content (DSC). 2009.http://www.ihe.net/Technical_Framework/upload/IHE_QRPH_TI_2009-08-10.pdf

21 IHE QRPH. IHE QRPH Data Element Exchange (DEX). http://www.ihe.net/Technical_Framework/upload/IHE_QRPH_Suppl_DSC_TI_2009-08-10.pdf (accessed 16 Jun2013).

22 Ouagne D, Hussain S, Sadou E, et al. The Electronic Healthcare Record for Clinical Research (EHR4CR) information model and terminology. *Stud Health Technol Inform* 2012;180:534–8.

25 ISO/IEC 11179. Home Page for ISO/IEC 11179 Information Technology -- Metadata registries. ISO/IEC 11179, Information Technology -- Metadata registries (MDR). 2013.http://metadata-standards.org/11179/ (accessed 13 Apr2013).