Building the Semantic Interoperability Architecture
Enabling Sustainable Proactive Post Market Safety Studies

Gokce B. Laleci Erturkmen\textsuperscript{1}, Asuman Dogac\textsuperscript{1}, Mustafa Yuksel\textsuperscript{1}, Sajjad Hussain\textsuperscript{2}, Gunnar Declerck\textsuperscript{2}, Christel Daniel\textsuperscript{2}, Hong Sun\textsuperscript{3}, Kristof Depraetere\textsuperscript{3}, Dirk Colaert\textsuperscript{3}, Jos Devlies\textsuperscript{4}, Tobias Krahn\textsuperscript{5}, Bharat Thakrar\textsuperscript{6}, Gerard Freriks\textsuperscript{7}, Tomas Bergvall\textsuperscript{8}, Ali Anil Sinaci\textsuperscript{1}

\textsuperscript{1}SRDC Ltd, Ankara, Turkey
\textsuperscript{2}INSERM, UMR_S 872, Eq. 20, Paris, France
\textsuperscript{3}AGFA Healthcare, Gent, Belgium
\textsuperscript{4}EUROREC Institute, France
\textsuperscript{5}OFFIS e.v., Oldenburg, Germany
\textsuperscript{6}Hoffmann-La Roche, Basel, Switzerland
\textsuperscript{7}Electronic Record Services B.V., Amsterdam, The Netherlands
\textsuperscript{7}WHO-Uppsala Monitoring Center, Uppsala, Sweden

\{gokce, asuman, mustafa, anil\}@srdc.com.tr
\{Sajjad.hussain, christel.daniel\}@crc.jussieu.fr
\{gunnar.declerck@utc.fr
\{hong.sun, kristof.depraetere, dirk.colaert\}@agfa.com
\{jos.devlies@eurorec.org
\{tobias.krahn@offis.de
\{bharat.thakrar@roche.com
\{g.freriks@e-RecordServices.EU
\{tomas.bergvall\}@who-umc.org

Abstract. Pre-approval clinical trials cannot guarantee that drugs will not have serious side effects after they are marketed. Post-approval drug safety data studies aim to address this problem, however, their effectiveness is started to be discussed especially after recent examples of drug withdrawals. This is due to the fact that, current post market safety studies largely depend on the submission of spontaneous case reports where underreporting is a major problem. The need for a more proactive approach is apparent, where safety data from multiple sources are actively monitored, linked and analyzed. Effective integration and utilization of electronic health records (EHR) can help to improve post-market safety activities on a proactive basis. SALUS aims to facilitate this through providing functional interoperability profiles and supporting open source toolsets enabling EHR systems and clinical research systems to communicate and

\* 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).

© Springer-Verlag Berlin Heidelberg 2011
exchange EHR data; implementing semantic interoperability solutions enabling meaningful interpretation of the exchanged EHR data; implementing security and privacy mechanisms and open source toolsets ensuring that clinical information is shared in an ethical and safe way and providing a novel exploratory analysis framework for open-ended temporal pattern discovery for safety studies on top of disparate, distributed, heterogeneous EHR Systems.

**Keywords:** Interoperability of health data, Patient safety, Re-use of Electronic health records

1 **Introduction**

The process of approving pharmaceutical agents for use in humans usually hinges on establishing the efficacy of the agent. This is usually achieved through appropriately designed and rigorously analyzed randomized clinical trials. Whilst certain safety aspects are established though the evaluation of the development program ranging from animal studies to dedicated large trials, for most approved agents however, there are limited data for ensuring the safety of the product at approval. These include, for example, the risk of rare events which are not identified in the clinical trials due to their limited size, or delayed effects of the drug due to the limited duration of the trials. Furthermore, clinical trials usually have exclusion criteria, e.g. the elderly, or pregnant women and therefore there are little or no data on certain groups available prior to approval, but who may ultimately use the agent. Moreover, the pattern of drug use in clinical trials may not necessarily reflect the real-world use once the drug reaches the population and therefore may impact safety.

For these reasons, while pre-market safety analysis through clinical trials remains vital, there is considerable attention towards improving the reporting and collection of post-market data. Current post-market safety surveillance and reporting activities are largely based on reports of suspected adverse drug reactions sent to the regulatory bodies by medical professionals, and in some countries by patients themselves. While spontaneous reporting remains a cornerstone of pharmacovigilance in the regulator environment, and is indispensable for signal detection, recent examples of drug withdrawals [1] due to uncommon adverse events after millions of patients were exposed, the need for a more effective and proactive surveillance is reinforced. One of the main problems of the current drug surveillance system is underreporting: It has been estimated that only around 5% of adverse drug events (ADEs) are reported through spontaneous reporting [2]. This is partially due to fact that overloaded medical personnel do not always see reporting as a priority. Another issue is that detecting adverse events may not always be straightforward, hence can be overlooked. Also, as it is not possible to estimate the number of patients taking a drug from spontaneous reports, they do not readily enable the quantification of any risk. Therefore the benefit-risk profile of the product cannot be determined.

It is evident that whilst more aggressive post market surveillance methods are needed, they should not create additional burden for the health care provider or the patient.
We believe that an effective integration and utilization of electronic health record (EHR) data can help to improve post-market safety activities will result in:

- Strengthening the spontaneous reporting process by automated ADE detection tools screening EHRs in a hospital so that ADE reporting burden can be overcome within a clinical institute. This can increase data accuracy as it eliminates manual screening of clinical care data for identifying ADEs.
- Enabling ADE reporting by extracting the available information from the EHRs into the individual case safety reports to avoid double data entry. This ensures delivering timely feedback to the regulatory bodies via automatic EHR supported adverse event reporting.
- Strengthening the current signal detection processes in SRS centers for tracing case reports to their corresponding patient records to allow actual incidence rates to be computed, and to provide additional information on extended parts of the underlying medical history of the patient.
- Enabling real time screening of multiple, distributed, heterogeneous EHRs for early detection of ADE signals. This facilitates proactive safety monitoring as a complementary approach to reactive signal detection based on spontaneous reports.
- Enabling sustainable and scalable EHR re-use facilitating wide scale outcome and effectiveness research, to be able to observe selected cohorts of patients over an extended period of time screening multiple, distributed, heterogeneous EHR systems.

Another important advantage of EHR is that they offer the potential for quantification and contextualization of any risk though safety studies and therefore help determine the benefit-risk profile of the drug, which ultimately determines the use of the drug in population. To facilitate these wide scale proactive post market safety studies, there is a need for a new capacity enabling access to the data locked in multiple different heterogeneous EHR systems: an interoperability architecture. This interoperability architecture should enable execution of proactive post market surveillance studies by the pharmaceutical companies and regulatory bodies in cooperation with the heterogeneous, distributed EHR systems.

SALUS Project [3], an R&D project co-financed by the European Commission's 7th Framework Programme (FP7), aims to create the necessary semantic and functional interoperability infrastructure to enable secondary use of EHR data in an efficient and effective way for reinforcing the post market safety studies, addressing the challenges that are presented above. The SALUS approach in building this interoperability architecture is briefly introduced in Section 2. Section 3 presents the related work, and finally Section 4 concludes the paper.

2  SALUS Interoperability Architecture

One of the first challenges to address is achieving syntactic and functional interoperability between EHR systems and clinical research systems. In SALUS, our approach is to provide functional interoperability profiles and open source tools to query EHR
data for ADE identification, ADE reporting and signal follow-up studies and to subscribe to clinical data for a selected cohort of patients for signal detection and outcome research over distributed EHRs. These interoperability profiles aim to enable the EHR systems and clinical research systems used for running post market safety studies to communicate and exchange data.

In order to enable the collaborating systems to automatically interpret the queries and the resulting clinical data exchanged meaningfully and accurately for producing useful results, there is a need for a semantic interoperability layer built upon these functional interoperability profiles. In SALUS we address this through a Semantic Web Framework. The Semantic Web (SW) framework allows knowledge representation via SW ontologies by providing constructs that represent shared vocabularies from multiple domains, such as medicine and health care [4]. Aiming towards semantic interoperability among heterogeneous knowledge sources in the health care domain, the main challenges to be addressed [5] are: i) Context-awareness: Identifying context-specific components from multiple knowledge sources relevant to the clinical problem at hand; ii) Modularity: Reusing relevant fragments from multiple knowledge sources; iii) Profile and Policy Management: Treating internal policies or profiles distinctively; iv) Correspondence Expressiveness: Relating heterogeneous knowledge, either within or between contexts; v) Dealing with Inconsistencies: Repairing or tolerating incompatibilities or inconsistencies, within or between contexts. In SALUS, we are utilizing all these facets for enabling semantic interoperability except Profile and Policy Management for the time being.

One of the first pre-requisites for such a semantic interoperability architecture is to have a common set of data elements as the core data set that is of interest to data analysis tools used for post marketing safety studies. The role of SALUS semantic interoperability framework is semantically mediating the clinical data represented through syntactically different but semantically equivalent EHR content models to one another, by making use of the core set of common data elements as the common denominator. This is facilitated through an ontological framework.

In the following subsections, we will further analyze the need for functional and semantic interoperability solutions for enabling post market safety studies on top of EHRs, and present SALUS approach to realize these.

### 2.1 Providing Functional Interoperability Profiles between EHR Systems and Clinical Research Systems Enabling Proactive Post Market Safety Studies

Achieving syntactic and functional interoperability between the EHR systems and clinical research is the necessary condition for further enabling semantic interoperability.

The requirements for syntactic and functional interoperability profiles for enabling proactive post market safety studies can be summarized as follows:

- Different rule based intelligent data analysis algorithms can be plugged on top of available EHRs to detect ADEs by checking the administered drugs, laboratory test results, vital signs, findings and diagnoses and report them to the medical practi-
tioners in order to facilitate reporting of ADEs. Standard based protocols are needed to specify the data to be screened in a machine processable manner, to feed these data to the intelligent data analysis algorithms, and to send the suspected ADE list back to the medical practitioners seamlessly within an EHR session. In this way it will be much easier to integrate several different ADE detection tools with heterogeneous EHR systems in order to conduct scalable distributed post market safety studies.

- Upon detection of such ADEs through these enabling mechanisms, if confirmed by the physician, these should be reported to regulatory bodies through standard based individual case safety reports. Most of the data that needs to be reported within these case reports are mostly available in the EHRs. In order to avoid duplication of effort to fill in these data to the case safety reports once again manually, there is a need for interoperability profiles to fill in these standard based forms by extracting the required information from the underlying EHR system, and sending the completed forms to the respective regulatory bodies.

- Spontaneous reports only report the ADE incidents. The information related to the percentage of other patients who used the drug but not experienced ADEs, i.e. the denominator data, is missing. On top of that, these reports may fail to take into account important information about the patient such as underlying medical conditions and the other drugs the patient may be taking. An ideal system for adverse drug reaction surveillance would combine the strength of case reports with those of EHRs [6]. Standard interfaces are needed for tracing case reports to their corresponding patient records, to query the EHRs to enable absolute reporting rates to be computed, and to retrieve additional information on extended parts of the underlying medical history of the patient.

- Strengthening available spontaneous reporting will already be a step forward for post marketing safety studies. However, in order to realize near-real time proactive post market safety studies, there needs to be a mechanism for screening the available heterogeneous and disparate EHRs for a specified time period for adverse event signal detection and also for conducting observational studies for validation of the suspected signals and for carrying out outcome research to see long term effects of drugs. There is a need for a machine processable mechanism to identify the target patient population for eligibility checking and the details of the clinical data to be collected, together with an exchange mechanism to feed the pseudonymized patient data to a common registry.

Syntactic and functional interoperability can be achieved if and when two or more systems are capable of communicating and exchanging data by specifying communication protocols and data formats. For addressing the requirements of the proactive post market safety studies summarized above, there is a need for standardized interfaces between EHR systems and clinical research systems:

- To specify the eligibility criteria to select the patients for the specific post market safety study
To specify the clinical data requested by an intelligent data analysis tool (ADE detection/signal detection/outcome research) to be fed to a clinical data registry for the selected group of patients

To transfer the specified clinical data to the clinical data registries for the selected patients

To query the underlying EHRs to support follow-up studies for signal detection

To query the underlying EHRs to fill in the ADE reports after an ADE is detected

To define such standardized interfaces, in the SALUS project we will follow a “profiling” approach. The profile concept aims to eliminate the need for a bilateral agreement between any two information exchange partners by defining a standard set of messages/documents, choreographies, business rules and constraints. By analyzing the selected use case scenarios the required transactions will be determined, and for each of them, the best suitable standards for enabling these transactions will be chosen. SALUS functional interoperability profiles will be based on the available initiatives for achieving syntactic interoperability for re-use of EHRs for clinical trial execution, such as IHE Retrieve Form for Data-Capture (RFD) [7], Clinical Research Data Capture (CRD) [8], Drug Safety Content Profile (DSC) [9] Profiles, HL7 Clinical Research Filtered Query Service Function Model (CRFQ SFM) [10], and where suitable by proposing the necessary extensions for enabling such a standard based interoperability architecture for post-market surveillance.

2.2 Providing Semantic Interoperability Framework Enabling Proactive Post Market Safety Studies

Beyond the ability of two or more computer systems to exchange information through addressing syntactic and functional interoperability requirements, semantic interoperability is needed to automatically interpret the information exchanged meaningfully and accurately in order to produce useful results [11].

Semantic interoperability is a prerequisite for enabling secondary use of EHRs in post market safety studies so that clinical data is consistently captured from the EHR systems and analyzed by intelligent data analysis algorithms.

The need for semantic interoperability between clinical research and clinical care systems may stem from the following facts:

- Clinical statements may be represented through different reference and domain information models, like HL7 RIM and CDA, CEN/ISO 13606 5 part (E.g. Reference Model plus Archetypes/Templates, Patient Mandate) or CDISC ODM/SDTM. CDISC [12] is a non-profit organization that develops standards covering almost all the steps within a regulated clinical research study including study design (SDM), study data collection (ODM and CDASH), study data analysis (ADAM) and submission to the regulatory bodies (SDTM).
- Although the same information model is selected, consider for example HL7 CDA, the same clinical information can be represented through different compositional aggregation of clinical statements. To address this problem, these information
models could be restricted more through CDA templates or CEN/ISO 13606 archetypes or CDISC ODMs based on selected CDASH data sets. Content models based on one or more of these possibilities can be created. Yet, the content models used by clinical research and care domains usually differ.

- Although the same content model (template/archetype) is chosen, clinical statements may be represented using coded terms from different terminology systems.

To address these problems, semantic interoperability framework should not only handle the structural mappings of two different information models, but also resolve the semantic mismatches due to the use of different terminology systems and different compositional aggregations to represent the same clinical concept differently even when a single information model is used.

In order to be able to detect such semantic similarities, there needs to be a common harmonized ontology that represents the semantics of reference information models, templates, archetypes and the terminology systems used. In SALUS Interoperability Framework we aim to provide tools to create and maintain such a harmonized ontology as a linked set of ontologies, and also to provide the semantic mediators that run on top of this harmonized model to map message payloads represented in one content profile to another in cooperation with terminology servers. SALUS common harmonized ontology aimed to act as a common denominator for exchanging clinical data required for proactive post market patient safety studies between clinical care and research systems. This ontology shall be based on the already existing standards used in clinical care and research domains and the already existing data sets. To be able to create this common harmonized ontology in a systematic and scalable way to serve the needs of SALUS use cases, the following activities are aimed to be carried out based on our methodology:

- Identifying the clinical information requirements of the selected SALUS use cases and the related transactions enabling safety research. For this, we aim to agree on and design content models as the message payloads to be exchanged based on the prominent clinical data exchange standards used in clinical care and research domains (such as CDISC ODM models, CDA templates, CEN/ISO 13606 archetypes). There is an important initiative in this respect, namely the intended CEN/ISO Semantic Interoperability Artefact Modeling Standard (SIAMS) [13]. SIAMS approach provides a generic, neutral, standard approach for defining models of use which is not specific to CEN/ISO 13606, openEHR or HL7 CDA syntax. This generic model of use can be translated into 13606 archetypes, openEHR archetypes or HL7v3 CDA based content templates when necessary. We aim to develop a generic editor for Clinical Information Models (CIMs) for defining models of use, which then can be translated to different standards like HL7 CDA, CDISC ODM and CEN/ISO 13606.

- Based on these information requirements expressed as content models, identifying the core common data element (CDE) set as meaningful fragments/building blocks to be used to create such content models. We will provide necessary tools to create, select, adapt and manage the CDEs in this core data set in conformance to ISO/IEC 11179 standard for metadata registries [14].
• Building a harmonized patient safety study ontology on top of this core CDE set. This evolving ontology will act as a common semantic dictionary of the clinical terms to be exchanged between EHR systems and clinical research systems. Supporting ontology fragments from domain specific ontologies (such as drug ontologies, specific disease ontologies) should also be added to this common ontology by providing the necessary semantic interlinkages between these ontology fragments. It should also cover the related fragments of the terminologies used in clinical care and research domains like MedDRA, LOINC, ATC, SNOMED CT, ICD-10, WHODD.

• On top of this harmonized ontology, SALUS Semantic Mediation Framework aims to enable the underlying clinical research and clinical care systems to communicate through their already existing domain information models. For this, we support two complementary approaches:
  — We enable the clinical research and clinical care systems to communicate via the functional interoperability profiles by exchanging message payloads in conformance to the specified content models. In this way, they become able to communicate through the existing well accepted standards they choose like archetypes, CDA templates and CDISC ODM models. SALUS provides tools to semantically mediate these semantically similar but syntactically different content models to one another by exploiting the harmonized SALUS ontology. (subsection 2.2.1)
  — We enable the development of semantic interfaces on top of their existing proprietary systems that will enable the EHRs and clinical research systems to communicate directly through the harmonized SALUS semantic model. For this, we aim to develop wrappers to translate these semantic interfaces enabled through semantic query languages like SPARQL to the native interfaces supported by the available systems, like SQL. (subsection 2.2.2)

In this way, we provide different approaches for clinical care and research systems that have varying levels of semantic awareness.

2.2.1 An initial prototype of Semantic Mediation in SALUS Architecture

An initial prototype of SALUS Semantic Interoperability platform is already available [15]. In this implementation, the harmonized ontology is based on an RDF representation of the BRIDG Domain Analysis Model [16], which already provides a coherent clinical research vocabulary that integrates established domain knowledge from existing standards developed at CDISC, HL7. The harmonized ontology is also semantically linked with the RDF representations of the terminologies and controlled vocabularies downloaded from BioPortal [4], and served over a triple store. The semantic models of two prominent EHR content models, namely HL7 Continuity of Care Record, and CEN/ISO 13606 EHR Extract for patient summaries are defined and semantically mapped to the harmonized ontology through ontology mapping tools. While defining the mappings between the corresponding ontology fragments, we utilize SPARQL rules. In this way it becomes possible to retrieve medical sum-
maries in HL7 CCD and CEN/ISO 13606 format and map them to the common model and serve them through a common knowledge base. A sample library of semantic queries in SPARQL is available to query the common knowledge base through CDASH variables. The semantic queries run through reasoners are also capable of addressing the semantic mismatches in the queries and the patient summaries available in knowledge base due to usage of different terminology systems. This is achieved by exploiting the mappings between terminology system ontologies and semantic linkage of terminology ontologies with the harmonized ontology. In this way, clinical researchers conducting post market safety studies can seamlessly collect the required pseudonymized patient medical data from heterogeneous EHR systems and semantically query the collected data for clinical analysis. Details of the architecture and an example scenario can be found in [15].

2.2.2 Building Semantic Interfaces on top of EHRs

In order to directly access EHRs from different EHR vendors with the harmonized ontology, each EHR system has to be extended with a semantic interface to expose their EHR content semantically. Such a semantic interface varies based on the characteristics of the EHR system it connects to as well as the performance requirement of the semantic interoperability platform. It could either statically dump the entire content of the system as RDF graphs, or allow dynamically execution of SPARQL queries on an EHR system. For the applications creating a mashup of EHR systems, as the patient records are continuously changing, building a SPARQL endpoint on top of EHR systems is preferred in order to provide up to date information.

A SPARQL endpoint maps attributes of an EHR to a semantic model, so that SPARQL queries could be translated to SQL (for relational database) queries to look up data in an EHR system, and return results as RDF using the ontology of the specified semantic model. Theoretically, it is possible to map different EHR systems directly to the harmonized core ontology, however, in practice, the big semantic gap between an EHR system and the harmonized ontology is hard to bridge in one step. Therefore, a two step approach is adopted to formalize the EHR systems to allow them to flexibly and explicitly express their records with the harmonized ontology. Local semantic models are developed for each EHR system respectively based on the data structure of their local systems, e.g. for a relational database, a table is mapped to an ontology class and a column is mapped to an ontology property. After retrieving RDF results represented in a local ontology, a conversion process is carried out to map the local ontology to the harmonized ontology, so that the RDF results from an EHR system are finally represented with the harmonized ontology and can be easily integrated with results from other systems as they share the same ontology.

3 Related Work

Given the reality that achieving broad-based, scalable interoperability across multiple domains will require the integration of multiple standards, it became apparent that an “integration organization” involving multiple stakeholders (including both vendor and
provider organizations) could serve a valuable role by defining – based on stakeholder dialogue – “real-world usage scenarios” – that, in turn, could then be instantiated using existing standards. Integrating the Healthcare Enterprise (IHE) has, in fact, emerged as such an organization and provides Integration Profiles based on “real-world usage scenarios”. One of such profiles is the Retrieve Form for Data-Capture (RFD) Profile [7] which is a joint initiative of CDISC and Integrating Healthcare Enterprise (IHE). It aims to facilitate data capture from EHR systems for clinical research. For this, it provides a standard way of transferring and displaying external data capture forms inside an EHR through XForms. On top of RFD, a content profile is proposed, namely the Clinical Research Data Capture (CRD) Profile [8], which defines a standardized content model in ASTM/HL7 Continuity of Care Document (CCD) format, and also in CDISC ODM/CDASH format. It also defines an XSLT mapping between these models. In this way, it aims to facilitate the pre-population of Case Report Forms by extracting this content from EHRs in CCD format, translating them to ODM/CDASH format and sending them to sponsors seamlessly. Similarly, IHE Drug Safety Content Profile (DSC) [9] facilitates exchange of ADEs.

Similarly, the HL7 Clinical Research Filtered Query Service Function Model (CRFQ SFM) [10] provides a service interface specification between clinical trial systems and EHR systems. In particular it covers three functionalities: searching a suitable trial protocol given patient parameters, searching suitable patients for trials given inclusion/exclusion criteria, and searching an EHR repository for a set of patients satisfying a particular safety event signal.

These standard based initiatives aim to enable the use of EHR systems during clinical trials effectively. However, these standards usually address the problem of technical interoperability: they specify the interfaces to communicate with EHR systems resolving structural interoperability (like HL7 CDA), and set of necessary functionalities that need to be implemented by EHR systems to address functional interoperability. When it comes to semantic interoperability, we face with more demanding challenges. The use of EHRs for patient safety requires semantic matching between the representations of clinical data in study design models on one side (eligibility criteria, data element of eCRFs) and in heterogeneous clinical systems on the other side.

CDISC SHARE initiative [18], a globally accessible electronic library is envisioned which enables precise and standardized data element definitions to be used by applications and studies to improve biomedical research and its link with healthcare. Similarly, National Cancer Institute in USA, developed a coherent set of Common Data Elements (CDEs), as well as the Cancer Data Standards Repository (caDSR) as a means to access, add to or modify CDEs [19]. The CDISC ASPIRE [20] Project is initiated to create a structured representation of a core set of encoded protocol eligibility criteria, using accepted medical terminology and vocabulary standards when available. Weng and coll. recently surveyed the literature about computable representations of eligibility criteria [17]. They analyzed the issues related to the expression language for representing eligibility rules, the encoding of eligibility concepts and the modeling of patient data. Although these initiatives set the interface as fixed as possible, the problem of translating clinical statements coded in local terminologies and local domain information models to these standard interfaces remains unsolved.
To demonstrate the value of Semantic Web specifications in bridging the divide between clinical practice and clinical research, a W3C task on Clinical Observations Interoperability (COI) is established [21]. As a use case for secondary use of EHRs for clinical research, checking clinical trial eligibility for patient recruitment is chosen and a prototype implementation is achieved. In this prototype, first of all, the eligibility criteria are specified as SPARQL queries based on an ontology derived from CDISC SDTM standard. Mappings are defined from SDTM Model to HL7 RIM Model, and from the HL7 RIM Model to EHR database schema. This work is an important effort to show the potential of the Semantic Web for bridging the clinical care and research domains. However, in this approach, a direct mapping between SDTM model to the HL7 RIM is proposed, which would not be practical when the number of standards to be harmonized increases. In Patel et. al [22], an ontology based method for matching patient records to clinical trials for cohort selection is presented. The medical records of Columbia University Medical Center which are coded through a MED taxonomy are transformed into SNOMED CT based ontology instances through custom written rules. The mappings between MED taxonomy and SNOMED terms are achieved semi-automatically through UMLS and NLP based mapping algorithms. The cohort selection is achieved through semantic queries executed on top of this knowledge base supported through a DL reasoner. Although terminology reasoning is included in this approach, the standards used by clinical research, such as CDISC have not been included in the study, i.e. the translation of eligibility criteria in CDISC standards to semantic queries has not been addressed.

Finally there are efforts to design, develop and validate a surveillance system that analyses data from EHRs and biomedical databases to detect adverse events. EU-ADR Project [23] is one of them, where data from eight hospital databases is used for ADE detection. Custom written queries have been used to access these eight different databases; data is retrieved through these queries and stored in a central database to perform signal detection through the proposed algorithm. Similarly in OMOP [24], a common data model is established and ETLs are defined to transfer the clinical data in different EHRs and claim databases to a central repository to run signal detection methods on top of them. These approaches do not propose semantic features to enable seamless collection of EHR data for safety analysis, where SALUS project complements these.

4 Conclusions

In SALUS Project we aim to create the necessary semantic and functional interoperability infrastructure to enable secondary use of EHR data in an efficient and effective way for reinforcing the post market safety studies. The project is initiated at February 2012, and currently the consortium is working on identifying pilot application and system requirements in parallel with the effort of building early prototypes. In this workshop the results of requirement analysis process and early SALUS prototypes will be presented together with the challenges and barriers in realizing the envisioned interoperability architecture.
References

10. HL7 Clinical Research Filtered Query Service Function Model (CRFQ SFM), http://www.hl7.org/Special/committees/rcrim/projects.cfm?action=edit&ProjectNumber=541
21. W3C task on Clinical Observations Interoperability (COI), http://esw.w3.org/HCLS/ClinicalObservationsInteroperability
23. EU-ADR, www.alert-project.org/