Demonstration of the SALUS Semantic Interoperability Framework for Case Series Characterization Studies

Mustafa Yuksel[1,2], Suat Gonul[1,2], Gokce Banu Laleci Erturkmen[1], Ali Anil Sinaci[1,2], Kristof Depraetere[3], Jos De Roo[3], Tomas Bergvall[4]

[1] SRDC Software Research & Development and Consultancy Ltd., Ankara, Turkey
{mustafa, suat, gokce}@srdc.com.tr
[2] Department of Computer Engineering, Middle East Technical University, Ankara, Turkey
[3]Advanced Clinical Applications Research Group, Agfa HealthCare, Gent, Belgium
{kristof.depraetere, jos.deroo}@agfa.com
Uppsala Monitoring Centre, WHO Collaborating Centre for International Drug Monitoring, Uppsala, Sweden
Tomas.Bergvall@who-umc.org

Abstract. This work aims to demonstrate the interoperability framework developed in the SALUS project which enables effective integration and utilization of EHR data to reinforce post-market safety activities.

Keywords: Secondary Use, Post Market Safety Studies, Electronic Health Records, Interoperability, Semantic Web

1 Introduction

Currently, pharmacovigilance activities are mainly based on signal detection studies run on voluntarily sent spontaneous reports. This process has several bottlenecks, only about 5% of adverse drug events (ADEs) are being reported [1]; the quality of the data collected through spontaneous reporting is low and finally spontaneous reports only report adverse incidents while the information related to other patients who used the drug but did not experience any adverse events, i.e. the denominator data, is missing. For these reasons there is a clear need for complementary pharmacovigilance activities. Compared to adverse drug event reports, Electronic Health Records (EHRs) cover extended parts of the underlying medical histories, include more complete information on potential risk factors, and are not restricted to patients who have experienced a suspected ADE [2]. Hence, there is a great potential in accessing EHRs for tracing safety reports back to medical summaries of patients, and also secondary use of EHRs for complementary pharmacoepidemiology studies for clinical signal evaluation and validation. For example, Uppsala Monitoring Centre (UMC) on behalf of the WHO International Programme for International Drug Monitoring analyzes the WHO global individual case safety reports (ICSR) database, VigiBase, for potential signals [3]. The main objective is to find new ADE previously unknown or incompletely documented. One aspect of this is to characterize the reported cases in comparison with a selected background population for checking whether there are other explana-

tions more likely to cause the reported adverse event than the exposure to the drug of interest. Yet, the data sets used for such studies are limited both in quantity and also considering the extent of medical information covered and geographical spread. Accessing a wide range of EHR sources seamlessly to collect the background information of any selected patient population, more importantly tracing the reported incidents back to their original EHRs, can provide major improvements for such clinical validation studies, as we demonstrate in this paper.

This work aims to demonstrate the interoperability framework developed in the SALUS project [4], which enables effective integration and utilization of EHR data to reinforce post-market safety activities. The objective is to enable safety analysts to seamlessly access EHR data from heterogeneous healthcare systems. We demonstrate that with the SALUS architecture it becomes possible to collect de-identified medical data sets of selected populations to run complementary safety analysis studies for adding meat to the bones of the potential signals.

2 SALUS Semantic Interoperability Approach for Case Series Characterization

The functionalities of the SALUS Interoperability framework will be demonstrated through an example scenario. In this scenario, a safety analyst at the UMC runs a case series characterization study for evaluating the validity of a potential signal, in particular the effect of nifedipine on myocardial infarction (MI) events. The safety analyst need to access medical data sets of selected populations (e.g. foreground population of patients having MI within two weeks of nifedipine intake, and the background population of all patients taking nifedipine) from disparate EHR systems to be able to check whether there are other explanations more likely to cause MI than the exposure to nifedipine.

The SALUS system provides the Case Series Characterization Tool (CSCT) Web application, which enables the safety analyst to formally define the characteristics of foreground and background populations. It is possible to define eligibility criteria by expressing several different clinical statements, such as conditions and medications. Such criteria are represented by selecting coded values from terminology systems; for example, the medical event of interest can be defined by selecting "myocardial infarction" MedDRA Preferred Term (PT), and medication of interest by selecting "nifedipine" from WHO-ATC (see Figure 1). The terminology systems to be used in these fields are configurable; e.g. another analyst may prefer to use SNOMED CT for defining problem codes. For enabling efficient type-ahead search functionality during code selection, the tool is integrated with a terminology server that indexes medical terminologies. It is also possible to define logical operators and temporal constraints among different criteria. The tool also enables the safety analyst to configure the statistics to be calculated for grouping and stratifying data sets of the eligible populations based on age, gender, common medications/events before/after medication/event of interest.

The coded data can be configured to be grouped under a preferred terminology system and level in the results, for example MedDRA High Level Group Terms (HLGT), no matter which local terminology system is used in the EHR sources. Finally, it is possible to define a number of coded risk factors to be specifically checked on both populations.

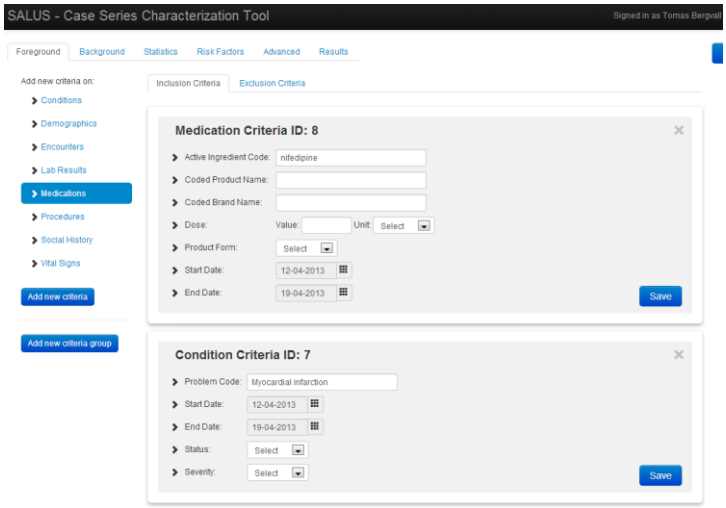

Figure 1 Case Series Characterization Tool-Eligibility Criteria definition

2.1 Challenges and how SALUS addresses them

First of all to be able to collect population data from multiple hospitals, the eligibility criteria need to be passed to disparate EHR sources, the coded eligibility criteria needs to be translated to the terminology systems used in EHRs and the de-identified medical data sets should be retrieved for the eligible patients. After aggregation, these medical data sets need to be analyzed to calculate the statistical information asked by the safety analyst. However, there are several challenges: i) divergent data models are used to represent EHRs, and ii) several different terminology systems are used to code structured patient data.

In our architecture, we address these problems by formalizing the local models of EHR sites and semantically aggregating them using a common model, which we call SALUS Common Information Model (CIM). CIM ontology forms the core of the SALUS Semantic Resource Set, with the aim of preventing n-to-n mappings among varying content models of data sources and requestors. CIM is constructed through ontological representation of the SALUS Common Data Elements required to be processed in our pilot cases and include elements to be present within a medical summary, such as patient demographics, encounter, condition (problem, diagnosis), allergy, family history, healthcare provider, and their sub-elements.

In SALUS we collect EHR data in the local models used by the EHR systems. In the demonstration case we have two data sources: one providing the medical summar-

ies of eligible patients in HL7 CCD/ Patient Care Coordination (PCC) templates, second one through a SPARQL interface implemented on top of the proprietary relational data model of the local EHR. In both cases, in order to proceed with semantic mediation, the first thing that has to be done is formalizing the retrieved EHR data by representing them as RDF entities in local ontologies corresponding to the local models. Then these are converted to SALUS CIM model through semantic mediation rules in Notation3 (N3) implemented in Euler Yap Engine (EYE).

Once we retrieve the medical summaries of the eligible patients in SALUS CIM format, the next step is to run the semantic queries on top of them to calculate the statistics requested for grouping and stratifying data sets of the eligible populations. Although all the patient data is represented using SALUS CIM, semantic interoperability has not been achieved yet as clinical statements in CIM can be expressed with codes from different terminology systems based on the preference of local systems. For example in our scenario we have the requirement of grouping the clinical conditions coded in ICD-9-CM and ICD-10-GM in source EHR systems as MedDRA HLGT terms, as the researcher prefers to see them in MedDRA, which is widely used in clinical research domain. SALUS CIM is linked with ontological representations of terminology systems; hence before the statistics are calculated on the aggregated data represented in CIM, terminology reasoning is handled to address not only structural but also semantic mismatches between data sources and the requestor.

3 Conclusion

We will demonstrate that the quantity and quality of the information provided by SALUS CSCT to the UMC safety analysts is a significant improvement compared to what they are able to access using traditional methods. Previously it was not possible to collect sufficient statistics about the underlying medical conditions of the background population, which becomes possible with SALUS by seamlessly collecting data from disparate EHR sources.

Acknowledgements. This work was supported by funding from the SALUS project (http://www.salus project.eu/). Grant agreement N° 287800.

References

1. Bates DW, Evans RS, Murff H, et al. Detecting adverse events using information technology. *J Am Med Inform Assoc* 2003;**10**(2):115-128.
2. Norén GN, Edwards IR. Opportunities and challenges of adverse drug reaction surveillance in electronic patient records. *PharmacoVigilance Review* 2010;**4**(1):17-20.
3. Norén GN, Edwards IR. Modern methods of pharmacovigilance: detecting adverse effects of drugs. *Clin Med* 2009;**9**(5):486-489.
4. SALUS Project, Scalable, Standard based Interoperability Framework for Sustainable Proactive Post Market Safety Studies, http://www.salusproject.eu