Semantic-sensitive extraction of EHR data to support adverse drug event reporting

Gunnar Declerck[1], Sajjad Hussain[1], Yves Parès[1], Christel Daniel[1], Mustafa Yuksel[2], Ali Anil Sinaci[2], Gokce Banu Laleci Erturkmen[2], Marie-Christine Jaulent[1]

[1] INSERM UMRS 872, Eq. 20, Paris, France
{gunnar.declerck, sajjad.hussain, yves.pares, christel.daniel, marie-christine.jaulent}@crc.jussieu.fr
[2] SRDC Ltd, Ankara, Turkey
{mustafa, anil, gokce}@srdc.com.tr

Abstract. The reasons behind adverse drug events (ADE) getting underreported by medical professionals are overlooking complex drug reactions and dealing with cumbersome manual process of reporting ADE based on patient profiles. We present an initial design of SALUS ICSR reporting tool that supports the reporting of ADE to regulatory authorities with services (i) enabling automatic prepopulation of reporting forms by extracting patient data from EHR and (ii) presenting pre-filled reporting forms to medical professionals for further completion and validation. The main objective of this tool is to ease the process of filling ADE reporting forms and increase the quality of reported data. To enable the compatibility of our reporting tool with heterogeneous EHR data models, SALUS interoperability platform supports the patient data extraction process and ensures the reporting of ADE in a standardized format expected by regulatory authorities.

Keywords: adverse drug event reporting, semantic interoperability, secondary use of EHR.

1 Introduction

Current post-market drug surveillance is largely based on reporting suspected adverse drug reactions to the regulatory bodies by medical professionals. This process is historically referred to as 'spontaneous reporting' because it relies on the active efforts by the reporter. One of the main problems spontaneous reporting systems (SRS) face is underreporting [1,2]. It has been estimated that only around 5% of adverse drug events (ADEs) are reported through SRS [3,4]. In the United States, less than 1% of ADEs are reported to the Food and Drug Administration (FDA), although they are frequently described in the electronic health record (EHR) systems [5]. This alarming situation is partially due to the fact that detecting ADE may not always be straightforward, hence can be overlooked. In addition, completing an individual case safety report (ICSR) is generally a costly operation in terms of time and labour needed. On the other hand, if medical professionals are not sufficiently aware of the importance

adfa, p. 1, 2011.

of reporting for patient safety, they will not see the benefit to devote their time to this activity.

There have been efforts in building automated systems for detecting ADE [3], as well as extracting patient data from EHR to prepopulate ICSRs [5]. In the present work we focus on the latter kind. One of the main technical challenges for building an ICSR reporting system dealing with secondary use of EHRs is establishing technical and semantic interoperability between EHR systems and regulatory bodies collecting ICSRs. Current EHR systems use heterogeneous information models to record patient data in local data warehouses, whereas ICSRs need to be in compliance with the ICH E2B(R2) standard [6]. In the context of the SALUS European project [7], we are developing an ICSR reporting tool that ensures the possibility of extracting patient data from EHR for prepopulating ICSR forms, by establishing semantic interoperability between the information model used in the local EHR and the E2B standard. The present paper describes this new approach and the semantic mediation platform it relies on.

2 Related Work

One of the recent attempts in solving the underreporting problem by reusing EHR data is ASTER proof of concept pilot project [5,8]. ASTER application enables automatic extraction of data from EHR to prepopulate an ADE report and direct electronic submission to the FDA. When the physician discontinues a drug due to an ADE in the EHR interface, a prepopulated report (demographics, product name and some other data elements are already filled-in) is automatically displayed. The physician only has to complete a small amount of additional information before sending the form. The physicians who tested the ASTER application agreed unanimously on its interest. However the ASTER application has limitations: (1) the extraction of EHR data is not built to be interoperable with several EHR information models, and (2) the form completed by the physician has to be processed manually by an intermediate instance, in charge of putting the form in the proper format for electronic reporting to FDA. SALUS ICSR reporting tool aims to overcome these limitations.

In addition, several initiatives have been made to specify how data should be extracted from EHRs to prepopulate ADE reporting forms. The most advanced one is the IHE Drug Safety Content profile (DSC) [9], an integration profile built as an add-on to the Retrieve Form for Data Capture profile (RFD) [10]. RFD specifies a generic protocol for handling information collected from electronic forms through a description of actors and transactions. It does not provide concrete details and specifications for actors' implementation and identifying the data to be transferred. To overcome such limitations, DSC focuses on the definition of the RFD retrieve form transaction [10], describes data needed to pre-fill the forms using HL7 Continuity of Care Document (CCD) data structure and how to convert it to the standard E2B data model used for ADE reporting. Nevertheless, fields and mappings to E2B are only partially defined. Pseudonymisation is also not covered by DSC and the profile is still in trial implementation supplement status.

3 SALUS Semantic Interoperability Approach for ADE Reporting

The SALUS platform is designed to ensure the possibility of extracting patient data needed to prepopulate ICSR forms, by converting the information model data elements used in the patient EHR to the requested ICSR data model, the ICH E2B(R2) in our case. ICH E2B(R2) [6] is a standard used by the World Health Organization (WHO) Collaborating Centre for International Drug Monitoring, which specifies an information model for ICSRs and a protocol for their electronic transmission.

3.1 SALUS Semantic Interoperability Approach through an Ontology of Common Data Elements

SALUS provides a semantic mediation framework based on ontologies, rather than defining structural mappings between information models through syntactic mapping mechanisms like XSLT. To avoid N*N mappings between several information models, a common ontology is used: SALUS Core ontology. This ontology has the role of representing the semantics of reference information models, templates, archetypes and the terminology systems used by the source EHR systems, and the target E2B(R2) information model. The SALUS Core ontology aims to act as a common denominator for exchanging clinical data, which is required for proactive post market patient safety studies, between clinical care and research systems, and hence shall be based on the already existing standards used in clinical care and research domains and the already existing data sets [11]. It is aimed to be built through a systematic approach by (i) examining the source and target content models of the selected pilot applications (in ICSR Reporting application these models are HL7 CCD templates and ISO/CEN EN 13606 based templates to represent medical summaries, and the E2B(R2) model), and also the available domain analysis models like BRIDG; (ii) extracting common data elements (CDEs) from these, and harmonizing these CDEs; (iii) representing the related terminology systems as ontologies and linking them with the CDEs in an ontological framework.

SALUS exposes its semantic interoperability platform to reconcile the information model and the terminology systems used to encode patient data in the EHR and the E2B data elements. Especially, the platform includes mappings between (a) standard EHR information model such as HL7 CCD, EN 13606 EHR Extracts, or local proprietary model used and the SALUS Core ontology; (b) E2B information model and SALUS Core ontology; and (c) terminologies used to encode data in EHR and ICSR such as MedDRA, CIM, LOINC and SNOMED-CT.

In summary, tools that would like to query disparate EHR systems for secondary use, like ICSR reporting tool, can communicate with the underlying EHR systems through the data services provided by Technical Interoperability and Semantic Interoperability Layers provided by SALUS Architecture; and, in this way, retrieve the required data sets in the format requested, disregarding to the heterogeneous interfaces and information models used by the local systems.

3.2 SALUS ICSR Reporting Tool: Initial Design

The ICSR reporting tool is designed to ensure (a) all transactions with other SALUS platform components necessary to extract patient data from the EHR to prepopulate the ICSR; (b) all operations necessary to complete the filling of the ICSR in compliance with E2B(R2) specifications and its correct transmission to the pharmacovigilance regulatory authorities. The ICSR reporting tool supports several additional functionalities: recording an ICSR to be completed and reported later; accessing previously sent and waiting to be completed ICSRs; updating and sending an ICSR reported in a previous session; finalizing and sending an ICSR. The ICSR prepopulation process is always triggered following the HP decision, but this can be done in two different circumstances: (i) the HP detects an ADE on the basis of his own expertise and decides to report it; (ii) the ADE notification tool, a complementary SALUS component performing real-time screening of EHR data, detects a potential ADE and displays an alert message to the HP, proposing him to report the case.

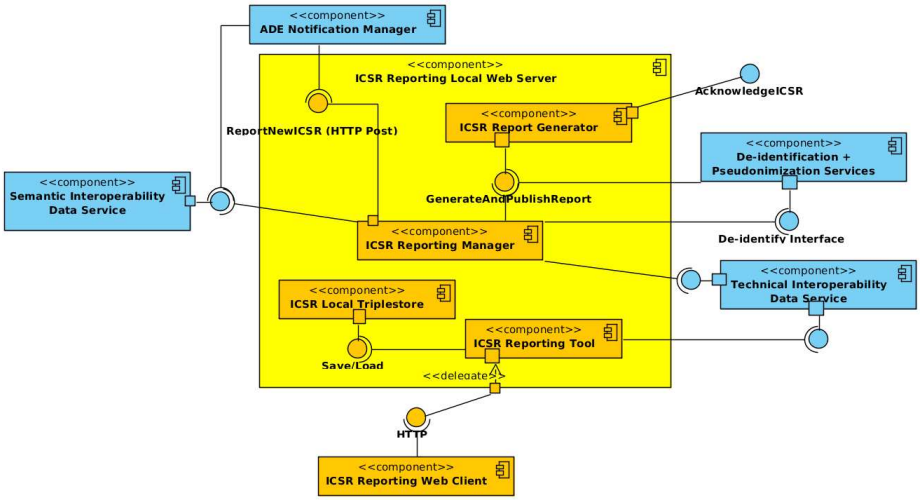

Fig. 1. Components of ICSR reporting tool

An overview of SALUS ICSR reporting system and high-level interactions among its components are shown in **Fig. 1**. *ICSR Reporting Manager* (IRM) is part of the SALUS platform *Semantic Services*, and gets invoked by the *ADE Notification Manager* if new ADE notifications need to be reported. The IRM is initialized with the goal to report detected ADEs to the regulatory body. The *ICSR Reporting Tool* (IRT) is warned by the IRM (through the *Technical Interoperability Data Service*) in order to prepare the ADE report and have the physician extend it with additional information using the *Reporting Web Client*. After the completion of the ADE report, the *ICSR Report Generator* generates the report in the mandated format and sends it to the regulatory bodies and/or pharmacovigilance centre(s). The IRT invokes the *ICSR Local Triplestore* service to save and load both sent and pending ICSR reports. The

IRM also interacts with the components of other systems. It invokes the *Semantic Interoperability Data Service* to retrieve relevant patient data for ICSR reporting from the local EHR in the form of RDF triples represented in SALUS Core ontology. It invokes the *De-identification Service*, which removes or replaces the patient identifiable data which then invokes *Pseudonymization Service* that generates a replacement identifier for the patient ID. It is then the task of the *Pseudonymization Service* to send the data to the *ICSR Report Generator* so that it can generate a pseudonymized report and send it.

4 Discussion and Conclusion

In this paper, we describe SALUS ICSR reporting tool, a new tool still in course of development that supports the reporting of ADEs to regulatory authorities with services (i) enabling automatic prepopulation of ICSRs by extracting patient data from EHR systems and (ii) assisting medical professionals in their manual completion and validation. Several challenges needing to be addressed for successful implementation of this tool have emerged during the design phase. One of the problems is dealing with unstructured EHR data. A first assessment of our pilot-sites (Technical University of Dresden, Germany and Lombardy region, Italy) has shown that only some of the patient data are available in a structured form, the other being only available in free text [12]. Second is dealing with the heterogeneity between E2B data elements and local EHR data. For solving such heterogeneity, the process of defining mappings between E2B data elements and EHR information model resulted into only a partial mapping: some E2B sections are simply not present in the EHR data model (e.g. "Seriousness of the ADE" or "Recurrence of ADE on readministration" have no corresponding section in CCD templates) or value sets only partially overlap. Conversion mechanisms need to be used for collecting these. Another problem is dealing with heterogeneity among EHRs of different pilot-sites. EHR systems use different terminologies to describe patient data, for example LOINC, ICD10 or SNOMED-CT. Since medical data (ADE, reported cause(s) of death in case of patient decease, relevant medical episode in patient medical history, etc.) must be described with MedDRA in the E2B ICSR, mapping those terminologies is necessary to ensure prepopulation. However, most terminologies present different granularity levels, so that mappings can only be approximated. Secondly, since terminologies are evolving, those mappings need to be regularly updated. In SALUS we address this problem by reusing existing mapping sources, such as OMOP CDM Vocabulary and BioPortal, and fine-tuning them. Last but not least, access to the hospital data warehouses storing EHR also poses some ethico-legal difficulties. Patient data must generally be de-identified before being accessed and cannot leave the Clinical Care Zone, i.e. the zone where identified data is maintained and accessed locally. For patient privacy reasons, the ICSR has also to be de-identified and pseudonymized before being sent to regulatory authorities, which is addressed in the SALUS architecture. In some circumstances, this phase can be skipped, but this remains exceptional and depends on national regulatory policies.

The above challenges show that prepopulation can only make part of the job, if data submitted through ICSR must be of quality. It cannot be fully automatic. Most of the time, the physician will need to validate and/or complete what the tool has prepopulated, or select data from a list of propositions. A big challenge for the conception of the tool will consequently be to find equilibrium between blind automation and manual expertise based completion of data.

Acknowledgements. 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 (http://www.salusproject.eu/).

References

1. Hazell, L., Shakir, S.A.: Under-reporting of adverse drug reactions: a systematic review. Drug Saf. 29(5), 385-96 (2006)
2. van der Heijden, P.G., Puijenbroek, E.P., van Buuren, S., van de Hofstede, J.W.: On the assessment of adverse drug reactions from spontaneous reporting systems: The influence of under-reporting on odds ratios. Stat Med 21, 2027-2044 (2002)
3. Bates, D.W., Evans, R.S., Murff, H., Stetson, P.D., Pizziferri, L., Hripcsak, G.: Detecting adverse events using information technology. Journal of the American Medical Informatics Association 10(2), 115-128 (2003)
4. Cullen, D.J., Bates, D.W., Small, S.D., Cooper, J.B., Nemeskal, A.R., Leape, L.L.: The incident reporting system does not detect adverse drug events: a problem for quality improvement. Jt. Comm. J. Qual. Improv. 21, 541-548 (1995)
5. Linder, J. A., Haas, J. S., Iyer, A., Labuzetta, M. A., Ibara, M., Celeste, M., Getty, G., Bates, D. W.: Secondary use of electronic health record data: spontaneous triggered adverse drug event reporting. Pharmacoepidemiol Drug Saf. 19(12), 1211-5 (2010)
6. ICH guideline 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.
7. SALUS Project, Scalable, Standard based Interoperability Framework for Sustainable Proactive Post Market Safety Studies, http://www.salusproject.eu
8. Brajovic, S., Piazza-Hepp, T., Swartz, L., Dal Pan, G.: Quality assessment of spontaneous triggered adverse event reports received by the Food and Drug Administration. Pharmacoepidemiol Drug Saf. 21(6), 565-70 (2012)
9. IHE Quality, Research and Public Health (QRPH) Technical Framework Supplement Drug Safety Content Profile (DSC), Integrating the Healthcare Enterprise (IHE), Trial Implementation Supplement, Aug. 2010.
10. IHE IT Infrastructure Technical Framework Supplement, Retrieve Form for Data Capture (RFD), Trial Implementation, Aug. 19, 2011.
11. Laleci, G.B., Yuksel, M., Dogac, A.: Providing Semantic Interoperability between Clinical Care and Clinical Research Domains. Accepted for publication in IEEE TITB, available from http://www.srdc.com.tr/publications/2012/SALUSSemanticInteroperability.pdf
12. SALUS Deliverable D8.1.1: Pilot Application Scenario and Requirement Specifications of the Pilot Application. May 28, 2012. Available in the Public Document section of http://www.salusproject.eu/