AN EXTENSIBLE SECURITY INFRASTRUCTURE FOR THE SECONDARY USE OF ELECTRONIC HEALTH RECORDS IN CLINICAL RESEARCH

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ELİF ERYILMAZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

JUNE 2013

Approval of the thesis:

AN EXTENSIBLE SECURITY INFRASTRUCTURE FOR THE SECONDARY USE OF ELECTRONIC HEALTH RECORDS IN CLINICAL RESEARCH

submitted by **ELİF ERYILMAZ** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Adnan Yazıcı
Head of Department, **Computer Engineering** _____

Prof. Dr. İsmail Hakkı Toroslu
Supervisor, **Computer Engineering Department, METU** _____

Prof. Dr. Asuman Doğaç
Co-supervisor, **SRDC Ltd.** _____

Examining Committee Members:

Prof. Dr. Özgür Ulusoy
Computer Engineering Department, Bilkent University _____

Prof. Dr. İsmail Hakkı Toroslu
Computer Engineering Department, METU _____

Assoc. Prof. Dr. Ahmet Coşar
Computer Engineering Department, METU _____

Assist. Prof. Dr. Aybar Can Acar
Graduate School of Informatics, METU _____

Assist. Prof. Dr. İsmail Sengör Altıngövde
Computer Engineering Department, METU _____

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: ELİF ERYILMAZ

Signature :

ABSTRACT

AN EXTENSIBLE SECURITY INFRASTRUCTURE FOR THE SECONDARY USE OF ELECTRONIC HEALTH RECORDS IN CLINICAL RESEARCH

Eryılmaz, Elif

M.S., Department of Computer Engineering

Supervisor : Prof. Dr. İsmail Hakkı Toroslu

Co-Supervisor : Prof. Dr. Asuman Doğaç

June 2013, 103 pages

In order to facilitate clinical research studies re-using Electronic Health Records (EHR) has a great potential. Besides interoperability, safeguarding the security and privacy of the medical data in the context of secondary use for clinical research is one of the most important challenges in this respect. In order to ensure that the clinical information is shared among EHR systems and clinical research systems in an ethical and safe way, there needs to be standards-based and adaptable security and privacy mechanisms that can be used by both clinical care and clinical research parties in an interoperable manner by taking into account policies, consent and use agreements of the participating parties.

In this thesis, an extensible security infrastructure has been developed that supports re-use of the EHRs for strengthening the post-approval drug safety studies in the area of clinical research. This work involves the implementation of the security architecture, including novel data protection mechanisms applied to the queried clinical instances as well as additional security services compatible with standard profiles that guarantees the safe use of EHRs for the clinical research studies. In conformance to the selected standards, guidelines, and well-accepted methodologies, this thesis has addressed to find a balance between the privacy concerns for the use of personal data and the requirements of clinical research environments that aim to serve to the public good. In this respect, flexible security architecture is designed and made configurable for the Data Protection Offices of EHR sources according to their preferences.

Keywords: De-identification, Pseudonymization, Anonymization, Secondary Use of EHRs, Clinical instance

ÖZ

ELEKTRONİK SAĞLIK KAYITLARININ KLİNİK ARAŞTIRMALARDA İKİNCİL KULLANIMI İÇİN GENİŞLETİLEBİLİR GÜVENLİK ALTYAPISI

Eryılmaz, Elif

Yüksek Lisans, Bilgisayar Mühendisliği Bölümü

Tez Yöneticisi : Prof. Dr. İsmail Hakkı Toroslu

Ortak Tez Yöneticisi : Prof. Dr. Asuman Doğaç

Haziran 2013 , 103 sayfa

Klinik araştırma çalışmalarını kolaylaştırmada Elektronik Sağlık Kayıtlarının (ESK) yeniden kullanılması büyük bir potansiyele sahiptir. Birlikte çalışabilirlik yanında, klinik araştırmalar için ikincil kullanımda tıbbi verilerin güvenlik ve gizliliğini korumak bu alandaki en önemli sorunlardan biridir. Klinik bilginin ESK sistemleri ve klinik araştırma sistemleri arasında etik ve güvenli bir şekilde paylaşılmasını sağlamak amacıyla, klinik bakım ve klinik araştırma partileri tarafından kullanılabilen standartlara dayalı, birlikte çalışabilen ve uyarlanabilir güvenlik ve gizlilik mekanizmaları olması gereklidir.

Bu tez kapsamında, klinik araştırma alanında pazar sonrası ilaç güvenliği çalışmalarını güçlendirmek amacıyla ESK'ların yeniden kullanımını destekleyen genişletilebilir bir güvenlik altyapısı geliştirilmiştir. Bu çalışma, sorgu sonucu olarak alınan klinik örneklere uygulanan yeni veri koruma mekanizmaları ile birlikte standart profillere uygun ek güvenlik servislerini kapsayarak ESK'ların klinik araştırma çalışmalarında güvenli kullanımını garanti etmektedir.

Anahtar Kelimeler: Belirginsizleştirme, Farklı İsimle Değiştirme, Anonimleştirme, ESK'ların İkincil Kullanımı, Klinik örnek

To my dearest family...

ACKNOWLEDGMENTS

I would like to express my sincere gratitude and appreciation to Prof. Dr. Asuman Doğaç for her encouragement and support throughout this study. I would like to thank my supervisor Prof. Dr. İsmail Hakkı Toroslu for his constant support, guidance and friendship. I would also like to convey thanks to jury members for their valuable comments on this thesis.

I am deeply indebted to my collegues Gökçe Banu Laleci Ertürkmen, Mustafa Yüksel, Ali Anıl Sınacı and all the other colleagues at SRDC Ltd., whose help, stimulating suggestions and encouragement helped me in all the time of research for and writing of this thesis.

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, as a part of the SALUS Project (Scalable, Standard based Interoperability Framework for Sustainable Proactive Post Market Safety Studies). I would also like to thank SALUS project and project partners, especially UMC, LISPA and TUD for their collaboration while identifying the requirements of this study from both clinical and research perspectives.

I am deeply grateful to Yavuz Özyanık for his continued motivating support and welcomed presence. Without his encouragement, I would have never had the strength to complete this work.

I am also grateful to my parents, Hatice and Mustafa Eryılmaz; my sister and her husband, Aslı and Kerem Yaylalı, for their love, belief and continued support.

Although he is not able to read yet, I would like to thank a lot to the little member in the extended family, my cute nephew Çınar, who helped me get away from the stress of this study recently.

Finally, my special thanks go to all my friends that I cannot mention their names separately here for their help, support and cheerful presence through the course of this study. Thanks for giving me a shoulder to lean on whenever I need.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ADE	Adverse Drug Event
AES	Advance Encryption Standard
ARR	Audit Record Repository
ASTM	American Society for Testing and Materials
ATNA	Audit Trail and Node Authentication
CA	Certification Authority
CCD	Continuity of Care Document
CCR	Continuity of Care Record
CDA	Clinical Document Architecture
CDE	Common Data Element
CDM	Common Data Model
DOM	Document Object Model
DWH	Data Warehouse
E2B	Electronic Transmission of Individual Case Safety Reports
EHR	Electronic Health Record
EU	European Union
FP7	Framework Program 7
HIPAA	Health Insurance Portability and Accountability Act
HITSP	Healthcare Information Technology Standards Panel
HL7	Health Level Seven
HTTP	Hypertext Transfer Protocol
ICH	International Conference on Harmonisation
ICSR	Individual Case Safety Report
IDAT	Identifier Data
IHE	Integrating the Healthcare Enterprise
ISO	International Organization for Standardization
IT	Information Technology
MDAT	Medical Data
ML	Machine Learning
NLP	Natural Language Processing
OMOP	Observational Medical Outcomes Partnership
PCC	Patient Care Coordination

PHI	Protected Health Information
PID	Patient Identifier
PSN	Pseudonym
RDF	Resource Description Framework
REST	Representational State Transfer
RIM	Reference Information Model
SHA	Secure Hash Algorithm
SOA	Service Oriented Architecture
TLS	Transport Layer Security
TMF	Telematics Platform for the Medical Research Networks of the Federal Ministry of Education and Research
TTP	Trusted Third Party
TUD	Technical University of Dresden
UDP	User Datagram Protocol
UMC	Uppsala Monitoring Centre
URI	Uniform Resource Identifier
W3C	World Wide Web Consortium
XML	eXtensible Markup Language
XSD	XML Schema Definition

CHAPTER 1

INTRODUCTION

Re-using Electronic Health Records (EHR) through the effective integration and utilization offers several benefits to improve the results of the clinical research studies. On the other hand, sharing the medical data in an ethical and safe way is one of the most critical issues in these studies. While sharing the medical data with research parties to fulfill their requirements to conduct the research, clinical care parties should also keep the rights of individuals by not disclosing any identifiable information. To address this challenge from the security and privacy point of view, there needs to be interoperable, standards-based and extensible security mechanisms that can be used by both clinical care and clinical research parties.

In the literature, there are several efforts [1, 2, 3] describing generic frameworks in order to protect patient related data in research networks. Among these, the Pommerening approach [1] is selected in this thesis as a pioneering work that is a result of a study carried out by the TMF(the Telematics Platform for the Medical Research Networks of the Federal Ministry of Education and Research) of Germany. It provides the basic requirements of ensuring patient privacy in research studies, and five different models for pseudonymization (a particular type of anonymization) of patient data. However, in these scenarios, the purpose is to create a separate data warehouse (DWH) that is set up for the purpose of clinical research studies and the proposed pseudonymization services are used to create such a DWH. However, in this thesis, we claim that without the need of such separate clinical research data warehouses, EHR systems can be involved in clinical research studies, by accepting population based queries from trusted parties and sharing de-identified medical summaries of the eligible patients through secure channels. Therefore, we have designed data protection mechanisms on top of the clinical data instances shared as result sets of population based queries instead of securing all data elements in the DWH of the responsible parties. The mechanisms developed in this scope include de-identification and pseudonymization services providing that clinical information is shared securely within the interoperable parties in an effective way, complying with all necessary legal requirements to protect patient rights. Additional mechanisms such as auditing of events and message level security compliant with the Integrating the Healthcare Enterprise (IHE) standards complement this data level security approach in this security infrastructure.

This thesis work is a small part of the SALUS Project [4] co-financed by the European Commission within the 7th Framework Program (FP7) under grant agreement no ICT-287800. SALUS Project aims to create the necessary infrastructure to enable

secondary use of Electronic Health Records (EHRs) in an efficient and effective way for reinforcing the post market safety studies so that patient safety can be ensured through early detection of rare adverse events. In SALUS Project, functional interoperability profiles have been developed to query population based EHR data from distributed EHR systems for carrying out post market safety studies. As a result of these population based queries, a set of medical summaries of the eligible patients can be shared in standard based medical summary formats, two of which are Health Level Seven (HL7) Clinical Document Architecture (CDA) Release 2 [5] document format using PCC/CCD templates and HL7 CDA Resource Definition Format (RDF) Model templates. From the security point of view, we have developed novel data protection mechanisms to work directly on top of clinical data based on these templates (assuming to retrieve EHRs in a structured form) for the post-market safety studies with additional security services in the scope of this thesis work. As an additional implementation to be used in one of the SALUS pilot application scenarios, we have also applied data protection mechanisms on top of the Individual Case Safety Reports (ICSRs) in ICH E2B format [6].

Within the scope of the SALUS Project, collected medical data sets from EHR Systems are specialized for SALUS Pilot applications to run Adverse Drug Event (ADE) notification, safety analysis methods and Individual Case Safety Report (ICSR) reporting tools. On the research side, each of these applications and methods requires to retrieve medical data sets in different formats. Based on the initial analysis, Temporal Pattern Discovery, Temporal Association Screening and Patient History tools prefer to retrieve data in conformance to Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) [7], while ICSR Reporting tool will produce case safety reports in E2B(R2) [6] specifications along with local models like the ICSR template provided by Italian Medicines Agency (AIFA). By analyzing these models required to process in the pilot applications, the common core data element set has been developed as meaningful fragments to be used for enabling patient safety studies and the development of SALUS harmonized ontology based on this common data element set has been finalized. Having this underlying structure as common data element set at the end, in the scope of this thesis, we have designed data level protection mechanisms to be applied on top of the common core data element set that can be semantically mapped to the any other data elements used in the selected SALUS pilot application scenarios. As a result, each data element that are mapped to the relevant Common Data Elements (CDEs) are taken into account to be anonymized by data level protection mechanisms that constitutes the general infrastructure in this thesis work. In this architecture on the data level, we have designed the de-identification techniques and pseudonymization services on top of CDEs that may have in any other format used in the selected SALUS pilot scenarios and can be applied easily by mapping CDEs to the other formats.

In order to build a generic security infrastructure for this purpose, we have analyzed many different approaches for enabling safety of the medical data in the context of secondary use for clinical research by taking into account available policies and regulations within Europe. First of all, European Union (EU) Directive 95/46/EC [8] that defines the legal ground for the circulation and use of personal data in EU have been followed as a basis of this analysis. The Commission published the "Opinion 4/2007 on the concept of personal data" document [9] to present further clarification

for the definition of personal data and the processing of personal data for clinical research. Despite these clarifications, there are still gray areas, and National Data Protection Supervisory Authorities in the EU countries publish data protection guidelines to set the legal ground for the secondary use of medical data in the context of local regulatory frameworks. Among available standards and guidelines compliant with these European regulations, International Organization for Standardization (ISO/TS 25237:2008) Health Informatics - Pseudonymization [10] is accepted as an underlying standard in our security model presented along this thesis. In addition to this, Health Insurance Portability and Accountability Act (HIPAA) [11] has carefully examined the general rules for uses and disclosures of de-identified protected health information, which we have benefited from in our work. Lastly, Healthcare Information Technology Standards Panel (HITSP) guidelines [12] and IHE IT Infrastructure Healthcare Pseudonymization Handbook [13] are taken into account as additional guidance to design and implement our extensible security architecture.

Within the scope of this thesis, we have developed extensible, standards-based, interoperable security framework that is compliant with legal and ethical requirements analyzed by taking into account policies, consent and use agreements of the participating infrastructures at both national and European level. As a result of this work, we have provided a fully functional open source toolset that enables the developers to self-enhance this security infrastructure for their research purposes according to the rules and regulations valid in their sites. In order to implement such system, existing standards, solutions, and concepts of European e-Infrastructures has been analysed and selected to be used in an adequate way. In conformance to the selected standards, guidelines, and well-accepted methodologies, we have tried to find a balance between the privacy concerns for the use of personal data and the requirements of clinical research environments that aims to serve to the public good. In order to provide flexibility of this security infrastructure, we have designed and implemented it as configurable for both clinical care and clinical research parties involved in research studies that there needs to be convincing results from both sides' point of view. We will provide the design and implementation details of the architecture in the following chapters.

This thesis is organized as follows. Chapter 2 presents the background on enabling standards and technologies study including Legal Framework that constitutes the building blocks for the design of this security infrastructure as well as technologies used to implement such system. This Chapter also includes the summary of the related work on software solutions in the market in comparison with our architecture presented throughout this thesis. The architecture, design and implementation details of the proposed security infrastructure explained in detail in Chapter 3. Finally, Chapter 4 concludes the thesis supported with some discussions and suggests possible future research directions.

CHAPTER 2

BACKGROUND ON LEGAL FRAMEWORK AND ENABLING TECHNOLOGIES

In this chapter, first legal framework and standards are presented guiding mostly throughout the design of our security infrastructure. After that, main enabling technologies and standards that are used to implement this security infrastructure are provided in this chapter.

2.1 Legal Framework and Standards

2.1.1 Definitions

In order to better explain the Legal Framework that is valid among the EU countries, the formal definitions of many terms are provided in this section in reference to EU Directive 95/46/EC [8], ARTICLE 29 Data Protection Working Party [9], ISO/TS 25237:2008 [10] standard and Health Insurance Portability and Accountability Act (HIPAA)[11].

- **Anonymization:** Anonymization is a process to remove the relation between a data set and the data subject and this can be done by:
  - Removing or transforming characteristics that can definitely identify the data subject in the data set, therefore the relation becomes not unique and it can be associated to more than one data subject.
  - Increasing the population in the data subjects set so that the association between the data set and the data subject is not unique.
- **Anonymous data:** Anonymous data is any information relating to a person where the person cannot be identified directly. Therefore, anonymous data is the data that previously referred to an identifiable person, but where that identification is no longer possible.
- **Consent:** The data subject's consent means any given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed.

- **Data Controller:** A data controller is the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of processing personal data.

- **Data Processor:** A data processor is defined as the natural or legal person, public authority, agency or any other body which processes personal data on behalf of the controller.

- **De-identification:** De-identification is defined as a process of removing personal identity revealing attributes and replacing the required identifiers and attributes required for research purposes either with pseudonyms or when possible with more generalized categories (like year of birth instead of exact birth date). It should be noted that in some cases, such de-identification may not serve to the needs of a specific clinical research study, when such generalization makes the data unusable. In such cases, where indirectly identifiable data is needed, specific agreements with data controllers may be needed which often requires patient consents.

- **IDAT or Identification Data:** IDAT means personal data allowing a data subject to be directly identified. In the Pommerening approaches [1], the data sources keep identity data (IDAT) and the medical data (MDAT) including medical history of the patient separately.

- **Personal Data:** Personal data shall mean any information relating to an identified or identifiable person (data subject) who directly or indirectly identified in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.

- **Pseudonymization:** Pseudonymization is a particular type of anonymization that both removes the relation with a data subject and adds a relation between a particular set of characteristics relating to the data subject and one/more pseudonyms. It provides a means for information to be linked to the same person across multiple data records without revealing the identity of the person as a data subject which is often required for clinical research studies.

- **Sensitive Data:** Sensitive data shall mean personal data allowing the disclosure of racial or ethnic origin, religious, philosophical or other beliefs, political opinions, membership of parties, trade unions, associations or organizations of a religious, philosophical, political or trade-unionist character, as well as personal data disclosing health and sex life.

- **(Trusted) Third party:** Third party means any natural or legal person, public authority, agency or any other body other than the data subject, the controller, the processor and the persons who, under the direct authority of the controller or the processor, are authorized to process the data. Trusted Third Party (TTP) on the other hand is a security authority who is responsible for pseudonymization and re-identification of the data subjects. Often the TTP is the only owner of the key of the cryptographic coding algorithms used in pseudonymization. Trusted third parties also act as security authorities assigning unique secondary identifiers to data subjects given the identifying data. This is especially required

where data is collected from multiple research sites and needs to be linked for analysis.

2.1.2 Pommerening Approaches

As mentioned at the beginning in Chapter 1, the Pommerening approaches are analyzed and selected as a pioneering work for the design of our security infrastructure in the scope of this thesis. The Pommerening approaches in pseudonymization are the result of a study that was carried out by the TMF of Germany who supported a project to develop and implement generic models for pseudonymization that can be used in research networks, but in other health care scenarios as well [1]. In this study, the basic requirements are as follows:

- The Data Sources keep identity data (IDAT) and the medical data (MDAT) separately.
- Central data pools must only contain anonymous or at least pseudonymous data.
- A trusted third party ("Datentreuhänder") that is protected by law (e.g. a notary) should carry out the pseudonymization.
- The use of unique patient identifiers across distinct networks is not allowed.

Pommerening and Reng [1] identified five scenarios for secondary use of clinical data and developed five models of pseudonymization for these scenarios:

- Single Data Source, One-Time Secondary Use
- Overlapping Data Sources, One-Time Secondary Use
- One-Time Secondary Use with Re-Identification
- Pseudonymous Research Data Pool
- Central Clinical Data Base, Many Secondary Uses

We will present each of these scenarios in the order of evolution very briefly in this section in order to show our selection as a result of analyzing these approaches in this respect.

Scenario 1 is the simplest one and it is a typical and very intuitive case for anonymization. A simple statistical evaluation of EHR data can be given as an example for this scenario. In Scenario 2, data from diverse sources must be linked together. This is why the pseudonymization is handled by a Trusted Third Party (TTP). Re-identification is not necessary in this scenario, hence one-way pseudonyms are enough. The pseudonymization model for this scenario is presented in Figure 2.1.

In Model 2, medical data (MDAT) is encrypted with the public key of secondary user (i.e. the data target or requestor), hence the Pseudonym (PSN) service as the trusted

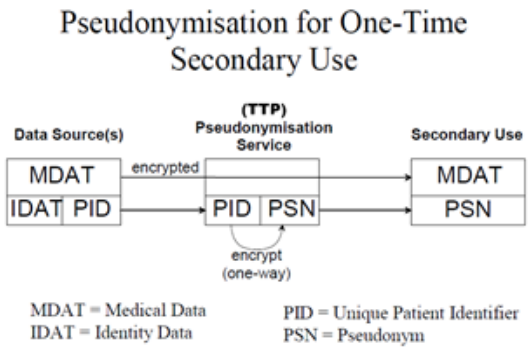


Figure 2.1: Pommerening Approach - Model 2

third party (TTP) cannot read the MDAT, but the secondary user is able to decrypt MDAT with its private key. The PSN is generated by the PSN service by doing a one-way encryption of the Patient Identifier (PID) with a secret key that is only known to the PSN service. The PSN service as the TTP does not store anything except the secret key.

Scenario 3 is similar to Scenario 2, but this time re-identification must be possible. Hence, Model 3 is an extension of Model 2, as shown in Figure 2.2.

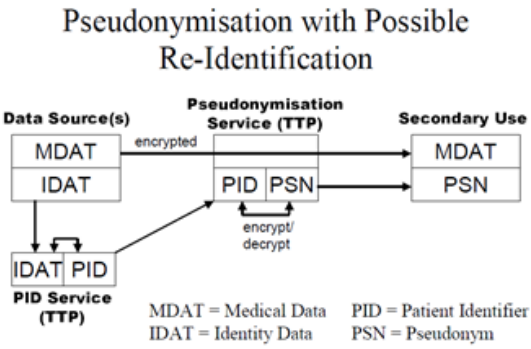


Figure 2.2: Pommerening Approach - Model 3

Model 3 involves a two-step procedure for pseudonymization and several keys and TTP services. First, there is a need for a PID that is not a “public” universal identifier (such as Patient Number, Insurance Number), but is project specific and is generated by a separate TTP service. Therefore, the PID TTP service stores the "patient list", i.e. the association between IDAT and PIDs. The PSN service works as a second TTP service and acts as in Model 2, but this time applies a reversible encryption procedure.

The PSN service does not store the association between PID and PSN, but can restore the PID from the PSN at any time with the help of its secret encryption key. For re-identification, the PID service is also involved in the process; it associates the PID with the identity data and notifies the data source. It should be noted that this model is very similar to the Pseudonymization process proposed by ISO [10].

Scenario 4 is almost identical with Scenario 3. The only difference is that on the secondary use side, the pseudonymized data have to be stored in a data pool with the need of long-term data accumulation. Hence, Model 4 is identical with Model 3; it only allows multiple secondary use. Model 4 is refereed as "Model B" accepted by German TMF. This data pool is available for several different research projects. What projects may get access, depends on the situation, but as a rule the projects must be associated to the specific health care or research network by contracts; i.e. the data pool must not be a self-service database for arbitrary projects.

Scenario 5 is the most complex scenario that better fits the needs of research networks with a "clinical focus". It supports the long-term observation of patients with chronic diseases, and facilitates the individual feedback of research results to the patient or to the responsible physician. Model 5 corresponding to this scenario introduces a central clinical database as a TTP service with online access for the treating clinician who is also responsible for the quality of the data. This central clinical database contains no identity data, but only the PID instead; the reference – in the case of authorised access – is established via the patient list. If a research project needs data from this central clinical database, the appropriate data set is exported after pseudonymization by a PSN service with a project specific key; which means, different projects get different pseudonyms. The main drawback with Model 5 is that it requires implementation of sophisticated communication procedures among many TTPs. This model is presented in Figure 2.3. Model 5 is referred as "Model A" of German TMF.

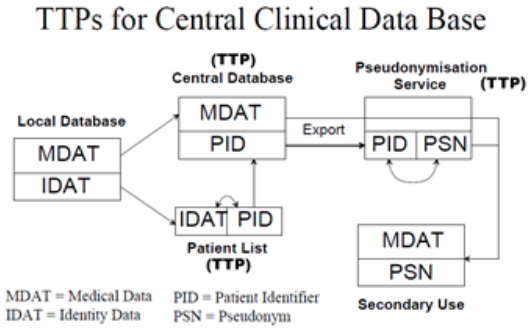


Figure 2.3: Pommerening Approach - Model 5

In our security architecture, we have made use of Scenario 2 for irreversible pseudonymization and Scenario 3 for the reversible pseudonymization. Normally, in reference to the ISO/TS 25237:2008 standard and Pommerening approaches, it is assumed that the data sources keep identity data (IDAT) and medical data (MDAT) separately. However, in our architecture, clinical instances include both IDAT and MDAT together requiring the two separate services as “De-identification” and “Pseudonymization” due to the fact that Pseudonymization service is only allowed to process MDAT (in encrypted format) to assign a specific pseudonym for the patient identifier. Therefore, de-identification process should take a part before pseudonymization in order to de-identify IDAT within the clinical instances. The details about our architecture will be provided in Chapter 3.

2.1.3 Concept of Personal Data Usage

The main purpose of EU Directive 95/46 of 24 October 1995 is to protect the fundamental rights and freedoms of natural persons and in particular their right to privacy, with regard to the processing of personal data. As presented in the previous section, personal data is the "data related to an identified or identifiable individual" and the directive sets the legal ground for the circulation and use of personal data along the following perspectives [8]:

- Fair and lawful processing
- Processing for limited purposes (no further incompatible processing)
- Adequate, relevant and not excessive
- Accurate and up to date
- Preservation no longer than is necessary
- Data subjects' rights (information and access)
- Secured processing (technically and organisationally)
- No transfer to third countries without adequate protection
- Notification to relevant regulator

The Directive also contains specific minimum requirements in terms of the processing of personal health information, which is categorised as a "special category of data" that requires special and additional protection in terms of obtaining, processing, security and disclosure (Article 8 of [8]). As a summary:

- Explicit consent of the data subject is available for data processing; or
- Processing is necessary for the purposes of carrying out the obligations and specific rights of the controller; or
- Processing is necessary to protect the vital interests of the data subject or of another person where the data subject is physically or legally incapable of giving his consent; or
- Processing is carried out in the course of its legitimate activities with appropriate guarantees by a foundation, association or any other non-profit-seeking body and that the data are not disclosed to a third party without the consent of the data subjects; or
- Processing is necessary for preventive medicine, medical diagnosis, treatment or healthcare services, with supervision by a health professional bound by professional secrecy.

The "Opinion 4/2007 on the concept of personal data" published by the Data Protection Working Party set up under Article 29 of Directive 95/46/EC presents further clarification for the definition of personal data and the processing of personal data under certain circumstances including clinical research. The aim is to establish the appropriate balance between protection of the data subject's rights on the one side, and on the other side the legitimate interests of data controllers, third parties and the public interest.

In [9], anonymous data is defined as it is any information about a person where the person cannot be identified by using any kinds of reasonable techniques. In line with this definition, it is clearly presented that, the protection rules cannot be applied to anonymous data related to a data subject no longer identifiable. In this respect, if within a clinical research study, subject data is collected in an anonymous manner inline with the anonymous data definition provided by Opinion 4/2007, it is clear that the data protection rules set in Directive 95/46/EC shall not apply, i.e. explicit consent of data subject is not mandatory in this case.

The Opinion 4/2007 also elaborates on the case of pseudonymization. The definition of pseudonymization process is inline with that of ISO/TS 25237:2008 as presented in the previous section. The definition of retraceable pseudonymisation (reversible pseudonymization) is provided where it is possible to re-identify the subject by using correspondence lists for identities and their pseudonyms or by using two-way cryptography algorithms. It is presented that retraceably pseudonymised data may be considered as information on individuals which are indirectly identifiable, and in this respect, data protection rules apply, yet it is stated in [9] that de-identification rules for such indirect identifiers can be more flexible than the application of these methods on the direct identifiers as they have lower risks than the direct identifiers.

Regarding irreversible pseudonymization where no re-identification is possible, it is stated in [9] that anonymization of data can be achieved by one-way cryptography algorithms. It is presented that in cases where re-identification is not required under any circumstances, applying appropriate technical measures (e.g. cryptographic, irreversible hashing) may not destroy the provisions of the related Directive [8].

As briefly summarized, the Opinion 4/2007 document provides further clarifications, yet there are still gray areas, especially related with the decision of whether a data can be considered as anonymous data and hence can be exempted from the data protection rules. In this respect, in the document the essential role of National Data Protection Supervisory Authorities is emphasized in the framework of their missions of monitoring the application of data protection law, which involves providing interpretation of legal provisions and concrete guidance to controllers and data subjects.

Based on these guidelines, the Data Protection Officers in respective EU countries publish guidelines on how clinical data can be used for research purposes. The alternatives to be pursued in terms of precedence during secondary use of personal medical data are as follows:

- Work on anonymous data,
- If impossible to achieve the scientific purpose with the previous, work on pseudonymized data (key-coded data),

- If impossible to achieve the scientific purpose with the previous, work on non-pseudonymized data (personal data).

As an example, the guidelines that are provided by the Data Protection Commissioner of Ireland [14] are also in line with Article 29 Working Party guidelines. The flowchart presented in Figure 2.4 by the Data Protection Commissioner of Ireland presents the steps to be followed more clearly.

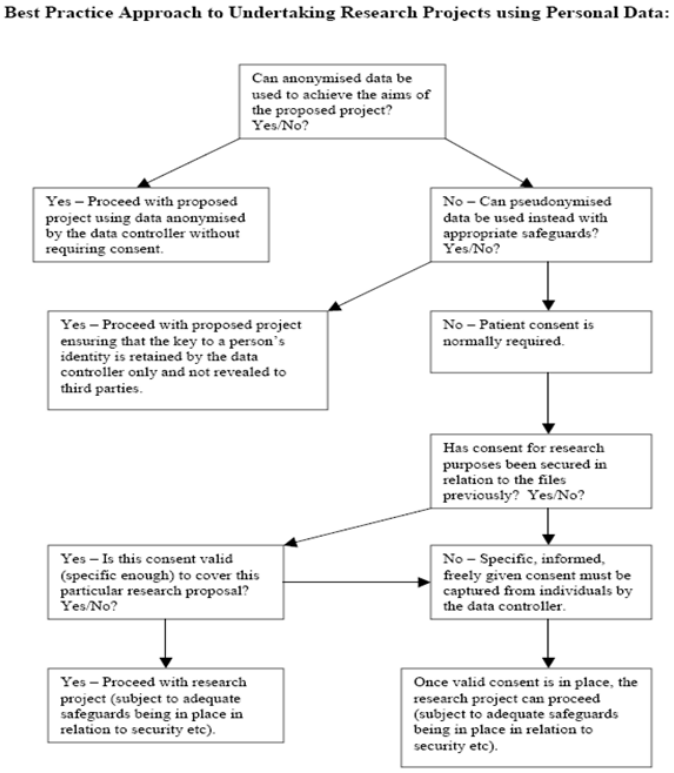


Figure 2.4: Guidelines provided by Data Protection Commissioner of Ireland

In these respect, to design our own security infrastructure, we have analysed many guidelines as regularity guidance with respect to legal framework following the rules and regulations defined by EU directives presented in this section. The details about how we have benefited from these guidelines are provided in Chapter 2.1.5.

2.1.4 ISO/TS 25237:2008 Health Informatics - Pseudonymization

ISO/TS 25237:2008 standard contains principles and requirements for privacy protection using pseudonymization services for the protection of personal health information. Briefly, ISO/TS 25237:2008 [10]:

- defines one basic concept for pseudonymization;

- gives an overview of different use cases for pseudonymization that can be both reversible and irreversible;
- defines one basic methodology for pseudonymization services including organizational as well as technical aspects;
- gives a guide to risk assessment for re-identification;
- specifies a policy framework and minimal requirements for trustworthy practices for the operations of a pseudonymization service;
- specifies a policy framework and minimal requirements for controlled re-identification;
- specifies interfaces for the interoperability of service interfaces.

In Figure 2.5, the basic ISO/TS 25237:2008 pseudonymization workflow is provided [10]. There are 4 parties in the workflow: 2 of them are Data Source and Data Target; the remaining 2 are the trusted third parties (TTP), namely Person Identification Service and the Pseudonymization Service. When Data Target requests data from the Data Source, first it only receives an acknowledgement that its request is being processed. Then, the Data Source sends a request to the Person Identification Service together with the data that the Data Target requests. In this case, it is assumed that the Data Source keeps the identity data (IDAT) and the medical data (MDAT), but not the patient identifier (PID). PID generation based on IDAT is the responsibility of the Person Identification Service. After generation of PID, IDAT is replaced with PID in the original data and this time forwarded to the Pseudonymization Service, which then replaces the PID with the pseudonym (PSN) it generates based on the PID. Finally, the pseudonymized data is sent to the Data Target by the Pseudonymization Service; this part is important, in the final step, it is not the Data Source delivers the pseudonymized data to the Data Target at first hand. For this reason, the Data Source has to pass the endpoint information to the Person Identification Service, which will then forward this information to the Pseudonymization Service.

Furthermore, in application, it is also common practice that the Data Source encrypts the MDAT before passing the data to the Person Identification Service. This way, the actual medical data travels in an encrypted way in the complete process. Although Person Identification Service and the Pseudonymization Service are trusted third parties, they do not need to know the medical details. The former only needs the IDAT, while the latter only needs the PID. Of course, the Data Target has to be able to decrypt the encrypted MDAT; i.e. it has to have the necessary decryption key. As a common practice, MDAT is encrypted by the Data Source with the public key of the Data Target, so that only the Data Target is able to decrypt with its private key.

After analyzing Pommerening approaches and related regularity rules, we have harmonised these results with this ISO/TS 25237:2008 standard workflow in our security architecture.

2.1.5 Regularity Guidance from Guidelines

In reference to selected standards compliant with EU rules and regulations, we have analyzed many guidelines to highlight the gray areas that may not be covered clearly by

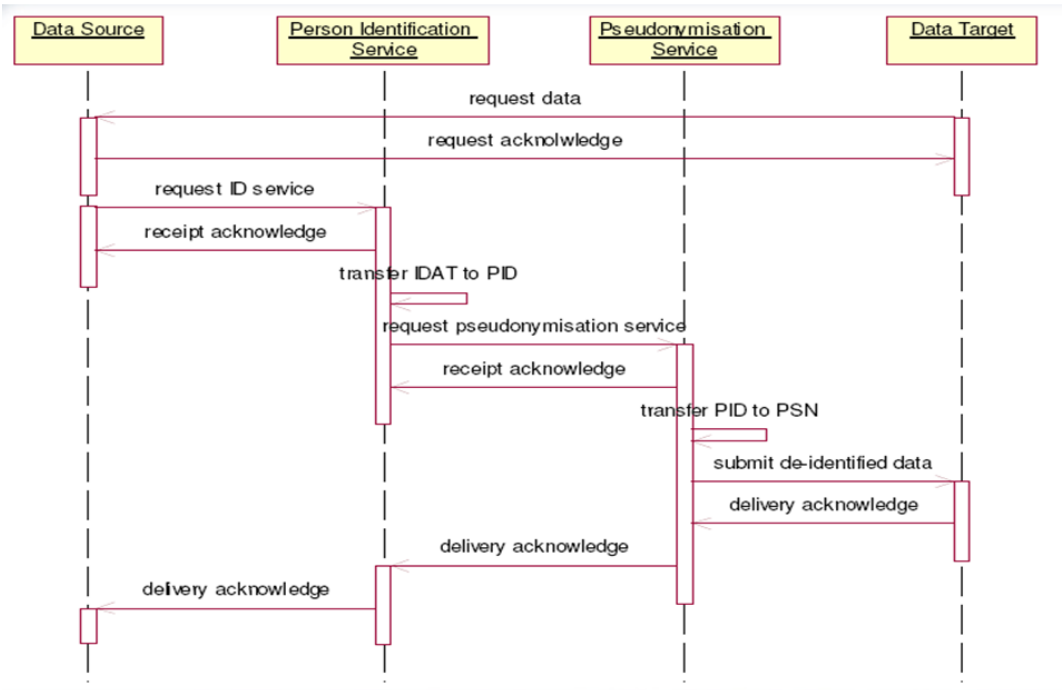


Figure 2.5: ISO/TS 25237:2008 Health Informatics – Pseudonymization Workflow

the laws. In the subsections of this Chapter, we have briefly summarized our findings from this analysis and how we have benefited from each of them while designing and implementing our security infrastructure.

2.1.5.1 Health Insurance Portability and Accountability Act (HIPAA) Regulations

The HIPAA Privacy Rule [11] defines the conditions for the usage of protected health information by third parties for research purposes. In this respect, HIPAA covers the de-identification process requirements at first as follows:

- A person with appropriate knowledge and experience applying generally acceptable statistically and scientific principles and methods for rendering information which is not individually identifiable.
- Removal of Protected Health Information (PHI)

HIPAA "Safe Harbor" De-Identification of Medical Record Information [11] requires that each of the following identifiers in the full data set presented in Table 2.1 of the individual or of relatives, employers, or household members of the individual must be removed from medical record information in order for the records to be considered de-identified.

HIPAA has also defined Limited Data Set referring to PHI (Protected Health Information) [11] that excludes 16 categories of direct identifiers from the below list (3. and

Table2.1: Removal of identifiers from the full data set

1. Names.	10. Account numbers.
2. All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP Code, and their equivalent geographical codes, except for the initial three digits of a ZIP Code if, according to the current publicly available data from the Bureau of the Census: a. The geographic unit formed by combining all ZIP Codes with the same three initial digits contains more than 20,000 people. b. The initial three digits of a ZIP Code for all such geographic units containing 20,000 or fewer people are changed to 000.	11. Certificate/license numbers.
3. All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older.	12. Vehicle identifiers and serial numbers, including license plate numbers.
4. Telephone numbers.	13. Device identifiers and serial numbers.
5. Facsimile numbers.	14. Web universal resource locators (URLs).
6. Electronic mail addresses.	15. Internet protocol (IP) address numbers.
7. Social security numbers.	16. Biometric identifiers, including fingerprints and voiceprints.
8. Medical record numbers.	17. Full-face photographic images and any comparable images.
9. Health plan beneficiary numbers.	18. Any other unique identifying number, characteristic, or code, unless otherwise permitted by the privacy Rule for re-identification.

18. are excluded) and may be used or disclosed, for purposes of research, public health, or health care operations, without obtaining either an individual's Authorization or a waiver or an alteration of Authorization for its use and disclosure, with a data use agreement. These excluded identifiers indicates that they may be used for rare cases and require special processing for de-identification.

Related to the HIPAA rules, ISO/TS 25237:2008 [10] standard defines the following level concepts with respect to the anonymity:

- Level 1 Anonymity: Removal of Clearly Identifying Data
  - A first, intuitive level of anonymity can be achieved by applying rules of thumb that provides a sufficient guarantee.
  - As an example of Level 1 Anonymity, the HIPAA rule is given. The HIPAA rule requires that for data to be considered de-identified, 18 enumerated identifiers presented earlier should be removed.
- Level 2 Anonymity: Static Model Based Re-identification Risk Analysis
  - This level includes a static risk analysis that checks for re-identification vulnerabilities by different interfaces. This level may for example include the removal of absolute time references. A reference time marker "T" is defined as the admission of a patient for an episode of care and other events; discharge is expressed with reference to this time marker.
- Level 3 Anonymity: Routine Resource Risk Analysis
  - An anonymized resource used for data mining must undergo a routine statistical evaluation for re-identification risks associated with the populated resource. Such risk analysis entails assessments of outliers and analytical linking with external information resources.

As indicated Level 1 anonymity, removing all unique identifiers specified by HIPAA does not entirely work to provide reasonable anonymity and requires further levels of anonymity. In the literature, there are many statistical disclosure control efforts addressing this problem such as k-anonymity [15], l- diversity [16] and t-closeness [17] and so on. These statistical approaches analyzed within the scope of this thesis work as a part of the state-of-the-art research will be explained briefly in Chapter 2.1.6.

2.1.5.2 Healthcare Information Technology Standards Panel (HITSP) Guideline

The HITSP Guideline [12] provides specific instructions for anonymizing data that are prepared for repurposing data created as part of routine clinical care delivery. This construct defines a guideline specification that provides the ability to anonymize patient identifiable information for Public Health Case Reporting. However, anonymization cannot be guaranteed by the use of this construct, and therefore a comprehensive risk assessment should be conducted in the implementation environment.

According to HITSP, when releasing data from which direct identifiers have been removed, there is always the potential for re-identification of individual-level records through a combination of variables that individually do not identify the data subject, but which, in combination, create a high risk of re-identification [12]. The most common such variables are:

- dates (e.g. birth, admission and discharge, procedures),
- geolocators (e.g. postal code, spatial data released on maps),
- gender, and
- diagnostic codes, especially when these refer to less common (though not necessarily rare) health conditions
- unusual education (e.g. PhD in statistical disclosure control procedures)
- unusual occupation (e.g. president of a major teaching hospital in Toronto).
- race, ethnicity, religion, and income or other socio-economic indicators

As a result, HITSP guideline addresses the rare cases that previously specified by the HIPAA rules at this point and suggest conducting risk analysis to deal with them. Within the scope of this thesis work, we have worked with Data Protection Officers of EHR data sources to deal with such kinds of variables by presenting our analysis on statistical disclosure control approaches to carry out such risk analysis in this scope. The details are briefly presented in Chapter 2.1.6.

2.1.5.3 IHE IT Infrastructure Healthcare Pseudonymization Handbook (White Paper)

IHE has provided a reasonably complete background on the topic of de-identification allowing it to be successfully specified and implemented compliant with the rules and regulations. This handbook [13] discusses a process, the current state of the art, and gaps with respect to the use of removing individually identifiable information de-identification of healthcare data including both anonymization and pseudonymization.

According to IHE, de-identification is a process that is used to lower privacy risks in the context of secondary-uses of data. In the handbook that IHE provided in this respect, a list of basic design considerations and a process that can be used to define a de-identification profile for specific use-cases are provided for the secondary use of data. As stated in [13], there is no single de-identification procedure fitting universally to the data needs of all environments. As a result, any security architecture addressing to protect patient related data, one should first determine what and whom should be protected from with respect to the applicable local laws, regulations and policies.

With this handbook, IHE has provided base guidance that can be re-used and improved by the other parties dealing with the protection of patient data. In this respect, IHE has specified a list of possible fields in Electronic Health Records (EHRs) and related risks as presented in Table 2.2.

Table2.2: List of Fields and Related Risks

Example Field	**Risk Characteristics**
Medical Record Number (MRN)	Direct identification of a patient in isolation this may be safe, however there are many exceptions both internally and externally. Is this identifier valuable? Can it be substituted? What are this risks of this being disclosed?
Social Security Number (SSN) - USA Realm	Direct identification
First, last, middle, other name fields	Direct identification, esp. when combined with other demographic elements
Mailing or physical address	Narrows set down to a specific location with a small number of people
Relatives (mother's maiden name, next of kin, parents, insurance guarantor)	Narrows set down to a specific location with a small number of people
Free form text fields (chief complaint, nursing observations, triage notes, test interpretation, susceptibility test interpretation, impressions, etc.)	Very difficult if not impossible to adequately scrub
Codified problems, medications, allergies, procedures	Provided that these data are not outliers, the risk of identifying a person is reasonably low
Medical activity ID (Lab assession, encounter)	Direct identification for those with access to a lab ordering system
Medical device identifiers (such as pump bar codes)	Identification of a patient for those with certain access to IT systems.

ISO/TS 25237:2008 standard previously has provided guidelines about how some particular data elements within EHRs can be de-identified. In the light of information presented in Table 2.2, IHE has reorganized this by grouping data types and related approaches as presented in Table 2.3.

Table2.3: Data Types and Possible De-identification Approaches

Data Types	**Approaches**
Person identifying direct identifiers	Should be removed where possible, or aggregated at a threshold specified by the domain or jurisdiction. Where these data need to be retained, risk assessment of unauthorized re-identification and appropriate mitigations to identified risks of the resulting data resource shall be conducted.
Aggregation variables	For statistical purposes, absolute data references should be avoided. Dates of birth are highly identifying. Ages are less identifying but can still pose a threat for linking observational data, therefore it is better to use age groups or age categories. In order to determine safe ranges, re-identification risk analysis should be run, which is outside the scope of this Technical Specification. Admission, discharge dates, etc. can also be aggregated into categories of periods, but events could be expressed relatively to a milestone (e.g. x months after treatment). Location data, if regional codes are too specific, should be aggregated. Where location codes are structured in a hierarchical way, the finer levels can be stripped, e.g. where postal codes or dialing codes contain 20 000 or fewer people, the code may be changed to 0001)
Demographic data are indirect identifiers	Should be removed where possible, or aggregated at a threshold specified by the domain or jurisdiction. Where these data need to be retained, risk assessment of unauthorized re-identification and appropriate mitigations to identified risks of the resulting data resource shall be conducted.

Continued on next page

Table 2.3 – continued from previous page

Outlier variables	Outlier variables should be removed based upon risk assessment. Age could also be identified as outlier variables (if a person is for instance 103 years old)
Persistent data resources claiming pseudonymity	Shall be subject to routine risk analysis for potentially identifying outlier variables. This risk analysis shall be conducted at least annually. The identified risks shall be coupled with a risk mitigation strategy.
Structured data variables	Structured data give some indication of what information can be expected and where it can be expected. It is then up to re-identification risk analysis to make assumptions about what can lead to (unacceptable) identification risks, ranging from simple rules of thumb up to analysis of populated databases and inference deductions. In "free text", as opposed to "structured", automated analysis for privacy purposes with guaranteed outcome is not possible.
Freeform text	Non-parsable data should be removed or special text processing approaches should be applied.
Text/voice data with non-parsable content	As with freeform text, non-parsable data should be removed.
Image data	Some medical data contain identifiable information within the data. Additional risk assessment shall be considered for identifiable characteristics of the image or notations that are part of the image.

Based on the category of the data item as presented in Table 2.3, IHE has specified different de-identification algorithms [13] that can be used in the de-identification process. These can be summarized as:

- Redaction: Removing an atomic data element
- Fuzzing: Adding "noise" to an atomic data element
- Generalization: Making an atomic data element less specific

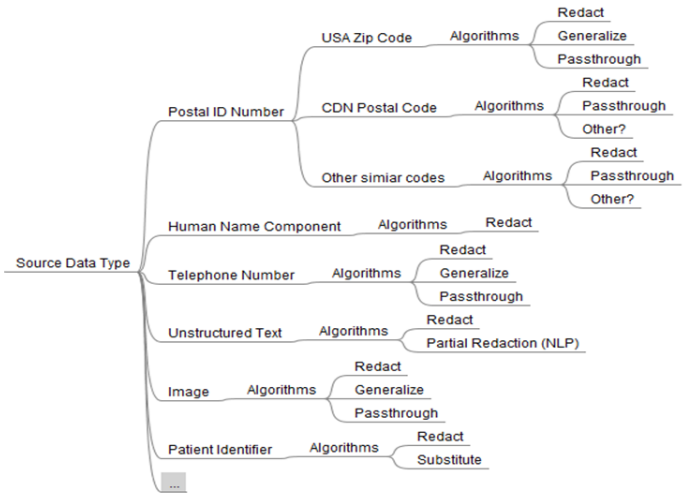


Figure 2.6: Decision tree for mapping de-identification algorithms

- Text Processing: Special considerations for free-format text
- (Recoverable) Substitution: Changing one data element into another data element
- Pass-through: No change

IHE has suggested providing a decision tree for mapping datum to appropriate possible pseudonymization and de-identification processing algorithms as presented in Figure 2.6.

Within the scope of this thesis work, in the design and implementation phase of our security architecture, we have guided mostly from this IHE whitepaper and followed the similar methodologies presented in this section. The details about how we have benefited from this work will be provided in Chapter 3.

2.1.6 Statistical Disclosure Approaches

As mentioned in Chapter 2.1.5.1, to protect individuals' identity when releasing data, data holders often remove or encrypt explicit identifiers specified by HIPAA, such as names and social security numbers. De-identifying data in this way, unfortunately, provide no guarantee for the anonymity. Released information often contains other data, such as race, birth date, sex, and ZIP code that can be linked to publicly available information to re-identify individuals and to infer information that was not intended for release.

In the scope of this thesis work, we have done state-of-the art research on the field of statistical disclosure control mechanisms addressing this problem. In the literature, there are several efforts trying to minimize disclosure risk by the use of the statistical

techniques. One of the preliminary concepts among them in the data protection field is "k-anonymity" [15]. In this study, the solution addressing the linking problem includes a formal protection model. K-anonymity indicates that if the information for each person contained in the dataset cannot be distinguished from at least k-1 individuals whose information also appears in the dataset, it can be referred as "k-anonymous" and provides the privacy protection formally. This protection model is important because it constitutes the basis on for the real-world systems known as Datafly [18], m-Argus [19] and k-Similar [20] in the scope of privacy protection.

In 2007, research study showed that [16], two simple attacks can destroy a k-anonymized dataset having severe privacy problems. The problem was due to the little diversity of sensitive attributes in dataset which results in the discovery of the values of these sensitive attributes. In order to address this problem, another formal privacy approach "l-diversity" is provided by the researchers. In this study [16], the main idea behind l-diversity approach is the requirement of making groups diverse enough with the well-represented values of the sensitive attributes in each group.

After the evolution of l-diversity approach, another study [17] showed that l-diversity has a number of limitations especially in preventing the attribute disclosure. This study proposed "t-closeness" privacy notion requiring that the distribution of a sensitive attribute in any equivalence class is close to the distribution of the attribute in the overall table (i.e., the distance between the two distributions should be no more than a threshold t).

Within the scope of this thesis work, we have mainly analyzed these three main statistical approaches as a starting point. After that, we have analyzed a research survey [21] about recent developments on top of these approaches within the scope of privacy-preserving data publishing field. We have also analyzed many open source tools for statistical disclosure control (ARX Data Anonymization Tool [22], UT Dallas Anonymization Toolbox [23], CAT: The Cornell Anonymization Toolkit [24]) with these existing approaches and algorithms. We have combined those efforts to explain the current risks for the disclosure of patient data to guide the Data Protection Offices of EHR sources in the SALUS Project. We have provided our knowledge from these studies to conduct risk analysis in order to share patient related data without destroying the privacy.

2.1.7 Related Work on Software Solutions

In the market, there are software tools [25, 26] like MeDS [25] aiming to remove all patient identifying information from clinical data documents. However, this tool basically removes the HIPAA-specified patient identifiers and has a lack of flexibility for rare conditions that may result in the identification of individuals indirectly. Therefore, to address this deficiency, we have developed our software flexible in order to configure the de-identification methods and set thresholds for uncommon cases based on the requirements of the respective stakeholders.

On the other hand, there are commercial tools such as CAT [27] and CATS [28] developed by Custodix to remove a large part of the medical data for the exchange purposes in an ethical way, compliant with governing legislation. However, we cannot know the

implementation details of these tools in order to configure them for our purpose in SALUS Project. Therefore, we have implemented our security infrastructure within the scope of this thesis work as open source software that will be publicly available at the end of the SALUS Project. Therefore, it will be further improved based on existing plans and end-user feedback. Additionally, we have used Service Oriented Architecture (SOA) methodologies to provide modularity for further developments easily by the other parties as well.

As a result, the comparison between our security infrastructure and the other software tools available in the market shows that our extensible security infrastructure have addressed the lacks of these available tools and provided additional features in the scope of this thesis work.

2.2 Enabling Technologies and Standards

After mentioning background information about legal framework with the existing standards, guidelines and regulations used in the design phase of our security architecture, in this Chapter, we have provided brief information about enabling technologies and standards used in the implementation phase.

2.2.1 HL7 Clinical Document Architecture

Clinical Document Architecture (CDA), developed by Health Level Seven (HL7) is a document mark-up standard that specifies the structure and semantics of a clinical document for the exchange of clinical data [5]. A valid CDA document is encoded in XML and conforms to the CDA XML Schema Definition (XSD) and it enables the formal representation of clinical statements through CDA entry classes. CDA is derived from HL7's central Reference Information Model (RIM) [29], thereby enabling data reusability - with lab or pharmacy messages, with claims attachments, clinical trials, etc.

The CDA has mainly two parts as a header defining the context of the document and body including the clinical report comprised of sections. A CDA section contains one narrative block and zero-to-many CDA entries. These main components of CDA document to be used to represent clinical statements are shown in Figure 2.7.

2.2.2 HL7/ASTM Continuity of Care Document

HL7/ASTM Continuity of Care Document (CCD) [30] is defined by HL7 and ASTM (American Society for Testing and Materials) International to integrate two complementary healthcare data specifications ASTM's Continuity of Care Record (CCR) [31] and HL7's Clinical Document Architecture (CDA). HL7/ASTM CCD is an implementation guide for sharing CCR patient clinical information summary data including patient demographics, problems, medications and allergies using HL7 CDA. The CCD is an XML-based standard that specifies the structure and encoding of a patient summary clinical document.

Figure 2.7: Main Components of a CDA Document

Normally, CCR also defines an XML schema as a content template for the exchange purposes of electronic medical information without loss of meaning to support continuity of care. However, Continuity of Care Document (CCD), an alternate implementation of the CCR using CDA syntax and format of the proprietary CCR format, is more widely-accepted than CCR's own XML schema. CCD defines a single document template with several section templates and clinical statement templates to be used for this purpose.

2.2.3 IHE Patient Care Coordination Templates

In order to address specific clinical need in support of optimal patient care, Integrating the Healthcare Enterprise (IHE) develops integration profiles that provide precise definitions of how standards can be implemented to meet specific clinical needs. IHE Patient Care Coordination (PCC) template is developed by IHE as an alternative content template to CCD detailing the CCD templates at the document, section and clinical statement levels.

PCC has six main document templates (as well as many section and entry level templates) different from the single CCD document template including Discharge Summary, Medical Document, Medical Summary, PHR Extract, PHR Update and Scanned Document. In short, by developing PCC, IHE defines more detailed templates than in CCD as a specific implementation of established standard to coordinate the optimal patient care.

In SALUS Project, while collecting the medical summaries from underlying EHR Systems, HL7 Clinical Document Architecture Release 2 (CDA) based templates, one of the well-defined EHR interface standards, are chosen. Entry level CDA content

modules (templates) that can carry the information required in the SALUS Pilot application scenarios are modeled through this standard. The specifications of the entry level content modules are based on the IHE Patient Care Coordination (PCC) templates and ASTM/HL7 Continuity of Care Document (CCD) templates. As a result, from the security point of view, we have first applied our data protection mechanisms developed in the security infrastructure on top of the clinical instances in the format of HL7 CDA PCC/CCD templates.

2.2.4 Resource Description Framework

The Semantic Web is a web of data that can be processed directly and indirectly by machines [32]. In order to create Semantic Web resources to be processed in this respect, Resource Description Framework (RDF) [33] developed by the World Wide Web Consortium (W3C) is one of the main languages. RDF is used to represent information on the web by extending the linking structure of the Web.

RDF uses URIs to name the relationship between things as well as the two ends of the link (referred as a "triple") [33]. A triple consists of subject (s), predicate (p) and object (o) scheme, where "p" is a property relationship between "s" which has a value of "o". In the triple structure, "s" and "p" can be represented via URIs; on the other hand, "o" can either be a URI (referring to another resource) or a literal value. This linking structure forms a directed, labeled graph, where the edges represent the named link between two resources, represented by the graph nodes.

SPARQL [34] is a query language for retrieval and manipulation RDF data. SPARQL can be used to express queries across diverse RDF data sources. The results of SPARQL queries can be results sets or RDF graphs. A simple SPARQL query to return ID of all patients in SALUS clinical instance in RDF format is presented in Figure 2.8.

```
SELECT ?ID
WHERE {
?pt a salus:Patient.
?pt salus:ID ?ID.
}
```

Figure 2.8: SPARQL query to retrieve Patient ID

In SALUS Project, besides the HL7 CDA PCC/CCD templates, clinical instances can be retrieved from EHR sources in the RDF format. Therefore, another application area of the data level protection mechanisms is on top of the RDF instances as well.

2.2.5 Electronic Transmission of Individual Case Safety Reports

The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) Expert Working Group has developed

guideline for implementing ICH requirements for the electronic transmission of Individual Case Safety Reports (ICSRs) describing adverse event(s) / reaction(s) experienced by an individual patient according to the ICH E2B (R3) message standard [6]. The ICH E2B (R3) message standard is built upon the Health Level 7 (HL7) ICSR Release 3 standard which is a particular message based on the HL7 Version 3 messaging standard for health care information transfer. This implementation guideline for the E2B (R3) message provides technical recommendations on the way to encode the E2B information model using the XML Schema of the ICH ICSR message.

In SALUS Project, as mentioned in the previous chapters, we have retrieved clinical instances either in HL7 CDA PCC/CCD templates or RDF format. Besides the clinical instances, we have retrieved ICSRs from the EHR sources to be used in one of the SALUS Pilot application scenarios as well. Therefore, ICSRs in ICH E2B (R3) are also taken into account for the additional application of the data level protection mechanisms within the scope of this thesis.

2.2.6 IHE Audit Trail and Node Authentication Profile

IHE has established the Audit Trail and Node Authentication (ATNA) Integration Profile for the security purposes in order to provide patient information confidentiality, data integrity and user accountability [35]. This profile provides access control with the limitation of both on network access level and node level.

On the technical front, ATNA requires the use of TLS to cater for the transport layer integrity, confidentiality, and (service) authentication for both the server and the client. The accountability requirement is supported by the introduction of an Audit Record Repository that is a central repository for log messages. For this repository the ATNA profile proposes the use of TLS transported SYSLOG messages (RFC 5425) [36].

Within the scope of the SALUS Project and as a part of the additional security services in this security infrastructure, open source implementation of an Audit Record Repository (ARR) OpenATNA [37] supporting RFC 3881 audit messages [38] over BSD Syslog as well as RFC 5424-5426 (UDP and TLS) is used. Besides the open source ARR, each SALUS component created their own digital certificates to communicate securely between each other.

CHAPTER 3

THE SECURITY INFRASTRUCTURE

In order to explain the security infrastructure that we have developed within the scope of this thesis work in a more systematic way, we have separated the data protection mechanisms including De-identification and Pseudonymization services, where we deal with the privacy of clinical data represented at the data level, from the additional security mechanisms supporting this approach such as auditing of events and message level security compliant with the IHE standards.

In the composition diagram as presented in Figure 3.1, overall security services developed as a part of the SALUS project with these two separate parts are shown.

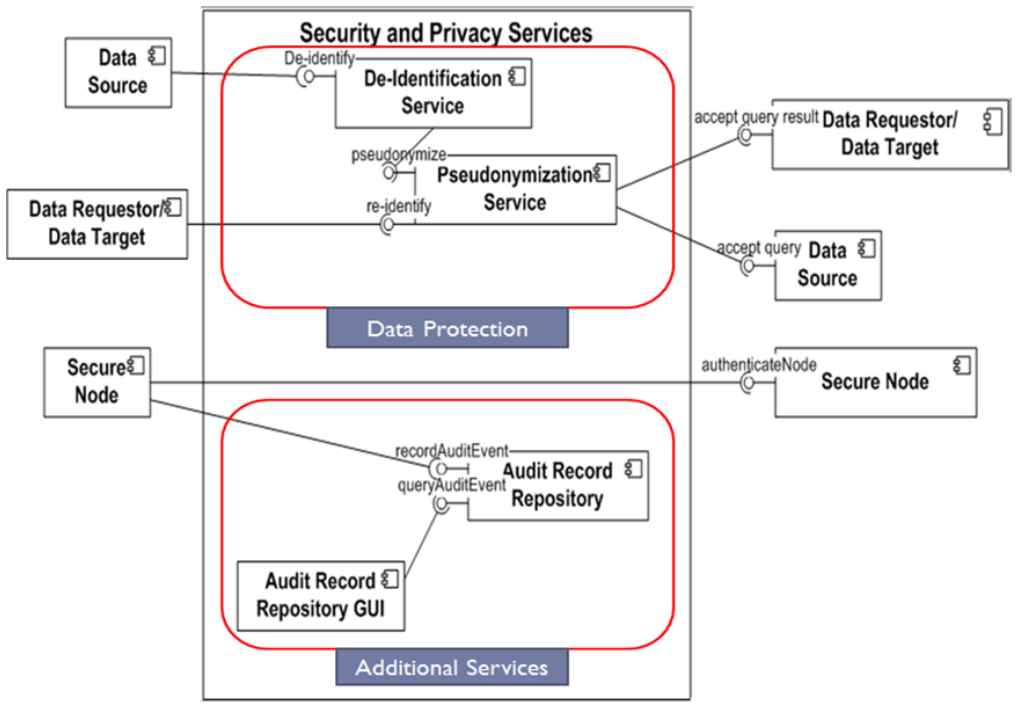


Figure 3.1: Composition diagram for the Security and Privacy Services in SALUS

As pointed out in Figure 3.1, regarding the data level protection, De-Identification Service is called at the Data Source side, after collecting the queried clinical data instances

from the underlying Data Source throughout the SALUS interoperability services. After de-identifying the clinical data set based on the user specific configurations, this data set is passed to the Pseudonymization Service for the selected data items (patient identifiers) to be pseudonymized.

Regarding the additional mechanisms to support the data level protection, our security architecture includes the implementation of IHE Audit Trail and Node Authentication (ATNA) Profile for the secure exchange of healthcare information and the auditing of events related to the access, production or modification of healthcare information. These mechanisms ensure the secure exchange of clinical data at the message level and provide the audit records in conformance to existing interoperability standards.

In Chapter 3.1, firstly, we have provided the details of data protection mechanisms developed in our security infrastructure. Then, the details about the additional security mechanisms can be found in Chapter 3.2.

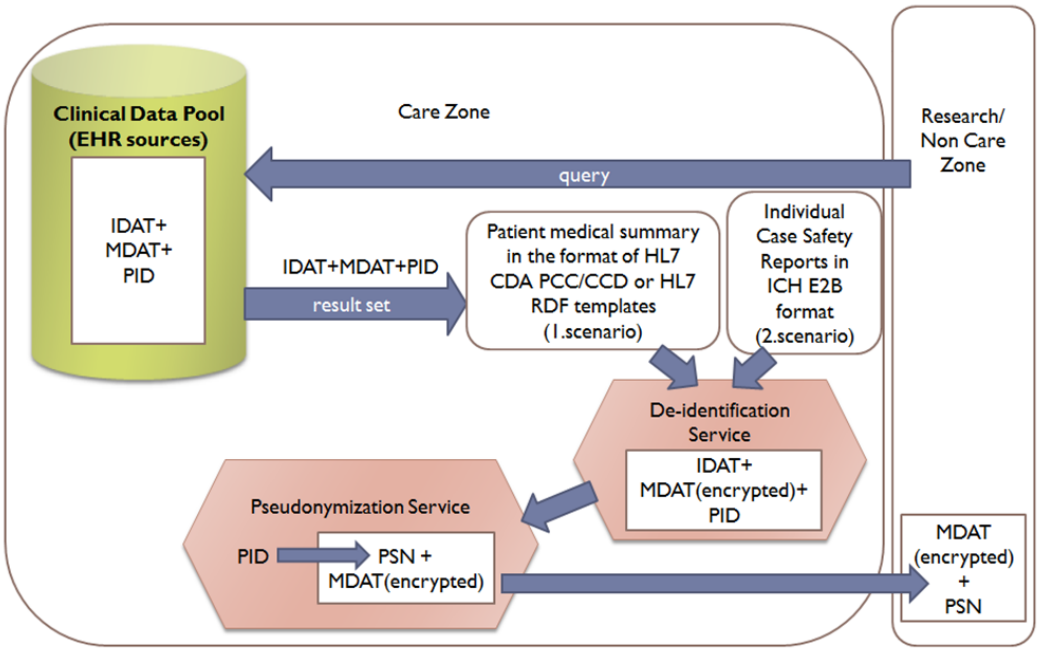
3.1 Data Protection Mechanisms

In the analysis of data protection architectures for each SALUS scenario, the notion of "zones" is adapted. We have defined three zones: Clinical Care Zone, responsible for the care of the patient where identified data is maintained and accessed locally; Non-Care Zone, includes the Pharmacovigilance Centers collecting Individual Case Safety Reports (ICSRs) and doing safety analysis in the context of SALUS pilot scenarios; and Clinical Research Zone, which includes Pharmacovigilance Centers and Pharmaceutical companies carrying out post market safety studies with EHR data. The overall SALUS data level security and privacy infrastructure in the context of Care and Research/Non-care Zones is depicted in the Figure 3.2.

As presented in Figure 3.2, when the data is queried to retrieve the data from the Research/Non-care Zone, the result set is passed to the De-identification Service including patient Medical Data (MDAT), the patient identifier (PID) as well as Identifier Data (IDAT) in the Clinical Zone. Our approach is now starting to differ from the Pommerening approach as it assumes that MDAT and IDAT are separated in clinical instances. However, in SALUS, when we have queried the clinical data pool, we are able to retrieve the IDAT of the related patient. Therefore, we have separated De-identification and Pseudonymization Services to deal with IDAT first and then passed the MDAT in an encrypted form to the Pseudonymization Service as a second step.

As pointed out in Figure 3.2, in SALUS pilot scenarios, the result set of query which needs to be passed to the security infrastructure including De-identification and Pseudonymization Services can be in the format of HL7 CDA PCC/CCD or HL7 CDA RDF template for patient medical summary; or ICH E2B format for the Individual Case Safety Reporting (ICSR). We have applied our data level protection mechanisms on all of them to be safely passed them to the Research zone. We will provide the implementation differences for those types in the De-identification and Pseudonymization Services sub-sections.

When the De-identification Service receives this medical summary or individual case safety report, it processes the medical data (MDAT) and Identifier Data (IDAT) to



IDAT: Identifier Data, MDAT: Medical Data, PID: Patient Identifier, PSN: Pseudonym

Figure 3.2: The Security and Privacy Infrastructure on Data Level

totally de-identify it to be sent to the Research/Non-care Zone. Then, de-identified MDAT in the encrypted format and PID are transferred to the Pseudonymization Service to replace PID with a pseudonym (PSN). Finally, the pseudonymized data is sent to the Research/Non-care Zone to be processed for the data analysis based on the SALUS predefined scenarios in a secure way.

3.1.1 De-identification Service

De-identification is a process of removing personal identity revealing attributes and replacing the required identifiers and attributes for research purposes either with pseudonyms or when possible with more generalized categories (like year of birth instead of exact birth date).

Unfortunately, there is no single de-identification procedure that will meet the diverse needs of all the medical uses while ensuring privacy of patient data. Therefore, we have developed flexible de-identification architecture. De-identification methods are developed in a modular way, which can be extended by implementing new methods when necessary. Also configuration architecture is developed so that for each element a different de-identification method can be chosen by collaborating with the respective stakeholders after analyzing and assessing the risks for each data element set that needs to be exchanged in a de-identified manner. As a result, it is possible to set different methods and contextualized thresholds for the specific cases, which are not very common and may result with identification of individuals when commonly suggested

de-identification methods are applied.

To be compliant with the existing methodologies at the European level, we have used guidelines as mentioned in Chapter 2.1.5 about how some particular data elements within Electronic Health Records (EHRs) can be de-identified. Before the implementation of De-Identification Service for clinical content based on the algorithms described by IHE as presented in Chapter 2.1.5.3, we have first conducted an analysis for each SALUS content model, which specifies the data elements to be exchanged within the scope of each SALUS use case. After this phase, for each data element in the content model, the table presented in Figure 3.3 (small part of it) is filled, indicating the type of the transformation algorithm (redaction, generalization, pass through etc.) that can be applied to each data element as a part of de-identification.

Transformation	Algorithm	Demographics	Patient Name or Initials	ID	Date of Birth	Gender	Race	Birth Place (Region or City)	Patient registration date	Patient de-registration date	Specialist Reco... Numbe...
Redaction	Delete Value		X	X				X			X
Substitution	Fixed Length Substitute Meaningless Value		X	X				X			X
	Original Length Substitute Meaningless Value		X	X				X			X
Recoverable Substitution	Pseudorandom Values			X							X
Fuzzing	Numeric (statistical algorithm)										
	Code Set (Random)										
Generalization	DOB to age				X						
	Shift by randomoffsets				X				X	X	
	Remove day/month/year				X				X	X	
	Geographical locations							X			
Pass-through	N/A					X	X				
Text Processing	N/A										

Figure 3.3: Possible De-identification algorithms for the common data elements

To explain the details of each de-identification method clearly, we have prepared examples for each of them as presented in Table 3.1.

After explaining de-identification methods with these explanations, we have proposed the following methods for each data element in the common data set of SALUS to be agreed on with the Data Protection Offices of EHR sources in the project. We have first grouped the similar data elements to be de-identified in a similar way with the possible de-identification methods mentioned in Table 3.1. There are five groups to apply the related techniques for each of them in general:

- **Category 1: Identifier (ID) values**
  - Related Data Elements: Demographics (ID, Specialist Record Number, GP medical record number, Hospital record number, Provider ID, Provider Organisation ID, Investigation number), Healthcare Provider (Provider ID), Organisation (Organisation ID)
  - Possible De-identification Techniques: Redaction (Delete Value), Substitution (Fixed Length Substitute Meaningless Value, Original Length Substitute Meaningless Value), Recoverable Substitution (Pseudorandom Values)

Table3.1: De-identification Methods with Examples and Explanations

Techniques	Variations	Before De-identificaton	After De-identification	Explanations
Redaction	Delete Value	<patientName> Elif Eryilmaz </patientName>	<patientName></patientName>	Delete only the value of the attribute from the instance
Substitution	Fixed Length Substitute Meaningless Value	<patientName> Elif Eryilmaz </patientName>	<patientName>XXXX </patientName>	Replace value with another value (meaningful or meaningless) having the fixed length defined by user
Recoverable Substitution	Pseudorandom Values	<telecom> 03122101763 </telecom>	<telecom> 26375757903 </telecom>	Substitute pseudorandom value with the choice of length (the length is chosen as 8 in the example)
Fuzzing	Numeric (statistical algorithm)	<telecom> 00390239112233 </telecom>	<telecom> 83785349589271 </telecom>	Replace the value with the one produced as a result of statistical algorithm
Generalization	Remove day/month /year	<effectiveTime>20070806</effectiveTime>	<effectiveTime>200708</effectiveTime> <effectiveTime>200706</effectiveTime> <effectiveTime>0806</effectiveTime>	Day value is removed from the date Month value is removed from the date Year value is removed from the date
	Shift by random offsets	<effectiveTime>20070806</effectiveTime>	<effectiveTime>20071204</effectiveTime>	The date value is shifted by some amount (120 days in example)

- **Category 2: Name and Provider values**
  - Related Data Elements: Demographics (Patient Name or Initials), Past Medical History (Treating Provider), Active Problems/Symptoms (Treating Provider), Lab Results (Result Provider), Procedures (Procedure Provider), Medications (Prescribing Provider), Encounters (Encounter Performer (Provider)), Vital Signs (Result Provider), Data Reporter (Reporter title, Reporter given name, Reporter family name, Reporter organization), Healthcare Provider (Organisation), Organisation (Organisation Name)
  - Possible De-identification Techniques: Redaction (Delete Value), Substitution (Fixed Length Substitute Meaningless Value, Original Length Substitute Meaningless Value)
- **Category 3: Date values**
  - Related Data Elements: Demographics (Date of Birth*, Patient registration date, Patient de-registration date), Pregnancy (Delivery Date, Last Menstrual Period Date), Past Medical History (Start Date, End Date, Date of Entry), Active Problems/Symptoms (Start Date, End Date, Date of Entry), Allergies/Intolerance (Start Date and Time, Date of end of reaction/event, Date of Entry), Family History (Age at Onset*), Procedures (Procedure Date), Medications (Start Date, End Date, Order Date, Date of Entry), Encounters (Start Date, End Date), Vital Signs (Result Date), Social History (Social history dates), Death (Date of Death), Lab Results (Result Date)
  - Possible De-identification Techniques: Substitution (Substitute meaningful value) Generalization (DOB to age, Shift by random offsets, Remove day/month/year)
  - Data elements: Date of birth* (Demographics), Age at onset* (Family history), Additional "DOB to age" option
- **Category 4: Address and Location values**
  - Related Data Elements: Demographics (Birth Place (Region or City), Address), Encounters (Care Provider Location (Organisation)), Data Reporter (Reporter address), Organisation (Organisation Address)
  - Possible De-identification Techniques: Redaction (Delete Value), Substitution (Fixed Length Substitute Meaningless Value, Original Length Substitute Meaningless Value), Generalization (Geographical locations)
- **Category 5: Free-text values**
  - Related Data Elements: Past Medical History (Comments/text describing Problem), Active Problems/Symptoms (Comments / text describing Problem), Procedures (Comments / text describing Procedure)
  - Possible De-identification Techniques: Redaction(Delete value), Text processing (not in the scope of our work)

Besides the data elements to be de-identified based on these techniques compliant with the IHE Pseudonymization Guideline, 78 data elements over the total data set

Demographics	Family History	Encounters
Gender	Kinship Type	Encounter Type
Race	Family History Observation name	Reason for Encounter
Ethnicity	Family History Observation code	**Vital Signs**
Pregnancy	**Lab Results**	Result Type (text)
YES/NO	Result Type (text)	Result Type (Coded)
Pregnancy Status	Result Type (Coded)	Result Value (Numeric)
Past Medical History	Result Value (Numeric)	Result Value (Coded)
Problem Name	Result Value (Coded)	Result Value (String)
Problem code	Result Value (String)	Unit
Problem Status	Unit	Result Reference range
Related Encounter	Result Reference range	Result Interpretation
Severity	Result Interpretation	Related Encounter
Active Problems/Symptoms	Related Encounter	Related Condition
Problem Name	Related Condition	**Social History**
Problem code	**Medications**	Social history type
Problem Status	Product Name	Social history status
Related Encounter	Product Code	**Death**
Severity	Brand name	Cause(s) of death
Allergies/Intolerance	Brand code	Was autopsy done?
Adverse Event Type (text)	Active Ingredient Name	Autopsy-determined cause(s) of death
Adverse Event Type (coded)	Active Ingredient Code	**Healthcare Provider**
Product Code	Product Form (of administration)	Provider type
Product Name	Dose	**Organisation**
Reaction	Frequency	Organisation type
Severity	Route (Coded)	
Outcome of reaction/event at the time of last observation	Duration	
Procedures	Indication	
Procedure Type (text)	Refills	
Procedure Type (coded)	Quantity	
Body Site	Days Supply	
Procedure Status	Related Encounter	
Indication	Fulfillment Instructions	
Related Encounter	Stop reason	

Figure 3.4: Data elements to be pass-through without any de-identification

are remained to be passed through without applying any de-identification technique as presented in Figure 3.4.

In reference to HIPAA rules and HITSP guideline, there is a potential risk for re-identification of individual-level records through a combination of these pass-through variables. These variables individually do not identify the data subject, but which, in combination, create a high risk of re-identification. In this respect, we have carried out state-of-the-art statistical analysis as presented in detail in Chapter 2.1.6 to highlight the potential risks for such cases to our end users. At the end of these discussions, our end users have identified the rare cases, for example they have identified "rare diseases" in a list, and then this list is used by the de-identification service to generalize these conditions so they cannot be used to identify the patient in such uncommon conditions. As a result, our security infrastructure have made it possible to specify thresholds for the uncommon cases based on the risk analysis conducted with EHR Data Protection Officers of EHR data sources in the scope of this work.

Within the scope of pilot application scenarios, clinical data instances will be passed to Research Zone in the format of HL7 CDA PCC/CCD or HL7 CDA RDF templates. Besides these two types of templates, in another scenario (Scenario 2 in Figure 3.2), Individual Case Safety Report (ICSR) in ICH E2B format needs to be passed to the research zone requiring the de-identification and pseudonymization. As mentioned before, we have made an analysis with the possible de-identification techniques on top of the Common Data Elements (CDE) that can be mapped to the other data elements in different document templates. After deciding each CDE and related method mentioned above, in the implementation phase, we have configured our security infrastructure to

map these CDEs to the elements in HL7 CDA PCC/CCD, HL7 CDA RDF and ICH E2B formats.

This approach provides that when these common data elements are mapped with the configuration to CDA, RDF or E2B data elements in the clinical data instances, the data level security mechanisms will remain as they are. Therefore, these mechanisms are made independent from the underlying format of data instances requiring only change in the configuration. In other words, we have made data level protection mechanisms on these types of medical instances possible with the same de-identification methods by only arranging the configuration files accordingly.

After arranging the basis of the infrastructure as configurable, we have used the concept of Service-Oriented Architecture (SOA) with RESTful implementations [39] in Java on top of HTTP to implement de-identification algorithms that enables extensible development within or across the current infrastructure. We have defined our de-identification functions to get specific data element as a source parameter and returning de-identified version of this element as a target. Then, we have processed these source strings according to the defined methodologies identified with the guidance of de-identification methods suggested in this section. We have defined different services for different types of the document formats using the same de-identification techniques but differing the parsing of the documents with configurations.

3.1.1.1 Implementation Details for HL7 CDA RDF templates

Before working on the HL7 CDA RDF templates, we have provided data elements in RDF that needs to be de-identified to our end users and they agreed on possible de-identification mechanisms proposed as presented in Appendix A.

After agreeing on de-identification methods, we have designed this part of our security infrastructure configurable to be updated later easily according to the preferences of our end users. In this respect, we have created configuration file in sample XML format including each related data elements to be de-identified, SPARQL query to retrieve and manipulate this data elements stored in the clinical instance and de-identification technique to be applied on it. Sample configuration entry for the "Patient ID" data element ("extension" attribute) is presented in Figure 3.5.

```
<!--Patient.ID.Extension.String -->
<Patient.ID.Extension.String>
    <sparql>SELECT ?ID ?extension WHERE {?pt a salus:Patient.?pt salus:ID ?ID.?ID salus:extension ?extension.}</sparql>
    <method>Substitute Pseudonym</method>
</Patient.ID.Extension.String>
```

Figure 3.5: Sample configuration entry for "Patient ID" needs to be de-identified

The full configuration file to de-identify all required data elements in the HL7 CDA RDF instances can be found in Appendix C.

When we have successfully parsed this configuration in Java with XML DOM Parser

[40] to retrieve data element, related sparql query and de-identification method, we have used Apache Jena API [41] to process sparql queries on top of the clinical instance in the RDF format in order to retrieve the data element to be de-identified. After retrieving this data element from the clinical instance as a result of sparql query, we passed this data element to De-identification Service to de-identify it with respect to the de-identification method configured in the configuration file. When we retrieve the de-identified data element from the De-identification Service, finally, we have successfully updated the clinical instance RDF model to store de-identified instance in another file.

In Figure 3.6 and Figure 3.7, sample parts of the RDF instances before and after the de-identification are presented.

Figure 3.6: An extract from RDF before de-identification

As seen in these figures above, we have applied a number of the de-identification methods (substitute pseudorandom value, shift date values by random offset, generalize geographical location, substitute meaningless value) to the clinical instance in RDF format.

3.1.1.2 Implementation Details for HL7 CDA PCC/CCD templates

Regarding the work on top of the HL7 CDA PCC/CCD clinical instances, we have used the same data set for RDF de-identification as presented in Appendix A. The

```
[]    rdf:type salus:patient ;
      salus:ID
              [ rdf:type salus:ii ;
                salus:extension "17689263734" ;
                salus:root "2.16.840.1.113883.2.9.4.3.2"
              ] ;
      salus:address
              [ rdf:type salus:addr ;
                salus:city "" ;
                salus:country "Italy" ;
                salus:postalCode "12345" ;
                salus:state "MI" ;
                salus:streetAddressLine
                        ""
              ] ;
      salus:allergy
              [ rdf:type salus:Allergy ;
                salus:adverseEventDate
                        [ rdf:type salus:ivlTs ;
                          salus:high "2010-05-01T00:00:00"^^xsd:dateTime ;
                          salus:low "2009-08-24T00:00:00"^^xsd:dateTime
                        ] ;
                salus:adverseEventType
                        [ rdf:type salus:cd ;
                          salus:code "235719002" ;
                          salus:codeSystem "2.16.840.1.113883.6.96" ;
                          salus:codeSystemName
                                  "SNOMED CT" ;
                          salus:displayName "Food intolerance"
                        ] ;
                salus:product
                        [ rdf:type salus:cd ;
                          salus:code "303300008" ;
                          salus:codeSystem "2.16.840.1.113883.6.96" ;
                          salus:codeSystemName
                                  "SNOMED CT" ;
                          salus:displayName "Egg protein"
                        ] ;
```

Figure 3.7: An extract from RDF after de-identification

only difference is we have changed sparql queries where we have used to find the data element in the clinical instance with the XPATH expressions [42] within the specified CDA sections to navigate through elements and attributes in the HL7 CDA PCC/CCD instances.

As we did for RDF instances, we have also created configuration file in sample XML format including each related data elements to be de-identified, XPATH expression to retrieve and manipulate this data elements stored in the clinical instance and de-identification technique to be applied on it. Sample configuration entry for the "Patient ID" data element ("extension" attribute) is presented in Figure 3.8.

The full configuration file to de-identify all required data elements in the HL7 CDA PCC/CCD instances can be found in Appendix D.

```
<!--Patient.ID.Extension.String -->
<Patient.ID.Extension.String>
    <xpath>/cda:ClinicalDocument/cda:recordTarget/cda:patientRole/cda:id/cda:extension/</xpath>
    <method>Substitute Pseudonym</method>
</Patient.ID.Extension.String>
```

Figure 3.8: Sample configuration entry for "Patient ID" needs to be de-identified

As we did for RDF instances, we have successfully parsed this configuration in Java to retrieve data element, related xpath expression and de-identification method. We have used javax.xml.xpath package [43] to process XPATH expressions on top of the clinical instance in the CDA format in order to retrieve the data element to be de-identified. After retrieving this data element from the clinical instance as a result of XPATH expression, we passed this data element to De-identification Service to de-identify it with respect to the de-identification method configured in the configuration file. When we retrieve the de-identified data element from the De-identification Service, finally, we have successfully updated the clinical instance and able to store de-identified instance in another file.

In Figure 3.9 and Figure 3.10, sample parts of the CDA instances before and after the de-identification are presented.

```
<recordTarget>
    <patientRole>
        <id extension="54321678901" root="2.16.840.1.113883.2.9.4.3.2" assigningAuthorityName="MEF - Italian Citizen Identity Authority"/>
        <id extension="54321678902" root="2.16.840.1.113883.2.9.4.3.2.1" assigningAuthorityName="Hospital Identity Authority"/>
        <id extension="54321678903" root="2.16.840.1.113883.2.9.4.3.2.2" assigningAuthorityName="GP Identity Authority"/>
        <id extension="54321678904" root="2.16.840.1.113883.2.9.4.3.2.3" assigningAuthorityName="Specialist Identity Authority"/>
        <addr use="HP">
            <streetAddressLine>Via Lago di Nemi, No 4</streetAddressLine>
            <city>Milan</city>
            <state>MI</state>
            <country>Italy</country>
            <postalCode>20142</postalCode>
        </addr>
        <telecom use="HP" value="tel:+390239112233"/>
        <patient>
            <name>
                <given>Sabina</given>
                <family>Cremona</family>
            </name>
            <birthTime value="19560708"/>
            <birthplace>
                <place>
                    <addr>
                        <city>Biella</city>
                        <country>Italy</country>
                    </addr>
                </place>
            </birthplace>
        </patient>
    </patientRole>
</recordTarget>
```

Figure 3.9: An extract from CDA before de-identification

As seen in these figures above, we have applied the same de-identification methods (substitute pseudorandom value, shift date values by random offset, generalize geographical location, substitute meaningless value) to the clinical instance in CDA format as well and successfully de-identified this clinical instance.

Figure 3.10: An extract from CDA after de-identification

3.1.1.3 Implementation Details for ICH E2B format for ICSRs

The implementation details for the de-identification of ICSRs in ICH E2B format are similar to the HL7 CDA PCC/CCD templates as they both have the structure as a variant of XML. We have only changed the XPATH definitions from the configuration of CDA according to the E2B structure. Additionally, data elements to be de-identified in ICSRs are different than the other two types of clinical instances. Before working on the E2B instances, we have provided data elements set that needs to be de-identified to our end users and they agreed on possible de-identification mechanisms proposed as presented in Appendix B.

As we did for CDA instances, we have created similar configuration file in sample XML format including each related data elements to be de-identified, XPATH expression to retrieve and manipulate this data elements stored in the ICSRs and de-identification technique to be applied on it. Sample configuration entry for the "Patient ID (Patient GP Medical Record Number)" data element is presented in Figure 3.11.

```
<!--Patient GP Medical Record Number-->
<patientgpmedicalrecordnumb>
    <xpath>ichicsr/safetyreport/patient/patientgpmedicalrecordnumb</xpath>
    <method>Substitute Pseudonym</method>
</patientgpmedicalrecordnumb>
```

Figure 3.11: Sample configuration entry for "Patient ID" needs to be de-identified

The full configuration file to de-identify all required data elements in the E2B instances can be found in Appendix E.

As we did for CDA instances, we have successfully parsed this configuration in Java

to retrieve data element, related xpath expression and de-identification method. After retrieving related data element from the ICSR as a result of XPATH expression, we passed this data element to De-identification Service to de-identify it with respect to the de-identification method configured in the configuration file. When we retrieve the de-identified data element from the De-identification Service, finally, we have successfully updated the ICSR and able to store de-identified E2B instance in another file.

In Figure 3.12 and Figure 3.13, sample parts of the E2B instances before and after the de-identification are presented.

```
<patient>
    <patientinitial>PH</patientinitial>
    <patientgpmedicalrecordnumb>7777</patientgpmedicalrecordnumb>
    <patientspecialistrecordnumb>6666</patientspecialistrecordnumb>
    <patienthospitalrecordnumb>5555</patienthospitalrecordnumb>
    <patientinvestigationnumb>4444</patientinvestigationnumb>
    <patientbirthdateformat>102</patientbirthdateformat>
    <patientbirthdate>19800101</patientbirthdate>
    <patientonsetage></patientonsetage>
    <patientonsetageunit></patientonsetageunit>
    <gestationperiod></gestationperiod>
    <gestationperiodunit></gestationperiodunit>
    <patientagegroup></patientagegroup>
    <patientweight>60</patientweight>
    <patientheight>170</patientheight>
    <patientsex>2</patientsex>
    <lastmenstrualdateformat>102</lastmenstrualdateformat>
    <patientlastmenstrualdate>20111130</patientlastmenstrualdate>
    <medicalhistoryepisode>
        <patientepisodenamemeddraversion>14.1</patientepisodenamemeddraversion>
        <patientepisodename>10027183</patientepisodename>
        <patientmedicalstartdateformat>602</patientmedicalstartdateformat>
        <patientmedicalstartdate>2010</patientmedicalstartdate>
        <patientmedicalcontinue>1</patientmedicalcontinue>
        <patientmedicalenddateformat></patientmedicalenddateformat>
        <patientmedicalenddate></patientmedicalenddate>
        <patientmedicalcomment>meniere</patientmedicalcomment>
        <patientmedicalhistorytext>meniere</patientmedicalhistorytext>
    </medicalhistoryepisode>
    <patientpastdrugtherapy>
        <patientdrugname>Meclozine</patientdrugname>
        <patientdrugstartdateformat>602</patientdrugstartdateformat>
        <patientdrugstartdate>2010</patientdrugstartdate>
        <patientdrugenddateformat>602</patientdrugenddateformat>
        <patientdrugenddate>2011</patientdrugenddate>
        <patientindicationmeddraversion>14.1</patientindicationmeddraversion>
        <patientdrugindication>10027183</patientdrugindication>
        <patientdrgreactionmeddraversion></patientdrgreactionmeddraversion>
        <patientdrugreaction></patientdrugreaction>
    </patientpastdrugtherapy>
    <patientdeath>
        <patientdeathdateformat></patientdeathdateformat>
        <patientdeathdate></patientdeathdate>
        <patientautopsyyesno></patientautopsyyesno>
        <patientdeathcause>
```

Figure 3.12: An extract from E2B before de-identification

As seen in these figures above, we have applied a number of the de-identification methods (substitute fixed length meaningless value, shift date values by random offset, generalize dates to year) to the ICSR in E2B format.

Figure 3.13: An extract from E2B after de-identification

As a next step, de-identified instances in all formats are ready to be passed to the Pseudonymization Service. As mentioned before, in our architecture, the instances retrieved from EHR sources include both IDAT and MDAT together requiring the two separate services as "De-identification" and "Pseudonymization" due to the fact that Pseudonymization service is only allowed to process MDAT (in encrypted format) to assign a specific pseudonym for the patient identifier. Therefore, de-identification process should take a part before pseudonymization in order to de-identify IDAT within these instances.

As an outcome of the De-Identification process, the de-identified medical data together with a study specific PID are passed to the Pseudonymization Service, so that a unique Pseudonym can be assigned replacing this PID. Finally, in reference to the ISO/TS 25237:2008 standard, only totally de-identified and pseudonymized medical data can be passed to the Research Zone.

3.1.2 Pseudonymization Service

According to the ISO/TS 25237:2008 standard as presented in Chapter 2.1.1, pseudonymization is a particular type of anonymization that both removes the relation with a data subject and adds an association between a particular set of characteristics relating to the data subject and one or more pseudonyms [10]. There are two types of pseudonymization which take place in the literature. These are:

- **Irreversible pseudonymization:** The pseudonymized data do not contain information that allows the re-establishment of the link between the pseudonymized data and the data subject. ISO reports that, if these conditions are met, the resulting data can be considered anonymous data in the sense of Directive 95/46/EC.

This would necessitate the data to be coded through one way coding/ cryptography algorithms: it is not possible to recalculate the direct identifiers from the pseudonyms replacing direct identifiers. In the [9], it is presented that pseudonymization achieved by one-way cryptography algorithms generally creates anonymous data. For the data which is pseudonymized with an irreversible pseudonymization algorithm to be considered as anonymous data (often called coded anonymous data), it should also be ensured that, the data should also be de-identified, i.e. it should not be possible to indirectly identify the data subject by linking the pseudonymized data with external data sets.

- **Reversible pseudonymization:** The pseudonymized data can be linked with the data subject by applying procedures restricted to Trusted Third parties. This process is often called Re-identification. This can be achieved through two way coding/cryptography algorithms. In most of the cases, reversible pseudonymization would require consent of the data subject.

In the scope of this thesis work, we have implemented both irreversible (hash-based) and reversible pseudonymization algorithms (based on two way cryptography algorithms) on top of the retrieved input from the De-identification Service in order to replace patient identifiers (PIDs) with pseudonyms.

3.1.2.1 Implementation Details for Irreversible Pseudonymization

To implement hash-based pseudonymization, we have used the SHA256 Message-Digest Algorithm a widely used cryptographic hash function that produces a 256-bit (32-byte) hash value. This algorithm has a cryptographic hash function taking an arbitrary block of data and returning a fixed-size bit string, the hash value. Java security package java.security [44] provides certain useful classes to generate Hash values with the functionality of a message digest algorithm, such as MD5 or SHA. At first, we have generated SHA256 Hash values by using Java.

However, it is possible to perform a dictionary attack (by accessing lookup table) based on hash values in case of putting all of them in a database table after hashing. Therefore, instead of directly hashing the values, we append a random string called "salt" to it before hashing. Thus the benefit provided by using a salted value is making a lookup table assisted dictionary attack against the stored values impractical, provided the salt is large enough. That is, an attacker would not be able to create a precomputed lookup table of hashed values (value + salt), because it would take too much space. A simple dictionary attack is still very possible, although much slower since it cannot be precomputed.

3.1.2.2 Implementation Details for Reversible Pseudonymization

To implement reversible pseudonymization, we have used two-way cryptography algorithm including encryption and decryption with the generated private key. In the Pseudonymization service, as only Trusted Third Party (TTP) is only allowed to encrypt and decrpt the message, we have used symmetric encryption technique in which

a single private key is used to both encrypt and decrypt the related message so that it arrives securely. In order to encrypt a part of the message that will be passed to Research zone, private key should be created by the TTP. For the time being, we have created our private key in the X.509 format to test our implementation. After that, we have used AES (Advance Encryption Standard) as an encryption algorithm with 256-bit keys available in java.security [44] package.

As a result of Pseudonymization Service, we are now able to generate pseudonyms (PSNs) for patient identifiers (PIDs) in both reversible and irreversible ways. The usage of these facilities in the Pseudonymization Service can be configured by the end-users based on their requirements in each different SALUS pilot application scenarios.

3.2 Additional Security Mechanisms

In the security architecture, the data level protection mechanisms are supported by additional mechanisms compliant with IHE Audit Trail and Node Authentication (ATNA) profile to ensure the clinical data security with the following principles:

- Each transaction should be audited to an audit repository to ensure accountability
- Each node should be mutually authenticated through X509 certificates
- Transport Layer Security (TLS) should be used for each transaction

We present the applicability of the IHE ATNA profile in two separate categories for these additional security mechanisms as "Auditing" and "Message Level Security" in the following subsections.

3.2.1 Auditing (Audit Trail)

In our security and privacy architecture, user accountability is provided through audit trail mechanism. The audit trail enables to audit activities in order to facilitate detection of improper creation, access, modification and deletion of personal health information.

Regarding this phase, security architecture includes an Audit Record Repository that implements IHE ATNA "Audit Record Repository Role". Each Actor in SALUS architecture that implements the Secure Node Actor in IHE ATNA profile, communicates with the Audit Record Repository through the "ITI-20: Record Audit Event" transaction of the IHE IT Infrastructure Technical Framework. In this security architecture, the client side implementation of ITI-20: Record Audit Event is also provided for the use of other components.

For the ITI-20 Record Audit Event transaction, we have analyzed Audit Record format compatible with IHE Audit Trail XML format covering RFC 3881 standard. This format has the following structure shown in Figure 3.14.

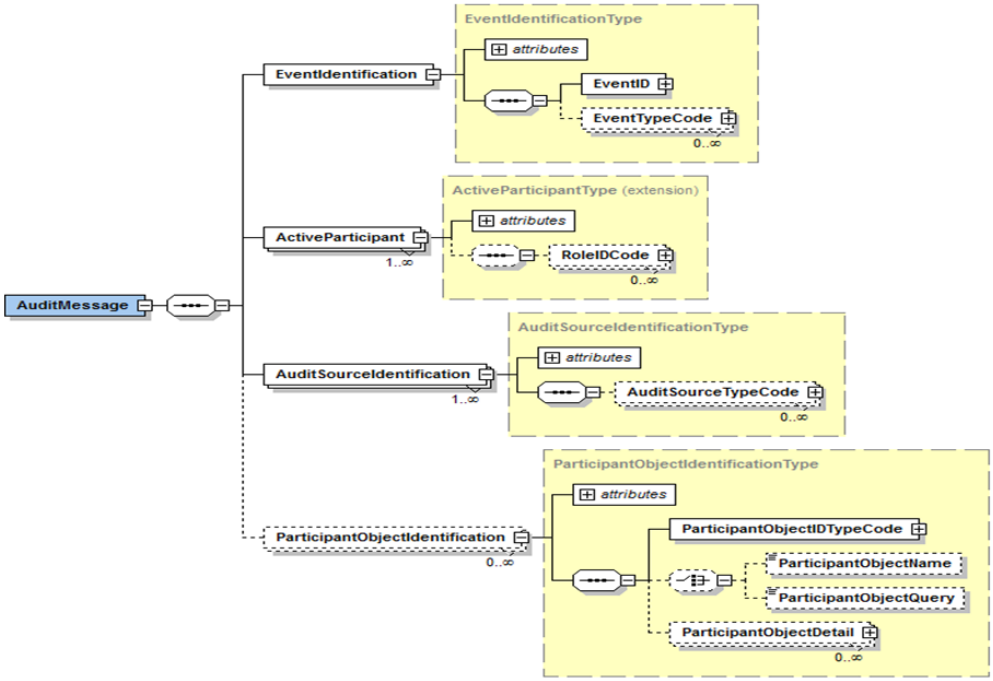


Figure 3.14: RFC 3881 Audit Record format

In this format, Event Action Code and Event Outcome Indicator attributes in Event Identification can be the one of the predefined values as presented in Table 3.2 and Table 3.3.

Table3.2: Event Action Code values

Value	**Meaning**	**Explanation**
C	Create	Create a new database object
R	Read/View/Print/Query	Display or print data
U	Update	Update data
D	Delete	Delete items
E	Execute	Perform a system or an application function, program execution or use of an object's method

In addition to these, to define Event Identification properly, we need to define Event IDs and Event Type Codes for auditing transactions between research and care zones in SALUS architecture with a suitable manner. We have identified 13 transactions taking place from care zone to the research zone to be audited in SALUS pilot application at the premises of SALUS end users.

To audit these transactions to the Audit Record Repository (ARR), we have created audit messages for each transaction between care and research zones. Figure 3.15 presents a sample audit message for one of the pilot application scenarios (ICSR re-

Table3.3: Event Outcome Indicator values

Value	Meaning
0	Success
4	Minor failure
8	Serious failure
12	Major failure

porting case).

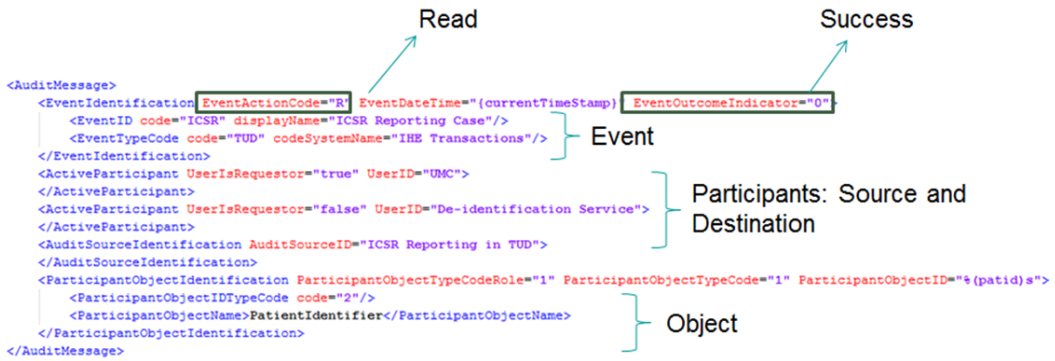


Figure 3.15: Sample Audit message in SALUS

After preparing suitable Audit Record messages, we have dealed with the sending of these messages to the Audit Record Repository. The transport of the transaction of Audit Record messages can be done in 2 ways:

- Syslog Messages (RFC 5424) [45] over TLS (RFC 5425) [36]
- Syslog messages (RFC 5424) over UDP (RFC 5426) [46]

TLS on top of TCP is reliable and secure but also less efficient. On the other side, UDP is easier (no certificates etc) and faster. In this security architecture, we have implemented both UDP and TLS to be used for the transmission of Audit Messages. Therefore, both UDP and TLS are tested for the SALUS components to send Audit messages to Audit Record Repository (ARR).

When the related messages are sent to the Audit Record Repository, these logs can be seen from both ARR server from terminal and simple graphical user interface of ARR by filtering any criteria as presented in Figure 3.16.

3.2.2 Message Level Security (Node Authentication)

This part of the IHE ATNA profile allows each secure node to use the access control to authenticate users. It also requires the use of bi-directional certificate-based node

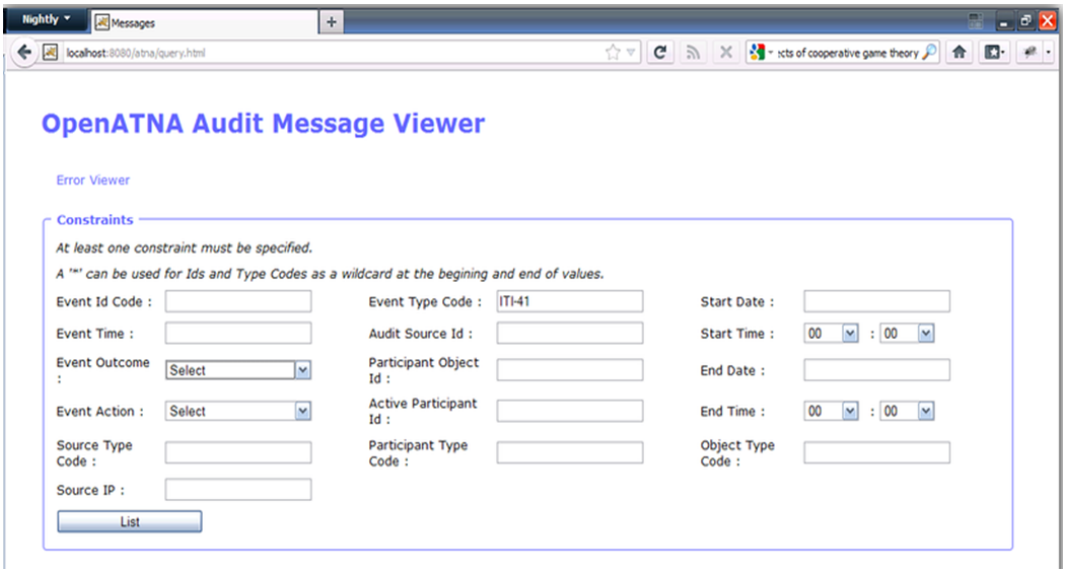


Figure 3.16: Sample view of Audit Message Viewer

authentication for connections to and from each node. For this purpose, X.509 certificates for node identity and keys, TCP/IP Transport Layer Security Protocol (TLS) for node authentication with an optional encryption are the basis to implement this part of IHE ATNA Profile.

In the scope of message level security, each Actor in SALUS architecture that is involved in transactions carrying Personal Health Information will implement the IHE Secure Node Actor. Each Secure Node will implement the "ITI-19 Authenticate Node" transaction, in other words, provide an interface named "Authenticate Node". In the Authenticate Node transaction, the local Secure Node will present its identity to a remote Secure Node, and authenticate the identity of the remote node. After this mutual authentication other secure transactions can take place through this secure pipe between the two nodes. This transaction uses RFC 2246 Transport Layer Security (TLS) 1.0 standard for node authentication which includes X.509 certificates for node identity and keys. These certificates are usually issued by a certification authority (CA) and contain identification information, a validity period, a public key, a serial number, and the digital signature of the issuer.

For the time being, we have created our own digital certificate to test the secure communication of our system. However, in order to fully test this part of the work, all actors should implement ITI-19 Node Authentication transaction to communicate each other and all certificates should be issued by CA. The system will be configured accordingly with the collaboration of the other partners in SALUS consortium. In this part of the security architecture, we have implemented our part to be integrated or re-used by the other partners according to their configurations.

CHAPTER 4

CONCLUSIONS AND FUTURE WORK

Within the scope of this thesis work, we have provided the design and implementation details of extensible security infrastructure on both data level and supporting it with additional security services. We have implemented de-identification and pseudonymization methods by using service-oriented methodology providing extensible approach to implement additional methods with respect to the requested features of the end-users. As a result, we have achieved a novel extensible security infrastructure for the secondary use of EHRs in clinical research. The novel aspects can be summarized as:

- De-identification is processed on top of the queried clinical data instances instead of all data elements in a data warehouse,
- An extensible de-identification framework is created based on SOA principles using RESTful implementations in a modular way that makes further development easy based on the needs of the end-users,
- A flexible de-identification framework is created where the de-identification method can be configured for each data element after analyzing and assessing the risks for each data element set that needs to be exchanged in a de-identified manner with the respective stakeholders.

Regarding the data level protection mechanisms in case of the rare conditions, we have also conducted state-of-the-art research on the statistical data protection approaches and provided the results related to the usage of clinical data in the scope of the SALUS Project. We have analyzed many open source tools for statistical disclosure as well as existing approaches and algorithms. However, as mentioned before, the assumption for this part of the work is that, our end-users have provided a list of rare cases that they want to de-identify and not to share research parties for any purposes. Therefore, we do not need to implement any mechanism for the statistical disclosure control in SALUS project from scratch right now.

After finalizing the work in the data level protection side, we have implemented additional security services compliant with IHE ATNA Profile for auditing and node authentication purposes for the secure communication between care and research zones in the scope of SALUS project.

As a result, within the scope of this thesis work, we have achieved extensible security infrastructure that can be used by both clinical care parties as well as clinical research parties at the same time by preserving the patient privacy. This thesis work covers almost all security infrastructure of the SALUS project.

Overall, the main contributions of this thesis can be summarized as follows:

- We have created novel data protection mechanisms in the scope of this extensible security infrastructure to enable the secondary use of EHRs for research purposes.
- Data level security mechanisms are supported by additional security services compliant with IHE Audit Trail and Node Authentication Profile (ATNA).
- We have created this security framework based on Service Oriented Architecture (SOA) methodologies in a modular way that gives a chance for further developments. Additionally, as all our implementation will be provided as open source software, it will be further improved based on existing plans and end-user feedback.
- We have designed and implemented our security infrastructure configurable and independent from the underlying structure to be updated according to the requests from end users easily.

As a future work in the data level data protection, in order to apply de-identification methods easily according to the requirements of the users, an easy to use Graphical User Interface (GUI) can be provided to configure the methods for specific data elements from the clinical instances. Additionally, this interface can also enable the users to set the thresholds for the specific rare cases instead of making manual configuration. However, this user-centric configuration also requires the suggestion of the necessary techniques based on the types of the data elements in order to ensure compliance with the European regulations and ISO guidelines as end-users may not totally know the related rules and regulations relying on statistical analysis in detail.

Regarding the additional security services, as mentioned, we have implemented our part supporting the both Audit Trail and Node Authentication parts of the IHE ATNA Profile. However, to fully test these services, all SALUS components should implement the related actors to securely communicate in the infrastructure as well as to audit the related events. This part of the work will be revisited after all integrations are done in the SALUS architecture.

REFERENCES

[1] K. Pommerening and M. Reng, "Secondary Use of the EHR via Pseudonymisation," *Studies in Health Technology and Informatics*, vol. 103, no. Medical and Care Compunetics 1, pp. 441–46, 2004.

[2] C. Thielscher, M. Gottfried, S. Umbreit, F. Boegner, J. Haack, and N. Schroeders, "Data processing system for patient data," in *Patent, WO 03/034294 A2.* Int. Patent, 2005.

[3] R. Noumeir, A. Lemay, and L. J.M., "Pseudonymization of radiology data for research purposes," *J Digit Imaging*, vol. 20, no. 3, pp. 284–95, Sep 2007.

[4] SALUS: Scalable, Standard based Interoperability Framework for Sustainable Proactive Post Market Safety Studies. [Online]. Available: http://www.salusproject.eu Last visited on June 2013.

[5] Clinical Document Architecture (CDA), Release 2. HL7. [Online]. Available: http://www.hl7.org/implement/standards/cda.cfm Last visited on June 2013.

[6] *Electronic Transmission of Individual Case Safety Reports Message Specification - E2B(R2)*, International Conference On Harmonisation Of Technical Requirements For Registration Of Pharmaceuticals For Human Use (ICH) Std., 2001.

[7] OMOP CDM Content Entity Model Ontology. [Online]. Available: http://www.salusproject.eu/ontology/omop-cdm-ontology.n3 Last visited on June 2013.

[8] Directive 95/46/EC of the European Parliament. [Online]. Available: http://ec.europa.eu/justice/policies/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf Last visited on May 2013.

[9] Article 29 Data Protection Working Party, 01248/07/EN WP 136, Opinion 4/2007 on the concept of personal data. [Online]. Available: http://ec.europa.eu/justice_home/fsj/privacy/index_en.htm Last visited on May 2013.

[10] *ISO/TS 25237:2008 Health Informatics – Pseudonymization*, ISO/TS Std., 2008. [Online]. Available: http://www.iso.org/iso/catalogue_detail?csnumber=42807

[11] Health Insurance Portability and Accountability Act (HIPAA) Code of Federal Regulations (CFR) Title 45, Part 164, Section 502(d) (CFR§164.502(d)) Uses and disclosures of protected health information: general rules. [Online]. Available: http://www.hhs.gov/ocr/privacy/hipaa/understanding/special/research/index.html Last visited on May 2013.

[12] Healthcare Information Technology Standards Panel (HITSP) Anonymize Public Health Case Reporting Data Component: HITSP/C87. [Online]. Available: http://www.hitsp.org/Handlers/HitspFileServer.aspx?FileGuid=9d2f55e2-a856-48a9-9c87-9dc345f02d80 Last visited on May 2013.

[13] "IHE IT Infrastructure White Paper - Healthcare Pseudonymization Handbook," Integrating the Healthcare Enterprise (IHE)," Trial Implementation Supplement, 2012. [Online]. Available: ftp://ftp.ihe.net/IT_Infrastructure/iheitiyr10-2012-2013/Technical_cmte/WorkItems/PseudonymizationWhitePaper/IHE-Healthcare-Data-Pseudonymization-White-Paper-InternalDraft-v033.docx

[14] Data Protection Guidelines on research in the Health Sector. [Online]. Available: http://www.dataprotection.ie/documents/guidance/Health_research.pdf Last visited on May 2013.

[15] L. Sweeney, "K-anonymity: A model for protecting privacy," *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, vol. 10, no. 5, pp. 557–70, 2002.

[16] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L-diversity: Privacy beyond k-anonymity," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 1, no. 1, pp. 3–es, Mar 2007.

[17] L. Ninghui, L. Tiancheng, and S. Venkatasubramanian, "T-Closeness: Privacy Beyond k-Anonymity and l-Diversity," in *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on.* IEEE Conference Publishing Services, Apr 2007, pp. 106–115.

[18] L. Sweeney, "Guaranteeing anonymity when sharing medical data, the Datafly system," in *Proceedings, Journal of the American Medical Informatics Association.* Hanley and Belfus, Inc., 1997.

[19] A. Hundepool and L. Willenborg, "m- and t-argus: Software for statistical disclosure control," in *Third International Seminar on Statistical Confidentiality.* Bled, 1996.

[20] L. Sweeney, "Towards the optimal suppression of details when disclosing medical data, the use of sub-combination analysis," in *Proceedings, MEDINFO 98.* International Medical Informatics Association, 1998.

[21] C. M. B. Fung, K. Wang, R. Chen, and P. Yu, "Privacy-preserving data publishing: A survey of recent developments," *ACM Computing Surveys (CSUR)*, vol. 42, no. 4, pp. 1–53, Jun 2010.

[22] ARX Data Anaymization Tool. [Online]. Available: http://arx.deidentifier.org/ Last visited on June 2013.

[23] UTD Anonymization Toolbox. [Online]. Available: http://cs.utdallas.edu/dspl/cgi-bin/toolbox/index.php Last visited on June 2013.

[24] Cornell Anonymization Toolkit. [Online]. Available: http://sourceforge.net/projects/anony-toolkit/ Last visited on June 2013.

[25] F. J. Friedlin and C. J. McDonald, "A Software Tool for Removing Patient Identifying Information from Clinical Documents," *J Am Med Inform Assoc.*, vol. 15, no. 5, pp. 601–10, Sep-Oct 2008.

[26] O. Ferrández, B. R. South, S. Shen, and S. M. Meystre, "A Hybrid Stepwise Approach for De-identifying Person Names in Clinical Documents," in *Proceedings of the 2012 Workshop on Biomedical Natural Language Processing (BioNLP 2012)*. Association for Computational Linguistics, June 2012, pp. 65–72.

[27] Custodix Anonymisation Tool. [Online]. Available: https://www.custodix.com/index.php/products/data-privacy/cat Last visited on June 2013.

[28] Custodix Anonymisation Services. [Online]. Available: https://www.custodix.com/index.php/products/data-privacy/cats Last visited on June 2013.

[29] Reference Information Model (RIM). HL7. [Online]. Available: http://www.hl7.org/implement/standards/rim.cfm Last visited on May 2013.

[30] Continuity of Care Document (CCD). HL7 / ASTM International. [Online]. Available: http://wiki.hl7.org/index.php?title=Product_CCD Last visited on May 2013.

[31] *ASTM E2369 - 05e2 Standard Specification for Continuity of Care Record (CCR)*, ASTM International Std., 2003. [Online]. Available: http://www.astm.org/Standards/E2369.htm Last visited on May 2013.

[32] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific American*, vol. 284, no. 5, pp. 34–43, May 2001.

[33] *Resource Description Framework (RDF)*, W3C Std., 1999. [Online]. Available: http://www.w3.org/RDF/ Last visited on May 2013.

[34] *SPARQL Query Language for RDF*, W3C Std., 2008. [Online]. Available: http://www.w3.org/TR/rdf-sparql-query/ Last visited on May 2013.

[35] Audit Trail and Node Authentication (ATNA) Profile. IHE. [Online]. Available: http://wiki.ihe.net/index.php?title=Audit_Trail_and_Node_Authentication Last visited on May 2013.

[36] Transport Layer Security (TLS) Transport Mapping for Syslog. IETF. [Online]. Available: http://tools.ietf.org/html/rfc5425 Last visited on May 2013.

[37] OpenATNA Project. [Online]. Available: https://www.projects.openhealthtools.org/sf/projects/openatna/ Last visited on May 2013.

[38] Security Audit and Access Accountability XML Message Data Definitions for Healthcare Applications. [Online]. Available: http://tools.ietf.org/html/rfc3881 Last visited on May 2013.

[39] Web Service technologies with REST architecture. [Online]. Available: http://www.w3.org/TR/ws-arch/#relwwwrest Last visited on May 2013.

[40] XML Parsing for Java . [Online]. Available: http://docs.oracle.com/cd/B28359_01/appdev.111/b28394/adx_j_parser.htm Last visited on May 2013.

[41] Apache Jena API. [Online]. Available: http://jena.apache.org/ Last visited on May 2013.

[42] XPath Tutorial. [Online]. Available: http://www.w3schools.com/xpath/ Last visited on May 2013.

[43] Package javax.xml.xpath. [Online]. Available: http://docs.oracle.com/javase/1.5.0/docs/api/javax/xml/xpath/package-summary.html Last visited on May 2013.

[44] Java SE Security. [Online]. Available: http://www.oracle.com/technetwork/java/javase/tech/index-jsp-136007.html Last visited on May 2013.

[45] The Syslog Protocol. [Online]. Available: http://tools.ietf.org/html/rfc5424 Last visited on May 2013.

[46] Transmission of Syslog Messages over UDP. [Online]. Available: http://tools.ietf.org/html/rfc5426 Last visited on May 2013.

APPENDIX A

DATA ELEMENTS IN HL7 CDA RDF AND IN HL7 CDA PCC/CCD TEMPLATES WITH POSSIBLE DE-IDENTIFICATION TECHNIQUES

TableA.1: Data elements in HL7 CDA RDF and in HL7 CDA PCC/CCD templates with possible De-identification Techniques

Field (CDE)	**Direct identifier**	**Quasi identifier**	**De-identification Method**
Patient. ID	Yes	-	Substitute Pseudonym (PS)
Patient. Allergy. AdverseEventType	No	No	Pass
Patient. Allergy. TimeInterval	No	No	Shift date
Patient. Allergy. Product	No	No	Pass
Patient. Allergy. Reaction	No	No	Pass
Patient. Allergy. Status	No	No	Pass
Patient. Allergy. Severity	No	No	Pass
Patient. Allergy. Comment	No	No	Remove
Patient. Condition. TimeInterval	No	No	Shift date
Patient. Condition. ProblemType	No	No	Pass
Patient. Condition. ProblemName	No	No	Pass/Generalize in the case of Rare Conditions

Continued on next page

Table A.1 – continued from previous page

Field (CDE)	Direct identifier	Quasi identifier	De-identification Method
Patient. Condition. ProblemCode	No	No	Pass/Generalize in the case of Rare Conditions
Patient. Condition. ProblemStatus	No	No	Pass
Patient. Condition. ProblemSeverity	No	No	Pass
Patient. Condition. TimeOfDeath	No	No	Shift date
Patient. Condition. Comment	No	No	Remove
Patient. Immunization. AdministeredDate	No	No	Shift date
Patient. Immunization. MedicationSeriesNumber	No	No	Pass
Patient. Immunization. Route	No	No	Pass
Patient. Immunization. Dose	No	No	Pass
Patient. Immunization. Site	No	No	Pass
Patient. Immunization. Reaction	No	No	Pass
Patient. Immunization. MedicationInformation. ProductName	No	No	Pass/Generalize in the case of rare drugs
Patient. Immunization. MedicationInformation. ProductName	No	No	Pass/Generalize in the case of rare drugs
Patient. Immunization. MedicationInformation. ActiveIngredient	No	No	Pass/Generalize in the case of rare drugs
Patient. Immunization. MedicationInformation. BrandName	No	No	Pass/Generalize in the case of rare drugs
Patient. Immunization. MedicationInformation. BrandName	No	No	Pass/Generalize in the case of rare drugs
Patient. Immunization. Comment	No	No	Remove

Continued on next page

Table A.1 – continued from previous page

Field (CDE)	Direct identifier	Quasi identifier	De-identification Method
Patient. Medication. TimeInterval	No	No	Shift date
Patient. Medication. AdministeredTiming	No	No	Shift date
Patient. Medication. Route	No	No	Pass
Patient. Medication. Dose	No	No	Pass
Patient. Medication. Site	No	No	Pass
Patient. Medication. DoseRestriction	No	No	Pass
Patient. Medication. ProductForm	No	No	Pass
Patient. Medication. DeliveryMethod	No	No	Pass
Patient. Medication. MedicationInformation. ProductName	No	No	Pass/Generalize in the case of rare drugs
Patient. Medication. MedicationInformation. ProductName	No	No	Pass/Generalize in the case of rare drugs
Patient. Medication. MedicationInformation. ActiveIngredient	No	No	Pass/Generalize in the case of rare drugs
Patient. Medication. MedicationInformation. BrandName	No	No	Pass/Generalize in the case of rare drugs
Patient. Medication. MedicationInformation. BrandName	No	No	Pass/Generalize in the case of rare drugs
Patient. Medication. Indication. TimeInterval	No	No	Shift date
Patient. Medication. Indication. ProblemType	No	No	Pass
Patient. Medication. Indication. ProblemName	No	No	Pass/Generalize in the case of Rare Conditions
Patient. Medication. Indication. ProblemCode	No	No	Pass/Generalize in the case of Rare Conditions

Continued on next page

Table A.1 – continued from previous page

Field (CDE)	Direct identifier	Quasi identifier	De-identification Method
Patient. Medication. Indication. ProblemStatus	No	No	Pass
Patient. Medication. Indication. ProblemSeverity	No	No	Pass
Patient. Medication. Indication. TimeOfDeath	No	No	Shift date
Patient. Medication. Indication. Comment	No	No	Remove
Patient. Medication. PatientInstructions	No	No	Pass
Patient. Medication. Reaction	No	No	Pass
Patient. Medication. Order. Number	No	No	Pass
Patient. Medication. Order. FillNumber	No	No	Pass
Patient. Medication. Order. QuantityOrdered	No	No	Pass
Patient. Medication. Order. ExpirationDateTime	No	No	Shift date
Patient. Medication. Order. DateTime	No	No	Shift date
Patient. Medication. FulfillmentInstructions	No	No	Pass
Patient. Medication. FulfillmentHistory. PrescriptionNumber	No	No	Pass
Patient. Medication. FulfillmentHistory. DispenseDate	No	No	Shift date
Patient. Medication. FulfillmentHistory. QuantityDispensed	No	No	Pass
Patient. Medication. FulfillmentHistory. FillNumber	No	No	Pass

Continued on next page

Table A.1 – continued from previous page

Field (CDE)	Direct identifier	Quasi identifier	De-identification Method
Patient. Medication. FulfillmentHistory. FillStatus	No	No	Pass
Patient. Medication. Comment	No	No	Remove
Patient. Pregnancy. ObservationDate	No	No	Shift date
Patient. Pregnancy. LastMenstrualPeriodDate	No	No	Shift date
Patient. Pregnancy. DeliveryDate	No	No	Shift date
Patient. Pregnancy. Comment	No	No	Remove
Patient. Procedure. TimeInterval	No	No	Shift date
Patient. Procedure. Type	No	No	Pass/Generalize in case of rare procedures
Patient. Procedure. Type	No	No	Pass/Generalize in case of rare procedures
Patient. Procedure. Status	No	No	Pass
Patient. Procedure. Site	No	No	Pass
Patient. Procedure. Indication. TimeInterval	No	No	Shift date
Patient. Procedure. Indication. ProblemType	No	No	Pass
Patient. Procedure. Indication. ProblemName	No	No	Pass/Generalize in the case of Rare Conditions
Patient. Procedure. Indication. ProblemCode	No	No	Pass/Generalize in the case of Rare Conditions
Patient. Procedure. Indication. ProblemStatus	No	No	Pass
Patient. Procedure. Indication. ProblemSeverity	No	No	Pass

Continued on next page

Table A.1 – continued from previous page

Field (CDE)	Direct identifier	Quasi identifier	De-identification Method
Patient. Procedure. Indication. TimeOfDeath	No	No	Shift date
Patient. Procedure. Indication. Comment	No	No	Remove
Patient. Procedure. Comment	No	No	Remove
Patient. Result. TimeInterval	No	No	Shift date
Patient. Result. Type	No	No	Pass/Generalize in case of rare tests
Patient. Result. Value	No	No	Pass
Patient. Result. Value	No	No	Pass
Patient. Result. Value	No	No	Pass
Patient. Result. Interpretation	No	No	Pass
Patient. Result. ReferenceRange	No	No	Pass
Patient. Result. RelatedCondition. TimeInterval	No	No	Shift date
Patient. Result. RelatedCondition. ProblemType	No	No	Pass
Patient. Result. RelatedCondition. ProblemName	No	No	Pass/Generalize in the case of Rare Conditions
Patient. Result. RelatedCondition. ProblemCode	No	No	Pass/Generalize in the case of Rare Conditions
Patient. Result. RelatedCondition. ProblemStatus	No	No	Pass
Patient. Result. RelatedCondition. ProblemSeverity	No	No	Pass
Patient. Result. RelatedCondition. TimeOfDeath	No	No	Shift date
Patient. Result. RelatedCondition. Comment	No	No	Remove
Patient. Result. Comment	No	No	Remove

Continued on next page

Table A.1 – continued from previous page

Field (CDE)	Direct identifier	Quasi identifier	De-identification Method
Patient. VitalSign. TimeInterval	No	No	Shift date
Patient. VitalSign. Type	No	No	Pass/Generalize in case of rare tests
Patient. VitalSign. Value	No	No	Pass
Patient. VitalSign. Value	No	No	Pass
Patient. VitalSign. Value	No	No	Pass
Patient. VitalSign. Interpretation	No	No	Pass
Patient. VitalSign. ReferenceRange	No	No	Pass
Patient. VitalSign. RelatedCondition. TimeInterval	No	No	Shift date
Patient. VitalSign. RelatedCondition. ProblemType	No	No	Pass
Patient. VitalSign. RelatedCondition. ProblemName	No	No	Pass/Generalize in the case of Rare Conditions
Patient. VitalSign. RelatedCondition. ProblemCode	No	No	Pass/Generalize in the case of Rare Conditions
Patient. VitalSign. RelatedCondition. ProblemStatus	No	No	Pass
Patient. VitalSign. RelatedCondition. ProblemSeverity	No	No	Pass
Patient. VitalSign. RelatedCondition. TimeOfDeath	No	No	Shift date
Patient. VitalSign. RelatedCondition. Comment	No	No	Remove
Patient. VitalSign. Comment	No	No	Remove

APPENDIX B

DATA ELEMENTS IN ICH E2B WITH POSSIBLE DE-IDENTIFICATION TECHNIQUES

TableB.1: Data elements in ICH E2B with possible De-identification Techniques

E2B data element	**E2B XML tag**	**Direct identifier**	**Quasi identifier**	**De-identification Method**
Message Type	<messagetype>	No	No	Pass
Message Format Version	<messageformatversion>	No	No	Pass
Message Format Release	<messageformatrelease>	No	No	Pass
Message Number	<messagenumb>	No	No	Pass
Message Sender Identifier	<messagesenderidentifier>	No	No	Pass
Message Receiver Identifier	<messagereceiveridentifier>	No	No	Pass
Message Date Format	<messagedateformat>	No	No	Pass
Message Date	<messagedate>	No	No	Pass
Sender's (case) safety report unique identifier	<safetyreportid>	No	No	Pass

Continued on next page

Table B.1 – continued from previous page

E2B data element	E2B XML tag	Direct identifier	Quasi identifier	De-identification Method
Identification of the country of the primary source	<primarysourcecountry>	No	No	Pass
Identification of the country where the reaction/event occurred	<occurcountry>	No	No	Pass
Date of this transmission Format	<transmissiondateformat>	No	No	Pass
Date of this transmission	<transmissiondate>	No	No	Pass
Type of report	<reporttype>	No	No	Pass
Serious	<serious>	No	No	Pass
Results in death	<seriousnessdeath>	No	No	Pass
Life threatening	<seriousnesslifethreatening>	No	No	Pass
Caused/ prolonged hospitalization	<seriousnesshospitalization>	No	No	Pass
Disabling/ Incapacitating	<seriousnessdisabling>	No	No	Pass
Congenital anomaly/ birth defect	<seriousnesscongenitalanomali>	No	No	Pass
Other medically important condition	<seriousnessother>	No	No	Pass

Continued on next page

Table B.1 – continued from previous page

E2B data element	E2B XML tag	Direct identifier	Quasi identifier	De-identification Method
Date report was first received from source Format	<receivedateformat>	No	No	Pass
Date report was first received from source	<receivedate>	No	No	Pass
Date of receipt of the most recent information for this report Format	<receiptdateformat>	No	No	Pass
Date of receipt of the most recent information for this report	<receiptdate>	No	No	Pass
Are additional documents available?	<additionaldocument>	No	No	Pass
List of documents held by sender	<documentlist>	No	No	Pass
Does this case fulfill the local criteria for an expedited report?	<fulfillexpeditecriteria>	No	No	Pass
Regulatory authority's case report number	<authoritynumb>	No	No	Pass

Continued on next page

Table B.1 – continued from previous page

E2B data element	E2B XML tag	Direct identifier	Quasi identifier	De-identification Method
Other sender's case report number	<companynumb>	No	No	Pass
Other case identifiers in previous transmissions	<duplicate>	No	No	Pass
Source(s) of the case identifier	<duplicatesource>	No	No	Pass
Case identifiers	<duplicatenumb>	No	No	Pass
Identification number of the report which is linked to this report	<linkreportnumb>	No	No	Pass
Report nullification	<casenullification>	No	No	Pass
Reason for nullification	<nullificationreason>	No	No	Pass
Was the case medically confirmed, if not initially from health professional?	<medicallyconfirm>	No	No	Pass
Reporter title	<reportertitle>	No	No	Pass
Reporter given name	<reportergivename>	No	No	Pass
Reporter middle name	<reportermiddlename>	No	No	Pass
Reporter family name	<reporterfamilyname>	No	No	Pass

Continued on next page

Table B.1 – continued from previous page

E2B data element	E2B XML tag	Direct identifier	Quasi identifier	De-identification Method
Reporter organization	<reporterorganization>	No	No	Pass
Reporter department	<reporterdepartment>	No	No	Pass
Reporter street	<reporterstreet>	No	No	Pass
Reporter city	<reportercity>	No	No	Pass
Reporter state or province	<reporterstate>	No	No	Pass
Reporter postcode	<reporterpostcode>	No	No	Pass
Reporter country code	<reportercountry>	No	No	Pass
Reporter qualification	<qualification>	No	No	Pass
Literature reference(s)	<literaturereference>	No	No	Pass
Study name	<studyname>	No	No	Pass
Sponsor study number	<sponsorstudynumb>	No	No	Pass
Study type in which the reaction(s)/ event(s) were observed	<observestudytype>	No	No	Pass
Sender Type	<sendertype>	No	No	Pass
Sender organization	<senderorganization>	No	No	Pass
Sender department	<senderdepartment>	No	No	Pass
Sender title	<sendertitle>	No	No	Pass
Sender given name	<sendergivename>	No	No	Pass

Continued on next page

Table B.1 – continued from previous page

E2B data element	E2B XML tag	Direct identifier	Quasi identifier	De-identification Method
Sender middle name	<sendermiddlename>	No	No	Pass
Sender family name	<senderfamilyname>	No	No	Pass
Sender Street address	<senderstreetaddress>	No	No	Pass
Sender City	<sendercity>	No	No	Pass
Sender State or Province	<senderstate>	No	No	Pass
Sender Postcode	<senderpostcode>	No	No	Pass
Sender Country Code	<sendercountrycode>	No	No	Pass
Sender Telephone	<sendertel>	No	No	Pass
Sender Telephone extension	<sendertelextension>	No	No	Pass
Sender Telephone country code	<sendertelcountrycode>	No	No	Pass
Sender Fax	<senderfax>	No	No	Pass
Sender Fax extension	<senderfaxextension>	No	No	Pass
Sender Fax country code	<senderfaxcountrycode>	No	No	Pass
Sender E-mail address	<senderemailaddress>	No	No	Pass
Receiver Type	<receivertype>	No	No	Pass
Receiver organization	<receiverorganization>	No	No	Pass
Receiver department	<receiverdepartment>	No	No	Pass
Receiver title	<receivertitle>	No	No	Pass

Continued on next page

Table B.1 – continued from previous page

E2B data element	E2B XML tag	Direct identifier	Quasi identifier	De-identification Method
Receiver given name	<receivergivename>	No	No	Pass
Receiver middle name	<receivermiddlename>	No	No	Pass
Receiver family name	<receiverfamilyname>	No	No	Pass
Receiver Street address	<receiverstreetaddress>	No	No	Pass
Receiver City	<receivercity>	No	No	Pass
Receiver State or Province	<receiverstate>	No	No	Pass
Receiver Postcode	<receiverpostcode>	No	No	Pass
Receiver Country Code	<receivercountrycode>	No	No	Pass
Receiver Telephone	<receivertel>	No	No	Pass
Receiver Telephone extension	<receivertelextension>	No	No	Pass
Receiver Telephone country code	<receivertelcountrycode>	No	No	Pass
Receiver Fax	<receiverfax>	No	No	Pass
Receiver Fax extension	<receiverfaxextension>	No	No	Pass
Receiver Fax country code	<receiverfaxcountrycode>	No	No	Pass
Receiver E-mail address	<receiveremailaddress>	No	No	Pass

Continued on next page

Table B.1 – continued from previous page

E2B data element	E2B XML tag	Direct identifier	Quasi identifier	De-identification Method
Patient name or initials	<patientinitial>	Yes	-	Remove
GP medical record number	<patientgpmedicalrecord numb>	Yes	-	Substitute Pseudonym (PS)
Specialist record number	<patientspecialistrecord numb>	Yes	-	Substitute Pseudonym (PS)
Hospital record number	<patienthospitalrecord numb>	Yes	-	Substitute Pseudonym (PS)
Investigation number	<patientinvestigation numb>	Yes	-	Substitute Pseudonym (PS)
Date of birth Format	<patientbirthdateformat>	No	No	Pass
Date of birth	<patientbirthdate>	No	Yes	Generalize to year (0-1, <95)
Age at time of onset of reaction/event	<patientonsetage>	No	No	Pass
Age at time of onset of reaction/event Unit	<patientonsetageunit>	No	No	Pass
Gestation period when reaction/event was observed in the fetus	<gestationperiod>	No	No	Pass

Continued on next page

Table B.1 – continued from previous page

E2B data element	E2B XML tag	Direct identifier	Quasi identifier	De-identification Method
Gestation period when reaction/event was observed in the fetus Unit	<gestationperiodunit>	No	No	Pass
Patient age group	<patientagegroup>	No	No	Pass
Weight (kg)	<patientweight>	No	No	Pass
Height (cm)	<patientheight>	No	No	Pass
Sex	<patientsex>	No	No	Pass
Last menstrual period date Format	<lastmenstrualdateformat>	No	No	Pass
Last menstrual period date	<patientlastmenstrualdate>	No	No	Shift date
MedDRA version for Medical History	<patientepisodenamemeddraversion>	No	No	Pass
Relevant medical episode (disease / surgical procedure / etc.)	<patientepisodename>	No	No	Pass
Start Date Format	<patientmedicalstartdateformat>	No	No	Pass
Start Date	<patientmedicalstartdate>	No	No	Shift date
Continuing	<patientmedicalcontinue>	No	No	Pass
End Date Format	<patientmedicalenddateformat>	No	No	Pass
End Date	<patientmedicalenddate>	No	No	Shift date
Comments	<patientmedicalcomment>	No	No	Pass

Continued on next page

Table B.1 – continued from previous page

E2B data element	E2B XML tag	Direct identifier	Quasi identifier	De-identification Method
Text for relevant medical history and concurrent conditions	<patientmedicalhistory text>	No	No	Pass
Name of Drug as Reported	<patientdrugname>	No	No	Pass
Start Date Format	<patientdrugstartdate format>	No	No	Pass
Start Date	<patientdrugstartdate>	No	No	Shift date
End Date Format	<patientdrugenddate format>	No	No	Pass
End Date	<patientdrugenddate>	No	No	Shift date
MedDRA version for indication	<patientindicationmeddraversion>	No	No	Pass
Indication	<patientdrugindication>	No	No	Pass
MedDRA version for reaction	<patientdrgreactionmeddraversion>	No	No	Pass
Reaction	<patientdrugreaction>	No	No	Pass
Date of death Format	<patientdeathdate format>	No	No	Pass
Date of death	<patientdeathdate>	No	No	Shift date
MedDRA version for reported cause(s) of death	<patientdeathreportmeddraversion>	No	No	Pass
Reported cause(s) of death	<patientdeathreport>	No	No	Pass
Parent identification	<parentidentification>	Yes	-	Substitute Pseudonym (PS)
Date of birth of parent Format	<parentbirthdateformat>	No	No	Pass

Continued on next page

Table B.1 – continued from previous page

E2B data element	E2B XML tag	Direct identifier	Quasi identifier	De-identification Method
Date of birth of parent	<parentbirthdate>	No	Yes	Generalize to year (0-1, <95)
Age of parent	<parentage>	No	No	Pass
Age of parent Unit	<parentageunit>	No	No	Pass
Last menstrual period date Format	<parentlastmenstrual dateformat>	No	No	Pass
Last menstrual period date	<parentlastmenstrual date>	No	No	Shift date
Weight (kg) of parent	<parentweight>	No	No	Pass
Height (cm) of parent	<parentheight>	No	No	Pass
Sex of parent	<parentsex>	No	No	Pass
MedDRA version for Medical History	<parentmdepisodemed draversion>	No	No	Pass
Relevant medical episode (disease / surgical procedure / etc.)	<parentmedicalepisode name>	No	No	Pass
Start Date Format	<parentmedicalstart dateformat>	No	No	Pass
Start Date	<parentmedicalstart date>	No	No	Shift date
Continuing	<parentmedicalcontinue>	No	No	Pass
End Date Format	<parentmedicalenddate format>	No	No	Pass
End Date	<parentmedicalend date>	No	No	Shift date
Comments	<parentmedical comment>	No	No	Pass

Continued on next page

Table B.1 – continued from previous page

E2B data element	E2B XML tag	Direct identifier	Quasi identifier	De-identification Method
Text for relevant medical history and concurrent conditions of parent (not including reaction/ event)	<parentmedicalrelevant text>	No	No	Pass
Name of Drug as Reported	<parentdrugname>	No	No	Pass
Start Date Format	<parentdrugstartdate format>	No	No	Pass
Start Date	<parentdrugstartdate>	No	No	Shift date
End Date Format	<parentdrugenddate format>	No	No	Pass
End Date	<parentdrugenddate>	No	No	Shift date
MedDRA version for indication	<parentindicationmed draversion>	No	No	Pass
Indication	<parentdrugindication>	No	No	Pass
MedDRA version for reaction	<parentdrgreactionmed draversion>	No	No	Pass
Reactions (if any and known)	<parentdrugreaction>	No	No	Pass
Reaction/ event as reported by primary source	<primarysourcereaction>	No	No	Pass
MedDRA version for reaction/ event term LLT	<reactionmeddraversion llt>	No	No	Pass

Continued on next page

Table B.1 – continued from previous page

E2B data element	E2B XML tag	Direct identifier	Quasi identifier	De-identification Method
Reaction/ event in MedDRA (LLT)	<reactionmeddrallt>	No	No	Pass
MedDRA version for reaction/ event term PT	<reactionmeddra versionpt>	No	No	Pass
Reaction/ event in MedDRA (PT)	<reactionmeddrapt>	No	No	Pass
Term highlighted by the reporter	<termhighlighted>	No	No	Pass
Date of start of reaction/ event Format	<reactionstartdate format>	No	No	Pass
Date of start of reaction/ event	<reactionstartdate>	No	No	Shift date
Date of end of reaction/ event Format	<reactionenddate format>	No	No	Pass
Date of end of reaction/ event	<reactionenddate>	No	No	Shift date
Duration of reaction/ event	<reactionduration>	No	No	Pass
Duration of reaction/ event Unit	<reactionduration unit>	No	No	Pass

Continued on next page

Table B.1 – continued from previous page

E2B data element	E2B XML tag	Direct identifier	Quasi identifier	De-identification Method
Time interval between beginning of suspect drug administration and start of reaction/ event	<reactionfirsttime>	No	No	Pass
Unit	<reactionfirsttime unit>	No	No	Pass
Time interval between last dose and start of reaction/ event	<reactionlasttime>	No	No	Pass
Unit	<reactionlasttimeunit>	No	No	Pass
Outcome of reaction/event at the time of last observation	<reactionoutcome>	No	No	Pass
Date format	<testdateformat>	No	No	Pass
Date	<testdate>	No	No	Shift date
test	<testname>	No	No	Pass
Result	<testresult>	No	No	Pass
Unit	<testunit>	No	No	Pass
Normal low range	<lowtestrange>	No	No	Pass
Normal high range	<hightestrange>	No	No	Pass
More information available	<moreinformation>	No	No	Pass
Description (free text)	<resultstestsprocedures>	No	No	Pass
Characterization of drug role	<drugcharacterization>	No	No	Pass

Continued on next page

Table B.1 – continued from previous page

E2B data element	E2B XML tag	Direct identifier	Quasi identifier	De-identification Method
Proprietary medicinal product name	<medicinalproduct>	No	No	Pass
Active Drug substance names	<activesubstance name>	No	No	Pass
Identification of the country where the drug was obtained	<obtaindrugcountry>	No	No	Pass
Batch/ lot number	<drugbatchnumb>	No	No	Pass
Authorization / Application Number	<drugauthorization numb>	No	No	Pass
Country of authorization/ application	<drugauthorization country>	No	No	Pass
Name of holder/ applicant	<drugauthorization holder>	No	No	Pass
dose (number)	<drugstructure dosagenumb>	No	No	Pass
dose (unit)	<drugstructure dosageunit>	No	No	Pass
number of separate dosages	<drugseparatedosage numb>	No	No	Pass
number of units in the interval	<drugintervaldosage unitnumb>	No	No	Pass
definition of the interval	<drugintervaldosage definition>	No	No	Pass
cumulative dose to first reaction (number)	<drugcumulativedosage numb>	No	No	Pass

Continued on next page

Table B.1 – continued from previous page

E2B data element	E2B XML tag	Direct identifier	Quasi identifier	De-identification Method
cumulative dose to first reaction (unit)	<drugcumulativedosage unit>	No	No	Pass
Dosage text	<drugdosagetext>	No	No	Pass
Pharmaceutical form (Dosage form)	<drugdosageform>	No	No	Pass
Route of administration	<drugadministration route>	No	No	Pass
Parent route of administration (in case of a parent child/fetus report)	<drugparadministration>	No	No	Pass
Gestation period at time of exposure	<reactiongestation period>	No	No	Pass
Gestation period at time of exposure Unit	<reactiongestationperiod unit>	No	No	Pass
MedDRA version for indication	<drugindicationmeddraversion>	No	No	Pass
Indication for use in the case	<drugindication>	No	No	Pass
Date of start of drug Format	<drugstartdate format>	No	No	Pass
Date of start of drug	<drugstartdate>	No	No	Shift date

Continued on next page

Table B.1 – continued from previous page

E2B data element	E2B XML tag	Direct identifier	Quasi identifier	De-identification Method
Time interval between beginning of drug administration and start of reaction/event	<drugstartperiod>	No	No	Pass
Time unit	<drugstartperiodunit>	No	No	Pass
Time interval between last dose of drug and start of reaction/event	<druglastperiod>	No	No	Pass
Time unit	<druglastperiodunit>	No	No	Pass
Date of last administration Format	<drugenddateformat>	No	No	Pass
Date of last administration	<drugenddate>	No	No	Shift date
Duration of drug administration	<drugtreatmentduration>	No	No	Pass
Duration of drug administration Unit	<drugtreatmentdurationunit>	No	No	Pass
Action(s) taken with drug	<actiondrug>	No	No	Pass
Did reaction recur on readministration?	<drugrecurreadministration>	No	No	Pass

Continued on next page

Table B.1 – continued from previous page

E2B data element	E2B XML tag	Direct identifier	Quasi identifier	De-identification Method
If yes, which reaction(s)/ event(s) recurred?	<drugrecuraction>	No	No	Pass
MedDRA version for Reaction assessed	<drugreactionassesmeddraversion>	No	No	Pass
Reaction assessed	<drugreactionasses>	No	No	Pass
Source of assessment (e.g., initial reporter, investigator, regulatory agency, company)	<drugassessmentsource>	No	No	Pass
Method of assessment (e.g., global introspection, algorithm, Bayesian calculation)	<drugassessment method>	No	No	Pass
Result	<drugresult>	No	No	Pass
Additional information on drug	<drugadditional>	No	No	Pass
Case narrative including clinical course, therapeutic measures, outcome and additional relevant information	<narrativeinclude clinical>	No	No	Pass

Continued on next page

Table B.1 – continued from previous page

E2B data element	**E2B XML tag**	**Direct identifier**	**Quasi identifier**	**De-identification Method**
Reporter's comments	<reportercomment>	No	No	Pass
MedDRA Version for Sender's diagnosis	<senderdiagnosismedraversion>	No	No	Pass
Sender's diagnosis/ syndrome and/ or reclassification of reaction/ event	<senderdiagnosis>	No	No	Pass
Sender's comments	<sendercomment>	No	No	Pass

APPENDIX C

HL7 CDA RDF TEMPLATE CONFIGURATION FILE

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<rdfconf>
    <!--Patient.ID.Extension.String -->
    <Patient.ID.Extension.String>
        <sparql>SELECT ?ID ?extension WHERE {?pt a salus:Patient.
        ?pt salus:ID ?ID.?ID salus:extension ?extension.}</sparql>
        <method>Substitute Pseudonym</method>
    </Patient.ID.Extension.String>
    <!--Patient.Allergy.TimeInterval-->
    <Patient.Allergy.TimeInterval.High.Datetime>
        <sparql>SELECT ?adverseEventDate ?high WHERE {?pt a salus:Patient.
        ?pt salus:allergy ?allergy.?allergy salus:adverseEventDate
        ?adverseEventDate.?adverseEventDate salus:high ?high.}</sparql>
        <method>Shift Date RDF</method>
    </Patient.Allergy.TimeInterval.High.Datetime>
    <Patient.Allergy.TimeInterval.Low.Datetime>
        <sparql>SELECT ?adverseEventDate ?low WHERE {?pt a salus:Patient.
        ?pt salus:allergy ?allergy.?allergy salus:adverseEventDate
        ?adverseEventDate.?adverseEventDate salus:low ?low.}</sparql>
        <method>Shift Date RDF</method>
    </Patient.Allergy.TimeInterval.Low.Datetime>
    <Patient.Allergy.TimeInterval.Value.Datetime>
        <sparql>SELECT ?adverseEventDate ?value WHERE {?pt a salus:Patient.
        ?pt salus:allergy ?allergy.?allergy salus:adverseEventDate
        ?adverseEventDate.?adverseEventDate salus:value ?value.}</sparql>
        <method>Shift Date RDF</method>
    </Patient.Allergy.TimeInterval.Value.Datetime>
    <!--Patient.Allergy.Comment-->
    <Patient.Allergy.Comment.String>
        <sparql>SELECT ?allergy ?comment WHERE {?pt a salus:Patient.
        ?pt salus:allergy ?allergy.?allergy salus:comment ?comment.}</sparql>
        <method>Delete Value</method>
    </Patient.Allergy.Comment.String>
    <!--Patient.Condition.TimeInterval-->
    <Patient.Condition.TimeInterval.High.Datetime>
        <sparql>SELECT ?problemDate ?high WHERE {?pt a salus:Patient.?pt
        salus:condition ?condition.?condition salus:problemDate ?problemDate.
        ?problemDate salus:high ?high.}</sparql>
        <method>Shift Date RDF</method>
    </Patient.Condition.TimeInterval.High.Datetime>
    <Patient.Condition.TimeInterval.Low.Datetime>
        <sparql>SELECT ?problemDate ?low WHERE {?pt a salus:Patient.?pt
        salus:condition ?condition.?condition salus:problemDate
        ?problemDate.?problemDate salus:low ?low.}</sparql>
        <method>Shift Date RDF</method>
    </Patient.Condition.TimeInterval.Low.Datetime>
    <Patient.Condition.TimeInterval.Value.Datetime>
        <sparql>SELECT ?problemDate ?value WHERE {?pt a salus:Patient.?pt
        salus:condition ?condition.?condition salus:problemDate ?problemDate.
        ?problemDate salus:value ?value.}</sparql>
        <method>Shift Date RDF</method>
    </Patient.Condition.TimeInterval.Value.Datetime>
    <!--Patient.Condition.TimeOfDeath-->
```

```
<Patient.Condition.TimeOfDeath.Datetime>
    <sparql>SELECT ?condition ?timeOfDeath WHERE {?pt a salus:Patient.
    ?pt salus:condition ?condition.?condition salus:timeOfDeath
    ?timeOfDeath.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Condition.TimeOfDeath.Datetime>
<!--Patient.Condition.Comment-->
<Patient.Condition.Comment.String>
    <sparql>SELECT ?condition ?comment WHERE {?pt a salus:Patient.?pt
    salus:condition ?condition.?condition salus:comment ?comment.}
    </sparql>
    <method>Delete Value</method>
</Patient.Condition.Comment.String>
<!--Patient.Immunization.AdministeredDate-->
<Patient.Immunization.AdministeredDate.Datetime>
    <sparql>SELECT ?immunization ?administeredDate WHERE {?pt a
    salus:Patient.?pt salus:immunization ?immunization.?immunization
    salus:administeredDate ?administeredDate.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Immunization.AdministeredDate.Datetime>
<!--Patient.Immunization.Comment-->
<Patient.Immunization.Comment.String>
    <sparql>SELECT ?immunization ?comment WHERE {?pt a salus:Patient.?pt
    salus:immunization ?immunization.?immunization salus:comment ?comment.}
    </sparql>
    <method>Delete Value</method>
</Patient.Immunization.Comment.String>
<!--Patient.Medication.TimeInterval-->
<Patient.Medication.TimeInterval.High.Datetime>
    <sparql>SELECT ?indicateMedicationStartStop ?high WHERE {?pt a
    salus:Patient. ?pt salus:medication ?medication.?medication
    salus:indicateMedicationStartStop ?indicateMedicationStartStop.
    ?indicateMedicationStartStop salus:high ?high.}
    </sparql>
    <method>Shift Date RDF</method>
</Patient.Medication.TimeInterval.High.Datetime>
<Patient.Medication.TimeInterval.Low.Datetime>
    <sparql>SELECT ?indicateMedicationStartStop ?low WHERE {?pt a
    salus:Patient.?pt salus:medication ?medication.?medication
    salus:indicateMedicationStartStop ?indicateMedicationStartStop.
    ?indicateMedicationStartStop salus:low ?low.}
    </sparql>
    <method>Shift Date RDF</method>
</Patient.Medication.TimeInterval.Low.Datetime>
<Patient.Medication.TimeInterval.Value.Datetime>
    <sparql>SELECT ?indicateMedicationStartStop ?value WHERE {?pt a
    salus:Patient.?pt salus:medication ?medication.?medication
    salus:indicateMedicationStartStop ?indicateMedicationStartStop.
    ?indicateMedicationStartStop salus:value ?value.}
    </sparql>
    <method>Shift Date RDF</method>
</Patient.Medication.TimeInterval.Value.Datetime>
<!--Patient.Medication.AdministeredTiming-->
<Patient.Medication.AdministeredTiming.Value.Datetime>
    <sparql>SELECT ?administrationTiming ?value WHERE {?pt a salus:Patient.
    ?pt salus:medication ?medication.?medication salus:administrationTiming
    ?administrationTiming.?administrationTiming salus:value ?value.}
    </sparql>
    <method>Shift Date RDF</method>
</Patient.Medication.AdministeredTiming.Value.Datetime>
<Patient.Medication.AdministeredTiming.Phase.High.Datetime>
    <sparql>SELECT ?phase ?high WHERE {?pt a salus:Patient.?pt
    salus:medication ?medication.?medication salus:administrationTiming
    ?administrationTiming.?administrationTiming salus:phase ?phase.
    ?phase salus:high ?high.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Medication.AdministeredTiming.Phase.High.Datetime>
<Patient.Medication.AdministeredTiming.Phase.Low.Datetime>
    <sparql>SELECT ?phase ?low WHERE {?pt a salus:Patient.?pt
    salus:medication ?medication.?medication salus:administrationTiming
```

```
    ?administrationTiming. ?administrationTiming salus:phase ?phase.?phase
    salus:low ?low.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Medication.AdministeredTiming.Phase.Low.Datetime>
<Patient.Medication.AdministeredTiming.Phase.Value.Datetime>
    <sparql>SELECT ?phase ?value WHERE {?pt a salus:Patient.?pt
    salus:medication ?medication.?medication salus:administrationTiming
    ?administrationTiming.?administrationTiming salus:phase
    ?phase.?phase salus:value ?value.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Medication.AdministeredTiming.Phase.Value.Datetime>
<!--Patient.Medication.Indication.TimeInterval-->
<Patient.Medication.Indication.TimeInterval.High.Datetime>
    <sparql>SELECT ?problemDate ?high WHERE {?pt a salus:Patient.?pt
    salus:medication ?medication.?medication salus:indication ?indication.
    ?indication salus:problemDate ?problemDate.?problemDate salus:high
    ?high.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Medication.Indication.TimeInterval.High.Datetime>
<Patient.Medication.Indication.TimeInterval.Low.Datetime>
    <sparql>SELECT ?problemDate ?low WHERE {?pt a salus:Patient.?pt
    salus:medication ?medication.?medication salus:indication
    ?indication. ?indication salus:problemDate ?problemDate.
    ?problemDate salus:low ?low.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Medication.Indication.TimeInterval.Low.Datetime>
<Patient.Medication.Indication.TimeInterval.Value.Datetime>
    <sparql>SELECT ?problemDate ?value WHERE {?pt a salus:Patient.
    ?pt salus:medication ?medication.?medication salus:indication
    ?indication. ?indication salus:problemDate ?problemDate.
    ?problemDate salus:value ?value.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Medication.Indication.TimeInterval.Value.Datetime>
<!--Patient.Medication.Indication.TimeOfDeath-->
<Patient.Medication.Indication.TimeOfDeath.Datetime>
    <sparql>SELECT ?indication ?timeOfDeath WHERE {?pt a salus:Patient.
    ?pt salus:medication ?medication.?medication salus:indication
    ?indication.?indication salus:timeOfDeath ?timeOfDeath.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Medication.Indication.TimeOfDeath.Datetime>
<!--Patient.Medication.Indication.Comment-->
<Patient.Medication.Indication.Comment.String>
    <sparql>SELECT ?indication ?comment WHERE {?pt a salus:Patient.?pt
    salus:medication ?medication.?medication salus:indication
    ?indication. ?indication salus:comment ?comment.}</sparql>
    <method>Delete Value</method>
</Patient.Medication.Indication.Comment.String>
<!--Patient.Medication.Order.ExpirationDateTime-->
<Patient.Medication.Order.ExpirationDateTime.Datetime>
    <sparql>SELECT ?orderInformation ?orderExpirationDateTime
    WHERE {?pt a salus:Patient.?pt salus:medication ?medication.
    ?medication salus:orderInformation ?orderInformation.
    ?orderInformation salus:orderExpirationDateTime
    ?orderExpirationDateTime.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Medication.Order.ExpirationDateTime.Datetime>
<!--Patient.Medication.Order.DateTime-->
<Patient.Medication.Order.DateTime.Datetime>
    <sparql>SELECT ?orderInformation ?orderDateTime WHERE {
    ?pt a salus:Patient.?pt salus:medication ?medication.?medication
    salus:orderInformation ?orderInformation.?orderInformation
    salus:orderDateTime ?orderDateTime.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Medication.Order.DateTime.Datetime>
<!--Patient.Medication.FulfillmentHistory.DispenseDate-->
<Patient.Medication.FulfillmentHistory.DispenseDate.Datetime>
    <sparql>SELECT ?fulfillmentHistory ?dispenseDate WHERE {?pt a
    salus:Patient.?pt salus:medication ?medication.
    ?medication salus:fulfillmentHistory ?fulfillmentHistory.
    ?fulfillmentHistory salus:dispenseDate ?dispenseDate.}</sparql>
```

```
    <method>Shift Date RDF</method>
</Patient.Medication.FulfillmentHistory.DispenseDate.Datetime>
<!--Patient.Medication.Comment-->
<Patient.Medication.Comment.String>
    <sparql>SELECT ?medication ?comment WHERE {?pt a salus:Patient.?pt
    salus:medication ?medication.?medication salus:comment ?comment.}
    </sparql>
    <method>Delete Value</method>
</Patient.Medication.Comment.String>
<!--Patient.Pregnancy.ObservationDate-->
<Patient.Pregnancy.ObservationDate.Datetime>
    <sparql>SELECT ?pregnancy ?pregnancyObservationDate WHERE {?pt a
    salus:Patient.?pt salus:pregnancy ?pregnancy.?pregnancy
    salus:pregnancyObservationDate ?pregnancyObservationDate.}
    </sparql>
    <method>Shift Date RDF</method>
</Patient.Pregnancy.ObservationDate.Datetime>
<!--Patient.Pregnancy.LastMenstrualPeriodDate-->
<Patient.Pregnancy.LastMenstrualPeriodDate.Datetime>
    <sparql>SELECT ?pregnancy ?lastMenstrualPeriodDate WHERE {?pt a
    salus:Patient.?pt salus:pregnancy ?pregnancy.?pregnancy
    salus:lastMenstrualPeriodDate ?lastMenstrualPeriodDate.}
    </sparql>
    <method>Shift Date RDF</method>
</Patient.Pregnancy.LastMenstrualPeriodDate.Datetime>
<!--Patient.Pregnancy.DeliveryDate-->
<Patient.Pregnancy.DeliveryDate.Datetime>
    <sparql>SELECT ?pregnancy ?deliveryDate WHERE {?pt a salus:Patient.
    ?pt salus:pregnancy ?pregnancy.?pregnancy salus:deliveryDate
    ?deliveryDate.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Pregnancy.DeliveryDate.Datetime>
<!--Patient.Pregnancy.Comment-->
<Patient.Pregnancy.Comment.String>
    <sparql>SELECT ?pregnancy ?comment WHERE {?pt a salus:Patient.?pt
    salus:pregnancy ?pregnancy.?pregnancy salus:comment ?comment.}
    </sparql>
    <method>Delete Value</method>
</Patient.Pregnancy.Comment.String>
<!--Patient.Procedure.TimeInterval-->
<Patient.Procedure.TimeInterval.High.Datetime>
    <sparql>SELECT ?procedureDateTime ?high WHERE {?pt a salus:Patient.
    ?pt salus:procedure ?procedure.?procedure salus:procedureDateTime
    ?procedureDateTime.?procedureDateTime salus:high ?high.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Procedure.TimeInterval.High.Datetime>
<Patient.Procedure.TimeInterval.Low.Datetime>
    <sparql>SELECT ?procedureDateTime ?low WHERE {?pt a salus:Patient.
    ?pt salus:procedure ?procedure.?procedure salus:procedureDateTime
    ?procedureDateTime.?procedureDateTime salus:low ?low.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Procedure.TimeInterval.Low.Datetime>
<Patient.Procedure.TimeInterval.Value.Datetime>
    <sparql>SELECT ?procedureDateTime ?value WHERE {?pt a salus:Patient.
    ?pt salus:procedure ?procedure.?procedure salus:procedureDateTime
    ?procedureDateTime.?procedureDateTime salus:value ?value.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Procedure.TimeInterval.Value.Datetime>
<!--Patient.Procedure.Indication.TimeInterval-->
<Patient.Procedure.Indication.TimeInterval.High.Datetime>
    <sparql>SELECT ?problemDate ?high WHERE {?pt a salus:Patient.?pt
    salus:procedure ?procedure.?procedure salus:indication ?indication.
    ?indication salus:problemDate ?problemDate.?problemDate salus:high
    ?high.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Procedure.Indication.TimeInterval.High.Datetime>
<Patient.Procedure.Indication.TimeInterval.Low.Datetime>
    <sparql>SELECT ?problemDate ?low WHERE {?pt a salus:Patient.?pt
    salus:procedure ?procedure.?procedure salus:indication ?indication.
    ?indication salus:problemDate ?problemDate.?problemDate salus:low
```

```
    ?low.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Procedure.Indication.TimeInterval.Low.Datetime>
<Patient.Procedure.Indication.TimeInterval.Value.Datetime>
    <sparql>SELECT ?problemDate ?value WHERE {?pt a salus:Patient.?pt
    salus:procedure ?procedure.?procedure salus:indication ?indication.
    ?indication salus:problemDate ?problemDate.?problemDate salus:value
    ?value.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Procedure.Indication.TimeInterval.Value.Datetime>
<!--Patient.Procedure.Indication.TimeOfDeath-->
<Patient.Procedure.Indication.TimeOfDeath.Datetime>
    <sparql>SELECT ?indication ?timeOfDeath WHERE {?pt a salus:Patient.?pt
    salus:procedure ?procedure.?procedure salus:indication ?indication.
    ?indication salus:timeOfDeath ?timeOfDeath.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Procedure.Indication.TimeOfDeath.Datetime>
<!--Patient.Procedure.Indication.Comment-->
<Patient.Procedure.Indication.Comment.String>
    <sparql>SELECT ?indication ?comment WHERE {?pt a salus:Patient.?pt
    salus:procedure ?procedure.?procedure salus:indication ?indication.
    ?indication salus:comment ?comment.}</sparql>
    <method>Delete Value</method>
</Patient.Procedure.Indication.Comment.String>
<!--Patient.Procedure.Comment-->
<Patient.Procedure.Comment.String>
    <sparql>SELECT ?procedure ?comment WHERE {?pt a salus:Patient.?pt
    salus:procedure ?procedure.?procedure salus:comment ?comment.}
    </sparql>
    <method>Delete Value</method>
</Patient.Procedure.Comment.String>
<!--Patient.Result.TimeInterval-->
<Patient.Result.TimeInterval.High.Datetime>
    <sparql>SELECT ?resultDateTime ?high WHERE {?pt a salus:Patient.?pt
    salus:result ?result.?result salus:resultDateTime ?resultDateTime.
    ?resultDateTime salus:high ?high.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Result.TimeInterval.High.Datetime>
<Patient.Result.TimeInterval.Low.Datetime>
    <sparql>SELECT ?resultDateTime ?low WHERE {?pt a salus:Patient.
    ?pt salus:result ?result.?result salus:resultDateTime ?resultDateTime.
    ?resultDateTime salus:low ?low.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Result.TimeInterval.Low.Datetime>
<Patient.Result.TimeInterval.Value.Datetime>
    <sparql>SELECT ?resultDateTime ?value WHERE {?pt a salus:Patient.?pt
    salus:result ?result.?result salus:resultDateTime ?resultDateTime.
    ?resultDateTime salus:value ?value.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Result.TimeInterval.Value.Datetime>
<!--Patient.Result.RelatedCondition.TimeInterval-->
<Patient.Result.RelatedCondition.TimeInterval.High.Datetime>
    <sparql>SELECT ?problemDate ?high WHERE {?pt a salus:Patient.?pt
    salus:result ?result.?result salus:relatedCondition ?relatedCondition.
    ?relatedCondition salus:problemDate ?problemDate.?problemDate
    salus:high ?high.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Result.RelatedCondition.TimeInterval.High.Datetime>
<Patient.Result.RelatedCondition.TimeInterval.Low.Datetime>
    <sparql>SELECT ?problemDate ?low WHERE {?pt a salus:Patient.?pt
    salus:result ?result.?result salus:relatedCondition ?relatedCondition.
    ?relatedCondition salus:problemDate ?problemDate.?problemDate
    salus:low ?low.}</sparql>
    <method>Shift Date RDF</method>
</Patient.Result.RelatedCondition.TimeInterval.Low.Datetime>
<Patient.Result.RelatedCondition.TimeInterval.Value.Datetime>
    <sparql>SELECT ?problemDate ?value WHERE {?pt a salus:Patient.?pt
    salus:result ?result.?result salus:relatedCondition ?relatedCondition.
    ?relatedCondition salus:problemDate ?problemDate.?problemDate
    salus:value ?value.}</sparql>
```

```
    <method>Shift Date RDF</method>
</Patient.Result.RelatedCondition.TimeInterval.Value.Datetime>
<!--Patient.Result.RelatedCondition.TimeOfDeath-->
<Patient.Result.RelatedCondition.TimeOfDeath.Datetime>
    <sparql>SELECT ?relatedCondition ?timeOfDeath WHERE {?pt a
    salus:Patient.?pt salus:result ?result.?result salus:relatedCondition
    ?relatedCondition.?relatedCondition salus:timeOfDeath ?timeOfDeath.}
    </sparql>
    <method>Shift Date RDF</method>
</Patient.Result.RelatedCondition.TimeOfDeath.Datetime>
<!--Patient.Result.RelatedCondition.Comment-->
<Patient.Result.RelatedCondition.Comment.String>
    <sparql>SELECT ?relatedCondition ?comment WHERE {?pt a salus:Patient.?pt
    salus:result ?result.?result salus:relatedCondition
    ?relatedCondition.?relatedCondition salus:comment ?comment.}</sparql>
    <method>Delete Value</method>
</Patient.Result.RelatedCondition.Comment.String>
<!--Patient.Result.Comment-->
<Patient.Result.Comment.String>
    <sparql>SELECT ?result ?comment WHERE {?pt a salus:Patient.?pt
    salus:result ?result.?result salus:comment ?comment.}</sparql>
    <method>Delete Value</method>
</Patient.Result.Comment.String>
<!--Patient.VitalSign.TimeInterval-->
<Patient.VitalSign.TimeInterval.High.Datetime>
    <sparql>SELECT ?resultDateTime ?high WHERE {?pt a salus:Patient.
    ?pt salus:vitalSign ?vitalSign.?vitalSign salus:resultDateTime
    ?resultDateTime.?resultDateTime salus:high ?high.}</sparql>
    <method>Shift Date RDF</method>
</Patient.VitalSign.TimeInterval.High.Datetime>
<Patient.VitalSign.TimeInterval.Low.Datetime>
    <sparql>SELECT ?resultDateTime ?low WHERE {?pt a salus:Patient.
    ?pt salus:vitalSign ?vitalSign.?vitalSign salus:resultDateTime
    ?resultDateTime.?resultDateTime salus:low ?low.}</sparql>
    <method>Shift Date RDF</method>
</Patient.VitalSign.TimeInterval.Low.Datetime>
<Patient.VitalSign.TimeInterval.Value.Datetime>
    <sparql>SELECT ?resultDateTime ?value WHERE {?pt a salus:Patient.
    ?pt salus:vitalSign ?vitalSign.?vitalSign salus:resultDateTime
    ?resultDateTime.?resultDateTime salus:value ?value.}</sparql>
    <method>Shift Date RDF</method>
</Patient.VitalSign.TimeInterval.Value.Datetime>
<!--Patient.VitalSign.RelatedCondition.TimeInterval-->
<Patient.VitalSign.RelatedCondition.TimeInterval.High.Datetime>
    <sparql>SELECT ?problemDate ?high WHERE {?pt a salus:Patient.?pt
    salus:vitalSign ?vitalSign.?vitalSign salus:relatedCondition
    ?relatedCondition.?relatedCondition salus:problemDate
    ?problemDate.?problemDate salus:high ?high.}</sparql>
    <method>Shift Date RDF</method>
</Patient.VitalSign.RelatedCondition.TimeInterval.High.Datetime>
<Patient.VitalSign.RelatedCondition.TimeInterval.Low.Datetime>
    <sparql>SELECT ?problemDate ?low WHERE {?pt a salus:Patient.?pt
    salus:vitalSign ?vitalSign.?vitalSign salus:relatedCondition
    ?relatedCondition.?relatedCondition salus:problemDate
    ?problemDate.?problemDate salus:low ?low.}</sparql>
    <method>Shift Date RDF</method>
</Patient.VitalSign.RelatedCondition.TimeInterval.Low.Datetime>
<Patient.VitalSign.RelatedCondition.TimeInterval.Value.Datetime>
    <sparql>SELECT ?problemDate ?value WHERE {?pt a salus:Patient.
    ?pt salus:vitalSign ?vitalSign.?vitalSign salus:relatedCondition
    ?relatedCondition.?relatedCondition salus:problemDate ?problemDate.
    ?problemDate salus:value ?value.}</sparql>
    <method>Shift Date RDF</method>
</Patient.VitalSign.RelatedCondition.TimeInterval.Value.Datetime>
<!--Patient.VitalSign.RelatedCondition.TimeOfDeath-->
<Patient.VitalSign.RelatedCondition.TimeOfDeath.Datetime>
    <sparql>SELECT ?relatedCondition ?timeOfDeath WHERE {?pt a
    salus:Patient.?pt salus:vitalSign ?vitalSign.?vitalSign
    salus:relatedCondition ?relatedCondition.?relatedCondition
    salus:timeOfDeath ?timeOfDeath.}</sparql>
```

```
    <method>Shift Date RDF</method>
</Patient.VitalSign.RelatedCondition.TimeOfDeath.Datetime>
<!--Patient.VitalSign.RelatedCondition.Comment-->
<Patient.VitalSign.RelatedCondition.Comment.String>
    <sparql>SELECT ?relatedCondition ?comment WHERE {?pt a
    salus:Patient.?pt salus:vitalSign ?vitalSign.?vitalSign
    salus:relatedCondition ?relatedCondition.?relatedCondition
    salus:comment ?comment.}</sparql>
    <method>Delete Value</method>
</Patient.VitalSign.RelatedCondition.Comment.String>
<!--Patient.VitalSign.Comment-->
<Patient.VitalSign.Comment.String>
    <sparql>SELECT ?vitalSign ?comment WHERE {?pt a salus:Patient.?pt
    salus:vitalSign ?vitalSign.?vitalSign salus:comment ?comment.}</sparql>
    <method>Delete Value</method>
</Patient.VitalSign.Comment.String>
<!--Pass through or Generalize for Rare Conditions-->
<!--Patient.Condition.ProblemName-->
<Patient.Condition.ProblemName.String>
    <sparql>SELECT ?condition ?problemName WHERE {?pt a salus:Patient.?pt
    salus:condition ?condition.?condition salus:problemName ?problemName.}
    </sparql>
    <method>Rare Process</method>
</Patient.Condition.ProblemName.String>
<!--Patient.Condition.ProblemCode-->
<Patient.Condition.ProblemCode.Code.String>
    <sparql>SELECT ?problemCode ?code WHERE {?pt a salus:Patient.?pt
    salus:condition ?condition.?condition salus:problemCode ?problemCode.
    ?problemCode salus:code ?code.}</sparql>
    <method>Rare Process</method>
</Patient.Condition.ProblemCode.Code.String>
<Patient.Condition.ProblemCode.DisplayName.String>
    <sparql>SELECT ?problemCode ?displayName WHERE {?pt a salus:Patient.
    ?pt salus:condition ?condition.?condition salus:problemCode
    ?problemCode.?problemCode salus:displayName ?displayName.}
    </sparql>
    <method>Rare Process</method>
</Patient.Condition.ProblemCode.DisplayName.String>
<!--Patient.Immunization.MedicationInformation.ProductName-->
<Patient.Immunization.MedicationInformation.ProductName.Code.String>
    <sparql>SELECT ?codedProductName ?code WHERE {?pt a salus:Patient.?pt
    salus:immunization ?immunization.?immunization
    salus:medicationInformation ?medicationInformation.
    ?medicationInformation salus:codedProductName ?codedProductName.
    ?codedProductName salus:code ?code.}</sparql>
    <method>Rare Process</method>
</Patient.Immunization.MedicationInformation.ProductName.Code.String>
<Patient.Immunization.MedicationInformation.ProductName.DisplayName.String>
    <sparql>SELECT ?codedProductName ?displayName WHERE {?pt a salus:Patient.
    ?pt salus:immunization ?immunization.?immunization
    salus:medicationInformation ?medicationInformation.?medicationInformation
    salus:codedProductName ?codedProductName.?codedProductName
    salus:displayName ?displayName.}</sparql>
    <method>Rare Process</method>
</Patient.Immunization.MedicationInformation.ProductName.DisplayName.String>
<!--Patient.Immunization.MedicationInformation.ProductName-->
<Patient.Immunization.MedicationInformation.ProductName.String>
    <sparql>SELECT ?medicationInformation ?codedProductName WHERE {?pt a
    salus:Patient.?pt salus:immunization ?immunization.?immunization
    salus:medicationInformation ?medicationInformation.?medicationInformation
    salus:codedProductName ?codedProductName.}
    </sparql>
    <method>Rare Process</method>
</Patient.Immunization.MedicationInformation.ProductName.String>
<!--Patient.Immunization.MedicationInformation.ActiveIngredient-->
<Patient.Immunization.MedicationInformation.ActiveIngredient.Code.String>
    <sparql>SELECT ?codedActiveIngredient ?code WHERE {?pt a salus:Patient.
    ?pt salus:immunization ?immunization.?immunization
    salus:medicationInformation ?medicationInformation.?medicationInformation
    salus:codedActiveIngredient ?codedActiveIngredient.
```

```
    ?codedActiveIngredient salus:code ?code.}</sparql>
    <method>Rare Process</method>
</Patient.Immunization.MedicationInformation.ActiveIngredient.Code.String>
<Patient.Immunization.MedicationInformation.ActiveIngredient.DisplayName.String>
    <sparql>SELECT ?codedActiveIngredient ?displayName WHERE {?pt a
    salus:Patient.?pt salus:immunization ?immunization.?immunization
    salus:medicationInformation ?medicationInformation.?medicationInformation
    salus:codedActiveIngredient ?codedActiveIngredient.?codedActiveIngredient
    salus:displayName ?displayName.}</sparql>
    <method>Rare Process</method>
</Patient.Immunization.MedicationInformation.ActiveIngredient.DisplayName.String>
<!--Patient.Immunization.MedicationInformation.BrandName-->
<Patient.Immunization.MedicationInformation.BrandName.Code.String>
    <sparql>SELECT ?codedBrandName ?code WHERE {?pt a salus:Patient.?pt
    salus:immunization ?immunization.?immunization salus:medicationInformation
    ?medicationInformation.?medicationInformation salus:codedBrandName
    ?codedBrandName.?codedBrandName salus:code ?code.}</sparql>
    <method>Rare Process</method>
</Patient.Immunization.MedicationInformation.BrandName.Code.String>
<Patient.Immunization.MedicationInformation.BrandName.DisplayName.String>
    <sparql>SELECT ?codedBrandName ?displayName WHERE {?pt a salus:Patient.
    ?pt salus:immunization ?immunization.?immunization
    salus:medicationInformation ?medicationInformation.?medicationInformation
    salus:codedBrandName ?codedBrandName.?codedBrandName salus:displayName
    ?displayName.}</sparql>
    <method>Rare Process</method>
</Patient.Immunization.MedicationInformation.BrandName.DisplayName.String>
<!--Patient.Immunization.MedicationInformation.BrandName-->
<Patient.Immunization.MedicationInformation.BrandName.String>
    <sparql>SELECT ?medicationInformation ?codedBrandName WHERE {?pt a
    salus:Patient.?pt salus:immunization ?immunization.?immunization
    salus:medicationInformation ?medicationInformation.
    ?medicationInformation salus:codedBrandName ?codedBrandName.}</sparql>
    <method>Rare Process</method>
</Patient.Immunization.MedicationInformation.BrandName.String>
    <!--Patient.Medication.MedicationInformation.ProductName-->
<Patient.Medication.MedicationInformation.ProductName.Code.String>
    <sparql>SELECT ?codedProductName ?code WHERE {?pt a salus:Patient.?pt
    salus:medication ?medication.?medication salus:medicationInformation
    ?medicationInformation.?medicationInformation salus:codedProductName
    ?codedProductName.?codedProductName salus:code ?code.}</sparql>
    <method>Rare Process</method>
</Patient.Medication.MedicationInformation.ProductName.Code.String>
<Patient.Medication.MedicationInformation.ProductName.DisplayName.String>
    <sparql>SELECT ?codedProductName ?displayName WHERE {?pt a salus:Patient.
    ?pt salus:medication ?medication.?medication salus:medicationInformation
    ?medicationInformation.?medicationInformation salus:codedProductName
    ?codedProductName.?codedProductName salus:displayName ?displayName.}
    </sparql>
    <method>Rare Process</method>
</Patient.Medication.MedicationInformation.ProductName.DisplayName.String>
<!--Patient.Medication.MedicationInformation.ProductName-->
<Patient.Medication.MedicationInformation.ProductName.String>
    <sparql>SELECT ?medicationInformation ?codedProductName WHERE {?pt a
    salus:Patient.?pt salus:medication ?medication.?medication
    salus:medicationInformation ?medicationInformation. ?medicationInformation
    salus:codedProductName ?codedProductName.}</sparql>
    <method>Rare Process</method>
</Patient.Medication.MedicationInformation.ProductName.String>
<!--Patient.Medication.MedicationInformation.ActiveIngredient-->
<Patient.Medication.MedicationInformation.ActiveIngredient.Code.String>
    <sparql>SELECT ?codedActiveIngredient ?code WHERE {?pt a salus:Patient.
    ?pt salus:medication ?medication.?medication salus:medicationInformation
    ?medicationInformation.?medicationInformation salus:codedActiveIngredient
    ?codedActiveIngredient. ?codedActiveIngredient salus:code ?code.}
    </sparql>
    <method>Rare Process</method>
</Patient.Medication.MedicationInformation.ActiveIngredient.Code.String>
<Patient.Medication.MedicationInformation.ActiveIngredient.DisplayName.String>
    <sparql>SELECT ?codedActiveIngredient ?displayName WHERE {?pt a
```

```
    salus:Patient.?pt salus:medication ?medication.?medication
    salus:medicationInformation ?medicationInformation.
    ?medicationInformation salus:codedActiveIngredient ?codedActiveIngredient.
    ?codedActiveIngredient salus:displayName ?displayName.}</sparql>
    <method>Rare Process</method>
</Patient.Medication.MedicationInformation.ActiveIngredient.DisplayName.String>
<!--Patient.Medication.MedicationInformation.BrandName-->
<Patient.Medication.MedicationInformation.BrandName.Code.String>
    <sparql>SELECT ?codedBrandName ?code WHERE {?pt a salus:Patient.?pt
    salus:medication ?medication.?medication salus:medicationInformation
    ?medicationInformation.?medicationInformation salus:codedBrandName
    ?codedBrandName.?codedBrandName salus:code ?code.}</sparql>
    <method>Rare Process</method>
</Patient.Medication.MedicationInformation.BrandName.Code.String>
<Patient.Medication.MedicationInformation.BrandName.DisplayName.String>
    <sparql>SELECT ?codedBrandName ?displayName WHERE {?pt a salus:Patient.
    ?pt salus:medication ?medication.?medication salus:medicationInformation
    ?medicationInformation.?medicationInformation salus:codedBrandName
    ?codedBrandName.?codedBrandName salus:displayName ?displayName.}</sparql>
    <method>Rare Process</method>
</Patient.Medication.MedicationInformation.BrandName.DisplayName.String>
<!--Patient.Medication.MedicationInformation.BrandName-->
<Patient.Medication.MedicationInformation.BrandName.String>
    <sparql>SELECT ?medicationInformation ?codedBrandName WHERE {?pt a
    salus:Patient.?pt salus:medication ?medication.?medication
    salus:medicationInformation ?medicationInformation.?medicationInformation
    salus:codedBrandName ?codedBrandName.}
    </sparql>
    <method>Rare Process</method>
</Patient.Medication.MedicationInformation.BrandName.String>
<!--Patient.Medication.Indication.ProblemName-->
<Patient.Medication.Indication.ProblemName.String>
    <sparql>SELECT ?indication ?problemName WHERE {?pt a salus:Patient.?pt
    salus:medication ?medication.?medication salus:indication ?indication.
    ?indication salus:problemName
    ?problemName.}</sparql>
    <method>Rare Process</method>
</Patient.Medication.Indication.ProblemName.String>
<!--Patient.Medication.Indication.ProblemCode-->
<Patient.Medication.Indication.ProblemCode.Code.String>
    <sparql>SELECT ?problemCode ?code WHERE {?pt a salus:Patient.?pt
    salus:medication ?medication.?medication salus:indication
    ?indication. ?indication salus:problemCode
    ?problemCode.?problemCode salus:code ?code.}</sparql>
    <method>Rare Process</method>
</Patient.Medication.Indication.ProblemCode.Code.String>
<Patient.Medication.Indication.ProblemCode.DisplayName.String>
    <sparql>SELECT ?problemCode ?displayName WHERE {?pt a salus:Patient.
    ?pt salus:medication ?medication.?medication salus:indication ?indication.
    ?indication salus:problemCode ?problemCode.?problemCode
    salus:displayName ?displayName.}</sparql>
    <method>Rare Process</method>
</Patient.Medication.Indication.ProblemCode.DisplayName.String>
<!--Patient.Procedure.Type-->
<Patient.Procedure.Type.Code.String>
    <sparql>SELECT ?procedureType ?code WHERE {?pt a salus:Patient.?pt
    salus:procedure ?procedure.?procedure salus:procedureType
    ?procedureType.?procedureType salus:code ?code.}
    </sparql>
    <method>Rare Process</method>
</Patient.Procedure.Type.Code.String>
<Patient.Procedure.Type.DisplayName.String>
    <sparql>SELECT ?procedureType ?displayName WHERE {?pt a
    salus:Patient.?pt salus:procedure ?procedure.?procedure
    salus:procedureType ?procedureType.?procedureType salus:displayName
    ?displayName.}</sparql>
    <method>Rare Process</method>
</Patient.Procedure.Type.DisplayName.String>
<!--Patient.Procedure.Type-->
<Patient.Procedure.Type.String>
```

```
    <sparql>SELECT ?procedure ?procedureType WHERE {?pt a salus:Patient.
    ?pt salus:procedure ?procedure.?procedure salus:procedureType
    ?procedureType.}</sparql>
    <method>Rare Process</method>
</Patient.Procedure.Type.String>
<!--Patient.Procedure.Indication.ProblemName-->
<Patient.Procedure.Indication.ProblemName.String>
    <sparql>SELECT ?indication ?problemName WHERE {?pt a salus:Patient.
    ?pt salus:procedure ?procedure.?procedure salus:indication ?indication.
    ?indication salus:problemName ?problemName.}
    </sparql>
    <method>Rare Process</method>
</Patient.Procedure.Indication.ProblemName.String>
<!--Patient.Procedure.Indication.ProblemCode-->
<Patient.Procedure.Indication.ProblemCode.Code.String>
    <sparql>SELECT ?problemCode ?code WHERE {?pt a salus:Patient.?pt
    salus:procedure ?procedure.?procedure salus:indication ?indication.
    ?indication salus:problemCode ?problemCode.
    ?problemCode salus:code ?code.}</sparql>
    <method>Rare Process</method>
</Patient.Procedure.Indication.ProblemCode.Code.String>
<Patient.Procedure.Indication.ProblemCode.DisplayName.String>
    <sparql>SELECT ?problemCode ?displayName WHERE {?pt a salus:Patient.
    ?pt salus:procedure ?procedure.?procedure salus:indication
    ?indication. ?indication salus:problemCode ?problemCode.
    ?problemCode salus:displayName ?displayName.}</sparql>
    <method>Rare Process</method>
</Patient.Procedure.Indication.ProblemCode.DisplayName.String>
<!--Patient.Result.Type-->
<Patient.Result.Type.Code.String>
    <sparql>SELECT ?resultType ?code WHERE {?pt a salus:Patient.?pt
    salus:result ?result.?result salus:resultType ?resultType.
    ?resultType salus:code ?code.}</sparql>
    <method>Rare Process</method>
</Patient.Result.Type.Code.String>
<Patient.Result.Type.DisplayName.String>
    <sparql>SELECT ?resultType ?displayName WHERE {?pt a salus:Patient.
    ?pt salus:result ?result.?result salus:resultType ?resultType.
    ?resultType salus:displayName ?displayName.}</sparql>
    <method>Rare Process</method>
</Patient.Result.Type.DisplayName.String>
<!--Patient.Result.RelatedCondition.ProblemName-->
<Patient.Result.RelatedCondition.ProblemName.String>
    <sparql>SELECT ?relatedCondition ?problemName WHERE {?pt a
    salus:Patient.?pt salus:result ?result.?result salus:relatedCondition
    ?relatedCondition.?relatedCondition salus:problemName
    ?problemName.}</sparql>
    <method>Rare Process</method>
</Patient.Result.RelatedCondition.ProblemName.String>
<!--Patient.Result.RelatedCondition.ProblemCode-->
<Patient.Result.RelatedCondition.ProblemCode.Code.String>
    <sparql>SELECT ?problemCode ?code WHERE {?pt a salus:Patient.?pt
    salus:result ?result.?result salus:relatedCondition ?relatedCondition.
    ?relatedCondition salus:problemCode ?problemCode.
    ?problemCode salus:code ?code.}</sparql>
    <method>Rare Process</method>
</Patient.Result.RelatedCondition.ProblemCode.Code.String>
<Patient.Result.RelatedCondition.ProblemCode.DisplayName.String>
    <sparql>SELECT ?problemCode ?displayName WHERE {?pt a salus:Patient.
    ?pt salus:result ?result.?result salus:relatedCondition
    ?relatedCondition. ?relatedCondition salus:problemCode ?problemCode.
    ?problemCode salus:displayName ?displayName.}</sparql>
    <method>Rare Process</method>
</Patient.Result.RelatedCondition.ProblemCode.DisplayName.String>
<!--Patient.VitalSign.Type-->
<Patient.VitalSign.Type.Code.String>
    <sparql>SELECT ?resultType ?code WHERE {?pt a salus:Patient.?pt
    salus:vitalSign ?vitalSign.?vitalSign salus:resultType
    ?resultType.?resultType salus:code ?code.}</sparql>
    <method>Rare Process</method>
```

```
    </Patient.VitalSign.Type.Code.String>
    <Patient.VitalSign.Type.DisplayName.String>
        <sparql>SELECT ?resultType ?displayName WHERE {?pt a salus:Patient.
        ?pt salus:vitalSign ?vitalSign.?vitalSign salus:resultType
        ?resultType.?resultType salus:displayName
        ?displayName.}</sparql>
        <method>Rare Process</method>
    </Patient.VitalSign.Type.DisplayName.String>
    <!--Patient.VitalSign.RelatedCondition.ProblemName-->
    <Patient.VitalSign.RelatedCondition.ProblemName.String>
        <sparql>SELECT ?relatedCondition ?problemName WHERE {?pt a
        salus:Patient.?pt salus:vitalSign ?vitalSign.?vitalSign
        salus:relatedCondition ?relatedCondition.
        ?relatedCondition salus:problemName ?problemName.}</sparql>
        <method>Rare Process</method>
    </Patient.VitalSign.RelatedCondition.ProblemName.String>
    <!--Patient.VitalSign.RelatedCondition.ProblemCode-->
    <Patient.VitalSign.RelatedCondition.ProblemCode.Code.String>
        <sparql>SELECT ?problemCode ?code WHERE {?pt a salus:Patient.?pt
        salus:vitalSign ?vitalSign.?vitalSign salus:relatedCondition
        ?relatedCondition. ?relatedCondition
        salus:problemCode ?problemCode.?problemCode salus:code ?code.}
        </sparql>
        <method>Rare Process</method>
    </Patient.VitalSign.RelatedCondition.ProblemCode.Code.String>
    <Patient.VitalSign.RelatedCondition.ProblemCode.DisplayName.String>
        <sparql>SELECT ?problemCode ?displayName WHERE {?pt a salus:Patient.
        ?pt salus:vitalSign ?vitalSign.?vitalSign salus:relatedCondition
        ?relatedCondition. ?relatedCondition
        salus:problemCode ?problemCode.?problemCode salus:displayName
        ?displayName.}</sparql>
        <method>Rare Process</method>
    </Patient.VitalSign.RelatedCondition.ProblemCode.DisplayName.String>
</rdfconf>
```

APPENDIX D

HL7 CDA PCC/CCD TEMPLATES CONFIGURATION FILE

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<cdaconf>
    <Patient.ID.Extension.String>
        <xpath>/ClinicalDocument/recordTarget/patientRole/id/@extension
        </xpath>
        <method>Substitute Pseudonym</method>
    </Patient.ID.Extension.String>
    <Patient.Allergy.TimeInterval.High.Datetime>
        <xpath>/ClinicalDocument/component/structuredBody/component/
        section/entry/act/entryRelationship/observation/effectiveTime/high
        /@value</xpath>
        <method>Shift Date CDA</method>
    </Patient.Allergy.TimeInterval.High.Datetime>
    <Patient.Allergy.TimeInterval.Low.Datetime>
        <xpath>/ClinicalDocument/component/structuredBody/component/section
        /entry/act/entryRelationship/observation/effectiveTime/low/@value
        </xpath>
        <method>Shift Date CDA</method>
    </Patient.Allergy.TimeInterval.Low.Datetime>
    <Patient.Allergy.Comment.String>
        <xpath>/ClinicalDocument/component/structuredBody/component/section
        /entry/act/entryRelationship/observation/text/reference/@value
        </xpath>
        <method>Delete Value</method>
    </Patient.Allergy.Comment.String>
    <Patient.Condition.TimeInterval.High.Datetime>
        <xpath>/ClinicalDocument/component/structuredBody/component/section
        /entry/act/entryRelationship/observation/effectiveTime/high/@value
        </xpath>
        <method>Shift Date CDA</method>
    </Patient.Condition.TimeInterval.High.Datetime>
    <Patient.Condition.TimeInterval.Low.Datetime>
        <xpath>/ClinicalDocument/component/structuredBody/component/section
        /entry/act/entryRelationship/observation/effectiveTime/low/@value
        </xpath>
        <method>Shift Date CDA</method>
    </Patient.Condition.TimeInterval.Low.Datetime>
    <Patient.Condition.TimeOfDeath.Datetime>
        <xpath>/ClinicalDocument/component/structuredBody/component/section
        /entry/act/entryRelationship/observation/entryRelationship/observation
        /effectiveTime/@value</xpath>
        <method>Shift Date CDA</method>
    </Patient.Condition.TimeOfDeath.Datetime>
    <Patient.Condition.Comment.String>
        <xpath>/ClinicalDocument/component/structuredBody/component/section
        /entry/act/entryRelationship/observation/text/reference/@value
        </xpath>
        <method>Delete Value</method>
    </Patient.Condition.Comment.String>
    <Patient.Immunization.AdministeredDate.Datetime>
```

```
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/entryRelationship/observation
    /effectiveTime/@value</xpath>
    <method>Shift Date CDA</method>
</Patient.Immunization.AdministeredDate.Datetime>
<Patient.Immunization.Comment.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/text/reference/@value</xpath>
    <method>Delete Value</method>
</Patient.Immunization.Comment.String>
<Patient.Medication.TimeInterval.High.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/effectiveTime/high/@value
    </xpath>
    <method>Shift Date CDA</method>
</Patient.Medication.TimeInterval.High.Datetime>
<Patient.Medication.TimeInterval.Low.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/effectiveTime/low/@value
    </xpath>
    <method>Shift Date CDA</method>
</Patient.Medication.TimeInterval.Low.Datetime>
<Patient.Medication.AdministeredTiming.Value.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/entryRelationship/observation
    /effectiveTime/@value</xpath>
    <method>Shift Date CDA</method>
</Patient.Medication.AdministeredTiming.Value.Datetime>
<Patient.Medication.AdministeredTiming.Phase.High.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section/
    entry/act/entryRelationship/observation/effectiveTime/high/@value
    </xpath>
    <method>Shift Date CDA</method>
</Patient.Medication.AdministeredTiming.Phase.High.Datetime>
<Patient.Medication.AdministeredTiming.Phase.Low.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/effectiveTime/low/@value
    </xpath>
    <method>Shift Date CDA</method>
</Patient.Medication.AdministeredTiming.Phase.Low.Datetime>
<Patient.Medication.Indication.TimeInterval.High.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/effectiveTime/high/@value
    </xpath>
    <method>Shift Date CDA</method>
</Patient.Medication.Indication.TimeInterval.High.Datetime>
<Patient.Medication.Indication.TimeInterval.Low.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/effectiveTime/low/@value
    </xpath>
    <method>Shift Date CDA</method>
</Patient.Medication.Indication.TimeInterval.Low.Datetime>
<Patient.Medication.Indication.TimeOfDeath.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/entryRelationship/observation
    /effectiveTime/@value</xpath>
    <method>Shift Date CDA</method>
</Patient.Medication.Indication.TimeOfDeath.Datetime>
<Patient.Medication.Indication.Comment.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/text/reference/@value
    </xpath>
    <method>Delete Value</method>
</Patient.Medication.Indication.Comment.String>
<Patient.Medication.Order.ExpirationDateTime.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/entryRelationship/
    observation/effectiveTime/@value</xpath>
    <method>Shift Date CDA</method>
</Patient.Medication.Order.ExpirationDateTime.Datetime>
```

```
<Patient.Medication.Order.DateTime.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/entryRelationship/
    observation/effectiveTime/@value</xpath>
    <method>Shift Date CDA</method>
</Patient.Medication.Order.DateTime.Datetime>
<Patient.Medication.FulfillmentHistory.DispenseDate.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/entryRelationship
    /observation/effectiveTime/@value</xpath>
    <method>Shift Date CDA</method>
</Patient.Medication.FulfillmentHistory.DispenseDate.Datetime>
<Patient.Medication.Comment.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/text/reference/@value</xpath>
    <method>Delete Value</method>
</Patient.Medication.Comment.String>
<Patient.Pregnancy.ObservationDate.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/entryRelationship/
    observation/effectiveTime/@value</xpath>
    <method>Shift Date CDA</method>
</Patient.Pregnancy.ObservationDate.Datetime>
<Patient.Pregnancy.LastMenstrualPeriodDate.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/entryRelationship/
    observation/effectiveTime/@value</xpath>
    <method>Shift Date CDA</method>
</Patient.Pregnancy.LastMenstrualPeriodDate.Datetime>
<Patient.Pregnancy.DeliveryDate.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/entryRelationship/
    observation/effectiveTime/@value</xpath>
    <method>Shift Date CDA</method>
</Patient.Pregnancy.DeliveryDate.Datetime>
<Patient.Pregnancy.Comment.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/text/reference/@value
    </xpath>
    <method>Delete Value</method>
</Patient.Pregnancy.Comment.String>
<Patient.Procedure.TimeInterval.High.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/effectiveTime/high/@value
    </xpath>
    <method>Shift Date CDA</method>
</Patient.Procedure.TimeInterval.High.Datetime>
<Patient.Procedure.TimeInterval.Low.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/effectiveTime/low/@value
    </xpath>
    <method>Shift Date CDA</method>
</Patient.Procedure.TimeInterval.Low.Datetime>
<Patient.Procedure.Indication.TimeInterval.High.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/effectiveTime/high/@value
    </xpath>
    <method>Shift Date CDA</method>
</Patient.Procedure.Indication.TimeInterval.High.Datetime>
<Patient.Procedure.Indication.TimeInterval.Low.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/effectiveTime/low/@value
    </xpath>
    <method>Shift Date CDA</method>
</Patient.Procedure.Indication.TimeInterval.Low.Datetime>
<Patient.Procedure.Indication.TimeOfDeath.Datetime>
    <xpath></xpath>
    <method>Shift Date CDA</method>
</Patient.Procedure.Indication.TimeOfDeath.Datetime>
<Patient.Procedure.Indication.Comment.String>
```

```
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/text/reference/@value
    </xpath>
    <method>Delete Value</method>
</Patient.Procedure.Indication.Comment.String>
<Patient.Procedure.Comment.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/entryRelationship/
    observation/effectiveTime/@value</xpath>
    <method>Delete Value</method>
</Patient.Procedure.Comment.String>
<Patient.Result.TimeInterval.High.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/effectiveTime/high/@value
    </xpath>
    <method>Shift Date CDA</method>
</Patient.Result.TimeInterval.High.Datetime>
<Patient.Result.TimeInterval.Low.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/effectiveTime/low/@value
    </xpath>
    <method>Shift Date CDA</method>
</Patient.Result.TimeInterval.Low.Datetime>
<Patient.Result.RelatedCondition.TimeInterval.High.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/effectiveTime/high/@value
    </xpath>
    <method>Shift Date CDA</method>
</Patient.Result.RelatedCondition.TimeInterval.High.Datetime>
<Patient.Result.RelatedCondition.TimeInterval.Low.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/effectiveTime/low/@value
    </xpath>
    <method>Shift Date CDA</method>
</Patient.Result.RelatedCondition.TimeInterval.Low.Datetime>
<Patient.Result.RelatedCondition.TimeOfDeath.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/entryRelationship
    /observation/effectiveTime/@value</xpath>
    <method>Shift Date CDA</method>
</Patient.Result.RelatedCondition.TimeOfDeath.Datetime>
<Patient.Result.RelatedCondition.Comment.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/text/reference/@value
    </xpath>
    <method>Delete Value</method>
</Patient.Result.RelatedCondition.Comment.String>
<Patient.Result.Comment.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/text/reference/@value
    </xpath>
    <method>Delete Value</method>
</Patient.Result.Comment.String>
<Patient.VitalSign.TimeInterval.High.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/effectiveTime/high/@value
    </xpath>
    <method>Shift Date CDA</method>
</Patient.VitalSign.TimeInterval.High.Datetime>
<Patient.VitalSign.TimeInterval.Low.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/effectiveTime/low/@value
    </xpath>
    <method>Shift Date CDA</method>
</Patient.VitalSign.TimeInterval.Low.Datetime>
<Patient.VitalSign.RelatedCondition.TimeInterval.High.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/effectiveTime/high/@value
    </xpath>
    <method>Shift Date CDA</method>
```

```
</Patient.VitalSign.RelatedCondition.TimeInterval.High.Datetime>
<Patient.VitalSign.RelatedCondition.TimeInterval.Low.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/effectiveTime/low/@value
    </xpath>
    <method>Shift Date CDA</method>
</Patient.VitalSign.RelatedCondition.TimeInterval.Low.Datetime>
<Patient.VitalSign.RelatedCondition.TimeOfDeath.Datetime>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/entryRelationship/
    observation/effectiveTime/@value</xpath>
    <method>Shift Date CDA</method>
</Patient.VitalSign.RelatedCondition.TimeOfDeath.Datetime>
<Patient.VitalSign.RelatedCondition.Comment.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/text/reference/@value
    </xpath>
    <method>Delete Value</method>
</Patient.VitalSign.RelatedCondition.Comment.String>
<Patient.VitalSign.Comment.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/text/reference/@value
    </xpath>
    <method>Delete Value</method>
</Patient.VitalSign.Comment.String>
<Patient.Condition.ProblemName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName
    </xpath>
    <method>Rare Process</method>
</Patient.Condition.ProblemName.String>
<Patient.Condition.ProblemCode.Code.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@code</xpath>
    <method>Rare Process</method>
</Patient.Condition.ProblemCode.Code.String>
<Patient.Condition.ProblemCode.DisplayName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Condition.ProblemCode.DisplayName.String>
<Patient.Immunization.MedicationInformation.ProductName.Code.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@code</xpath>
    <method>Rare Process</method>
</Patient.Immunization.MedicationInformation.ProductName.Code.String>
<Patient.Immunization.MedicationInformation.ProductName.DisplayName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Immunization.MedicationInformation.ProductName.DisplayName.String>
<Patient.Immunization.MedicationInformation.ProductName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Immunization.MedicationInformation.ProductName.String>
<Patient.Immunization.MedicationInformation.ActiveIngredient.Code.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@code</xpath>
    <method>Rare Process</method>
</Patient.Immunization.MedicationInformation.ActiveIngredient.Code.String>
<Patient.Immunization.MedicationInformation.ActiveIngredient.DisplayName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Immunization.MedicationInformation.ActiveIngredient.DisplayName.String>
<Patient.Immunization.MedicationInformation.BrandName.Code.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@code</xpath>
    <method>Rare Process</method>
```

```
</Patient.Immunization.MedicationInformation.BrandName.Code.String>
<Patient.Immunization.MedicationInformation.BrandName.DisplayName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Immunization.MedicationInformation.BrandName.DisplayName.String>
<Patient.Immunization.MedicationInformation.BrandName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Immunization.MedicationInformation.BrandName.String>
<Patient.Medication.MedicationInformation.ProductName.Code.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Medication.MedicationInformation.ProductName.Code.String>
<Patient.Medication.MedicationInformation.ProductName.DisplayName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/
    section/entry/act/entryRelationship/observation/code/@displayName
    </xpath>
    <method>Rare Process</method>
</Patient.Medication.MedicationInformation.ProductName.DisplayName.String>
<Patient.Medication.MedicationInformation.ProductName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Medication.MedicationInformation.ProductName.String>
<Patient.Medication.MedicationInformation.ActiveIngredient.Code.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@code</xpath>
    <method>Rare Process</method>
</Patient.Medication.MedicationInformation.ActiveIngredient.Code.String>
<Patient.Medication.MedicationInformation.ActiveIngredient.DisplayName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Medication.MedicationInformation.ActiveIngredient.DisplayName.String>
<Patient.Medication.MedicationInformation.BrandName.Code.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@code</xpath>
    <method>Rare Process</method>
</Patient.Medication.MedicationInformation.BrandName.Code.String>
<Patient.Medication.MedicationInformation.BrandName.DisplayName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Medication.MedicationInformation.BrandName.DisplayName.String>
<Patient.Medication.MedicationInformation.BrandName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Medication.MedicationInformation.BrandName.String>
<Patient.Medication.Indication.ProblemName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Medication.Indication.ProblemName.String>
<Patient.Medication.Indication.ProblemCode.Code.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section/
    entry/act/entryRelationship/observation/code/@code</xpath>
    <method>Rare Process</method>
</Patient.Medication.Indication.ProblemCode.Code.String>
<Patient.Medication.Indication.ProblemCode.DisplayName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Medication.Indication.ProblemCode.DisplayName.String>
<Patient.Procedure.Type.Code.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@code</xpath>
```

```
    <method>Rare Process</method>
</Patient.Procedure.Type.Code.String>
<Patient.Procedure.Type.DisplayName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Procedure.Type.DisplayName.String>
<Patient.Procedure.Type.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@code</xpath>
    <method>Rare Process</method>
</Patient.Procedure.Type.String>
<Patient.Procedure.Indication.ProblemName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Procedure.Indication.ProblemName.String>
<Patient.Procedure.Indication.ProblemCode.Code.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@code</xpath>
    <method>Rare Process</method>
</Patient.Procedure.Indication.ProblemCode.Code.String>
<Patient.Procedure.Indication.ProblemCode.DisplayName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Procedure.Indication.ProblemCode.DisplayName.String>
<Patient.Result.Type.Code.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@code</xpath>
    <method>Rare Process</method>
</Patient.Result.Type.Code.String>
<Patient.Result.Type.DisplayName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Result.Type.DisplayName.String>
<Patient.Result.RelatedCondition.ProblemName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Result.RelatedCondition.ProblemName.String>
<Patient.Result.RelatedCondition.ProblemCode.Code.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@code</xpath>
    <method>Rare Process</method>
</Patient.Result.RelatedCondition.ProblemCode.Code.String>
<Patient.Result.RelatedCondition.ProblemCode.DisplayName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.Result.RelatedCondition.ProblemCode.DisplayName.String>
<Patient.VitalSign.Type.Code.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@code</xpath>
    <method>Rare Process</method>
</Patient.VitalSign.Type.Code.String>
<Patient.VitalSign.Type.DisplayName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.VitalSign.Type.DisplayName.String>
<Patient.VitalSign.RelatedCondition.ProblemName.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@displayName</xpath>
    <method>Rare Process</method>
</Patient.VitalSign.RelatedCondition.ProblemName.String>
<Patient.VitalSign.RelatedCondition.ProblemCode.Code.String>
    <xpath>/ClinicalDocument/component/structuredBody/component/section
    /entry/act/entryRelationship/observation/code/@code</xpath>
```

```
        <method>Rare Process</method>
    </Patient.VitalSign.RelatedCondition.ProblemCode.Code.String>
    <Patient.VitalSign.RelatedCondition.ProblemCode.DisplayName.String>
        <xpath>/ClinicalDocument/component/structuredBody/component/section
        /entry/act/entryRelationship/observation/code/@displayName</xpath>
        <method>Rare Process</method>
    </Patient.VitalSign.RelatedCondition.ProblemCode.DisplayName.String>
</cdaconf>
```

APPENDIX E

ICH E2B TEMPLATES CONFIGURATION FILE

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<icsrconf>
    <patientinitial>
        <xpath>ichicsr/safetyreport/patient/patientinitial</xpath>
        <method>Write Anonymized</method>
    </patientinitial>
    <!--Patient GP Medical Record Number-->
    <patientgpmedicalrecordnumb>
        <xpath>ichicsr/safetyreport/patient/patientgpmedicalrecordnumb
        </xpath>
        <method>Substitute Pseudonym</method>
    </patientgpmedicalrecordnumb>
    <patientspecialistrecordnumb>
        <xpath>ichicsr/safetyreport/patient/patientspecialistrecordnumb
        </xpath>
        <method>Substitute Pseudonym</method>
    </patientspecialistrecordnumb>
    <patienthospitalrecordnumb>
        <xpath>ichicsr/safetyreport/patient/patienthospitalrecordnumb
        </xpath>
        <method>Substitute Pseudonym</method>
    </patienthospitalrecordnumb>
    <patientinvestigationnumb>
        <xpath>ichicsr/safetyreport/patient/patientinvestigationnumb
        </xpath>
        <method>Substitute Pseudonym</method>
    </patientinvestigationnumb>
    <patientbirthdate>
        <xpath>ichicsr/safetyreport/patient/patientbirthdate</xpath>
        <method>Generalize Date to Year</method>
    </patientbirthdate>
    <patientlastmenstrualdate>
        <xpath>ichicsr/safetyreport/patient/patientlastmenstrualdate
        </xpath>
        <method>Shift Date ICSR</method>
    </patientlastmenstrualdate>
    <patientmedicalstartdate>
        <xpath>ichicsr/safetyreport/patient/medicalhistoryepisode/
        patientmedicalstartdate</xpath>
        <method>Shift Date ICSR</method>
    </patientmedicalstartdate>
    <patientmedicalenddate>
        <xpath>ichicsr/safetyreport/patient/medicalhistoryepisode/
        patientmedicalenddate</xpath>
        <method>Shift Date ICSR</method>
    </patientmedicalenddate>
    <patientdrugstartdate>
        <xpath>ichicsr/safetyreport/patient/patientpastdrugtherapy/
        patientdrugstartdate</xpath>
        <method>Shift Date ICSR</method>
    </patientdrugstartdate>
    <patientdrugenddate>
        <xpath>ichicsr/safetyreport/patient/patientpastdrugtherapy/
```

```
    patientdrugenddate</xpath>
    <method>Shift Date ICSR</method>
</patientdrugenddate>
<patientdeathdate>
    <xpath>ichicsr/safetyreport/patient/patientdeath/patientdeathdate
    </xpath>
    <method>Shift Date ICSR</method>
</patientdeathdate>
<parentidentification>
    <xpath>ichicsr/safetyreport/patient/parent/parentidentification
    </xpath>
    <method>Substitute Pseudonym</method>
</parentidentification>
<parentbirthdate>
    <xpath>ichicsr/safetyreport/patient/parent/parentbirthdate</xpath>
    <method>Generalize Date to Year</method>
</parentbirthdate>
<parentlastmenstrualdate>
    <xpath>ichicsr/safetyreport/patient/parent/parentlastmenstrualdate
    </xpath>
    <method>Shift Date ICSR</method>
</parentlastmenstrualdate>
<parentmedicalstartdate>
    <xpath>ichicsr/safetyreport/patient/parent/parentmedicalhistoryepisode
    /parentmedicalstartdate</xpath>
    <method>Shift Date ICSR</method>
</parentmedicalstartdate>
<parentmedicalenddate>
    <xpath>ichicsr/safetyreport/patient/parent/parentmedicalhistoryepisode
    /parentmedicalenddate</xpath>
    <method>Shift Date ICSR</method>
</parentmedicalenddate>
<parentdrugstartdate>
    <xpath>ichicsr/safetyreport/patient/parent/parentpastdrugtherapy/
    parentdrugstartdate</xpath>
    <method>Shift Date ICSR</method>
</parentdrugstartdate>
<reactionstartdate>
    <xpath>ichicsr/safetyreport/patient/reaction/reactionstartdate
    </xpath>
    <method>Shift Date ICSR</method>
</reactionstartdate>
<reactionenddate>
    <xpath>ichicsr/safetyreport/patient/reaction/reactionenddate</xpath>
    <method>Shift Date ICSR</method>
</reactionenddate>
<testdate>
    <xpath>ichicsr/safetyreport/patient/test/testdate</xpath>
    <method>Shift Date ICSR</method>
</testdate>
<drugstartdate>
    <xpath>ichicsr/safetyreport/patient/drug/drugstartdate</xpath>
    <method>Shift Date ICSR</method>
</drugstartdate>
<drugenddate>
    <xpath>ichicsr/safetyreport/patient/drug/drugenddate</xpath>
    <method>Shift Date ICSR</method>
</drugenddate>
<!--Pass through, control for rare diseases-->
<patientepisodename>
    <xpath>ichicsr/safetyreport/patient/medicalhistoryepisode
    /patientepisodename</xpath>
    <method>Pass</method>
</patientepisodename>
<patientdrugname>
    <xpath>ichicsr/safetyreport/patient/patientpastdrugtherapy
    /patientdrugname</xpath>
    <method>Pass</method>
</patientdrugname>
<patientdrugindication>
```

```
        <xpath>ichicsr/safetyreport/patient/patientpastdrugtherapy/
        patientdrugindication</xpath>
        <method>Pass</method>
    </patientdrugindication>
    <patientdeathreport>
        <xpath>ichicsr/safetyreport/patient/patientdeath/patientdeathreport
        </xpath>
        <method>Pass</method>
    </patientdeathreport>
    <parentmedicalepisodename>
        <xpath>ichicsr/safetyreport/patient/parent/parentmedicalhistoryepisode
        /parentmedicalepisodename
        </xpath>
        <method>Pass</method>
    </parentmedicalepisodename>
    <parentmedicalcomment>
        <xpath>ichicsr/safetyreport/patient/parent/parentmedicalhistoryepisode/
        parentmedicalcomment</xpath>
        <method>Pass</method>
    </parentmedicalcomment>
    <parentmedicalrelevanttext>
        <xpath>ichicsr/safetyreport/patient/parent/parentmedicalhistoryepisode/
        parentmedicalrelevanttext</xpath>
        <method>Pass</method>
    </parentmedicalrelevanttext>
    <parentdrugname>
        <xpath>ichicsr/safetyreport/patient/parent/parentpastdrugtherapy/
        parentdrugname</xpath>
        <method>Pass</method>
    </parentdrugname>
    <parentdrugindication>
        <xpath>ichicsr/safetyreport/patient/parent/parentpastdrugtherapy/
        parentdrugindication</xpath>
        <method>Pass</method>
    </parentdrugindication>
    <testname>
        <xpath>ichicsr/safetyreport/patient/test/testname</xpath>
        <method>Pass</method>
    </testname>
    <moreinformation>
        <xpath>ichicsr/safetyreport/patient/test/moreinformation</xpath>
        <method>Pass</method>
    </moreinformation>
    <resultstestsprocedures>
        <xpath>ichicsr/safetyreport/patient/test/resultstestsprocedures</xpath>
        <method>Pass</method>
    </resultstestsprocedures>
    <drugadditional>
        <xpath>ichicsr/safetyreport/patient/drug/drugadditional</xpath>
        <method>Pass</method>
    </drugadditional>
    <narrativeincludeclinical>
        <xpath>ichicsr/safetyreport/patient/summary/narrativeincludeclinical
        </xpath>
        <method>Pass</method>
    </narrativeincludeclinical>
    <reportercomment>
        <xpath>ichicsr/safetyreport/patient/summary/reportercomment</xpath>
        <method>Pass</method>
    </reportercomment>
    <senderdiagnosis>
        <xpath>ichicsr/safetyreport/patient/summary/senderdiagnosis</xpath>
        <method>Pass</method>
    </senderdiagnosis>
    <sendercomment>
        <xpath>ichicsr/safetyreport/patient/summary/sendercomment</xpath>
        <method>Pass</method>
    </sendercomment>
</icsrconf>
```