# SEMANTIC TRANSFORMATION FRAMEWORK TO ENABLE STATISTICAL RESEARCH ON MEDICAL SUMMARIES

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

BY

ANIL PAÇACI

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

JUNE 2015

Approval of the thesis:

**SEMANTIC TRANSFORMATION FRAMEWORK TO ENABLE STATISTICAL RESEARCH ON MEDICAL SUMMARIES**

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

Prof. Dr. Gülbin Dural Ünver _____
Dean, Graduate School of **Natural and Applied Sciences**

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

Assist. Prof. Dr. İsmail Sengör Altıngövde _____
Supervisor, **Computer Engineering Dept., METU**

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

**Examining Committee Members:**

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

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

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

Assoc. Prof. Dr. Pınar Karagöz
Computer Engineering Department, METU _____

Assist. Prof. Dr. Aybar Can Acar
Informatics Institute, 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: ANIL PAÇACI

Signature :

# ABSTRACT

SEMANTIC TRANSFORMATION FRAMEWORK TO ENABLE STATISTICAL RESEARCH ON MEDICAL SUMMARIES

Paçacı, Anıl
M.S., Department of Computer Engineering
Supervisor : Assist. Prof. Dr. İsmail Sengör Altıngövde
Co-Supervisor : Prof. Dr. Asuman Doğaç

June 2015, 91 pages

One of the important aspects of the clinical research studies carried out in the pharmacovigilance and pharmacoepidemiology is the postmarketing drug surveillance. Utilization of the available Electronic Health Record (EHR) data is key to complement and strengthen the postmarketing safety studies. In addition, EHRs provide a huge, but still under-utilized source of information for the observational studies in clinical research. However, use of different EHR data models and vocabularies presents an important interoperability challenge. Predominant solution to this problem is to transform the data from these disparate EHR datasets into a common data model (CDM) in order to enable large-scale systematic analyses.

Existing transformation practices depend on proprietarily developed Extract - Transform - Load (ETL) procedures. It requires a significant amount of expertise in both source and target models, as well as detailed technical knowledge on the

underlying database implementations. Moreover, the experience gained during the transformation of one source is not readily transferable to other domains.

In this thesis, we address these challenges and develop the necessary semantic transformation machinery to translate the EHR data available in SALUS Common Information Model to the Observational Medical Outcomes Partnership (OMOP) CDM. It enables pharmacovigilance researchers to seamlessly run existing safety analysis methods defined in the OMOP project on top of disparate EHR sources. The semantic materialization technique is adopted with the use of semantic mapping rules for data conversion on EYE reasoner. Accuracy and feasibility of the proposed framework have been evaluated in real-world settings together with pharmacovigilance researchers.

Keywords: Postmarketing Safety Study, Common Data Model, OMOP CDM, Semantic Mapping Rules, Data Transformation, Pharmacovigilance

# ÖZ

HASTA KAYITLARI ÜZERİNDE İSTATİSTİKİ ANALİZLER İÇİN
ANLAMSAL DÖNÜŞÜM SİSTEMİ

Paçacı, Anıl
Yüksek Lisans, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi : Yrd. Doç. Dr. İsmail Sengör Altıngövde
Ortak Tez Yöneticisi : Prof. Dr. Asuman Doğaç

Haziran 2015, 91 sayfa

Farmakoepidemiyoloji ve farmakovijilans alanında yürütülen klinik araştırmaların en önemli boyutlarından biri de pazarlama sonrası ilaç gözetimleridir. Halihazırdaki Elektronik Sağlık Kayıtlarından (ESK) faydalanılması pazar sonrası güvenlik çalışmalarında tamamlayıcı ve güçlendirici bir faktördür. Ek olarak, ESK-lar klinik araştırmalardaki gözlemsel çalışmalar için devasa ama yeterli oranda kullanılamayan bilgi kaynaklarıdır. Fakat farklı veri modellerinin ve kod sistemlerinin kullanımı birlikte-işlerlik sorunlarını ortaya çıkarmaktadır. Büyük ölçekli sistematik çalışmların yapılmasına olanak sağlayan en yaygın çözüm, farklı kaynaklarda bulunan verinin ortak bir veri modeline dönüştürülmesidir.

Günümüzde kullanılan dönüşüm teknikleri çoğunlukla, kullanılan veri modeline göre özel olarak tasarlanan ETL prosedürlerine dayanmaktadır. Bu yöntem hem kullanılan modeller hem de kullanılan veri tabanı uygulamaları hakkında detaylı uzmanlık gerektirmektedir. Önemli bir diğer sorun ise bir kaynağın dönüşümü

sırasında kazanılan uzmanlığın diğer kaynaklara kolay bir şekilde aktarılamamasıdır.

Bu tez çalışmasında, yukarıda bahsedilen zorlukları çözmeyi amaçlayan, halihazırda SALUS CIM formatında tanımlı ESKların Observational Medical Outcomes Partnership (OMOP) Ortak Veri Modeli'ne çevrilmesini sağlayacak gerekli anlamsal dönüşüm sistemleri geliştirilmiştir. Bu sayede farmakovijilans araştırmacılarının OMOP projesi çerçevesinde tanımlanan güvenlik analiz metodlarını farklı veri kaynakları üzerinde kolayca çalıştırması sağlanmıştır. Veri dönüşümü esnasında EYE reasoner ve anlamsal veri dönüştürme kurallarının yardımıyla anlamsal maddeleştirme teknikleri uygulanmıştır. Sunulan sistemin uygulanabilirliği ve doğruluğu, farmakovijilans araştırmacılarının yardımı ile gerçek bir ESK sistemi üzerinde test edilmiştir.

Anahtar Kelimeler: Pazar Sonrası İlaç Güvenlik Çalışmaları, Ortak Veri Modeli, OMOP OVM, Anlamsal Eşleştirme Kuralları, Veri Dönüşümü, Farmakovijilans

*To my dearest parents*

*Mukadder Paçacı, Fahrettin Paçacı*

# 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 Assist. Prof. Dr. İsmail Sengör Altıngövde 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, Ali Anıl Sınacı, Suat Gönül 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 Agfa, UMC, LISPA and TUD for their collaboration while identifying the requirements of this study from both clinical and research perspectives.

I would also thank the Scientific and Technological Research Council of Turkey (TÜBITAK) for providing the financial means throughout this study through TÜBİTAK-BİDEB graduate scholarship (2228-a).

I am also grateful to my parents, Mukadder and Fahrettin Paçacı; my sister Yasemin Paçacı, for their love, belief and continued support.

Finally, very special thanks go to my dear friends Damla Ezgi Küçükkayalar, Cenk Köknar, Doruk Aksoy, Serbay Arslanhan and Mehmet Kubilay Kahveci for their help, support and cheerful presence through the course of this study.

# TABLE OF CONTENTS

# LIST OF TABLES

TABLES

# LIST OF FIGURES

FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ADaM | Analysis Data Model |
| ADE | Adverse Drug Event |
| ADR | Adverse Drug Reaction |
| AMI | Acute Myocardial Infarction |
| BRIDG | Biomedical Research Integrated Domain Group |
| CDISC | Clinical Data Interchange Standards Consortium |
| CIM | Common Information Model |
| CDM | Common Data Model |
| CPRD | Clinical Practice Research Datalink |
| DB | Database |
| DBLP | Digital Bibliography & Library Project |
| DOI | Drug Outcomes of Interest |
| DWH | Data Warehouse |
| EHR | Electronic Health Record |
| ER | Entity - Relationship |
| EYE | Euler yet another proof Engine |
| ETL | Extract - Transform - Load |
| FDA | Food and Drug Administration |
| GROUCH | Generalized Review of OSCAR Unified Checking |
| Health-ITUES | Health IT Usability Evaluation Scale |
| HL7 | Health Level Seven International |
| HOI | Health Outcomes of Interest |
| HTTP | Hyper-Text Transfer Protocol |
| ICD | International Classification of Diseases |
| ICT | Information and Communication Technologies |
| ID | Identifier |
| IEC | International Electrotechnical Commission |
| ISO | International Organization for Standardization |

| | |
|---|---|
| IT | Information Technologies |
| LISPA | Lombardia Informatica Spa |
| LOD | Linked Open Data |
| LOINC | Logical Observation Identifiers Names and Codes |
| MedDRA | MedDRA or Medical Dictionary for Regulatory Activities |
| NATHAN | Utility of Natural History Information |
| N3 | Notation 3 |
| OMOP | Observational Medical Outcomes Partnership |
| OSCAR | Observational Source Characteristics Analysis Report |
| OWL | Web Ontology Language |
| RDBMS | Relational Database Management System |
| RDF | Resource Description Framework |
| RDFS | RDF Schema |
| RIM | Reference Information Model |
| SALUS | Scalable, Standard based Interoperability Framework for Sustainable Proactive Post Market Safety Studies |
| SNOMED CT | Systematized Nomenclature of Medicine–Clinical Terms |
| SPARQL | SPARQL Protocol and RDF Query Language |
| SQL | Structured Query Language |
| SQuaRE | Software product Quality Requirements and Evaluation |
| SRDC | Software Research and Development Consultancy |
| SUS | System Usability Scale |
| TAS | Temporal Association Screening |
| THIN | The Health Improvement Network |
| TPC | Temporal Pattern Characterization |
| TUD | Technical University of Dresden |
| UMC | Uppsala Monitoring Centre |
| UML | Unified Modelling Language |
| UMLS | Unified Medical Language System |
| URI | Uniform Resource Identifier |
| US | United States |
| WHO | World Health Organization |
| XML | eXtensible Markup Language |
| XSD | XML Schema Definition |

# CHAPTER 1

# INTRODUCTION

It is a well-accepted fact that drugs may still have serious side effects even after they are marketed, called Adverse Drug Events (ADEs). Postmarketing drug surveillance plays an important role in capturing such ADEs, since scope and duration of clinical trials are limited by their nature. Pharmacovigilance is the science focuses on the detection, assessment and prevention of the ADEs and any other drug-related problems [26]. Historically, drug safety surveillance research in pharmacovigilance has been depending mostly on the mandatory reports produced by randomized trials of the industry and the case reports submitted to the regulatory bodies such as FDA Adverse Event Reporting System. In literature, many studies discuss the deficiencies of these approaches. Clinical trials are criticized for their limitations on population range and trial durations [29]. Case report analysis, on the other hand; suffers from differential reporting, uneven quality and underreporting problems [47, 31, 53, 36]. Although these approaches are still predominant in the market, new research area that uses already available electronic health data for clinical research purposes is emerging, which is referred as the secondary use of Electronic Health Records (EHRs). EHRs provide a huge, but still under-utilized source of information on the real world use of drugs for the observational studies.

Primary focus of the observational studies is to identify the risk factors and the prognostic indicators in situations where conducting a randomized trial would be impossible or unethical [29]. Observational studies have several advantages compared to the experimental studies where randomized controlled experiments

are conducted. In addition to being costly, randomized clinical trials can only cover small populations in a shorter time span. Observational studies, on the other hand; are not only cost effective but also provide major advantages in terms of broad range of patient population and greater timespan. Despite their drawbacks such as the potential bias [38, 41], new approaches which aims to utilize increasing number of well structured, large, healthcare datasets greatly increase the potential of observational studies.

EHR data available in observational healthcare datasets covers extended part of the patient medical history and includes more complete information about the risk factors compared to spontaneous case reports. This broad range clinical information could be highly beneficial for surveillance studies which makes it a great potential to complement and strengthen existing postmarketing safety studies [52, 33].

Data in healthcare information system which is collected during the care service enables researchers to assess the impact of real clinical practices. However, there are challenges against fully exploitation of this huge amount of data such as the interoperability of data coming from disparate sources. Lack of shared information infrastructure hinders integration of data coming from heterogeneous systems and locations.

In general, healthcare datasets are stored in different data models. These different data models often organize the data in different ways and make use of local terminologies. This is the basis of the interoperability problems for healthcare researchers. It is difficult to analyse the data in disparate sources in a systematic way by using same tools if these sources have different structures. Therefore researchers usually develop custom analyses for specific data source. Use of different coding schemes and local terminologies only make the situation worse. This is another reason why most of the analyses focus on single data source. In order to extend analyses to multiple heterogeneous data sources, either analysis should be tailored for each data model and terminology, or conversion to a common data model (CDM) must be performed [46]. A CDM basically defines a standardized data structure, transparent transformation rules, develops and

specifies a common vocabulary to represent concepts. More importantly, it enables establishment of library routines over this infrastructure. Hence, a CDM not only facilitates understanding of the analysts, but also enables large scale systematic analyses over disparate sources.

Tailoring analysis for each data model in an effort to conduct experiments over heterogeneous data models presents more significant limitations [32]. For each analysis detailed knowledge about all the underlying models would be required. More importantly, analyses would be source specific which makes it harder to reproduce on other observational data sources. On the other hand, if data is normalized within a CDM and a common vocabulary is used, library of standard analytic routines that were developed based on a common format [48] enables systematic analyses over disparate sources and produces comparable results.

In order to perform normalization of an healthcare dataset into a CDM, detailed knowledge about the underlying model is still necessary. However, once conversion is done, same analyses and experiments could be performed homogeneously. Selection of the common terminology must be made carefully so that all the local codes can be covered. Moreover, source data model can have specific characteristics, richness and granularity which can be lost in CDM. In other words, CDM may not allow representing all the data or the relationships within the data. Therefore it is important to assess the feasibility of the CDM on the use cases. These two stand as possible limitations of the using CDMs in normalizing observational health care data for active safety surveillance.

There are important initiatives for building large data pools from the EHRs to benefit from the available longitudinal observational data such as BRIDG [2] and ADaM [1] from the CDISC and OMOP CDM [3] from OMOP. Common objective of these initiatives is to address both syntactic and semantic interoperability issues in utilization of healthcare data from various sources so that researchers can develop their analyses for a single CDM instead of developing different methods per data source. In line with this vision, a new research area for signal detection and safety monitoring based on data mining applications over such data pools has emerged [52, 42].

CDM based interoperability approach has been validated to be viable and reliable methodology for conducting postmarketing surveillance over heterogeneous healthcare datasets. Among other data models, OMOP CDM is the most commonly adopted since it has been found to be superior particularly for addressing diverse requirements of both researchers and data owners. Ogunyemi et al. present a comparative analyses between BRIDG, ADaM, Mini-Sentinel and OMOP CDM to assess their appropriateness for comparative effectiveness research [45]. A source data warehouse has been transformed into these target models, and their extensibility, adoption of standardized terminologies, availability of library of analytic routines and loss of information during transformation have been evaluated. Authors conclude that although each model has its own benefits and limitations, OMOP CDM is the one which meets the broadest set of needs with regards to usage objectives. Due to these reasons, we have concentrated our focus on the OMOP Common Data Model.

In literature, number of studies aim to validate the CDM based interoperability approach for postmarketing surveillance by transforming disparate healthcare datasets into a common data model [46, 48, 57, 40]. In all of these efforts, it is underlined that the transformation process requires significant amount of effort and expertise on both source and target models. In two cases [46, 57], it is reported that complete process including transformation of records, mapping source terminologies etc. required approximately 24 PMs. Moreover, expertise gained in transforming one data source is not readily transferable to other domains, since transformation is based on proprietary ETL (Extract - Transform - Load) procedures and the terminology mappings are usually source specific. Although there are some efforts to ease the transformation process such as the ETL Guidelines published by OMOP [37], it still presents a big challenge for pharmacovigilance researchers.

We address this problem by proposing a semantic transformation framework for linked EHR data. We have developed a semantic representation for OMOP CDM based on Linked Open Data (LOD) principles, named OMOP Content Entity Model and built the necessary machinery to handle population of CDM instance from this OMOP Content Entity Model considering the terminology mappings.

Proposed framework enables data sources to be transformed into OMOP CDM by only providing abstract mapping rules to OMOP Content Entity Model and rules out the effort spent to address RDBMS specific details.

This rule based approach enables defining the mappings between two data models without dealing with implementation details. As a result, any RDF based data model can be seamlessly translated to target OMOP CDM only by developing the semantic mapping rules, without the extra burden causing from the details of underlying database implementations. Being modular makes semantic transformation framework a scalable, verifiable and sustainable approach.

Validity and applicability of the proposed framework has been evaluated in a case study, which includes the Lombardy Region of Italy and the Hospital of Technical University of Dresden, in the scope of FP7-ICT-287800 SALUS Project. Data retrieved in conformant to SALUS Common Information Model has been transformed into OMOP Content Entity Model by means of semantic mapping rules and OMOP CDM instances have been populated. In addition to the comparison of descriptive statistics between native and transformed databases, Temporal Pattern Discovery method from OMOP Methods Library [11] has been executed over populated instances in order to demonstrate the validity of the proposed approach.

## 1.1 Objectives

CDM based interoperability approach is proven to be successful in facilitating the secondary use of observational healthcare datasets for postmarketing safety studies. It requires data owners to normalize their data into a predefined CDM, so that researchers and data analysts can develop their methods on this CDM independent from the schema of underlying data sources. However, data transformation process itself presents important challenges. First of all, transformation requires significant expertise on both source and target data models. Second, detailed technical knowledge about the underlying relational database and ETL procedures is required since CDMs are implemented on relational database man-

agement systems. Last but not least, ETL procedures developed for a specific model are not scalable to others and expertise gained from one domain is not easily transferable to other domains.

We believe that there is a need for more coordinated approach to address the limitations of the current CDM transformation practices. Schema transformation should be handled on a higher level than population of the underlying database which requires to handle implementation specific details. In this way syntactic and semantic parts of the transformation process can be separated in favour of scalability and verifiability. In this thesis, we propose a semantic representation for commonly adopted OMOP CDM so that schema of the source data models can be mapped to OMOP CDM through abstract mapping rules. On top of this we propose a machinery to populate the underlying OMOP CDM instance automatically hiding all the cumbersome technical details.

## 1.2 Summary of Contributions

Having described the objectives in Section 1.1, contributions of this thesis work can be summarized as follows:

- OMOP Content Entity Model, which is the semantic representation of the OMOP Common Data Model, has been developed based on semantic web technologies such as RDFS [17] and OWL [15]. It constitutes a standard, machine-processable format for the OMOP CDM.

- OMOP DB Adapter, which automatically populates the underlying OMOP CDM instance from the RDF data available in OMOP Content Entity Model format, has been fully developed. OMOP DB Adapter seamlessly performs the era calculations on the fly, from condition occurrences and drug exposures. As a consequence, any standardized method from OMOP Methods Library [11] can be readily executed.

- Temporal Pattern Discovery Tool enables pharmacovigilance researchers to analyse temporal associations between condition - medication pairs by

executing standard OMOP methods without dealing implementation specific details (i.e. database connection parameters, local vocabularies of the data source).

- Set of semantic rules have been developed to be executed in EYE Reasoning Engine [4]. Based on the pilot application scenario defined in Section 4.1, these mapping rules derive OMOP Content Entity Model constructs from entities represented in SALUS Common Information Model.

- A case study covering two different EHR sources, namely LISPA and TUD, has been conducted. Based on the results obtained from this case study, validity and applicability of the proposed framework has been evaluated.

The rest of this thesis is structured as follows: Chapter 2 presents brief information about the technologies and the standards used in the realization of this thesis. Chapter 3 goes into details of the proposed semantic transformation framework and its components. In Chapter 4, pilot application scenario and the results obtained from the pilot study are presented. Chapter 5 outlines a survey on similar studies in the literature. Finally, Chapter 6 concludes the thesis by giving final remarks and presents the future research directions.

# CHAPTER 2

# BACKGROUND ON ENABLING TECHNOLOGIES

In this chapter, main technologies and standards which have been used in the context of this thesis are presented.

## 2.1 OMOP Project

Observational Medical Outcomes Partnership (OMOP) is a public-private partnership funded and managed through the Foundation for the National Institutes of Health, with the overall aim to improve the safety monitoring of medical products. The partnership is conducting a multi-year comparative evaluation of analytical methods for safety surveillance of longitudinal observational databases across a spectrum of disparate (administrative claims as well as electronic health records) data sources.

OMOP conducts empirical evaluation on the performance of various analytic and statistic methods in order to estimate the correlation between the treatment and the outcome. OMOP conducts this evaluation over multiple heterogeneous sources by creating network out of them. In-line with its mission, OMOP produced stack of tools and technologies which are freely shared within the research community. This stack includes a data model, experimental protocols, guidelines, database evaluation tools and standardized methods library. Following subsections present more detailed information about the Common Data Model and OMOP Methods Library respectively.

### 2.1.1 Common Data Model

No single observational data source can meet all expected outcome analysis needs, so there is a demand for assessing and analysing multiple data sources concurrently. The OMOP Common Data Model (CDM) can be used for that purpose to standardize the data format used by generating a separate CDM instance for each source dataset.

The CDM needs to support research to identify and evaluate associations between interventions (drug exposure, procedures, healthcare policy changes etc.) and outcomes caused by these interventions (condition occurrences, procedures, drug exposure etc.).

The CDM is designed to store observational data under the following principles:

- Data protection. The CDM aims at providing data storage optimal for analysis. In addition, all data that might jeopardize the identity and protection of patients, such as names, precise birthdays etc. are limited. Exceptions are possible where the research expressly requires more detailed information, such as precise birth dates for the study of infants.
- Reuse of existing models. In designing the CDM, industry-leading data modeling efforts are leveraged, such as HL7 RIM, the HIMSS EHR Definitional Model, the i2b2 Hive framework, the HMORN Virtual Data Warehouse, etc.
- Design of domains. The domains are modeled in a person-centric relational data model, where for each record the identity of the person and a date is captured as a minimum.
- Standard vocabulary. To standardize the content of those records, the CDM relies on a Standard Vocabulary containing all necessary and appropriate corresponding standard healthcare concepts.
- Technology neutrality. The CDM does not require a specific technology. It can be realized in any relational database, such as Oracle, MySQL etc., or as SAS analytical datasets.

- Scalability. The CDM is optimized for data processing and computational analysis to accommodate data sources that vary in size, up to and including databases with tens of millions of persons and billions of clinical observations.

The CDM defines table structures for each of the data in a Person-centric model. Almost all tables have foreign keys into the Person table and a date. This allows for a longitudinal view on all the healthcare-relevant events. In addition, Providers carrying out healthcare are linked to many of the events as well. Figure 2.1 shows the complete entity-relationship diagram of the OMOP CDM.

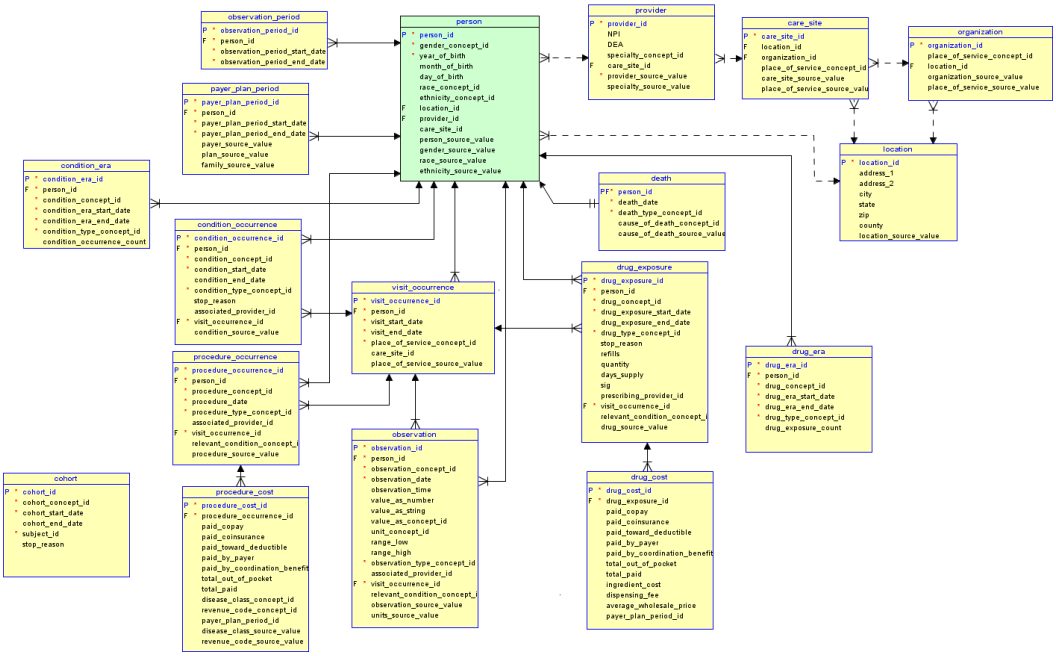



Figure 2.1: Entity-Relationship Diagrams for the OMOP Common Data Model Tables

### 2.1.2 OMOP Methods Library

There are many researchers and data analysts in pharmacovigilance with considerable experience. Due to lack of interoperability solutions, those experts have different statistical and epidemiological approaches depending on the data source. Common Data Model based interoperability approach addresses this problem of pharmacovigilance and increase the consistency within the data

sources. More importantly, it enables researchers to conduct similar analyses on disparate data sources and obtain comparable results.

A library of standardized analytic methods are being developed by the OMOP to analyse the association between any drug and outcomes and their performance are evaluated empirically. OMOP team freely shares these methods with public so as to increase transparency and consistency within the pharmacovigilance research community. All these analysis routines was developed for the CDM, so that this common set of procedures is applicable across participating data sources.

In 2012, originally defined set of procedures were published by OMOP, which includes:

- Automated Pharmacoepidemiology Evaluation eXplorer (APEX) Cohort Method
- Adapted Self-Controlled Case Series for Accumulated Exposure - Erasmus University Medical Center Rotterdam
- Observational Screening (OS) - UBC and OMOP Research Team
- Self-Controlled Case Series (SCCS) - Columbia University
- Cohort Method - OMOP Research Team
- Disproportionality Analysis - OMOP Research Team
- IC Temporal Pattern Discovery - Uppsala Monitoring Centre
- Case Control - Columbia University
- Longitudinal Gamma Poisson Shrinker (LGPS) & Longitudinal Evaluation of Observational Profiles of Adverse events Related to Drugs (LEOPARD) - Erasmus University Medical Center Rotterdam

IC Temporal Pattern Discovery from the Uppsala Monitoring Centre of the World Health Organization, which is based on the observed-to-expected incidence of a medical around a pivot event, is employed in the context of this thesis. Section 3.4 describes the method and its usage in detailed manner.

## 2.2 SALUS Project

SALUS (Scalable, Standard based Interoperability Framework for Sustainable Proactive Post Market Safety Studies) is an research and development project supported and co-financed by the European Commission in the scope of 7th Framework Programme (FP7). Overall objective of the SALUS project is to enable secondary use of existing Electronic Health Record (EHR) data in patient care domain by providing a ready-to-use solution stack. SALUS Project focuses on:

- Reinforcing spontaneous reporting process by automating the ADE detection process on heterogeneous healthcare databases
- Reinforcing spontaneous reporting process by populating case reports with the data extracted from EHRs automatically
- Enabling postmarketing safety analysis and effectiveness queries on different cohorts from disparate EHR systems
- Strengthen signal detection process
- Facilitating large scale comparative effectiveness research on distributed, heterogeneous healthcare data sources in an attempt to identify long term safety issues of a medical product

### 2.2.1 Common Information Model

SALUS Project delivers a technical and semantic interoperability infrastructure which connect heterogeneous healthcare data sources through a semantic layer. For this purpose, available content models have been analysed in the context of SALUS project and set of abstract data element definitions have been delivered. On top this common data elements an RDF based information model has been developed [56]. This SALUS Common Information Model (CIM) plays the mediator role and fosters syntactic and semantic interoperability. Apart from the semantic capabilities such as inference of implicit facts through semantic reasoning, it is yet another content model.

In SALUS Project, EHR data from distributed systems is queried through functional interoperability profile and results of these queries is shared as a set of medical summaries. On research site, applications and methods requires this set of medical summaries in various formats. For instance, Temporal Pattern Discovery and Patient History tools process data in OMOP CDM [3] format, while ICSR Reporting tool produces the case safety reports in E2B(R2) [34] format. CIM is the core of the SALUS interoperability architecture and prevents n-to-n mappings among these various content models from clinical care and research domains.

Based on the SALUS Common Data Elements, patient centric CIM contains 13 different clinical entities, namely Patient demographics, Patient family history, Patient social history, Laboratory results, Vital signs, Medical condition: diagnosis, Medical condition: adverse events, Pregnancy, Medical procedure, Encounter: visit, Encounter: stay, Medication, and Immunotherapy. Figure 2.2 presents a small portion of SALUS CIM covering the Medical condition entity, in XSD [25].

## 2.3 Medical Terminologies and Terminology Servers

In healthcare, terminologies and controlled vocabularies are used formal representation of concepts and they define unique meaning for concepts. For instance, terms "hearth attack", "acute myocardial infarction" and "myocardial infarction" are all the same semantically and used to refer the same concept by the healthcare professionals, which is stoppage of the blood flow to the part of the heart. Unfortunately, situation is a bit more complex in the context of healthcare applications. Unless these concepts are represented with a unique code for a controlled vocabulary, there is no way for a machine to understand all these terms are semantically same. This is why healthcare applications depend on clinical terminology stems to express the semantics of the clinical concepts. For example, if two applications use "Acute Myocardial Infarction" - "10000891" from MedDRA vocabulary, semantics can be shared in a consistent and automated way with no ambiguity.

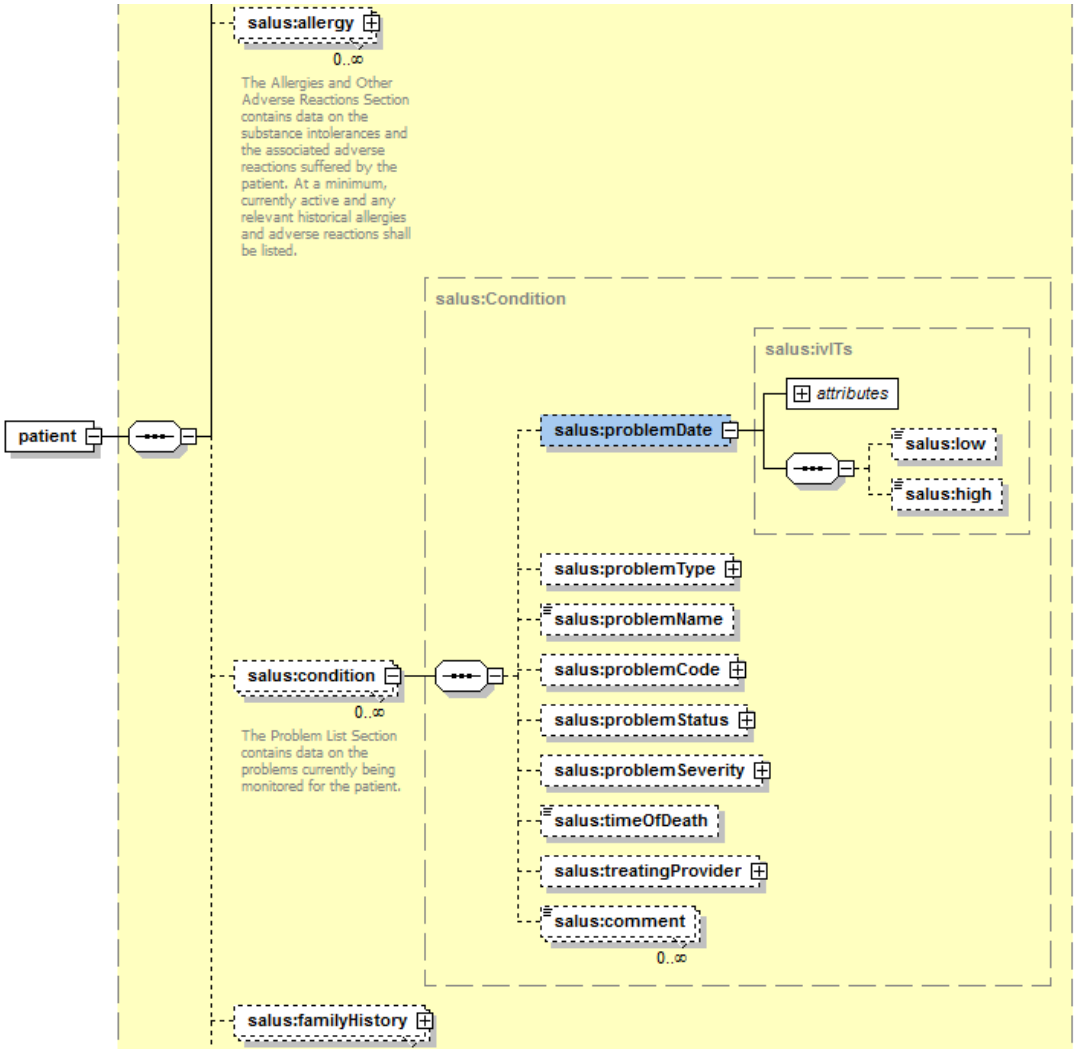



Figure 2.2: An Extract from SALUS Common Information Model represented in XSD

In literature, there exist number of different medial terminologies used to represent the entities in patient summaries. Among them, Systemized Nomenclature of Medicine Clinical Terms (SNOMED CT) [21], Medical Dictionary for Regulatory Activitie (MedDRA) [10], International Statistical Classification of Diseases and Related Health Problems 9th and 10th Revision (ICD9-CM, ICD10) [6, 7] and The Anatomical Therapeutic Chemical Classification System with Defined Daily Doses (ATC/DDD) [27] are the most commonly used ones. In the SALUS project, ATC is adopted in both clinical care and research site for medical products. On the other hand, multiple terminology systems are used to define conditions in SALUS Project. MedDRA is used by the researchers while

different versions of ICD is adopted in SALUS data sources, i.e. ICD9-CM in LISPA and German Modification of ICD10 in TUD.

In the context of SALUS Project, a Terminology Server together with a terminology reasoning service has been developed [55]. Terminology reasoning service provides mapping between different terminology systems used in different contexts. Terminology Server maintains the terminology systems and the mappings among them. It provides a RESTful interface for querying terms and term mappings. Terminology systems maintained in the SALUS Terminology Server are presented in Table 2.1

Table 2.1: Terminology Systems in SALUS Terminology Server and Object Identifiers (OID)

| **Label** | **Name** | **Object Identifier** |
|---|---|---|
| ATC | Anatomical Therapeutic Chemical Classification System | 2.16.840.1.113883.6.73 |
| AdministrativeGender | HL7 Administrative Gender Codes | 2.16.840.1.113883.5.1 |
| ICD9CM | International Classification of Diseases, Ninth Revision, Clinical Modification | 2.16.840.1.113883.6.2 |
| ICD10 | International Classification of Diseases, Tenth Revision | 2.16.840.1.113883.6.3 |
| ICD10GM | International Classification of Diseases, Tenth Revision, German Modification | 1.2.276.0.76.5.{changes based on the yearly version [2004-2013]} |
| MedDRA | Medical Dictionary for Regulatory Activities | 2.16.840.1.113883.6.163 |
| SNOMEDCT | SNOMED Clinical Terms | 2.16.840.1.113883.6.96 |
| LOINC | Logical Observation Identifiers Names and Codes | 2.16.840.1.113883.6.1 |
| TUDLabTest | Codes for Laboratory Tests used locally in Technical University of Dresden | NA |

SALUS Terminology Server has been integrated with the Temporal Pattern Discovery Tool in the scope of this thesis. On the query definition screen, Termi-

nology Server integration enables researchers to define their queries in MedDRA independent from the terminology system used in underlying data source. More detail is given in Section 3.4.

## 2.4 Semantic Web & Linked Data

Semantic Web is defined as "a web of data that can be processed directly and indirectly by machines" by its creator, Tim Berners-Lee [30]. It can be considered an extension to the existing Web system with better methodologies and it enables representation of the knowledge in a standardized way by defining ontologies [51]. In this way, meaning of the entities and their relationships can be formally defined. This explicit formal specification of the terms and their relationships in a domain is considered as an ontology in the Semantic Web. In other saying, ontology defines the schema of the knowledge in a particular domain. In Semantic Web, using existing well-known vocabularies for knowledge representation is a well accepted practice.

To foster the standardized knowledge representation, Semantic Web requires ontologies to be linked with other ontologies so that same concepts can be represented with common, well-known terms. On top of this, actual entities should be connected with other entities and ontologies so that these explicit links between the entities can be traversed by semantic web applications. Linked Data is "a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF" [9]. Tim Berners-Lee defines four basic principles for Linked Data:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs. so that they can discover more things.

Objective of the Linked Open Data Project [9] is to extend the available open datasets which are connected to each other based on linked data principles. DBLP bibliography, Geonames, Wikibooks and Wikipedia are examples of such open datasets. As of August 2014, in the Linking Open Data initiative, there are 1014 interlined data sets as displayed in the cloud diagram in Figure 2.3

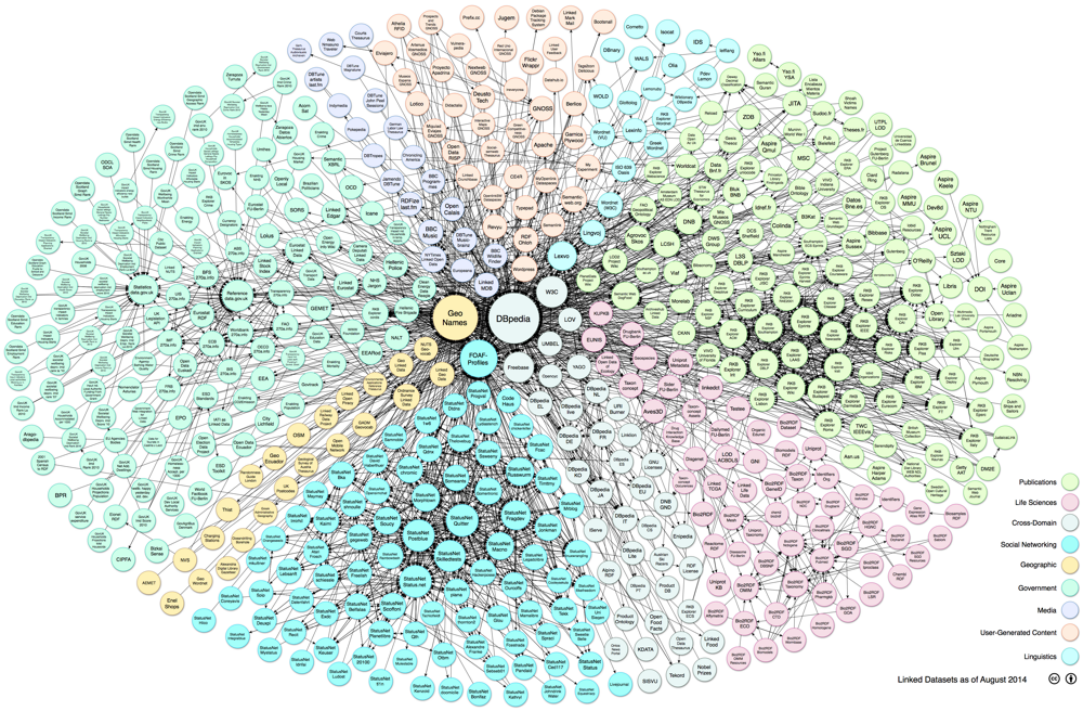



Figure 2.3: Visualization of Linked Open Data Cloud as of August 2014

## 2.5 EYE Reasoning Engine

EYE (Euler Yet another proof Engine) is a backward-forward-backward chaining reasoner engine. AGFA, one of the beneficiaries of the SALUS Project, develops and maintains the open source EYE project. It is distinguished by providing logic based proofs and performing Euler path detection to avoid vicious cycles. EYE performs reasoning grounded in First Order Logic. The backward-forward-backward chaining of the reasoning process is realized in three steps:

- Prolog backward chaining
- Forward meta level reasoning

- Backward proof construction

As the Resource Description Framework (RDF) [16] syntax, EYE Reasonin engine uses the Notation3 (N3) [13]. N3 is a language developed by semantic web community as an alternative non-XML, human readable serialization for RDF models. It is extended to allow greater expressiveness by adding variables, logical implication, functional predicates and allowing graphs to be represented as literals. Thus it is a superset of the RDF. RDF data, rules and queries have to conform N3 syntax, since EYE consumes the input only in N3. On the other hand, EYE Reasoning engine can be used via Java, C-sharp, Python, Javascript, Prolog programming languages or the command line.

In the context of SALUS Project, EYE Reasoning Engine is used in almost all kinds of semantic processing and reasoning operations. In the proposed semantic transformation framework, EYE is responsible of processing the semantic mapping rules and deriving the OMOP Ontology statements.

# CHAPTER 3

# ARCHITECTURE OF SEMANTIC TRANSFORMATION FRAMEWORK

Overview of the proposed semantic transformation framework is presented in Figure 3.1. Initially, OMOP Ontology instances are generated through semantic mapping rules from RDF based medical summaries available in underlying EHR systems. Then, OMOP DB Adapter generates SQL Insert statements which will be used to populate OMOP CDM instance. Meanwhile, drug and condition eras, which are the main CDM constructs of interest, are calculated on the fly. Once OMOP CDM instance is populated, any standard analysis method from OMOP Method Library can be seamlessly executed.

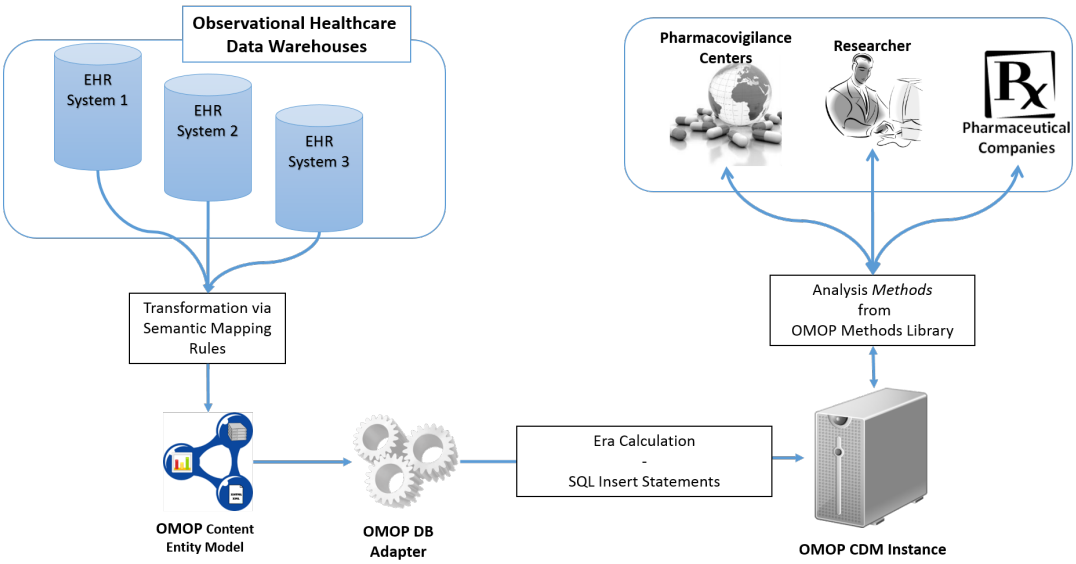



Figure 3.1: Architecture of the Proposed Semantic Transformation Framework

## 3.1 OMOP Content Entity Model

OMOP is based on a model where data from various sources is extracted and transformed to a common structure and framework for further analysis. This is referred as the OMOP Common Data Model. Unlike HL7 CDA which is based on an XML Schema defining the structure, OMOP CDM is defined through human readable tables and SQL scripts as defined in Section 2.1.1 in detail. In other saying, it is a relational model presented by entity – relationship diagrams provided by OMOP team. Figure 3.2 depicts the simplified version of the OMOP Common Data Model, where only tables and their relations are presented.

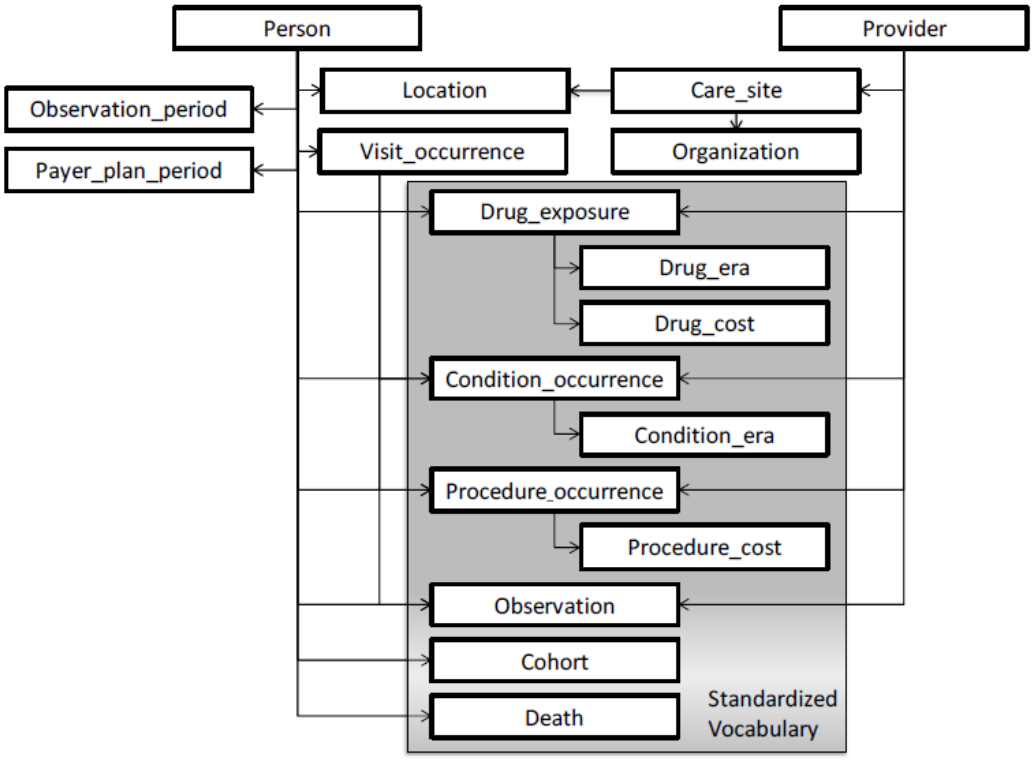



Figure 3.2: Tables of OMOP Common Data Model and their inter-relations

Even though relational database management systems provide many benefits such as the maturity and robustness, there are number of drawbacks using an RDBMS. Most importantly, RDBM systems are considered neither flexible nor adaptive because of its structural and flat nature. As an alternative data modelling paradigm, Resource Description Framework (RDF) [16] is considered su-

perior to classic relational models. Despite its shorter history, RDF stack has been an open standard managed by World Wide Web Consortium. As a result, it has been widely adopted by data modelers and there have been numerous standards and tools developed over the RDF stack. Having adopted Linked Data Principles [9], RDF data models can benefit from being decentralized and distributed. More importantly, RDF provides more flexibility compared to relational models and its capability to capture all kind of data - unstructured (text documents), semi-structured (HTML Documents) or structured (standard databases) - makes it a universal standard for data interoperability. In order to better exploit semantic standards utilize the aforementioned advantages, we have decided to create a machine processable, semantic representation for the OMOP CDM based on the RDF stack, called OMOP Content Entity Model.

Since the structure of the OMOP CDM is uncomplicated, we have created the content entity model manually. In order to obtain the semantic and standardized representation of the OMOP CDM in a human - readable way, OMOP Content Entity Model has been developed by using Web Ontology Language (OWL) [15] based on Linked Data principles [9]. Construction of the OWL Model has been performed based upon the entity-relationship diagrams provided by OMOP team. Constructed model is not a 100% direct correspondence of the original OMOP CDM, since some further information such as the code system OID for concept codes are needed. For each construct in the entity-relationships diagram of OMOP CDM, an OWL construct has been created according to pre-defined mappings as given in Table 3.1.

Table 3.1: ER Constructs' Mappings to OWL Resources

| **ER Constructs** | **OWL Resources** |
|---|---|
| Entity | owl:Class |
| Attribute | owl:DatatypeProperty |
| Relationship | owl:ObjectProperty |

Development of the OMOP Content Ontology Model consists of three basic steps, based on the mappings defined between OWL resources and ER constructs:

1. An OWL class is used to represent each table in OMOP CDM. Thus, each row of the table can be represented as the individuals or the instances of corresponding class.

2. Inter-relations between the tables are represented as owl:ObjectProperty. In other saying, foreign key references are mapped to owl:ObjectProperty. rdfs:domain and rdfs:range properties are used to designate the direction.

3. Rest of the columns are transformed to owl:DatatypeProperty. Proper RDF datatype is designated by rdfs:range property according to the data type of the column.

Visualization of the OMOP Content Entity Model is given in Figure 3.3.

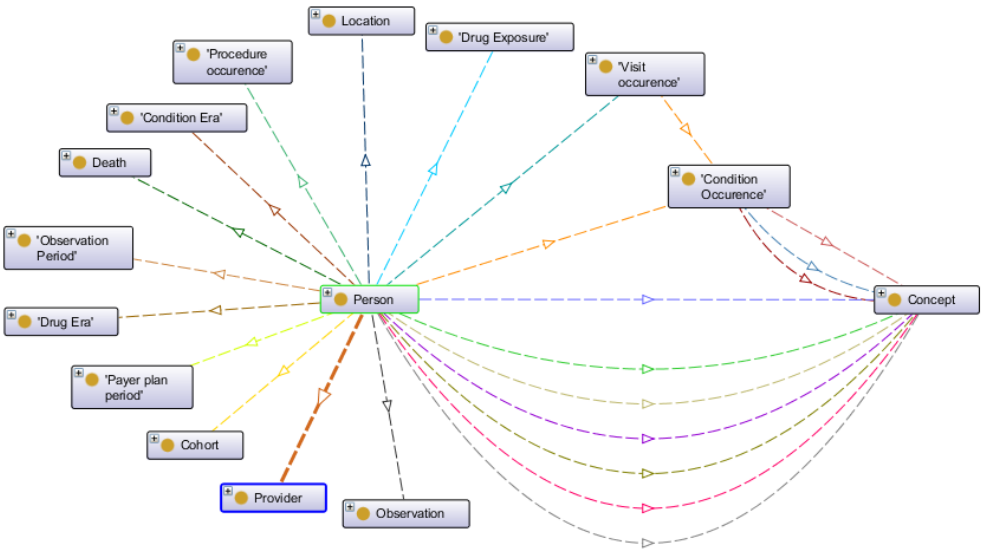



Figure 3.3: Visualization of the OMOP Ontology

An example of minimal OMOP Content Entity Model instance in N3 syntax [13] is presented in Figure 3.4. Person – Location relationship from original CDM is represented as owl:ObjectProperty while other fields of the Person Table are reflected as simple owl:DatatypeProperty.

Generated OMOP Content Entity Model Ontology, in other saying OMOP Ontology, introduces a level of abstraction and serves as a middle layer in transformation process of source model into OMOP CDM. Simply, first, patient data

```
@prefix omop: <http://www.salusproject.eu/ontology/omop-cdm#> .

[    a omop:Person ;
     omop:care_site_id "53254132423" ;
     omop:day_of_birth "24" ;
     omop:location
     [    a omop:Location ;
          omop:address_1 "Via Don Giovanni Minzoni" ;
          omop:address_2 "No 34" ;
          omop:city "Milan" ;
          omop:state "MI" ;
          omop:zip "56123"
     ] ;
     omop:month_of_birth "04" ;
     omop:year_of_birth "1967"
].
```

Figure 3.4: An Example OMOP Content Entity Model Instance

available in RDF based source model is converted into OMOP Ontology instances through high level semantic mapping rules, then OMOP CDM instance is populated from OMOP Ontology. Semantic rule based approach makes the transformation process easier to validate compared to existing ETL process, which is the prevalent method to integrate data from multiple systems into the target database. Correctness of the semantic mapping rules and the transformation itself is verified through standard SPARQL Queries which are executed on OMOP Ontology, in contrast to dealing with technology specific details as in ETL approach.

In addition to serving as a middle layer, created ontology constitutes basis for the machine processable, semantic representation of the data in OMOP CDM format. Following Linked Open Data Principles [9], OMOP CDM becomes ready to be plugged into LOD cloud and linked with other ontologies for sharing and querying, such as biomedical ontologies [37].

## 3.2 Semantic Mapping Rules

In order to achieve overall goal of populating OMOP CDM instance with the RDF based patient data, mapping between two data models is essential. This transformation involves two semantically different levels, first structure of the data should be transformed into the OMOP CDM, after then SQL insert scripts should be generated from RDF data since OMOP CDM is implemented on a relational database. Content entity model – OMOP ontology – has been presented as a middle layer in this transformation process. This approach clearly enables to handle semantics and implementation specific details of the transformation independently.

In the scope of this work, patient data provided by the SALUS project has been used to populate OMOP CDM. SALUS Semantic Interoperability Layer enables retrieval of patient data conformant to RDF based SALUS CIM format, which is described in Chapter 2.2.1 in detail. Therefore rest of this chapter focuses on the methodology for mapping the SALUS CIM to the OMOP ontology.

As the first step of transformation process, it is necessary to convert RDF data in SALUS CIM to OMOP Content Entity Model developed in RDF/OWL. One of the most widespread approaches for transforming one schema to another is use of inference rules. By using these rules, facts and statements in one model is used to derive corresponding statements in target model.

We have conducted an extensive study to map all SALUS CIM constructs to OMOP Content Entity Model. only after we have developed a set of transformation rules in Notation3 (N3) using the Euler Yap Engine (EYE) [4]. EYE has been the choice of inference engine since it is scalable, high performance inference engine. Moreover, it is freely available under an open-source license.

EULER YAP Engine executes the set of semantics mapping rules, which are designed specifically for SALUS CIM, to derive the statements of the OMOP ontology instance. An example OMOP Ontology instance is presented in Figure 3.4 in Chapter 3.1. This instance has been generated by the Euler YAP Engine from the SALUS CIM instance, presented in Figure 3.5.

```
@prefix :        <http://www.srdc.com.tr/ontmalizer/instance#> .
@prefix rdfs:    <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl:     <http://www.w3.org/2002/07/owl#> .
@prefix xsd:     <http://www.w3.org/2001/XMLSchema#> .
@prefix rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix salus:   <http://www.salusproject.eu/ontology/common-information-model#> .


[     a       salus:Patient ;
      salus:address
      [
          a       salus:addr ;
          salus:city "Milan" ;
          salus:country "Italy" ;
          salus:nullFlavor "NP"^^salus:nullFlavorTypeDatatype ;
          salus:postalCode "56123" ;
          salus:state "MI" ;
          salus:streetAddressLine "Via Don Giovanni Minzoni No 34" ;
          salus:use "String" .
      ];
      salus:dateOfBirth "1967-04-24"^^xsd:date ;
      salus:providerOrganization
      [
          a       salus:OrganizationInformation ;
          salus:organizationID
          [
              a       salus:ii ;
              salus:extension "53254132423" ;
              salus:nullFlavor "NP"^^salus:nullFlavorTypeDatatype ;
              salus:root "" .
          ];
          salus:organizationName "String" ;
      ];
].
```

Figure 3.5: An Example SALUS Common Information Model Instance

Complete set of semantic mapping rules consists of 17 rules. Most of the rules are defined to transform SALUS CIM classes into OMOP ontology classes, while two of the rules are responsible of mapping SALUS specific datatypes to simple RDF datatypes, namely ivlTs2start and ii2string. Exhaustive list of semantic mapping rules with corresponding SALUS CIM and OMOP ontology classes is presented in Table 3.2.

It is worth to mention that there is not a one-to-one correspondence between SALUS CIM and OMOP ontology constructs because of the nature of the two data models. These models use different levels of granularity to define same concepts. For instance, both Allergy and Condition from the SALUS CIM are mapped to OMOP CONDITION, whereas Condition can also be mapped into

OMOP DEATH depending on the cardinality of a pre-defined criteria. Specifically, if a Condition instance includes a property "salus:timeOfDeath", then this condition is treated as an OMOP DEATH instead of the generic OMOP CONDITION. Different rules are executed on this instance in an attempt to extract the additional information regarding the death of the person.

Actual Condition2Death rule, the one responsible of generating OMOP DEATH instance based on a predefined criteria, is presented in Figure 3.6. EYE transformation rules for most common OMOP constructs person, drug exposure and condition occurrence are given in Appendix B, Appendix C and Appendix D for interested reader, respectively.

Even though drug and condition eras are essentials components of the OMOP CDM, they are not part of the transformation rule set. Because an era simply defines the persistence exposure or occurrence and is not readily available in SALUS CIM conformant data. Era information is calculated from the exposure and occurrence data once those tables of the OMOP CDM instance are populated. Era calculation process is presented in Chapter 3.3.1 in detail.

This rule based approach presented enables to define the mappings between two data models without dealing with implementation details. As a result, any RDF based data model can be seamlessly translated to target OMOP CDM only by developing semantic mapping rules, without the extra burden causing from the details of underlying database implementation. Being modular makes transformation rules a scalable, verifiable and sustainable approach.

## 3.3 OMOP DB Adapter

Transformation rule set developed for EYE Inference Engine is responsible of generating OMOP ontology instances from the SALUS patient data. Standardized methods from OMOP Methods Library [11] are designed for OMOP CDM, which specifies a relational database schema to define the structure as well as the standardized vocabulary to define the content. Therefore, a relational database instance implementing the OMOP CDM needs populating from the RDF based

```
#Condition2Death
{?CONDITION :mapDeathGraph {
    ?CONDITION :mapDeath ?DEATH.
    ?DEATH a omop:Death;
        omop:person ?op;
        omop:death_date ?odd;
        omop:cause_of_death_source_value ?ocodsv.
}}
<=
{
    ?CONDITION a salus:Condition.
    ?SCOPE e:optional {
        ?sp salus:condition ?CONDITION.
        ?sp :mapPatient ?op.
    },{
        ?op e:tuple (salus:condition ?CONDITION)
    }.
    #death date
    ?CONDITION salus:timeOfDeath ?sctod.
    ?sctod salus:value ?odd.

    #cause of death source value
    ?SCOPE e:optional {
        ?CONDITION salus:problemCode ?spc.
        ?spc :mapCode ?ocodsv.
    },{
        ?ocodsv e:tuple (salus:problemCode ?CONDITION).
    }.
    ?DEATH e:tuple (?CONDITION).
}.
```

Figure 3.6: An Example EYE Rule to transform Condition to DEATH

OMOP Content Entity Model. OMOP DB Adapter is the corresponding component of the proposed framework, responsible from the population of OMOP CDM instance.

OMOP DB adapter performs the second phase of the transformation process. It

consumes the OMOP ontology instances and generates the corresponding SQL Insert statements. These SQL statements are used to populate the underlying OMOP CDM Instance. While SQL statements are generated, original concept codes in the source model need to be converted to OMOP concept IDs from the Standard Vocabulary [24], since OMOP CDM accommodates an internal mechanism to represent concepts from both standard and local terminologies.

The Standard Vocabulary developed by OMOP team is a semantic network containing all of the Concepts, Concept-to-Concept Relationships, Source to concepts mappings and other meta-data necessary to describe the meanings and structures of the data within the CDM. The Standard Vocabulary accommodates concepts for each of the entities of interest relative to drugs, conditions, procedures, visits, demographics, etc. OMOP Standard Vocabulary enables to represent concepts in their original terminology, as well as providing mappings to other terminologies. Structure of the OMOP Standard Vocabulary is a standardized format designed to integrate and standardize terminologies for observational analysis represented by entity-relationship diagram in Figure 3.7.

OMOP DB adapter is also responsible from initialization of the OMOP CDM implemented on a relational database. During the initialization phase, vocabulary tables are populated with the OMOP Standard Vocabulary. This data includes many standardized terminology system, such as the SNOMED-CT, ICD9-CM and ATC. Since one of the SALUS data sources, i.e. TUD, uses a localized version of the ICD10 medical classification system, this additional terminology is inserted into database.

In order to correctly reference concepts from the OMOP Standard Vocabulary, OMOP DB Adapter queries Standard Vocabularies to look up for the OMOP 'concept_id' while generating SQL Insert statements.

### 3.3.1 Era Calculation

In addition to the descriptive characteristics of a person such as date of birth, gender etc., observational databases contain two main types of healthcare records,

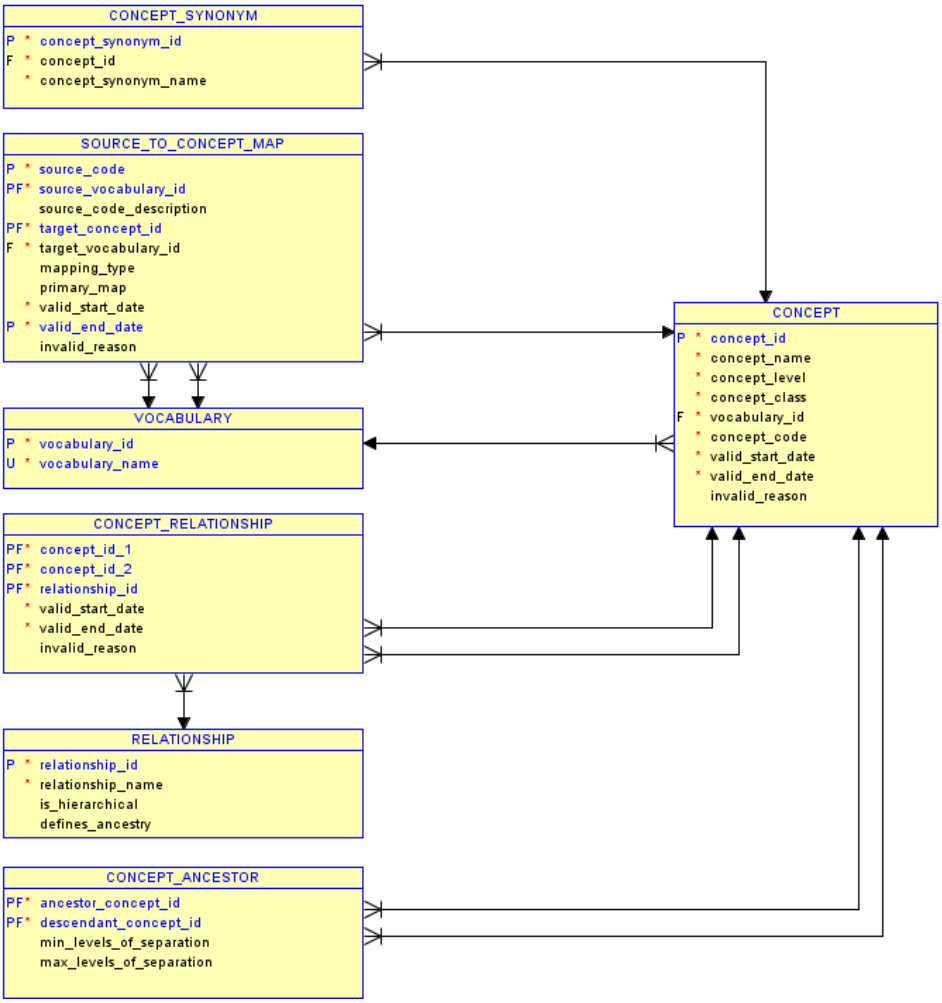



Figure 3.7: Entity-Relationship Diagram of the OMOP Vocabulary

medication exposures and condition occurrences. Therefore, OMOP CDM is designed to combine independent exposures of same medication or occurrence of same condition into a single era through a persistence window. Persistence windows represent the allowable timespan between the exposure of same drug and occurrence of same condition. Figure 3.8 visualizes the calculation of eras trough persistence windows.

As specified in OMOP CDM specification [35], ideal length of a persistence window is 30 days. In the case of medications, if there is less than 30 days between consecutive exposures of same drug, then a drug era is created. Both exposures are merged into this new drug era. If there is no other drug exposure

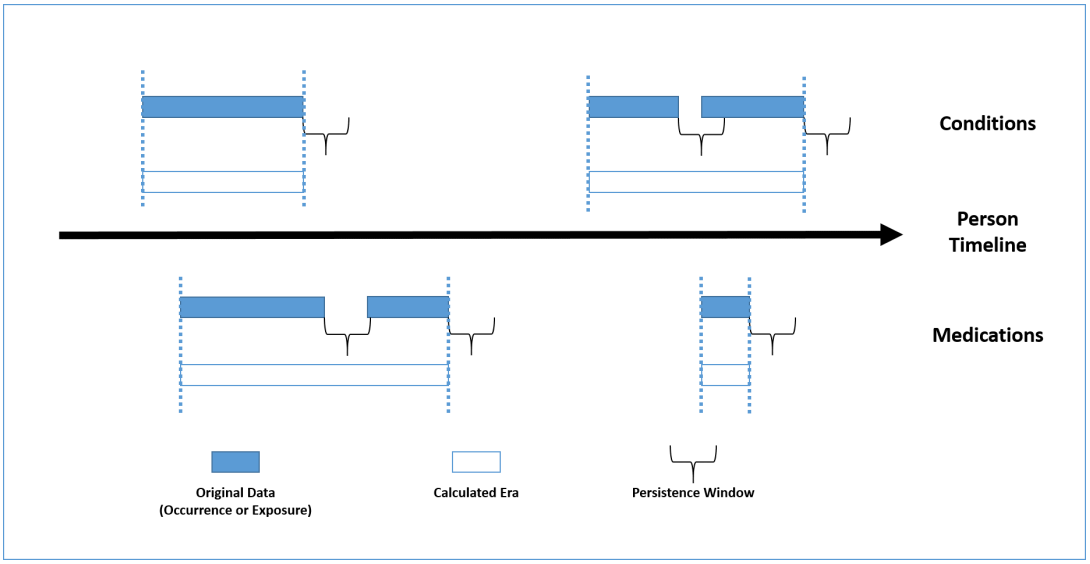



Figure 3.8: Visualization of Condition and Drug Era Calculation

in the persistence window, then an era is formed from a single exposure. All drug exposures are either merged into an already existing drug era or constitutes a drug era itself. Consequently, every single drug exposure is associated with a drug era. Similarly, conditions eras are calculated from condition occurrences using the same persistence window approach.

In proposed framework, OMOP DB Adapter performs the era calculation. First, all condition occurrences and drug exposures are generated using the records from the source database. Later, era calculation is performed on the fly during the generation of SQL Insert statements from the OMOP Ontology instances.

## 3.4 Temporal Pattern Discovery Tool

Semantic transformation framework implements the Temporal Pattern Discovery Tool, which serves as the user interface, on the purpose of providing easy-to-use analysis tool for pharmacovigilance researchers. Requirements of such a tool can be listed as follows;

- Execution of standard OMOP Methods on underlying OMOP CDM instance

- Hiding unnecessary details from user such as RDBMS connection parameters, thus enabling users to focus on the analysis itself
- Modification of the analysis based on user defined parameters, in other words enabling users to define their own queries based on standard OMOP methods
- Use of any desired terminologies independent of underlying data during query definition
- Persistence storage for queries and results in an attempt to increase sharing and re-use
- User friendly interfaces for analysis results, i.e. visualization

In the scope of this work, Temporal Pattern Discovery Tool has been developed as a web application in order to address above requirements. The tool consists of three main parts: the client, the web service and the method. Client component, which is the main interface with user interaction, has been implemented with the latest high performance web technologies incorporating HTML5 design principles and RESTful client-server communication so that it can be run on any modern browser without distributing and installing any software.

The tool is mainly composed of two methods, Temporal Association Screening (TAS) and Temporal Pattern Characterization (TPC). TAS enables broad scale screening for signals of the electronic health record (EHR) data available through SALUS. A statistical measure is used as a threshold for what can be suspected to be a causal association between a drug and a potential ADR. TPC gives the users a visual representation of the temporal pattern of a specific drug and event in the patient population. The visual representation is called a chronograph and is especially good for the detection of potential confounding factor like the indication for treatment.

Web based user interface of the Temporal Pattern Discovery Tool provides three main capabilities in three different tabs. First two tabs, namely Association Screening and Pattern Characterization tabs, allow criteria definition for TAS and TPC methods respectively. Final Results tab displays persisted queries and

results. In the rest of the chapter, descriptions of these tabs together with details of TAS and TPC methods are presented.

### 3.4.1 Temporal Association Screening

TAS tool is based on a standard method from the OMOP Methods Library [11], IC Temporal Pattern Discovery. It is an R method designed specifically for OMOP CDM on the purpose of measuring disproportionality in EHR data by IC$\Delta$ metric proposed by Noren [44].

In Association Screening tab, user can specify group of drugs and conditions as well as set of other parameters such as the name of the era tables or threshold values for the IC$\Delta$ metric. IC Temporal Pattern Discovery written in R language is modified based on the user-specified criteria by the server component. Finally, modified R script is executed over the underlying OMOP database in order to calculate the disproportionality between all drug-condition pairs specified by the users. If no condition and/or medication is specified, then method considers all conditions and/or medication available in the database.

Temporal Pattern Discovery Tool enables users to define condition and medication criteria in any terminology of choice, independent from the underlying data source. By using mappings from the SALUS Terminology Server, specified concepts are mapped to target terminology system of the underlying data source. SALUS pilot scenario can be given as an example to illustrate the role of terminology server in the criteria selection. In care site, different terminologies are used to represent conditions, such as ICD-9 and ICD-10 German Modification whereas MedDRA is used in the research side. Temporal Pattern Discovery Tool enables user to define their conditions on interest in MedDRA and performs the terminology mapping seamlessly depending on the underlying data source. Furthermore, tool provides auto-complete feature on the condition and medication names to increase the usability. Association Screening tab and the criteria definition interface are presented in Figure 3.9.

R script executed on the OMOP database produces result files in text format.

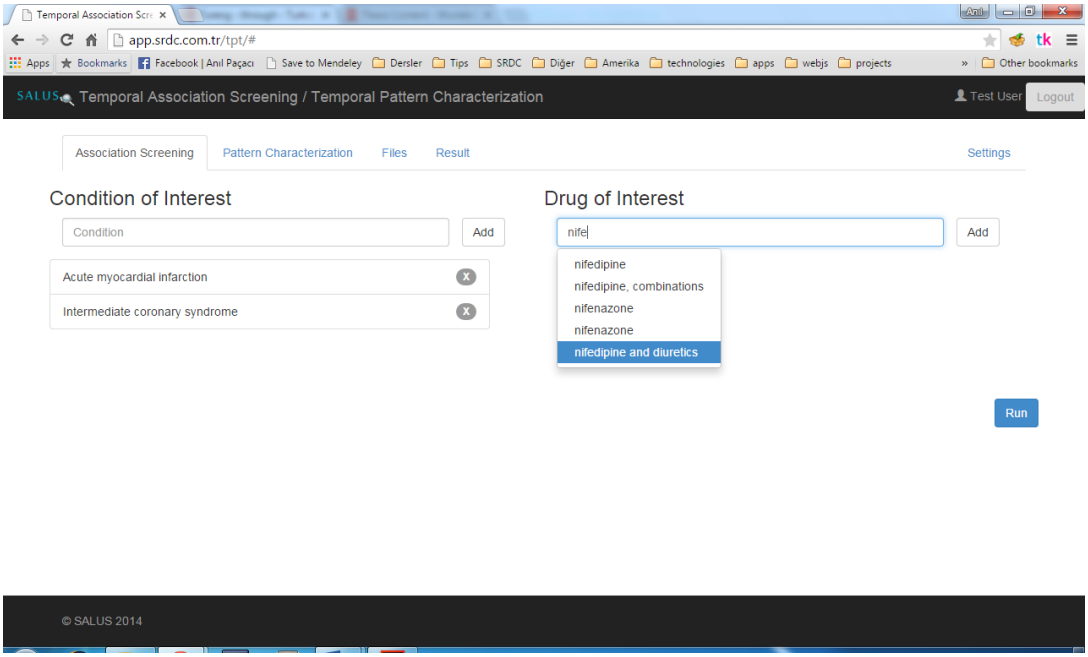


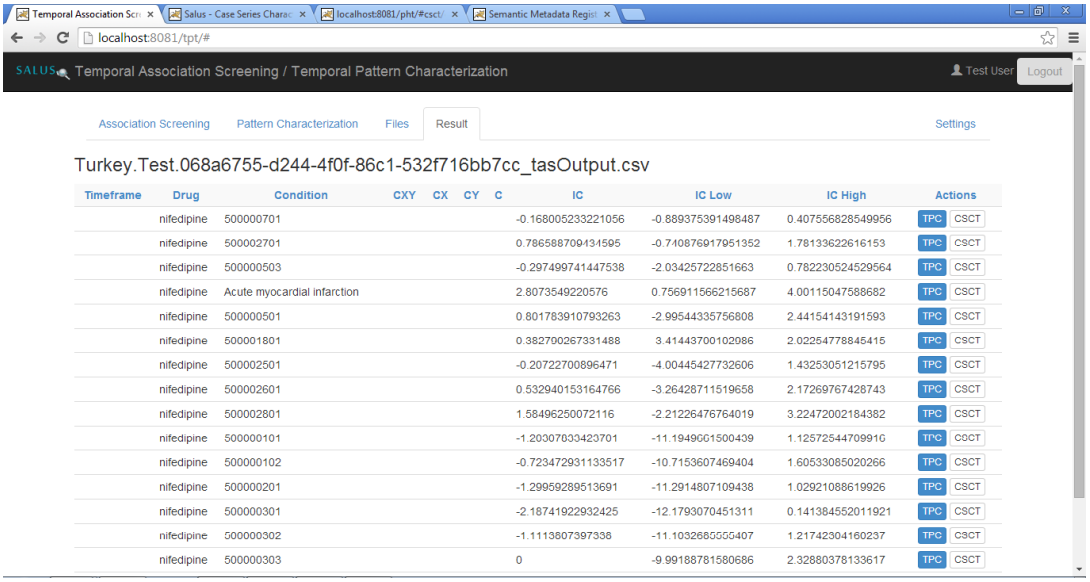

Figure 3.9: Criteria Definition Interface

Temporal Pattern Discovery Tool process those text files and presents in a tabular view. Through the pagination mechanism, users can navigate within different drug-condition associations. Users can also sort the associations based on their attributes, i.e. names of conditions and medications or calculated IC values. Figure 3.10 presents the result interface for the TAS method. In this example, suspected associations of the nifedipine with all conditions are presented.

### 3.4.2 Temporal Pattern Characterization

The statistical measure derived from the Temporal Association Screening tool is a measure of disproportionality taking some confounders into account. To further analyze a specific drug-event pair a visual representation of the temporal pattern was introduced by Norén et al [43] called chronographs. Objective of the Temporal Pattern Characterization is to produce chronographs based on specific drug-condition pair.

TPC starts with specification of eligibility criteria which may include different drug-event combinations as in TAS. As well as this process can be initiated from




| Timeframe | Drug | Condition | CXY | CX | CY | C | IC | IC Low | IC High | Actions |
|---|---|---|---|---|---|---|---|---|---|---|
| | nifedipine | 500000701 | | | | | -0.168005233221056 | -0.889375391498487 | 0.407556828549956 | TPC CSCT |
| | nifedipine | 500002701 | | | | | 0.786588709434595 | -0.740876917951352 | 1.78133622616153 | TPC CSCT |
| | nifedipine | 500000503 | | | | | -0.297499741447538 | -2.03425722851663 | 0.782230524529564 | TPC CSCT |
| | nifedipine | Acute myocardial infarction | | | | | 2.8073549220576 | 0.756911566215687 | 4.00115047588682 | TPC CSCT |
| | nifedipine | 500000501 | | | | | 0.801783910793263 | -2.99544335756808 | 2.44154143191593 | TPC CSCT |
| | nifedipine | 500001801 | | | | | 0.382790267331488 | 3.41443700102986 | 2.02254778845415 | TPC CSCT |
| | nifedipine | 500002501 | | | | | -0.207227008964 71 | -4.00445427732606 | 1.43253051215795 | TPC CSCT |
| | nifedipine | 500002601 | | | | | 0.532940153164766 | -3.26428711519658 | 2.17269767428743 | TPC CSCT |
| | nifedipine | 500002801 | | | | | 1.58496250072116 | -2.21226476764019 | 3.22472002184382 | TPC CSCT |
| | nifedipine | 500000101 | | | | | -1.20307633423701 | -11.1949601500439 | 1.12572544709916 | TPC CSCT |
| | nifedipine | 500000102 | | | | | -0.723472931133517 | -10.7153607469404 | 1.60533085020266 | TPC CSCT |
| | nifedipine | 500000201 | | | | | -1.29959289513691 | -11.2914807109438 | 1.02921088619926 | TPC CSCT |
| | nifedipine | 500000301 | | | | | -2.18741922932425 | -12.1793070451311 | 0.141384552011921 | TPC CSCT |
| | nifedipine | 500000302 | | | | | -1.1113807397338 | -11.1032685555407 | 1.21742304160237 | TPC CSCT |
| | nifedipine | 500000303 | | | | | 0 | -9.99188781580686 | 2.32880378133617 | TPC CSCT |

Figure 3.10: Tabular Presentation of the Temporal Association Screening (TAS) Result

TAS results, it is possible to define from scratch i.e. specifying the condition and medication of interests from the Pattern Characterization tab. The criteria definition is almost identical to TAS, the only difference is that TPC requires exactly one drug-condition pair.

Once the eligibility criteria is defined by the safety analyst, the method with specified parameters is executed and the chronograph is generated from the calculated statistics as illustrated in Figure 3.11. The chronograph image is returned to the interface to be displayed to the user.

The lower half of shows a frequency diagram of occurrences of medical events in different time periods in relation to the drug prescription. In this example there are almost 4000 medical events in the first time period after drug prescription. The time periods can be of a custom defined length but the standard implementation only allows for time periods of 30 days. The graph in the upper half show the IC value in the different time periods and allows for easier identification of changes in the ratio of observed to expected number of cases.

Analyzing a chronograph gives the safety analyst a visual representation of the empirical basis for a possible association between a drug prescription and an

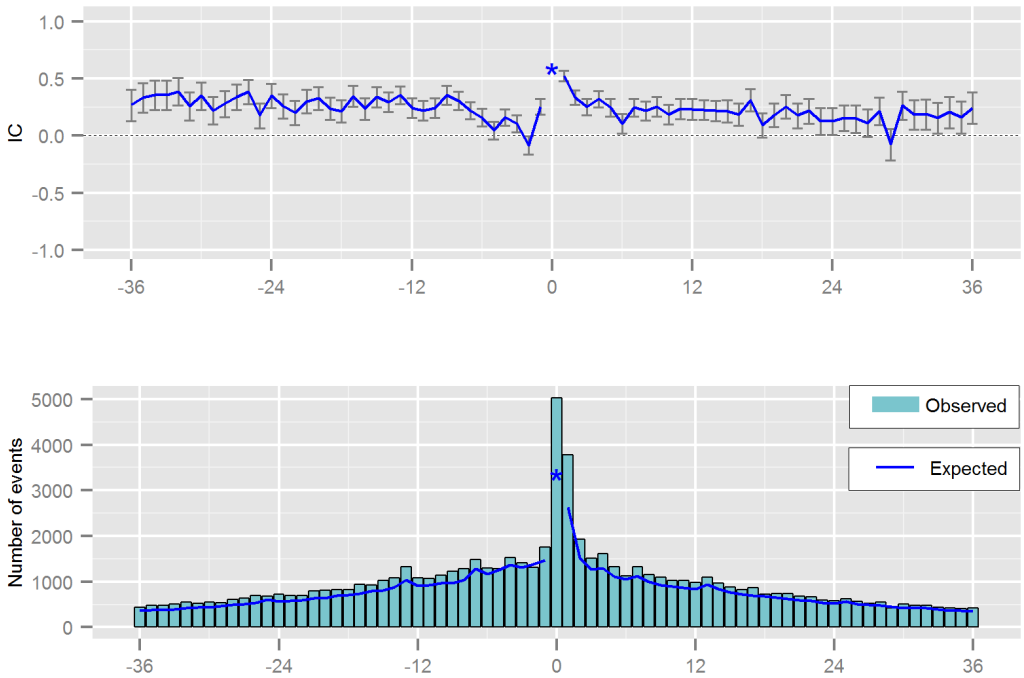



Figure 3.11: Chronograph visualizing the temporal pattern between a prescription of a drug and the occurrence of medical event

event. A consistently high IC value indicates a higher incidence of the event in this population compared to a background. A temporary increase of the IC value in the months prior to drug prescription could indicate a reversed causality i.e. the event I causing the prescription of the drug e.g. indications. An increase of the IC value in the months after the drug prescription could be a sign of a potential signal. However, such patterns do not have to imply causality since a number of reasons can exist that can explain the pattern.

Table 3.2: Complete EYE Transformation Rule Set. Names of datatypes in both content model has been given in lowercase in order to distinguish from class names

| **SALUS CIM Construct** | **OMOP CEM Equivalent** | **EYE Function (Rule) Name** |
|---|---|---|
| Patient | PERSON | Patient2Person |
| Encounter | VISIT_ OCCURANCE | Encounter2VisitOccurance |
| Allergy | CONDITION_ OCCURANCE | Allergy2 ConditionOccurance |
| Condition | CONDITION_ OCCURANCE | Condition2 ConditionOccurance |
| Medication | DRUG_EXPOSURE | Medication2DrugExposure |
| Immunization | DRUG_EXPOSURE | Immunization2 DrugExposure |
| Procudure | PROCEDURE_ OCCURANCE | Procedure2 ProcedureOccurance |
| Result | OBSERVATION | Result2Observation |
| Pregnancy | OBSERVATION | Pregnancy2Observation |
| SocialHistory | OBSERVATION | SocialHistory2Observation |
| Condition | DEATH | Condition2Death |
| HealthCareProvider | PROVIDER | HealthCareProvider2 Provider |
| OrganizationInformation | CARE_SITE & ORGANIZATION | OrganizationInformation2 CareSite |
| CD | CONCEPT | cd2concept |
| Addr | LOCATION | addr2Location |
| Ivlts | start & end dates | ivlTs2start/end dates |
| Ii | String | ii2string |

# CHAPTER 4

# PILOT STUDY: RUNNING EXPLORATORY ANALYSIS STUDIES OVER EHRS FOR SIGNAL DETECTION

One of the overall objectives defined in the scope of SALUS Project, which is described in Section 2.2 in detail, is to support detection and clinical evaluation of potential signals through standardized analyses on top of the available data sources. In this thesis, a semantic transformation framework is proposed in an attempt to address this objective. A pilot study has been conducted to evaluate and validate the proposed framework in the scope of the SALUS Project. In this chapter, scenario of the pilot study and the result obtained during the pilot are presented respectively.

## 4.1 Pilot Scenario

As described previously, one of the main objectives of SALUS Project is to enable querying and subscription of subsets of medical summaries from EHR Systems and supporting data warehouses with the intent of reinforcing postmarketing surveillance. To achieve this, a central clinical data repository is set up that specifies the clinical data collected from the underlying EHRs, namely LISPA (A Regional Health Data Warehouse (DWH) is maintained in Lombardy Region in Italy) and TUD (EHR system at Uniklinikum Dresden (UKD)). LISPA regional database includes anonymized data of 16 million patients with over 10 years longitudinal data on the average. TUD EHR System on the other hand,

contains records belonging approximately a million patients. This repository subscribes to these clinical data sources through standard based SALUS interoperability profiles. The clinical data fed to this repository through the semantic transformation framework is converted to the native data model. In SALUS settings, OMOP CDM [3] is the choice of target data model since it is employed by the corresponding SALUS end-user, WHO-UMC (World Health Organization - Uppsala Monitoring Centre). In this way, available algorithms such as the ones from OMOP Methods Library [11] developed in the OMOP Project for signal detection can be continuously executed against EHRs. It becomes possible to automatically identify emerging patterns of temporal association between drug prescriptions and medical events for detailed clinical review. Figure 4.1 clearly visualizes this scenario and the setup.

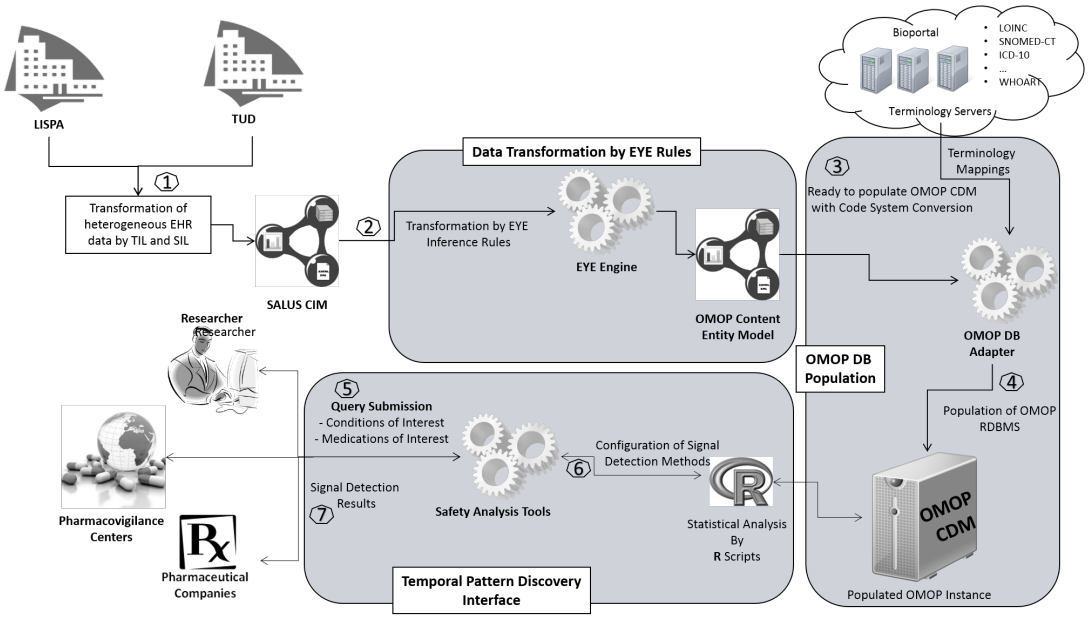



Figure 4.1: Deployment of the Semantic Transformation Framework on top SALUS Electronic Health Record Sources

One common scenario in pharmacovigilance centres such as WHO-UMC starts with analysis of the whole patient database of potential causally related drug-adverse drug event pairs. In this step, analysts may choose one the following paths:

- Consider all drugs against all adverse drug events

- Consider one or more drugs against one or more adverse drug events
- Consider a single drug against a single adverse drug event

By following these paths, analysts may ask questions such as "I am interested in the safety of vancomycin, can you give all the events that are likely to be temporally related?" or "I would like to know all drugs that could cause acidosis". Temporal Association Screening (TAS) method of the Temporal Pattern Discovery tool has been designed to address all of these functionalities. Steps of an example TAS scenario can be described as follows:

1. Safety analyst defines the inclusion criteria whether it be a specific drug, a specific event, all drugs and events or a specific combination of a drug and event.
2. Safety analyst clicks on the run button and waits for the results (The execution times can be long if the inclusion criteria include many drugs and events).
3. The query is sent to the Temporal Association Screening web service, which handles the input and creates the necessary configuration files for the TAS method.
4. The web service starts the execution of the method on top of the Clinical Data Repository in OMOP CDM Format.
5. The method generates the statistical measure, IC$\Delta$, for all of the selected drug and event combinations.
6. Safety analyst can look at and decide whether or not the drug and event combination might have the potential of becoming a signal.

After choosing a particular pair, analysts most often needs to substantiate the validity of the selected pair. It is possible to progress the scenario in one of the following ways:

- Temporal Pattern Characterization

- Case Series Characterization

The latter analysis is not in scope of this thesis. The former TPC method enables analysts to see the occurrence probability of the adverse drug event considering the drug intake time relative to the event time. At this point, Temporal Pattern Characterization (TPC) of the Temporal Pattern Discovery Tool comes into stage. Actually, TAS and TPC methods are integrated in the same application on purpose such that the analysts are able to initialize TPC after the TAS. However, it is also possible to initiate these processes separately through the same user interface. Steps of an example TPC scenario can be described as follows:

1. Safety analyst specifies a single drug-event pair as an eligibility criteria. This pair can be chosen either from a former TAS result or from scratch by using the Pattern Characterization tab of the Temporal Pattern Discovery tool.
2. Safety analyst clicks on the run button and waits for the results (The execution times can be long depending on the size of underlying data source).
3. The query is sent to the Temporal Pattern Characterization web service, which handles the input and creates the necessary configuration files for the TPC method.
4. The web service starts the execution of the method on top of the Clinical Data Repository in OMOP CDM Format.
5. The chronograph image is returned to the interface to be displayed to the safety analyst.

## 4.2 Results

In order to asses the effectiveness and validity of the proposed semantic transformation framework, a comprehensive validation study has been planned and

performed based on the given pilot scenario definition. Quality and the accuracy of the populated data have been compared with native database in a detailed comparative analysis. Additionally, efficacy and the usability of the proposed framework have been evaluated by a set of end-users. More detailed information about these activities along with the results obtained is presented in following subsections.

### 4.2.1 Results of Comparative Analysis

Since this thesis claims to enable easy-to-use, scalable transformation methodology through high level semantic mappings, we have validated the proposed approach's usefulness by evaluating its performance in transforming real-world observational datasets into OMOP CDM. Although data warehouses of TUD and LISPA differ in terms of the schema and the vocabularies used, SALUS technical interoperability framework enables retrieval of SALUS CIM conformant data from both databases. Therefore, transformation processes are identical in two pilot sites. In this section, evaluation results obtained from TUD pilot are presented.

OMOP team created the Observational Source Characteristics Analysis Report (OSCAR) [14] to characterize the data once they were converted to the OMOP CDM. This program allows comparison of the data between the native and transformed database so that the accuracy of the transformation processes can be assessed. Running OSCAR over OMOP CDM requires Statistical Analysis Software (SAS) [19] which is not available in TUD settings. Since original TUD data warehouse and OMOP CDM are installed on Oracle Database, SQL queries were developed for both schema to calculate similar statistics to OSCAR. Those statistics were extracted and compared for two patient cohorts as well as the overall data, namely patient with acute myocardial infarction occurrence and patients with nifedipine exposure.

The number of total patients in the original TUD data warehouse is 893,870 and 95.66% of those patient were moved to OMOP CDM instance during the transformation process. Number of patients in nifedipine cohort accounts for 0.013%

of the total TUD patients whereas number of patients in acute myocardial infarction (AMI) cohort corresponds to 0.287% of the population. All 113 patients in original nifedipine cohort are preserved in the OMOP CDM instance. On the other hand, 6 of the patients in the original AMI cohort are not in the OMOP CDM instance. Table 4.1 provides the statistics on number of patients obtained from both native and transformed databases and it shows that the statistics are well aligned.

Table 4.1: Population Counts in Native TUD and Populated OMOP Databases

| | **Native TUD Database** | **Populated OMOP CDM Instance** |
|---|---|---|
| Number of patients | 893870 | 855101 |
| Number of patients in nifedipine cohort | 113 | 113 |
| Number of patients in AMI cohort | 2562 | 2556 |

Detailed analysis on patients who could not be transferred reveals that those patients were left out by OMOP DB Adapter intentionally. Those are the patients with inadequate demographic information, particularly date of birth. Since the age information is essential to targeted analyses, those records simply were not considered for the population.

Demographic characteristics of the underlying patient population have been preserved to a great extend during the transformation process. Average age statistics have been calculated for both gender groups and whole population and presented in Table 4.2. Figure 4.2 presents the gender distribution of the nifedipine and AMI cohorts as well as the overall population in both native TUD and populated OMOP databases. Table 4.3 compares the statistics calculated for nifedipine exposures and AMI occurrences in those databases. In line with the previous findings, both results demonstrate substantial amount of similarity between two native and populated databases.

99.93% of the all medication records in drug exposure table were transformed to OMOP CDM instance. Top 100 most frequent medications accounts for 52.55%

Table 4.2: Patients' Age Statistics in Native TUD and Populated OMOP Databases

| | **Native TUD Database** | **Populated OMOP CDM Instance** |
|---|---|---|
| Average age (overall) | 50.723 | 50.877 |
| Average age (male) | 49.917 | 49.879 |
| Average age (female) | 51.791 | 51.848 |

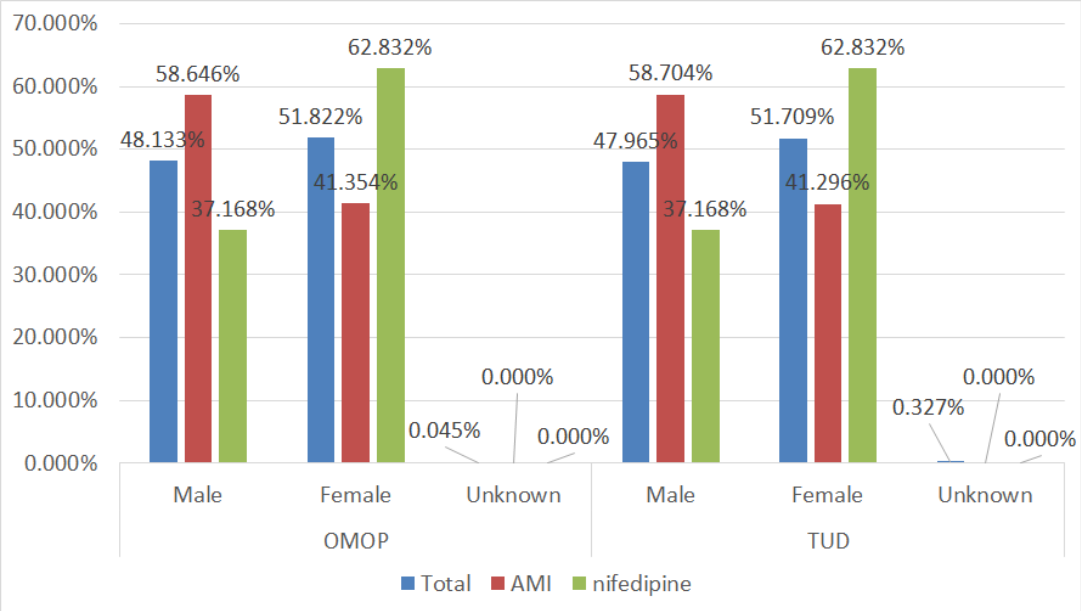



Figure 4.2: Demographic summary of AMI and nifedipine cohorts in original TUD and populated OMOP database

of the drug exposure data and 96.87% of those were correctly reflected in the populated OMOP CDM instance. Analysis on the unmapped drug exposure shows that those records were not transformed on purpose since they miss essential data like the medication code or the medication start data. In Figure 4.3, 50 most frequent medications with their incidence rates in native TUD and populated OMOP database is presented. It shows that incidence rates of medication in both databases is well aligned.

Similar analysis were conducted for condition occurrences. This time, 99.59% of the condition occurrences were successfully transformed. Top 100 most occurring conditions made up 29.96% of all data and 94.55% of these occurrences

Table 4.3: Performance of Data Transformation Process on nifedipine and AMI Cohorts

| | **Native TUD Database** | **Populated OMOP CDM Instance** |
|---|---|---|
| Nifedipine | | |
| Total exposures | 494 | 494 |
| Average exposure per patient | 4.371 | 4.371 |
| Acute Myocardial Infarction | | |
| Total occurrences | 2562 | 2556 |
| Average occurrence per patient | 2.618 | 2.623 |

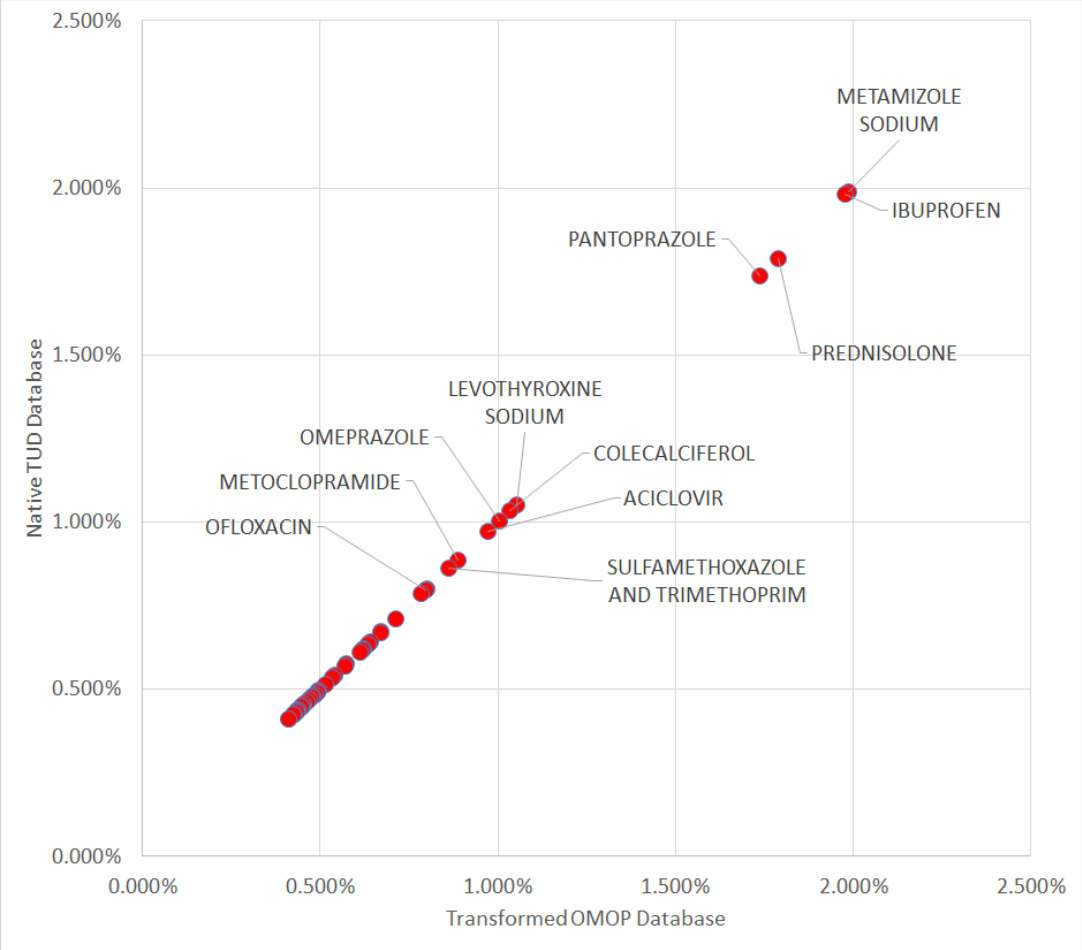



Figure 4.3: Proportion of Top 50 Medications to Whole Exposures in Native TUD and Populated OMOP Databases

were mapped to OMOP CDM instance. Incidence rates 50 most occurring conditions in both databases is presented in Figure 4.4. It also presents an almost linear trend line as medication incidence rates except few conditions. Cause of these abnormalities is the differences between the vocabularies of source and target databases. In TUD a local version of ICD10, ICD10-GM (German modification), is used to annotate conditions which is not part of the standardized OMOP Vocabulary. Therefore ICD10-GM was loaded to OMOP CDM instance manually before the transformation and any conflict between these code-lists were handled by the OMOP DB Adapter during the transformation. For instance, number of conditions which are not in the top 100 in original data were mapped to 'Essentielle (primäre) Hypertonie' – 'I.10-' and created inconsistency original and transformed data. However, this problem effects negligible portion of the condition records as described previously.

### 4.2.2 Results of End-user Evaluation

In order to assess whether the proposed framework fulfills the intended use from an end-user point of view or not, it has been tested and evaluated by real end-users from WHO-UMC. An evaluation and validation framework based on the ISO/IEC 25040 Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) – Evaluation reference model and guide [8] has been developed. This analysis covers the validation of functionalities provided by Temporal Pattern Discovery Tool, since it is the part of the proposed framework which end-users interacts with.

Total of 4 end-users have taken part in the validation activities from WHO-UMC. Three of these end users are research pharmacists with professional experience in pharmacovigilance ranging from 2-4 years; and the other one is a senior researcher with more than 15 years of experience in pharmacovigilance and signal detection. This group of pharmacovigilance researchers (end users) obtained training in using the tools until sufficiently proficient. They were then instructed to use the tools based on the scenario presented in section 4.1. Objective is to investigate the research questions that the system was designed to


Figure 4.4: Proportion of Top 50 Conditions to Whole Occurrences in Native TUD and Populated OMOP Databases

be able to provide answers to.

A specific questionnaire for Temporal Pattern Discovery Tool has been designed based on the Health IT Usability Evaluation Scale (Health-ITUES) [54] on the purpose of assessing the effectiveness. This Health-ITUES tool is developed by the Department of Biomedical Informatics, The Ohio State University, Columbus, Ohio, USA. An on-line version of the designed questionnaire has been provided to evaluators from WHO-UMC. As a result, Temporal Pattern Discovery Tool is found to be a useful and effective for finding suspicious condition - medication pairs, analysing association screening and pattern characterization results. HEALTH-ITUES items which evaluators agree on can be listed as:

- Using the tool makes it easier to add/set Condition of Interest and Drug of Interest values for the Association Screening Pattern Characterization processes.
- Using the tool makes it easier to see and analyse the results of the Association Screening and Pattern Characterization processes.
- Using the tool makes it easier to execute Pattern Characterization or Case Series Characterization from the Association Screening Results.
- Using the tool is useful for seeing and analysing the results of the Association Screening and Pattern Characterization processes.

In order to evaluate the usability characteristics of the Temporal Pattern Discovery Tool, System Usability Scale (SUS) [28] has been used. SUS is a widely adopted measuring mechanism for the usability of computer software. It provides a well-established scale so that the results can be compared within a global perspective. SUS score calculated for Temporal Pattern Discovery Tool is 64, which is merely below the expected average score 68.

In addition to questionnaires, end-users have also participated in a focus group interview. Focus group is a form of qualitative research in which a group of people are asked about their perceptions, opinions, beliefs, and attitudes towards a product, service, concept, advertisement, idea, or packaging. Comments and feedbacks from participants have been obtained as an output of this interview. Different aspects regarding the usability of the tool have been highlighted in the focus group discussions, which explains the low SUS score. These aspects are:

- It is not possible to do real time searches since the queries took long to execute.
- Some error messages did not contain sufficient information to understand the real problem.

Apart from these limitations, evaluators indicated that the possibility of selecting different credibility intervals in certain statistical measures is an important

plus. Mora importantly, they highlighted that the tool indeed has the potential to provide useful information in signal detection work.

# CHAPTER 5

# RELATED WORK

## 5.1 Existing Common Data Models

There is an extensive research in literature in an attempt to utilize huge amount of data available in heterogeneous healthcare dataset for the postmarketing surveillance. As introduced in Chapter 1, there are different data models developed by different organizations, each having its own advantages and limitations.

Analysis Data Model (ADaM) which is developed by Clinical Data Interchange Standards Consortium (CDISC) is one of the most prevalent CDM implementations. CDISC provides number of data standards for exchange, storage and submission of clinical trials to the United States Food and Drug Administration (FDA). Objective of the ADaM is to specify fundamental principles and standards for the creation of analysis dataset and its related meta-data for the clinical trials [1]. Normally, data collected during the clinical trials is submitted to FDA in Study Data Tabulation Model (SDTM)[20] from CDISC which is built around the concept observation and organizes data into clinical domains. However, SDTM is not favourable for statistical analysis. Purpose of the ADaM is to collect all these data from different SDTM domains into a single database so that analysis can be executed homogeneously with little or no additional effort. Set of well-defined analytic methods is available to serve this purpose. However, ADaM does not specify the use of standardized vocabulary, which creates an obstacle to achieve the semantic interoperability.

Another important initiative in the literature is Biomedical Research Integrated

Domain Group Model (BRIDG) [2], which is a joint effort of CDISC, FDA, National Cancer Institution and HL7 Regulated Clinical Research Information Management Technical Committee (RCRIM). Objective behind the BRIDG model is to represent the semantics of the data in the clinical trials and pre-clinical research. For the clinical research community, shared semantics are crucial component to achieve computable semantic interoperability in order to enable machine-to-machine meaning transfer. BRIDG Model is represented as visual representations in the form of UML Diagram, Hl7 Reference Information Model [18] and Web Ontology Language [15]. In addition to suffering from absence of library of methods, BRIDG model is criticized not specifying a standardized vocabulary [45].

Last but not least, Observational Medical Outcomes Partnership also focuses on use of observational health data for drug safety research. OMOP's focus is to assess feasibility and utility of the observational healthcare datasets for identification and evaluation of drug – condition associations. OMOP has defined a common data model and developed a set of standard analytical and statistical methods to identify drug - condition pairs over this common data model. OMOP CDM not only defines the structure and the organization of the data, but also defines the content of the data with use of standardized terminologies in order to harmonize different data sources, in contrast to its counterparts. OMOP is also unique in a way that it promotes transparency by serving all information of interest to the public. OMOP has the possibility of being an important instrument for the analyses of real world data, with appropriate processes and training to ensure continual use of the CDM and standardized methods by different users and organizations.

## 5.2 Survey on Existing Common Data Model Transformation Efforts

In literature, there also exist number of studies which employ the CDM based approach on various healthcare datasets. Their purpose is to validate effectiveness, usefulness and performance of different CDMs for drug safety surveillance by systematical analysis and comparing descriptive statistics extracted from the

native EHR database and transformed CDM database. Among others, OMOP CDM has been found to be superior particularly for addressing diverse requirements of both researchers and data owners. Ogunyemi et al. performed transformation into ADaM, BRIDG, Mini-Sentinel and OMOP CDM with the purpose of realizing comparative analyses of these most common CDMs [45]. As comparison criteria, authors evaluated the extensibility of the model, adoption of standardized terminologies, availability of library of analytic routines and loss of information during transformation. Although each model has its own benefits and limitations, OMOP CDM is found to be one which meets the broadest set of needs with regards to the usage objectives.

In a study conducted in US, Reisinger et al. [48] have developed a CDM and applied on an administrative claim and an EHR database and performed evaluations. They have adopted the person-centric design, which is inspired by OMOP CDM, by organizing healthcare records in person time-line as drug and condition eras. They perform the transformation of raw data into CDM in three steps; respectively one-to-one mapping of source tables to CDM constructs, mapping condition and drug concepts using terminology dictionary and aggregation of drug and condition eras for subsequent healthcare encounters. For the validation of the CDM, authors compared the descriptive statistics obtained from both raw and transformed data. In consequence of this comparative evaluation, authors deduce that developed CDM is a feasible and necessary tool enabling integration of disparate data sources and analysis into a coordinated system. Authors point out the importance of systematic mappings of terminologies as a tool to maximize the semantic interoperability.

In another study conducted by Overhage et al. in [46], the OMOP CDM has been implemented on 10 different US medical databases in order to validate its usefulness for large scale safety analysis. After transformation of local databases into CDM instances, they assessed the suitability and performance by executing same methods on all instances. Terminology mappings retrieved from Unified Medical Language System (UMLS) [23] is used to enable common vocabulary across the databases. Descriptive statistics about the underlying data are retrieved by using OMOP tools OSCAR [14] and GROUCH [5] with the intent of

evaluating the quality of transformation process. As a result, authors report that 93.2% to 99.7% of condition records and 88.8% to 97.6% of medication records were transformed successfully. A set of standard analyses could be executed on all instances successfully without any modification. As a result of this empirical study authors claim that OMOP CDM and the standardized terminologies offer compelling benefits for safety surveillance applications although CDM based approach can cause loss of granularity and richness compared to the source data.

Generally, CDM transformation and validation efforts is related to transformation of US EHR databases like the two examples described previously. There have also been some efforts to validate CDM approach in European data sources, such as the study held by Zhou et al. [57]. Zhou et al. investigates the appropriateness of OMOP CDM structure for the UK medical records and validates the data transformation by using different statistical methods. The Health Improvement Network (THIN) [22] is an electronic health record database developed in UK, which has been used in pharmacoepidemiological studies and validated extensively [39]. Transformation methodology proposed by Zhou et al. [57] is based on 4 main steps; transformation of raw THIN contents in to CDM tables, enhancement of the data with concept codes from OMOP vocabulary, aggregation of drug and condition eras by using persistence windows of 30 days and finally documentation of all the transformation process using ETL specifications provided by OMOP team [37]. For the validation of the CDM THIN, authors choose 9 Health Outcomes of Interest (HOI) and 10 Drugs of interest (DOI) and apply several analyses against them in order to compare the raw THIN and the CDM THIN. These analyses include a dedicated programming as well as standard OMOP methods OSCAR [14], NATHAN [12]and GROUCH [5]. Except some incomplete mappings, which only have little effects on selected HOIs and DOIs, CDM THIN produces almost identical results to original data while performance of analyses improves up to two times. In the light of these analyses, authors conclude that the OMOP CDM is a valuable tool for pharmacoepidemiological research although structure of the THIN presents challenges to effectively benefit this tool.

In a more recent study held by Matcho et al. [40], previous CDM transformation

efforts are criticized. Transformation of US claim databases and EHR data into OMOP CDM performed by Overhage et al. [46] is described as one of the successful examples with high range of mapped condition and drug records. On the other hand, insufficient amount of records mapped in the previously described THIN transformation effort [57] shows that structure of the THIN creates obstacles for successful pharmacoepidemiological analyses. Clinical Practice Research Datalink (CPRD) is a population based UK observational data warehouse, which is similar to THIN in terms of content and structure. Thereby, Matcho et al. [40] addresses these limitations by performing a transformation into OMOP CDM and assessing its adequacy by replicating the previous study of Schlienger et al. [50]. By designing ETL for transforming database records and providing mappings between source and target terminologies, they have managed to transfer 99.9% of condition records and 89.7% of medication records into CDM instance accurately. In case of medication records, information loss is higher than the THIN study [57]. However, by conducting a case study; it has been proven that those unmapped codes did not hinder the analyses significantly since results obtained from the CPRD raw data and the CDM data were comparable. Authors stress that while CDM provides normalization mechanism for all codes into standard terminology, it also maintains the original source codes so that specific analyses also with the original codes can be possible.

# CHAPTER 6

# CONCLUSIONS AND FUTURE WORK

Today, postmarketing studies occupy an important role in detection of possible Adverse Drug Events (ADE) to ensure safety of medical products. Compared to the pre-market clinical trials, postmarketing studies have solid advantages such as covering longer timespan and being cost-effective. Historically, postmarketing surveillance activities depend on spontaneous case reports which suffers from problems such as under-reporting and low content quality. Recent trend in pharmacovigilance is to facilitate the secondary use of Electronic Health Records (EHR). By its very nature, EHR data includes more comprehensive information regarding the patient history than spontaneous case reports. More importantly, this data is already available at no additional cost.

Use of existing EHR data for postmarketing surveillance studies presents its own challenges. The most important among them is the fact that the available EHR data is spread out to different systems with disparate data models. It hinders researchers from analysing the data in a systematic way by using the same methods. Common Data Model (CDM) based interoperability approach, which defines a standardized data structure, transparent transformation rules, develops and specifies a common vocabulary to represent the concepts in a specific domain, addresses this challenge. CDM enables researchers to develop their analysis independent from underlying data source and produce comparable results. In the field of medical informatics, there exists number of CDMs proposed by different organization such as BRIDG, ADaM and OMOP CDM.

In literature, there are number of studies which aim to validate CDM based

interoperability approach for postmarketing surveillance and transform different observational datasets into a common data model. A common problem reported in all of these studies is that adopting this approach requires significant amount of expertise, which is not easily transferable to other domains. In addition to details related to semantics of underlying data model, great deal of implementation-specific, technical knowledge about the underlying database is required since transformation is based on proprietary ETL (Extract - Transform - Load) procedures.

Within the scope of this thesis, design and implementation of a novel data transformation framework to address aforementioned challenges have been presented. We have created a semantic representation for the OMOP CDM, the most commonly adopted CDM, and designed a set of abstract, semantic mapping rules to handle the schema transformation. We have developed the OMOP DB Adapter to handle automatic population of the OMOP database from OMOP Content Entity Models. In this way, semantic representation of the OMOP CDM can be used as a middle layer and transformation of schemas can be realized independent from the implementation specific details of underlying databases. On top of this, we have developed the Temporal Pattern Discovery Tool as a proof of concept which seamlessly executes standardized OMOP methods over the populated OMOP database. Novel aspects of the proposed framework can be summarized as follows:

- Semantic representation for the OMOP Common Data Model, called OMOP Content Entity Model, has been created using the Web Ontology Language (OWL). It constitutes a universal, standardized and machine-processable representation for the original OMOP CDM.

- On top of the OMOP Content Entity Model, a novel approach for CDM transformation has been proposed and implemented. Unlike previous ETL based efforts, syntax and semantics of the transformation are addressed in separate levels. Semantic mapping rules defines how the original data can be mapped to OMOP CDM constructs while OMOP CD Adapter is responsible of populating the underlying database instance.

- Temporal Pattern Discovery Tool based on pre-defined analyses from OMOP Methods Library [11] has been designed and implemented. It enables researchers and data analysts to define queries in their vocabulary of choice independent from underlying data source.

Semantic transformation framework has been deployed on SALUS pilot sites, namely TUD and LISPA, in order to evaluate the correctness and the efficacy of the proposed approach in real world settings. Based on the guidelines provided by OMOP [14], several statistics on patient populations of both source and target databases have been calculated to validate the correctness of the data transformation. As a result, it is concluded that the semantic transformation framework correctly populates the OMOP CDM with a very high level accuracy. Based on the pilot application scenarios defined in the scope of the SALUS Project, group of pharmacovigilance researchers used the Temporal Pattern Discovery Tool and provided their experience through standardized surveys and focus group meetings. It is indicated by end-users that the tool indeed has the potential to provide useful information in signal detection research.

As a future work, we plan to increase the utilization of semantic web technologies. As the first step, condition and medication concepts in OMOP Ontology can be linked to existing linked biomedical ontologies [49] and vocabulary mappings can be explored in a more natural manner. This could eliminate the need for a separate terminology server. In addition, set of semantic mapping rules for different data sources can be developed in order to validate the proposed approach in different healthcare settings.

# REFERENCES


[1] Analysis Data Model (ADaM) | CDISC. http://www.cdisc.org/adam. Last visited on May 2015.

[2] Biomedical Research Integrated Domain Group (BRIDG) Model | CDISC. http://cdisc.org/bridg. Last visited on May 2015.

[3] Common Data Model | Observational Medical Outcomes Partnership. http://omop.org/CDM. Last visited on May 2015.

[4] Euler Yet another proof Engine. http://eulersharp.sourceforge.net/. Last visited on May 2015.

[5] Generalized Review of OSCAR Unified Checking | Observational Medical Outcomes Partnership. http://omop.org/GROUCH. Last visited on May 2015.

[6] ICD - ICD-9-CM - International Classification of Diseases, Ninth Revision, Clinical Modification.

[7] International Statistical Classification of Diseases and Related Health Problems 10th Revision. http://apps.who.int/classifications/icd10/browse/2015/en. Last visited on June 2015.

[8] ISO/IEC 25040:2011 - Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Evaluation process. http://www.iso.org/iso/catalogue_detail.htm?csnumber=35765. Last visited on May 2015.

[9] Linked Data | Linked Data - Connect Distributed Data across the Web. http://linkeddata.org/. Last visited on May 2015.

[10] MedDRA - Medical Dictionary for Regulatory Activities. http://www.meddra.org/. Last visited on June 2015.

[11] Methods Library | Observational Medical Outcomes Partnership. http://omop.org/MethodsLibrary. Last visited on May 2015.

[12] NATHAN - Utility of Natural History Information | Observational Medical Outcomes Partnership. http://omop.org/NATHAN. Last visited on May 2015.

[13] Notation3 (N3): A readable RDF syntax. http://www.w3.org/TeamSubmission/n3/. Last visited on May 2015.

[14] OSCAR - Observational Source Characteristics Analysis Report (OSCAR) Design Specification and Feasibility Assessment | Observational Medical Outcomes Partnership. http://omop.org/OSCAR. Last visited on May 2015.

[15] OWL Web Ontology Language Overview. http://www.w3.org/TR/owl-features/. Last visited on May 2015.

[16] RDF - Semantic Web Standards. http://www.w3.org/RDF/. Last visited on May 2015.

[17] RDF Schema 1.1. http://www.w3.org/TR/rdf-schema/. Last visited on June 2015.

[18] Reference Information Model (RIM) Downloads. http://www.hl7.org/implement/standards/rim.cfm. Last visited on May 2015.

[19] Statistical Analysis Software, SAS/STAT | SAS. http://www.sas.com/en_us/software/analytics/stat.html. Last visited on May 2015.

[20] Study Data Tabulation Model (SDTM) | CDISC. http://www.cdisc.org/sdtm. Last visited on May 2015.

[21] Systemized Nomenclature of Medicine Clinical Terms. http://www.ihtsdo.org/snomed-ct. Last visited on June 2015.

[22] THIN - The Health Improvement Network | THIN is an unobtrusive medical data collection scheme that collects anonymised patient information from GP practices using the INPS ViSion software. http://www.thin-uk.net/. Last visited on May 2015.

[23] Unified Medical Language System (UMLS) - Home.

[24] Vocabularies | Observational Medical Outcomes Partnership. http://omop.org/Vocabularies. Last visited on May 2015.

[25] W3C XML Schema. http://www.w3.org/XML/Schema. Last visited on June 2015.

[26] WHO | Pharmacovigilance.

[27] WHO | The Anatomical Therapeutic Chemical Classification System with Defined Daily Doses (ATC/DDD).

[28] A. S. f. P. Affairs. System Usability Scale (SUS). Sept. 2013.

[29] K. Benson and A. J. Hartz. A Comparison of Observational Studies and Randomized, Controlled Trials. *New England Journal of Medicine*, 342(25):1878–1886, 2000.

[30] T. Berners-Lee, J. Hendler, O. Lassila, et al. The semantic web. *Scientific american*, 284(5):28–37, 2001.

[31] T. Brewer and G. A. Colditz. Postmarketing Surveillance and Adverse Drug Reactions. *JAMA*, 281(9):824, Mar. 1999.

[32] R. A. Bright and R. C. Nelson. Automated support for pharmacovigilance: a proposed system. *Pharmacoepidemiology and drug safety*, 11(2):121–5, Mar. 2002.

[33] P. Coorevits, M. Sundgren, G. O. Klein, A. Bahr, B. Claerhout, C. Daniel, M. Dugas, D. Dupont, A. Schmidt, P. Singleton, G. De Moor, and D. Kalra. Electronic health records: new opportunities for clinical research. *Journal of internal medicine*, 274(6):547–60, Dec. 2013.

[34] Food, H. Drug Administration, et al. International conference on harmonisation; e2b (r3) electronic transmission of individual case safety reports; data elements and message specification; appendix on backwards and forwards compatibility; availability. notice. *Federal register*, 79(35):9908, 2014.

[35] D. For and P. Comment. Observational Medical Outcomes Partnership Common Data Model Version 3 Specifications. (April), 2011.

[36] C. D. Furberg, A. A. Levin, P. A. Gross, R. S. Shapiro, and B. L. Strom. The FDA and drug safety: a proposal for sweeping changes. *Archives of internal medicine*, 166(18):1938–42, Oct. 2006.

[37] M. Khayter. OMOP Common Data Model ( CDM V4 . 0 ) ETL Mapping Specification Template. 2012.

[38] R. Kunz and A. D. Oxman. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ*, 317(7167):1185–1190, Oct. 1998.

[39] J. D. Lewis, R. Schinnar, W. B. Bilker, X. Wang, and B. L. Strom. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. *Pharmacoepidemiology and Drug Safety*, 16(4):393–401, Apr. 2007.

[40] A. Matcho, P. Ryan, D. Fife, and C. Reich. Fidelity Assessment of a Clinical Practice Research Datalink Conversion to the OMOP Common Data Model. *Drug Safety*, 37(11):945–959, 2014.

[41] L. E. Moses. Measuring effects without randomized trials? options, problems, challenges. *Medical Care*, pages AS8–AS14, 1995.

[42] J. C. Nelson, A. J. Cook, O. Yu, S. Zhao, L. A. Jackson, and B. M. Psaty. Methods for observational post-licensure medical product safety surveillance. *Statistical methods in medical research*, 24(2):177–93, Apr. 2015.

[43] G. N. Norén, A. Bate, J. Hopstadius, K. Star, and I. R. Edwards. Temporal pattern discovery for trends and transient effects. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08*, page 963, New York, New York, USA, Aug. 2008. ACM Press.

[44] G. N. Norén, J. Hopstadius, and A. Bate. Shrinkage observed-to-expected ratios for robust and transparent large-scale pattern discovery. *Statistical methods in medical research*, 22(1):57–69, Feb. 2013.

[45] O. I. Ogunyemi, D. Meeker, H. E. Kim, N. Ashish, S. Farzaneh, and a. Boxwala. Identifying appropriate reference data models for comparative effectiveness research (CER) studies based on data from clinical information systems. *Medical Care*, 51(8 Suppl 3):S45–52, 2013.

[46] J. M. Overhage, P. B. Ryan, C. G. Reich, a. G. Hartzema, and P. E. Stang. Validation of a common data model for active safety surveillance research. *Journal of the American Medical Informatics Association*, 19(1):54–60, 2012.

[47] T. Piazza-Hepp and D. Kennedy. Reporting of adverse events to MedWatch. *Am. J. Health Syst. Pharm.*, 52(13):1436–1439, July 1995.

[48] S. J. Reisinger, P. B. Ryan, D. J. O'Hara, G. E. Powell, J. L. Painter, E. N. Pattishall, and J. a. Morris. Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases. *Journal of the American Medical Informatics Association : JAMIA*, 17(6):652–662, 2010.

[49] M. Salvadores, P. R. Alexander, M. A. Musen, and N. F. Noy. BioPortal as a Dataset of Linked Biomedical Ontologies and Terminologies in RDF. *Semantic web*, 4(3):277–284, Jan. 2013.

[50] R. G. Schlienger, H. Jick, and C. R. Meier. Use of nonsteroidal anti-inflammatory drugs and the risk of first-time acute myocardial infarction. pages 327–332, 2002.

[51] N. Shadbolt, W. Hall, and T. Berners-Lee. The semantic web revisited. *Intelligent Systems, IEEE*, 21(3):96–101, 2006.

[52] M. Suling and I. Pigeot. Signal detection and monitoring based on longitudinal healthcare data. *Pharmaceutics*, 4(4):607–40, Jan. 2012.

[53] D. K. Wysowski and L. Swartz. Adverse drug event surveillance and drug withdrawals in the United States, 1969-2002: the importance of reporting suspected reactions. *Archives of internal medicine*, 165(12):1363–9, June 2005.

[54] P.-Y. Yen, D. Wantland, and S. Bakken. Development of a Customizable Health IT Usability Evaluation Scale. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, 2010:917–21, Jan. 2010.

[55] M. Yuksel. *A Semantic Interoperability Framework for Reinforcing Post Market Safety Studies.* PhD thesis, Middle East Technical University, Ankara, Turkey, May 2013.

[56] M. Yuksel, S. Gönül, G. B. L. Erturkmen, A. A. Sinaci, K. Depraetere, J. De Roo, and T. Bergvall. Demonstration of the SALUS Semantic Interoperability Framework for Case Series Characterization Studies. In *SWAT4LS*. Citeseer, 2013.

[57] X. Zhou, S. Murugesan, H. Bhullar, Q. Liu, B. Cai, C. Wentworth, and A. Bate. An evaluation of the THIN database in the OMOP common data model for active drug safety surveillance. *Drug Safety*, 36(2):119–134, 2013.

# APPENDIX A

# OMOP CONTENT ENTITY MODEL IN WEB ONTOLOGY LANGUAGE (OWL)


```
<?xml version="1.0"?>

<!DOCTYPE rdf:RDF [
    <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
    <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
    <!ENTITY xml "http://www.w3.org/XML/1998/namespace" >
    <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
    <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
<!ENTITY omop "http://www.salusproject.eu/ontology/omop-cdm#" >
]>

<rdf:RDF xmlns="http://www.salusproject.eu/ontology/omop-cdm#"
     xml:base="http://www.salusproject.eu/ontology/omop-cdm"
     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
     xmlns:owl="http://www.w3.org/2002/07/owl#"
     xmlns:xml="http://www.w3.org/XML/1998/namespace"
     xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
     xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
 xmlns:omop="http://www.salusproject.eu/ontology/omop-cdm#">
    <owl:Ontology rdf:about="http://www.salusproject.eu/ontology/omop-cdm">
        <owl:versionInfo rdf:datatype="&xsd;string">Created with TopBraid Composer</owl:versionInfo>
    </owl:Ontology>

    <rdfs:Datatype rdf:about="&xsd;date"/>

    <owl:ObjectProperty rdf:about="&omop;care_site">
        <rdfs:label rdf:datatype="&xsd;string">Care site</rdfs:label>
        <rdfs:range rdf:resource="&omop;CareSite"/>
        <rdfs:domain rdf:resource="&omop;Provider"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;cause_of_death_concept">
        <rdfs:label rdf:datatype="&xsd;string">Cause of death concept</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Death"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;cause_of_death_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Cause of death source value</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Death"/>
```

```
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;cohort">
        <rdfs:label rdf:datatype="&xsd;string">Cohort</rdfs:label>
        <rdfs:range rdf:resource="&omop;Cohort"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;cohort_concept">
        <rdfs:label rdf:datatype="&xsd;string">Ccohort concept</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Cohort"/>
        <rdfs:range rdf:resource="&omop;Concept"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;condition_concept">
        <rdfs:label rdf:datatype="&xsd;string">Condition concept</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;ConditionEra"/>
        <rdfs:domain rdf:resource="&omop;ConditionOccurence"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;condition_era">
        <rdfs:label rdf:datatype="&xsd;string">Condition era</rdfs:label>
        <rdfs:range rdf:resource="&omop;ConditionEra"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;condition_era_end_date">
        <rdfs:label rdf:datatype="&xsd;string">Condition era end date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;ConditionEra"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;condition_era_start_date">
        <rdfs:label rdf:datatype="&xsd;string">Condition era start date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;ConditionEra"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;condition_occurence">
        <rdfs:label rdf:datatype="&xsd;string">Condition occurence </rdfs:label>
        <rdfs:range rdf:resource="&omop;ConditionOccurence"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
        <rdfs:domain rdf:resource="&omop;VisitOccurence"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;condition_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Condition source value</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;ConditionOccurence"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;condition_type_concept">
        <rdfs:label rdf:datatype="&xsd;string">Condition type concept</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;ConditionEra"/>
        <rdfs:domain rdf:resource="&omop;ConditionOccurence"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;death">
        <rdfs:label rdf:datatype="&xsd;string">Death</rdfs:label>
        <rdfs:range rdf:resource="&omop;Death"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
```

```
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;death_type_concept">
        <rdfs:label rdf:datatype="&xsd;string">Death type concept</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Death"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;disease_class_concept">
        <rdfs:label rdf:datatype="&xsd;string">Disease class concept</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;ProcedureCost"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;disease_class_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Disease class source value</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;ProcedureCost"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;drug_concept">
        <rdfs:label rdf:datatype="&xsd;string">Drug concept</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;DrugEra"/>
        <rdfs:domain rdf:resource="&omop;DrugExposure"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;drug_cost">
        <rdfs:label rdf:datatype="&xsd;string">Drug cost</rdfs:label>
        <rdfs:range rdf:resource="&omop;DrugCost"/>
        <rdfs:domain rdf:resource="&omop;DrugExposure"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;drug_era">
        <rdfs:label rdf:datatype="&xsd;string">Drug era</rdfs:label>
        <rdfs:range rdf:resource="&omop;DrugEra"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;drug_exposure">
        <rdfs:label rdf:datatype="&xsd;string">Drug exposure</rdfs:label>
        <rdfs:range rdf:resource="&omop;DrugExposure"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
        <rdfs:domain rdf:resource="&omop;VisitOccurence"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;drug_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Drug source value</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;DrugExposure"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;drug_type_concept">
        <rdfs:label rdf:datatype="&xsd;string">Drug type concept</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;DrugEra"/>
        <rdfs:domain rdf:resource="&omop;DrugExposure"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;ethnicity_concept">
        <rdfs:label rdf:datatype="&xsd;string">Ethnicity concept</rdfs:label>
```

```
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;ethnicity_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Ethnicity source value</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;gender_concept">
        <rdfs:label rdf:datatype="&xsd;string">Gender concept</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;gender_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Gender source value </rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;location">
        <rdfs:label rdf:datatype="&xsd;string">Location</rdfs:label>
        <rdfs:domain rdf:resource="&omop;CareSite"/>
        <rdfs:range rdf:resource="&omop;Location"/>
        <rdfs:domain rdf:resource="&omop;Organization"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;observation">
        <rdfs:label rdf:datatype="&xsd;string">Observation</rdfs:label>
        <rdfs:range rdf:resource="&omop;Observation"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
        <rdfs:domain rdf:resource="&omop;VisitOccurence"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;observation_concept">
        <rdfs:label rdf:datatype="&xsd;string">Observation concept</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Observation"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;observation_period">
        <rdfs:label rdf:datatype="&xsd;string">Observation period</rdfs:label>
        <rdfs:range rdf:resource="&omop;ObservationPeriod"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;observation_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Observation source value</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Observation"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;observation_type_concept">
        <rdfs:label rdf:datatype="&xsd;string">Observation type concept</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Observation"/>
    </owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:about="&omop;organization">
        <rdfs:label rdf:datatype="&xsd;string">Organization</rdfs:label>
        <rdfs:domain rdf:resource="&omop;CareSite"/>
        <rdfs:range rdf:resource="&omop;Organization"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;payer_plan_period">
        <rdfs:label rdf:datatype="&xsd;string">Payer plan period</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugCost"/>
        <rdfs:range rdf:resource="&omop;PayerPlanPeriod"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;person_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Person source value</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;place_of_service_concept">
        <rdfs:label rdf:datatype="&xsd;string">Place of service concept</rdfs:label>
        <rdfs:domain rdf:resource="&omop;CareSite"/>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Organization"/>
        <rdfs:domain rdf:resource="&omop;VisitOccurence"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;place_of_service_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Place of service source value</rdfs:label>
        <rdfs:domain rdf:resource="&omop;CareSite"/>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Organization"/>
        <rdfs:domain rdf:resource="&omop;VisitOccurence"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;prescribing_provider">
        <rdfs:label rdf:datatype="&xsd;string">Prescribing provider</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugExposure"/>
        <rdfs:range rdf:resource="&omop;Provider"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;procedure_concept">
        <rdfs:label rdf:datatype="&xsd;string">Procedure concept</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;ProcedureOccurence"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;procedure_cost">
        <rdfs:label rdf:datatype="&xsd;string">Procedure cost</rdfs:label>
        <rdfs:range rdf:resource="&omop;ProcedureCost"/>
        <rdfs:domain rdf:resource="&omop;ProcedureOccurence"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;procedure_occurence">
        <rdfs:label rdf:datatype="&xsd;string">Procedure occurence</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Person"/>
        <rdfs:range rdf:resource="&omop;ProcedureOccurence"/>
        <rdfs:domain rdf:resource="&omop;VisitOccurence"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;procedure_source_value">
```

```
        <rdfs:label rdf:datatype="&xsd;string">Procedure source value</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;ProcedureOccurence"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;procedure_type_concept">
        <rdfs:label rdf:datatype="&xsd;string">Procedure type concept</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;ProcedureOccurence"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;provider">
        <rdfs:label rdf:datatype="&xsd;string">Provider</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Person"/>
        <rdfs:range rdf:resource="&omop;Provider"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;race_concept">
        <rdfs:label rdf:datatype="&xsd;string">Race concept </rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;race_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Race source value</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;relevant_condition_concept">
        <rdfs:label rdf:datatype="&xsd;string">Relevant condition concept</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;DrugExposure"/>
        <rdfs:domain rdf:resource="&omop;Observation"/>
        <rdfs:domain rdf:resource="&omop;ProcedureOccurence"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;revenue_code_concept">
        <rdfs:label rdf:datatype="&xsd;string">Revenue code concept</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;ProcedureCost"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;revenue_code_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Revenue code source value</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;ProcedureCost"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;speciality_concept">
        <rdfs:label rdf:datatype="&xsd;string">Speciality concept</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Provider"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;speciality_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Speciality source value</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Provider"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;unit_source_value">
```

```
        <rdfs:label rdf:datatype="&xsd;string">Unit source value</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Observation"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;value_as_concept">
        <rdfs:label rdf:datatype="&xsd;string">Value as concept</rdfs:label>
        <rdfs:range rdf:resource="&omop;Concept"/>
        <rdfs:domain rdf:resource="&omop;Observation"/>
    </owl:ObjectProperty>

<owl:ObjectProperty rdf:about="&omop;visit_occurence">
        <rdfs:label rdf:datatype="&xsd;string">Visit occurence</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Person"/>
        <rdfs:range rdf:resource="&omop;VisitOccurence"/>
    </owl:ObjectProperty>

    <owl:DatatypeProperty rdf:about="&omop;address_1">
        <rdfs:label rdf:datatype="&xsd;string">Address 1</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Location"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;address_2">
        <rdfs:label rdf:datatype="&xsd;string">Address 2</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Location"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;associated_provider_id">
        <rdfs:label rdf:datatype="&xsd;string">Associated provider</rdfs:label>
        <rdfs:domain rdf:resource="&omop;ConditionOccurence"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;average_wholesale_price">
        <rdfs:label rdf:datatype="&xsd;string">Average wholesale price</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugCost"/>
        <rdfs:range rdf:resource="&xsd;float"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;care_site_id">
        <rdfs:label rdf:datatype="&xsd;string">Care site id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;CareSite"/>
        <rdfs:domain rdf:resource="&omop;Person"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;care_site_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Care site source value</rdfs:label>
        <rdfs:domain rdf:resource="&omop;CareSite"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;city">
        <rdfs:label rdf:datatype="&xsd;string">City</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Location"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;code">
```

```
        <rdfs:label rdf:datatype="&xsd;string">Code</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Concept"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;code_system">
        <rdfs:label rdf:datatype="&xsd;string">Code system</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Concept"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;code_system_name">
        <rdfs:label rdf:datatype="&xsd;string">Code system name</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Concept"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;code_system_version">
        <rdfs:label rdf:datatype="&xsd;string">Code system version</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Concept"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;cohort_end_date">
        <rdfs:label rdf:datatype="&xsd;string">Cohort end date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Cohort"/>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;cohort_id">
        <rdfs:label rdf:datatype="&xsd;string">Cohort id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Cohort"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;cohort_start_date">
        <rdfs:label rdf:datatype="&xsd;string">Cohort start date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Cohort"/>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;concept_id">
        <rdfs:label rdf:datatype="&xsd;string">Concept id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Concept"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;condition_end_date">
        <rdfs:label rdf:datatype="&xsd;string">Condition end date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;ConditionOccurence"/>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;condition_era_end_date">
        <rdfs:label rdf:datatype="&xsd;string">Condition era end date</rdfs:label>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;condition_era_id">
        <rdfs:label rdf:datatype="&xsd;string">Condition era id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;ConditionEra"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;condition_era_start_date">
```

```
        <rdfs:label rdf:datatype="&xsd;string">Condition era start date</rdfs:label>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;condition_occurence_count">
        <rdfs:label rdf:datatype="&xsd;string">Condition occurence count</rdfs:label>
        <rdfs:domain rdf:resource="&omop;ConditionEra"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;condition_occurence_id">
        <rdfs:label rdf:datatype="&xsd;string">Condition occurence id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;ConditionOccurence"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;condition_start_date">
        <rdfs:label rdf:datatype="&xsd;string">Condition start date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;ConditionOccurence"/>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;county">
        <rdfs:label rdf:datatype="&xsd;string">County</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Location"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;day_of_birth">
        <rdfs:label rdf:datatype="&xsd;string">Day of birth</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Person"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;days_supply">
        <rdfs:label rdf:datatype="&xsd;string">Days supply</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugExposure"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;dea">
        <rdfs:label rdf:datatype="&xsd;string">DEA</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Provider"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;death_date">
        <rdfs:label rdf:datatype="&xsd;string">Death date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Death"/>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;dispensing_fee">
        <rdfs:label rdf:datatype="&xsd;string">Dispensing fee</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugCost"/>
        <rdfs:range rdf:resource="&xsd;float"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;display_name">
        <rdfs:label rdf:datatype="&xsd;string">Display name</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Concept"/>
```

```
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;drug_cost_id">
        <rdfs:label rdf:datatype="&xsd;string">Drug cost id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugCost"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;drug_era_end_date">
        <rdfs:label rdf:datatype="&xsd;string">Drug era end date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugEra"/>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;drug_era_id">
        <rdfs:label rdf:datatype="&xsd;string">Drug era id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugEra"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;drug_era_start_date">
        <rdfs:label rdf:datatype="&xsd;string">Drug era start date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugEra"/>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;drug_exposure_count">
        <rdfs:label rdf:datatype="&xsd;string">Drug exposure count</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugEra"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;drug_exposure_end_date">
        <rdfs:label rdf:datatype="&xsd;string">Drug exposure end date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugExposure"/>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;drug_exposure_id">
        <rdfs:label rdf:datatype="&xsd;string">Drug exposure id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugExposure"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;drug_exposure_start_date">
        <rdfs:label rdf:datatype="&xsd;string">Drug exposure start date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugExposure"/>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;family_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Family source value</rdfs:label>
        <rdfs:domain rdf:resource="&omop;PayerPlanPeriod"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;ingredient_cost">
        <rdfs:label rdf:datatype="&xsd;string">Ingredient cost</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugCost"/>
        <rdfs:range rdf:resource="&xsd;float"/>
    </owl:DatatypeProperty>
```

```
<owl:DatatypeProperty rdf:about="&omop;location_id">
        <rdfs:label rdf:datatype="&xsd;string">Location id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Location"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;location_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Location source value</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Location"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;month_of_birth">
        <rdfs:label rdf:datatype="&xsd;string">Month of birth</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Person"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;npi">
        <rdfs:label rdf:datatype="&xsd;string">NPI</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Provider"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;observation_date">
        <rdfs:label rdf:datatype="&xsd;string">Observation date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Observation"/>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;observation_id">
        <rdfs:label rdf:datatype="&xsd;string">Observation id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Observation"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;observation_period_end_date">
        <rdfs:label rdf:datatype="&xsd;string">Observation period end date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;ObservationPeriod"/>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;observation_period_id">
        <rdfs:label rdf:datatype="&xsd;string">Observation period id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;ObservationPeriod"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;observation_period_start_date">
        <rdfs:label rdf:datatype="&xsd;string">Observation period start date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;ObservationPeriod"/>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;observation_time">
        <rdfs:label rdf:datatype="&xsd;string">observation time</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Observation"/>
        <rdfs:range rdf:resource="&xsd;dateTime"/>
    </owl:DatatypeProperty>
```

```
<owl:DatatypeProperty rdf:about="&omop;organization_id">
        <rdfs:label rdf:datatype="&xsd;string">Organization id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Organization"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;organization_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Organization source value</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Organization"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;paid_by_coordination_benefits">
        <rdfs:label rdf:datatype="&xsd;string">Paid by coordination benefits</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugCost"/>
        <rdfs:domain rdf:resource="&omop;ProcedureCost"/>
        <rdfs:range rdf:resource="&xsd;float"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;paid_by_payer">
        <rdfs:label rdf:datatype="&xsd;string">Paid by payer</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugCost"/>
        <rdfs:domain rdf:resource="&omop;ProcedureCost"/>
        <rdfs:range rdf:resource="&xsd;float"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;paid_coinsurance">
        <rdfs:label rdf:datatype="&xsd;string">paid coinsurance</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugCost"/>
        <rdfs:domain rdf:resource="&omop;ProcedureCost"/>
        <rdfs:range rdf:resource="&xsd;float"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;paid_copay">
        <rdfs:label rdf:datatype="&xsd;string">Paid copay</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugCost"/>
        <rdfs:domain rdf:resource="&omop;ProcedureCost"/>
        <rdfs:range rdf:resource="&xsd;float"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;paid_toward_deductible">
        <rdfs:label rdf:datatype="&xsd;string">Paid toward deductible</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugCost"/>
        <rdfs:domain rdf:resource="&omop;ProcedureCost"/>
        <rdfs:range rdf:resource="&xsd;float"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;payer_plan_period_end_date">
        <rdfs:label rdf:datatype="&xsd;string">Payer plan period end date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;PayerPlanPeriod"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;payer_plan_period_id">
        <rdfs:label rdf:datatype="&xsd;string">Payer plan period id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;PayerPlanPeriod"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;payer_plan_period_start_date">
        <rdfs:label rdf:datatype="&xsd;string">Payer plan period start date</rdfs:label>
```

```
        <rdfs:domain rdf:resource="&omop;PayerPlanPeriod"/>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;payer_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Payer source value</rdfs:label>
        <rdfs:domain rdf:resource="&omop;PayerPlanPeriod"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;person_id">
        <rdfs:label rdf:datatype="&xsd;string">person id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Person"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;plan_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Plan source value</rdfs:label>
        <rdfs:domain rdf:resource="&omop;PayerPlanPeriod"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;procedure_cost_id">
        <rdfs:label rdf:datatype="&xsd;string">Procedure cost id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;ProcedureCost"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;procedure_date">
        <rdfs:label rdf:datatype="&xsd;string">Procedure date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;ProcedureOccurence"/>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;procedure_occurence_id">
        <rdfs:label rdf:datatype="&xsd;string">Procedure occurence id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;ProcedureOccurence"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;provider_id">
        <rdfs:label rdf:datatype="&xsd;string">Provider id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Provider"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;provider_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Provider source value</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Provider"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;quantity">
        <rdfs:label rdf:datatype="&xsd;string">Quantity</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugExposure"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;range_high">
        <rdfs:label rdf:datatype="&xsd;string">Range high</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Observation"/>
```

```
        <rdfs:range rdf:resource="&xsd;float"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;range_row">
        <rdfs:label rdf:datatype="&xsd;string">Range low</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Observation"/>
        <rdfs:range rdf:resource="&xsd;float"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;refills">
        <rdfs:label rdf:datatype="&xsd;string">Refills</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugExposure"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;sig">
        <rdfs:label rdf:datatype="&xsd;string">Sig</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugExposure"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;speciality_source_value">
        <rdfs:label rdf:datatype="&xsd;string">Speciality source value</rdfs:label>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;state">
        <rdfs:label rdf:datatype="&xsd;string">State</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Location"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;stop_reason">
        <rdfs:label rdf:datatype="&xsd;string">Stop reason</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Cohort"/>
        <rdfs:domain rdf:resource="&omop;ConditionOccurence"/>
        <rdfs:domain rdf:resource="&omop;DrugExposure"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;total_out_of_pocket">
        <rdfs:label rdf:datatype="&xsd;string">Total out of pocket</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugCost"/>
        <rdfs:domain rdf:resource="&omop;ProcedureCost"/>
        <rdfs:range rdf:resource="&xsd;float"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;total_paid">
        <rdfs:label rdf:datatype="&xsd;string">Total paid</rdfs:label>
        <rdfs:domain rdf:resource="&omop;DrugCost"/>
        <rdfs:domain rdf:resource="&omop;ProcedureCost"/>
        <rdfs:range rdf:resource="&xsd;float"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;unit_concept_id">
        <rdfs:label rdf:datatype="&xsd;string">Unit concept id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Observation"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;value_as_number">
```

```
        <rdfs:label rdf:datatype="&xsd;string">Value as number</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Observation"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;value_as_string">
        <rdfs:label rdf:datatype="&xsd;string">Value as string</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Observation"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;visit_end_date">
        <rdfs:label rdf:datatype="&xsd;string">Visit end date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;VisitOccurence"/>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;visit_occurence_id">
        <rdfs:label rdf:datatype="&xsd;string">Visit occurence id</rdfs:label>
        <rdfs:domain rdf:resource="&omop;VisitOccurence"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;visit_start_date">
        <rdfs:label rdf:datatype="&xsd;string">Visit start date</rdfs:label>
        <rdfs:domain rdf:resource="&omop;VisitOccurence"/>
        <rdfs:range rdf:resource="&xsd;date"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;year_of_birth">
        <rdfs:label rdf:datatype="&xsd;string">Year of birth</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Person"/>
        <rdfs:range rdf:resource="&xsd;integer"/>
    </owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="&omop;zip">
        <rdfs:label rdf:datatype="&xsd;string">Zip</rdfs:label>
        <rdfs:domain rdf:resource="&omop;Location"/>
        <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>

    <owl:Class rdf:about="&omop;CareSite">
        <rdfs:label rdf:datatype="&xsd;string">Care Site</rdfs:label>
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>

<owl:Class rdf:about="&omop;Cohort">
        <rdfs:label rdf:datatype="&xsd;string">Cohort</rdfs:label>
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>

<owl:Class rdf:about="&omop;Concept">
        <rdfs:label rdf:datatype="&xsd;string">Concept</rdfs:label>
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>

<owl:Class rdf:about="&omop;ConditionEra">
        <rdfs:label rdf:datatype="&xsd;string">Condition Era</rdfs:label>
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>
```

```
<owl:Class rdf:about="&omop;ConditionOccurence">
        <rdfs:label rdf:datatype="&xsd;string">Condition Occurence</rdfs:label>
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>

<owl:Class rdf:about="&omop;Death">
        <rdfs:label rdf:datatype="&xsd;string">Death</rdfs:label>
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>

<owl:Class rdf:about="&omop;DrugCost">
        <rdfs:label rdf:datatype="&xsd;string">Drug Cost</rdfs:label>
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>

<owl:Class rdf:about="&omop;DrugEra">
        <rdfs:label rdf:datatype="&xsd;string">Drug Era</rdfs:label>
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>

<owl:Class rdf:about="&omop;DrugExposure">
        <rdfs:label rdf:datatype="&xsd;string">Drug Exposure</rdfs:label>
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>

<owl:Class rdf:about="&omop;Location">
        <rdfs:label rdf:datatype="&xsd;string">Location</rdfs:label>
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>

<owl:Class rdf:about="&omop;Observation">
        <rdfs:label rdf:datatype="&xsd;string">Observation</rdfs:label>
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>

<owl:Class rdf:about="&omop;ObservationPeriod">
        <rdfs:label rdf:datatype="&xsd;string">Observation Period</rdfs:label>
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>

<owl:Class rdf:about="&omop;Organization">
        <rdfs:label rdf:datatype="&xsd;string">Organization</rdfs:label>
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>

<owl:Class rdf:about="&omop;PayerPlanPeriod">
        <rdfs:label rdf:datatype="&xsd;string">Payer plan period</rdfs:label>
        <rdfs:subClassOf rdf:resource="&omop;Organization"/>
    </owl:Class>

<owl:Class rdf:about="&omop;Person">
        <rdfs:label rdf:datatype="&xsd;string">Person</rdfs:label>
    </owl:Class>

<owl:Class rdf:about="&omop;ProcedureCost">
        <rdfs:label rdf:datatype="&xsd;string">Procedure cost</rdfs:label>
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>

<owl:Class rdf:about="&omop;ProcedureOccurence">
        <rdfs:label rdf:datatype="&xsd;string">Procedure occurence</rdfs:label>
```

```
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>

<owl:Class rdf:about="&omop;Provider">
        <rdfs:label rdf:datatype="&xsd;string">Provider</rdfs:label>
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>

<owl:Class rdf:about="&omop;VisitOccurence">
        <rdfs:label rdf:datatype="&xsd;string">Visit occurence</rdfs:label>
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>
</rdf:RDF>
```

# APPENDIX B

# N3 RULE FOR OBTAINING OMOP PERSON FROM SALUS CIM PATIENT

```
{?PATIENT :mapPatientGraph{
    ?PATIENT :mapPatient ?PERSON.
    ?PERSON a omop:Person;
        omop:person_id ?pid;
        omop:location ?oa;
        omop:year_of_birth ?oyear;
        omop:month_of_birth ?omonth;
        omop:day_of_birth ?oday;
        omop:provider ?op;
        omop:care_site_id ?oosv;
        omop:person_source_value ?opsv;
        omop:gender_source_value ?og;
        omop:race_source_value ?or;
        omop:ethnicity_source_value ?oe.
}}
<=
{
    ?PATIENT a salus:Patient;
        salus:ID ?sid.
    ?sid :mapII ?pid.

    ?SCOPE e:optional {?PATIENT salus:gender ?sg. ?sg :mapCode ?og.},
                      {?og e:tuple (salus:gender ?PATIENT)}.
    ?SCOPE e:optional {
        ##month and date values will be parsed in the Java code
        ?PATIENT salus:dateOfBirth ?oyear.
        ("" 1) func:substring ?omonth.
        ("" 1) func:substring ?oday.
    },{
        ?oyear e:tuple (salus:dateOfBirth ?PATIENT).
        ?omonth e:tuple (salus:dateOfBirth ?PATIENT).
        ?oday e:tuple (salus:dateOfBirth ?PATIENT).
    }.
    ?SCOPE e:optional {?PATIENT salus:race ?sr. ?sr :mapCode ?or.},
                      {?or e:tuple (salus:race ?PATIENT)}.
    ?SCOPE e:optional {?PATIENT salus:ethnicity ?se. ?se :mapCode ?oe.},
                      {?oe e:tuple (salus:ethnicity ?PATIENT)}.
    ?SCOPE e:optional {?PATIENT salus:address ?sa. ?sa :mapAddress ?oa.},
                      {?oa e:tuple (salus:address ?PATIENT)}.
    ?SCOPE e:optional {
                       ?PATIENT salus:healthCareProvider ?shcp.
```

```
                     ?shcp :mapHealthCareProvider ?op.},
                    {?op e:tuple (salus:healthCareProvider ?PATIENT)}.
    ?SCOPE e:optional {
        ?PATIENT salus:providerOrganization ?spo.
        ?spo salus:organizationID ?soi.
        ?soi :mapII ?oosv.
    }, {
        ?oosv e:tuple (salus:providerOrganization ?PATIENT).
    }.
    ?SCOPE e:optional {?PATIENT salus:ID ?si. ?si :mapII ?opsv.},
                    {?opsv e:tuple (salus:ID ?PATIENT)}.
    ?PERSON e:tuple (?PATIENT).
}.
```

# APPENDIX C

# N3 RULE FOR OBTAINING OMOP DRUG EXPOSURE FROM SALUS CIM MEDICATION

```
{?MEDICATION :mapMedicationGraph {
    ?MEDICATION :mapMedication ?DRUGEXPOSURE.
    ?DRUGEXPOSURE a omop:DrugExposure;
        omop:person ?op;
        omop:drug_exposure_start_date ?odesd;
        omop:drug_exposure_end_date ?odeed;
        omop:refills ?or;
        omop:quantity ?oq;
        omop:sig ?os;
        omop:prescribing_provider_id ?oppi;
        omop:drug_source_value ?ocsv;
        omop:relevant_condition_concept ?orcc.
}}
<=
{
    ?MEDICATION a salus:Medication.
    ?SCOPE e:optional {
        ?sp salus:medication ?MEDICATION.
        ?sp :mapPatient ?op.
    },{
        ?op e:tuple (salus:medication ?MEDICATION)
    }.
    #start and end dates
    ?SCOPE e:optional {
        ?MEDICATION salus:indicateMedicationStartStop ?smiss.
        ?smiss :mapIvlTs (?odesd ?odeed).
    },{
        ?odesd e:tuple (salus:indicateMedicationStartStop ?MEDICATION).
        ?odeed e:tuple (salus:indicateMedicationStartStop ?MEDICATION).
    }.
    #refills
    ?SCOPE e:optional {
        ?MEDICATION salus:fulfillmentHistory ?smfh.
        ?smfh salus:fillNumber ?or.
    },{
        ?or e:tuple (salus:fulfillmentHistory ?MEDICATION)
    }.
    #quantity
    ?SCOPE e:optional {
        ?MEDICATION salus:fulfillmentHistory ?smfh.
        ?smfh salus:quantityDispensed ?smfh_qd.
```

```
        ?smfh_qd salus:value ?oq.
    },{
        ?MEDICATION salus:orderInformation ?smoi.
        ?smoi salus:quantityOrdered ?smoi_qo.
        ?smoi_qo salus:value ?oq.
    },{
        ?oq e:tuple (salus:fulfillmentHistory ?MEDICATION)
    }.
    #sig
    ?SCOPE e:optional {
        ?MEDICATION salus:fulfillmentInstructions ?os.
    },{
        ?os e:tuple (salus:fulfillmentInstructions ?MEDICATION)
    }.
    #prescribing provider id
    ?SCOPE e:optional {
        ?MEDICATION salus:orderInformation ?smoi.
        ?smoi salus:orderingProvider ?smoi_op.
        ?smoi_op salus:providerID ?smoi_op_pid.
        ?smoi_op_pid :mapII ?oppi.
    },{
        ?oppi e:tuple (salus:orderInformation ?MEDICATION).
    }.
    #drug source value
    ?SCOPE e:optional {
        ?MEDICATION salus:medicationInformation ?smi.
        ?smi salus:codedProductName ?smi_cpn.
        ?smi_cpn :mapCode ?ocsv.
    },{
        ?ocsv e:tuple (salus:medicationInformation ?MEDICATION).
    }.
    #relevant condition concept
    ?SCOPE e:optional {
        ?MEDICATION salus:indication ?si.
        ?si salus:problemCode ?si_pc.
        ?si_pc :mapCode ?orcc.
    },{
        ?orcc e:tuple (salus:indication ?MEDICATION).
    }.
    ?DRUGEXPOSURE e:tuple (?MEDICATION).
}.
```

# APPENDIX D

# N3 RULE FOR OBTAINING OMOP CONDITION EXPOSURE FROM SALUS CIM ALLERGY

```
#Allergy2ConditionOccurence
{?ALLERGY :mapAllergyGraph {
    ?ALLERGY :mapAllergy ?CONDITIONOCCURENCE.
    ?CONDITIONOCCURENCE a omop:ConditionOccurence;
        omop:person ?op;
        omop:condition_start_date ?ocsd;
        omop:condition_end_date ?oced;
        omop:stop_reason ?osr;
        omop:condition_source_value ?ocsv.
}}
<=
{
    ?ALLERGY a salus:Allergy.
    ?SCOPE e:optional {
        ?sp salus:allergy ?ALLERGY.
        ?sp :mapPatient ?op.
    },{
        ?op e:tuple (salus:allergy ?ALLERGY)
    }.
    #start and end dates
    ?SCOPE e:optional {
        ?ALLERGY salus:adverseEventDate ?saedt.
        ?saedt :mapIvlTs (?ocsd ?oced).
    },{
        ?ocsd e:tuple (salus:encounterDateTime ?ALLERGY).
        ?oced e:tuple (salus:encounterDateTime ?ALLERGY).
    }.
    #stop reason
    ?SCOPE e:optional {
        ?ALLERGY salus:status ?sas.
        ?sas salus:displayName ?osr.
    },{
        ?osr e:tuple (salus:status ?ALLERGY).
    }.
    #condition source value
    ?SCOPE e:optional {?ALLERGY salus:adverseEventType ?saet. ?saet :mapCode ?ocsv.},
                  {?ocsv e:tuple (salus:status ?ALLERGY).}.
    ?CONDITIONOCCURENCE e:tuple (?ALLERGY).
}.
```

# APPENDIX E

# RELATED PUBLICATIONS

1. Anil Pacaci, Suat Gonul, A. Anil Sinaci, Gokce B. Laleci Erturkmen, Hong Sun, Jos De Roo, "A Semantic Transformation Approach for the Secondary Use of Observational Healthcare Data in Postmarketing Safety Studies", *Journal of American Medical Informatics Association*, submitted for publication.

2. A. Anil Sinaci, Gokce B. Laleci Erturkmen, Suat Gonul, Mustafa Yuksel, Paolo Invernizzi, Bharat Thakrar, Anil Pacaci, H. Alper Cinar and Nihan Kesim Cicekli, "Post Marketing Safety Study Tool: A web based, dynamic and interoperable system for postmarketing drug surveillance studies", Spec. issue of *BioMed Research International*, Forthcoming 2015.