Name:	Peter R. Rijnbeek, PhD
Affiliation:	Erasmus MC Rotterdam, The Netherlands
Email:	p.rijnbeek@erasmusmc.nl
Presentation type (select one):	Poster

Implementation of the OMOP CDM and OHDSI tools in the European Medical Information Framework (EMIF)

Peter R. Rijnbeek, PhD[1], Michel van Speybroeck[2], Lara Tramontan, PhD[3,4], Leonardo Méndez-Boo, MD MPH[5,6], Rients van Wijngaarden[7], Rosa Gini, PhD[8], Miguel A. Mayer PhD[9], Lars Pedersen, PhD[10], Alessandro Pasqua Msc[11], José L. Oliveira[12], PhD, Johan van der Lei, PhD[1]

[1]Erasmus MC, Rotterdam, The Netherlands; [2]Janssen Pharmaceutica NV, Beerse, Belgium [3]Arsenàl.IT, TV, Italy; [4]SoSePe, PD, Italy; [5]Direcció de Sistemes d'Informació, Institut Català de la Salut, Spain;[6] Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Barcelona, Spain; [7]STIZON, Utrecht, The Netherlands; [8]Agenzia regionale di sanità della Toscana, Florence, Italy; [9]Research Programme on Biomedical Informatics (IMIM-UPF), Spain; [10]Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus N, Denmark; [11]Genomedics Srl, Florence, Italy; [12]University of Aveiro, Aveiro, DETI/IEETA, Portugal

Abstract

The European Medical Information Framework (EMIF) project's main objective is to develop a sustainable platform that enables data discovery and unprecedented use of data sources of a wide variety of types, including electronic health records (EHR) in both primary and hospital care. EMIF adopted the OMOP Common Data Model (CDM) and is currently actively mapping nine EHR databases. Furthermore, EMIF aims to incorporate the OHDSI tools to support data discovery and querying. This work describes the current status and initial findings. EMIF aims to strengthen the collaboration with the OHDSI community to achieve our common goals.

Introduction

Given Europe's diversity and emphasis on local governance, retaining control is of paramount importance to the data custodians if long term sustainability and maximum impact is to be achieved. The EMIF-Platform is therefore based on a federation of the EHR data sources, rather than a centralized database containing all data. Each data source retains control over its data and its delivery to (and, potentially, withdrawal from) the federation. The EHR data sources differ considerably with respect to the underlying population and/or the type of data collected. Each data source has its own underlying data model which makes the use of common tools more difficult and the extraction process less transparent. Differences in clinical terminology systems (ICD9, READ, etc.) and differences in extraction algorithms further complicate the interpretation of the results.

Mapping to the OMOP CDM

To improve both the semantic as well as the syntactic operability, EMIF is currently actively mapping nine European data sources to the OMOP CDM and OHDSI vocabularies: Denmark (AUH[1]), Italy (ARS[2], IMS HEALTH LPD[3], PEDIANET[4]), Spain (IMASIS[5], SIDIAP[6]), UK (THIN[7]), The Netherlands (IPCI[8], PHARMO[9]). For each data source a multi-disciplinary team is formed consisting of local domain experts, data managers, data administrators, and OMOP CDM specialist, that follow a workflow as shown in Figure 1. The workflow starts with a profiling step where the White Rabbit tool is used to gain insight in the data structure and contents. Additionally, the local domain experts create a data dictionary that gives background information on each table and a translation of the fields to the English language. A useful new feature for White Rabbit would be an import facility of the data dictionary so the tool can display the fields and table names in English.

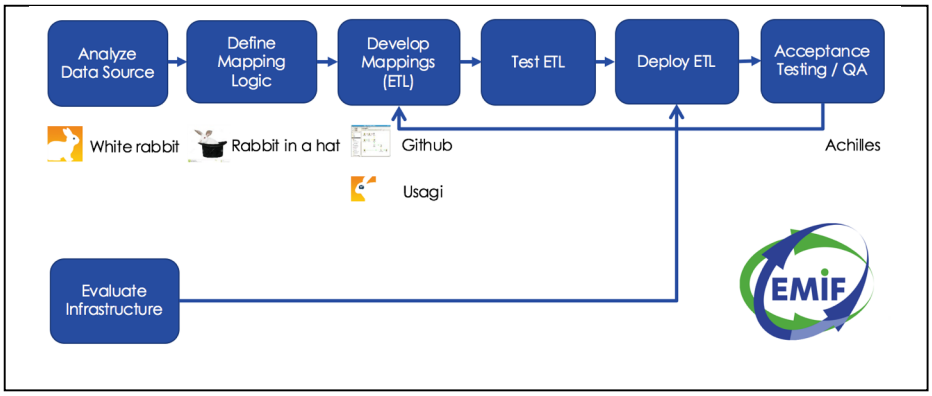

Figure 1. Workflow to map EMIF EHR data sources to the OMOP CDM, leveraging OHDSI tools and experience.

The mapping logic is defined in a workshop at the data source using the Rabbit in a Hat tool. Challenges were encountered in the following areas:

- Mapping of local terminology systems to standard vocabularies: local terminology systems are often at a different level of granularity (especially for drugs and measurements) compared to the CDM standard vocabularies. This task can be resource-consuming.
- Heterogeneity of the included populations: criteria for inclusion and exclusion in the databases differ which makes the interpretation of the observation period in the CDM more difficult.
- Mapping of drugs: some European drugs are not included in the RxNorm vocabulary.
- Mapping of vaccinations: vaccinations are often separately captured from regular drugs with their own coding system.
- Use of concept types: concept types are sometimes missing in particular for visits and procedures.
- Capturing geographic information: current location entity is based on US geographies, a more abstract geographical representation would accommodate a broader usage.

After the development and deployment of the Extract Transform Load (ETL) procedure in the local infrastructure, acceptance testing will be performed. This will be done through ACHILLES HEEL and study replication.

Utilization of OHDSI Tools

The EMIF architecture contains layers to support data discovery through community based catalogues, dashboards functionality, database querying tools, and tools for central analysis in a remote research environment. Currently, the use of the OHDSI tools is being evaluated in this full pipeline. For example, ACHILLES is implemented as part of the EMIF-Catalogue (see Figure 2) which gives a detailed overview of the available databases. EMIF currently plans to evaluate ACHILLES to test the functionality in European context and help develop new features. An important requirement is the ability to view the data in the local terminology and to have an integrated security layer to allow access control by the DC. Also ATLAS, to facility database querying in the distributed network, is being evaluated.

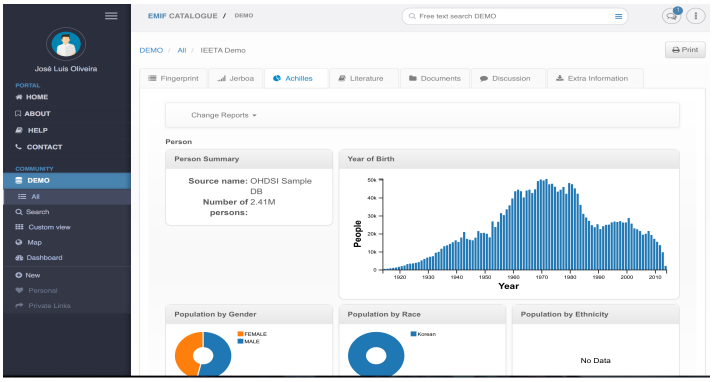

Figure 2. Achilles integrated in the EMIF-Catalogue.

Conclusion

The OMOP CDM and the OHDSI tools are a natural fit to the main objectives of EMIF, i.e. to enable data discovery and re-use. EMIF is actively mapping nine European data sources to the OMOP CDM and will be evaluating the mapping process and the OHDSI tools in the upcoming period. EMIF aims to strengthen the collaboration with the OHDSI community to achieve our common goals.

References

1. Ehrenstein V, Antonsen S, Pedersen L. Existing data sources for clinical epidemiology: Aarhus University Prescription Database. Clin Epidemiol. 2010;2:273-9.
2. Gini R, Francesconi P, Mazzaglia G, Cricelli I, Pasqua A, Gallina P, et al. Chronic disease prevalence from Italian administrative databases in the VALORE project: a validation through comparison of population estimates with general practice databases and national survey. BMC Public Health. 9 gennaio 2013;13(1):15.
3. Cricelli C, Mazzaglia G, Samani F, Marchi M, Sabatini A, Nardi R, Ventriglia G, Caputi AP. Prevalence estimates for chronic diseases in Italy: exploring the differences between self-report and primary care databases. J Public Health Med 2003; 25:254-7

4. Avillach P, Coloma PM, Gini R, Schuemie M, Mougin F, et al. EU-ADR consortium. Harmonization process for the identification of medical events in eight European healthcare databases: the experience from the EU-ADR project. J Am Med Inform Assoc. 2013 Jan 1;20(1):184-92. doi: 10.1136/amiajnl-2012-000933. Epub 2012 Sep 6.
5. Mayer MA, Furlong LI, Torre P, Planas I, Cots F, Iquierdo E et al. Reuse of EHRs to support clinical research in a hospital of reference. Stud Health Technol Inform 2015; 210:224-6.
6. Bolíbar B, Fina F, Morros R, Garcia MD, Hermosilla E, Ramos R, Rosell M, Rodríguez J, Medina M, Calero S, Prieto D. SIDIAP database: electronic clinical records in Primary Care as a source of information for epidemiologic research. Med Clin (Barc). 2012 May 19;138(14):617-21.
7. Blak BT, Thompson M. Dattani H and Bourke A. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. Informatics in Primary Care 2012,19(4)
8. Vlug AE, van der Lei J, Mosseveld BM, van Wijk MA, van der Linden PD, Sturkenboom MC, van Bemmel JH (1999) Postmarketing surveillance based on electronic patient records:the IPCI project. Methods Inf Med 38(4–5):339–344.
9. Herings R, Pedersen L. Pharmacy-based Medical Record Linkage Systems. In: Strom B, Kimmel S, editors. Pharmacoepidemiology 5ed: John Wiley & Sons, Ltd.;2012. P 270-86