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, MSc[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), Barcelona, 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

Background

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. EMIF aims to incorporate the OHDSI tools to support data discovery and querying. This work describes the current status and initial findings.

Database	Country / Region	Population Size	Type
Agenzia regionale di sanità della Toscana (ARS)	Italy / Tuscany	$5 \cdot 10^6$	Administrative database of Tuscan population
Aarhus University Hospital Database	Denmark / Northern region	$2.3 \cdot 10^6$	Administrative database for population Central and North Jutland
Health Search - IMS HEALTH LPD	Italy	$1.6 \cdot 10^6$	Primary care data of GP's using the Health Search System
Integrated Primary Care Information (IPCI)	Netherlands	$2.8 \cdot 10^6$	Primary care database
Pedianet	Italy	$400 \cdot 10^3$	Primary care pediatric database
Pharmo	Netherlands	$8.4 \cdot 10^6$	Primary care database
Information System of Parc de Salut Mar (IMASIS)	Spain / Barcelona	$1.4 \cdot 10^6$	Hospital database
The Information System for the Development of Research in Primary Care (SIDIAP)	Spain / Catalonia	$6.4 \cdot 10^6$	Primary care database
The Health Informatics Network (THIN)	United Kingdom	$12 \cdot 10^6$	Primary care database

Methods

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 outlined below.
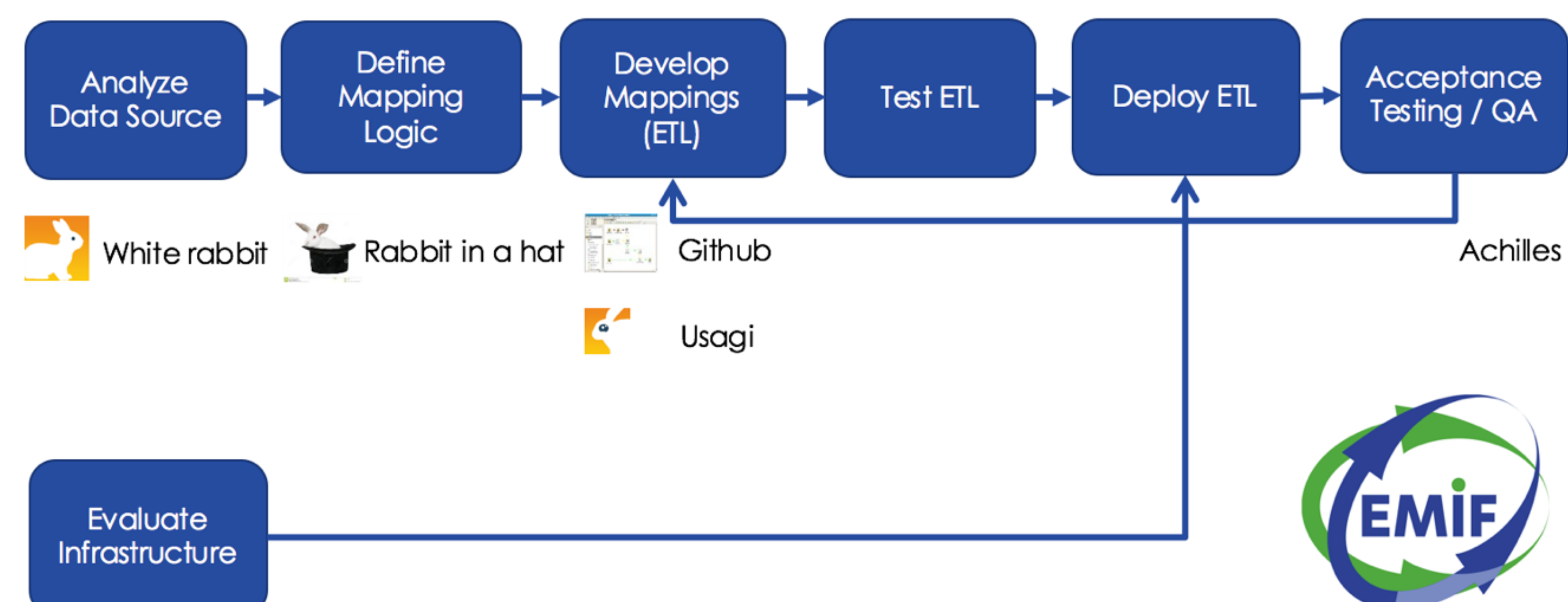
Mapping logic is defined in a workshop and ETL is subsequently developed using different OHDSI tools and programming languages. Achilles heel is used to perform a data quality assessment.

Results

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. An initiative is started to add missing European drugs.
- **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.

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. For example, ACHILLES is implemented as part of the EMIF-Catalogue which gives a detailed overview of the available databases.

EMIF Catalogue
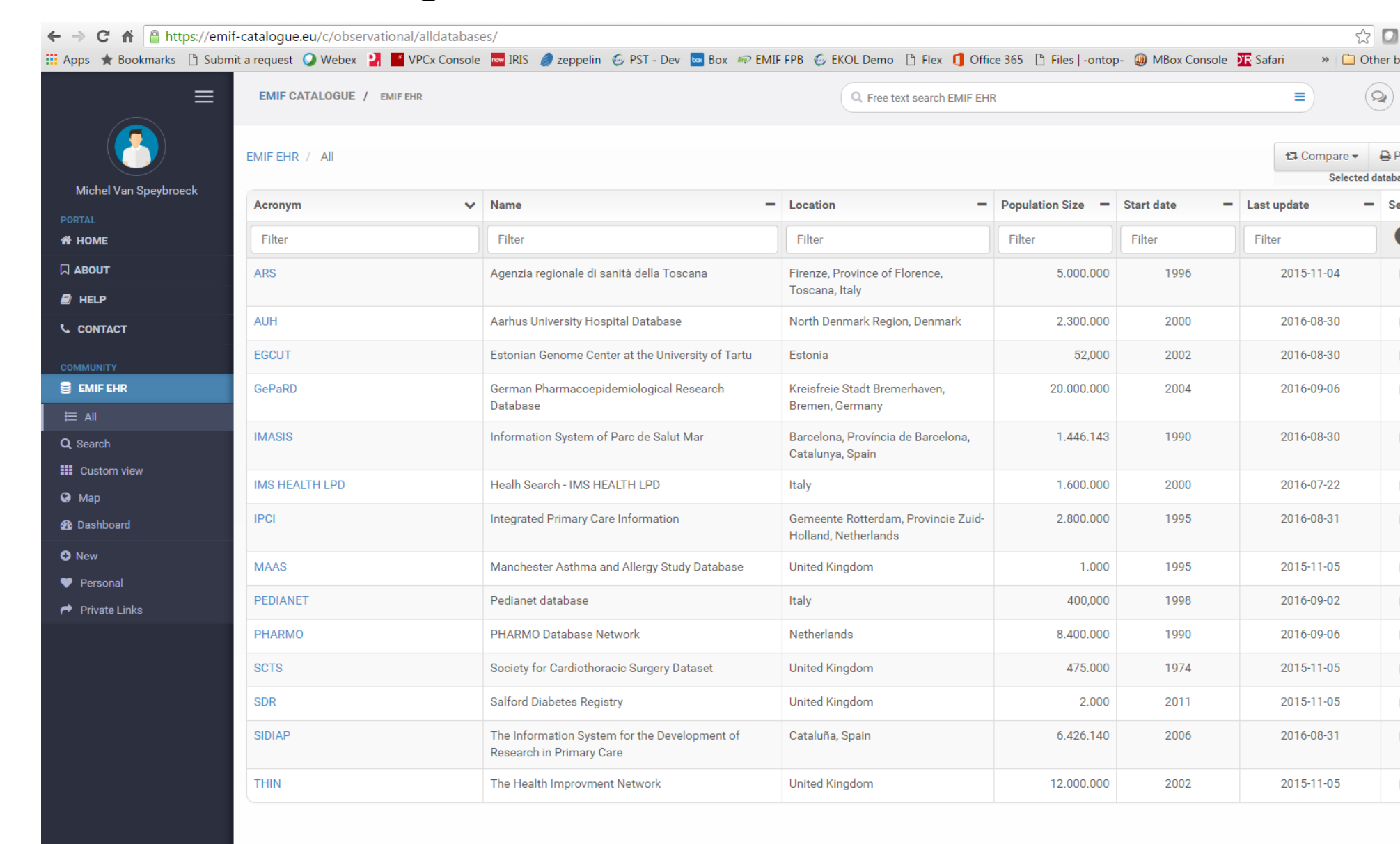
Achilles integration
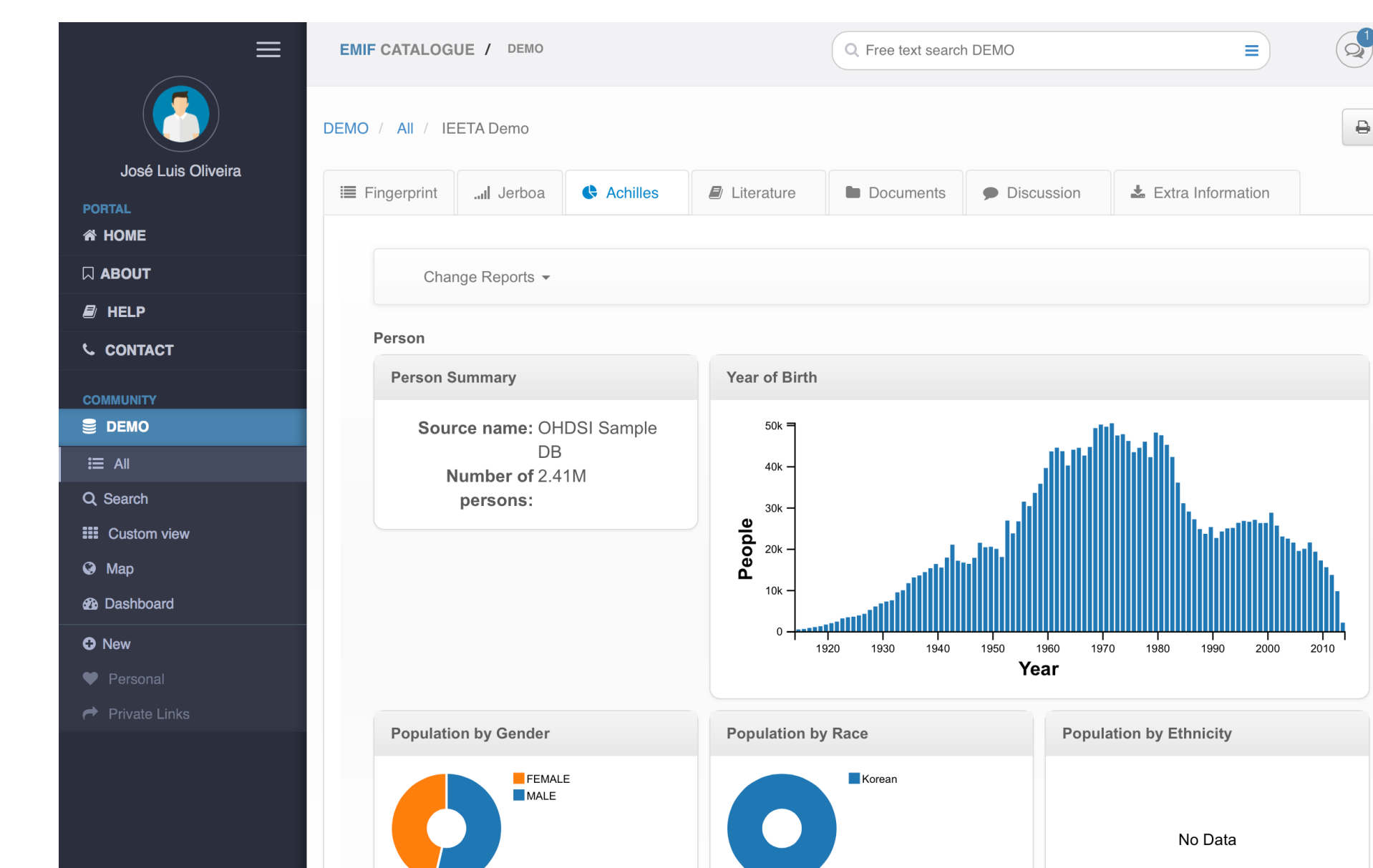
Conclusions

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 will collaborate with OHDSI on data standards, vocabulary mappings and further development of tools.