

Background

- The Information System for Research in Primary Care (SIDIAP; www.sidiap.org) contains data of anonymized patients' healthcare records for nearly **six million** people (approximately 80% of the Catalan population) registered in 287 primary care practices throughout Catalonia (Figure 1) since **2005**.
- It includes **data collected by health professionals** during routine visits in **primary care**, such as clinical diagnoses and measurements (weight, blood pressure, etc.), laboratory tests, treatments, hospital referrals, demographic and lifestyle information. SIDIAP can be linked to other sources of data such as hospital discharge, population registries, etc.
- **Objective:** To convert the SIDIAP data base to the Observational Medical Outcomes Partnership - Common Data Model (**OMOP-CDM**) as part of **The European Medical Information Framework (EMIF)**.

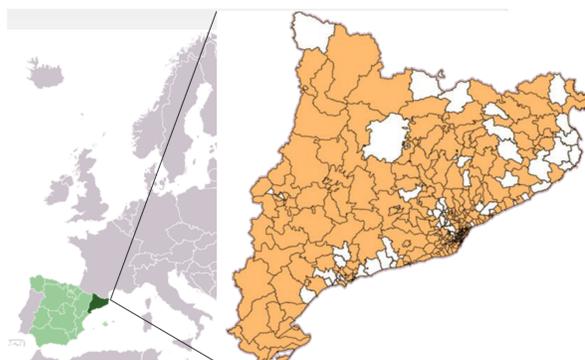
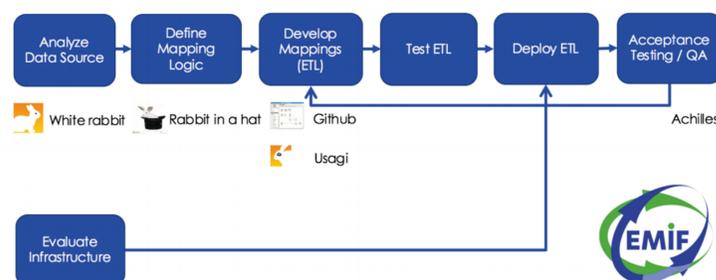


Figure 1. SIDIAP population coverage in Catalonia, Spain.

Methods

- SIDIAP has used the workflow proposed by EMIF for the conversion to the OMOP-CDM (Figure 2).
- First, the **White Rabbit** tool was used to profile the data structure and contents.
- Second, the mapping logic was defined using the **Rabbit in a Hat** tool, followed by the automatic mapping of codes.
- A manual validation of the automatic mapping was performed by two clinicians using the **Usagi** tool.
- The mapping of concepts with the highest number of registers was prioritized.
- Finally, the development and deployment of the **Extract Transform Load (ETL)** was performed for the clinical domains shown in the results (mapped terms and mapped number of registers).

Figure 2. Workflow to map EMIF data sources to the OMOP CDM, leveraging OHDSI tools and experience. EMIF



Results

- SIDIAP data source has 14 tables that were mapped to 11 tables of the OMOP-CDM (Figure 3).
- More than 24.700 terms and 1.400.000.000 registers **were mapped** for the following clinical domains: procedure occurrence, drug exposure, condition occurrence and measurement.
- More than >90% of registers were mapped for all domains, except for drug exposure for which 84,3% of registers were mapped.
- Regarding the domain of drugs, 75% of source registers could be mapped to the class Clinical Drug, which is important to retain information on drug substance, drug strength and form.
- For most domains, only a small % of terms (i.e. 25,3% of tests) needed mapping in order to obtain a high % of registers mapped (i.e. 95% for tests).

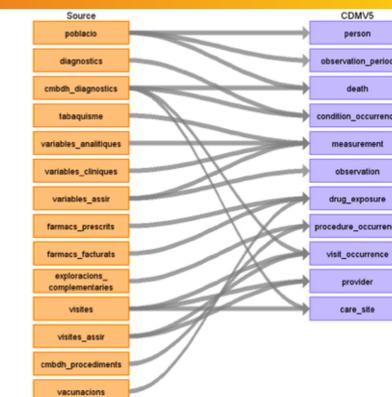


Figure 3. Data source analysis using the White Rabbit tool
* Source names are in Catalan.

Table 1. Information mapped in SIDIAP: number of terms and registers by source table

Source Table	OMOP	Source terms	Mapped Terms	Mapped Terms(%)	Source Registers	Mapped Registers	Mapped Registers (%)
Tests		510	129	25,3%	10.629.950	10.096.950	95,0%
Hospital Procedures	PROCEDURE_OCCURRENCE	3.689	560	15,2%	2.963.300	2.699.566	91,0%
Prescribed Drugs							
Billed Drugs	DRUG_EXPOSURE	27.389	4.345	15,9%	942.464.823	794.619.660	84,3%
Vaccinations							
Primary care Diagnoses Hospital diagnoses	CONDITION_OCCURRENCE	20.671	19.572	94,7%	94.699.661	93.938.584	99,2%
Laboratory Measurements	MEASUREMENT	99	99	100%	537.387.710	537.360.195	100%

Conclusions

- The majority of clinical domains in SIDIAP have been successfully converted to the OMOP-CDM, with over 90% of registered terms mapped for most domains. The prioritization of mapping of the most registered concepts was an useful strategy to effectively convert SIDIAP to the OMOP-CDM. The combination of automatic and manual mapping techniques by clinicians, as well as the participation of a multidisciplinary team were required for a successful conversion to the OMOP-CDM.
- **To be done:** 1) to finalize mapping of remaining domains; and 2) to perform the acceptance testing using the Achilles heel and study replication.
- **Challenges encountered** in the mapping process included:
 - The Spanish authorized drugs may not be fully represented in the RxNorm vocabulary or RxNorm current extension.
 - Decisions must be made to accommodate similar information from complementary sources such as prescribed and billed drugs to correctly represent the drug prescribing-dispensing-billing continuum.
 - Local extensions, modifications, translations and other aids in international classifications such as those deployed in ICD10 usage in Catalonia should be accounted to preserve the original intended meaning of the recording physician.