

Inception to ConcePTION: a conceptual framework for characterizing pharmacoepidemiologic data sources. A contribution from the ConcePTION project

Romin Pajouheshnia¹ (R.Pajouheshnia@uu.nl), Nicolas H. Thurin², Giuseppe Roberto³, Miriam Sturkenboom⁴, Rosa Gini³, on behalf of co-authors from IMI ConcePTION Work package 7

¹Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, The Netherlands; ²Bordeaux PharmacoEpi, INSERM CIC1401, Université de BORDEAUX, France; ³Agenzia regionale di sanità della Toscana, Osservatorio di epidemiologia, Florence, Italy; ⁴Julius Centre for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands

Introduction and objectives

In 2019, the Innovative Medicines Initiative (IMI) funded the ConcePTION project with the vision that there is a societal obligation to rapidly reduce uncertainty about the safety of medication use in pregnancy and breastfeeding.

The ConcePTION project's strategy was to design a tailored Common Data Model (CDM) with the capacity to preserve the granularity resulting from European data heterogeneity, be implemented within constrained timelines and budget, and conduct distributed analyses transparently and efficiently.

We introduce the set of concepts used to describe the European data sources involved in the ConcePTION project and illustrate the ConcePTION CDM, which serves as the keystone of the federated ConcePTION network.

Methods and results

20 Data Access Providers (DAPs) accessing 21 data sources were included in the study. A content analysis of the data sources was conducted following a series of structured 90-minute interviews with DAPs, in order to ascertain rich information on the data source(s) they provide access to. In parallel, the ConcePTION CDM was developed, drawing on the content and standards of existing CDMs, and was further informed by the findings of the content analysis.

Concept		Definition
Data source	Data bank	<i>Data bank</i> A collection of structured healthcare data, sustained by an organization, which mandates that a record is prompted whenever a specified class of events occurs to persons belonging to a specific population.
	<i>Originator of a data bank</i>	Organization which mandates and sustains the data bank.
	<i>Prompt of a data bank</i>	A specified class of events that prompt a record in the data bank.
	<i>Underlying population of a data bank</i>	Records are generated whenever the prompts happen to a specified population.
	Data source	A collection of data banks having the same underlying population, or overlapping populations, that can be linked to one another at an individual level.
Instance of a data source	Data source population	An organization may have access to entire data banks, or a subset thereof. The corresponding population is referred to as the data source population.
	Instance of a data source	Subset of a data source extracted for the purpose of conducting one or more studies.
	Instance population	The subpopulation of the data source population that is included in an instance of a data source.
Instance of a data source	Instance of a CDM	An instance of a data source converted to a CDM.

Table 1. ConcePTION conceptual framework.

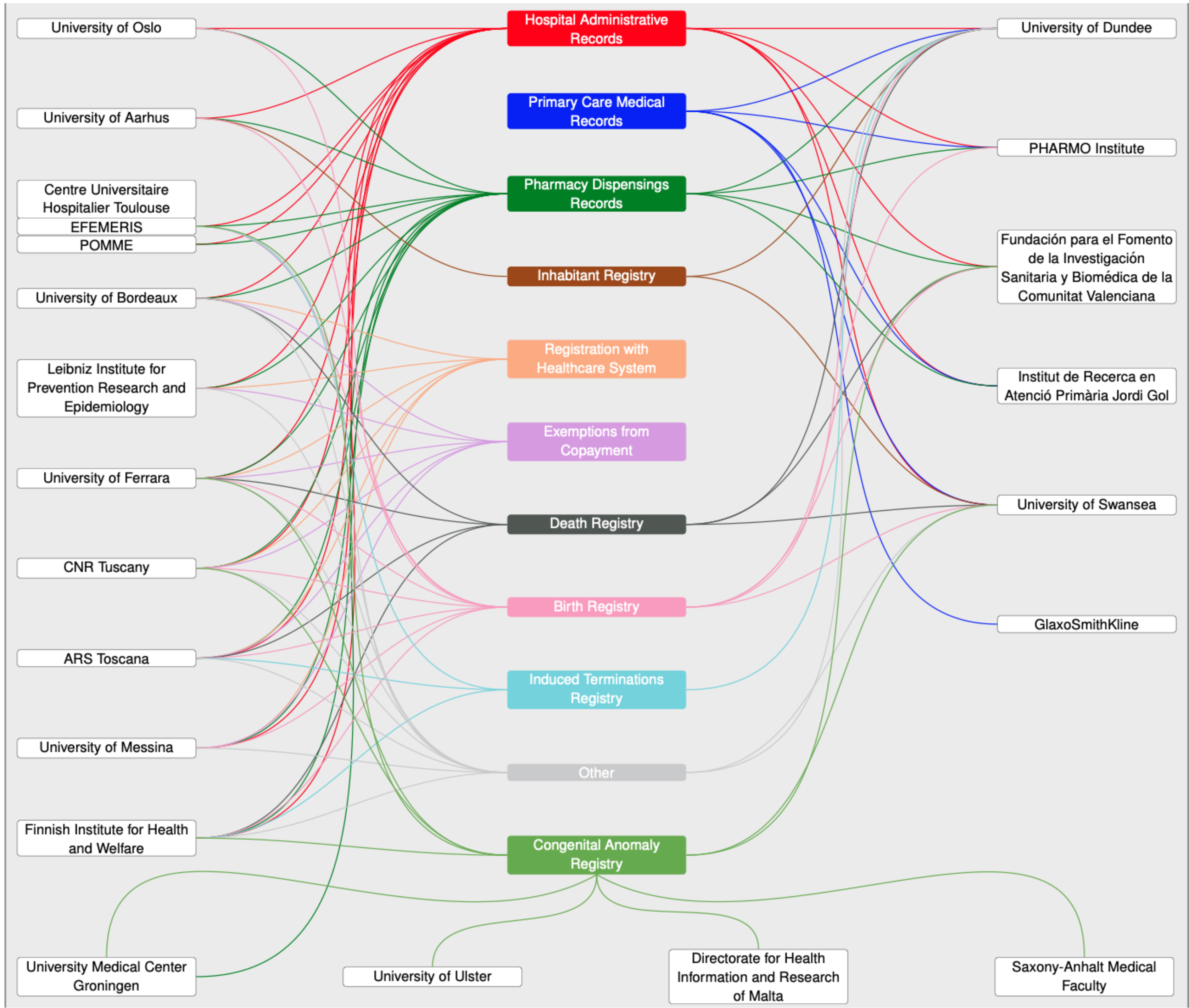


Figure 1. Data banks composing each data source

Conclusions

By its first anniversary, the ConcePTION CDM has enabled 13 data sources to run common scripts to contribute to major European projects, demonstrating its capacity to facilitate effective and transparent deployment of distributed analytics, and its potential to address effectiveness and safety questions about medicines exposure.

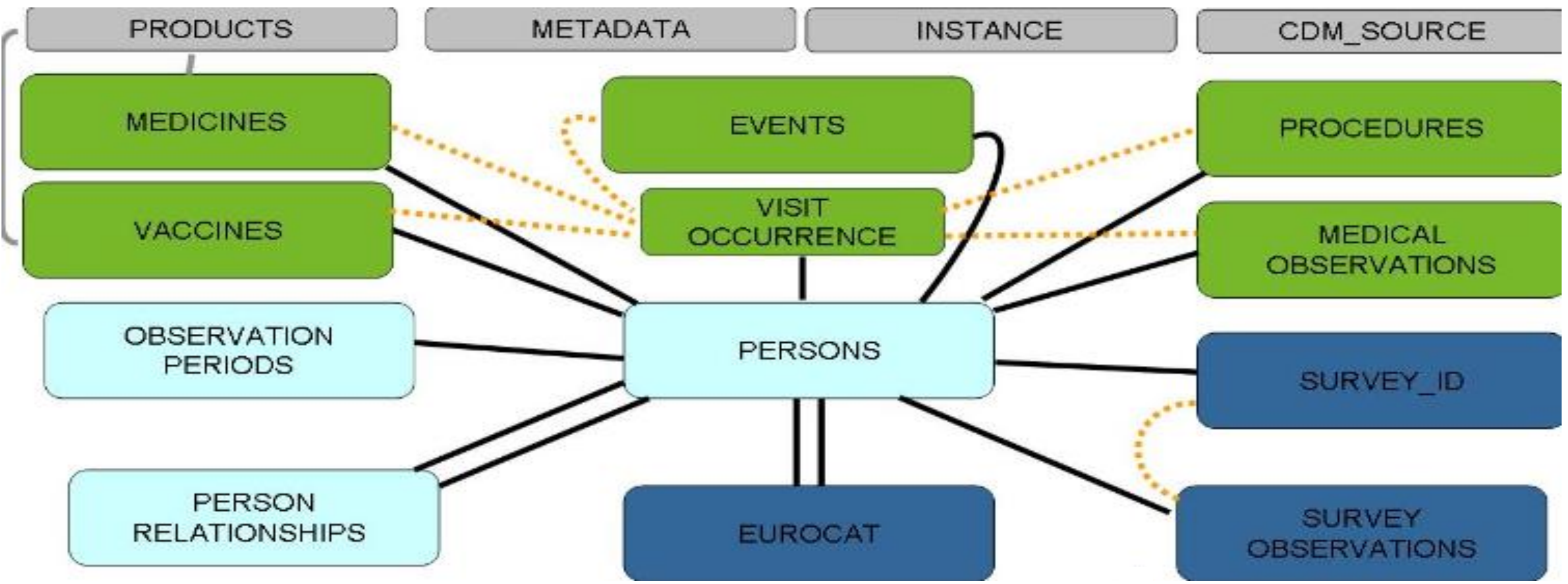


Figure 2. The ConcePTION CDM version 2.2.

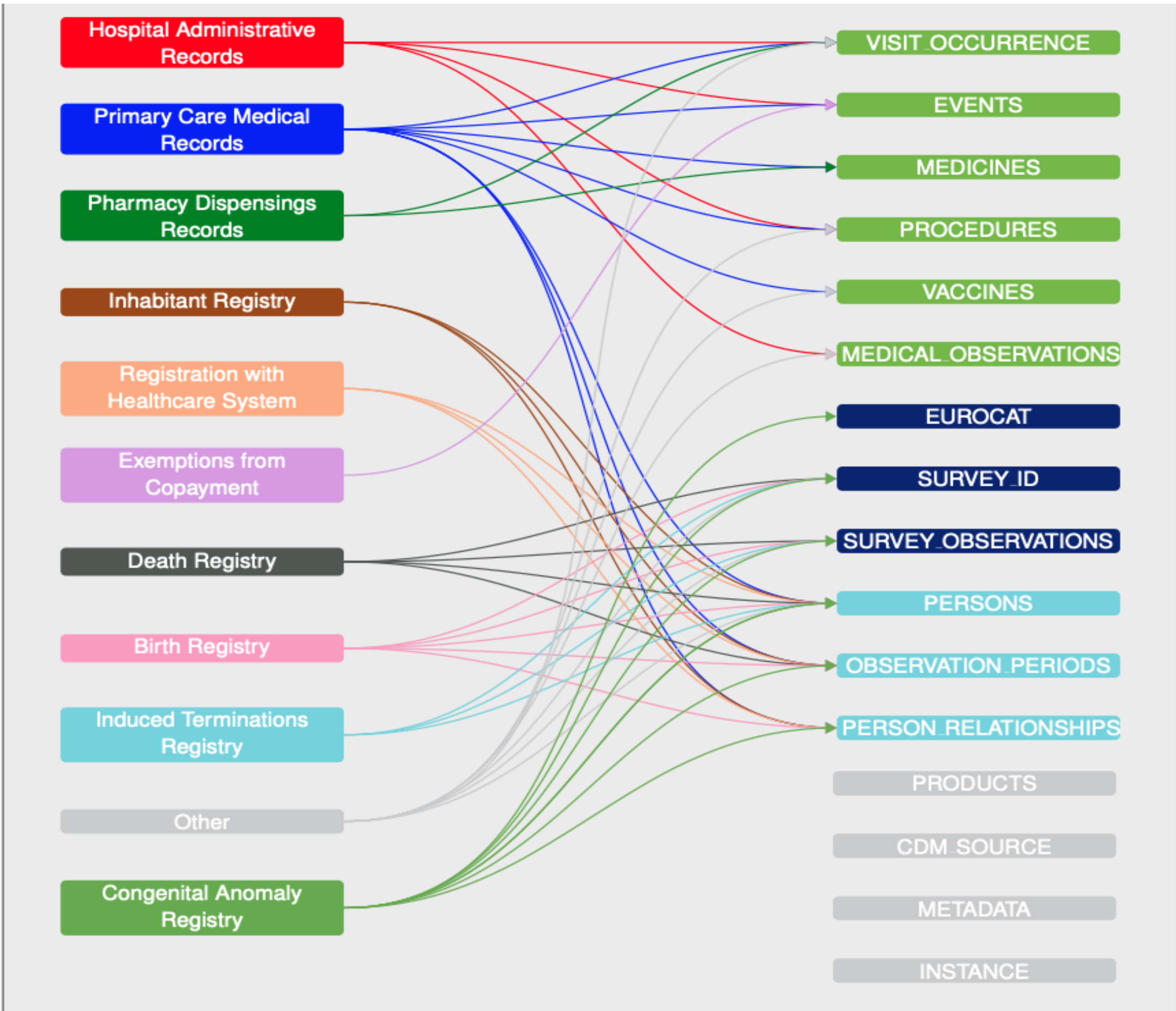


Figure 3. Common ETL between original families of data banks and tables of the ConcePTION CDM v2.2