



# Case study #1: OHDSI in Europe

Christian Reich  
QuintilesIMS



# Overview

- Experience so far
  - EMIF
  - QuintilesIMS
  - RxNorm Extension
- European OHDSI Chapter



# EMIF

## Facts & Figures

**Start Date** 01/01/2013

**End Date** 31/12/2017

### Contributions €

IMI funding 24 356 849

EFPIA in kind 24 124 503

Other 7 835 649

**Total cost** 56 317 001



## EMIF-Platform



Develop a framework for evaluating, enhancing and providing access to human health data across Europe, support EMIF-Metabolic and EMIF-AD as well as support research using human health data in general

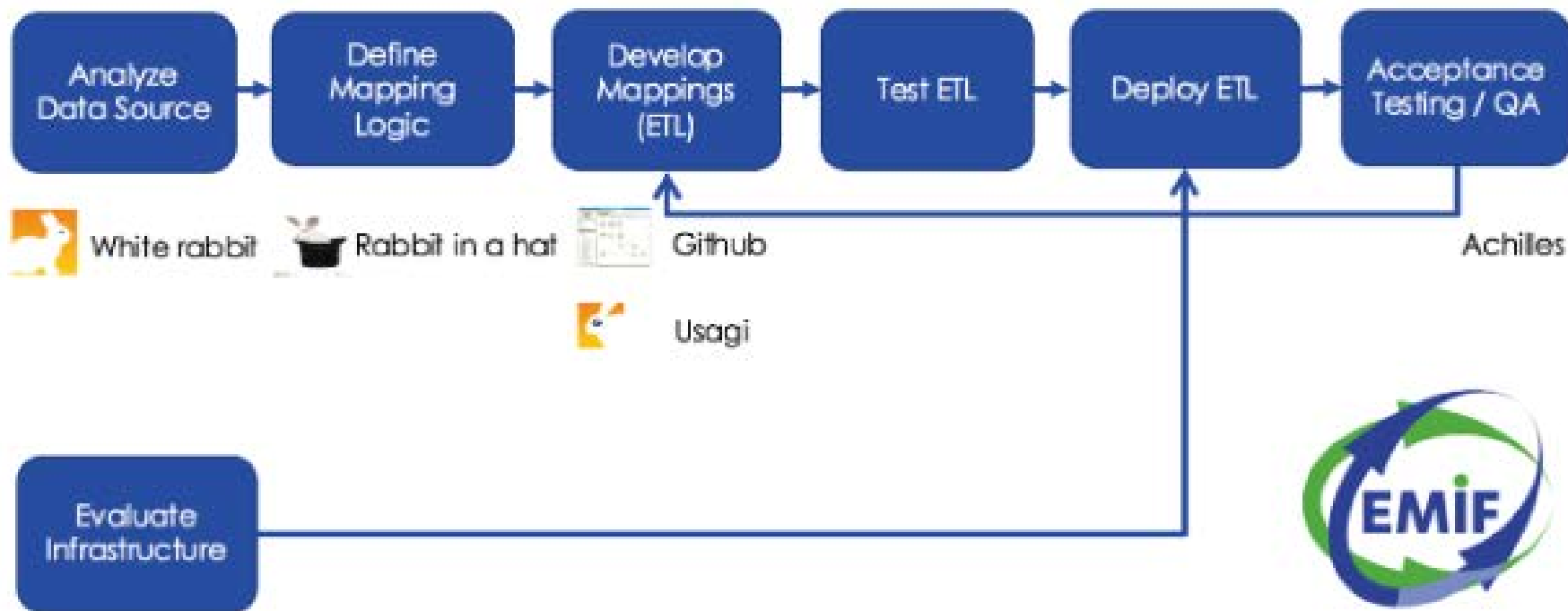


# EMIF: Participating Data Owners

Database	Country / Region	Population Size	Type
Agenzia regionale di sanità della Toscana (ARS)	Italy / Tuscany	5 10 <sup>6</sup>	Administrative database of Tuscan population
Aarhus University Hospital Database	Denmark / Northern region	2.3 10 <sup>6</sup>	Administrative database for population Central and North Jutland
Health Search - IMS HEALTH LPD	Italy	1.6 10 <sup>6</sup>	Primary care data of GP's using the Health Search System
Integrated Primary Care Information (IPCI)	Netherlands	2.8 10 <sup>6</sup>	Primary care database
Pedianet	Italy	400 10 <sup>3</sup>	Primary care pediatric database
Pharmo	Netherlands	8.4 10 <sup>6</sup>	Primary care database
Information System of Parc de Salut Mar (IMASIS)	Spain / Barcelona	1.4 10 <sup>6</sup>	Hospital database
The Information System for the Development of Research in Primary Care (SIDIAP)	Spain / Catalonia	6.4 10 <sup>6</sup>	Primary care database
The Health Informatics Network (THIN)	United Kingdom	12 10 <sup>6</sup>	Primary care database
Estonian Genome Center at the University of Tartu (EGCUT)	Estonia	50 10 <sup>3</sup>	Biobank



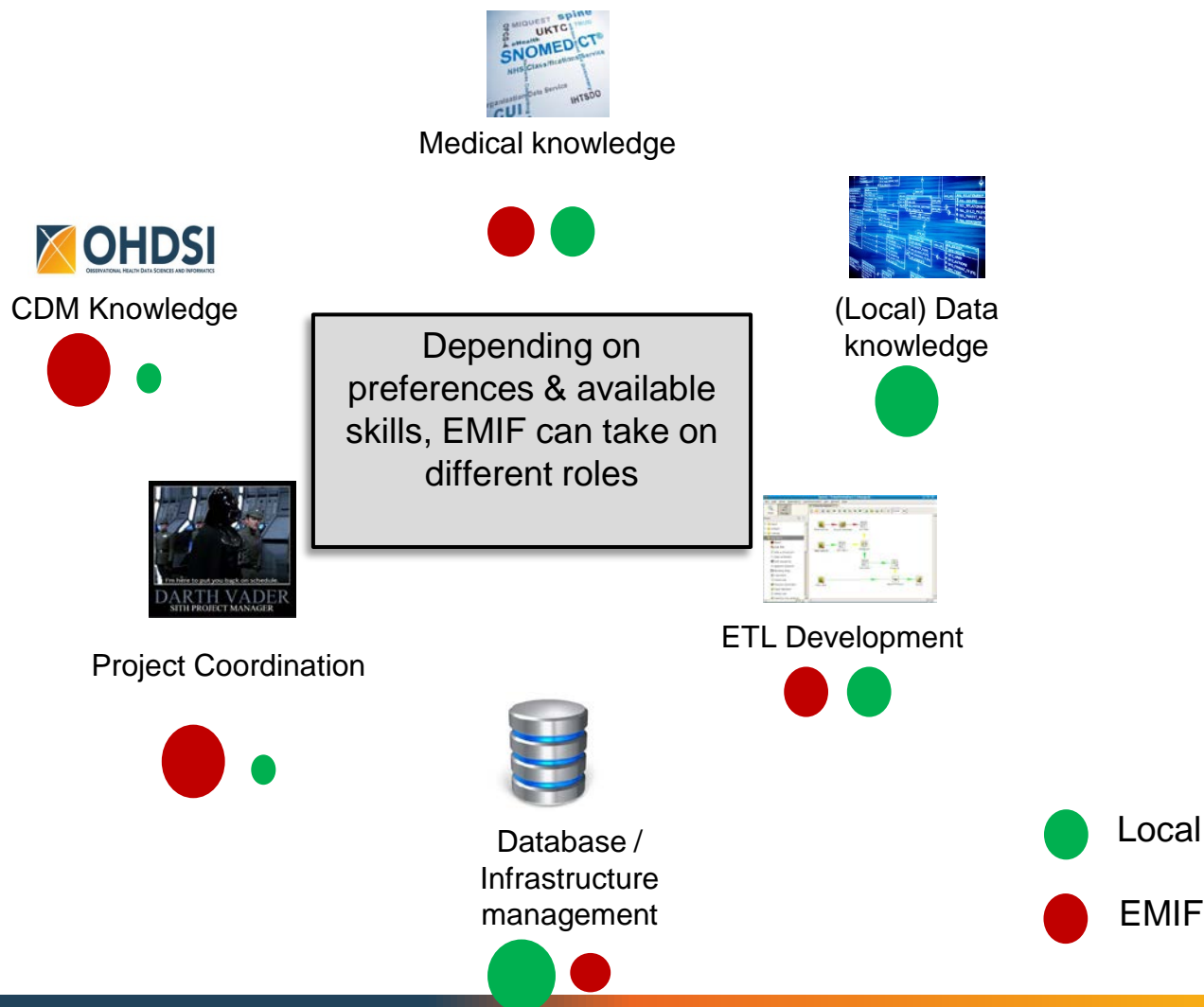
# EMIF: Standard ETL Process



- Local and CDM team together



# EMIF Support for data custodians





# ETL Best Practices shared within OHDSI community

The screenshot shows the OHDSI website interface. At the top left is the OHDSI logo. To its right is the text "Observational Health Data Sciences and Informatics". On the far right of the top bar is a "Log In" link. Below the top bar is a search bar and links for "Recent Changes", "Media Manager", and "Sitemap". A breadcrumb trail reads: "Trace: • ohdsi\_symposium\_2016\_poster\_submissions • ohdsi\_library • ohdsi\_symposium\_2016\_posters • data\_network • welcome • overview • **etl\_best\_practices**".

The main content area is titled "ETL creation best practices". It includes a paragraph: "This document describes some of the best practices we have developed over the years when trying to create an ETL (Extract, Transform, Load) process to convert data into the OMOP Common Data Model (CDM). We have found it best to split the process into four distinct activities:" followed by a numbered list:

1. Data experts and CDM experts together design the ETL
2. People with medical knowledge create the code mappings
3. A technical person implements the ETL
4. All are involved in quality control

Below this is a section titled "1. Data experts and CDM experts together design the ETL". It contains a paragraph: "Designing the ETL requires in-depth knowledge of the source data, but it also requires knowledge of the CDM, and having someone with experience in past ETLs to the OMOP CDM can speed up the design activity. Ideally, the data and CDM experts should sit down together at the same location in a one- or two-day session." and another paragraph: "We have developed two tools that have proven to be helpful for this activity: [White Rabbit](#) and [Rabbit-in-a-Hat](#)."

On the left side of the page is a sidebar with navigation links under the heading "Documentation". The links are: "Getting Started with OHDSI", "Common Data Model (CDM)" (with sub-links "CDM Specifications" and "CDM Vocabulary"), "Convert Database to CDM (ETL)" (with sub-links "ETL creation best practices", "Example ETLs", "ETL Tools", and "ETL Support"), and "Tool Specific Documentation" (with sub-links "ATLAS", "ACHILLES", "White Rabbit", "Usagi", "Methods Library", "WebAPI", and "Laertes").

On the right side of the page is a "Table of Contents" for the document "documentation:etl\_best\_practices". It lists the following items:

- ETL creation best practices
  - 1. Data experts and CDM experts together design the ETL
    - White Rabbit
    - Rabbit-in-a-Hat
  - 2. People with medical knowledge create the code mappings
  - 3. A technical person implements the ETL
  - 4. All are involved in quality control

# SIDIAP: Information System for Development of Research in Primary Care.

---



- **2010: IDIAP Jordi Gol** (UAB Primary Care Research Inst.) + **Catalan Institute of Health** (Government of Catalonia)
- **One primary care health provider** using one same e-records software (eCAP)
- **289** Primary Care Centres in Catalonia
- **>5,8 million** people (80% population)
- **>3,400** GPs
- **>40 million** person-years of research-usable data (2006 onwards)



# Integrated Primary Care Information (IPCI)



- Project started in 1989
- Data from general practitioners (GP)
- GP is gatekeeper for medical care in NL
- Mixture of city and country practices

475 practices

372 active practices

601 active general practitioners

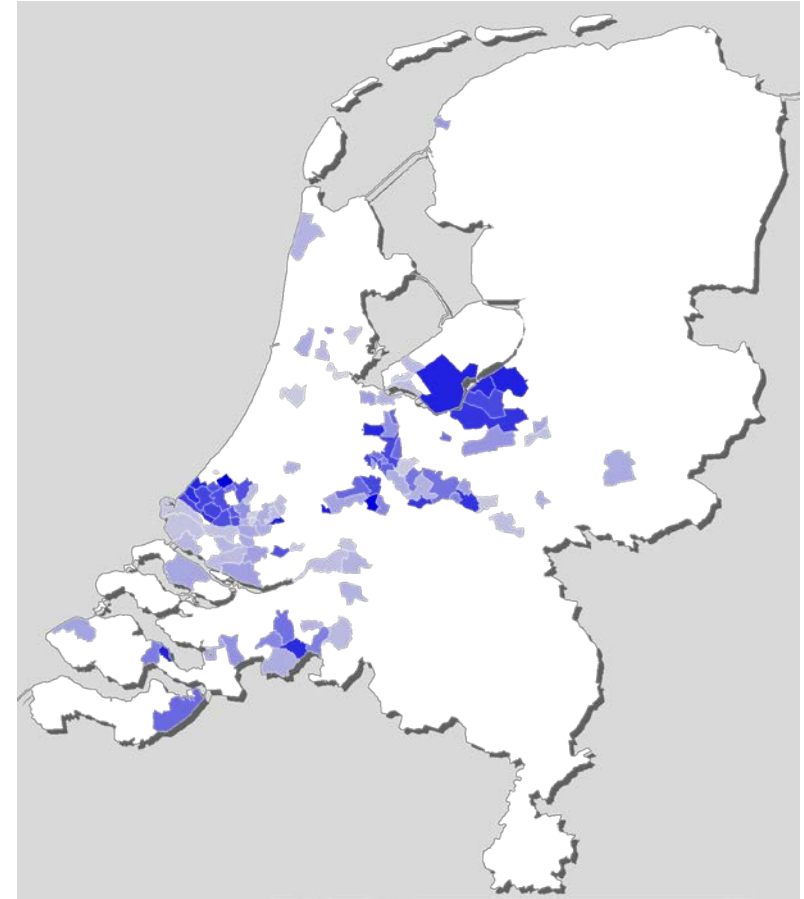
1.900.000 research patients

1.000.000 active patients

19.160.000 patient years

5.658.000 research quality years

*(16.800.000 citizens in The Netherlands)*





# Feedback from EMIF Experience

1. Mixed local/CDM teams
2. Achilles immediately
3. Create Data Catalog
4. Multiple languages: Create data dictionary in English
5. Large complexity of source data: Wide and long format tables are mixed, often redundancies
6. Local terminology: often at a different level of granularity
7. European drugs are often not included in the RxNorm vocabulary
8. Lack of concept types required for European Healthcare setting
9. Geographic information: current location entity is based on US geographies
10. Data owners have nice visualizations that can be made available to the entire OHDSI community



# New IMI Initiative: EHDN



## RESEARCH & INNOVATION

### Participant Portal

European Commission > Research & Innovation > Participant Portal > Opportunities

HOME

FUNDING OPPORTUNITIES

HOW TO PARTICIPATE

EXPERTS

SUPPORT ▾

Search



LOGIN



REGISTER

#### EU Programmes 2014-2020

Search Topics

Updates



Calls



H2020

3rd Health Programme

Asylum, Migration and  
Integration Fund

Consumer Programme

COSME

European Statistics

Hercule III Programme

Internal Security Fund - Borders

Internal Security Fund - Police

### TOPIC : European Health Data Network (EHDN)

**Topic identifier:** IMI2-2017-12-04  
**Publication date:** 19 July 2017

**Types of action:** IMI2-RIA Research and Innovation action  
**DeadlineModel:** two-stage  
**Opening date:** 19 July 2017  
**Deadline:** 24 October 2017 17:00:00  
**2nd stage Deadline:** 16 May 2018 17:00:00

Time Zone : (Brussels time)



Horizon 2020

[Call budget overview](#)

#### Topic Description

[- Less](#)

##### Specific Challenge:


Healthcare data are very fragmented. Even data within one healthcare centre are typically spread across different repositories. Across entities, different standards are used to code diagnosis, lab results, drugs or procedures. In most healthcare systems, a majority of the core clinical data is buried in unstructured (text) notes, making data analysis even more challenging.



# QuintilesIMS Data



# QuintilesIMS Data Assets




IMS LifeLink Data Asset	US 	CA 	DE 	UK 	FR 	IT 	ES 	CH 
Electronic Medical Records	Ambulatory & Oncology	Ambulatory	Ambulatory, Diabetes	Ambulatory	Ambulatory, Diabetes	Ambulatory	-	-
Longitudinal Prescriptions	Retail, Mail, Specialty, LTC	Retail	Retail	Retail	Retail	Retail & DPC (direct to patient)	Retail	Retail & self-dispensing doctors
Adjudicated Claims	Payer (all sites)	Payer (Drug Plans)	-	-	-	-	-	-
Pre-adjudicated Claims	Medical	-	-	-	-	-	-	-
Hospital	In & outpatient	-	-	In & outpatient (England only)	-	-	-	-
Other	Consumer, Retail Purchasing-OTC/CPG	-	-	-	-	-	-	-

IMS LifeLink Data Asset	NL 	BE 	HU 	PT 	PL 	AU 	JP 	KR 
Electronic Medical Records	-	-	-	-	-	-	-	-
Longitudinal Prescriptions	Retail	Retail	Retail	-	Retail	Retail	Retail	Retail
Adjudicated Claims	-	-	-	-	-	-	-	-
Pre-adjudicated Claims	-	-	-	-	-	-	-	-
Hospital	In & outpatient	In & outpatient	-	In & outpatient	-	-	-	-
Other	-	-	-	-	-	-	-	-



# European Data – to be OMOP CDMed

			
Patients	10.1 million	33millions	12 million
Drugs	FCC	ATC, ATC Ephemra, PZN	Gemscripts, dm&d, Read
Diagnosis	ICD10	ICD10, ICD10GM	Read
Procedure	ICD10	N/A	
Measurements	ICD10, free form text	ICD10, free form text	Read
Patients	Unique per refresh	Unique per practice/patient ID	Unique per patient ID
Region	French Departments	East/West	England, Wales, Scotland, Ireland
Provider profile	GP only	GP plus some specialist	GP only



# Source Table Structures




Table Name	No of Fields	Description 
AHD	25	Additional Health Data: lifestyle data, preventative healthcare, immunizations, test results and death details.
Consult	9	Practice consultations with date, time and duration.
Dosage	3	Prescribed dosage
Medical	22	Records of symptoms, diagnoses and interventions recorded by the primary care team.
Pack	8	Links to the Therapy table when a quantity have prescribed in packs.
Pack_size	2	Links to the Therapy table when a quantity have prescribed in packs.
Patient	18	Information on patient characteristics and re
Staff	4	Specific roles of the practice staff
Therapy	24	Prescriptions issued to patients
Thin_prac	9	Created for updates and provides a summary of practice

Table Name	No of Fields	Description 
DAFR, Back_Office_data	27	All data is stored in this one table
NDF Reference file	37	Reference file to drugs in back_office_data.
Reference file for tests	5	Results or tests for patients in back_office_data
Doctor Attributes	6	Provider level specifics

Source Data Tables	No of Fields	Description 
Practice	16	Practice level information including specialties.
Patient	19	All documented patients and their demographic data.
Diagnosis	22	ICD -10 codes included in the data
Diagnosis Event Text	3	Additional Diagnosis Event Information
Therapy	47	WHO ATC5 level
Problem Events	6	All documented diagnosis/problem level data, like other_event tables descriptions.
Prescription Events	21	All documented prescription level data.
Test Prevention Events	21	All documented test level with units and value data
Adipositas Smoking Events	8	All documented obesity and smoking status
Action Events	8	All documented "notes" from physician (i.e. referrals, sick leave, hospitalization, etc.



# Lesson Learned

- Date Shifting
  - Added logic to shift date of actual patient transactions
- Encrypt/De-identify Provider or Plan information within a link dataset
  - Encrypted provider ID information when linked claims with EMR dataset
- Privacy ICD9/10 Codes
  - Removal of ICD9/10 codes that are considered privacy issues, such as death or sexual abuse
  - Using “fake” date in Death table to indicate a death
- Data unable to leave a specific country
- Pilot Patients
  - Removal of patients that were are “dummy patients”
- Patients without transaction
  - Adding an observation period
- Local country vocabulary mapping
- Local knowledge of each countries health system
- Knowledge of local data and business rules
- Ability to extract patient level data
- Cleaning dirty data
- Standardize measurement and unit of measure
- Source field not transferring to OMOP CDM





# Drugs outside RxNorm – RxNorm Extension



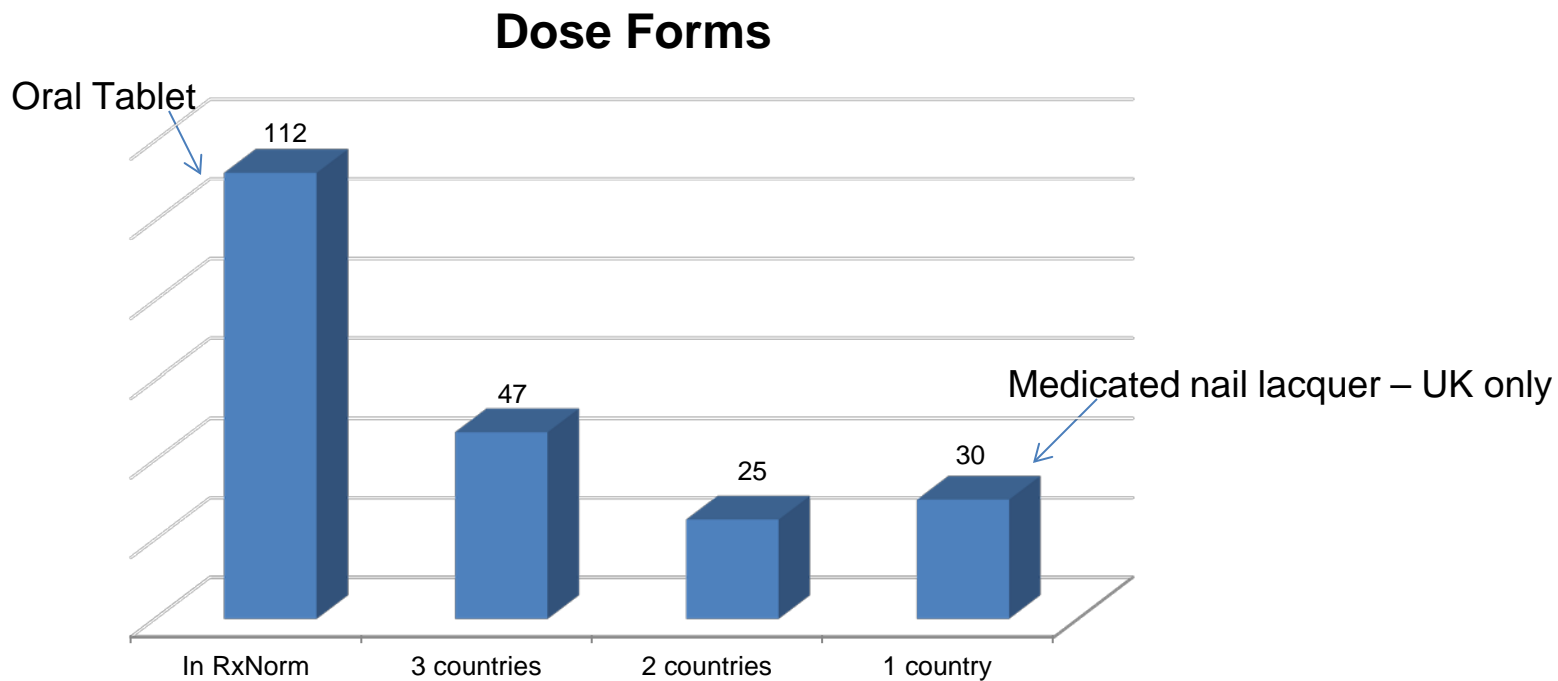
# Different Drug Markets – Different Formulations



Nicotine Chewing Gum [Nicorette]	RxNorm	France, Germany, UK
Nicotine Oral Lozenge [Nicorette]	RxNorm	France, Germany, UK
Nicotine Cartridge [Nicorette]		Germany
Nicotine Dry Powder Inhaler [Nicorette]		France
Nicotine Inhalant Solution [Nicorette]		UK
Nicotine Metered Dose Inhaler [Nicorette]		Germany
Nicotine Nasal Spray [Nicorette]		Germany, UK
Nicotine Oral Solution [Nicorette]		Germany
Nicotine Oral Spray [Nicorette]		France, Germany
Nicotine Sublingual Tablet [Nicorette]		France, Germany, UK
Nicotine Topical Solution [Nicorette]		Germany
Nicotine Transdermal System [Nicorette]		France, Germany, UK

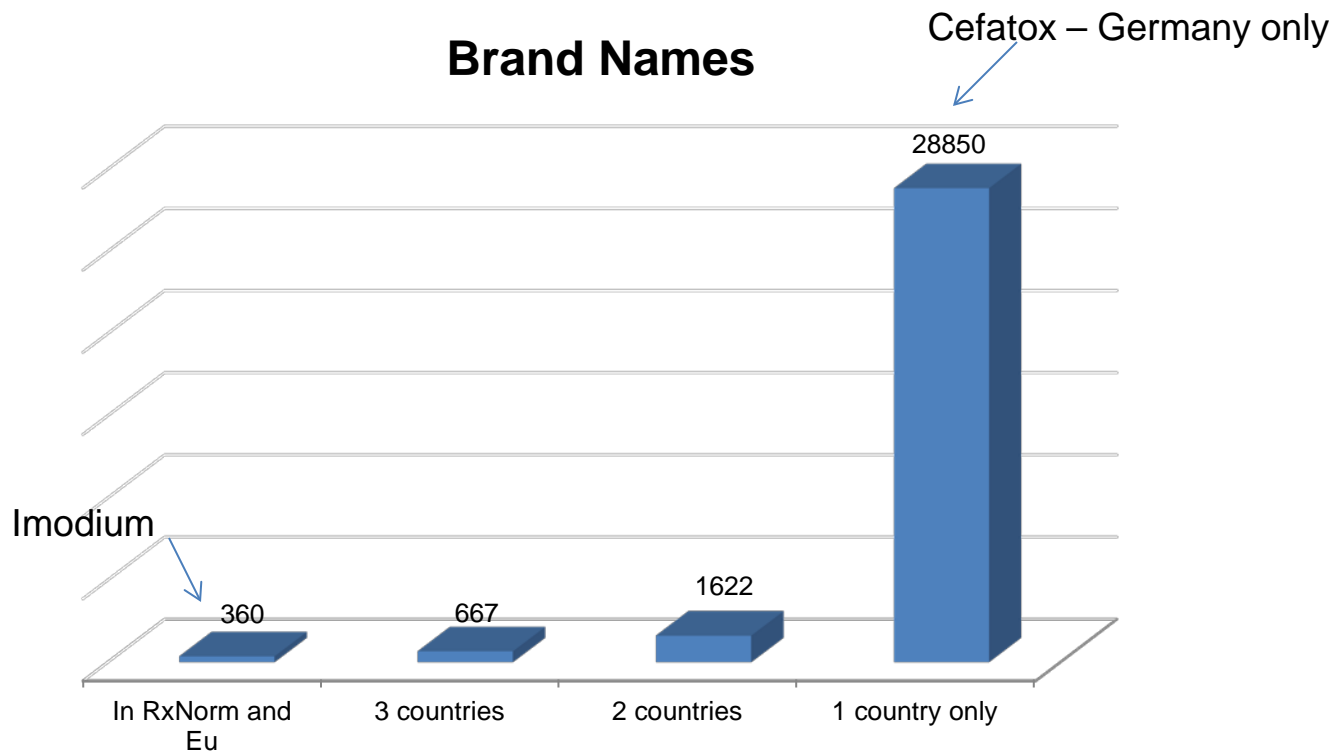


# Use of Dose Forms



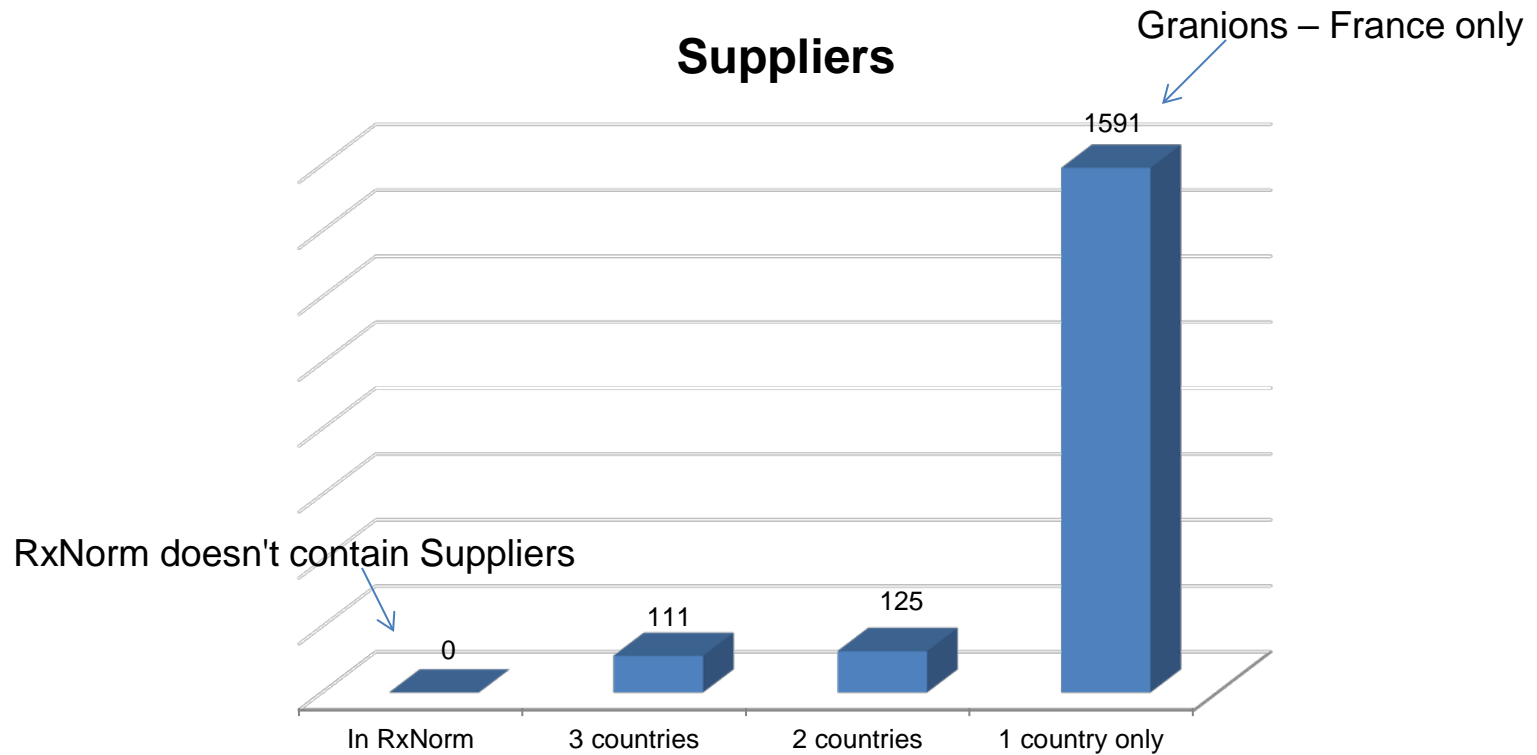


# Use of Brand Names



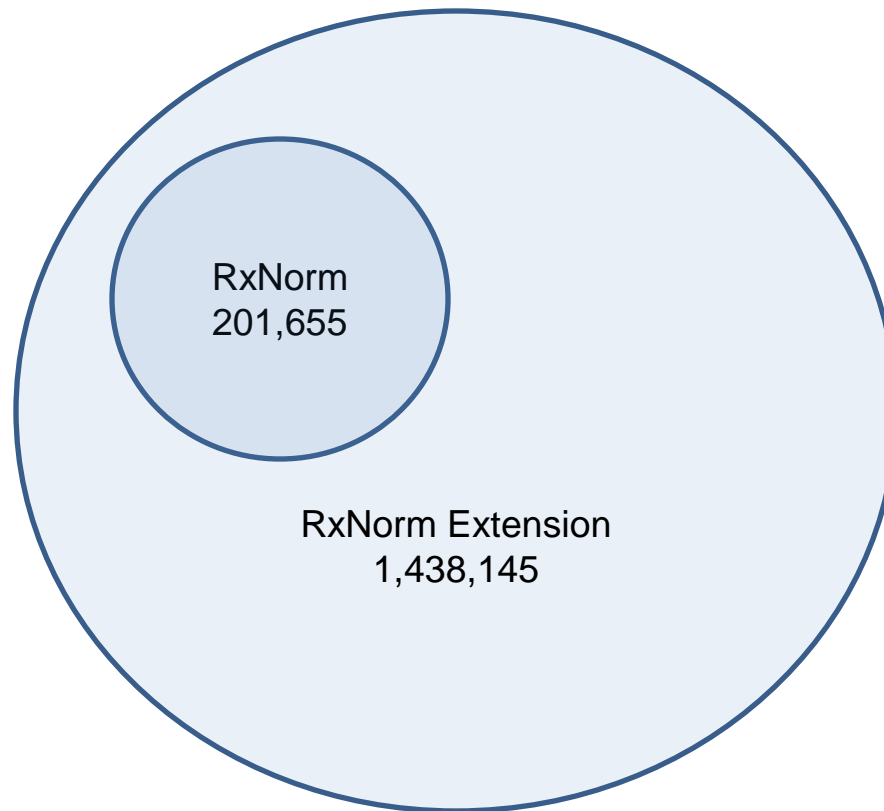


# Use of Suppliers





# Introducing RxNorm Extension: RxNorm Worldwide





# RxNorm Extension Hierarchy

“Nicotine 2 MG Sublingual Tablet [Nicorette] Box of 100 by Johnson & Johnson”

Classes	Attributes	Ingredient	Drug Strength	Dose Form	Brand Name	Box size	Supplier
Ingredient		Nicotine					
Clinical Drug Component		Nicotine	2 mg/mL				
Clinical Drug		Nicotine	2 mg/mL	Sublingual Tablet			
Branded Drug		Nicotine	2 mg/mL	Sublingual Tablet	Nicorette		
Quantified Branded Drug							
Quantified Branded Box		Nicotine	2 mg/mL	Sublingual Tablet	Nicorette	Box of 100	
Marketed Product		Nicotine	2 mg/mL	Sublingual Tablet	Nicorette	Box of 100	J&J



## European OHDSI Chapter





# European OHDSI Chapter

## Why a European coordinating center?

- We need to stimulate **transparent and reproducible** research in Europe.
- Interest in the OMOP-CDM is growing strongly, but **coordination is needed** to achieve high standards in data conversion.
- Improving operability is challenging in Europe because of the **large amount of vocabularies** -> major mapping effort is needed with local input.
- European **training** centers are needed to train all stakeholders.
- **Regulators interested** in OHDSI would like to have a European OHDSI office.
- A **European community** needs to be build that actively contributes to tool development, adoption, and study execution.
- European funding resources are available through **several program** to improve operability in the medical domain
- An **annual symposium** is needed to inform and enlarge the European OHDSI community
- etc...



**First Annual**

# **EUROPEAN OHDSI SYMPOSIUM**

**March 23th 2018**

**Tutorials March 24th**



## **Bridging Europe**

**Erasmus MC Rotterdam The Netherlands**



# When and where

Department of Medical Informatics  
Erasmus MC Rotterdam, The Netherlands

- March 23th 2018 OHDSI Symposium
- March 24th 2018 Tutorials
- Max 250 participants
- Poster sessions
- [www.ohdsi-europe.org](http://www.ohdsi-europe.org)
- email: [info@ohdsi-europe.org](mailto:info@ohdsi-europe.org)

Registration for the Symposium will open Sept 1th.

Registration for tutorials will be announced through the website and OHDSI communication channels.

Interested in sharing European results as a keynote speaker or on a poster? Let us know!

