



The European OHDSI Initiative: Why are we here?

Peter Rijnbeek

Department of Medical Informatics
Erasmus MC Rotterdam



Performing multi-database studies in Europe

Our Goals:

- We want to generate real-world evidence in a fully reproducible way
- We want to shorten the time from study conception till publication without sacrificing quality

Our Challenges:

- We have to deal with a large amount of different database structures
- We have many different coding systems in Europe (ICPC, READ, ICD9, ICD10, etc.)
- Our health care systems differ considerably across Europe

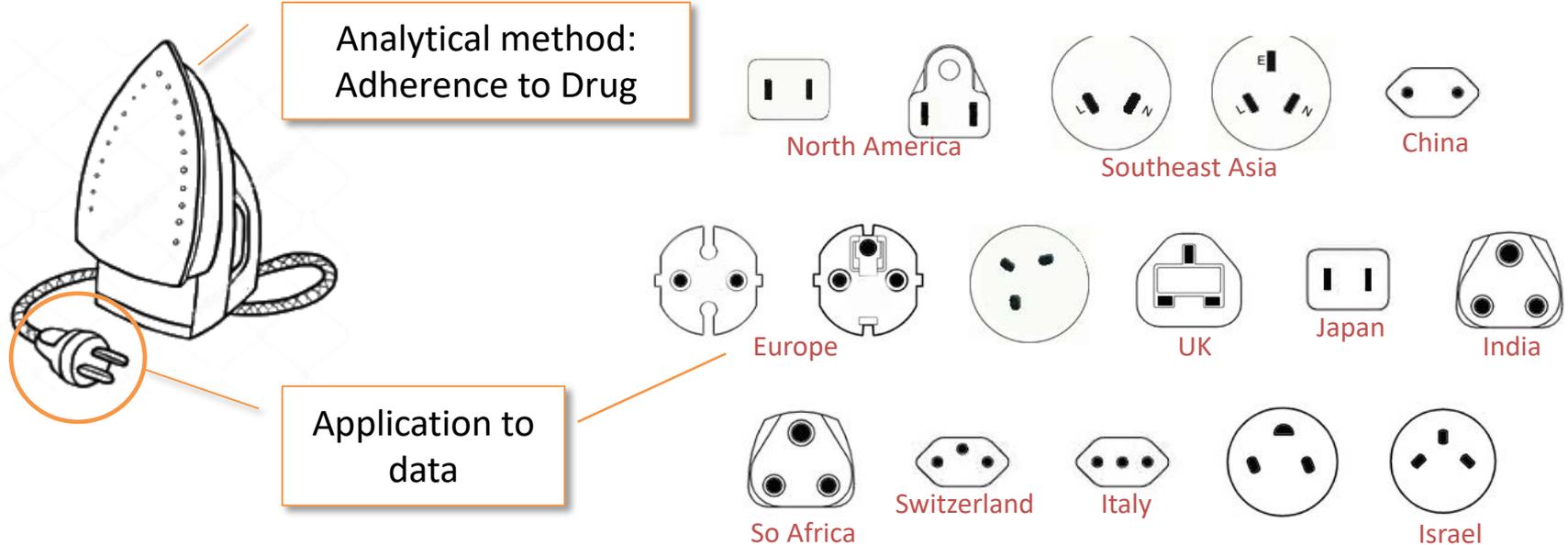


We need to find solutions to improve the inter-operability of data

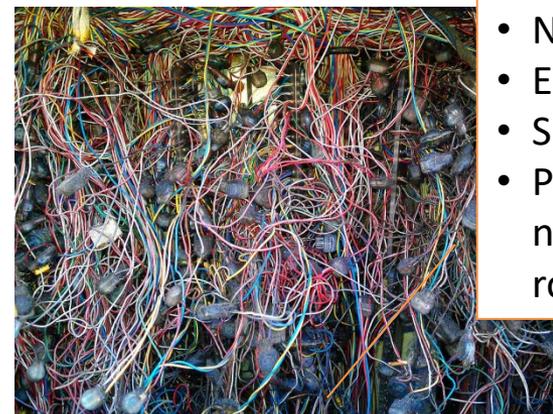
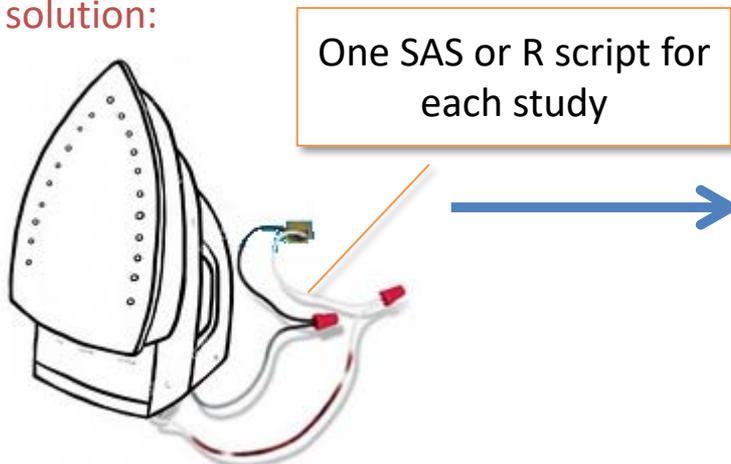
We need a framework for rapid access to and fully reproducible analysis of data

Current Approach: "One Study – One Script"

"What's the adherence to my drug in the data assets I own?"



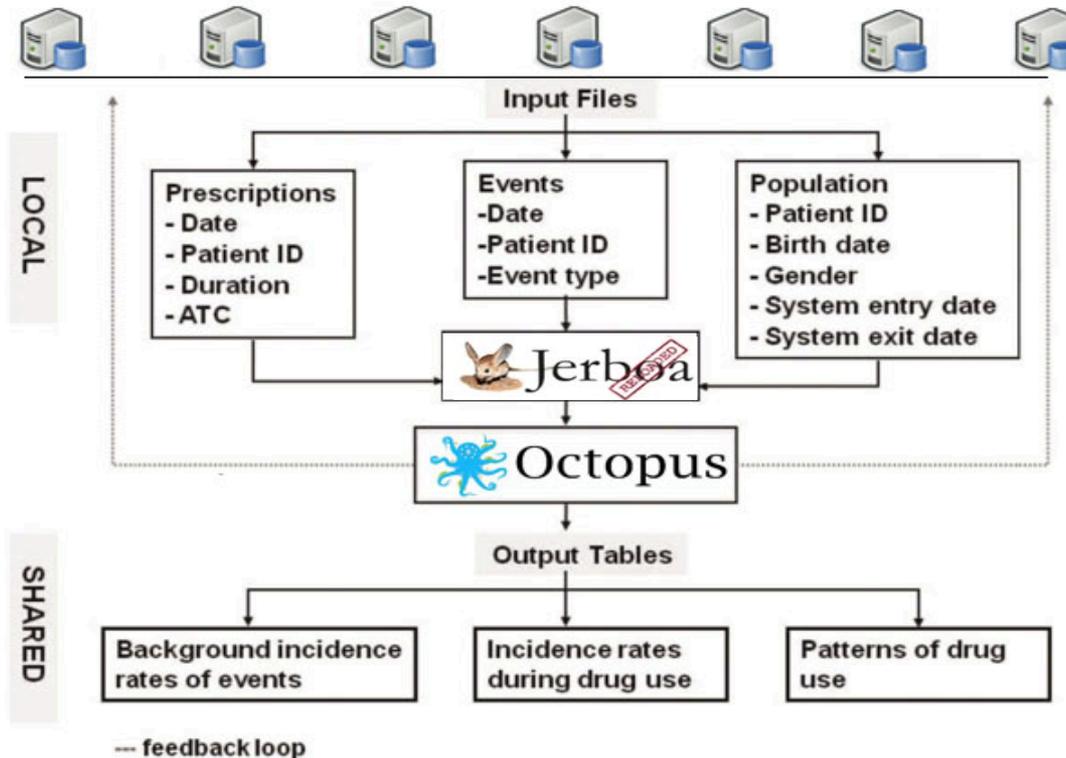
Current solution:



- Not scalable
- Not transparent
- Expensive
- Slow
- Prohibitive to non-expert routine use



We have made progress by developing tools to standardize part of the process





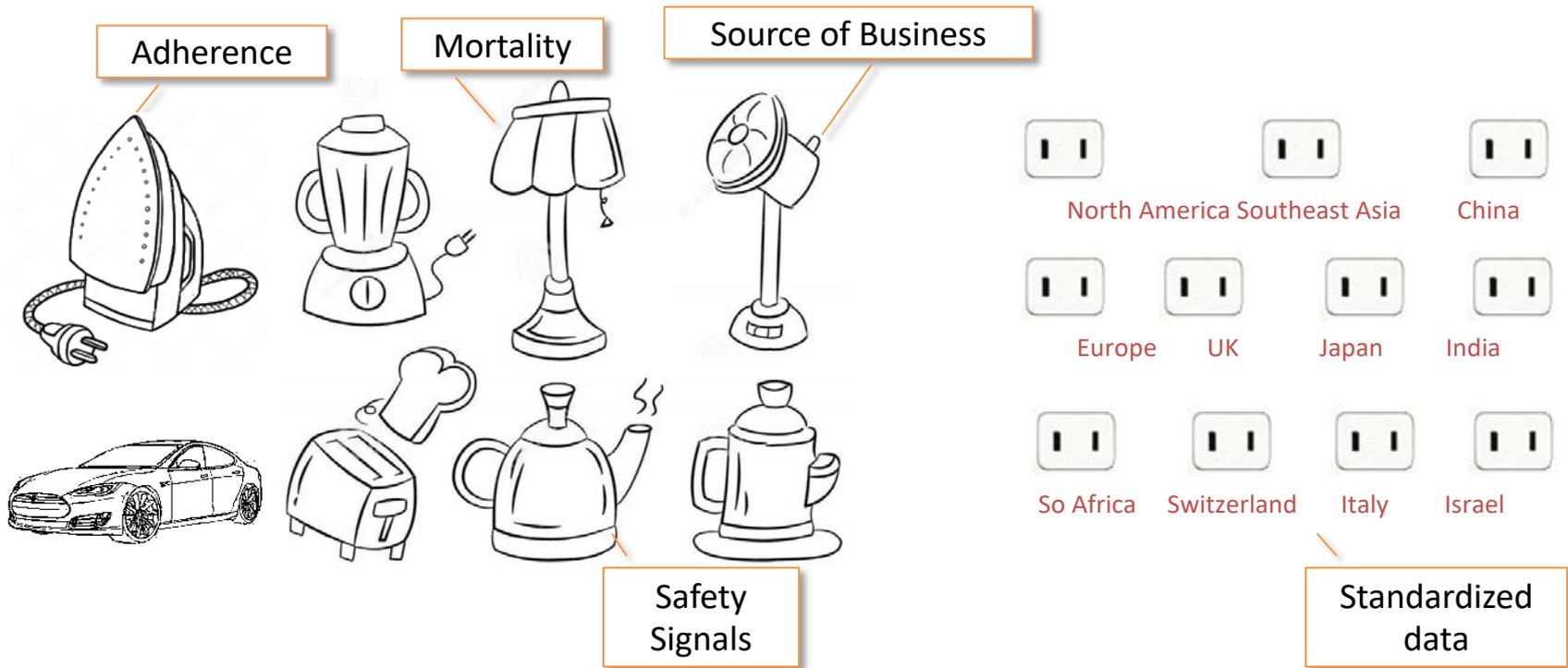
But we can do better...



- Each data custodian has to create custom data extraction scripts
- This step is often not transparent to the study lead
- It is time consuming to perform the semantic harmonization step



Solution: Data Standardization Enables Systematic Research



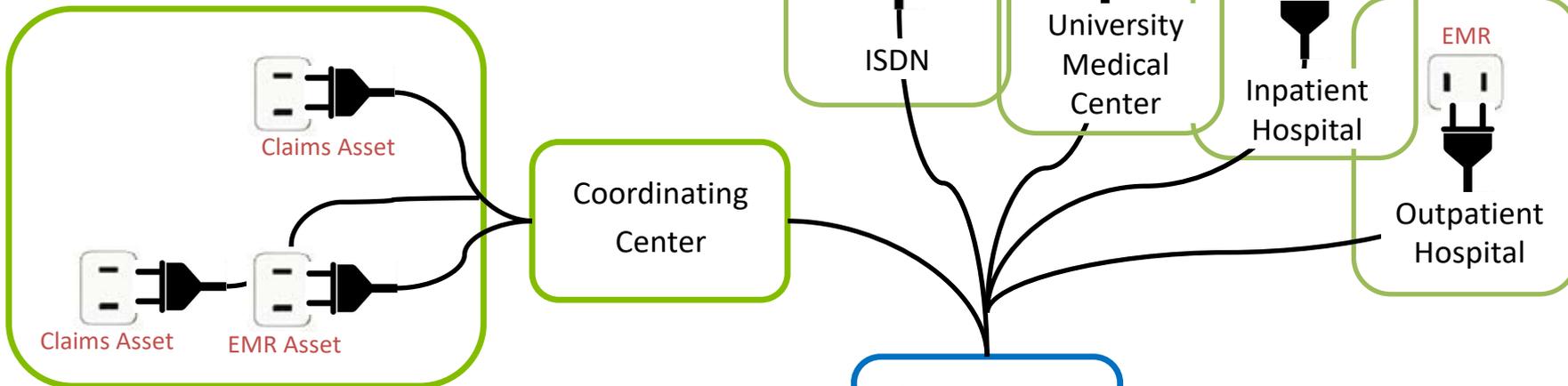
OHDSI Tools

OMOP CDM



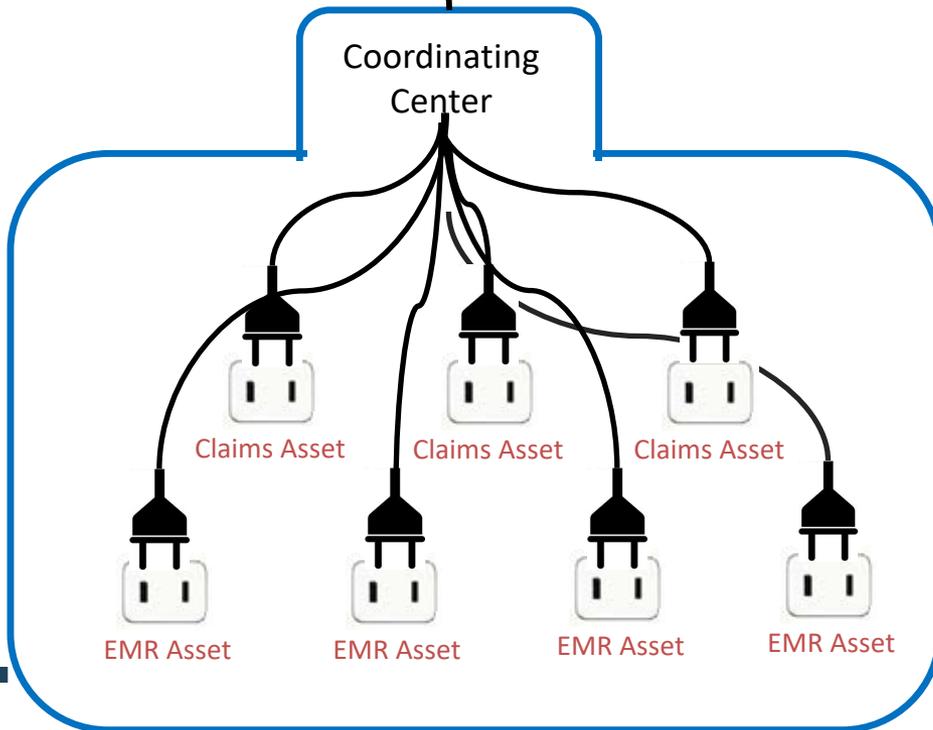
Network Studies

Networks of networks



Another Network

Network





The European Medical Information Framework (EMIF) Project

ACADEMIC PARTNERS

37



14 European countries combining
57 partners

€56 million worth of resources

3 projects in one

5 year project (2013–2018)

57
Partners

SME PARTNERS

9



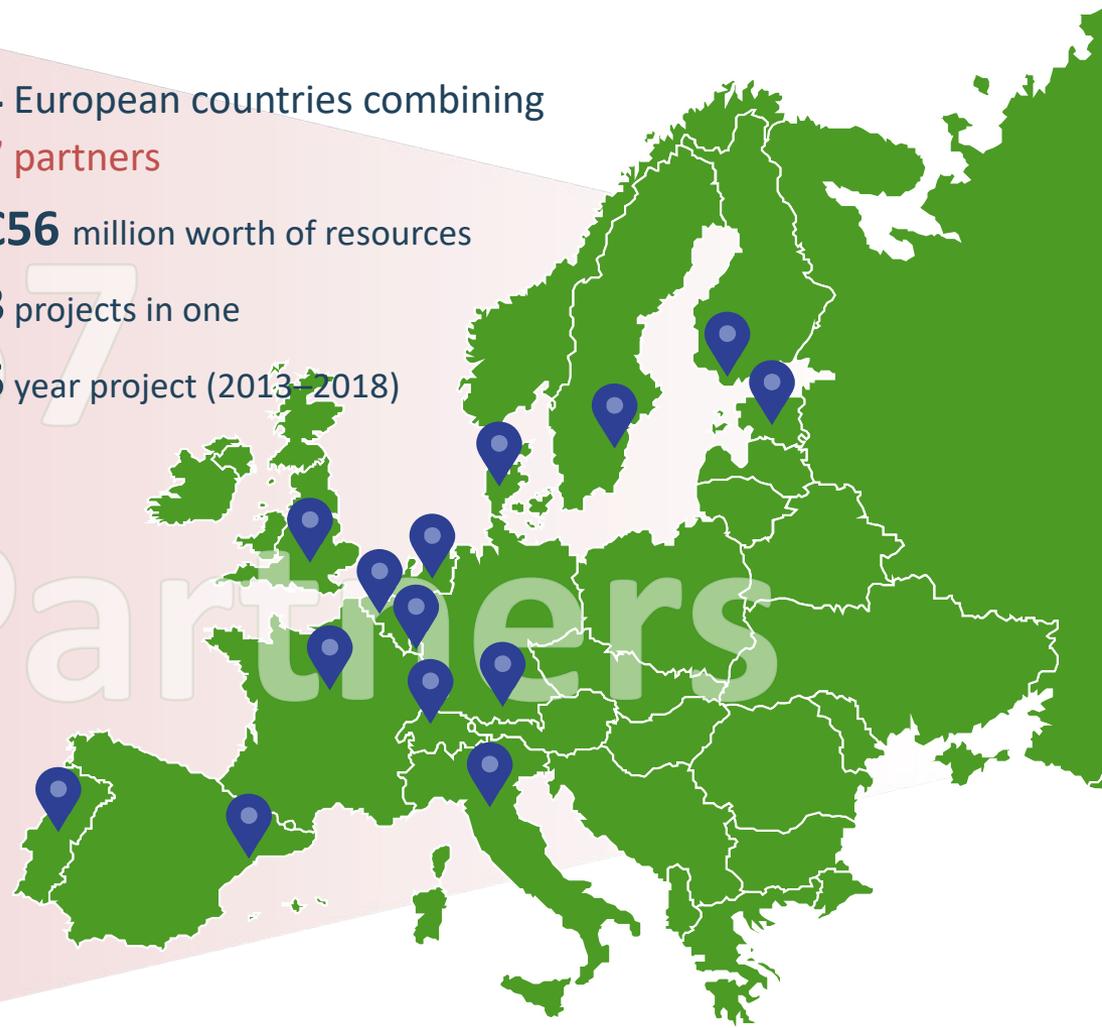
EFPIA PARTNERS

10



PATIENT ORGANISATION

1



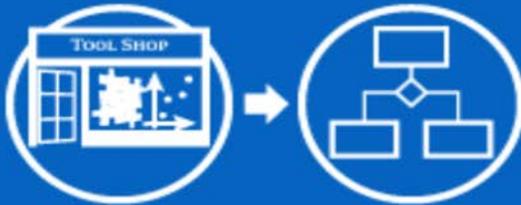


Goals of EMIF

EMIF-Platform Objective

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

Tool Development



Key tool developed—**EMIF Catalogue** as data “shop window” to support the platform architecture, also being utilised by other initiatives (DP-UK, EPAD)

Workflow, data extraction, harmonisation and aggregation software developed and being used to support several studies as use cases for research process

Common Data Model



Piloting the **OMOP common data model** in the platform databases to assess its potential towards data harmonization



Why the OMOP-CDM

Our requirements:

- Active open source community: mapping tools, analytical tools etc.
 - Solution for semantic harmonization by using standardized vocabularies
 - Need to store ALL source data including source vocabularies
 - Possibility to refine to database-specific queries
 - Multiple technical infrastructures supported
-



EMIF Adopted the OMOP-CDM and has worked on mapping 10 European databases

Table 1: The 10 European databases that are part of the EMIF initiative and that are mapped to the OMOP CDM.

Database	Country / Region	Population Size	Type
Agenzia regionale di sanita della Toscana (ARS)	Italy / Tuscany	5 10 ⁶	Administrative database of Tuscan population
Aarhus University Hospital Database	Denmark / Northern Region	2.3 10 ⁶	Administrative database of Central and North Jutland
Health Search IMS Health LPD	Italy	1.6 10 ⁶	Primary care data of GP's using the Health Search System
Integrated Primary Care Information (IPCI)	Netherlands	2.8 10 ⁶	Primary care database
Pedianet	Italy	0.4 10 ⁶	Pediatric database
Pharmo	Netherlands	8.4 10 ⁶	Primary care database
Information System of Parc de Salut Mar (IMASIS)	Spain	1.4 10 ⁶	Hospital database
The Information System for the Development of Research in Primary Care (SIDIAP)	Spain / Cataluna	6.4 10 ⁶	Primary care database
The Health Informatics Network (THIN)	United Kingdom	12 10 ⁶	Primary care database
Estonian Genome Center at the University of Tartu	Estonia	52 10 ³	Biobank

ETL requires multi-disciplinary team



Medical knowledge



CDM Knowledge



(Local) Data knowledge



Depending on preferences & available skills, EMIF can take on different roles



Project Coordination



ETL Development



Database / Infrastructure management

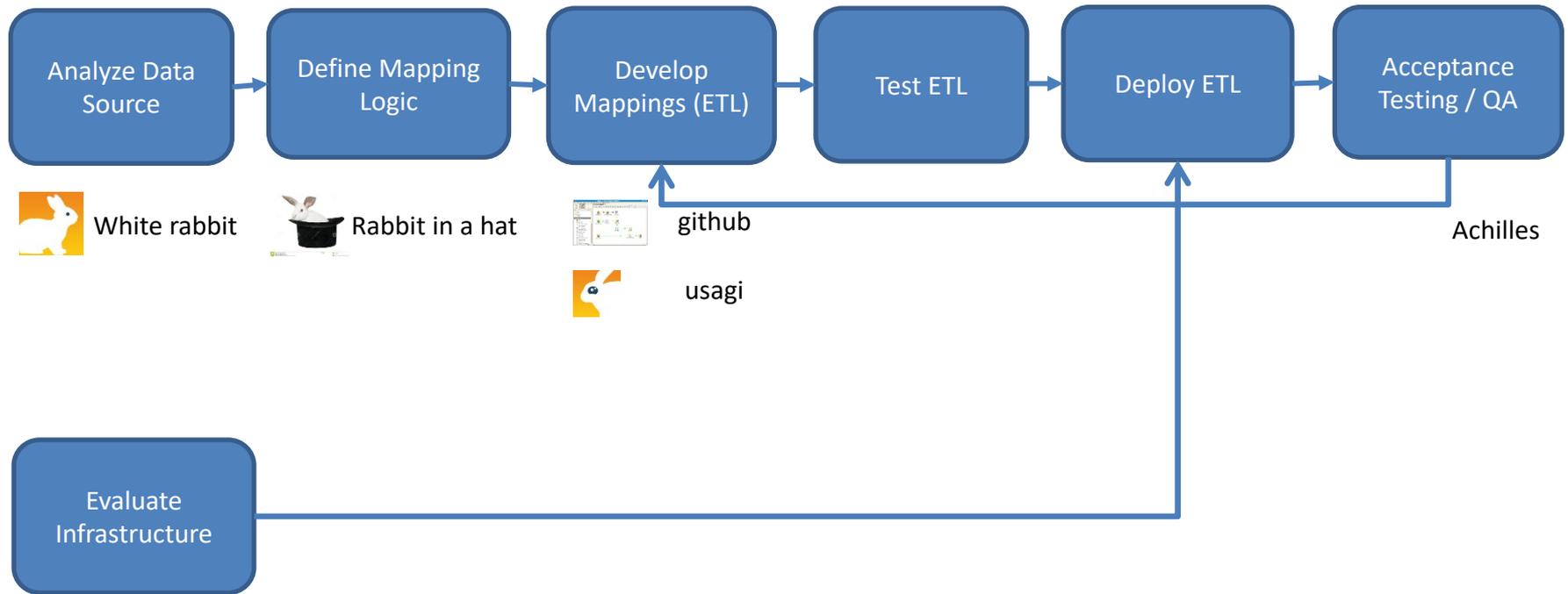


Local

EMIF



OHDSI tools supporting the process





Assessment of the conversion of ten European Databases to the OMOP CDM and evaluation of the use of OHDSI tools



Michel van Speybroeck¹, Lars Halvorsen¹, Myriam Alexander, PhD¹³, Glen James, PhD¹³, Lara Tramontan, PhD^{3,4}, Leonardo Méndez-Boo, MD MPH^{5,6}, Rients van Wijngaarden, MSc⁷, Rosa Gini, PhD⁸, Miguel A. Mayer PhD⁹, Lars Pedersen, PhD¹⁰, Alessandro Pasqua MSc¹, Sulev Reisberg MSc¹², Johan van der Lei, PhD², Peter R. Rijnbeek, PhD²

¹Janssen Pharmaceutica NV, Beerse, Belgium; ²Erasmus MC, Rotterdam, The Netherlands; ³Arsenal.IT, TV, Italy; ⁴SoSePe, PD, Italy; ⁵Direcció de Sistemes d'Informació, Institut Català de la Salut, Spain; ⁶Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Barcelona, Spain; ⁷STIZON, Utrecht, The Netherlands; ⁸Agenzia regionale di sanità della Toscana, Florence, Italy; ⁹Research Programme on Biomedical Informatics (IMIM-UPF), Barcelona, Spain; ¹⁰Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus N, Denmark; ¹¹Genomedics Srl, Florence, Italy; ¹²University of Tartu, Estonia; ¹³GSK, Uxbridge, United Kingdom

Background

The European Medical Information Framework (EMIF) aims to develop a sustainable platform for the (re)use of real world data sources, covering a wide variety of sources: regional healthcare systems, hospital data, primary care data and biobanks. The harmonization of data sources towards the OMOP CDM and the use of OHDSI tools are an important constituent of the EMIF platform. The population data sources that are part of the EMIF initiative are shown in Table 1.

Table 1: The 10 European databases that are part of the EMIF initiative and that are mapped to the OMOP CDM.

Database	Country / Region	Population Size	Type	Status
Agenzia regionale di sanità della Toscana (ARS)	Italy / Tuscany	5 10 ⁶	Administrative database of Tuscan population	Completed
Aarhus University Hospital Database	Denmark / Northern Region	2.3 10 ⁶	Administrative database of Central and North Jutland	Completed
Health Search IMS Health LPD	Italy	1.6 10 ⁶	Primary care data of GP's using the Health Search System	Completed
Integrated Primary Care Information (IPCI)	Netherlands	2.8 10 ⁶	Primary care database	Completed
Pedianet	Italy	0.4 10 ⁶	Pediatric database	In Progress
Pharmo	Netherlands	8.4 10 ⁶	Primary care database	Completed for cohort
Information System of Parc de Salut Mar (IMASIS)	Spain	1.4 10 ⁶	Hospital database	In Progress
The Information System for the Development of Research in Primary Care (SIDIAP)	Spain / Cataluna	6.4 10 ⁶	Primary care database	In Progress
The Health Informatics Network (THIN)	United Kingdom	12 10 ⁶	Primary care database	Completed
Estonian Genome Center at the University of Tartu	Estonia	52 10 ³	Biobank	Completed

Methods

Mapping to the OMOP CDM

The mapping to the OMOP CDM was based on the best practices as developed by the OHDSI community. Different technologies for the ETL (Java-jCDMBuilder / SQL / Kettle / Python) were used – depending on the party who developed the ETL and / or the technology that was acceptable for the data source

Assessment of the mapping

Following the mapping of the databases, there is a need to understand the overall 'quality' of the mappings and to assess the readiness of the mapped databases to support research questions. The process that is followed is illustrated below.

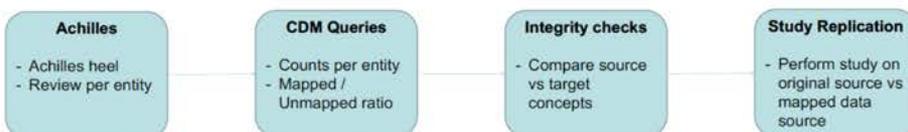


Figure 1: Proposed Mapping Assessment Flow for the 10 European Data Sources

Evaluation of Achilles

The standalone version of Achilles (version 1.3) was reviewed by 26 users, covering researchers as well as database owners through a structured assessment.

Results

Mapping to the OMOP CDM

Based on the experience in working with 10 data sources, the following factors were found to be most impactful on overall speed and quality of the mapping:

1. Source Database research readiness: The 'quality' of the input data structure – and the availability of internal knowledge on how the database is defined- are the primary driver of efficiency and quality of the CDM Mapping
2. Strong project management: superior results in terms of quality and speed can be achieved when resources are allocated and active project management is executed.
3. Vocabulary mappings: establishing the vocabulary mappings is the most resource intensive step. It's recommended to set realistic goals with associated timings (e.g. map the top 20% of lab tests, covering 80% of all occurrences)

Assessment of the mapping

Table 2, shows an overview of the drug level mappings. All data sources have a link of their drug coding system to ATC. Where available, a more granular mapping to clinical drug or form was performed. Unmapped category indicates records where no standard drug code could be found. Aside from missing concepts, this can also be attributed to the fact that the source tables can have a broader scope e.g. different OTC products.

Table 2: Drug level mapping. % based on record count

Data Source	Ingredient	Clinical Drug Comp	Clinical Drug Form	Quant Clinical Drug	Clinical Drug	Unmapped
ARS	80.5%	0.0%	0.0%	0.0%	0.0%	19.5%
AUH	5.5%	10.6%	12.0%	0.0%	0.0%	72.0%
IPCI	34.9%	3.5%	0.9%	0.0%	0.0%	56.3%
GENOMEDICS	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%
PHARMO	14.0%	7.6%	3.6%	0.0%	0.0%	74.7%
PEDIANET	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%
THIN	7.8%	0.0%	0.0%	1.5%	0.0%	69.7%
EGCUT	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Evaluation of Achilles

User experience was generally very positive with 66% qualifying it as good or excellent and 31% as OK and 4% as poor. Additional features of interest included the possibility to see the frequency distribution per person of a particular entity and the ability to search using local vocabularies. This has now been implemented in Atlas. The full report is available at <http://forums.ohdsi.org/t/emif-evaluation-of-achilles/1964>

Conclusions

The Achilles tool was well appreciated by our user group and suggestions to improve this tool have been made and implemented. Our work on the conversion of European databases to the OMOP-CDM showed that it is feasible but requires detailed quality assessments. Extensions of the Standardized Vocabularies are needed to capture all the European data adequately. This work is ongoing. The conversion of the databases will be further assessed and improvements will be proposed in the upcoming period.



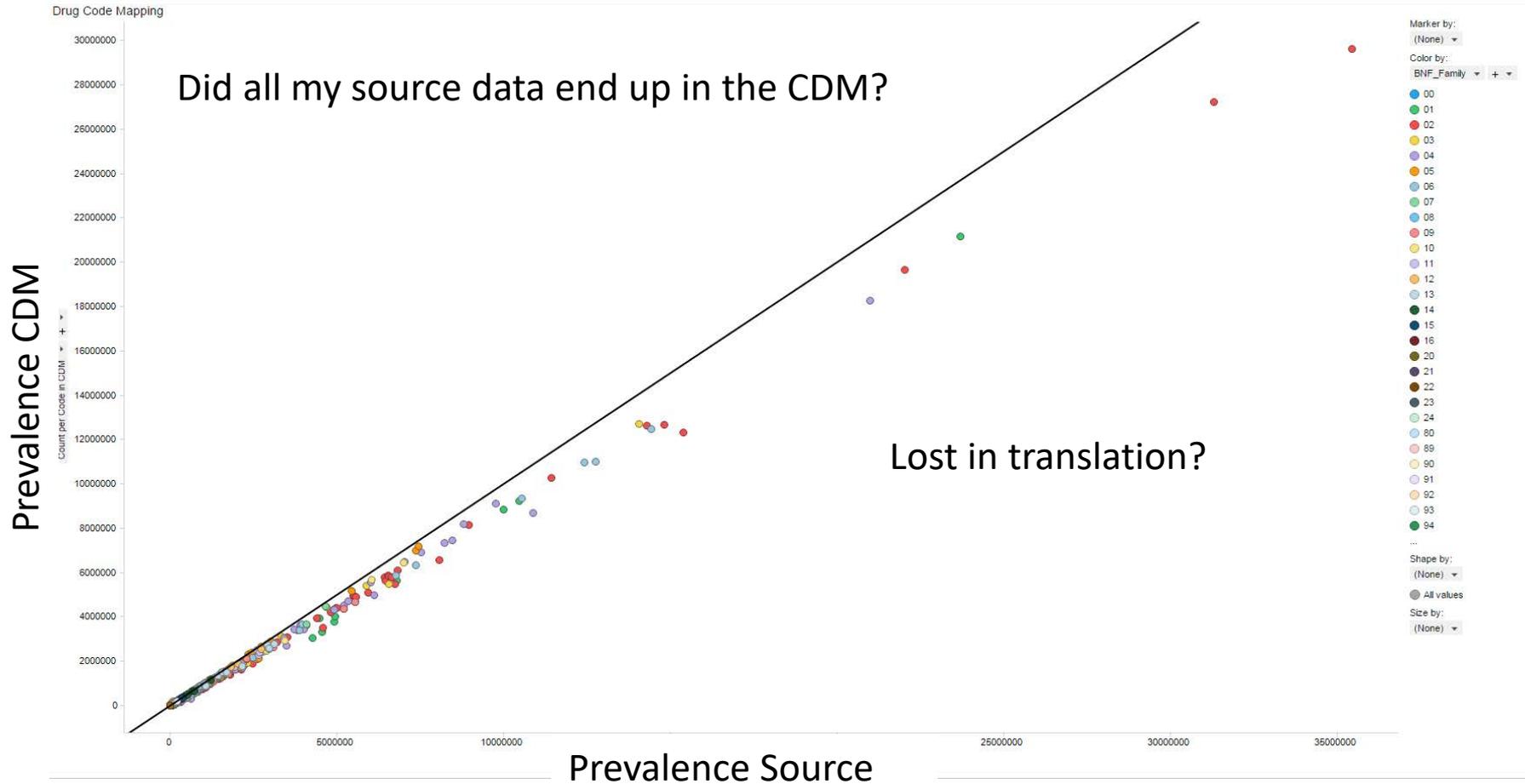
Current Challenges: ETL

The following factors were found to be most impactful on overall speed and quality of the ETL:

- 1. Source Database research readiness:** The 'quality' of the input data structure – and the availability of internal knowledge on how the database is defined- are the primary driver of efficiency and quality of the CDM Mapping.
- 2. Strong project management:** superior results in terms of quality and speed can be achieved when resources are allocated and active project management is executed.
- 3. Vocabulary mappings:** establishing the vocabulary mappings is the most resource intensive step. It's recommended to set realistic goals with associated timings (e.g. map the top 20% of lab tests, covering 80% of all occurrences).



Evaluation of translation: Structural Mapping

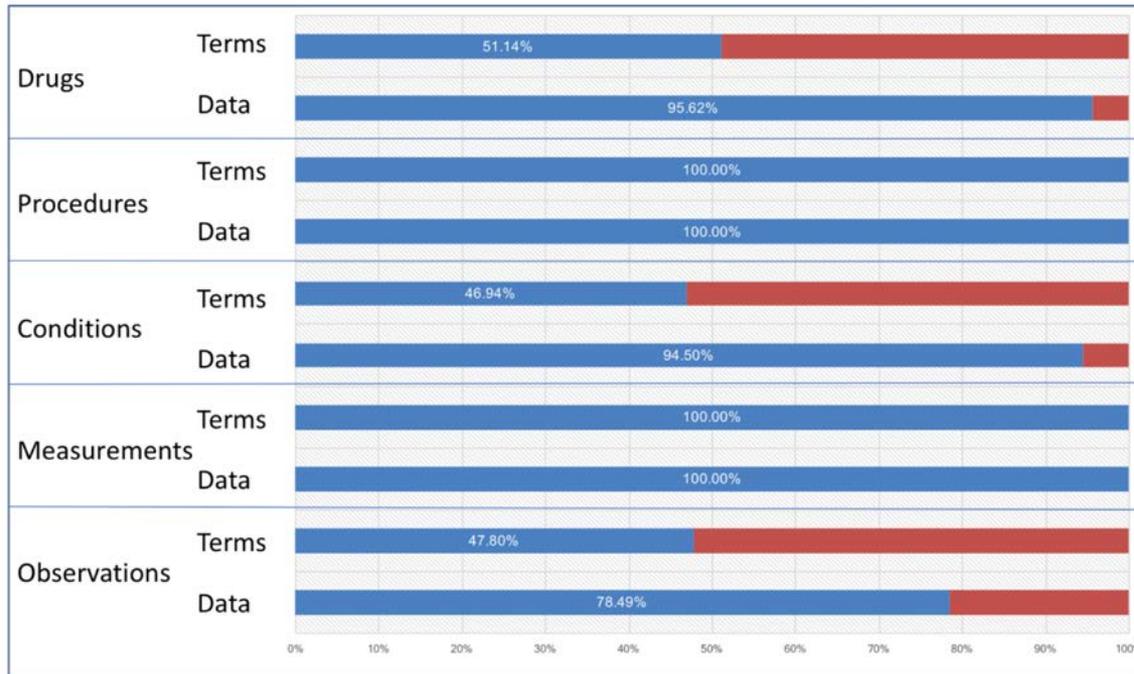


- Can be very good reason for differences: business rules assessment
- Iterative process to optimize the ETL
- No structural CDM limitations encountered so far



Evaluation of translation: Vocabulary Mapping

IPCI Database Example



- High data coverage.
- Term coverage is further improved by extending the Standard Vocabularies, e.g. RxNorm-Extension to accommodate European Drug market



Rationale for the European OHDSI Initiative

- 1) The need for data standardization through a common data model is recognized by many stakeholders, including regulators
- 2) Many projects are being initiated in Europe in which the OMOP-CDM is adopted
- 3) There is a large interest across Europe to learn more about the OMOP-CDM and OHDSI tools



We are in a perfect storm!!



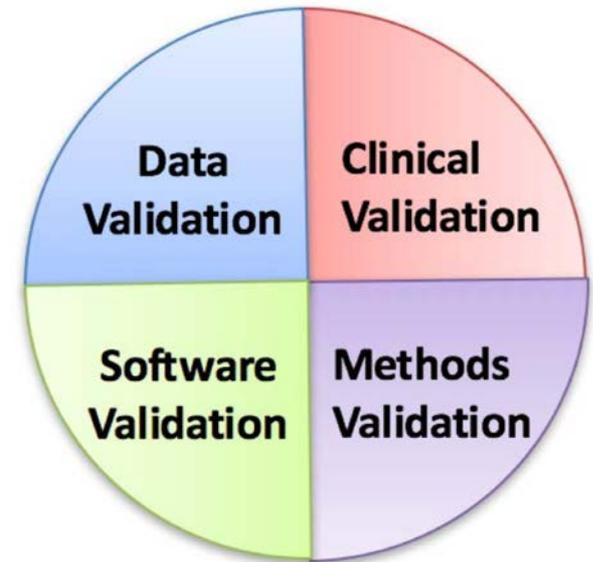
But...

- We should not all invent the same wheel but collaboratively perform open science
- An active European Community will be able to further improve CDM structure, and will improve vocabulary content to match our European needs



What do we need?

- Implement quality framework for full analytical pipeline to build trust
- Establish ETL conventions (business rules) to further improve the interoperability
- Create a European eco-system to support mapping, including certified SMEs
- Further extend the data network





Goddess of Divine and order

THEMIS

Goal of the Themis workgroup is to define business rules for ETL to further improve the inter-operability of the data.

For example:

- Patients with multiple values for sex, gender, race
- Potentially contradictory relationship between days supply, quantity, drug exposure end date and sig in Drug Exposure
- Duplicate procedures or visits at the same day
- Medical events after death date.
- Observation Period definitions for Claims and EHR records
- etc

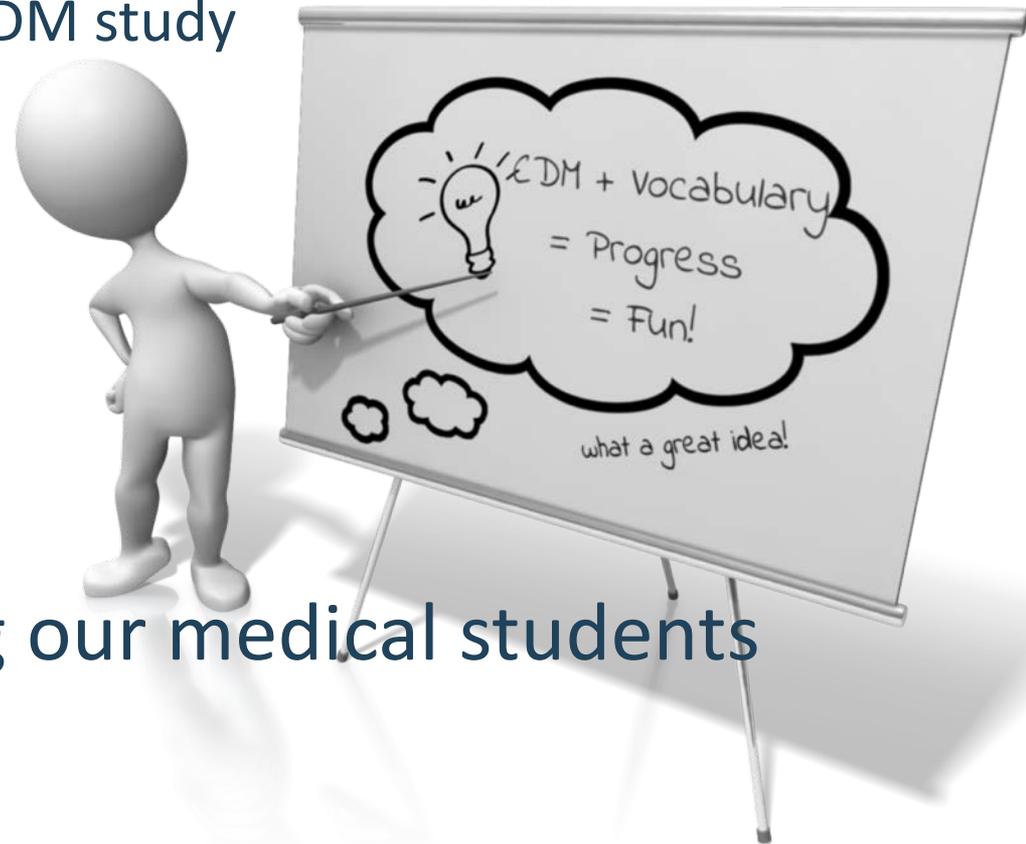


March 8th face-to-face
Themis meeting at
Amgen, U.S.



Training a crucial prerequisite for success!

- There is a lot of miss-understanding about the OMOP-CDM
 - A protocol-based versus CDM study
 - Loss of source data
 - Etc.
- We need to train all stakeholders including our medical students





Aims OHDSI-EUROPE

- Stimulate Community Building in Europe
 - Stimulate Adoption of the CDM
 - Create a platform for sharing ideas and results
 - Close collaboration with all stakeholders
 - Close collaboration with other chapters
 - Training of all stakeholders
-



Join the Journey

Our ambitions are high but achievable if all disciplines collaborate and share ideas!

We need all of you to keep the OHDSI train on track and running at speed in Europe!

