From EMIF to EHDEN: scaling up the big data ecosystem across Europe

Nigel Hughes
EHDEN Project Leader/Co-Coordinator
Scientific Director, JCI Patient Data for Research, Janssen
All too often real world research is a challenging journey....
IT’S GOING TO TAKE ALL SORTS OF DATA....

Probabilistic linkage to obtain new types of data

Examples of biomedical data
- Pharmacy data
- Health care center (electronic health record) data
- Claims data
- Registry or clinical trial data
- Data outside of health care system

Ability to link data to an individual
- Easier to link to individuals
- Harder to link to individuals
- Only aggregate data exists

Data quantity
- More
- Less

We need fully transparent and reproducible pipelines that enable large-scale federated analyses across Europe.
Why is this not current practice?

Analytical method

The data...

What will it require?

- Data interoperability
- Standardised analytics
- Data network
- Strong community
EHDEN will build on expertise and tools from prior IMI projects, such as EMIF, and will collaborate intensively with the global OHDSI community.
WHAT WE NEED IS INFRASTRUCTURE AND A NETWORK TO CONDUCT REAL WORLD RESEARCH IN THE 21ST CENTURY....
‘We have more pilots than an airline….’

Anon, Conference Participant, 2019
**LARGE-SCALE OBSERVATIONAL RESEARCH IS FEASIBLE**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Data Sources</th>
<th>Countries</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2 Diabetes Mellitus</td>
<td>11</td>
<td>&gt; 250 million</td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“Characterizing treatment pathways at scale using the OHDSI network.”
George Hripcsak et al. - PNAS (2016)27:7329–7336
Uncertainty remains about the optimal monotherapy for hypertension, and RCT data is lacking.

**LEGEND-HTN** — a systematic, large-scale study under a new-user cohort design; 3 primary outcomes (AMI, Heart Failure hospitalisation, CVA), 6 secondary outcomes & 46 safety outcomes

1st first line role of thiazide or thiazide-like diuretics under appreciated in current guidelines

Estimated 1.3 CVS events avoided for every 1000 patients THZ vs ACEi

"Comprehensive comparative effectiveness and safety of 1st line antihypertensive drug classes."
Carlos Moedas
EU commissioner for research, science and innovation

“Radical collaboration where multinational companies work together and share data instead of keeping it secret is helping to change the model of the pharmaceutical industry and solve problems more quickly.” (Source)

Accelerate research and development.

Speed up patient access to innovative treatments.

Improve patient outcomes and safety of medicines.
IMI EU Real World Project Timeline 2011 - 2025

Key European Infrastructure Projects
- EMIF
  - 2013-2018
- EHR4CR
  - 2011-2016
- EHDEN
  - 2018-2023

Future/Proposed Projects
- EU-PEARL
  - 2019-2023
- Patient Generated Data (H2O)
  - 2020-2025?

Foundational Projects
- EHR4CR
  - 2011-2016
- EMIF
  - 2013-2018
- EHDEN
  - 2018-2023
- DO-IT
- HARMONY
- ROADMAP
- BIGDATA@HEART
- PIONEER
- EHDEN
- ...
EMIF CONTRIBUTING TO SUSTAINABLE ELEMENTS IN OTHER PROJECTS

- EHDEN will utilise the technology framework from IMI1 EMIF, with key partners in both projects
- Innovation will continue to scale the architecture, develop new tools and incorporate the OMOP CDM and OHDSI tools more extensively
- Open science, open source technology is a main driver for IMI projects, inclusive of EHDEN

Elements of the EMIF EHR Platform will be used for the EHDEN architecture

EMIF-AD will be a spin-out from EMIF in 2019, supporting the AD research community
Vision

The European Health Data & Evidence Network (EHDEN) aspires to be the trusted observational research ecosystem to enable better health decisions, outcomes and care.

Mission

Our mission is to provide a new paradigm for the discovery and analysis of health data in Europe, by building a large-scale, federated network of data sources standardized to a common data model.
**EHDEN CONSORTIUM**

**Start date:** 1 Nov 2018  
**End date:** 30 Apr 2024  
**Duration:** 66 months

22 partners

Almost €29 million

Universities, public bodies and research organisations

- Erasmus MC
- Uppsala Monitoring Centre
- universidade de Aveiro
- NICE
- Karolinska Institutet
- Nuffield Trust
- University of Oxford
- ERF (European Research Framework)

**SME & Mid-sized companies**

- SYNAPSE
- ODYSSEUS Data Services Inc

**Non-profit organisations**

- thehyve

**EFPIA & Associated partners**

- EFPIA Lead
- Bayer
- Celgene
- Pfizer
- Lilly
- Abbvie
- Sanofi
- UCB
- Novartis
- Servier
**Why is EHDEN needed?**

**Regulators**
- Continuous effectiveness & safety assessment

**Payers**
- Assess cost effectiveness
- Determine reimbursement

**Patients**
- Understanding benefits and risks
- Improvement of care
- Better patient outcomes

**Research**
- Understanding diseases
- New indications

**Development**
- Trial design and feasibility
- Demonstrating value
EHDEN IS ABOUT ...

**Federation**
Creation of an EU-wide architecture for federated analyses of real world data

**Harmonisation**
Harmonise more than 100 million anonymised health records to the OMOP common data model

**Community**
Establish a self-sustaining open science collaboration in Europe, supporting academia, industry, regulators, payers, government, NGOs and others

**Outcomes**
Enabling outcomes driven healthcare at a European level

**Education**
The establishment of an EHDEN Academy, webinars and face-to-face training sessions to train all stakeholders
**FOUNDATIONAL PILLARS**

**Infrastructure**
- Creation of an EU-wide federated network architecture
- **Privacy** by design
- **Data harmonisation** to the OMOP common data model
- Training & certification of SMEs

**Research & Outcomes**
- **Use cases** to evaluate the EHDEN federated network
- Collaboration on consistent methodologies
- Collaboration with the global OHDSI research network
- Incorporation of the ICHOM health outcome standards

**Education & Community**
- Establishment of an EHDEN Academy
- Expansion of the OHDSI network in Europe
- Collaboration on collective memory for research use cases
Key Components for EHDEH

Build
Fuel
Drive
Technology
Engagement
Outreach
Outcomes
<table>
<thead>
<tr>
<th>WP1: Evidence Workflow Development</th>
<th>WP2: Outcome Driven Healthcare</th>
<th>WP3: Personalized Medicine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorporating the use cases for supporting development and validation of the EHDEN socio-technical approach, inclusive of BD4BO projects</td>
<td>Related to all activities specific to e.g. BD4BO projects outcome focus, and ICHOM standards incorporation</td>
<td>Focusing on the support of outcomes/value based healthcare, inclusive of clinical prediction models, with the incorporation of ‘novel’ patient data</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WP4: Technical Implementation</th>
<th>WP5: Data Workflow Implementation &amp; Service Deployment</th>
<th>WP6: Outreach and Sustainability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key priority is socio-technical development of the EHDEN federated framework and relevant services</td>
<td>Development, oversight and evaluation of the ecosystem development from SME qualification/certification to data source engagement, OMOP CDM mapping and evaluation</td>
<td>Ensuring the development of value propositions for key stakeholders, and developing the sustainable operational model for EHDEN during and post IMI phase</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WP7: Project Management and Dissemination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentrating on intra-project project management, internal communications and external dissemination, and responding to IMI deliverables</td>
</tr>
</tbody>
</table>
THE EHDEN FEDERATED DATA NETWORK

Local Governance

EMR  
LIMS  
Rx  
Dx  
Admin ...

Local Database  
OMOP Database

Analysis query

Aggregated results

The EHDEN platform

EMHDEN will develop new tools and dashboards.

Many different open source tools (cohort builder, estimation, incidence rate, ....)

ATLAS
THE OMOP COMMON DATA MODEL

Patient-centric
Tabular
Extendable
Built for analytics
Relational design

Standardized clinical data
- Person
  - Observation period
  - Specimen
  - Death
  - Visit occurrence
    - Procedure occurrence
    - Drug exposure
    - Device exposure
    - Condition occurrence
    - Measurement
    - Observation
    - Note
    - Note NLP
    - Fact relationship

Standardized health system data
- Location
- Care site
- Provider
- Payer plan period
- Cost
- Cohort
- Cohort attribute
- Condition era
- Drug era
- Dose era

Standardized meta-data
- CDM source
- Concept
- Vocabulary
- Domain
- Concept class
- Concept relationship
- Relationship
- Concept synonym
- Concept ancestor
- Source-to-concept map
- Drug strength
- Cohort definition
- Attribute definition

Standardized derived elements
- Standardized vocabularies
  - Concept
  - Vocabulary
  - Domain
  - Concept class
  - Concept relationship
  - Relationship
  - Concept synonym
  - Concept ancestor
  - Source-to-concept map
  - Drug strength
  - Cohort definition
  - Attribute definition

Standardized health economies
- Care site
- Provider
- Location
- Payer plan period
- Cost
- Cohort
- Cohort attribute
- Condition era
- Drug era
- Dose era

v 5.0.1
Main developers of many of the OHDSI tools are EHDEN partners

Collaborators
EHR Catalogue
OHDSI

Findable
White Rabbit
Rabbit In A Hat
Usagi
Jerboa
ARACHNE

Accessible
Ethical code of practice

Interoperable
OMOP cdm

Reusable
Private remote research environment
ATLAS
CALL PROCESS OVERVIEW

Data sources
- Tailored for project objectives and sustainability

Open calls
- Focusing on SMEs able to support mapping and sustainability

Grant awarding
- Evaluated via a pre-defined set of criteria by the Data source prioritisation committee

Data sources can choose the SME from the pool of EHDEN certified SMEs

SMEs are paid via grants from the harmonisation fund

Payments are milestone based

Supporting SMEs

Open calls
- Tailored for project objectives and sustainability

Grant awarding
- Evaluated via a pre-defined set of criteria by the Data source prioritisation committee

Data sources can choose the SME from the pool of EHDEN certified SMEs

SMEs are paid via grants from the harmonisation fund

Payments are milestone based

Harmonisation fund

Workshop

Source Data Evaluation

Mapping Cycle

Audit

Mapping

Share of Mapping Process

Mapped data sources are encouraged to be active members of the EHDEN community, participating in research studies.
**Open Call**
Submission of applications for the open call for SMEs from the 1st of April until the 1st of May (17h00) via the EHDEN website.

**Evaluation**
Following an eligibility check, applications were evaluated by the SME certification committee.

**Training & Certification**
Certification and training of selected SMEs in all necessary competencies.

- 34 SME profiles made
- 28 Eligible applications
- 11 SMEs initially selected
Certified SMEs:

- edenceHealth NV
- Easter-eggs
- Biomeris
- Clinerion Ltd
- B2i Healthcare
- ITTM S.A.
- MediSapiens Ltd
- P-95
- Imosphere
- Quretec OÜ
- Stacc OÜ
48 Data partner profiles made
29 Eligible applications
20 Selected Data Partners
2020 Call Timelines

**SME**
The second open call for SMEs.

**SME**
The third open call for SMEs.

**Data partner**
The second open call for Data partners.

**Data partner**
The third open call for Data partners.
The success of EHDEN will only be based on trust and relevance.

Openness, transparency and collaboration will be key drivers for trust.

Relevance needs to be based on relative usefulness of data, research outputs and accessible technology.

Ultimately, trust and relevance are based on successful outputs.
EHDEN will ensure compliance with e.g. GDPR

- Compliance with citizen and data protection regulations
- Further development of the EMIF Ethical Code of Practice
- Ethical Advisory Board
- Privacy by design (data remains local)
- Evolving framework for data protection
Use Case 1 – Drug utilisation study
Validate the OMOP CDM in European data sources comparing results from source versus OMOP data

Use Case 2 – Drug and device safety study
Test and further develop existing (OHDSI) population-level estimation and population-level prediction packages in EU Data sources

Use Case 3 – HTA study
Assess whether data commonly used for HTA purposes can easily be measured and analysed using the OMOP CDM
“To compare the risk of post-operative complications and mortality between unicompartmental vs total knee replacement.”

Monday
- Group consensus on the problem
- Draft cohort definitions

Tuesday
- Review clinical characterisation
- Draft patient-level prediction design

Wednesday
- Review patient-level prediction results
-Externally validate prediction model

Thursday
- Draft population-level effect estimation design
- Review population-level effect estimation diagnostics

Friday
- Review of results
- Plan for completing publications

(EHDEN 1st Study-a-thon, Oxford, December 2018)
Aim

To develop an e-learning environment to train all stakeholders on the use of the tools and processes that are being adopted within EHDEN and OHDSI.

Collaboration

Course development on the OMOP Common Data Model and the rich set of OHDSI tools that are being developed in collaboration with the OHDSI community.

Infrastructure

The EHDEN Academy is developed in Moodle and is hosted in the Amazon AWS cloud. We use virtual machines for assignments.
**The Development & Milestones of the EHDEN Project**

**IMI Duration**
Nov 2018 – Apr 2024

**OHDSI**
Symbiotic Relationship/Global Linkage

**EHDEN needs to be sustainable**
- European project supported by IMI
- Building research infrastructure for coming decades
- Interoperability & standardisation
- OHDSI framework/Open Science
- Collective ownership

**Inception**
Creation of IMI Project

**Pilot Year**
Initial:
- SME & Data Calls
- Research Use Cases
- Technology Architecture
- Stakeholder Engagement

**Scale Up**
Expansion:
- SME & Data Calls
- ETL/Mapping (OMOP CDM)
- Research Use Cases
- Technology Architecture
- Community Building
- Evidence Generation

**Sustainability**

Scalability
- Open Science Community
- OHDSI Framework & Network Evolution
- Research Use Cases
- Pharmaco-surveillance
- Evidence Generation
- Discovery & R&D
COVID-19, OHDSI & EHDEN and the Need for Data & Insights
COVID-19 Patient trajectory

Data elements captured

Medical history:
- Demographics
- Conditions
- Drugs
- Health service utilization

Plus...
- ‘recent’ health behavior measurement

Characterization:
- Medical history and presenting symptoms amongst patients tested for COVID-19
- Population-level Estimation:
  - Does exposure increase the risk of incidence of COVID-related symptoms?
- Patient-level Prediction:
  - Amongst all patients, who received COVID-19 test?
  - Amongst patients with flu-like symptoms, who received COVID-19 test?

Characterization:
- Medical history and presenting symptoms amongst patients tested positive for COVID-19
- Population-level Estimation:
  - Does prior drug exposure increase risk of COVID-19 incidence?
    (ACE susceptibility question could be answered, but substantial bias exists if testing isn’t complete)
- Patient-level Prediction:
  - Amongst patients tested for COVID-19, who tested positive?

Characterization:
- Medical history amongst patients hospitalized for COVID-19
- Treatment utilization among patients with COVID-19
- Population-level Estimation:
  - Does prior drug exposure increase risk of COVID-19 hospitalization? (ACE)
- Patient-level Prediction:
  - Amongst patients with COVID-19, who requires hospitalization?
  - Amongst patients hospitalized with COVID-19, who requires intensive services?

Characterization:
- Outcomes for patients with COVID-19
- Outcomes for patients hospitalized for COVID-19
- Population-level Estimation:
  - Comparative effects of interventions on COVID-19
  - Does prior drug exposure increase risk of COVID-19 severity?
- Patient-level Prediction:
  - Amongst patients with COVID-19, who die?

COVID-related questions that can be potentially answered:

Present with symptoms
Tested for COVID-19
Result obtained for COVID-19 test
Hospitalization
Hospitalization with intensive services
Death

*Note: testing may take place anytime before symptoms through after hospitalization, or may not occur at all in COVID patients

Follow-up observation period

End of 30-d risk window
End of 90-d risk window
End of 365-d risk window

Plus...
- inpatient services

Plus...
- mortality

*COVID-19*
EHDEN: 21ST CENTURY NETWORK FOR 21ST CENTURY REAL WORLD RESEARCH

• EHDEN is a bold step in recognising that a flagship project is needed to address conducting real world research for the 21st Century

• At its heart is the acknowledgement that we need to develop a community via a federated network, within an ecosystem, all based on a quid pro quo around data for research use

• The Open Science/Open Source community of OHDSI, and the OMOP CDM are critical enablers and partners for EHDEN in this endeavour

• Watch out for #DataSavesLives....
• EHDEN Coordinator: Assoc Prof Peter Rijnbeek, Erasmus Medical Centre: p.rijnbeek@erasmusmc.nl

• EHDEN Coordinator: Nigel Hughes, Janssen: nhughes@its.jnj.com

• EHDEN project website: https://www.ehden.eu
• IMI website: https://www.imi.europa.eu
• IMI2 BD4B0 programme: https://bd4bo.eu
• OHDSI website: https://ohdsi.org
• List of converted (‘OMOPd’) databases: http://www.ohdsi.org/web/wiki/doku.php?id=resources:2018_data_network
• OHDSI Europe website: https://www.ohdsi-Europe.org/
This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union’s Horizon 2020 research and innovation programme and EFPIA.