Enabling Large-Scale Analysis of Electronic Health Records in Europe: The EHDEN Project

Peter Rijnbeek $^1$, Nigel Hughes $^2$

$^1$Department of Medical Informatics
Erasmus MC Rotterdam
The Netherlands

$^2$Epidemiology,
Janssen R&D
Belgium
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.
THE CHALLENGES OF REAL WORLD DATA

What will it require?

Data interoperability  Standardised analytics  Data network  Strong community
A Common Data Model will enable standardised analytics to generate reliable evidence.
THE 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

SME & Mid-sized companies

Non-profit organisations

EFPIA & Associated partners

Academic coordinator

EFPIA Lead
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
## WORK PACKAGE STRUCTURE

### WP1: Evidence Workflow Development
Incorporating the use cases for supporting development and validation of the EHDEN socio-technical approach, inclusive of BD4BO projects

### WP2: Outcome Driven Healthcare
Related to all activities specific to e.g. BD4BO projects outcome focus, and ICHOM standards incorporation

### WP3: Personalized Medicine
Focusing on the support of outcomes/value based healthcare, inclusive of clinical prediction models, with the incorporation of ‘novel’ patient data

### WP4: Technical Implementation
Key priority is socio-technical development of the EHDEN federated framework and relevant services

### WP5: Data Workflow Implementation & Service Deployment
Development, oversight and evaluation of the ecosystem development from SME qualification/certification to data source engagement, OMOP CDM mapping and evaluation

### WP6: Outreach and Sustainability
Ensuring the development of value propositions for key stakeholders, and developing the sustainable operational model for EHDEN during and post IMI phase

### WP7: Project Management and Dissemination
Concentrating on intra-project project management, internal communications and external dissemination, and responding to IMI deliverables
**EHDEN: WHERE ARE WE NOW AFTER 26 MONTHS?**

### Data Partners
- 1st General: 18 Partners
- 2nd COVID-19: 25 Partners
- 3rd General: 26 Partners
- ~200mio records
- 16 countries
- Mapping work ongoing

### Pillar 1
- 26 SMEs certified by September 2020
- Data Partner Catalogue created
- ETL tooling
- Quality dashboards
- Security framework
- SME Directory

### Pillar 2
- DUS
- HTA
- Drug/Device Safety
- Methods Research
- Protocols developed
- Initial work with prior CDM-mapped datasets
- Initial work on ICHOM standards incorporation
- Study-a-thons

### Pillar 3
- Value proposition development
- Stakeholder engagement
- Sustainability planning
- EHDEN Academy launched
IMPROVING INTEROPERABILITY OF DATA: THE EHDEN CALLS

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

- SME certification committee prioritizes SMEs for training and certification

Training & Certification

- SMEs are paid via grants from the harmonisation fund

Workshop

- Mapped data sources are encouraged to be active members of the EHDEN community, participating in research studies.

Harmonisation fund

- 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

- Mapped data sources are encouraged to be active members of the EHDEN community, participating in research studies.
NEED FOR A SUPPORT NETWORK: OPEN CALLS FOR SMEs

1st SME Call (n=11)
2nd SME Call (n=15)
14 countries

New SME call opens in March, watch our website and Social Media Channels
Small to Medium-sized Enterprises (SMEs) have created business opportunities beyond the EHDEN Data Partner Calls!
ENABLING DATA PARTNERS TO JOIN THE NETWORK

• Three calls have been finalized, including one COVID-19 Call. Two calls per year are foreseen in the project for a total amount of 17 Million Euro financial support. For more information about the Open Calls see the EHDEN website: www.ehden.eu

<table>
<thead>
<tr>
<th>Activity</th>
<th>Max. amount per complexity level in Euro</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>A. Create new Data Transformation and Analytical Infrastructure</td>
<td>40.000</td>
</tr>
<tr>
<td>B. Revise Existing Data Transformation and Analytical Infrastructure</td>
<td>30.000</td>
</tr>
<tr>
<td>C. Inspect Completed Data Transformation and Analytical Infrastructure</td>
<td>10.000</td>
</tr>
</tbody>
</table>

• Data Partners from EU Member States and H2020 countries can apply through online application portal.

• New data partner call opens in April! Watch our website and Social Media Channels
AFTER THREE CALLS WE SELECTED 60 DATA PARTNERS IN 16 COUNTRIES

Geographic spread of data partners. The shade of blue indicates the # of data partners in that country (darker = more)
Taskforce to help data partners with data relevant to COVID to get on the OMOP-CDM

Taskforce contains EHDEN consortium partners that have extensive expertise in OMOP ETL, as well as certified SMEs
Huge opportunity to learn about our processes for ETL and quality control.

Interesting to see the diversity in competences, data readiness, complexity of data, infrastructure needs etc.
Infrastructure and Tool development
• Tool integration in single environment (Single Sign On, Authentication Mechanisms)

• Service Desk Development (Self-service portal, ticketing system, phone number)

• Continuous Security Checks, Server monitoring
CONTACT DETAILS

Database Acronym
PCI

Database Name
Integrated Primary Care Information

Institution name
Erasmus University Medical Center

Department name
Department of Medical Informatics

Street Address
Dr Molewaterplein 50

City
Rotterdam

Postal code
3000 DR

Country
Netherlands
• Based on Achilles
• Extracts all data needed for the visualisations in the data catalogue
• Can easily be extended with additional analysis
Network Dashboards
Dashboard Data Upload

Data Source: Integrated Primary Care Information
Database type: Gp
Country: Netherlands
Release date: 2020-09-20

Upload new database characteristics
For more information see the CatalogueExport R-package.

Results file
Choose File no file selected
- Upload

Upload history

<table>
<thead>
<tr>
<th>Upload Date</th>
<th>R Package Version</th>
<th>Generation Date</th>
<th>CDM Version</th>
<th>Vocabulary Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan. 6, 2021, 9:15 a.m.</td>
<td>1.0</td>
<td>20210105</td>
<td>5.3.1</td>
<td>v5.0 03-APR-20</td>
</tr>
<tr>
<td>Dec. 27, 2020, 10:20 p.m.</td>
<td>1.0</td>
<td>20201227</td>
<td>5.3.1</td>
<td>v5.0 03-APR-20</td>
</tr>
<tr>
<td>Dec. 27, 2020, 6:20 p.m.</td>
<td>1.0</td>
<td>20201227</td>
<td>5.3.1</td>
<td>v5.0 03-APR-20</td>
</tr>
<tr>
<td>Dec. 27, 2020, 10:40 a.m.</td>
<td>1.0</td>
<td>20201226</td>
<td>5.3.1</td>
<td>v5.0 03-APR-20</td>
</tr>
<tr>
<td>name</td>
<td>concept_id</td>
<td>concept_name</td>
<td>domain_id</td>
<td>drc</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------</td>
<td>---------------------------------------</td>
<td>-----------</td>
<td>--------</td>
</tr>
<tr>
<td>Integrated Primary Care Information</td>
<td>306666</td>
<td>Hypertensive disorder</td>
<td>Condition</td>
<td>5086700</td>
</tr>
<tr>
<td>Integrated Primary Care Information</td>
<td>261828</td>
<td>Type 2 diabetes mellitus</td>
<td>Condition</td>
<td>4483800</td>
</tr>
<tr>
<td>Integrated Primary Care Information</td>
<td>264761</td>
<td>Cough</td>
<td>Condition</td>
<td>2471700</td>
</tr>
<tr>
<td>Integrated Primary Care Information</td>
<td>81902</td>
<td>Urinary tract infectious disease</td>
<td>Condition</td>
<td>2283100</td>
</tr>
<tr>
<td>Integrated Primary Care Information</td>
<td>267011</td>
<td>Acute upper respiratory infection</td>
<td>Condition</td>
<td>2040300</td>
</tr>
<tr>
<td>Integrated Primary Care Information</td>
<td>4223659</td>
<td>Fatigue</td>
<td>Condition</td>
<td>1857100</td>
</tr>
<tr>
<td>Integrated Primary Care Information</td>
<td>104153</td>
<td>Low back pain</td>
<td>Condition</td>
<td>1700100</td>
</tr>
<tr>
<td>Integrated Primary Care Information</td>
<td>3170109</td>
<td>Asthma</td>
<td>Condition</td>
<td>1757000</td>
</tr>
<tr>
<td>Integrated Primary Care Information</td>
<td>4133101</td>
<td>Excessive cerumen in ear canal</td>
<td>Condition</td>
<td>1488000</td>
</tr>
<tr>
<td>Integrated Primary Care Information</td>
<td>4037322</td>
<td>C/O: a general symptom</td>
<td>Condition</td>
<td>1463000</td>
</tr>
</tbody>
</table>

By clicking on the concept_id, you can view additional information about the concept in Athena.
The concepts browser shows the total number of records (RC) and total number of records including all descendant (DRC) across the selected data sources. No exact numbers are provided but they are rounded up to the nearest multiple of 100.

## Concept Browser

<table>
<thead>
<tr>
<th>concept_name</th>
<th>acronym</th>
<th>AP-HM</th>
<th>IPCI</th>
<th>MHD</th>
<th>SIADIAP</th>
<th>UCCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>22q11.2 deletion syndrome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-methylglutaconic aciduria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46, XX true hermaphrodite</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4p partial monosomy syndrome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5p partial monosomy syndrome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABO incompatibility reaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal abscess</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal actinomycosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal aortic aneurysm without rupture</td>
<td>1.6k</td>
<td>1.6k</td>
<td>400</td>
<td>400</td>
<td>9.8k</td>
<td>9.8k</td>
</tr>
<tr>
<td>Abdominal aortic ectasia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal colic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain Filter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Concept Network Coverage table shows the total number of records (RC) and total number of records including all descendant (DRC) of a specific concept in the full data network.

By clicking on the concept_id you can view additional information about the concept in Athena.
### Concept Network Coverage

<table>
<thead>
<tr>
<th>concept_id</th>
<th>concept_name</th>
<th>domain_id</th>
<th>Databases</th>
<th>desc</th>
<th>rec</th>
</tr>
</thead>
<tbody>
<tr>
<td>260427</td>
<td>Common cold</td>
<td>Condition</td>
<td>3</td>
<td>7.63M</td>
<td>7.63M</td>
</tr>
<tr>
<td>316865</td>
<td>Hypertensive disorder</td>
<td>Condition</td>
<td>2</td>
<td>7.63M</td>
<td>5.89M</td>
</tr>
<tr>
<td>201826</td>
<td>Type 2 diabetes mellitus</td>
<td>Condition</td>
<td>4</td>
<td>5.17M</td>
<td>4.49M</td>
</tr>
<tr>
<td>254761</td>
<td>Cough</td>
<td>Condition</td>
<td>3</td>
<td>4.05M</td>
<td>4.05M</td>
</tr>
<tr>
<td>194133</td>
<td>Low back pain</td>
<td>Condition</td>
<td>4</td>
<td>4.59M</td>
<td>4.5M</td>
</tr>
<tr>
<td>81902</td>
<td>Urinary tract infectious disease</td>
<td>Condition</td>
<td>4</td>
<td>4.11M</td>
<td>3.9M</td>
</tr>
<tr>
<td>432705</td>
<td>Traumatic AND/OR non-traumatic injury</td>
<td>Condition</td>
<td>2</td>
<td>9.03M</td>
<td>3.12M</td>
</tr>
<tr>
<td>260159</td>
<td>Acute bronchitis</td>
<td>Condition</td>
<td>4</td>
<td>3.09M</td>
<td>3.06M</td>
</tr>
<tr>
<td>24660</td>
<td>Acute tonsillitis</td>
<td>Condition</td>
<td>4</td>
<td>2.95M</td>
<td>2.95M</td>
</tr>
<tr>
<td>316524</td>
<td>Hematuria - finding</td>
<td>Condition</td>
<td>1</td>
<td>2.8M</td>
<td>2.8M</td>
</tr>
</tbody>
</table>

**Disclaimer:** The presented data can be used for an initial feasibility assessment but should not be used for research without involvement of the data custodian.
• Maintenance and extension of the ETL Tools

• Tool development to support drug mapping

• Exploring the use of Usagi in a collaborative framework to distribute the mapping effort
# CdmInspection

**R Package to support quality control inspection of an OMOP-CDM instance**

## Introduction

The European Health Data and Evidence Network (EFPIA) released a multibar yearly Open Call for financial and other support.
Research Pillar
Clinical characterization: What happened to them?

Patient-level prediction: What will happen to me?

Population-level effect estimation: What are the causal effects?

observation

inference

causal inference

• Multiple Study-a-thons
• UC1 – Drug Utilisation and RMM Effectiveness
• UC2 – Regulatory / Safety
• UC3 – HTA
METHODS DEVELOPMENT: SOME EXAMPLES

• Pipeline development
  • Heterogeneity of Treatment Effect
  • Natural Language Processing
  • Disease Trajectories
  • Frequent Pattern Mining
  • Learning Curve Research

• Other topics
  • Patient Level Prediction Models Review
  • Explainable AI
Community Building and Sustainability
• Stakeholder Workshops, e.g. with HTA bodies
• EHDEN Webinars
• Collaboration with other European Projects
• Interaction with regulators
• Interaction with European Commission
• Providing educational resources
Learning for Anyone, Anywhere

On-demand training and development programs developed by the OHDSI and EHDEN community.

Tool Ecosystem
Learn to use our latest tools for observational research in your organization.

Meet Like-minded People
Share and develop your ideas with an interdisciplinary group of learners.

Join the Journey
Prepare to participate in worldwide face-to-face events such as our Study-A-Thon.

Connect with Experts
Learn from experienced instructors that help to drive the development of our tools forward.

FREE RESOURCE: ACADEMY.EHDEN.EU
Now

- Free education in observational research from the perspective of OHDSI
- Modular training along Learning Paths
- Link to the OHDSI global community
- Supports EHDEN’s SME training/certification & Data Partners programme

Opportunity

- Forum engagement and community development to extend EHDEN and OHDSI
- Engagement by potential partners and sponsors in Academy development and community
- Scientific standard communicated for observational research within OHDSI framework
- Insights generation on research community’s education and practice
- Future educational institution with clear remit and formal qualifications
As of 8th February ~800 participants enrolled with ~50% active in the last 14 days
- Major concentration in US, UK and Spain
- Focus driven by EHDEN SME Learning Pathway in Europe
  - European SME marketplace developing (sustainability)

Focus now is on content development in 2021
- Learning pathways developed from modules created by OHDA and Erasmus Medical Center
- https://academy.ehden.eu
SUSTAINABILITY: THE EHDEN INSTITUTE

Governance Structure

"ADAM"
- Disease specific research programs
- Regulatory research program
- One-off studies
- ...

"EVE"
- EHDEN Academy
- Vocabularies
- Community, Data Network
- Tools & platforms
- Certification

Management, communication, legal, etc.

Regulators / HTAs

Tenders, applications for funding

Specific RP funding

Data Partners

Structural funding

Other projects

SMEs
CLEAR OBJECTIVES IN EHDEN YEAR 3

- **Launch of EHDEN Portal**
  - Initial stages, starting with Catalogue

- **Research Use Cases with Data Partners**
  - Collaborating post-ETL on designated use cases – proof of principle of the EHDEN network

- **Promoting OHDSI & Community Education**
  - Standard adoption by key stakeholders in EU
  - EHDEN Academy expansion

- **Launch of EHDEN Institute sustainability**
  - New legal entity
  - Parallel to EHDEN project
  - Development of research operating model and future business model

- And much more…
EHDEN IS IN A DYNAMIC WORLD

• We are all cognisant of the impact of the COVID-19 pandemic

• OHDSI is evolving and expanding. Countries are adopting the OMOP CDM, e.g. Denmark, Germany as are other IMI projects (e.g. BD4BO, EU-PEARL, H20, et al)

• From a PEST analysis now:
  – Political: Instability
  – Economic: Global Shock
  – Social: Negative Trends
  – Technology: Networks Expanding

• EHDEN has grown as a thought leader in this dynamic environment

• EMA and EU Commission plans very convergent with EHDEN

• We are on track to support observational research in a sustainable, federated, open science network for the future

www.ehden.eu
@IMI_EHDEN
IMI_EHDEN
github.com/EHDEN

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.