



Empowering the OHDSI community: Multi-modal data handling and network study design with Data2Evidence



data4life

Mission-driven.

We digitalize health research to advance public health & personalized medicine.

Non-profit.

We are 100% nonprofit, always putting impact first.

Open source.

We contribute back to the open source community with key solutions.



Our co-created solutions address health researchers' specific needs

COLLECT.

Sensor Data Collection Platform

Collection of data from sensors, wearables, and questionnaires

DATA2EVIDENCE.

Health Data Management Solution

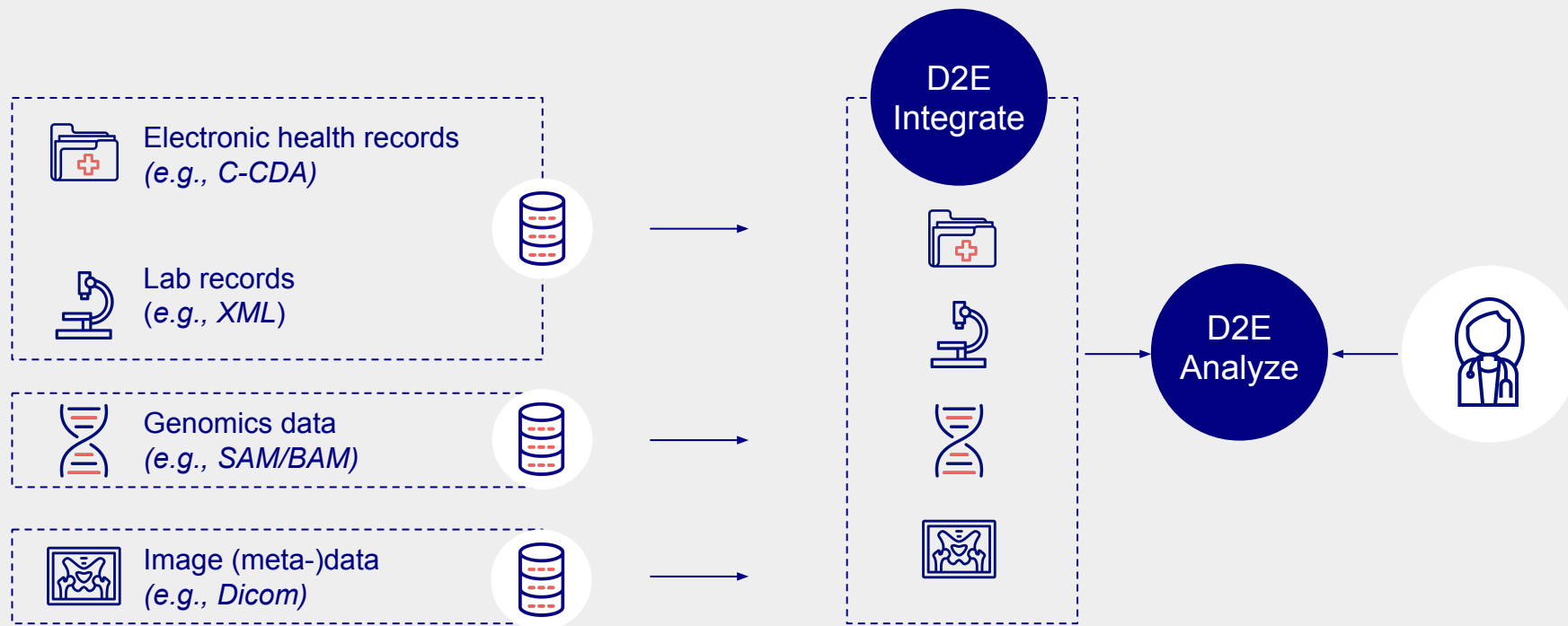
Extract, combine, and analyze data to generate evidence-based insights

FAIR.

Health Research Metadata Catalog

Easy identification of datasets suitable for secondary research

Enabling end to end access to real world data



AI-Powered multi-modal health data platform for enhanced discovery

OMOP CDM¹

~11M electronic health records

Customized OMOP extensions

~60K genomics records

~3.5M pathology records

~66M imaging records

~1.9M ECG data

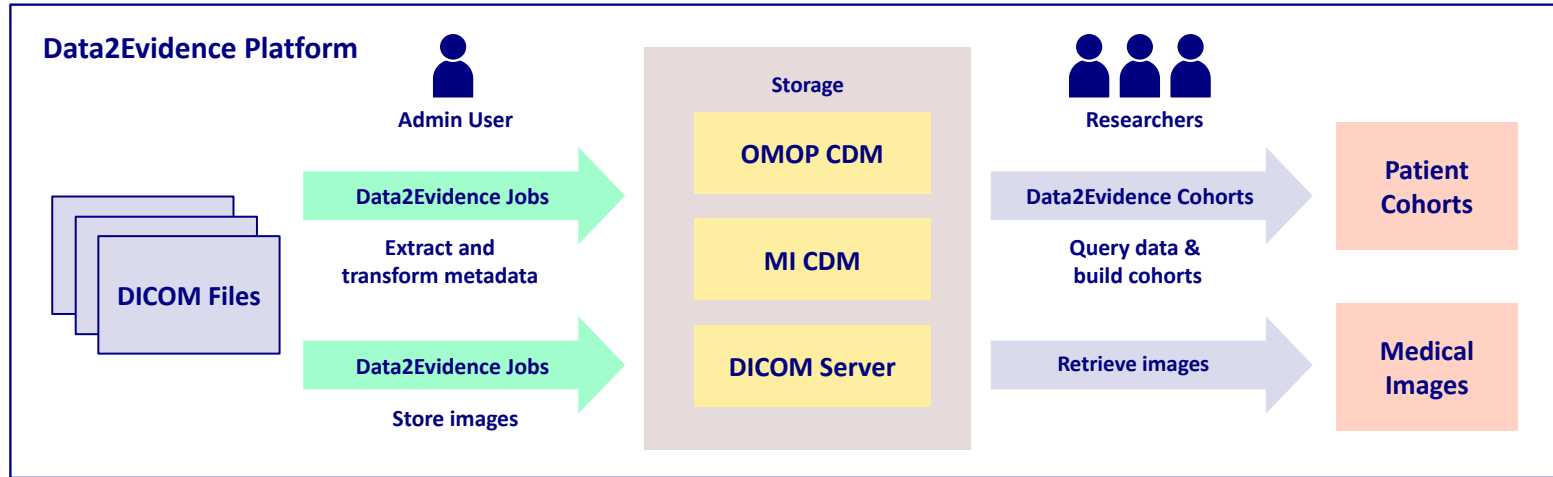
Hospital Platform

Harmonized data set
enabling research with
real-world data

- Modern development experience with a container orchestration framework based on Kubernetes
- Multi-modal data analysis, model training, and inference across siloed datasets
- Cross-institutional network studies
- Increased analysis quality due to a larger and broader pool of data
- Access to individual data sets consolidated on a single data platform

¹ OMOP CDM = Observational Medical Outcomes Partnership Common Data Model

Data2Evidence Imaging Data Support



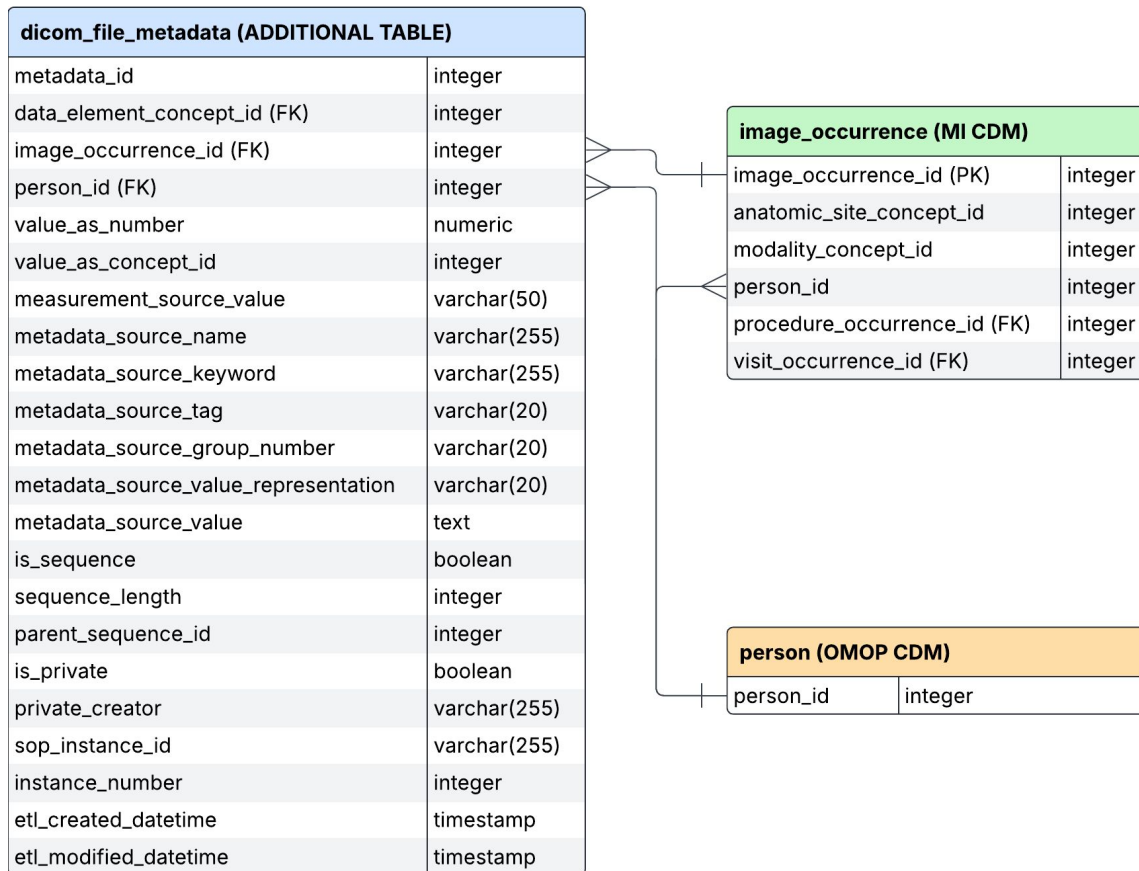
The Data2Evidence platform offers an end-to-end integration of medical imaging data, from ingestion of metadata extracted from the headers of DICOM files to the creation of patient cohorts.

Our solution is based on the work done by the OHDSI Medical Imaging Working Group.

Medical Imaging Extension

Our custom table *dicom_file_metadata* allows users to ingest all metadata values from DICOM files

Since the metadata sits in a separate schema from the OMOP CDM, admins can impose access controls based on schema.



Imaging Data Integration

A graphical tool for orchestrating ETL workloads with the use of modular, drag-and-drop nodes

Supports whole workflow of Data Integration:

- Concept Mapping (full text and semantic)
- Data Mapping (based on White Rabbit and Rabbit-in-a-Hat)
- Implementation of the transformation

The screenshot displays the DicomETLFlow interface. The top navigation bar includes 'Users', 'Datasets', 'Studies', 'Jobs', 'ETL', 'Setup', and 'Account'. The main workspace shows a workflow diagram with nodes like 'ExtractMetadata', 'MetadataID', 'ConceptTagID', 'MappedConceptID', 'TransformForIngestion', 'TransformForEATable', 'GetNumericRecords', 'GetNumericRecords', 'GetNumericRecords', 'GetImageOccurrence', and 'IngestImageOccurrence'. A 'Configure Concept Mapping' dialog is open in the foreground, showing a table of reference concepts from the 'Synup5pct' dataset.

Configure Concept Mapping

Reference concepts from dataset: Synup5pct

Populate concepts Download CSV Clear and Import another file

Status	Source	Name	Frequen...	Descrip...	Concept...	Concept...	Domain	Vocabul...
checked	A001	Cholera due to Vibrio c	A00	Cholera due to Vibrio c	200629	Cholera due to Vibrio c	Condition	
checked	A009	Cholera, unspecified	A00	Cholera, unspecified	198677	Cholera	Condition	SNOMED
checked	A0100	Typhoid fever, unspeci	A010	Typhoid fever, unspeci	192819	Typhoid fever	Condition	SNOMED
checked	A0101	Typhoid meningitis	A010	Typhoid meningitis	4100102	Meningitis due to typh	Condition	
checked	A0102	Typhoid fever with hea	A010	Typhoid fever with hea	440735	Rheumatic fever witho	Condition	
checked	A0103	Typhoid pneumonia	A010	Typhoid pneumonia	4186072	Pneumonia in typhoid	Condition	
checked	A0104	Typhoid arthritis	A010	Typhoid arthritis	192819	Typhoid fever	Condition	SNOMED
checked	A0105	Typhoid osteomyelitis	A010	Typhoid osteomyelitis	141663	Osteomyelitis	Condition	SNOMED
checked	A0109	Typhoid fever with oth	A010	Typhoid fever with oth	192819	Typhoid fever	Condition	
checked	A011	Paratyphoid fever A	A011	Paratyphoid fever A	442291	Paratyphoid C fever	Condition	

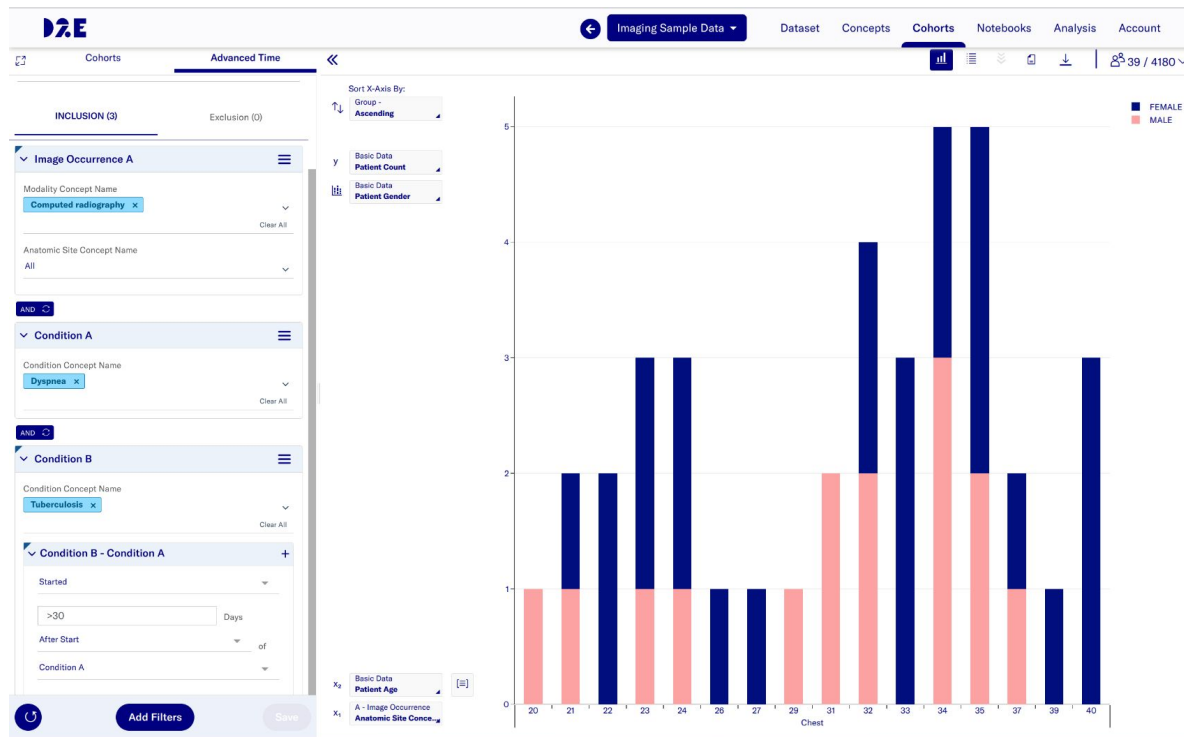
Rows per page 10 1-10 of 71,704

Apply

Data2Evidence Cohorts

A graphical tool for building patient cohorts

Users can apply filters on various attributes of the clinical and imaging data



This cohort show 39 patients, categorized by age and gender, who underwent Computed Radiography of the chest and were diagnosed with tuberculosis more than 30 days after the onset of dyspnea.

Imaging Data Integration

Demo

Types of OHDSI Network studies



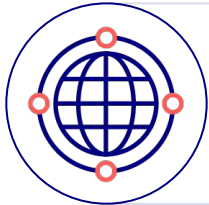
Clinical characterization

- Which treatment did patients obtain after diagnosis?
- How many patients experienced the outcome after treatment?



Patient-level prediction

- What is the probability a patient will develop a disease?
- What is the probability a patient will experience the outcome?



Population-level effect estimation

- Does one treatment cause an outcome?
- Does one treatment cause an outcome more than an alternative?

OHDSI Network Studies - Current approach vs Strategus

Traditional Approach

- Create an R package for executing the study
- Distribute this R package to network sites via [OHDSI Studies · GitHub](#)
- OHDSI network sites download the R package
- The sites configure an R installation and run the downloaded R study package
- Results are reviewed and contributed back to the study coordinator

Strategus approach

- Install & configure Strategus (HADES module)
- Create a JSON document to specify the study design choices
- Distribute the JSON document to network sites via [OHDSI Studies · GitHub](#)
- OHDSI network sites download the JSON document and execute the study via Strategus
- Results are reviewed and contributed back to the study coordinator

Source - https://www.ohdsi.org/wp-content/uploads/2022/11/2022_11_22_Strategus_OHDSI_Community_Call.pdf

OHDSI Network Studies - Strategus + Data2Evidence

- Jupyter notebook based interface for Strategus study design and execution
 - Researcher can focus on the study question without worrying about R infrastructure
 - Execution engine with all HADES & Strategus dependencies pre-installed
 - Docker based R kernels with all required modules installed and configured
- Pre-built study templates for commonly used network studies
 - Boilerplate code, loaded from github, including relevant HADES module(s) & default settings
- Decoupling of study design & execution
 - Researcher - create a study specification (JSON format), push to github repo
 - Data Partners - fetch the JSON and execute the study (custom execution settings)
- Integrated Shiny app based result viewer (**under development**)
 - Visualize study outcomes using shiny app

Clinical characterization

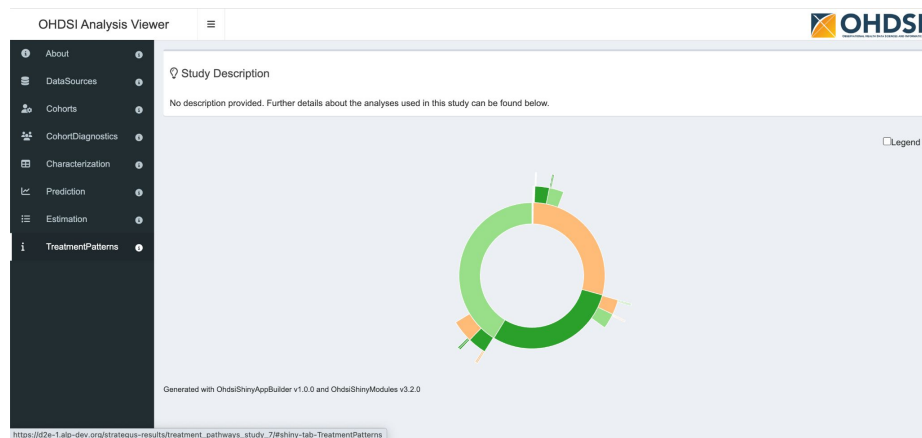
Treatment utilization: What treatments were done and in what sequence?

Question: In patients diagnosed with depression, what antidepressant treatments (SSRIs, SNRIs, and bupropion) were they exposed to, and in what order?

Why

Knowledge of current treatment practices to improve clinical practice:

- Better design of randomized control trials
- Give insight in choice of treatment and doctors preference



Clinical characterization

Demo

Data4Life and ATLAS

Data4Life is working together with the ATLAS team on a new UI to reduce the entry barrier of novice users and non-technical users to using ATLAS.

To this effect, we have envisioned three workstreams:

- User Interviews
- UX re-design
- Implementation of a new user interface

We plan to show a prototype of the new ATLAS UI as part of the ATLAS WG meeting at the Global OHDSI Symposium

The image displays a wireframe of the ATLAS user interface for Cohort Definition. The main window is titled 'Cohort Definition' and contains several sections:

- Cohort Entry Events:** Includes a 'Diabetes cohort' section with tabs for 'Earliest', 'All', and 'Latest'. It features an 'Add event' button and a list of condition concept sets for 'Diabetes Type 1' and 'Diabetes Type 2'.
- Evaluation Criteria:** Includes a 'Diabetes Type 2' section with tabs for 'Earliest', 'All', and 'Latest'. It features an 'Add event' button and a list of condition concept sets for 'Diabetes Type 2'.
- Cardiovascular disease:** Includes an 'Add event' button and a list of condition concept sets for 'Cardiovascular disease'.
- CAD-AF-Correlation:** Includes tabs for 'Continuous observation', 'Fixed duration to initial event', and 'Continuous drug exposure'. It features an 'Add event' button and a list of condition concept sets for 'Diabetes Type 2'.

A floating dialog box is overlaid on the left side of the main window, containing the following options:

- At least:** A dropdown menu with 'Exactly' selected.
- At most:** A dropdown menu with '1' selected.
- All:** A checkbox that is checked.
- Distinct concept:** A checkbox that is unchecked.
- Distinct start date:** A checkbox that is unchecked.
- OK:** A button at the bottom.

Initial wireframe of possible changes to the ATLAS user interface

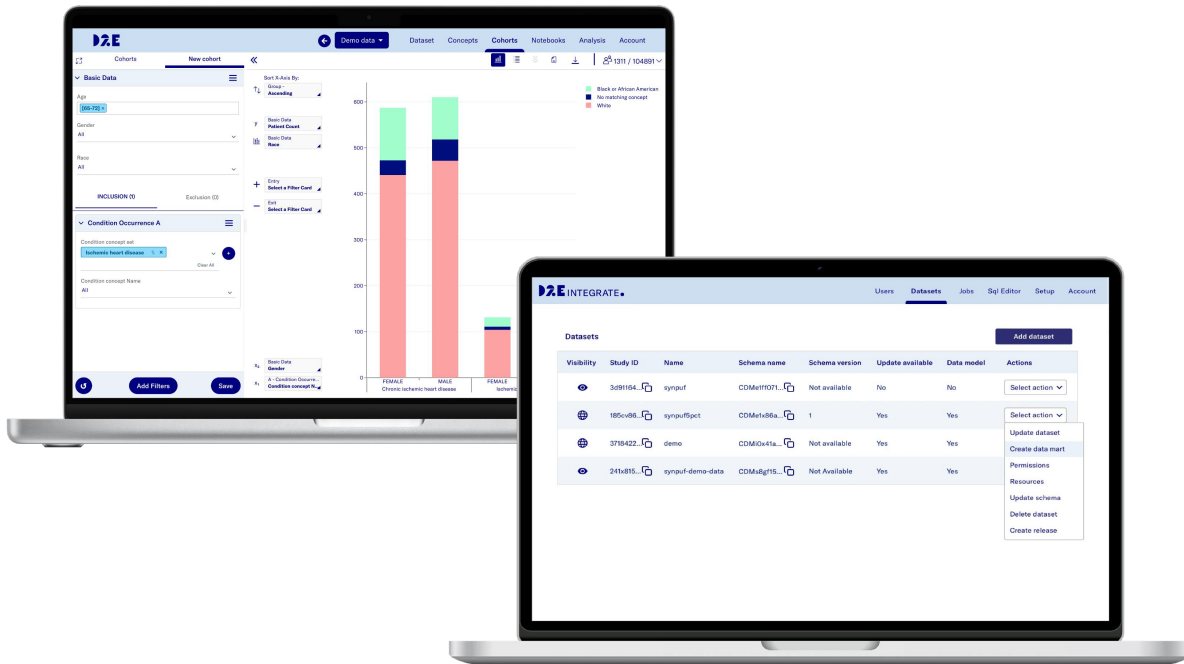
Data4Life at the OHDSI global symposium

Join us on **October 7** at the OHDSI Global Symposium for a hands-on collaborative session! Together, we'll design a network study using Data2Evidence and explore the power of real-world research in action.



Your research. Our platform.

D2E DATA2EVIDENCE • is open source!



Get started: www.d2e.sg

Github repo:
<https://github.com/OHDSI/D2E>



Thank you.

Peter Hoffmann
Co-CTO

peter.hoffmann@data4life-asia.care

Dr. Karthik Seetharaman
Lead PM (D2E and FAIR)

karthik.seetharaman@data4life-asia.care

Dr. Amit Sharma
Data Scientist

amit.sharma@data4life-asia.care