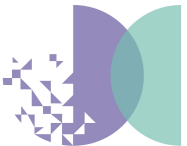


Creating standardised datasets in healthcare to aid federated analysis using the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM)

Solmaz Eradat Oskoui, PhD



Overview

1	Background
2	Data Interoperability: Global Perspective
3	Data Interoperability: Healthcare
4	OMOP CDM: Adoption, Benefits, Use Cases & Challenges
5	Conclusion

Background



Applied Mathematics, PhD

St.Andrews University

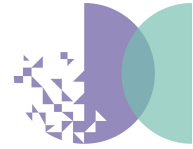
2014



Consultant/Senior Data Analyst/Data Engineer

Academia/IT/Banking/Healthcare

2014 - 2023



Founder

Data Harmonise Ltd

2024

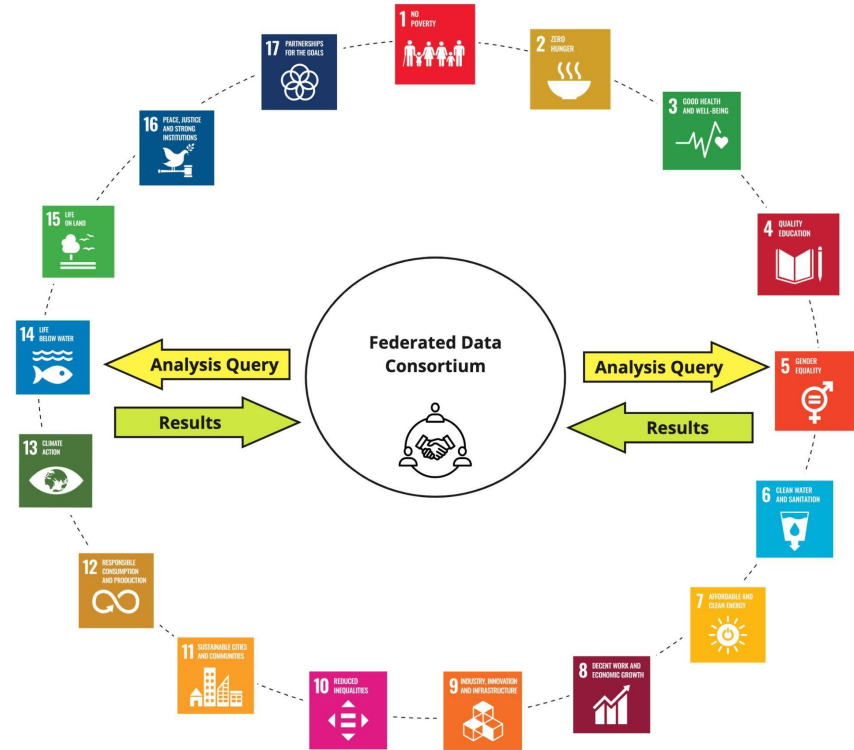
Data Interoperability: Global Perspective

Data and statistics at the center of national and global action to deliver the UN Sustainable Development Goals (SDG).

Collaboration and integration becomes achievable when decentralised data systems “work together” and are part of a network of a seamless data supply chain.

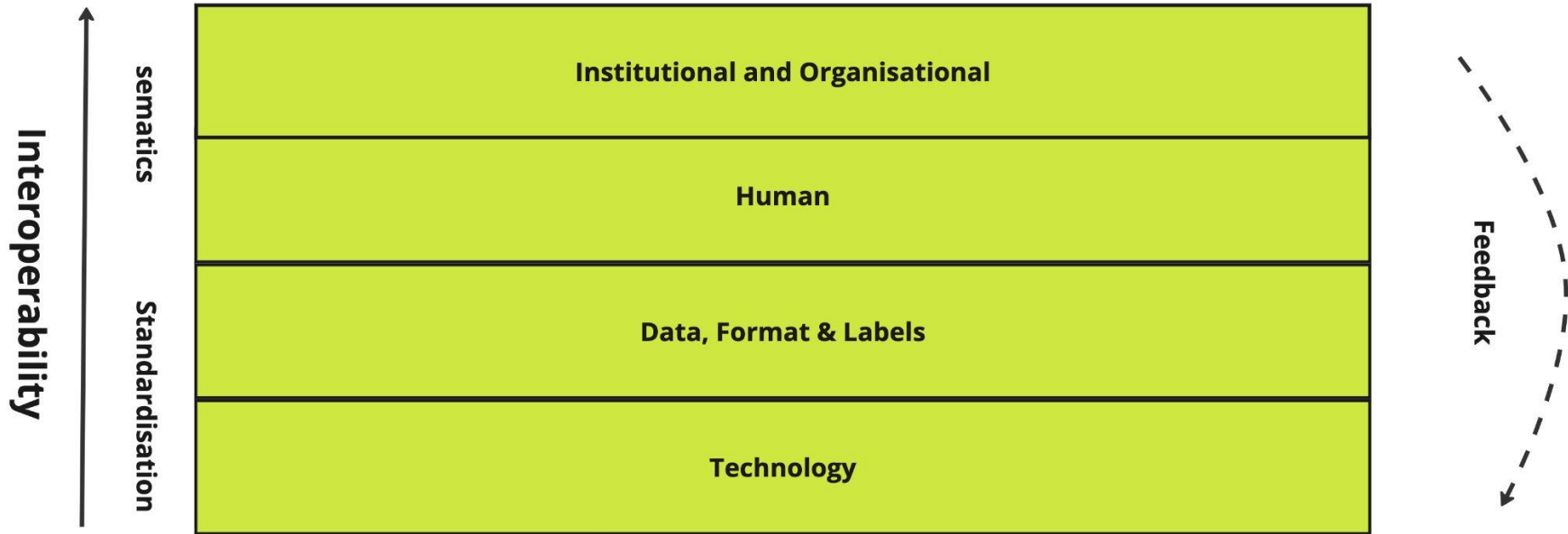
United Nations World Data Forum (UNWDF) - recognised to deliver these goals by 2030, good practises for enhancing data **interoperability** was needed.

“Interoperability is the ability to join-up and merge data without losing meaning.”



Key Components of Interoperability

Concept of interoperability broken down to layers that relate to standardisation and semantics



Source: [UN Data Interoperability Guide](#)

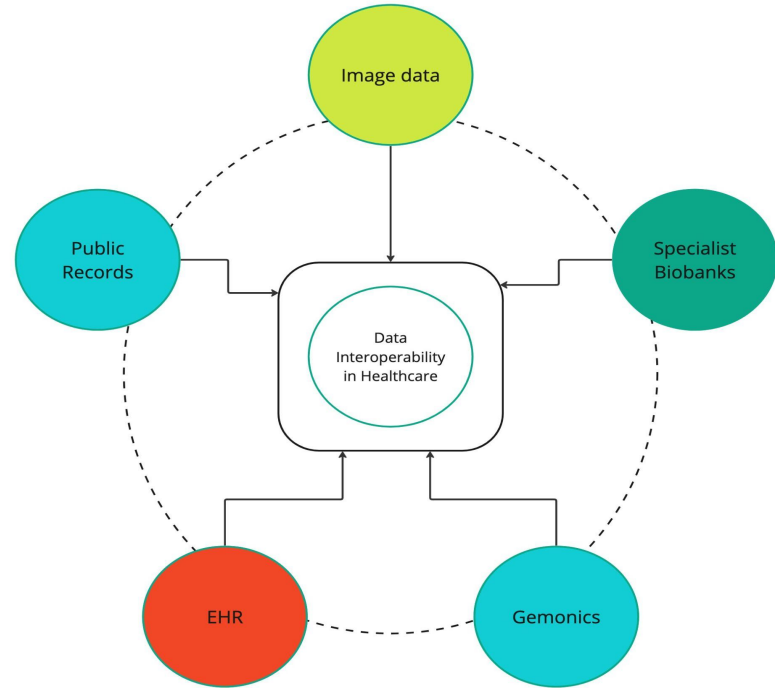
Data Interoperability: Healthcare

Diverse Sources of Healthcare Data: Data are collected for different purposes, stored in different formats using different database systems and information models.

Inconsistencies in Data Formats and Terminologies: Healthcare data can vary greatly both in structure and terminologies from one organisation to the next.

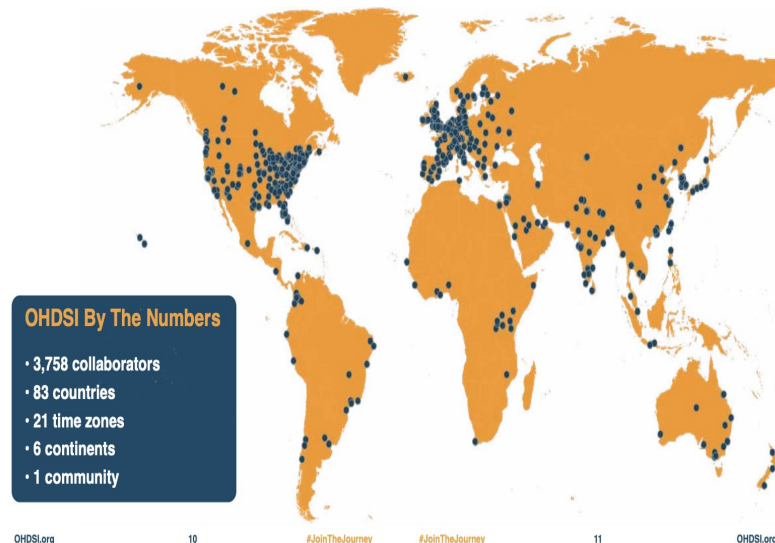
Impact on Research & Data Sharing: Difficulties in aggregating data for large-scale analysis.

Standardised Communication: To ensure seamless communication and data exchange between different healthcare systems (often from different vendors), requires standardising data both in structure and in terminologies.



OHDSI Community & OMOP CDM

- **Mission:** The Observational Health Data Sciences and Informatics (**OHDSI**) a multi-stakeholder, interdisciplinary community striving to improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.
- **Method:** Creating a standardised data model known as the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM).
- **Impact:**
 - Growing adoption over 10 years,
 - **534 data sources** from **49 countries** mapped,
 - Represents **956 million patients** (12% of world population).



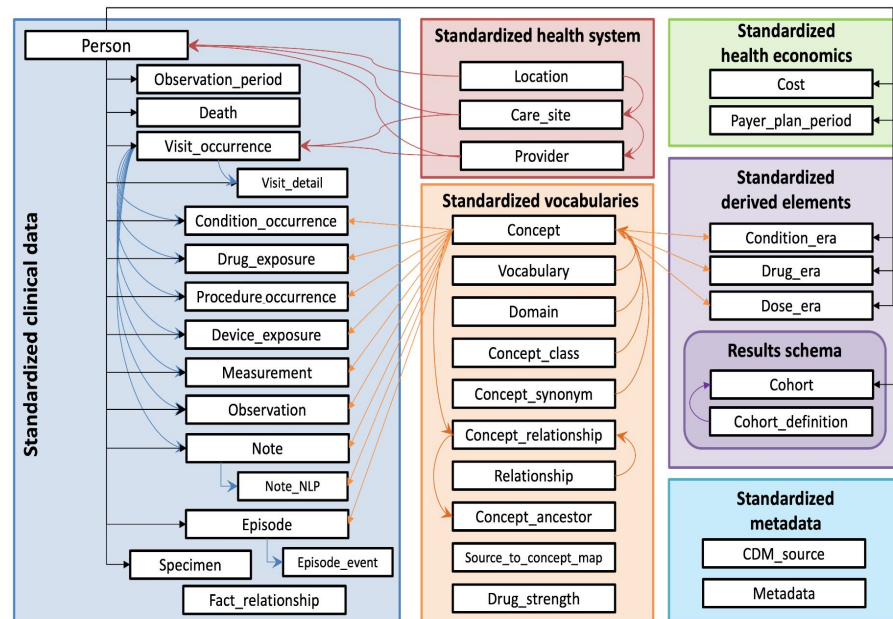
Source: [Where the OHDSI community has been and where are we going Annual report 2023](#)

Adopting the OMOP CDM

OMOP-CDM used to link observational healthcare databases.

Key features:

- Patient-centric
- Tabular/Relational Design
- Standardised Vocabulary
- Extendable
- Built for collaborative analytics through the creation of **federated data networks**.



Source: [OMOP-CDM V.5.4](#)

Source to Standard Mapping

Extract



[patients.csv](#)

Transform



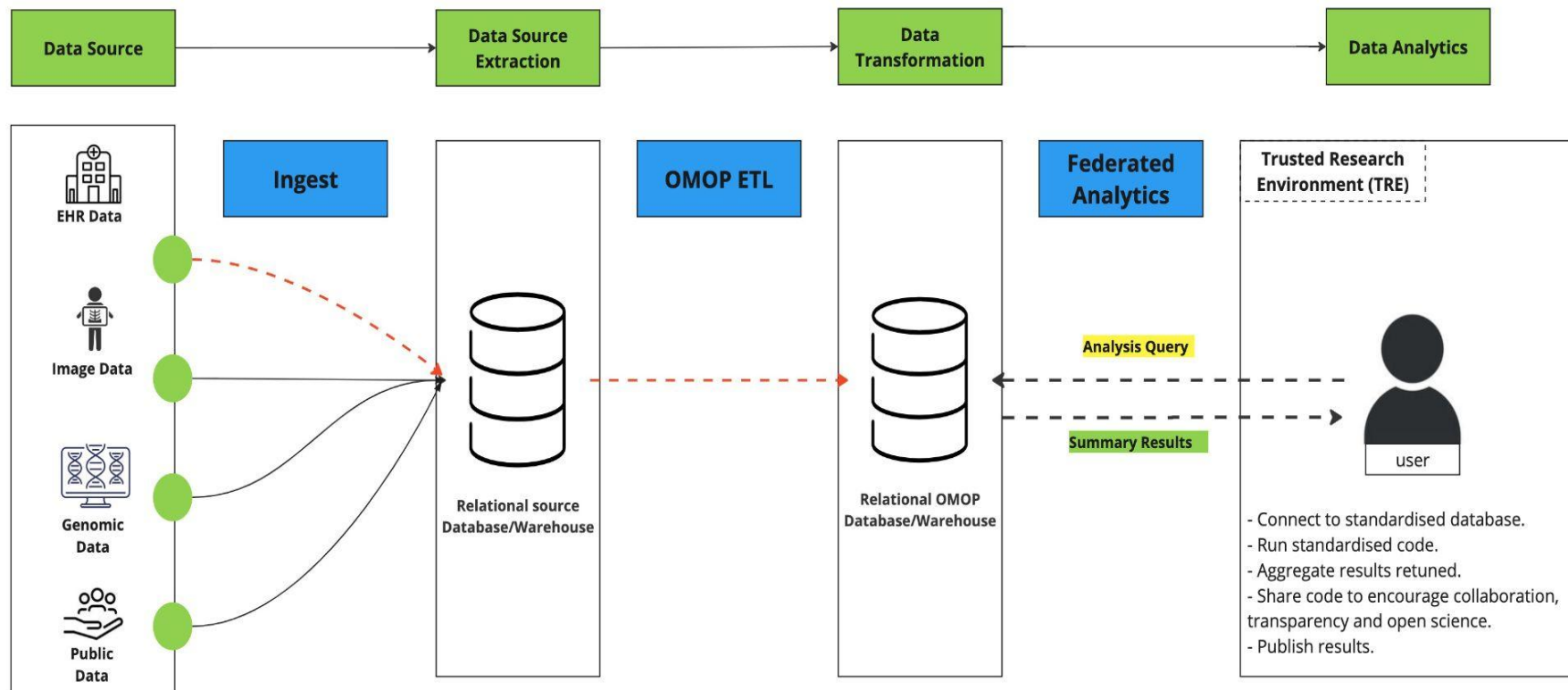
[person.sql](#)

Load



[person table](#)

Technical Architecture Overview



Benefits of adopting the OMOP CDM

Open-source and Collaborative:

- Actively developed and maintained tools on Github with a supportive community forum.
- Offers guidance and assistance for general and specific data conversion questions.

Standardised Representation:

- Creates a common language for healthcare data, simplifying understanding.

FAIR Data Principles:

- Supports **F**indability, **A**ccessibility, **I**nteroperability, and **R**eusability of data.
- Enables researchers and healthcare technology development for better patient and population health decisions.

Large-Scale Datasets:

- Enables creation of large, standardised datasets for national and international federated network studies.

Use Cases and Success Stories

Darwin EU:

- 5 year study will supports regulatory decision-making through use of real world evidence.
- 40 Data providers
- ~130M patient data
- 140+ studies delivered by 2025

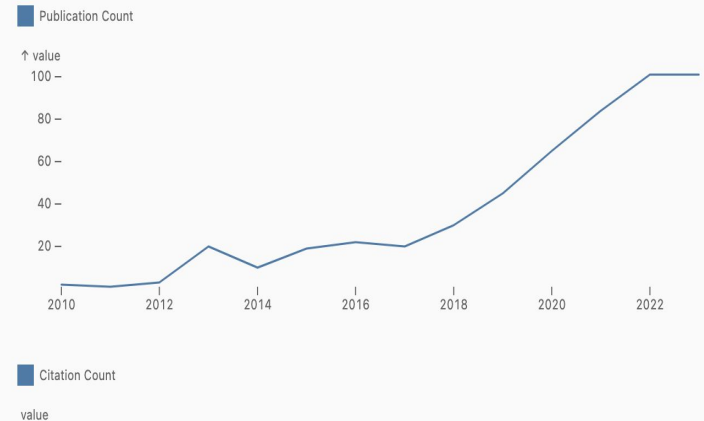
Pioneer

- Enabling collaboration between pharmaceutical industry, patient records and clinical trials in the area of prostate cancer.

OMOP CDM used in Covid 19 studies:

- [Impact of Covid-19 vaccinations on 24 millions patients across 26 global databases.](#)

Publications



Source: [Community Dashboard](#)

Challenges when adopting the OMOP CDM

Interpretation of Source Data: Requires well-defined source data dictionaries and input from clinical experts familiar with the source data.

Conversion Time: First iteration of mapping typically takes 6-12 months to complete.

Time Consuming Vocabulary Mapping: Identifying terminologies to be added to the OMOP vocabularies can be a time consuming and manual process.

Ongoing Maintenance: Regular updates are needed to adapt to changes in source data, data models, and vocabularies.

Conclusion

Interoperability:

- Enabler in the delivery of the UN SDGs by allowing sharing of right data, in the right format, at the right time across sectors.

OMOP CDM:

- Data integration from different healthcare systems.
- Supports FAIR principles through standardised dataset, codes and results.
- Cross-institutional and cross-country data sharing securely through federated analytics.
- Supports advanced analytics applications (AI, federated learning, cross-sector studies).
- Accelerates healthcare innovations and generation of insight from health data for all.

Future of OMOP CDM:

- Ongoing work to enhance OMOP CDM with new extensions (e.g., genomics, imaging, GenAI, and Animal Health).
- OMOP federated networks have the potential to positively impact patient treatment and healthcare outcomes for all.

THANK YOU FOR YOUR ATTENTION

Email: solmaz.eradat@dataharmonise.co.uk



Solmaz Eradat Oskoui, PhD
Founder | Data Engineer

