PHUSE SDE JAPAN
05 December 2025

Real-World Data Potential and Challenges for Regulatory Submissions

Content

- Introduction
- Source data selection
- Interoperability
- Lineage and traceability
- Conclusions

Introduction

- Real-world data
  - FDA: Data related to patient health status and/or the delivery of health care routinely collected from a variety of sources [1)].
  - EMA: The data that describes patient characteristics in routine clinical practice [2)].
  - PMDA: Data obtained in the actual medical environment [3)].
- Opportunities
  - External control arm
  - Pragmatic trials
  - Decentralized trials

1) FDA guidance: Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products - guidance for Industry. (July 2024).
2) EMA Data Quality Framework for EU medicines regulation: application to Real-World Data (Draft, Nov 2024).
3) PMDA Points to Consider for Ensuring the Reliability in Utilization of Registry Data for Applications. (Provisional Translation – March 2021)

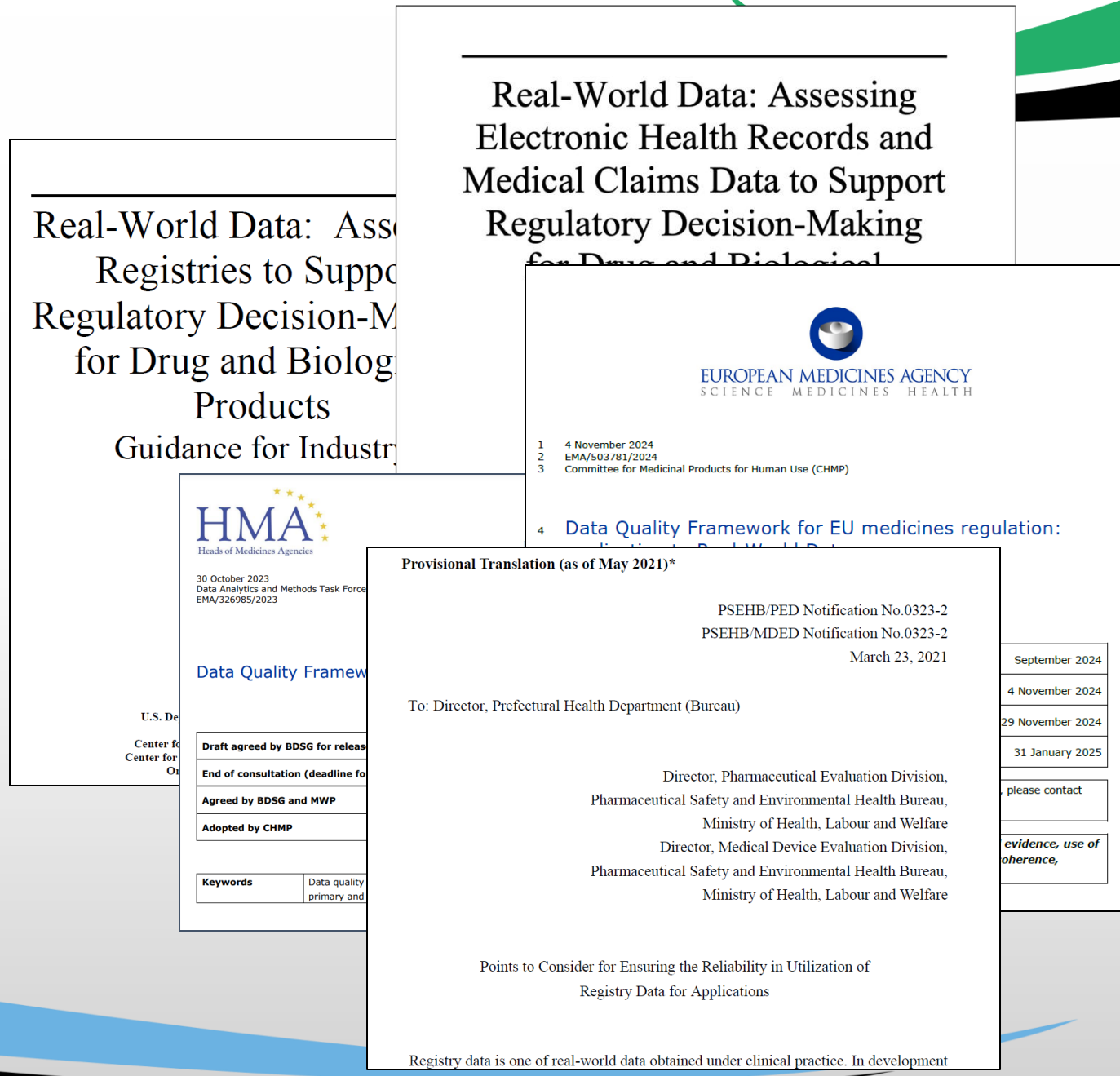
Introduction

- Reasons for using Real-World data
  - Efficiency
  - Less patient burden
  - Faster approval for rare disease treatments

- Before you start
  - Data source selection
  - Feasibility assessment
  - Regulatory alignment

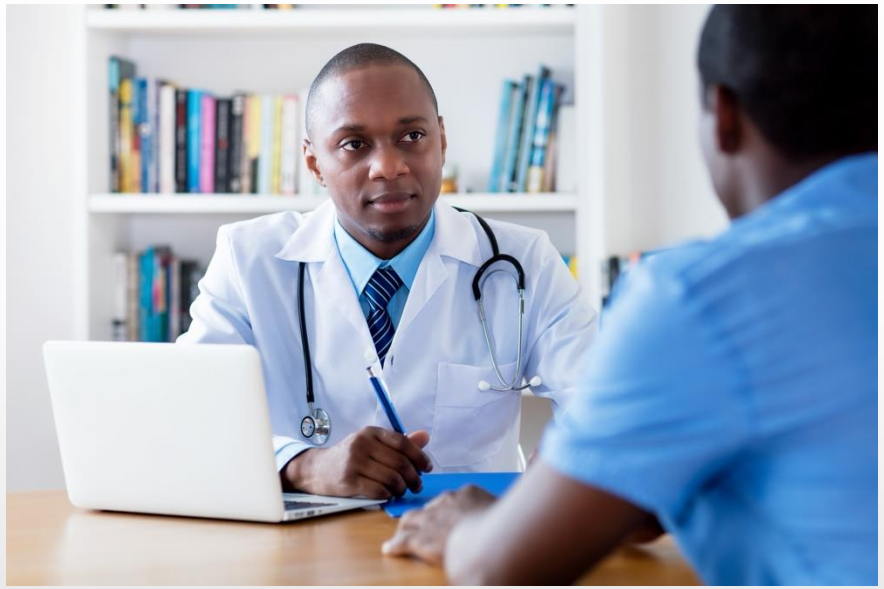
Regulatory guidance

- Data quality framework EMA
  - Extensiveness
  - Coherence
  - Timeliness
  - Relevance
  - Reliability
- FDA assessing EHR and Claims
  - Reliability
  - Relevance
- PMDA
  - Reliability



Source Data Selection

- Electronic Health Records
  - General Practitioner (GP)
  - Hospital/clinic data
- Claims
  - Pharmacy
  - GP, Hospital, Clinic
  - Insurance companies
- Registries
- Other



Source Data Selection

- Collection practice has effect on **reliability** and analysis
  - Data collected and stored in information system according to own specifications.
- What is collected has effect on **relevance** of data
- Standardization has effect on **reliability** and **relevance**
  - Improving interoperability
  - Traceability/lineage requirements
  - Reducing variety of information

FDA guideline: "Data in CDM-driven networks rarely contain all the source information present at the individual health care sites, and the data elements chosen for a given CDM network may not be sufficient for all research purposes or questions."

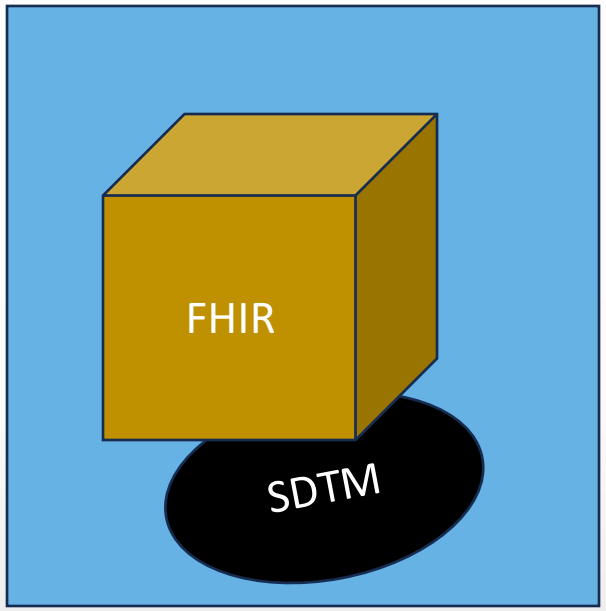
Source Data Selection

- Request for Information
  - Relevance: Data elements, Observation Periods, Coverage, …
  - Reliability: Provenance, traceability, quality, …
  - Other: Governance, Costs, Dictionaries, …
- PHUSE RWE Project: Quality and Reusability of Real-world data
  - White paper including RFI template: Public review expected in 2026 Q1
- SPFID-2:
  - Evaluate fit for purpose of sources based on information provided [1)].

1) A Structured Process to Identify Fit-for-Purpose Study Design and Data to Generate Valid and Transparent Real-World Evidence for Regulatory Uses. (Gatto et.al., Mar 2023).

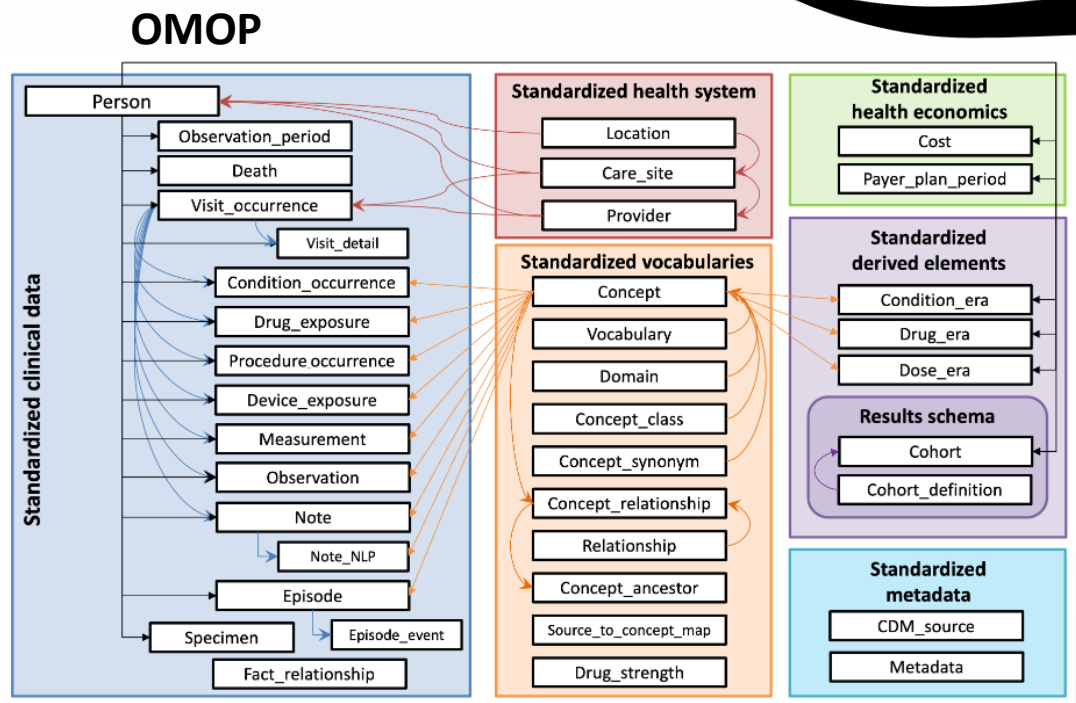
Interoperability

- Improves automation
- Improves clarity
- Improves efficiency
- Reduces errors and inconsistencies
- Needs
  - Mapping
  - Documentation
  - Validation



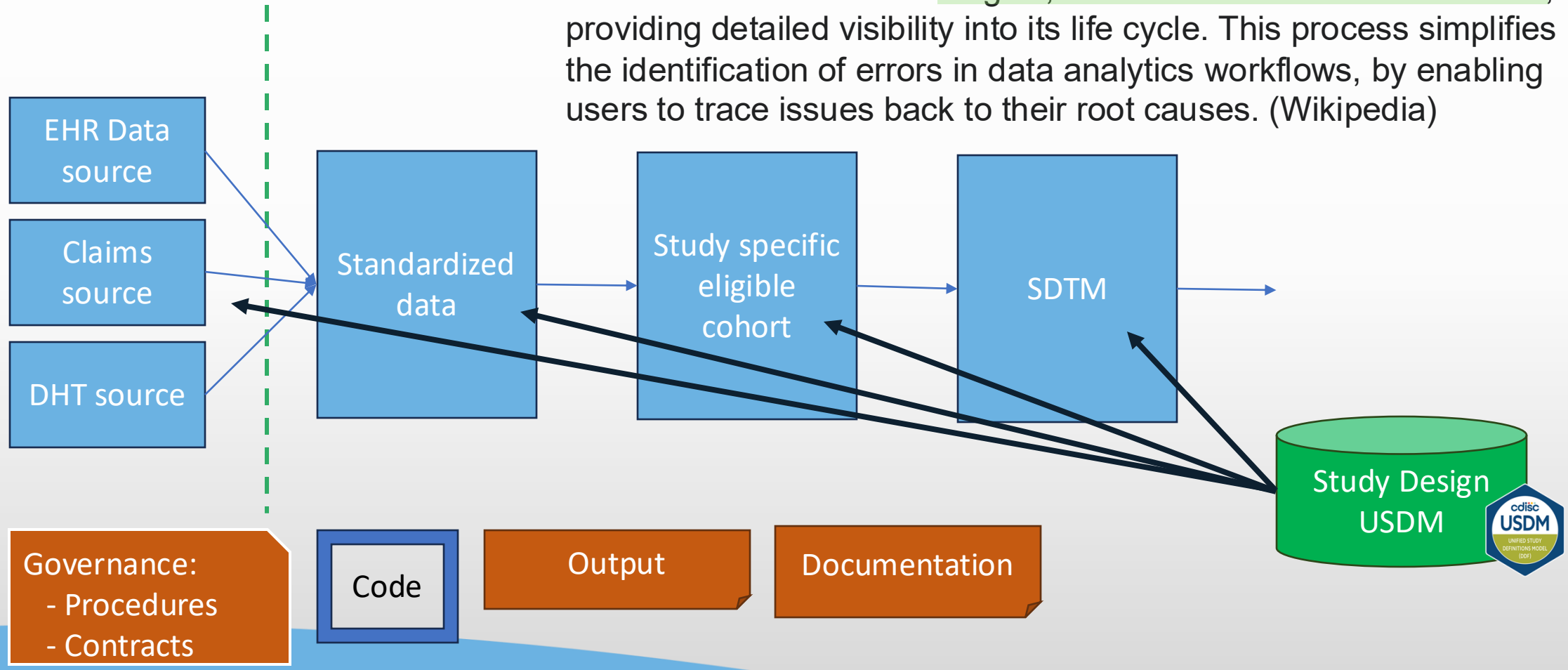
Interoperability

- Challenges
  - Data structure
  - Terminology (semantic interoperability)
  - Missing information
  - Governance
  - Usability
- In-between data layer needed for fit-for-purpose assessments.
  - OMOP(-like)
  - Extended/simplified SDTM
  - Data vendor structures

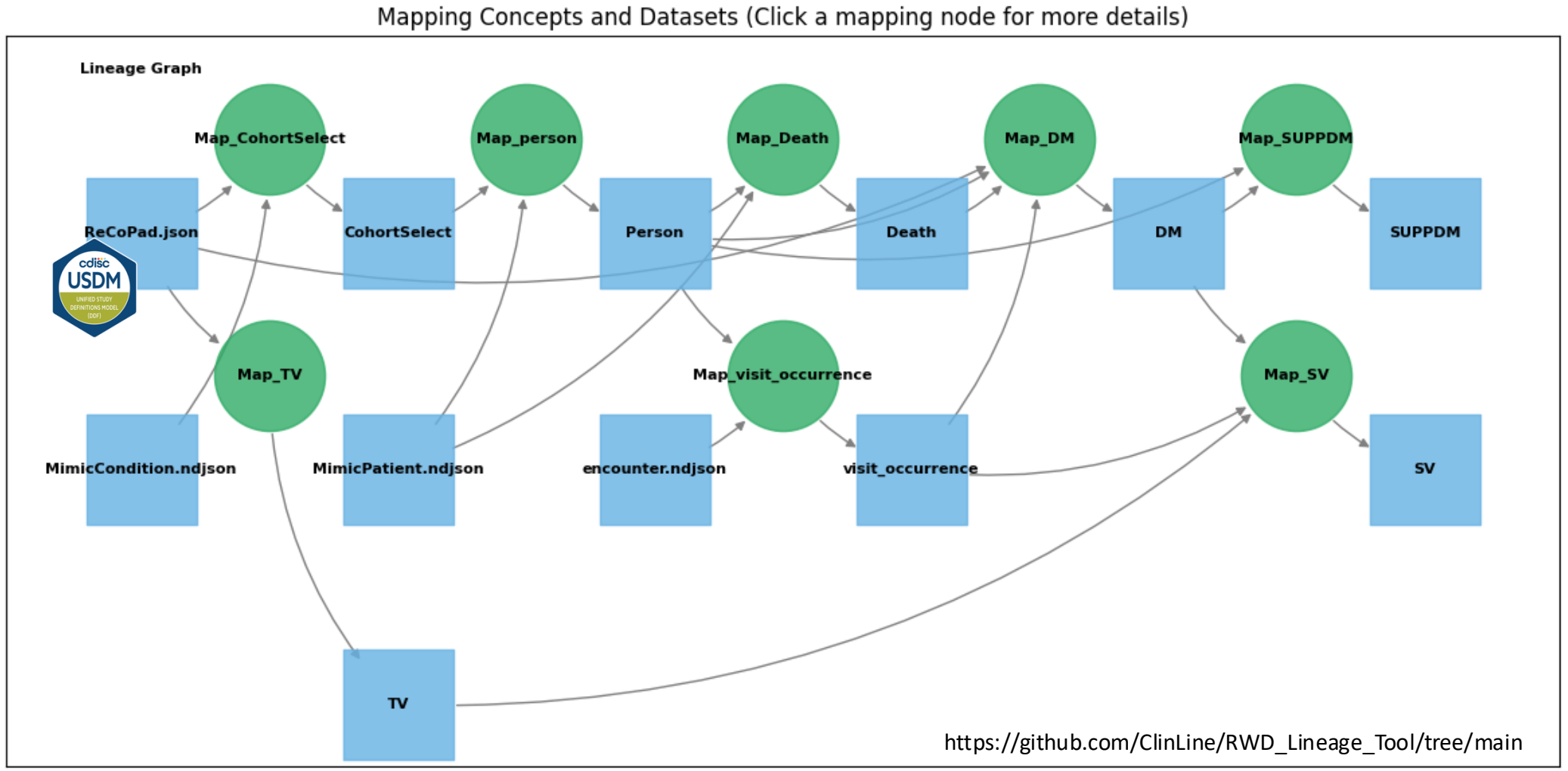


Data lineage

Data lineage refers to the process of tracking how data is generated, transformed, transmitted and used across a system over time. It documents data's origins, transformations and movements, providing detailed visibility into its life cycle. This process simplifies the identification of errors in data analytics workflows, by enabling users to trace issues back to their root causes. (Wikipedia)



Data Lineage and traceability

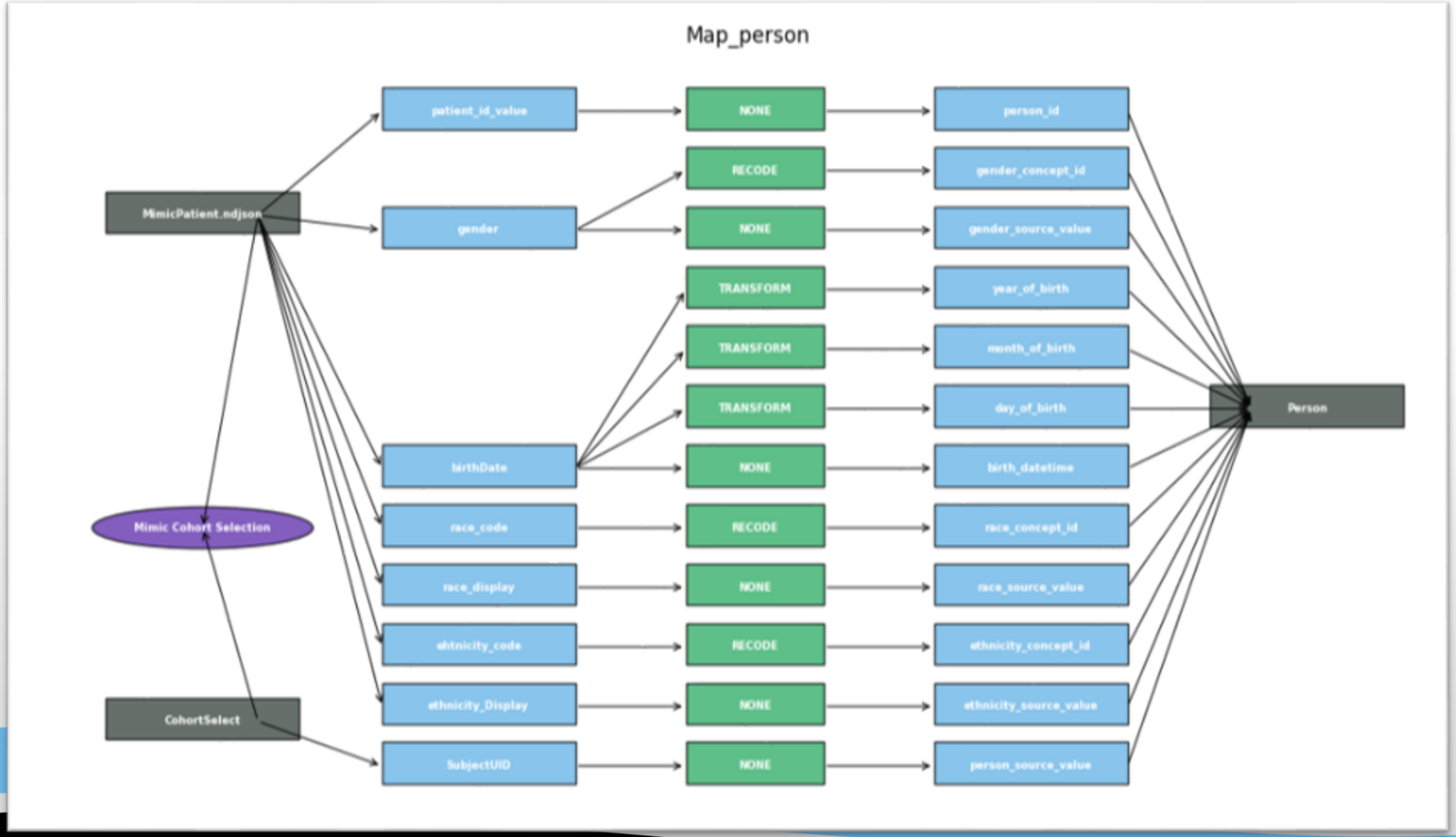


Transformation Nodes

- Transformation between two fixed states in data lineage
  - Data exchange
  - Validation
  - Mapping and data transformations
  - Data Selection (positive / negative)
  - Reporting
- Combines all information defining the transformation
  - Code
  - Input datasets
  - Input specifications
  - Procedures
  - Contracts

#ClearDataClearImpact

Transformation Nodes

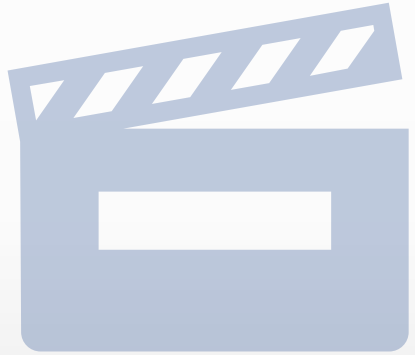


Security and verifiability

- Transformation nodes and all referred information:
  - Input and output
  - Documentation
  - Code and log / proof of execution
- Immutable
  - Protected environment
  - Proof that it is not changed
    - Date/time file/run / log verification
    - Rerun verification
    - Hashing / cryptography

Challenges

- Source data formats
- Mapping details
- Documentation!!
  - Availability
  - Findability
- Pseudonymization/Anonymization
- AI utilization
- Submission data requirements
- Current processes, procedures, and priorities



Unified Study Definition Model

- Holds full Study design in digitized format
- May hold the corresponding document with links to parameterized information (including M11)
- Includes model area's for
  - Population and eligibility criteria
  - Schedule of Activities (SoA) and other timelines
  - Activities and corresponding procedures and assessments
  - And more…
- Can include mapping information (Alias Code)
- Drives automation!!

https://www.cdisc.org/ddf

Projects and Solutions

- Use cases
- Open-source tool development
- USDM utilization
- Industry initiatives
  - CDISC
  - PHUSE
  - TransCelerate
- Requirements and implementation

Conclusions

- Account for high diversity in source data
- Ensure lineage and information readiness
  - Clear mapping methods and mapping overview
  - Inclusion of traceability information
  - Visualizations
- Automation
  - Standard and validated programs
  - Reduce manual tasks
  - Utilize defensive programming
- Future improvements
  - More Automation & Scalability
  - Open-source tool
  - USDM as a driver
  - Available coding and BC lists as a driver

Questions and contact

- Questions?

- Like to learn more?
  - Berber Snoeijer, ClinLine: b.snoeijer@clinline.eu