



PHUSE EMEA SDE25 Germany

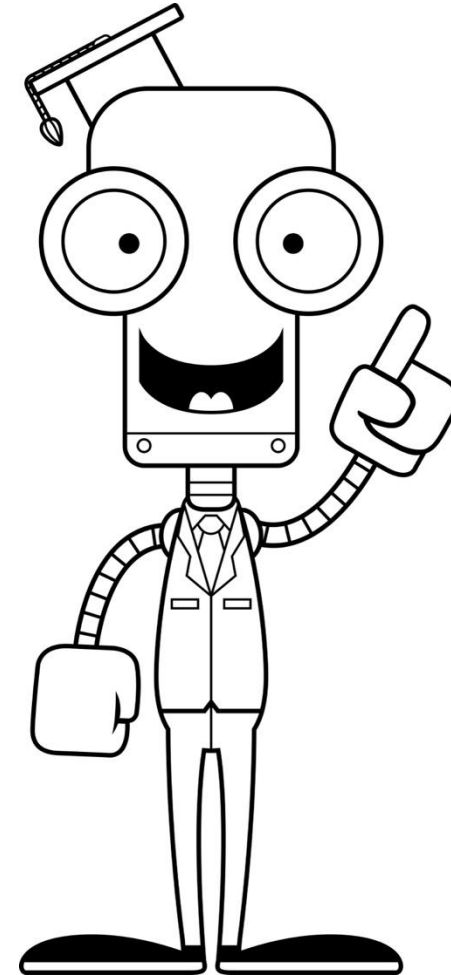
Bridging the standards: standardized mapping and lineage

Berber Snoeijer, ClinLine

Javier Rodriguez Soto, OCS Life Sciences

Content

- Introduction
- Use case description
- MIMIC data
- Mapping and automation
- Lineage tool
- Conclusions



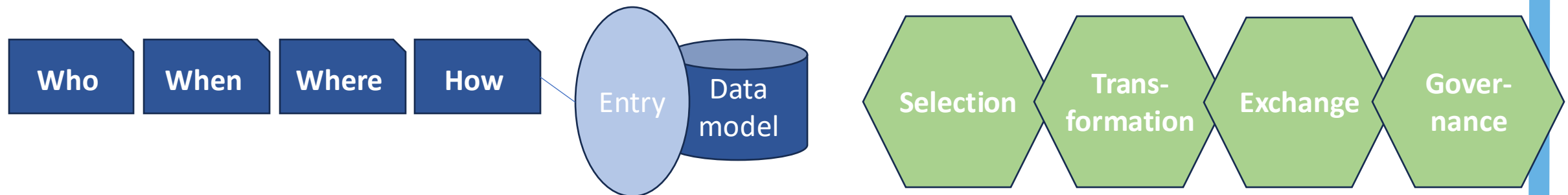
Introduction

See www.clinline.org/break-learn for the webinar recording (30 minutes):



■ What is data lineage?

- Wikipedia: **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.**



Regulatory requirements

- EMA Data Quality Framework
 - Reliability: fundamentally depends on the systems and process in place for the primary collection of data and its processing.
- FDA guideline: Data Standards for Drug and Biological Product Submissions Containing Real-World Data
 - During data curation and data transformation, adequate processes should be in place to ensure confidence in the resultant data.
 - Documentation of these processes may include but is not limited to electronic documentation (e.g., audit trails, quality control procedures, etc.) of data additions, deletions, or alterations from the source data system to the final study analytic data set(s).
- FDA guideline: Real World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products
 - In general, sponsors should address the procedures used to ensure completeness and accuracy of study data, as well as processes for data accrual, curation, and transformation over the data life cycle.

Use Case

1. Select a cohort of patients from a large, multi-center, multi-site, multi-national, multi-ethnic deidentified Real-World Data source
2. Transform FHIR data to a common data model (OMOP) for selection
3. Assess eligibility of patients based on selection criteria.
4. Transform eligible and selected patient data to SDTM submissible datasets
5. Select and Map the data with the OCS LS Mapping Engine
6. Show lineage for every step and transformation using a newly developed Open Source tool.

**Proof Of Concept 1:
Focus on patient data and
corresponding lineage to SDTM
DM dataset**

What is MIMIC?

- Medical Information Mart for Intensive Care (**MIMIC**)-IV
- **Deidentified dataset** of ~300,000 critical care patients (from Beth Israel Deaconess Medical Center (Boston, MA)).
- **Publicly available data:**
 - Database accessible with full credential on PhysioNet*
 - Demo database accessible with open credential
- **Widely used** – over 3,000 publications, including retrospective analyses and algorithm development.

**Access requires completion of a training course and approval through a data user agreement.*

General Overview of MIMIC-IV

MIMIC Follows the Patient journey, integrating data from three hospital systems

- Patient enters the Emergency Department
ED system: Triage event information, vital signs, medical history, diagnosis, etc.
- Transferred into the ICU
ICU-specific clinical system: Procedures, drug administrations, chartered events, stays, etc.
- Patient Tracking during the stay (Hospital)
Electronic Health Record (EHR) system: laboratory and microbiology tests, billing codes, transfers, procedures, etc.

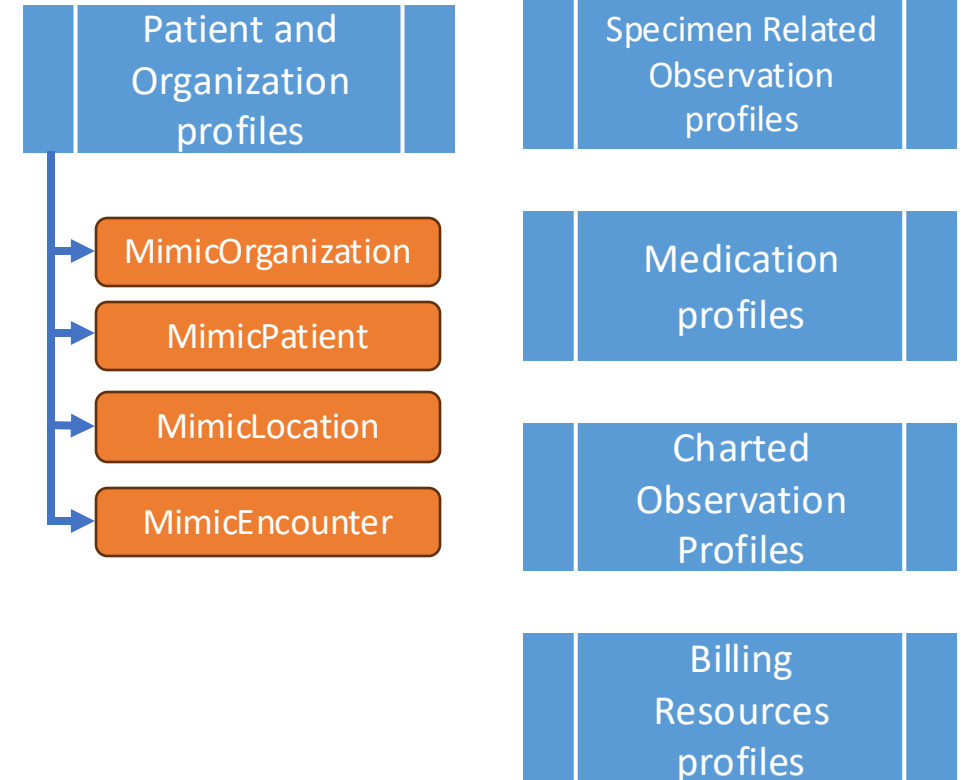


MIMIC-IV-on-FHIR

- **Proof of Concept:** MIMIC-IV-on-FHIR.
- **Fast Healthcare Interoperability Resources (FHIR)** provides a framework for structuring and exchanging health data.
- **FHIR adoption is expanding worldwide.** Growth driven by regulatory mandates.
- **Defines a set of resources** for common healthcare entities.
- **Uses "profiling"** to adapt base **resources** for local healthcare cases.

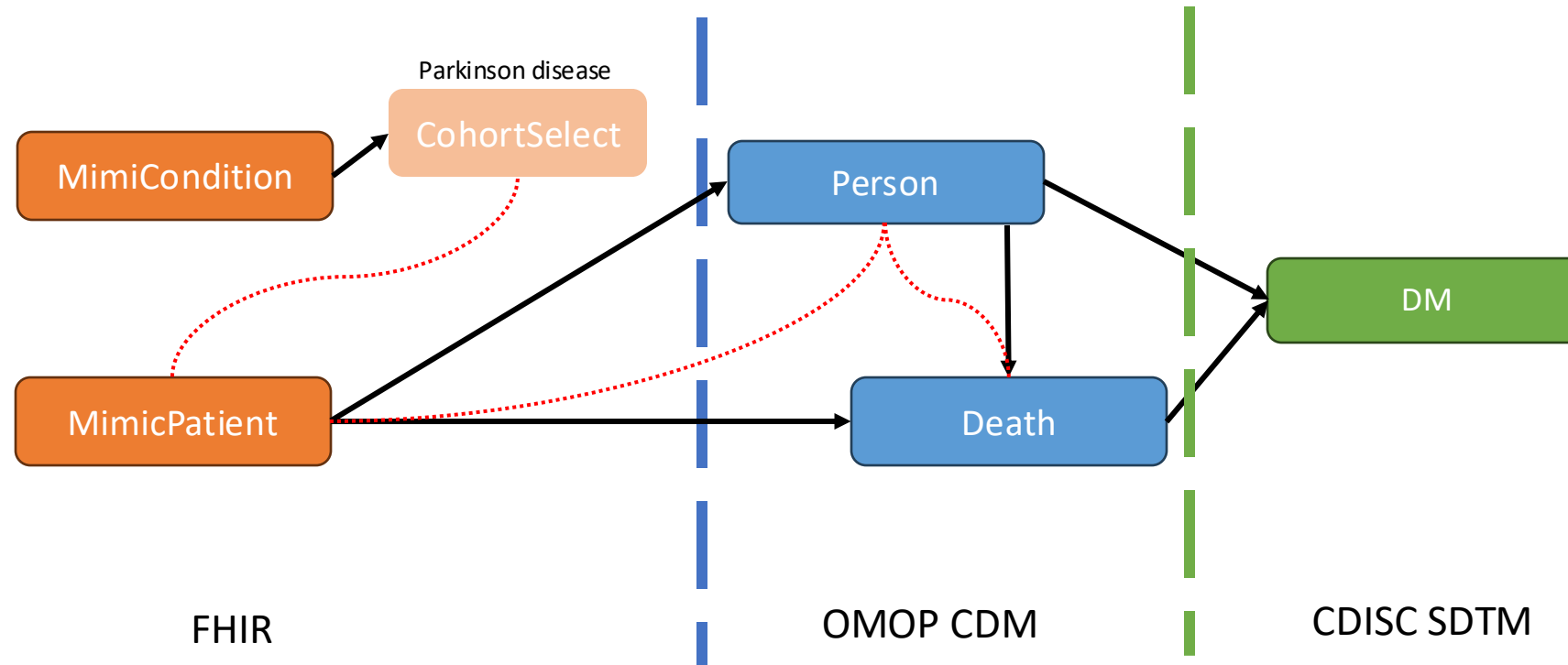
Profiles and Terminology

- MIMIC-IV-on-FHIR collection of new specialized profiles:
 - Total of 24 + 6 (ED) profiles.
 - Used to collect MIMIC data without modification.
 - Retains original terminology, no mapping was performed.
- **Supports FAIR principles** (Findable, Accessible, Interoperable, Reusable).



Proof of concept: Mapping Patient Data

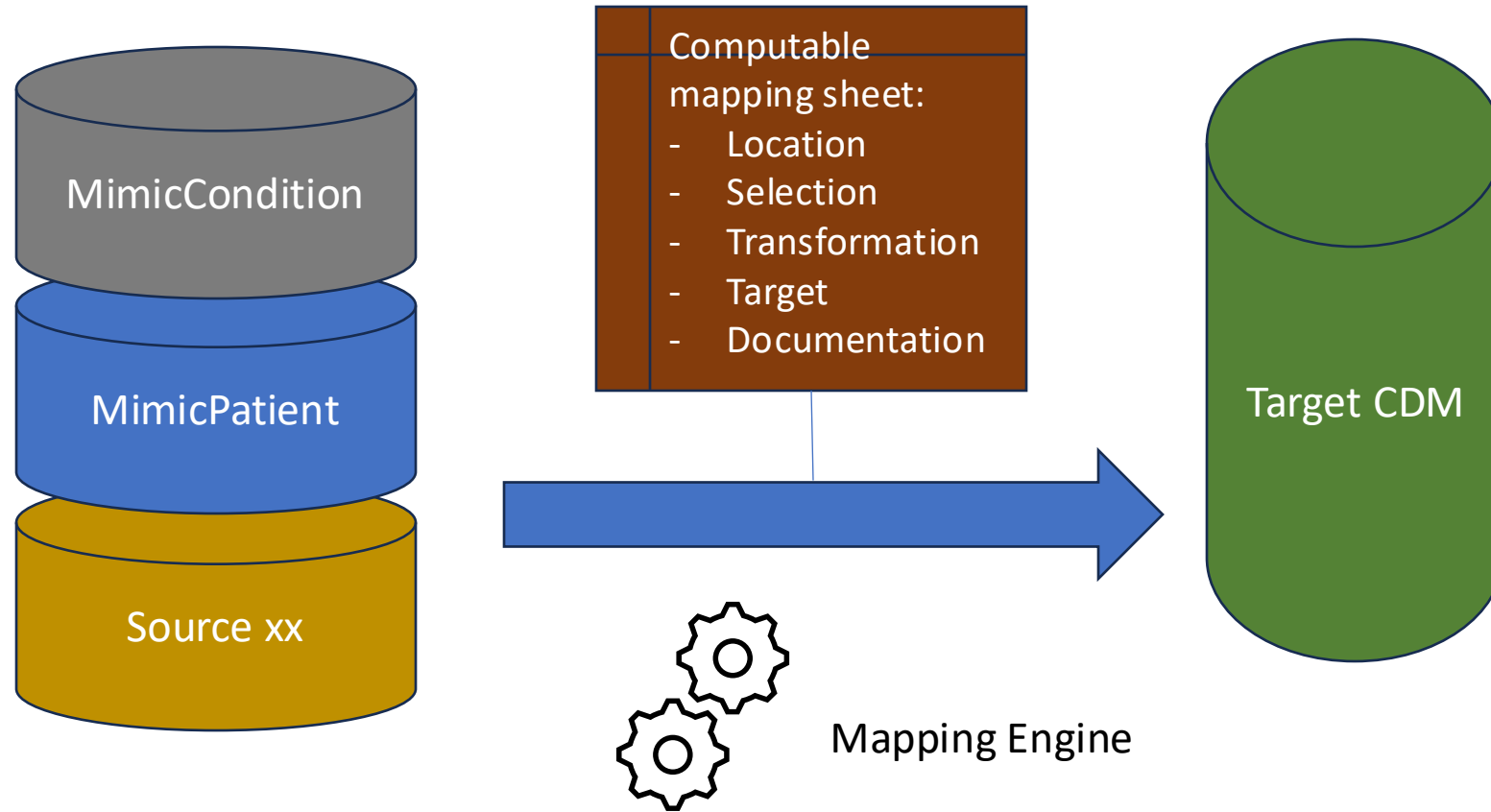
- FHIR Patient resource to CDISC Standard (SDTM Demographics)



Considerations

- Data Structure challenges when working with FHIR and SDTM.
 - Different purpose and structure
- Terminology Alignment:
 - FHIR stores terminology in two resources: CodeSystems and ValueSets
 - SDTM uses CDISC-controlled terminology (CT)
- Handling challenges like missing identifiers and standardizing formats
- Data mapping: Determine which information might be relevant for target data
- Ensuring compliance with CDISC standards

Data mapping automation and traceability

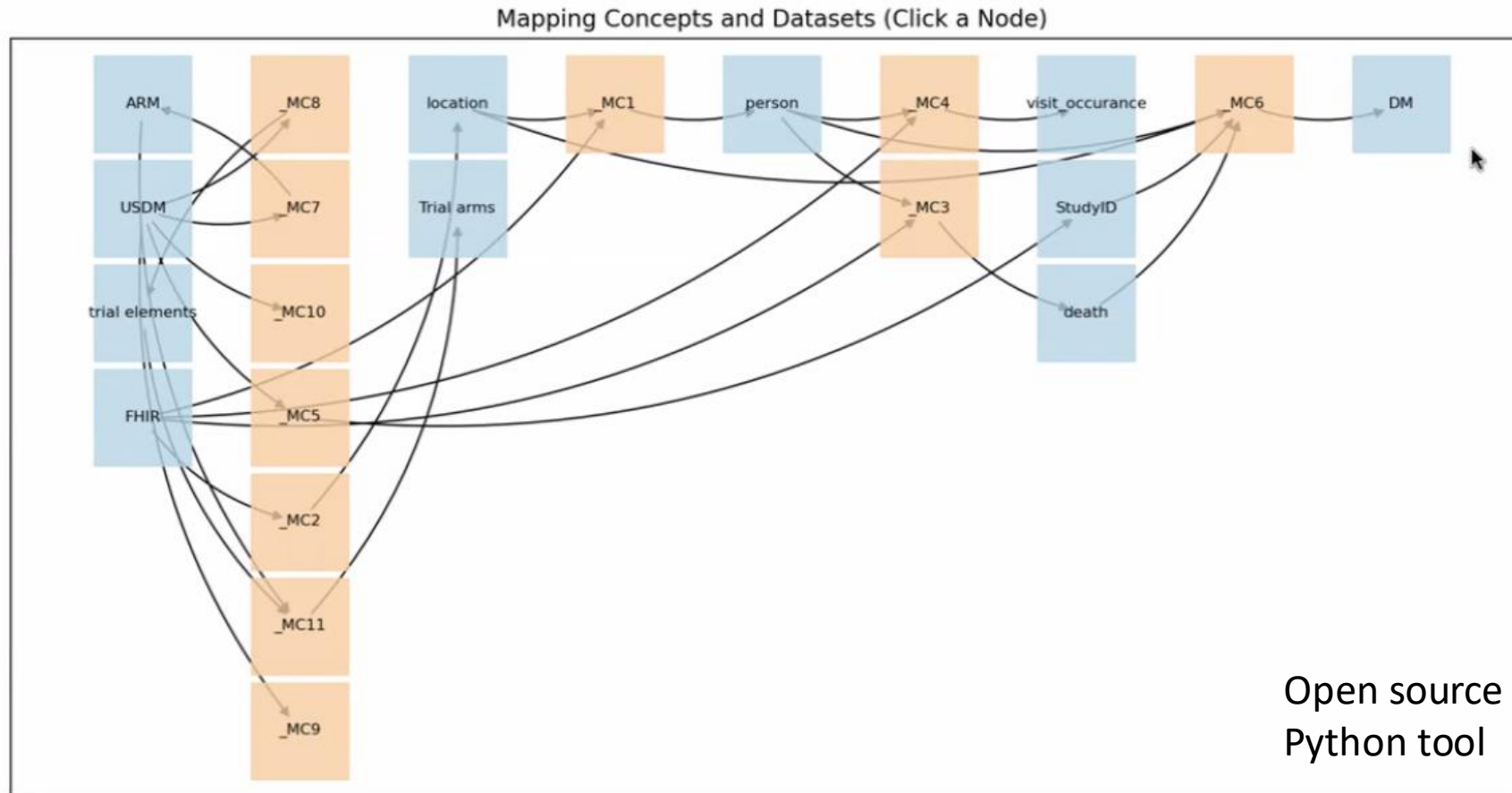


Defensive programming

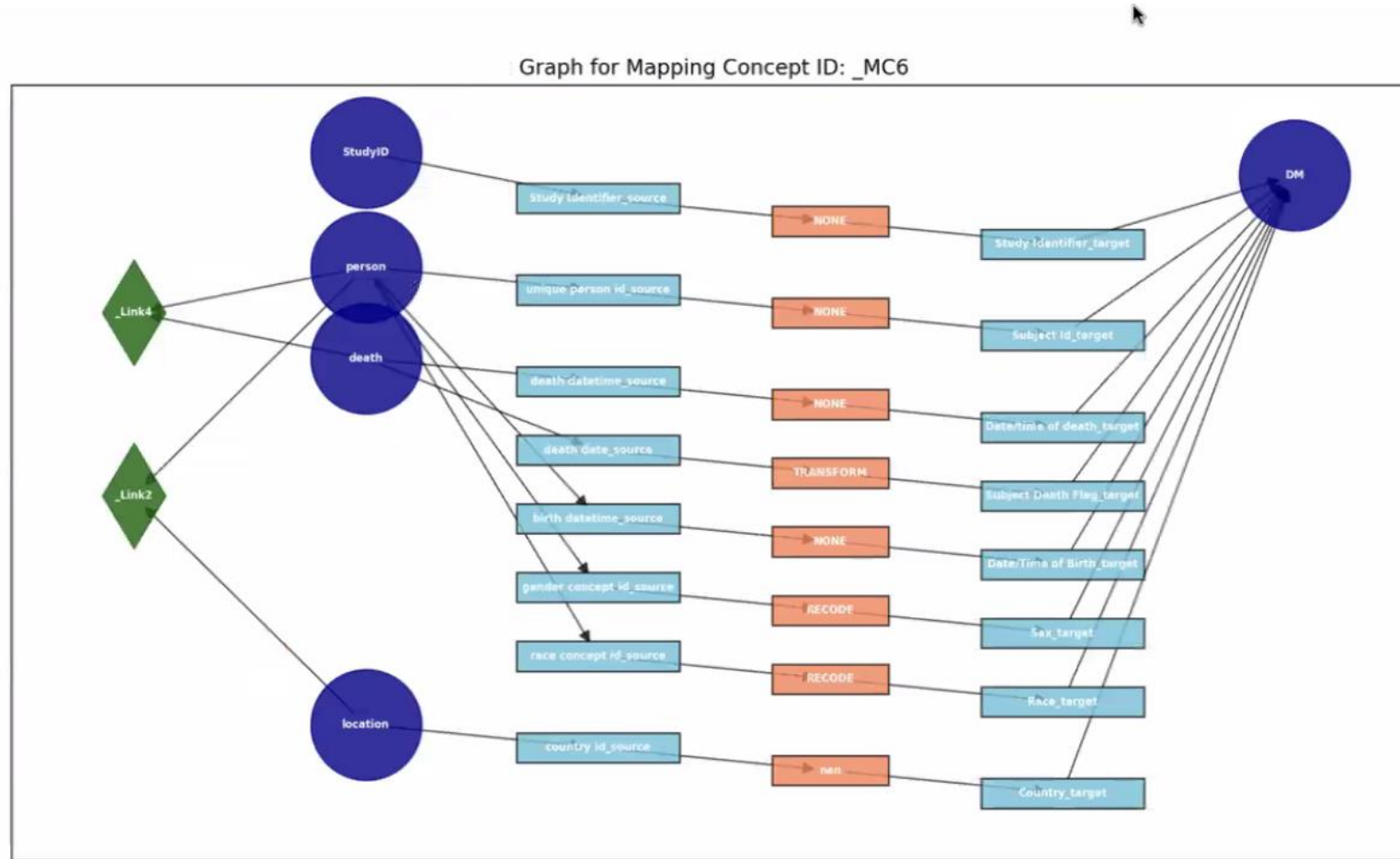
- Expect the worst of your source data
- Build in checks for the unexpected (e.g. incomplete or inconsistent data, out-of-range values, unpredictable data, etc.)
- Alerting for mismatches in code systems and terminologies (e.g. values that are not include in the code list)
- In summary: making robust programs



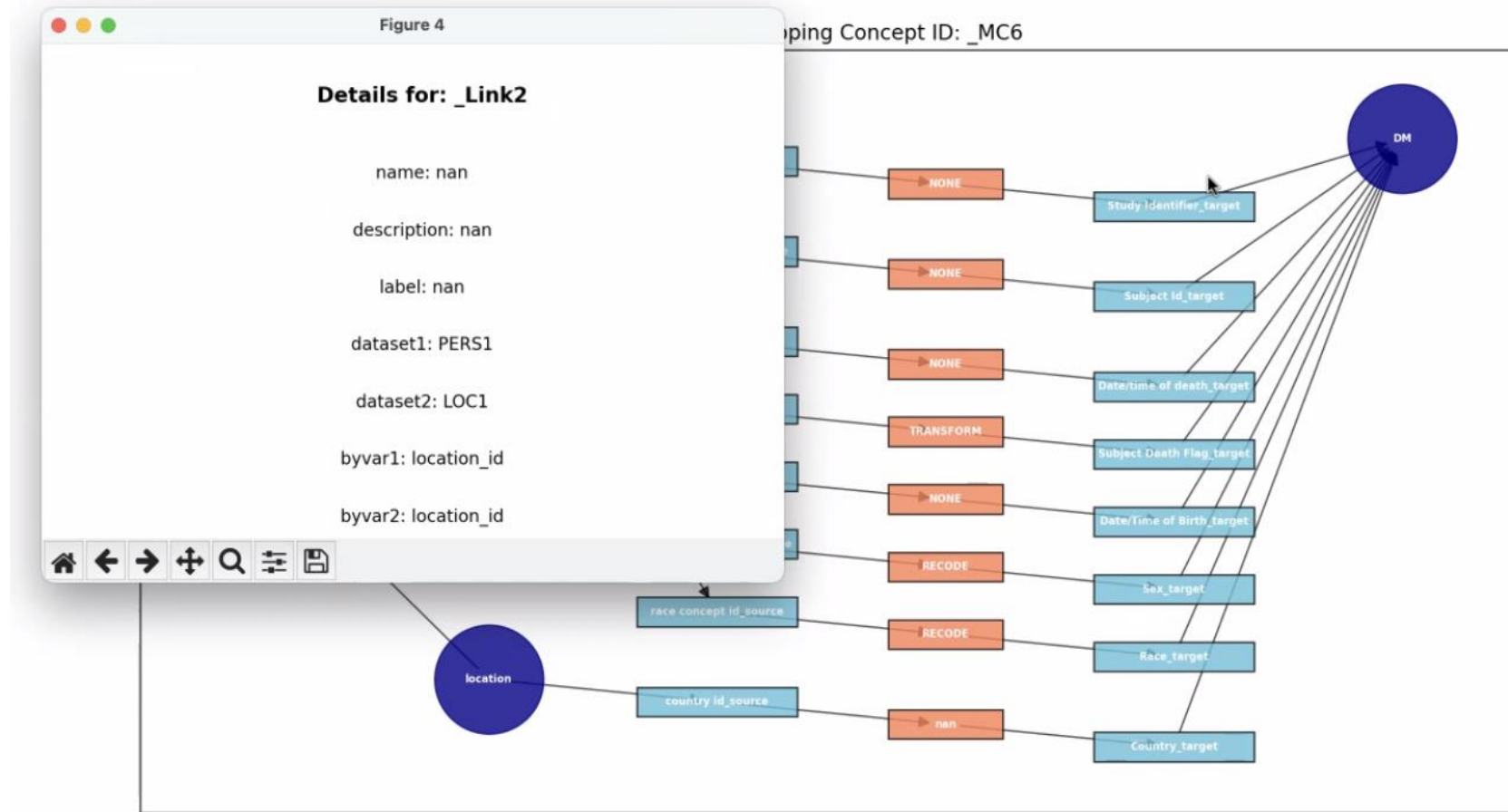
Lineage tool output



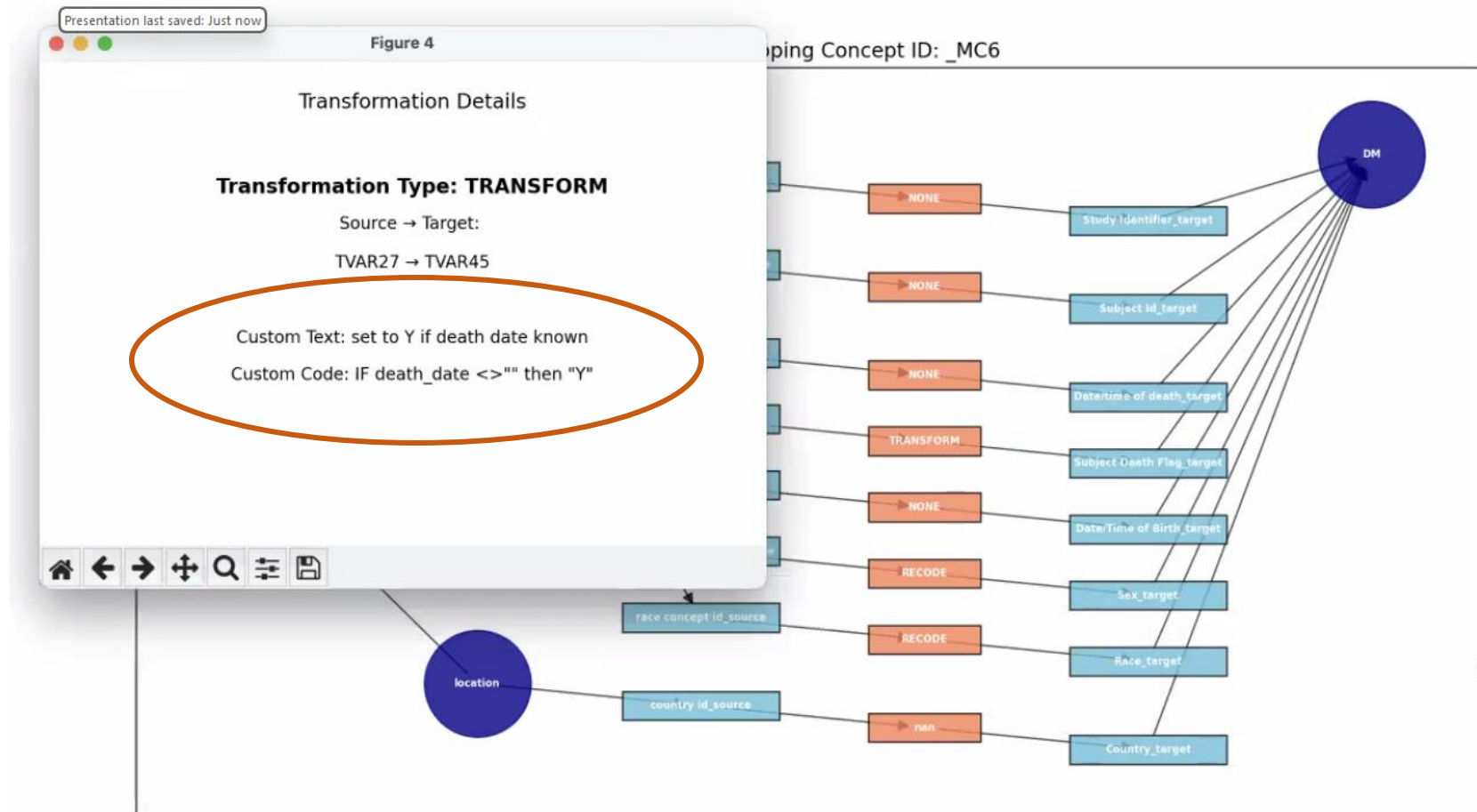
Lineage tool output



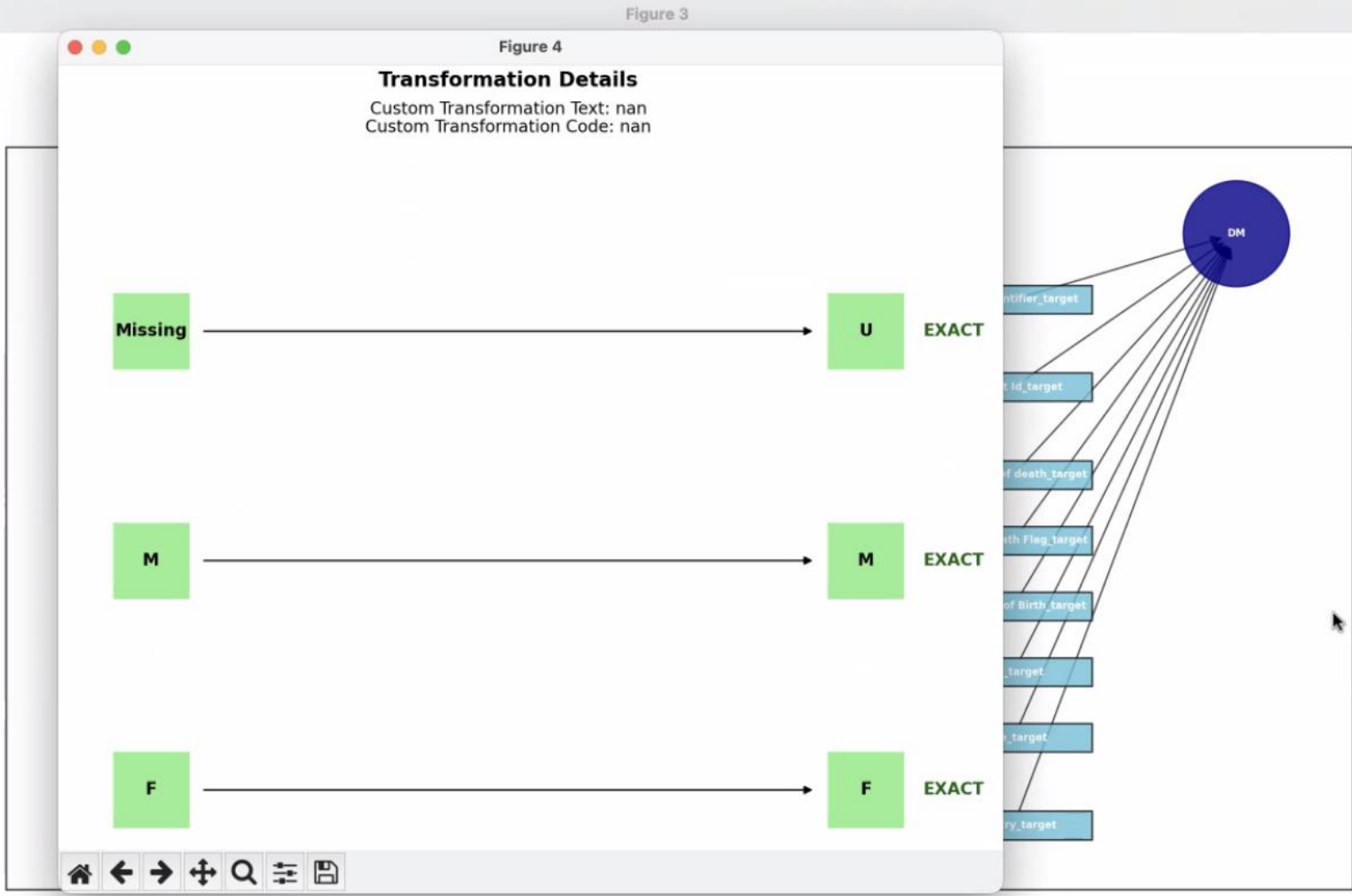
Lineage tool output



Lineage tool output



Lin



Next steps

- Lineage tool
 - Improve presentation
 - Add more mapping details
 - Make open source and available via Github
- Mapping
 - Add more OMOP domains to drive other SDTM domains
 - Add more SDTM domains
 - Add USDM as input to drive the process
 - Align with CDISC RWD lineage team on mapping logic and output
- Test
 - Test on other indications

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