

# A systematic review process enhances the reliability and utility of datasets

*Title: Lessons Learned from EHDEN Data Partner Reviews: Improving ETL Processes and Data Quality in OMOP CDM Conversions*

**Background:** As part of the European Health Data & Evidence Network (EHDEN) initiative to transform data sources to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), data partners undergo milestone assessments to evaluate their progress in implementing the ETL (Extract, Transform, Load) process and ensuring data quality.

Figure 1: Tools utilized in systematic review

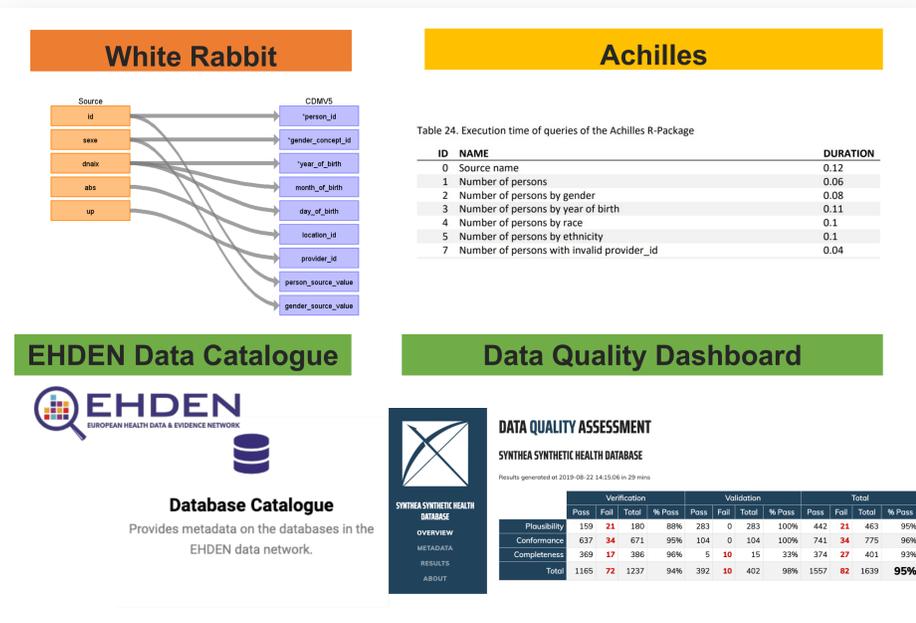
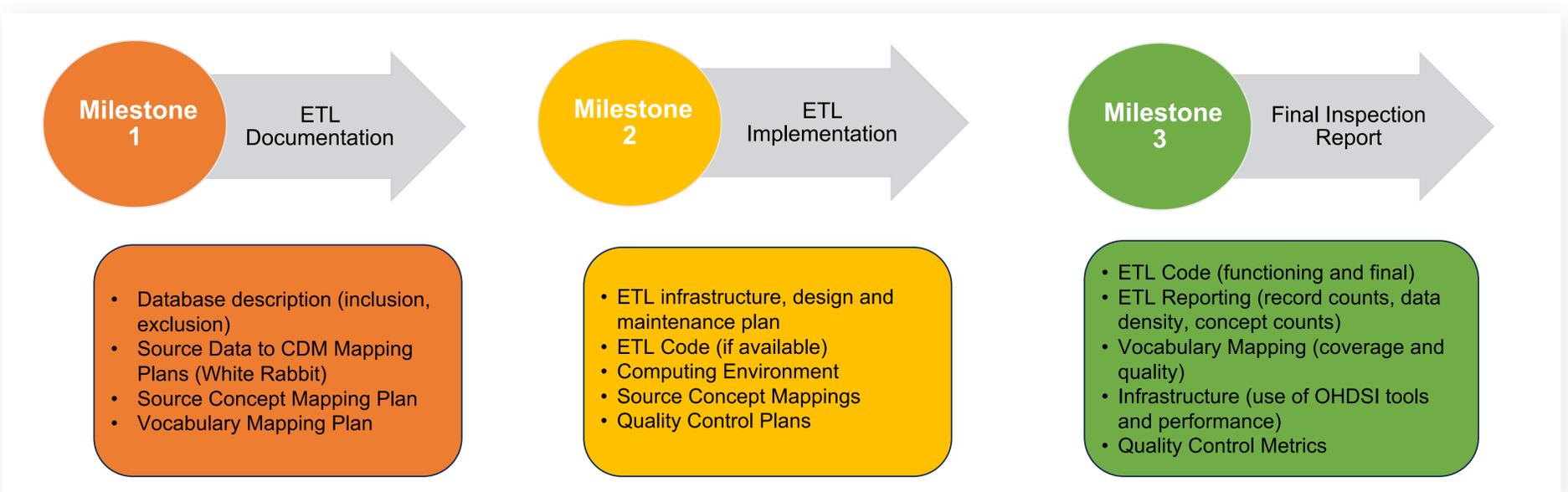


Table 1: Examples of identified areas of improvement

Improvement Area	Reviewer Suggestions
Data Loss Evaluation	Implement a process to compare source counts to target counts
Source Code Mapping Coverage	Evaluate unmapped source codes, retain valid records in the CDM and iterate on mappings over time
Source Code Mapping Quality	Evaluate the accuracy of source code mapping and implement a process to detect deprecated and/or updated codes
Implausible Records	Review potential erroneous records with future or implausible dates
Data Model Conformance	Review failures observed for data model conformance and ensure alignment with CDM specifications

Figure 2: Description of Milestone Assessment



**Key Takeaway:** The identified areas for improvement in the review process highlight the complexity and challenges inherent in transforming diverse healthcare datasets into a standardized format. By addressing the identified issues and improving data quality, the data partners enhanced the reliability and utility of their datasets for healthcare research and analysis within the EHDEN network.

