

# PREPARE project uses OHDSI tools for privacy-preserving clinical prediction

## A Unified ETL Strategy for Nine Clinical Datasets

### Experiences with OMOPing Rehabilitation Data

#### A Unified ETL Strategy for Nine Distinct Clinical Datasets

**Background:** The Rehabilitation research relies on real-world observational data to model the patients' progression, to stratify populations, and to support personalized care pathways. Different rehabilitation organizations have heterogeneous data, ranging from clinical measurements and PROMs to functional assessments and free-text clinical information. The PREPARE Project addresses these challenges across nine rehabilitation clinical. PREPARE is implemented aiming at integration of heterogeneous rehabilitation datasets, using a unified analytical framework to support predictive modelling and personalized rehabilitation planning. This achieved using OMOP Common Data Model and standardizing rehabilitation data using a unified ETL strategy and tool-supported methodology, highlighting how OHDSI standards were used to support rehabilitation research needs rather than generic data integration

#### Results - OMOP Standardization Progress

Clinical cases Completed to-date:  6 / 9 (67%)

ETL Iterations:  3-5 cycles per partner

Key Challenges:  High (heterogeneity, vocab gaps)

The unified ETL environment facilitated early identification of semantic inconsistencies and supported informed decisions on data inclusion, transformation, or exclusion, securing consistency on data quality principles.

#### Conclusion

The PREPARE project illustrates that the **OMOP CDM** can successfully facilitate rehabilitation research **when implemented through a Unified ETL Framework**. This approach incorporated the ETL tools to transform and unify rehabilitation data, **enabling Clinicians-Centered Mapping**, and embedded **Data Quality** principles, leading to a complete **OHDSI Alignment** in guiding targeted rehabilitation research.

**Unified ETL framework and tools bridges data standards and real-world rehabilitation practice**

#### Methods

#### Unified ETL strategy for rehabilitation data standardization



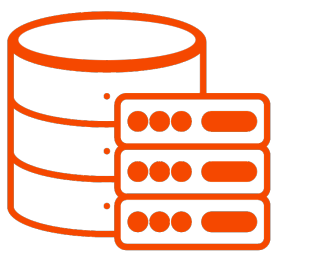
#### Data Context - Heterogeneous rehabilitation datasets

- Clinical data
- PROMs
- Functional assessments
- Free-text records



#### Unified ETL Strategy - Aligned with EU data quality principles:

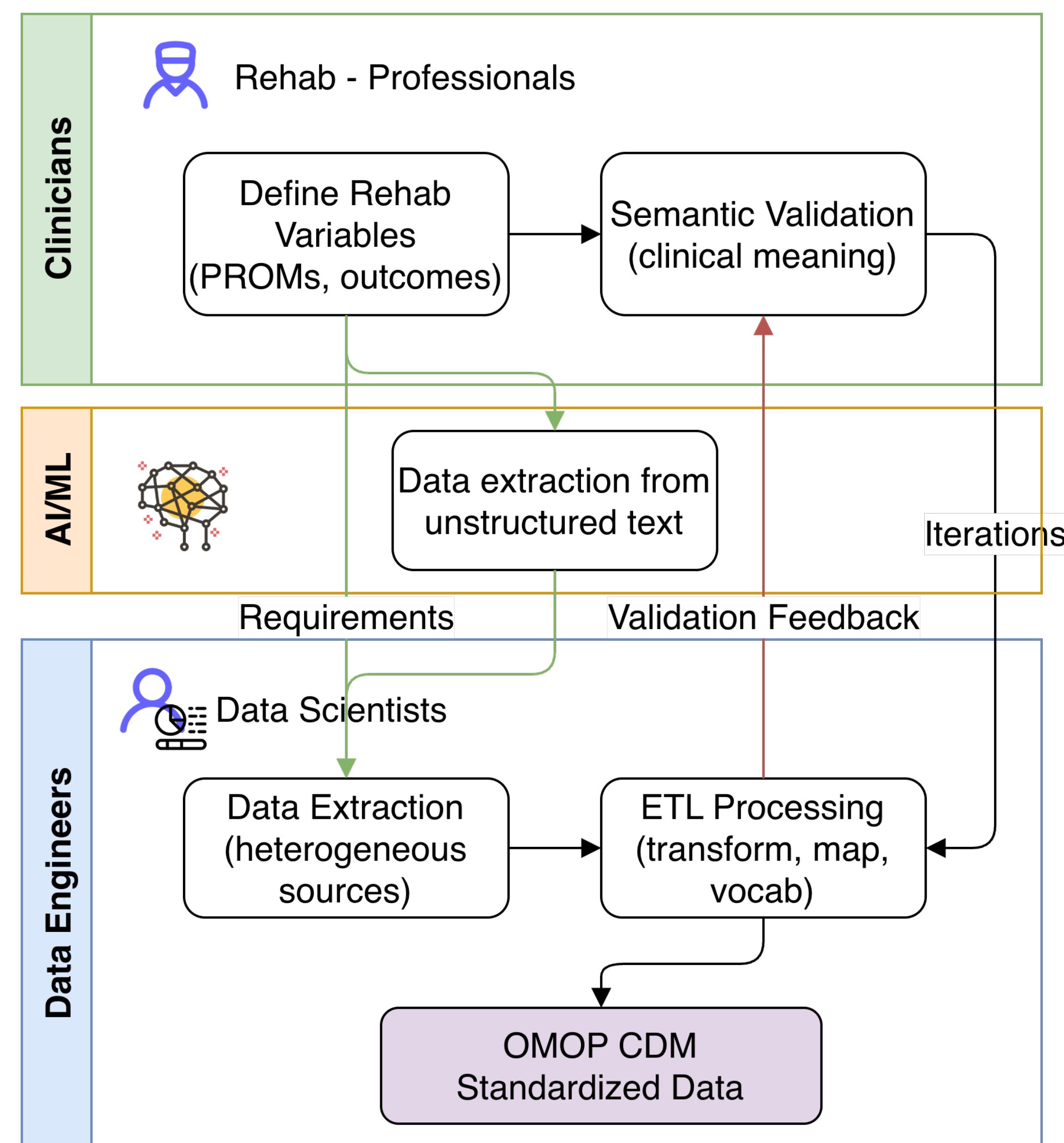
- Relevance
- Coherence
- Completeness
- Secondary-use utility



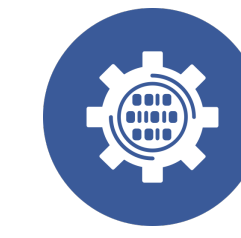
#### Minimum Viable OMOP Dataset - Defined per clinical case:

- Clinicians + Data Scientists Focus on:
- Predictive variables
  - Functional outcomes
  - Temporal events

#### Workflow



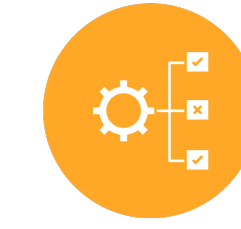
#### Tool & Methodology



#### Unified ETL Tool

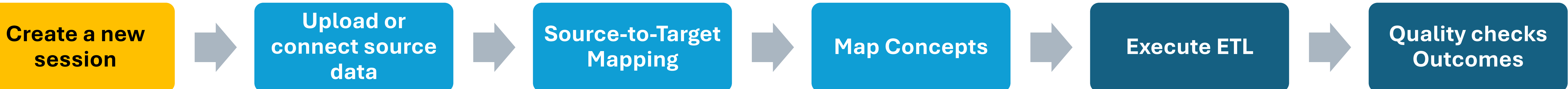
Supports:

- Schema mapping
- Data extraction & transformation
- Vocabulary mapping (ATHENA)
- Validation & iteration



#### Methodological Features

- Iterative process → session-based ETL
- Semantic mapping → ATHENA + clinician refinement - Custom concepts where needed
- Data quality feedback → OHDSI-aligned checks
- Traceability → Transparent & reusable mappings



**OMOP ETL Tool** Dashboard

**ETL Sessions Dashboard**

Session Name	Description	Created	Open	Delete
auth	No description	2025-03-18 08:45	Open	Delete
test	No description	2025-03-18 08:37	Open	Delete

**Create New ETL Session**

Session Name: test

Description: Spaces and special characters will be automatically converted to underscore. This will be used for database schema names.

What will be created:

- Dashboard schema: dashboard\_test
- Source data schema: source\_data\_test
- Column mapping schema: column\_mapping\_test
- Concept mapping schema: concept\_mapping\_test
- OMOP CDM schema: omop\_cdm\_test

Since multiple iterations during the data transformation will be applied, ETL environment provides creation and reusability of multiples sessions.

**Map Columns to OBSERVATION\_PERIOD**

INSTANCE NAME: person

SOURCE TABLE: pilot\_table1\_person

Existing Instances: No existing instances for this table

OMOP CDM Target Fields:

- observation\_period\_id \*
- person\_id \*

The Field Mapping interface allows you to create detailed mappings between your source data and OMOP CDM tables (drag & drop fields).

**Map Columns to PERSON**

INSTANCE NAME: person

SOURCE TABLE: pilot\_table1\_person

Existing Instances: 5 mappings

Source Columns:

- person\_id
- gender\_concept\_id
- year\_of\_birth
- month\_of\_birth
- day\_of\_birth

OMOP CDM Target Fields:

- person\_id \*
- gender\_concept\_id \*

If CDM required columns do not exist in source files (e.g. race, ethnicity etc.) the user can type a custom concept (e.g customRace) to later assign a static concept value.

**Concept Mapping**

1. Upload OMOP Vocabulary

2. Extract Values for Mapping

3. Concept Mappings

Source Value	Instance Name	Field	Domain	Mapped Concept	Actions
0	person	sex_concept_id	sex	0	No matching concept
M	person	gender_concept_id	gender	8507	Male
Outpatient Visit	visit	visit_concept_id	visit	9202	Outpatient Visit
32817	visit	visit_type_concept_id	visit	32817	Outpatient Visit

The concept mapping is used to map the source values to OMOP standard concepts, based on the USAGI tool.

It requires to upload the concepts' vocabulary from ATHENA. The tool proposes the most suitable concepts and the user can check, refine or change the suggestions

Petros Patias, Charalampos Georgiadis, Themistoklis Roustanis, George-Robert Patias, Alexandros Zacharegkas and the PREPARE Project Group