We made an ETL pipeline to convert rare disease data (CDE) to OMOP CDM

Objective

Rare Disease registries → CDE Semantic Model → OMOP CDM v5.3.1 → Patient registry

Methods

Terminology mapping
- Manual mapping
- SNOMED CT - 21, LOINC - 6, OMOP - 2, HCPCS - 1
- SKOS → relatedMatch

Data transformation
- Load RDF into a triplestore
- Use SPARQL to generate OMOP CDM compatible tables

Apply on use-case
- Apply the mapping to an ERN registry

ETL Pipeline

CDE RDF Files → SPARQL 1.1 Query → Triplestore → Terminology mappings RDF → Person id 420, Gender_concept_id 8532, Year_of_birth 1969

RESULTS:
- Terminology Mapping
  - 52 CDE terms to 30 Athena terms
  - SNOMED CT - 21, LOINC - 6, OMOP - 2, HCPCS - 1
  - SKOS → relatedMatch
- ETL Pipeline
  - Use SPARQL to generate OMOP CDM artefacts
  - Use curated mappings for transformation
- Future work:
  - Finalize and validate with use-cases from rare disease registries that collaborate with the EJP RD

Available resources:
- Link to the EJP RD website
- Link to the CDE semantic model on GitHub