

Transforming England Admitted Patient Care Data to the OMOP Common Data Model Using PIANO

Standardising England's CUREd+ Health Data for Research

Background

- The **CUREd+ Database** is a national dataset for England, with a population of 57 million people (2011 to 2023)
- It's composed of **11 different health datasets** that were linked using NHS England's patient tracing.
- All required **NHS ethics and Health Research Authority (HRA) approvals** are already in place for researchers.



CUREd+ An Urgent and Emergency Care Database

Problem

- Lack of data standardisation limits the use of clinical databases and their integration within secure data environments.
- It also restricts the applicability of this data for collaborative analysis and federated research projects.

Solution

- We're adopting the **Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM)**.
- This model provides a standardised structure for real-world health data.

Approach

- We used Evidentli's PIANO platform¹, an AI-supported Extract-Transform-Load (ETL) tool, to map England's Admitted Patient Care² (APC) data to the OMOP CDM 5.4.
- This dataset, one of NHS England's largest and most frequently requested, consists of 51GB in binary format (260GB in CSV format).



Secure and Validated Data Transformation

Security: We performed a risk assessment to securely install PIANO within our secure data environment (SDE).

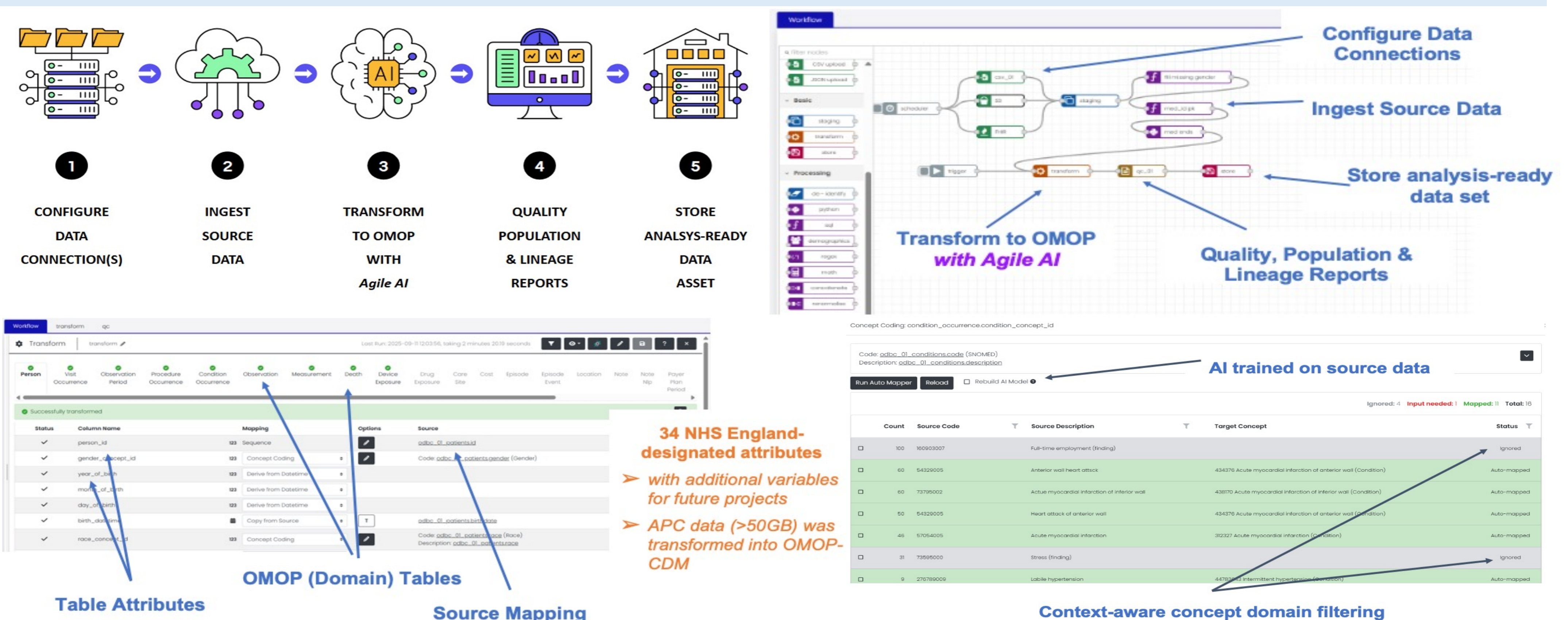
Three-Phase Pipeline to map APC dataset

- **Synthetic Data:** Designed the mapping pipeline using a synthetic APC dataset in PIANO's external hub.
- **Test Data:** Deployed PIANO in our SDE and tested the pipeline on a one-year subset of real APC data to check compatibility and resource needs.
- **Full Transformation:** Applied the pipeline to the entire APC dataset, processing it in three batches to manage the large scale. We focused on mapping **34 NHS England-designated OMOP-CDM attributes**³, plus additional variables for future collaborations.

Testing and Validation

- We used PIANO's integrated quality check tools to ensure the integrity and correctness of our resulting APC-OMOP database.
- To validate the reliability of our process, we replicated a previous study's analysis on a matched patient cohort.

Piano's Workflow for Data Transformation



Key Results and Conclusion

Data Processed: We standardised **234 million records** (260GB in CSV format) of APC data from 2011 to 2023.

Efficiency: The PIANO platform converted our APC data to OMOP-CDM in three months (1.25 FTE), which is faster than the estimated average of **6-9 month**. This duration includes time spent on the initial learning curve and overcoming technical challenges.

Automation: PIANO's **AI-auto-mapper** handled most of the vocabulary standardisation, with only a few requiring manual coding.

Conclusion

- We successfully and rapidly transformed large-scale APC data, overcoming challenges with data sensitivity and limited resources.
- The standardised data was integrated into the Yorkshire & Humber SDE, unlocking its potential for **federated health data research**.
- This project establishes a clear best practice, showing how automated tools like PIANO can help the NHS adopt the **OMOP model more widely** and enable more extensive and collaborative research.