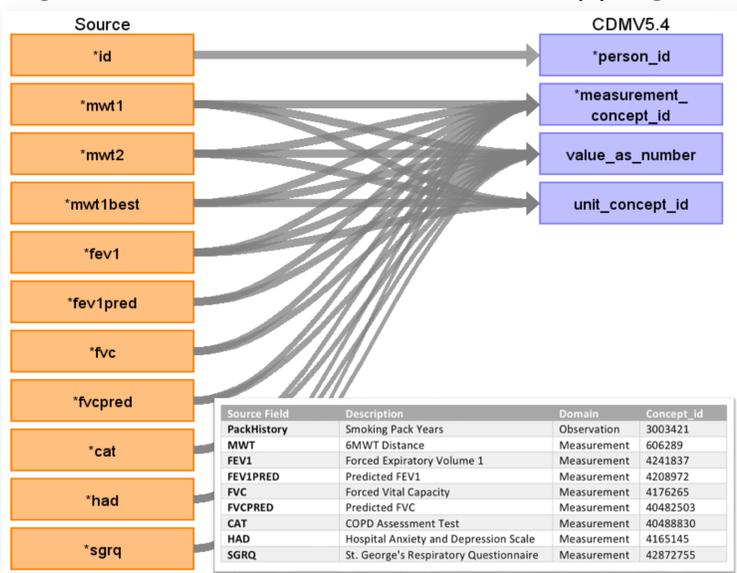


Open research datasets harmonized to the OMOP CDM can be used to evaluate COPD AI models.

Title: *Harmonizing Public Research Access Datasets for AI Advancements in Asthma/ COPD Diagnosis.*

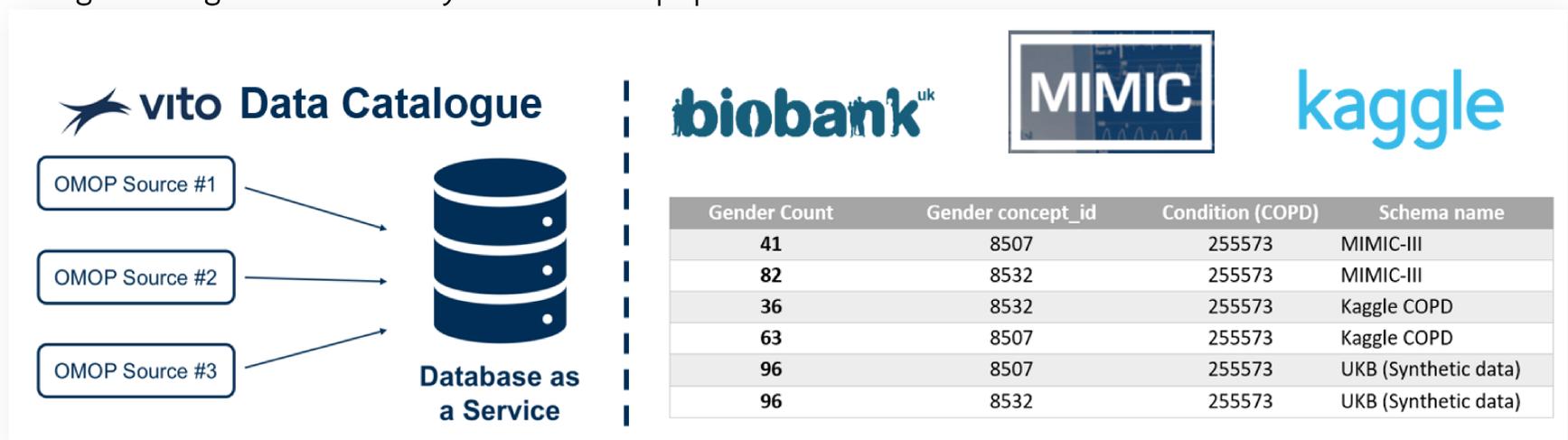
Background: Chronic Obstructive Pulmonary Disease (COPD) is a significant global health concern, causing substantial morbidity and mortality worldwide. With millions affected, COPD necessitates advanced diagnostic and therapeutic approaches. Leveraging **Artificial Intelligence (AI)** on standardized datasets offers promise in enhancing COPD management. Our study focuses on **evaluating three research datasets (Kaggle COPD datasets, MIMIC3, UK Biobank)**, harmonized to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), to develop and assess advanced AI solutions for COPD. Traqbeat, a medical technology company, provided two indicative COPD AI models for demonstration, illustrating the benefits and challenges of AI model training, testing, and assessment with OMOP.

Figure1: Structural and Semantic mapping.



The mapping approach undertaken in this study involves both structural and conceptual mapping to harmonize COPD-related datasets to the OMOP CDM (Figure 1) ensuring **consistency and interoperability**. The study focused on mapping specific **clinical measurements** and variables relevant to COPD diagnosis and management. Additionally, the implementation of **Extract, Transform, and Load (ETL)** processes was crucial in translating these mappings into practical implementation.

Figure2: High level summary of the COPD population in the VITO database.



Conclusion: Harmonizing datasets to OMOP ensures consistency in clinical data representation, facilitating cross-database analyses (Figure 2) and promoting AI solutions for COPD diagnosis and management. By utilizing public research datasets, we eliminate complex data access agreements, fostering collaboration and encouraging broader participation in developing solutions for COPD research



Frédéric Jung¹, Vangelis Sakkalis², Chang Sun³, Mahmoud Ibrahim³, Gökhan Ertaylan¹

1) VITO, Vlaamse Instelling voor Technologisch Onderzoek, Mol, Belgium
 2) Traqbeat technologies, Heraklion, Greece
 3) Institute of Data Science, Maastricht University, Maastricht, the Netherlands