SciDataCon 2025



Contribution ID: 112

Type: Presentation

Metadata Meets Standardization: Leveraging a Staging Database to Integrate African Longitudinal Mental Health Data into OMOP CDM 5.4

Wednesday 15 October 2025 14:55 (11 minutes)

Background: Longitudinal studies are necessary for tracking the progression of mental health disorders such as depression, anxiety, and psychosis. However, the integration of diverse mental health data from different sources and waves—especially in low- and middle-income countries—remains complex due to variability in instruments, socio-cultural expressions, and data structural formats. This study introduces a novel staging database framework developed under the INSPIRE Mental Health (INSPIRE MH) project, which bridges standard OMOP CDM concepts and non-standard INSPIRE concepts, using a harmonized metadata-driven approach. The database supports longitudinal, multi-source mental health datasets while preserving sociocontextual detail that would otherwise be lost in direct OMOP CDM conversions.

Methods: The staging database is structured using the DDILifecycle metadata standard and follows a snowflake schema, enabling seamless documentation of datawaves, instruments, instrumentitems, and social context variables across individual and household levels. This schema allows data from HDSS populations (Kilifi, Iganga Mayuge, Kagando) and secondary data accessed from African researchers to be captured uniformly. A dynamic ETL pipeline was created using R to map the data from the source to the staging database, in preparation for mapping to the OMOP CDM while preserving variable traceability. Importantly, the database accommodates non-standard vocabularies (INSPIRE Concepts) such as religion, income sources, and household size—variables essential for contextualizing mental health in African settings.

Results: Out of 14 datasets targeted for migration, 10 have been successfully transformed from the source to the staging database, representing over 163,000 individuals. The staging database allows for a preliminary understanding of the trends of the three outcomes of interest. The SD facilitates standardized ETL, has helped reduce data loss, and provides an intermediate layer for data visualization and harmonization quality checks. Automated quality assessments were run using the Data Quality Dashboard (DQD) in R, ensuring fidelity between source and staging database layers. This architecture also allows experimentation with HDSS-linked datasets and non-coded mental health concepts.

Conclusion: The INSPIREMH staging database represents a significant advancement in the standardization and integration of longitudinal mental health data across heterogeneous African contexts. It is more than a pre-processing step; it acts as a living metadata repository and research-ready platform that captures both clinical and socio-contextual variables. By combining the strengths of DDI Lifecycle, OMOP CDM, and a novel staging model, this architecture sets a precedent for future mental health data science pipelines in LMICs, ensuring both interoperability and local relevance.

Primary authors: Dr KIRAGGA, Agnes (African Population And Health Research Center); MUGOTITSA, Bylhah (African Population And Health Research Center)

Co-authors: Ms MAILOSI, Dorothy (Committee on Data of the International Science Council- CODATA); Dr OMONDI, Evans (African Population And Health Research Center); Prof. GREENFIELD, Jay; Prof. TODD, Jim (Catholic University of Health and Allied Sciences); Ms ANDESO, Pauline (African Population And Health Research Center); Mr MOMANYI, Reinpeter

Presenter: MUGOTITSA, Bylhah (African Population And Health Research Center)

Session Classification: Presentations Session 7: Open research through Interconnected, Interoperable, and Interdisciplinary Data

Track Classification: SciDataCon2025 Specific Themes: Open research through Interconnected, Interoperable, and Interdisciplinary Data