Research Data Warehouse / Datamart

Over the past several years, the Krembil Centre for Neuroinformatics has made significant improvements to the CAMH research data management platform, advancing our mission of using big data, AI, and brain modelling to change our understanding of mental health. The platform allows users to collect data using a standardized set of data capture tools, make connections in the integration layer, and distribute the knowledge through various channels. This enables researchers and clinicians to access unified data more easily. As the platform's capabilities have continued to develop, the team has updated and streamlined concordant operational processes, including online tutorials, and live workshops to help users navigate and access data.

Pictured left to right:
Nikola Bogetic, IT Specialist and **Ade el Ansari, Knowledge Engineer Lead**

KCNI Data Management Platform

Research data applications, including XNAT, LabKey, and REDCap, are used to collect and manage data of different modalities: medical imaging, genetic, and cognitive assessments. To support the need for interoperability, cross-database queries, and open science, the team has developed an integration layer that can also access restricted records from the CAMH electronic medical record.

The data from the above applications flow into the integration layer, adhering to quality control and standardization processes. The data are transformed to be compliant with HL7 FHIR - a global standard for describing and exchanging healthcare data. The use of HL7 FHIR ontology allows users to have a common vocabulary for querying and analyzing the data. This provides a shared understanding of the data for our internal users, as well as the open science community.

The unified data are loaded into the CAMH Knowledge Graph (Blue Brain Nexus), which provides core capabilities to serve high-quality data. In addition to typical database operations, the knowledge graph ensures validation and compliance with the HL7 FHIR ontology, manages data provenance, and administers user privileges. Once the data are in the knowledge graph, they can be accessed using graph queries, full-text search, and more recently, through data marts.

Data marts are an easy way to provide researchers, clinicians, and BI/Analytics teams with direct access to the required data cuts. Data are refreshed automatically and are available as tables in a relational database that can be queried using SQL (Structured Query Language). The use of data marts has enabled tools like Clinical Decision Support dashboards, which allow clinicians to view patient's history and progression at a glance. With growing demand for data analytics, visualization, and machine learning, data marts will be a key solution to a variety of data requirements.

In the last 18 months, the KCNI team has integrated data for 12 research studies, one clinical pathway, and over 12,000 research participants and patients. For the next stage, the team will be focusing on transforming genetic and imaging data to be compliant with prominent ontologies. In addition, the clinical notes will be stored within the knowledge graph to support AI and ML analysis pipelines. We look forward to onboarding new projects and expanding the platform with additional features.