Contribution ID: **70** Type: **Talk (15min + 5min)**

Multimodal Imaging in Neuropsychiatric Disorders dataset (MINDset)

Thursday 27 February 2025 11:40 (20 minutes)

Precision psychiatry faces significant challenges, including limited sample sizes and the generalizability of findings, variability in clinical phenotyping, and the need for robust biomarkers to guide personalized treatment approaches. Additionally, the integration of diverse data sources—such as multi-omics, electrophysiology, neuroimaging, clinical records, and cognitive-behavioural data—adds complexity to research efforts. To address these challenges, robust methodologies for data management, quality control, and interoperability are critical. In line with this, the Multimodal Imaging in Clinical Neuroscience research group has collected a comprehensive range of data from diverse studies, encompassing multimodal imaging (fMRI, EEG, fNIRS, PET), omics (microbiome, epigenetics), and cognitive and neuropsychiatric test scores from both psychiatric patients and healthy individuals. However, with approximately 800 datasets scattered across multiple storage systems, the lack of standardized and centralized data and storage poses a barrier to cross-modality analyses and large-scale studies. A unified approach to data management is essential to overcome these obstacles and advance precision psychiatry.

The Multimodal Imaging in Neuropsychiatric Disorders dataset (MINDset) project addresses significant challenges in the integration and analysis of complex datasets within neuropsychiatric research by developing a comprehensive database housing multimodal neuroimaging and electrophysiology data with metadata, and pre-processed features, organized in accordance with the Brain Imaging Data Structure (BIDS) (Gorgolewski *et al.* 2016) format. Additionally, a user-friendly query tool to facilitate data access is also being developed. The MINDset web server and database are hosted on a secure, password-protected cloud environment managed by the IT services of FZ Jülich, utilizing the Kubernetes architecture (Burns *et al.* 2016). This setup leverages Kubernetes container orchestration capabilities to ensure efficient management, scaling, and high availability of the service.

The metadata file containing demographics, cognitive, neuropsychological and psychopathological assessment results originally in Microsoft Excel format were incorporated into JSON file structure using Python scripts (Python Software Foundation 2024). Each row is parsed into values corresponding to keys from the first row, creating a dictionary, which is finally written into a JSON file. The nested JSON format preserves the hierarchy of the structured data, making it easy to access and ensuring accurate data representation.

The BEAVERDAM tool (More H and Denker M 2022) was used to build the database from the created JSON files. This tool facilitates the integration of multiple metadata files, structured in a JSON-like format, into a unified, browsable database repository. The BEAVERDAM tool utilizes MongoDB (Bradshaw *et al.* 2019) for storage, enabling more data scalability and flexibility, as well as supporting flexible deployment in cloud-based environments. The query tool enables users to apply filters across a range of parameters, dynamically updating various visualization charts, statistical data, and tabular representations.
Centralizing and standardizing our data via MINDset will greatly boost research capabilities in precision psychiatry, unlocking insights at the intersection of multiple modalities. Looking ahead, this unified approach to managing and integrating multimodal data can serve as a model for other researchers and facilities, fostering broader adoption and enhancing collaborative efforts across the field.

Acknowledgements: We thank Heather More and Volker Hofmann for their invaluable support on the project.

I want to participate in the youngRSE prize

no

Primary author: Dr RAJKUMAR, Ravichandran (Institute of Neuroscience and Medicine 4, INM-4, Forschungszentrum Jülich, Germany)

Co-authors: Mr ALI, Syed (Institute of Neuroscience and Medicine 4, INM-4, Forschungszentrum Jülich, Germany); Ms HEINEMANN, Beke (Institute of Neuroscience and Medicine 4, INM-4, Forschungszentrum Jülich, Germany); Ms WATHARE, Sanskruti (Institute of Neuroscience and Medicine 4, INM-4, Forschungszentrum Jülich, Germany); Mr RANDENI KORALALAGE, Lochana Chathuranatha (Institute of Neuroscience and Medicine 4, INM-4, Forschungszentrum Jülich, Germany); Ms RAMKIRAN, Shukti (Institute of Neuroscience and Medicine 4, INM-4, Forschungszentrum Jülich, Germany); Dr ZIMMERMANN, Markus (Institute of Neuroscience and Medicine 4, INM-4, Forschungszentrum Jülich, Germany); Prof. SHAH, N.Jon (Institute of Neuroscience and Medicine 4, INM-4, Forschungszentrum Jülich, Germany); Prof. NEUNER, Irene (Institute of Neuroscience and Medicine 4, INM-4, Forschungszentrum Jülich, Germany)

Presenter: Dr RAJKUMAR, Ravichandran (Institute of Neuroscience and Medicine 4, INM-4, Forschungszentrum Jülich, Germany)

Session Classification: Visualization with Research Software

Track Classification: Policies and Community Building: open source research software