Contribution ID: **43**   Type: **Talk**

A Transatlantic Infrastructure to Share and Process BigBrain Datasets

Tuesday 28 October 2025 12:30 (15 minutes)

The demand for interoperable data processing in neuroscience underscores persistent challenges in cross-border data sharing, secure access to distributed resources, and the portability of tools across organizations. Neuroscience datasets are particularly sensitive due to privacy regulations and their size and diversity, which complicate collaborative research. These requirements motivate the development of approaches that ensure secure access, portability, and reproducible workflows while harmonizing data interoperability across infrastructures. A promising solution lies in the "bring compute-to-data" paradigm, where applications are executed on local infrastructures without moving sensitive data, thereby fostering compliance with jurisdictional and institutional requirements.

Within this context, the BigBrain dataset and related ultra-high-resolution neuroanatomical resources provide a unique test case. These data are of great scientific value but demand specialized processing pipelines and scalable computing environments. European research infrastructures such as EBRAINS have advanced services for data access and HPC, yet integrating analysis workflows seamlessly across borders remains complex. The CBRAIN platform is a federated neuroinformatics environment for distributed computing and data management and offers complementary strengths: a flexible plugin-based tool integration model, user-friendly interfaces, and federated access to heterogeneous compute and storage resources. CBRAIN emphasizes cross-institutional interoperability and secure, auditable execution, making it well-suited for international collaborations around BigBrain.

The presentation will demonstrate the successful adoption and extension of two neuroscience applications on CBRAIN, deployed to resources at the Jülich Supercomputing Centre (JSC): HippUnfold, for automated hippocampal subfield segmentation, and a Cell Detection workflow for imaging analysis.
To enable this, CBRAIN's security model was extended to support service accounts for automated but controlled access, a critical feature for scaling collaborative analysis. In parallel, deployment of CBRAIN components was customized to align with JSC's policies and infrastructure requirements. Datalad was central to this integration, providing dataset versioning, provenance tracking, and reproducible data management across CBRAIN and JSC systems (Figure 1). This ensured seamless synchronization of BigBrain-derived datasets and enabled transparent workflows that can be rerun or extended by collaborators.

The results demonstrate the portability of the HippUnfold and Cell Segmentation pipelines, executed efficiently by CBRAIN across borders, with consistent outputs on BigBrain data. These use cases confirm that CBRAIN provides a robust framework for secure data access, tool sharing, and reproducible processing. More broadly, this work illustrates a pathway to harmonize access to computational resources internationally, enabling scalable neuroscience research that leverages high-value datasets such as BigBrain.

Primary authors: CARON, Bryan (McGill University); MEHMON, Mohammad Shahbaz (Forschungszentrum Jülich)

Co-authors: EVANS, Alan (McGill University); RIEDEL, Morris (Forschungszentrum Jülich and University of Iceland); BECK, Natacha (McGill University); GLATARD, Tristan (Centre for Addiction and Mental Health)

Presenter: CARON, Bryan (McGill University)

Session Classification: Session 1: Multiscale Data Integration & AI-based Processing