International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Learnings from Building Neuroinformatics Pipelines

Suyash Bhogawar

Health Informatics Solutions Architect, Rackspacem USA; Manasi Weginwar, California State University Eastbay, USA

Abstract: ***Neuroinformatics plays a pivotal role in advancing our understanding of the brain by integrating various data acquisition modalities and facilitating the dissemination of these datasets through data sharing platforms [1]. This paper presents an overview of the neuroinformatics pipeline, detailing the workflow from data acquisition to data sharing from my experience setting it up at the new Cornell MRI Facility and learning from being a data curator at openneuro [19]. It discusses the evolution of data acquisition methods, standard data processing procedures, and the emergence of cloud-based data sharing platforms. Furthermore, a comparison of prominent data sharing tools, including XNAT and Flywheel [18], is provided to highlight their respective features and functionalities.***

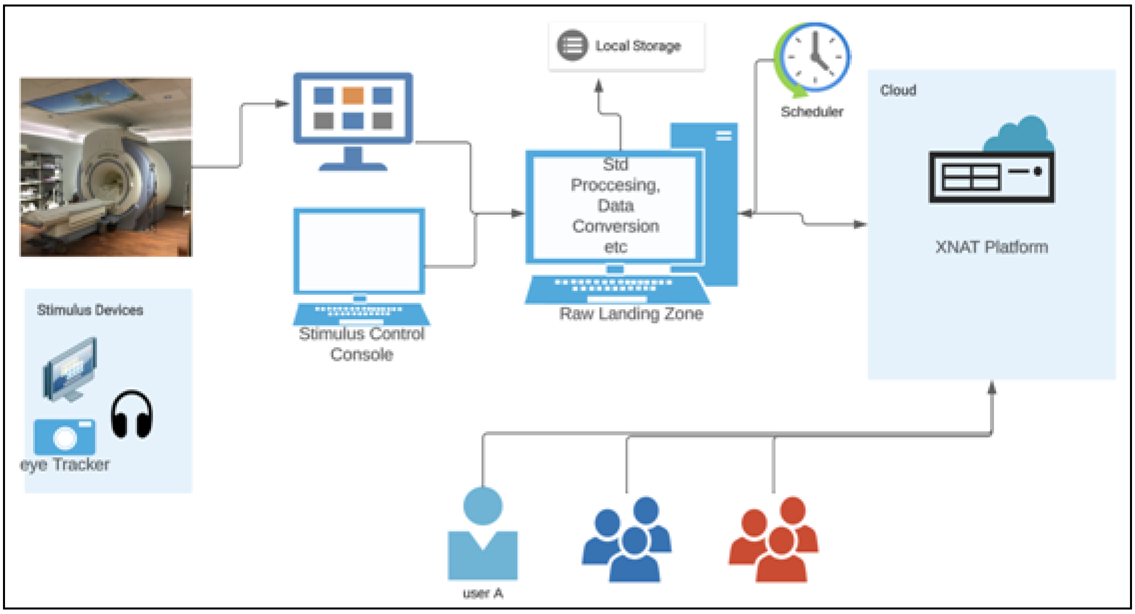
Keywords: Neuroinformatics, Brain Data Acquisition, Data Sharing Platforms, MRI Facility, Cloud-based Data Sharing

1. Introduction

Neuroinformatics involves the integration of neuroscience, informatics, and computational techniques to effectively manage and analyze complex neuroimaging data. This paper focuses on the comprehensive neuroinformatics pipeline, encompassing the entire spectrum from data acquisition to data sharing. The increasing availability of neuroimaging data has underscored the importance of efficient data acquisition, processing, and sharing to foster scientific collaboration and accelerate research progress.

2. Overall Workflow

The neuroinformatics pipeline comprises several stages, including data acquisition, preprocessing, analysis, and data sharing. This paper predominantly focuses on the initial stages of data acquisition and data sharing. Following diagram encapsulates various stages of data flow from data source to data platform. From the data platform, data captured at a research facility gets to respective labs and users.



3. Data Sources

3.1 MRI Scanner

Magnetic Resonance Imaging (MRI) scanners are fundamental instruments in neuroimaging studies, enabling the visualization of brain structure and function [20]. Over the years, advancements in MRI technology have led to improved spatial and temporal resolutions, resulting in more detailed and accurate images.

3.2 Physiological Devices

Physiological devices, such as electroencephalography (EEG)[21], magnetoencephalography (MEG) [21], and functional near-infrared spectroscopy (fNIRS) [22], provide complementary insights into brain activity. These devices capture neural signals and physiological responses, contributing to a comprehensive understanding of brain dynamics.

3.3 Stimulus Devices

Stimulus devices, ranging from visual displays to auditory systems, are essential for controlled experiments in neuroimaging. These devices elicit specific responses, enabling researchers to investigate brain activation patterns in response to various stimuli.[23]

4. MRI Console

MRI console is a specialized local workstation acting as control plane for the magnet. It processes data using manufacturer-provided software packages and allows technicians to view images. Data from the console is then pushed to the local compute for standard pre-processing and then pushed to the cloud (XNAT) [21].

Local compute resources are pivotal in neuroimaging, facilitating real-time processing and analysis of acquired data. The evolution of local compute capabilities has been transformative, enabling researchers to perform complex analyses with increased speed and accuracy.

5. Standard PreProcessing

Data preprocessing and analysis pipelines have evolved considerably. Previously, manual interventions were common in data preprocessing [6], introducing subjectivity and potential errors. Today, standardized preprocessing pipelines mitigate these issues, ensuring consistent and reproducible results across studies. Some of the preprocessing pipelines were for data format conversions such as dicom to nifti, denoising data, data reconstruction [8], meica etc.

To perform preprocessing in recent times, I would have gone with batch processing on cloud for ease and scalability of compute resources.

Though we had all standard processing applied to all datasets acquired, if we had to do it today, a more modular codebase could have been used, where a specific script shared by a lab or user can be plugged into the processing pipeline [12]. This will ensure adaptability of preprocessing pipeline per every researcher's requirements.

Another aspect which I would have focused on is data organization and standardization. I would have converted the dataset into BIDS or other data organization format at this juncture, which ensures adding all metadata at the source and data standardization at the organization level.

6. Cloud Data Sharing Platforms

6.1 Then vs Now

The emergence of cloud-based data sharing platforms has revolutionized how neuroimaging data is disseminated [18]. In the past, data sharing was often cumbersome due to limited storage capacities and slow data transfer speeds. Cloud platforms now offer efficient solutions, enabling researchers to store, share, and collaborate on large datasets seamlessly.

6.2 XNAT

XNAT (Extensible Neuroimaging Archive Toolkit) is a widely used open-source platform for managing and sharing neuroimaging data. It provides a flexible environment for storing diverse data types, metadata, and analysis results. XNAT's modular architecture allows customization to suit specific research needs [15].

6.3 Flywheel

Flywheel is another prominent cloud-based data sharing and management platform. It offers a user-friendly interface for organizing and sharing neuroimaging data, fostering collaboration among researchers. Additionally, Flywheel provides tools for data curation, quality control, and integration with various analysis pipelines [19].

6.4 Other Tool Comparison

Various other cloud-based data sharing platforms are available, each with its unique features and capabilities. Comparing these platforms based on factors like scalability, data security, and integration with analysis pipelines is crucial to selecting the most suitable platform for specific research requirements.

7. Conclusion

The neuroinformatics pipeline, encompassing data acquisition, preprocessing, and data sharing, is a critical component of modern neuroscience research. Advancements in data acquisition methods, local compute capabilities, standardized processing pipelines, and cloud-based data sharing platforms have collectively accelerated research progress and enabled global collaboration. As technology continues to evolve, the neuroinformatics landscape will continue to shape the future of brain research.

References

[1] Amunts, K., Ebell, C., Muller, J., Telefont, M., Knoll, A., & Lippert, T. (2016). The Human Brain Project: Creating a European Research Infrastructure to Decode the Human Brain. Neuron, 92(3), 574-581.
[2] Van Horn, J. D., & Toga, A. W. (2009). Human neuroimaging as a "Big Data" science. Brain Imaging and Behavior, 3(4), 350-361.
[3] Poldrack, R. A., Gorgolewski, K. J., & Varoquaux, G. (2017). Computational neuroimaging challenges to discover the brain's structure and function. Nature Neuroscience, 20(3), 349-353.
[4] Poline, J. B., Breeze, J. L., Ghosh, S., Gorgolewski, K. J., Halchenko, Y. O., Hanke, M., ... & Poldrack, R. A. (2012). Data sharing in neuroimaging research. Frontiers in Neuroinformatics, 6, 9.
[5] Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., ... & Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Scientific Data, 3, 160044.
[6] Kundu, P., Voon, V., Balchandani, P., Lombardo, M. V., Poser, B. A., & Bandettini, P. A. (2017). Multi-

echo fMRI: a review of applications in fMRI denoising and analysis of BOLD signals. *Neuroimage*, *154*, 59-80.

[7] Duval, T., & Lubrano, V. New neuroimaging technologies in SPM: BIDS, docker, boutique, and quality control.

[8] Duval, T., & Lubrano, V. New neuroimaging technologies in SPM: BIDS, docker, boutique, and quality control.

[9] Markiewicz, C. J., Gorgolewski, K. J., Feingold, F., Blair, R., Halchenko, Y. O., Miller, E., ... & Poldrack, R. (2021). The OpenNeuro resource for sharing of neuroscience data. *Elife*, *10*, e71774.

[10] Pedroni, A., Bahreini, A., & Langer, N. (2019). Automagic: Standardized preprocessing of big EEG data. *NeuroImage*, *200*, 460-473.

[11] Poldrack, R. A. (2019). The costs of reproducibility. *Neuron*, *101*(1), 11-14.

[12] McDougal, R. A., Bulanova, A. S., & Lytton, W. W. (2016). Reproducibility in computational neuroscience models and simulations. *IEEE Transactions on Biomedical Engineering*, *63*(10), 2021-2035.

[13] Herrick, R., Horton, W., Olsen, T., McKay, M., Archie, K. A., & Marcus, D. S. (2016). XNAT Central: Open sourcing imaging research data. *NeuroImage*, *124*, 1093-1096.

[14] Mori, S., Wu, D., Ceritoglu, C., Li, Y., Kolasny, A., Vaillant, M. A., ... & Miller, M. I. (2016). MRICloud: delivering high-throughput MRI neuroinformatics as cloud-based software as a service. *Computing in Science & Engineering*, *18*(5), 21-35.

[15] Jayapandian, C., Wei, A., Ramesh, P., Zonjy, B., Lhatoo, S. D., Loparo, K., ... & Sahoo, S. S. (2015). A scalable neuroinformatics data flow for electrophysiological signals using MapReduce. *Frontiers in neuroinformatics*, *9*, 4.

[16] Tapera, T. M., Cieslak, M., Bertolero, M., Adebimpe, A., Aguirre, G. K., Butler, E. R., ... & Satterthwaite, T. D. (2021). Flywheeltools: data curation and manipulation on the flywheel platform. *Frontiers in neuroinformatics*, *15*, 678403.

[17] Lopez-Novoa, U., Charron, C., Evans, J., & Beltrachini, L. (2019, August). The BIDS Toolbox: A web service to manage brain imaging datasets. In *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)* (pp. 378-382). IEEE.

[18] Holdgraf, C., Appelhoff, S., Bickel, S., Bouchard, K., D'Ambrosio, S., David, O., ... & Hermes, D. (2018). BIDS-iEEG: an extension to the brain imaging data structure (BIDS) specification for human intracranial electrophysiology.

[19] Markiewicz, C. J., Gorgolewski, K. J., Feingold, F., Blair, R., Halchenko, Y. O., Miller, E., ... & Poldrack, R. (2021). The OpenNeuro resource for sharing of neuroscience data. *Elife*, *10*, e71774.

[20] Poldrack, R. A. (2012). The future of fMRI in cognitive neuroscience. *Neuroimage*, *62*(2), 1216-1220.

[21] da Silva, F. L. (2013). EEG and MEG: relevance to neuroscience. *Neuron*, *80*(5), 1112-1128.

[22] Adorni, R., Gatti, A., Brugnera, A., Sakatani, K., & Compare, A. (2016). Could fNIRS promote neuroscience approach in clinical psychology?. *Frontiers in psychology*, *7*, 456.

[23] Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in neuroinformatics*, *2*, 343.