

DMP PLAN TEST

Data Type

Types and amount of scientific data expected to be generated in the project:

Although the project will not generate new data, we estimate that the proposed archive will store approximately 250TB within its first year. After the first year, we estimate up to 500TB will be deposited annually. These data will be decentralized such that up to 300 TB are stored at Texas Advanced Computing Center (TACC). We will utilize commercial cloud storage through Microsoft Azure and with an unlimited amount of storage at Amazon Web Service (AWS) through the Amazon Open Data Program. The types of data that will be stored include actigraph, human videos, primate videos, rodents videos, eye tracking, trials and events, sensory stimuli, virtual reality, eye tracking, trials and events, sensory stimuli, virtual reality, psychophysics, human pose estimation.

Scientific data that will be preserved and shared, and the rationale for doing so:

BEHIVE will provide a secure, cloud-based, and user-friendly data ecosystem for storing, tracking, analyzing, and sharing multimodal behavioral data. It will support diverse data modalities used in humans and non-human animal models primates (macaque and marmoset), rodents (rats, mice, and hamsters) and birds, such as videographic, audiographic, electrophysiologic, imaging, eye movements, vocalizations aligned with corresponding stimuli. By aligning with NIH data management and sharing policies, BEHIVE will facilitate the secure, cloud-based, and user-friendly storage and analysis of behavioral data.

Metadata, other relevant data, and associated documentation:

The proposed BEHIVE will be designed to work seamlessly with computational platforms such as Brainlife, NEMAR, CBRAIN, and DataJoint. Integrations will also extend to the existing data archives such as DABI, DANDI, BossDB, and OpenNeuro. BEHIVE Data Type metadata records will be synchronized and harmonized from the metadata records for the datasets and files (assets in DANDI archive terms) which potentially would follow their own schema. The effort will be made to converge on use of the same set of ontologies and metadata dictionaries to minimize the necessary amount of metadata harmonization, and instead benefit from centralized ontologies, e.g. as the one being developed within INCF Behavioral tasks and paradigms ontology. BEHIVE Data Types will be uniformly and sufficiently documented and readily available on the platform, with references to underlying used Ontologies where applicable. Detailed user-oriented documentation and walkthroughs will be provided on various aspects of metadata entry and editing, and released under a permissive Creative Commons license.

Related Tools, Software and/or Code

Support from the BRAINLIFE and DataJoint teams will enable the deployment of cutting-edge tools for data analysis and visualization, enabling the application of machine learning and artificial intelligence. The proposed BEHIVE will be designed to work seamlessly with computational platforms such as BRAINLIFE, NEMAR, CBRAIN, and DataJoint. Integrations will also extend to the existing data archives such as DABI, DANDI, BossDB, and OpenNeuro. These expansive integrations will rely on a decentralized data tracking technology such as DataLad. All of these tools, software, analysis code, and data archives are available online, and most of them (such as DANDI, DataLad, etc) are Free and Open Source Software released under permissive OSI-compliant licenses. The code developed in the scope of the BEHIVE project will also be released under OSI-compliant licenses as appropriate (primarily Apache-2). Additional visualization tools that collaborators will help contribute to the proposal include but are not limited to:

Standards

We will consider all standards that our advisory board and collaborators will suggest. Currently, we are listing some of the most prominent project standards for behavioral data that we will be considering, see Table XX. Dr. Pestilli is collaborator of the Hierarchical Event Descriptor (HED) and Dr. Halchenko of Neurodata Without Borders (NWB). The standards all have published documentation that can be accessed on

public websites or GitHub. The majority of these standards have been developed under BRAIN Initiative awards. We have enlisted collaborators that will help contribute the standards to the proposal (see LoCs).

Outputs

Sample Output 1, Type 1, 2021-01-01, Repository 1, Open Access

Sample Output 2, Type 2, 2022-02-02, Repository 2, Controlled Access