PSC
PITTSBURGH SUPERCOMPUTING CENTER

Carnegie Mellon University University of Pittsburgh

The Brain Image Library:

Data Submission Workshop – April 3, 2023

PSC

CBI Center for Biologic Imaging

Alexander Ropelewski (Contact PI)

Kathy Benninger (Networking)
Greg Hood (Image Analysis+ HPC)
Derek Simmel (Systems+Security)
Arthur Wetzel (Image Analysis)
Luke Tuite (User Support+Web)
Ivan Cao-Berg (Software+Support)
Elizabeth Pantalone (Web)
Mariah Kenney (Data Curator)

Alan Watson (PI)

Simon Watkins (Microscopy)
Iana Vasylieva (Post-Doc)

Alex Greg Derek Kathy Art Luke Ivan Liz Mariah Simon Alan Iana

This project is supported by the National Institute Of Mental Health of the National Institute of Health under Award Number R24MH114683.
The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
This work also used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Specifically, it used the Bridges system, which was supported by NSF award number ACI-1445606, and the Bridges-2 system, which is supported by NSF award number ACI-1928147.

Workshop Agenda - Today

Topic	Length (mins)	Schedule (EDT)
Introduction & Scope of BIL	20	1:00 - 1:20
Using the BIL Portal Data Transfer Rsync Globus SFTP Using OnDemand for cleanup/organization	45	1:20 - 2:05
File Organization	15	2:05 - 2:20
Break	15	2:20 - 2:35
Metadata File	30	2:35 - 3:05
Uploading Metadata Submitting for Publication	5	3:05 - 3:10
Validation Process & Timeline	10	3:10 - 3:20
Requesting a DOI	15	3:20 - 3:35
Questions/Discussion/Individual Help	25	3:35 – 4:00

Workshop Agenda - Tomorrow

Topic	Length	Schedule
How to find BIL data Website Inventory DOI	25	1:00 - 1:25
Tools Available at BIL	5	1:25 - 1:30
Jupyter Lab/Open OnDemand Neuron Visualization Registration	30	1:30 - 2:00
Break	15	2:00 – 2:15
BIL Analysis Ecosystem and Data Visualization	1:15	2:15 – 3:30
Questions/Discussion/Individual Help	30	3:30 – 4:00

PSC
PITTSBURGH SUPERCOMPUTING CENTER

Carnegie Mellon University University of Pittsburgh

The Brain Image Library:

Data Submission Workshop – April 3, 2023

PSC

CBI Center for Biologic Imaging

Alexander Ropelewski (Contact PI)

Kathy Benninger (Networking)
Greg Hood (Image Analysis+ HPC)
Derek Simmel (Systems+Security)
Arthur Wetzel (Image Analysis)
Luke Tuite (User Support+Web)
Ivan Cao-Berg (Software+Support)
Elizabeth Pantalone (Web)
Mariah Kenney (Data Curator)

Alan Watson (PI)

Simon Watkins (Microscopy)
Iana Vasylieva (Post-Doc)

This project is supported by the National Institute Of Mental Health of the National Institute of Health under Award Number R24MH114683.
The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
This work also used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Specifically, it used the Bridges system, which was supported by NSF award number ACI-1445606, and the Bridges-2 system, which is supported by NSF award number ACI-1928147.

In Memory of

Marcel Bruchez

1973 - 2022

Carnegie Mellon University | PSC | University of Pittsburgh

Goals and Agenda

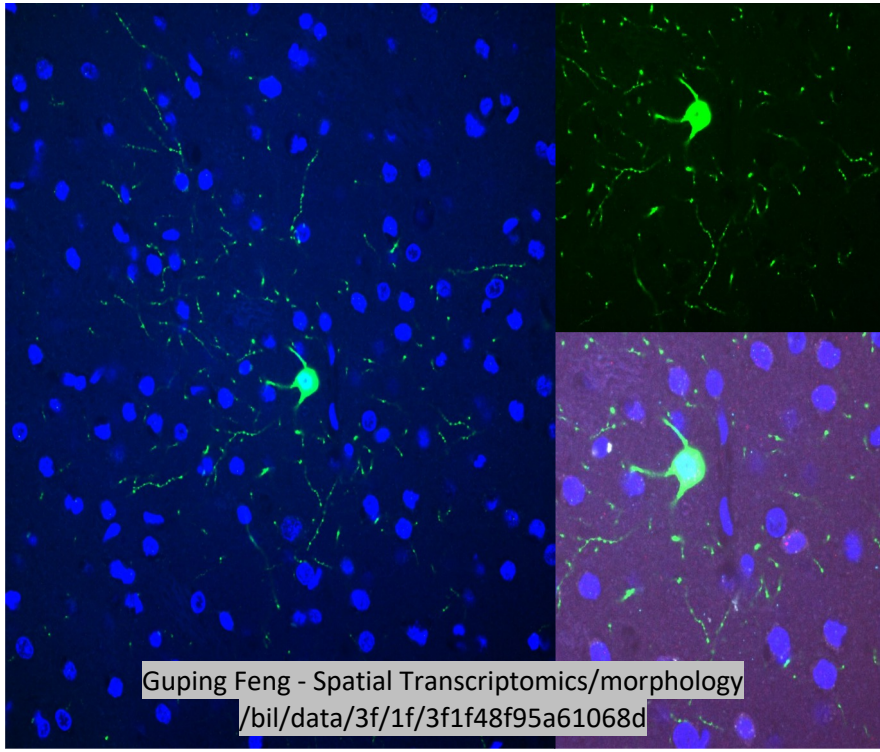
- What is the Brain Image Library
  - Scope, Types of Data and Data Coverage
- Additional Services Offered
- Coming Attractions
- How Do I get Started:
  - Pre-funding Stage
  - Funding has Been Awarded
- Data Submission

The Brain Image Library

Mission: National public resource enabling researchers to deposit, analyze, mine, share and interact with microscopy datasets of the brain.

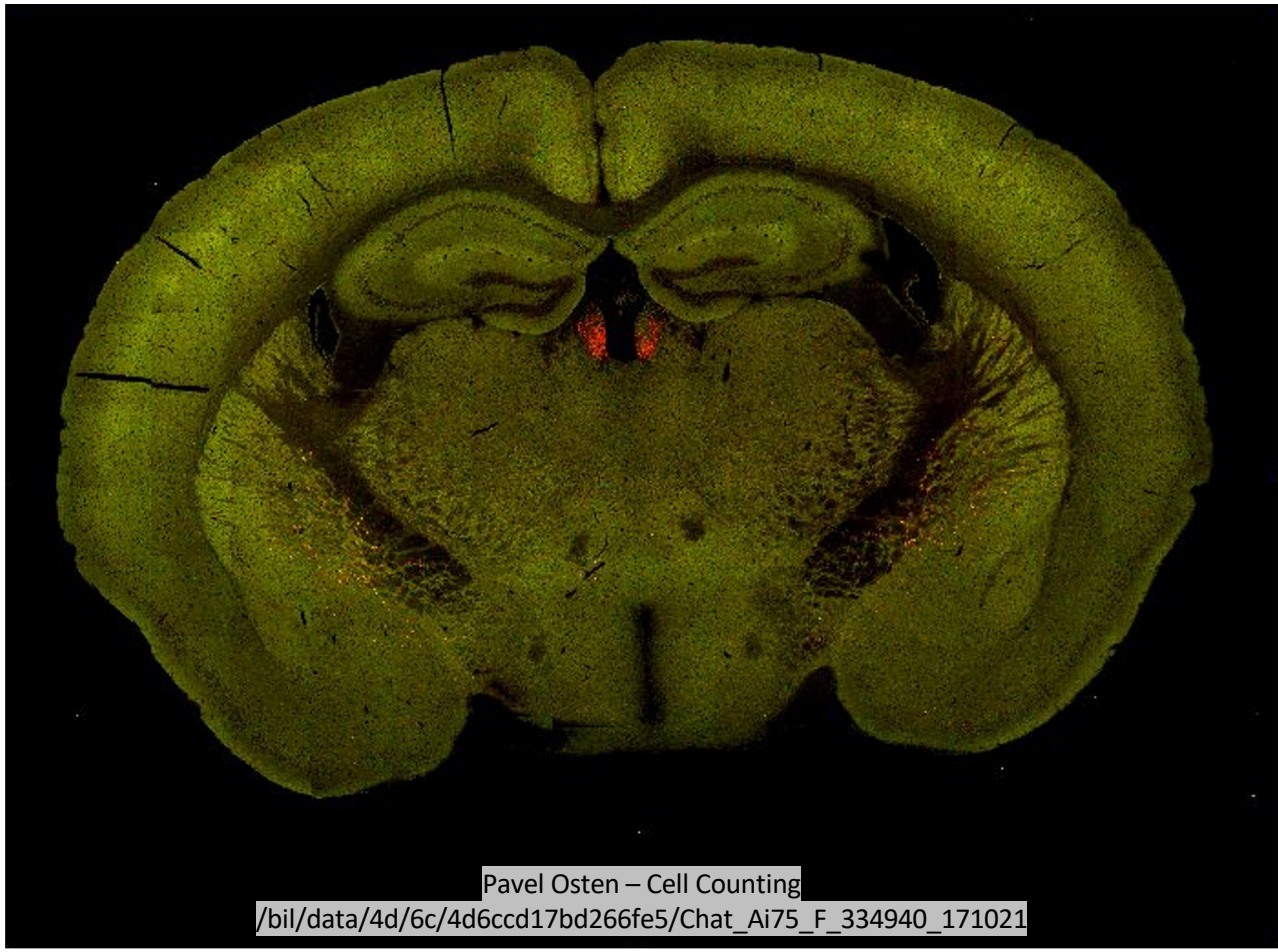
Scope:

- Permanent repository for high-quality brain microscopy datasets
  - *All NIH Investigators are required to deposit their data and make it publicly accessible.*
  - *NIH BRAIN Initiative funded investigators producing microscopy data are required to deposit their data in BIL.*
  - *Funding is not required to deposit data in BIL, but data deposited must be of interest to BRAIN Initiative.*
  - *Contributing data to BIL is free, there is no charge to contribute data.*
- Provide Analysis Ecosystem with desktop visualization and HPC computing capability for pre-submission data processing and post-submission exploration at no charge
- Provide user access and support including network path analysis



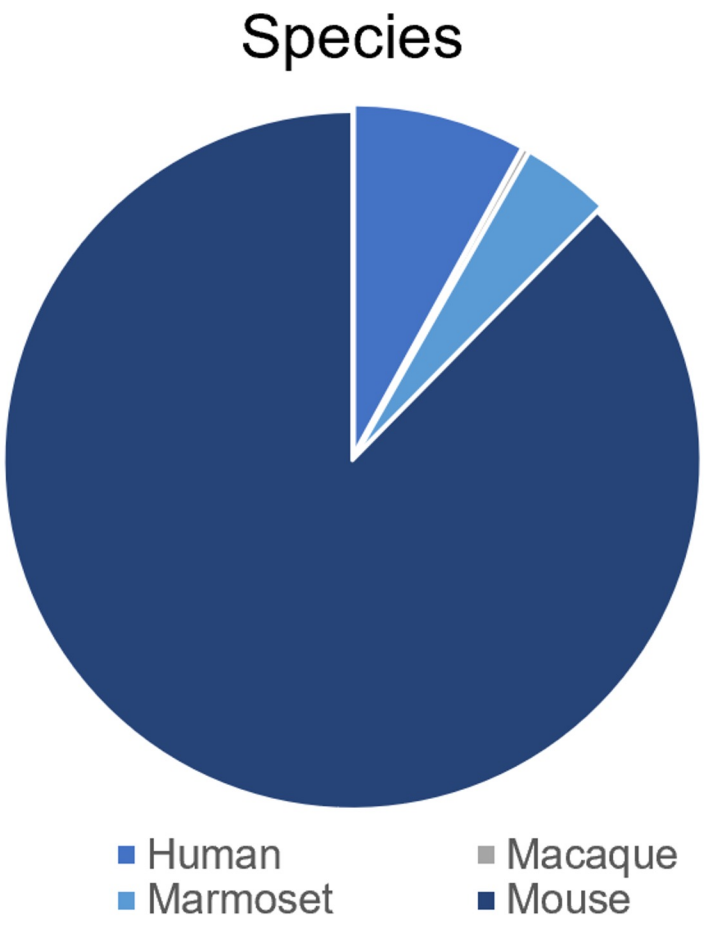
Breadth of Data

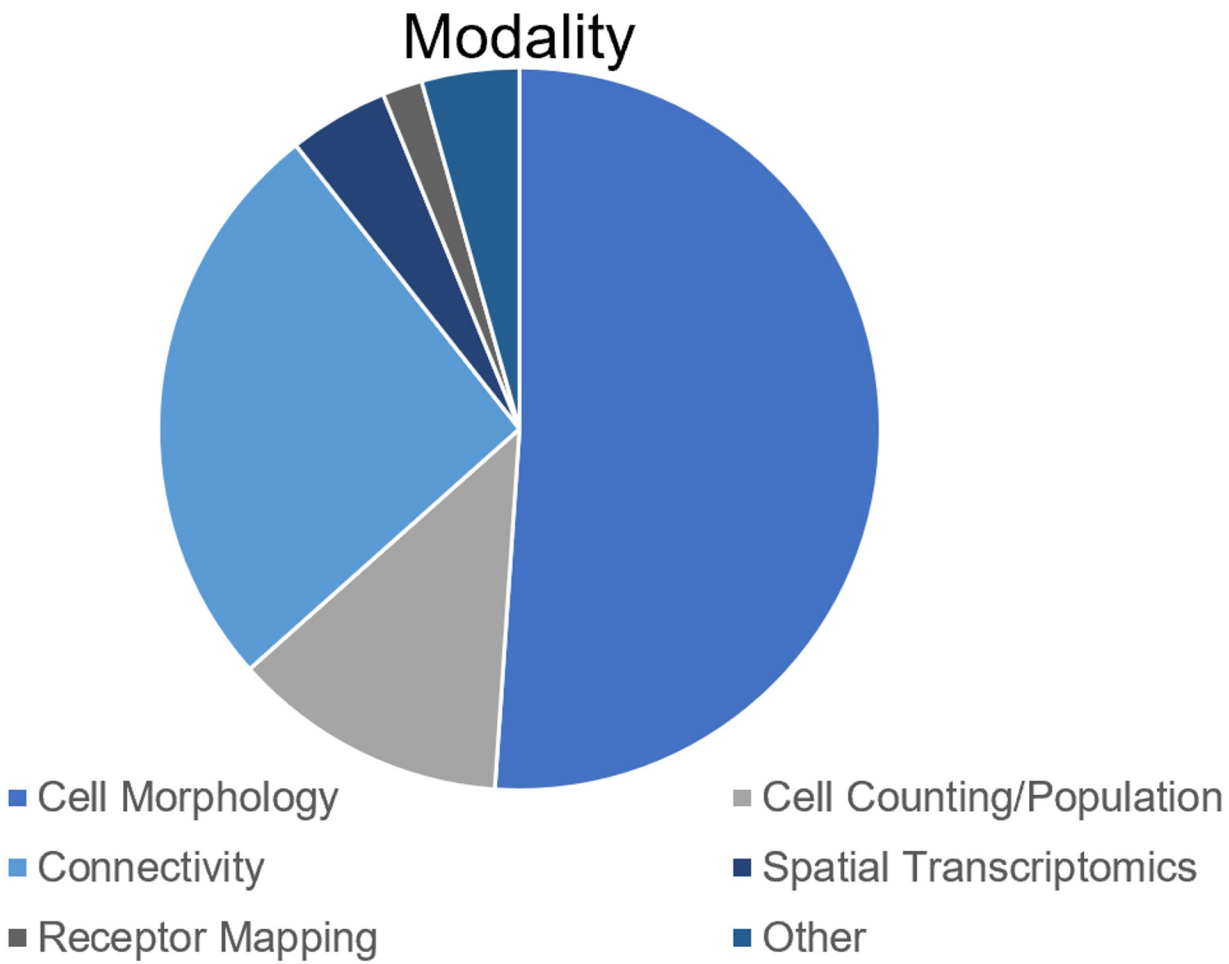
- Data Scope:
  - Whole (and partial) brain image datasets of mouse, rat, human, other mammals and model organisms along with their higher-level aligned and tracing data
  - Targeted experiments including connectivity between cells and spatial transcriptomics (*FISH)
  - Historical collections



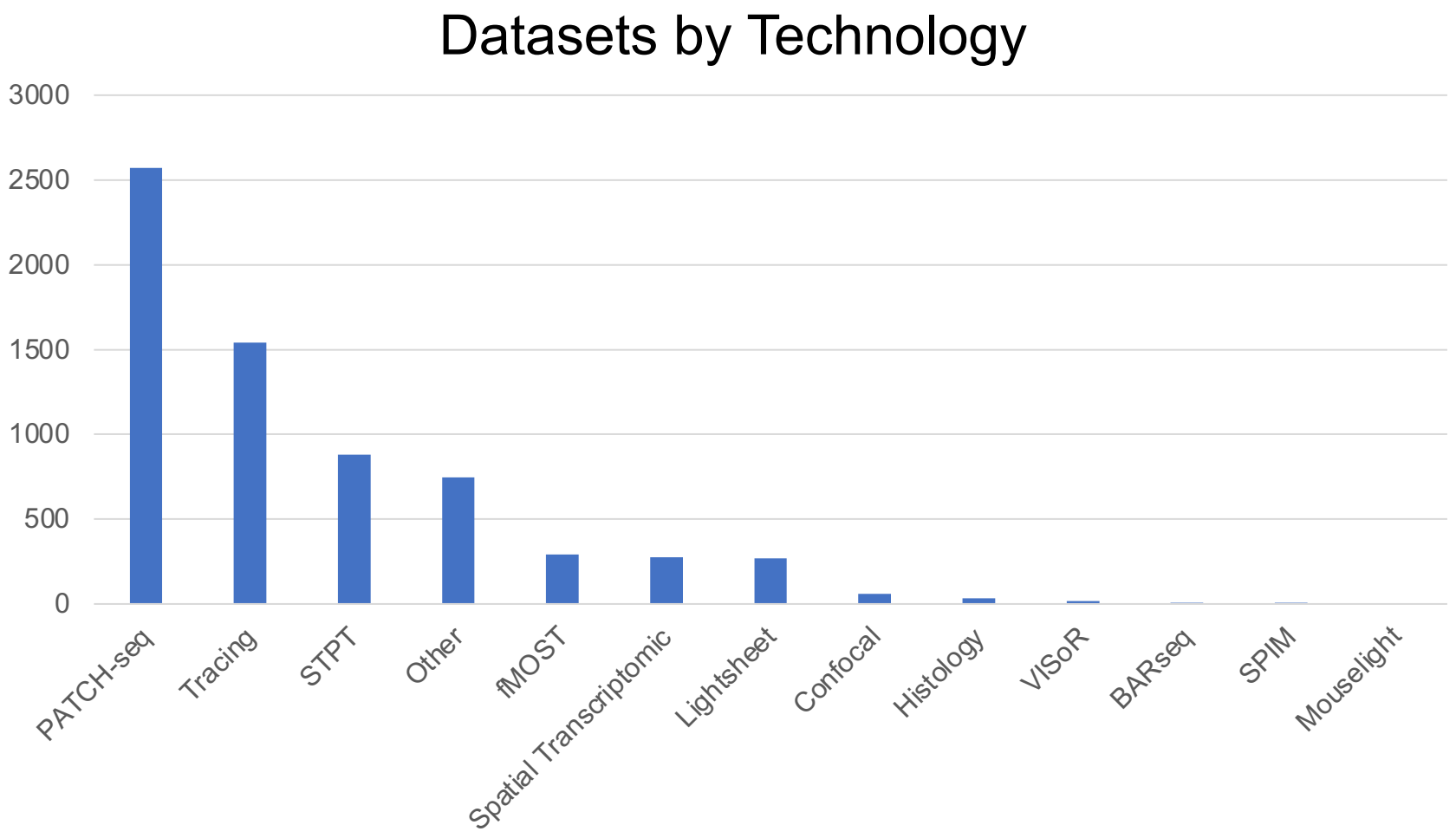
Pavel Osten – Cell Counting
/bil/data/4d/6c/4d6ccd17bd266fe5/Chat_Ai75_F_334940_171021

Deposited Data Coverage





Deposited Data Coverage

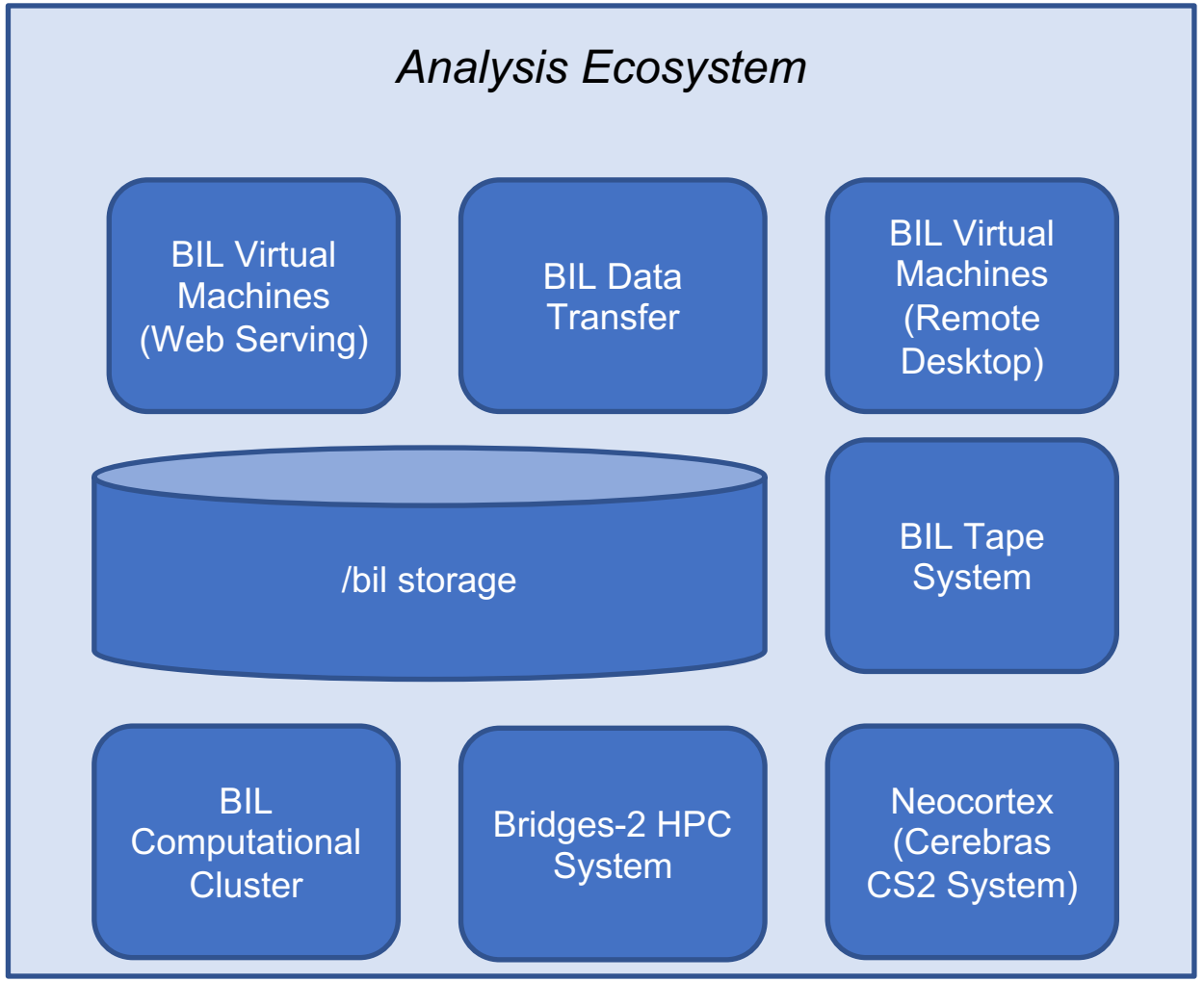


Additional Services BIL offers to Data Contributors

- Networking Support
  - Identify bottlenecks with data transfer
  - Recommendations to resolve last-mile issues
- Receive data via alternate media
- DOIs
  - Dataset DOIs (enable tracking of data usage)
  - Group/Collection DOIs
- Analysis Ecosystem
  - Pre Submission Processing
  - Public Data Exploration

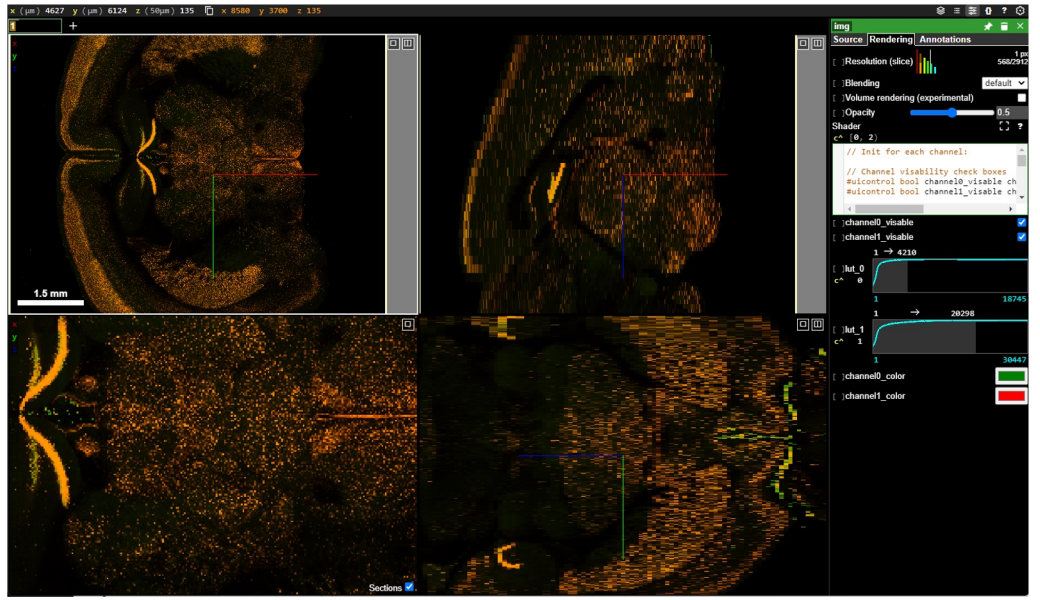
BIL Analysis Ecosystem

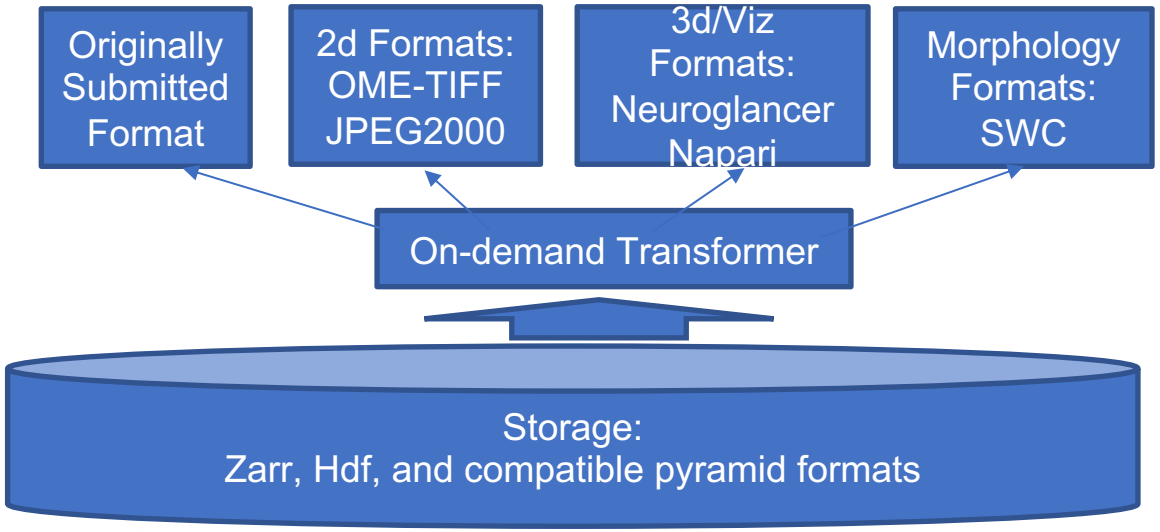
- BIL Computational, Visualization, and High-Speed Networking Systems to process and explore BIL data in-place available at no charge for open research and to support courses



Future: Making Data More Accessible

- On-Demand Transformer
  - Serving data in selected formats
- Greater coordination and linkage with large consortia including BICAN
- Metadata usability and greater accessibility
  - Metadata crosswalk
  - Links to related resources
- Manage project space and resources through ACCESS





How Do I Get Started?

Pre-Proposal

- Contact us at bil-support@psc.edu to discuss your project a few weeks prior to submission

- We can provide:
  - A letter of support stating BIL will accept your data
  - A template (NIH) Data Management and Sharing Plan outlining specific standards in use at BIL.

Beginning of Project

- Contact us at bil-support@psc.edu to discuss your data in detail:
  - The amount of data to be submitted and the timeframe
  - The anticipated submission timeline
  - Visualization

- We will advise you on:
  - The current metadata template most appropriate for your data.
  - How to structure your data to endure that it can be visualized
  - How to gain access to the BIL systems, including the Analysis Ecosystem
  - Upcoming training

PSC

January 1, 2023

RE: Funding Opportunity Announcement # RFA-AA-xx-xx

Dear Dr Watson,

It is our pleasure to write a letter in support of your application to NIH/NINDS entitled "Your Title Here".

As you know, the Brain Image Library (BIL) is a public resource serving the neuroscience community by providing a persistent centralized repository for microscopy data generated by brain researchers. BIL contains thousands of brain datasets including high-resolution data from the BRAIN Initiative Cell Census Network. Data in BIL includes: whole brain datasets and their accompanying neuron tracings; whole brain datasets from cell counting experiments; datasets from spatial transcriptomics; datasets arising from multimodal experiments such as patch-seq; and data from other experimental modalities. Data contributed to BIL is distributed under the Creative Commons Attribution-ShareAlike 4.0 International License, which will allow for the broadest use and reuse of data. Data in BIL uses the Standard metadata for 3D microscopy (https://doi.org/10.1038/s41597-022-01562-5) which was developed by a BRAIN Initiative standards project. The standard is aligned with both DataCite metadata, enabling DOIs to be issued for the data, and key OME metadata. The BIL is supported by the National Institutes of Mental Health of the National Institutes of Health under award number R24-MH-114793.

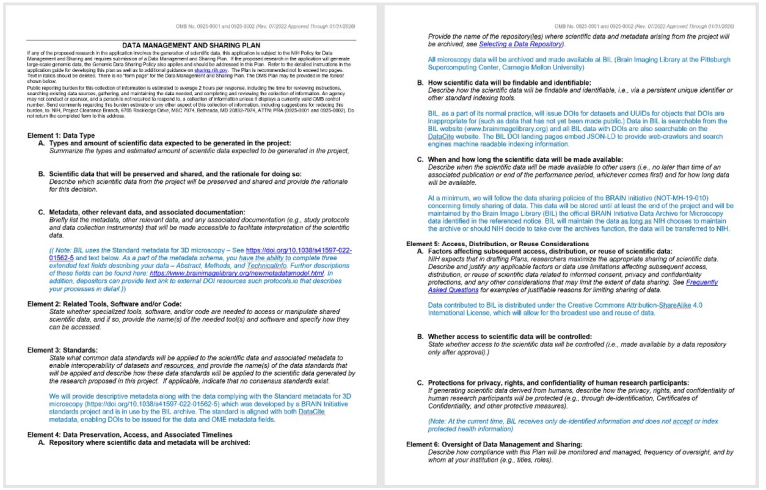
We have discussed your proposed project and are excited that we will be receiving your microscopy data to incorporate into the BIL. We understand that your project aims to ...what you intend to do along with a description of your data (e.g. confocal microscopy z-stack images of 25 mouse brains, Z-stacks will range between 500-1500 slices, and the total data size per brain is expected in the order of 1 Tb each.)

Thank you for your efforts to incorporate your data into the Brain Image Library, and best of luck with your grant application. I look forward to working with you on this exciting project!

Sincerely,

Alexander J. Ropelewski
Operations Director, Brain Image Library
Director Biomedical Applications Group
Pittsburgh Supercomputing Center

PSC | Carnegie Mellon University | University of Pittsburgh | 300 S. Craig Street, Pittsburgh, Pennsylvania 15213
www.psc.edu | Tel: 412.268.4960 | Fax: 412.268.5832



DATA MANAGEMENT AND SHARING PLAN

Element 1: Data Type
A. Types and amount of scientific data expected to be generated in the project:
Summarize the types and estimated amount of scientific data expected to be generated in the project.

B. Scientific data that will be preserved and shared, and the rationale for doing so:
Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.

C. Metadata, other relevant data, and associated documentation:
Briefly list the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

Element 2: Related Tools, Software and/or Code:
State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

Element 3: Standards:
State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources, and provide the name(s) of the data standards that will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project. If applicable, indicate that no consensus standards exist.

Element 4: Data Preservation, Access, and Associated Timelines
A. Repository where scientific data and metadata will be archived:

Provide the name of the repository(ies) where scientific data and metadata arising from the project will be archived, see Selecting a Data Repository).

B. How scientific data will be findable and identifiable:
Describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.

C. When and how long the scientific data will be made available:
Describe when the scientific data will be made available to other users (i.e., no later than time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

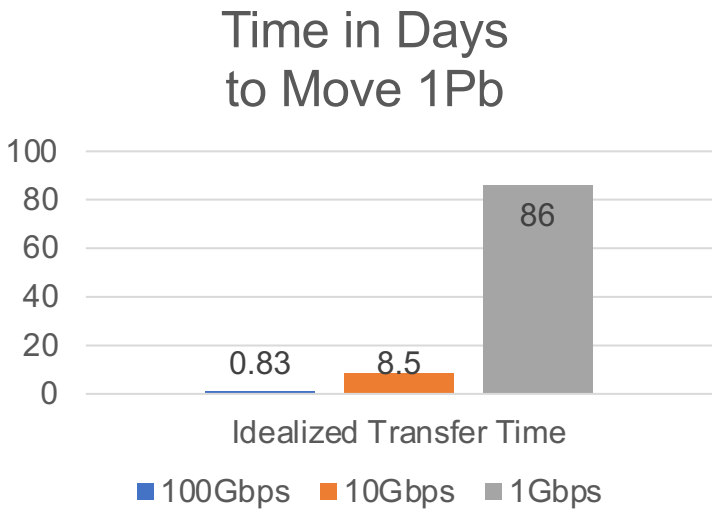
Element 5: Access, Distribution, or Reuse Considerations
A. Factors affecting subsequent access, distribution, or reuse of scientific data:
NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data. Describe and justify any applicable factors or data use limitations affecting subsequent access, distribution, or reuse of scientific data related to informed consent, privacy and confidentiality protections, and any other considerations that may limit the extent of data sharing. See Frequently Asked Questions for examples of justifiable reasons for limiting sharing of data.

B. Whether access to scientific data will be controlled:
State whether access to the scientific data will be controlled (i.e., made available by a data repository only after approval).)

C. Protections for privacy, rights, and confidentiality of human research participants:
If generating scientific data derived from humans, describe how the privacy, rights, and confidentiality of human research participants will be protected (e.g., through de-identification, Certificates of Confidentiality, and other protective measures).

Element 6: Oversight of Data Management and Sharing:
Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles).

Carnegie Mellon University PSC University of Pittsburgh

Plan Ahead for Data Submission

- Start the data submission process well-in-advance of any deadlines that you may have for making the data public
- Consider the time required to move data:
  - We can provide assistance locating where the transfer bottleneck is occurring
- Data validation time will be proportional to the number of files in the submission:
  - Smaller number of large files validate quicker
  - Smaller total number of datasets in submission will validate quicker
- Curation team may need your assistance to resolve issues