PSC
PITTSBURGH SUPERCOMPUTING CENTER

Carnegie Mellon University University of Pittsburgh

The Brain Image Library:

Data Submission Workshop – June 27, 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
Creating the Metadata File	20	2:35 - 2:55
Uploading Metadata Submitting for Publication	5	2:55 - 3:00
Validation Process & Timeline	10	3:00 - 3:10
Requesting a DOI	10	3:10 - 3:20
How to find BIL data • Website • Inventory • DOI	15	3:20 - 3:30
Tools available at BIL	15	3:30 – 3:45
Questions/Discussion/Individual Help	15	3:45 – 4:00

PSC
PITTSBURGH SUPERCOMPUTING CENTER

Carnegie Mellon University University of Pittsburgh

The Brain Image Library:

Data Submission Workshop – June 27, 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.

In Memory of

Marcel Bruchez

1973 - 2022

Carnegie Mellon University PSC University of Pittsburgh

Goals and Agenda

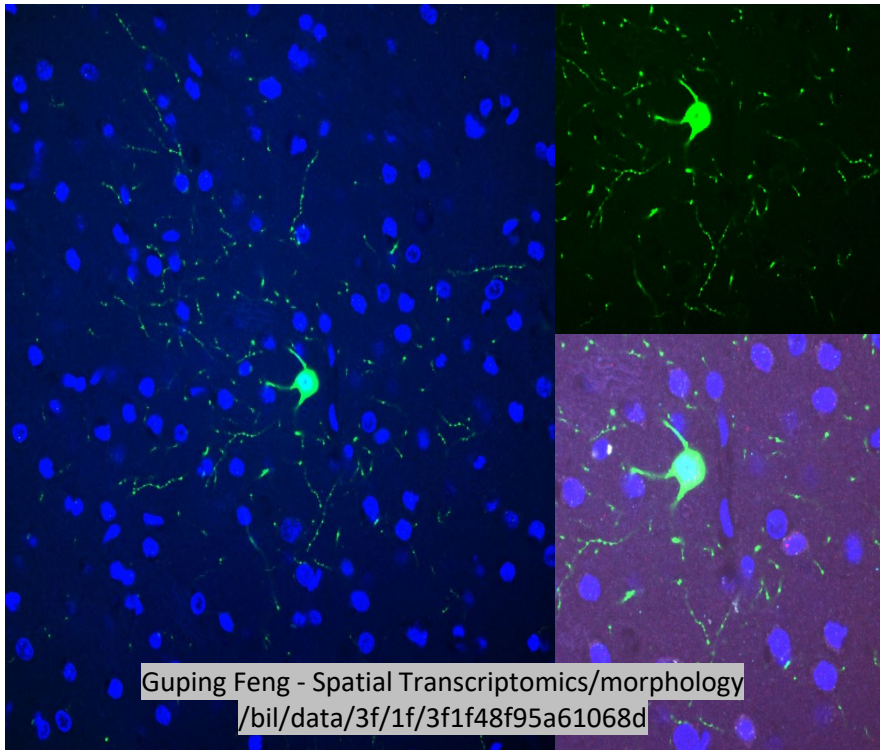
- What is the Brain Image Library
  - Scope, Types of Data and Data Coverage
- Services Offered
- Coming Attractions
- How Do I get Started:
  - Pre-funding Stage
  - Funding has Been Awarded
  - Data Submission

The Brain Image Library

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

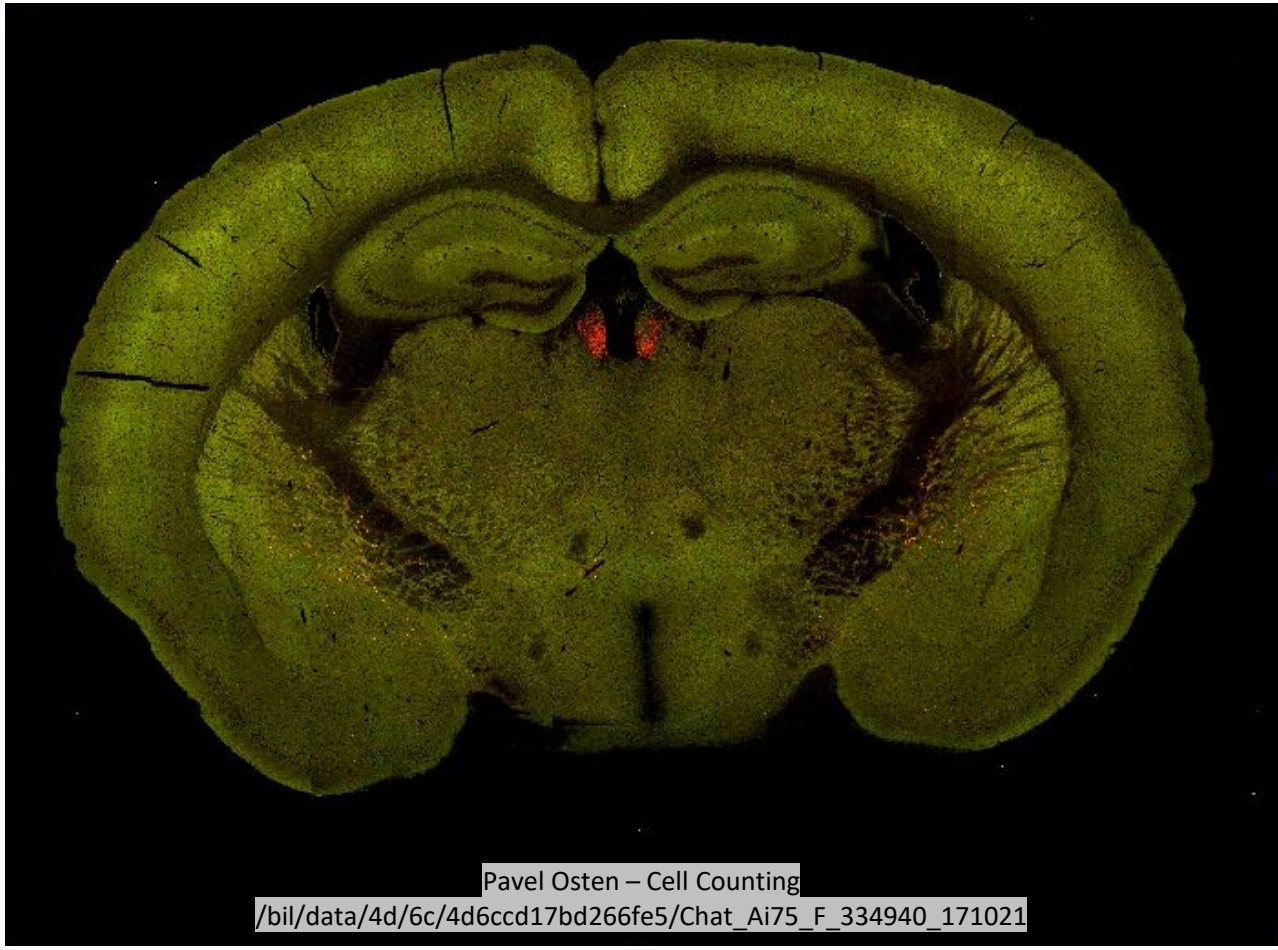
Scope:

- 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.*
- 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



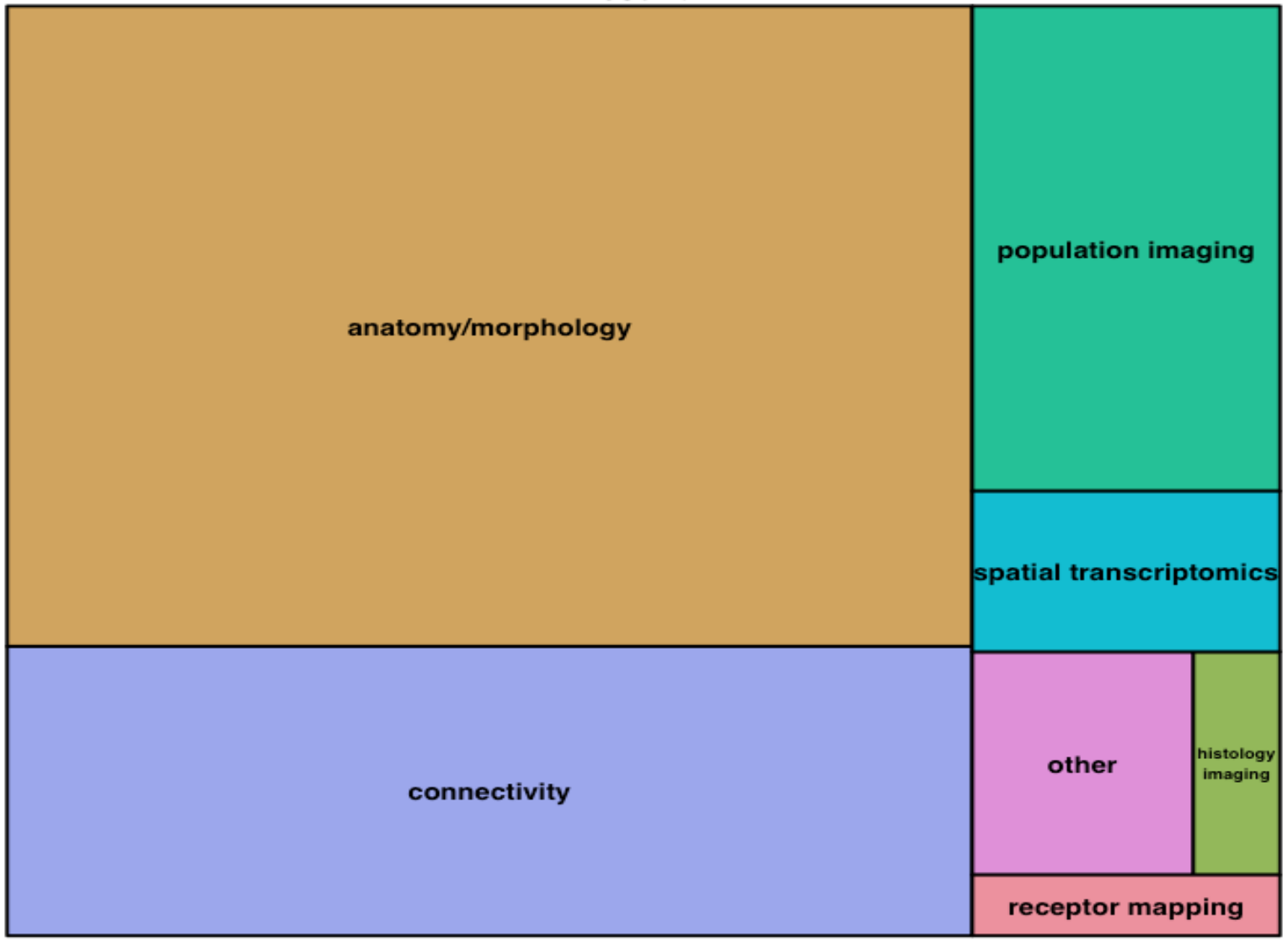
Breadth of Data

- 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

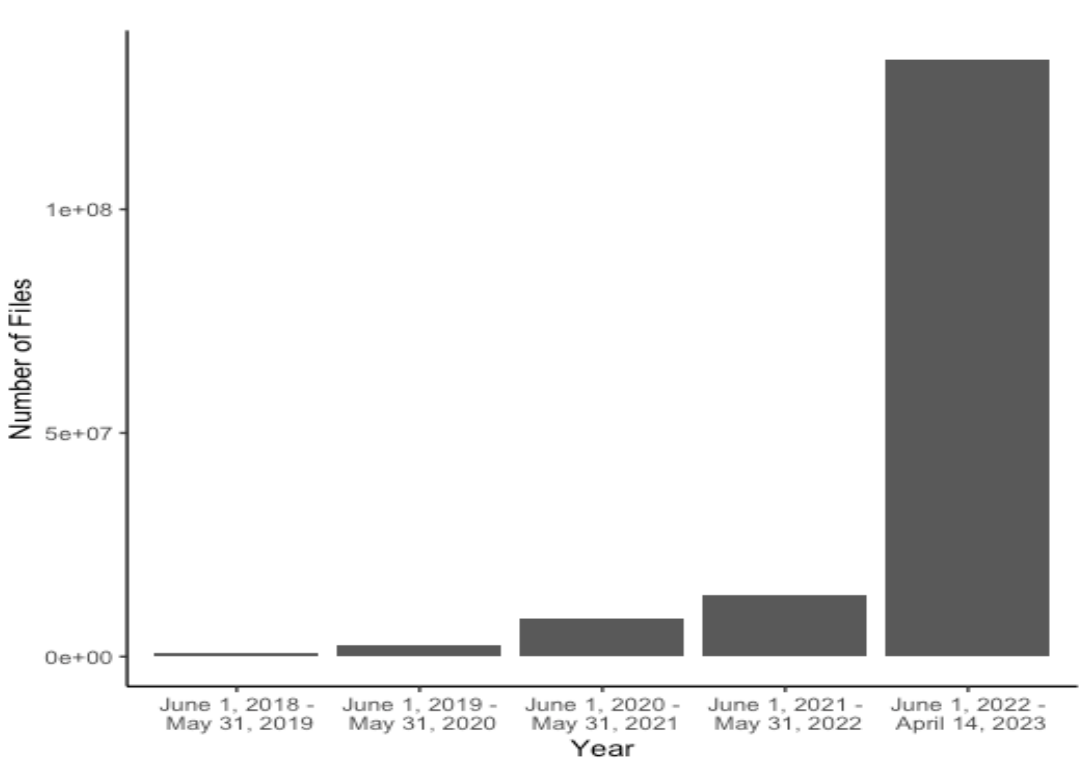


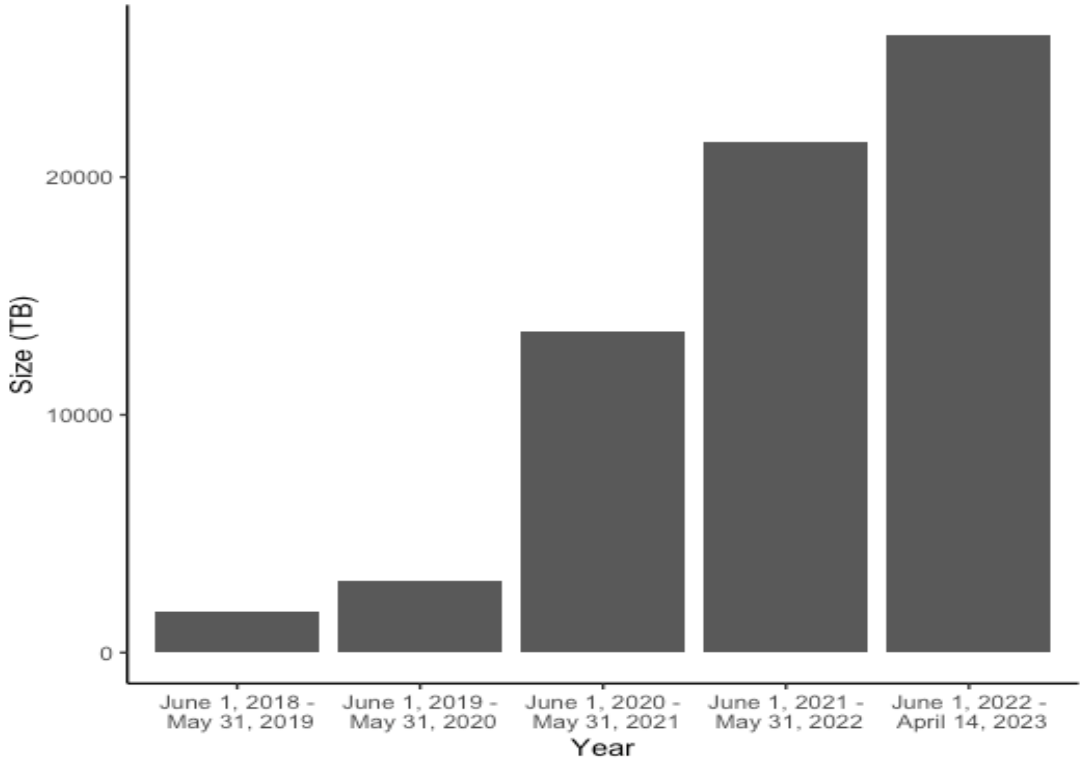
Pavel Osten – Cell Counting
/bil/data/4d/6c/4d6ccd17bd266fe5/Chat_Ai75_F_334940_171021

Deposited Data Coverage



Deposited Data Coverage





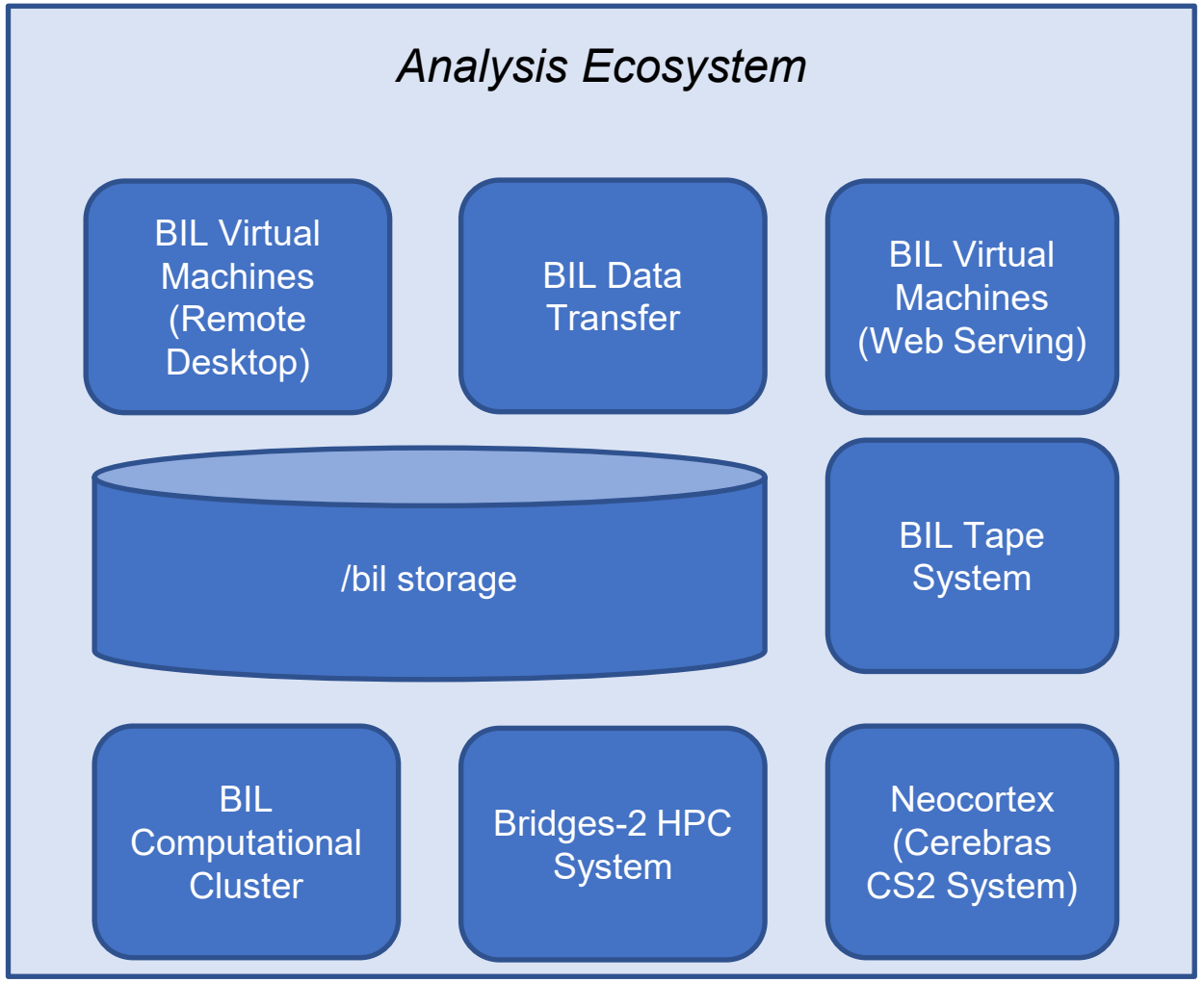
Carnegie Mellon University PSC University of Pittsburgh

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
- Analysis Ecosystem
  - Pre submission processing
  - Public data exploration

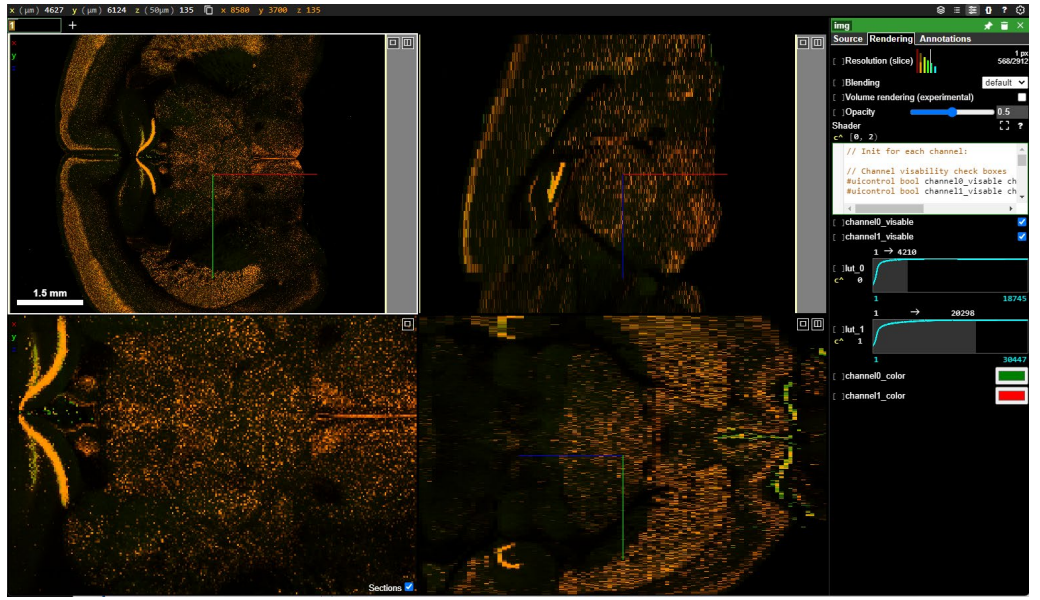
BIL Analysis Ecosystem

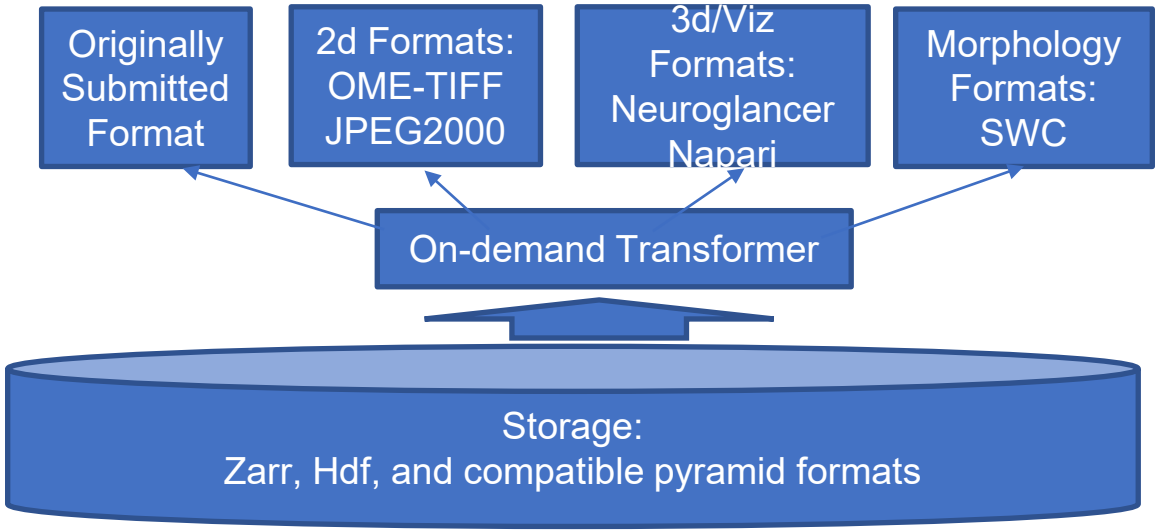
- BIL Computational, Visualization, and High-Speed Networking Systems to process and explore BIL data in-place available at no charge for open research and to support courses



Future: Making Data More Accessible

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





How Do I Get Started?

- Ideally, Before Proposal Submission:
  - Contact us at bil-support@psc.edu to discuss your project a few weeks prior to submission:
    - The anticipated types and amounts of data
    - The anticipated submission timeline
  - 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.
  - We can also discuss potential collaborations beyond data deposition

Funding Stage: Detailed Discussion

- When your proposal is funded:
  - Congratulations!
  - Contact us at bil-support@psc.edu to discuss your data in detail:
    - The funded amount of data to be produced.
    - The anticipated submission timeline
  - We will advise you on:
    - The current metadata template most appropriate for your data.
    - How to structure your data to ensure that it can be visualized
    - How to gain access to the BIL systems, including the Analysis Ecosystem
    - Upcoming training

Submission Stage: Please Plan Ahead

- 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:
  - Small number of large files preferred
  - Smaller submissions preferred.
- Curation team may need your assistance to resolve issues