

The Brain Image Library:

Data Submission Workshop – January 8, 2024



Alexander Ropelewski (Contact PI)

Kathy Benninger (Networking)

Rozita Laghaei (HPC+Support)

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



Rozita



Derek



Kathy



Art



Luke



Ivan



Liz



Mariah



Simon Alan



Iana

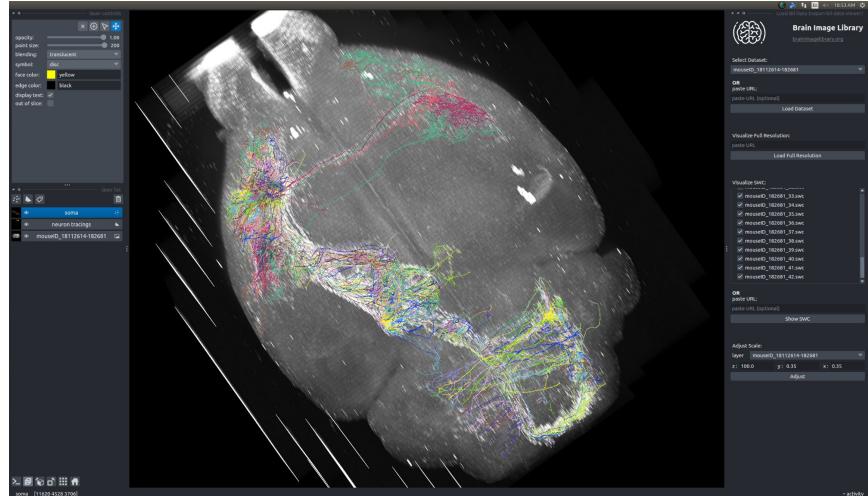


The Brain Image Library is supported by the National Institutes of Mental Health of the National Institutes of Health under award number R24-MH-114793. The napari visualization plugin is funded by the Chan Zuckerberg Initiative Donor Advised Fund grant #DAF2022-309651. This work utilized the Extreme Science and Engineering Discovery Environment (XSEDE), supported by National Science Foundation (NSF) award ACI-1548562, and allocation BIO210066 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, supported by NSF award #2138259, #2138286, #2138307, #2137603, and #2138296. Specifically, it used the Bridges system, supported by NSF award ACI-1445606, and the Bridges-2 system, supported by NSF award ACI-1928147. This content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies mentioned.

Topic	Length (mins)	Schedule (EDT)
Introduction & Scope of BIL	20	1:00 - 1:20
Using the BIL Portal	45	1:20 - 2:05
Data Transfer		
Using OnDemand for cleanup/organization		
File Organization	15	2:05 - 2:20
Break	15	2:20 - 2:35
Creating the Metadata File	20	2:35 - 2:55
Uploading Metadata	5	2:55 - 3:00
Submitting for Publication		
Validation Process & Timeline	10	3:00 - 3:10
Citing BIL Data	10	3:10 - 3:20
How to find BIL data	15	3:20 - 3:30
•API		
•Landing Pages for Datasets		
•Website		
•Inventory		
•DOI		
Tools available at BIL	15	3:30 – 3:45
•BIL Analysis Ecosystem		
•One Click Visualization		
Questions/Discussion/Individual Help	15	3:45 – 4:00

Introduction

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

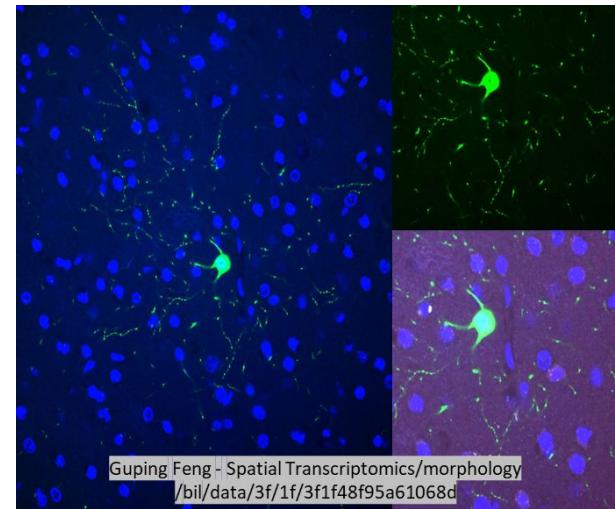


The Brain Image Library

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

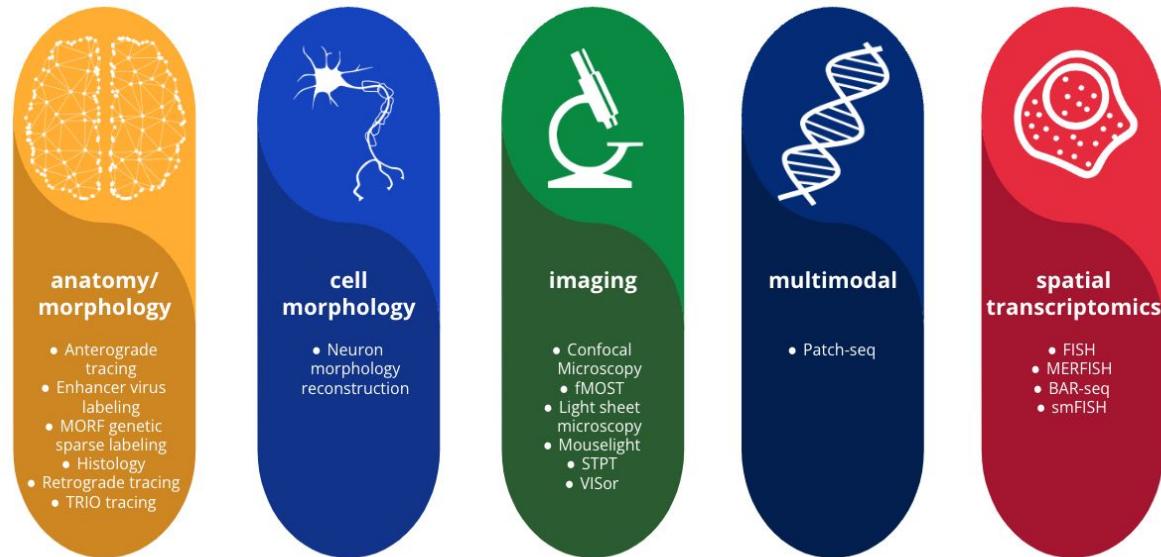
Scope:

- 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

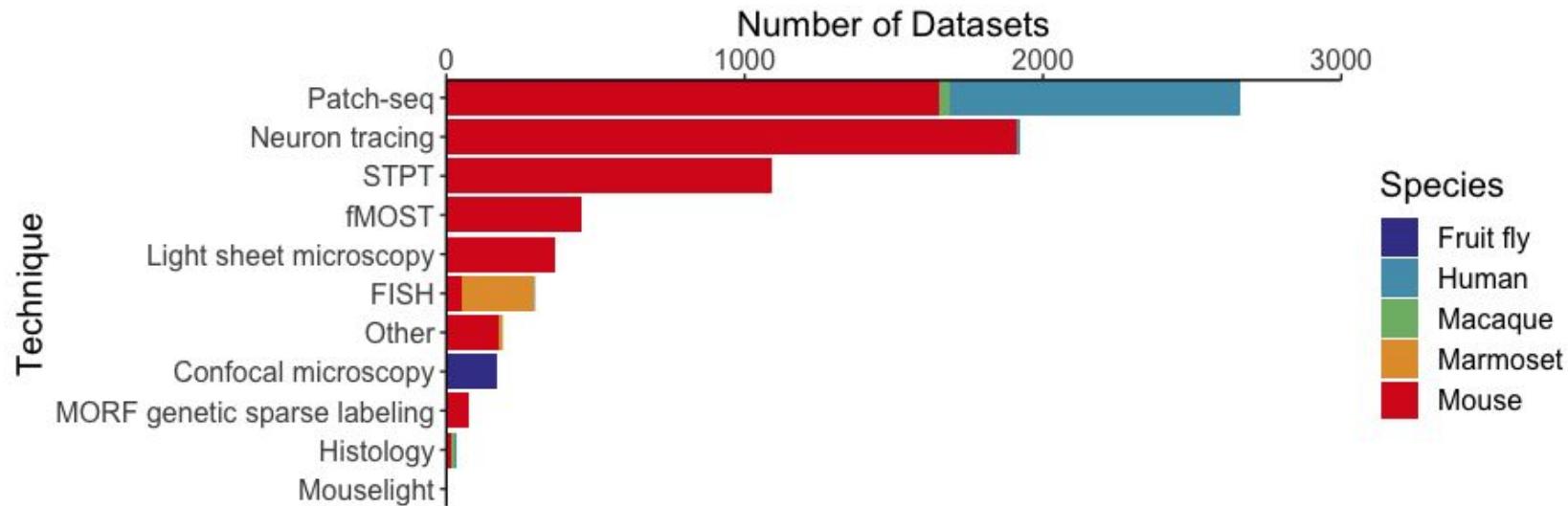


Breadth of Data

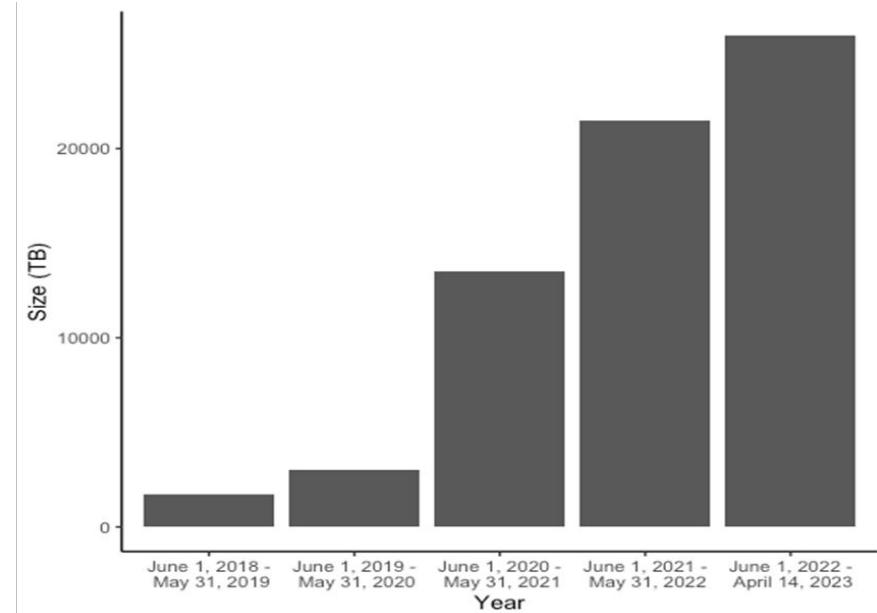
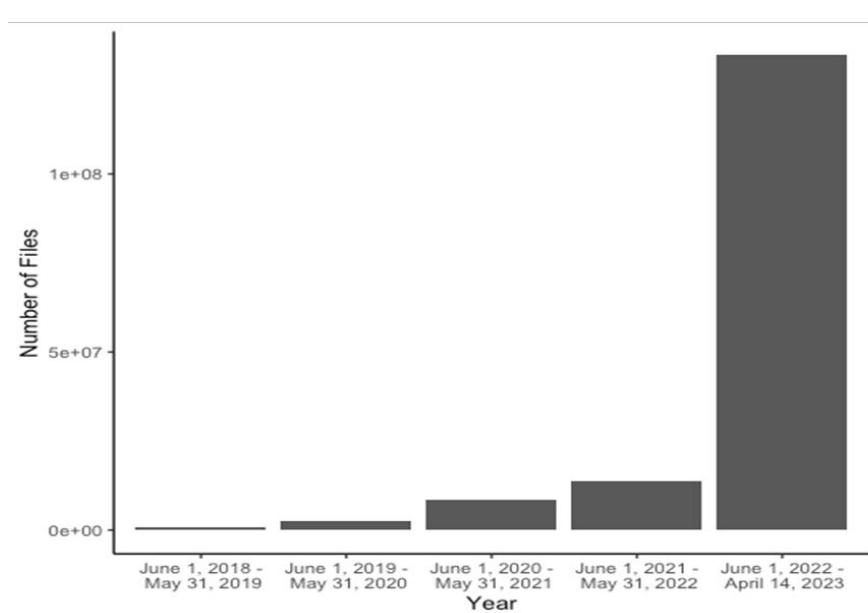
- Whole (and partial) optical microscopy brain datasets along with their higher-level data and annotations
- Targeted experiments including connectivity between cells and spatial transcriptomics
- Historical collections



Current Deposited Data Coverage



Deposited Data Coverage - Public Data



Additional Services BIL Provides

- 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
- Dataset DOIs and Group DOIs

Brain Image Library

About ▾ Data Submission ▾ Data Access ▾ Contact

Whole-brain LSFM imaging of a E15.5 mouse for DevCCF project
(Subject: E15.5_BB0428Technique: Background)

[BIL ID: ace-cab-ear]
Dataset Citation:
Zhang, Jiangyang, Lee, Choong Heon, Jiangyang Zhang Lab, Betty, Rebecca, Liwang, Josephine, Yongsoo Kim, Kronman, Fae, Manjila, Steffy Babu, Yongsoo Kim, Yongsoo Kim Lab, Minteer, Jennifer. (2022). Whole-brain LSFM imaging of a E15.5 mouse for DevCCF project (Subject: E15.5_BB0428Technique: Background) [Dataset/Microscopy]. Brain Image Library.

Abstract:
This dataset is part of a BICCN project to create developmental CCFs with associated ontology and true 3D anatomical labels while also demonstrating the application of our CCFs by generating quantitative mappings of GABAergic neurons in the developing mouse brain. This sample consists of LSFM imaging of a E15.5 mouse (Subject: E15.5_BB0428Technique: Background Autofluorescence).

Methods:
Images were processed using a modified LifeCanvas Protocol

Contributors and Roles:
Zhang, Jiangyang (Project Leader), New York University; Kronman, Fae (DataCurator), Pennsylvania State University ; Lee, Choong Heon (DataCollector), New York University; Jiangyang Zhang Lab (ResearchGroup); Betty, Rebecca (DataCollector), Pennsylvania State University; Liwang, Josephine (DataCollector), Pennsylvania State University; Yongsoo Kim (ProjectLeader), Pennsylvania State University; Kronman, Fae (ContactPerson), Pennsylvania State University; Manjila, Steffy Babu (DataCollector), Pennsylvania State University; Yongsoo Kim (ContactPerson), Pennsylvania State University; Yongsoo Kim Lab (ResearchGroup); Minteer, Jennifer (DataCollector), Pennsylvania State University; Beck, Kira (ProjectMember), Pennsylvania State University

Funding:
National Institutes of Health
(1-R01-MH124605-01) Establishing Common Coordinate Framework for Quantitative Cell Census in Developing Mouse Brains

Technical Information:
other (Modality) / light sheet microscopy (Technique)

Images were stitched and compressed using in house code with MATLAB and Python, respectively

Instrument Metadata

microscopytype	Light sheet microscope (LSFM)
microscopemanufacturerandmodel	Life Canvas Tech
objectivename	
objectiveimmersion	
objectivernova	

License:
Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

Data:

Data location on the Brain Image Library Analysis Ecosystem:
[/bil/data/91/aa/91aaad194ce577ebe/E15.5_BB0428/LSFM/stitched_00](#)

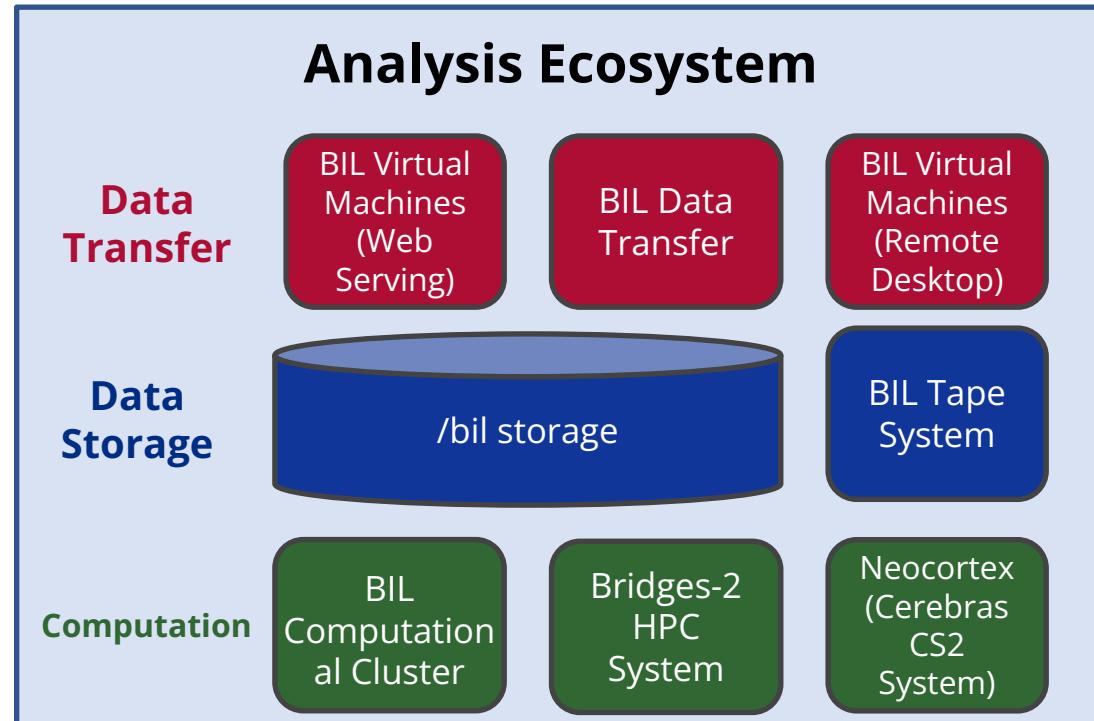
Dataset download link:
[https://download.brainimagedlibrary.org/91/aa/91aaad194ce577ebe/E15.5_BB0428/LSFM/stitched_00](#)

Metadata download link:
[https://api.brainimagedlibrary.org/retrieve/bilid=ace-cab-ear](#)

Manifest download link:
[https://download.brainimagedlibrary.org/inventory/122a14b8-dfba-59e4-8abf-7c1aaaf6c47b.json](#)

Analysis Ecosystem

Analysis Ecosystem provided to process, store and explore BIL data in-place
available at no charge for open research and to support courses



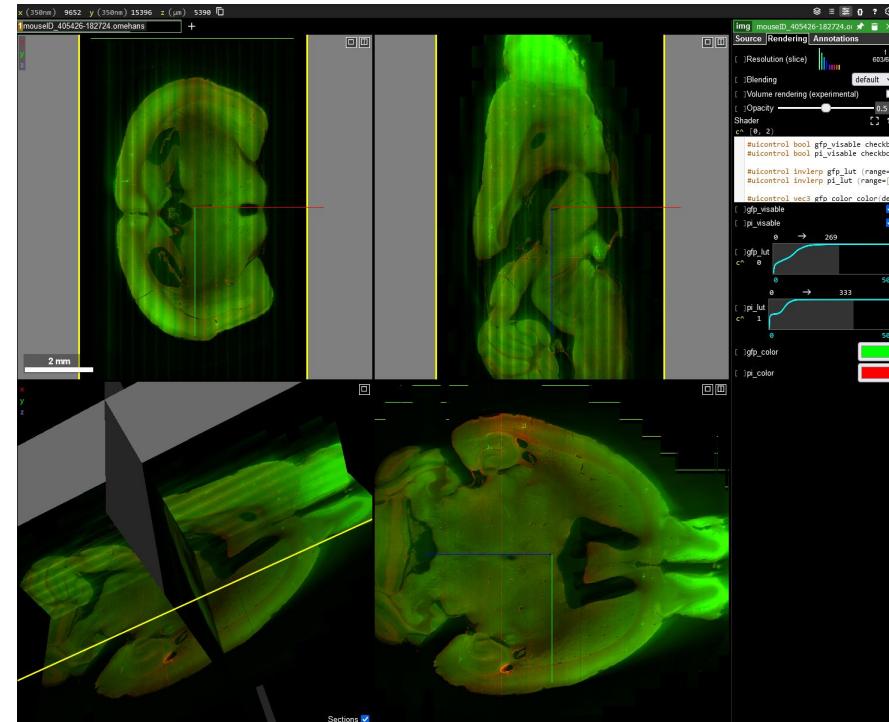
New Features

New Features:

- Metadata API: Enables programmatic searching/retrieval of BIL metadata for public datasets
- Visualization through neuroglancer/napari (currently, limited number of datasets)
- Dataset ids in addition to submission ids.

Coming Soon:

- Ecosystem allocations through ACCESS
- Enhanced dataset linkage & organization
- Controlled-access visualization of non-public, in-process datasets
- On-Demand Transformer



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

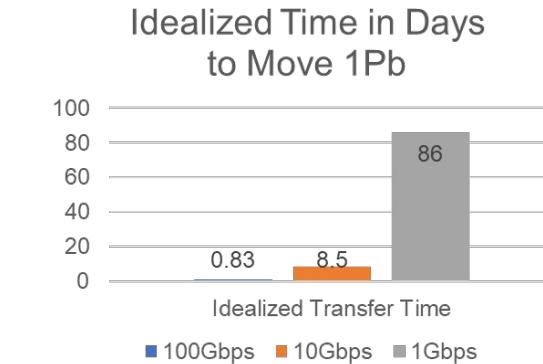
When your proposal is funded:

- 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: 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 - email: bil-support@psc.edu
- Data validation time is generally 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





Carnegie
Mellon
University



Using the BIL Data Submission Portal

Luke Tuite

Brain Image Library Data Submission Process

Workflow for
depositing data
at BIL

- Creating submissions
- Structuring your data
- Transferring data to BIL
- Making your data public

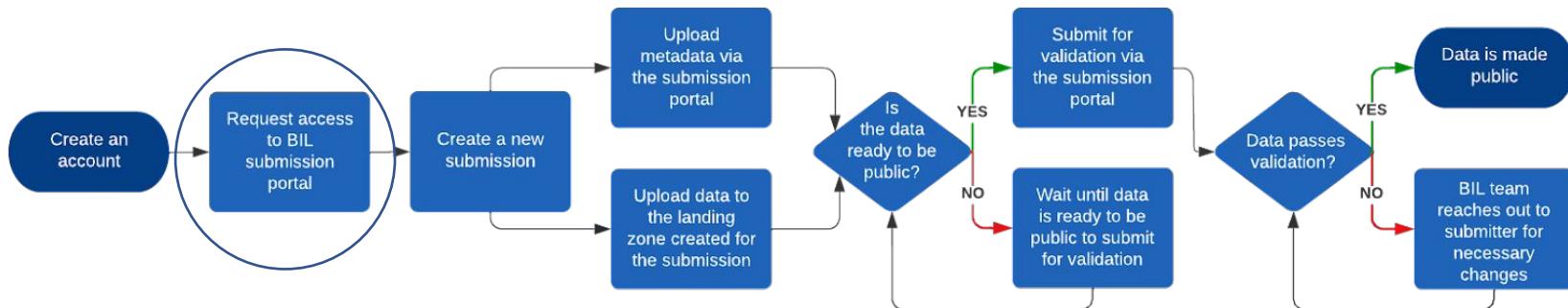
Providing
metadata for
your submissions

- Structuring your metadata
- Proper inclusions

BIL PI Dashboard

- Creating your group's Project
- View all submissions to your Project

Submission Overview



Submitting Your Data to BIL: Setting up your Account

Visit the ACCESS Portal Website:
<https://identity.access-ci.org/new-user>

Enter the Required Information

Verify your ACCESS Portal Account

<http://www.brainimagelibrary.org/account.html>



Identity Management

[Acceptable Use](#) [FAQ](#) [Help](#) [Privacy](#) [Security](#)

ACCESS User Registration

Avoid Creating Duplicate Accounts

If you already have an XSEDE or ACCESS account, please do not create another one. You can [request a username reminder](#) and/or [reset your password](#) to continue using an existing account. If you're having trouble accessing an existing account, please [contact us](#) rather than creating another one.

Two Options for New User Registration

If you don't already have an XSEDE or ACCESS account, there are two registration options:

1. [Register with an existing identity](#): Using an existing GitHub, Google, Microsoft, ORCID, or University account when registering with ACCESS simplifies the sign-up process and enables you to log in to ACCESS using that existing account. With this option, creating an ACCESS-specific password is optional during registration, and you will also have the option to create an ACCESS-specific password later if needed.
2. [Register without an existing identity](#): With this option, you'll be prompted to enter all your registration info and select an ACCESS-specific password. You can [link](#) a GitHub, Google, Microsoft, ORCID, or University account later if desired.

Finalizing Account Creation

Request access to the data submission portal Send email to bil-support@psc.edu along with your ACCESS Portal Account. You will receive an email message once access has been enabled. Please allow 24 hours for access to be granted.

Setting Password: Visit the Website: apr.psc.edu

PSC

PSC Password Change Utility

Use this automated process to change your PSC password. Please note that changing your PSC password **does not** change your password on the XSEDE User Portal.

See more information on [PSC password policies](#).

IMPORTANT

- After 6 hours, the password change process times out. If the password change is not completed you will have to restart.
- Your IP address will be logged for security reasons. If you want to avoid this, you can [contact PSC User Services](#) to request that your password be reset. Please note that User Services will only reset your password to its initial value.

Click **Start** to change your PSC password.

Start

Finalizing Account Creation

Please Respond: PSC User Password Change



From grants@psc.edu
to ropelews@pscuxb.psc.edu

12:19:28 PM

Dear Alexander J. Ropelewski,

A PSC User Password Change request has been submitted from the PSC Password Change Tool.

Return to the page where you initiated the password change and enter the following security code:

CuDD0T2f

IMPORTANT: Each password change request is valid for 6 hours, please complete the request within the allocated time. If expired, you will have to restart the request. Please note the security code is case sensitive.

If you have not submitted a password change request please contact PSC User Services at <http://www.psc.edu/index.php/about/contact-us> immediately.

PSC

PSC Password Change Utility

A unique security code has been sent to all registered email addresses we have on file for you at PSC. At least one address matches the address you entered on the previous screen. Enter the security code below and click **Submit**

Note:

- Do not close this page until you have completed the password change process
- The security code is case sensitive. Enter it exactly as it appears in your email. You will have 3 chances to enter the correct security code after which your account will be put on hold for 15 minutes.
- Remember the password change process times out after 6 hours. If you exceed the allotted time you must restart this process.

Security Code

© Pittsburgh Supercomputing Center, Carnegie Mellon University, University of Pittsburgh
300 S. Craig Street, Pittsburgh, PA 15213 Phone: [412.268.4960](tel:412.268.4960) Fax: 412.268.5832

PSC

PSC Password Change Utility

Complete the form below to finish the PSC password change process.

Note: Your new password must:

- be at least 8 characters long
- contain characters from at least 3 of the following groups:
 - lower-case letters
 - upper-case letters
 - digits
 - special characters [excluding apostrophe (') and quotes (" ")]
- be different from the last three passwords that you have used.

Username:

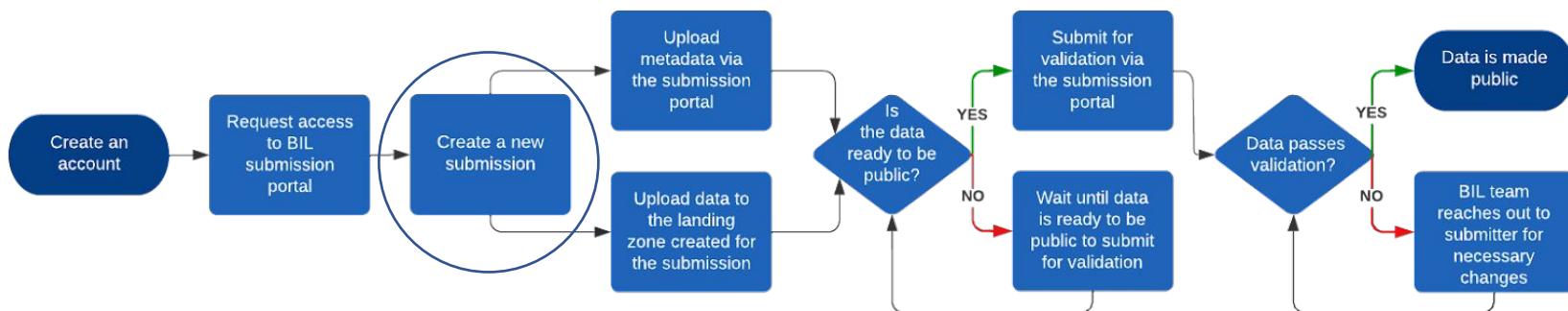
New PSC password:

Confirm PSC password:

© Pittsburgh Supercomputing Center, Carnegie Mellon University, University of Pittsburgh
300 S. Craig Street, Pittsburgh, PA 15213 Phone: [412.268.4960](tel:412.268.4960) Fax: 412.268.5832

© Pittsburgh Supercomputing Center, All Rights Reserved

Submission Overview



Submission Portal

The Data submission portal is located at
submit.brainimagelibrary.org

Portal allows you to:

- Create a data submission *Take note of your submission ID for following exercises*
- Create a landing space for your data to live
- Access detailed view of your data submission information

Submitting Data: Create Your Data Submission

Create data submission through the submit.brainimagelibrary.org submission portal

Creates a new directory under your landing zone page on the BIL Filesystem

The screenshot shows the Brain Image Library submission interface. At the top, there's a dark header bar with the BIL logo, user info (lunis6), and a 'Log out' button. Below it is a secondary navigation bar with 'New', 'View', and 'Submit Publish Request' dropdowns. A 'Return to PI Dashboard' button is also present. The main content area is titled 'New Submission' and is described as 'Step 1 of 2: Create new submission and then upload metadata'. It contains several input fields: 'Name:' (with a red asterisk), 'Description:' (with a red asterisk), 'Organization name:' (with a red asterisk), 'Lab name:' (with a red asterisk), 'Project funder:' (with a red asterisk, containing 'NIH'), 'Project funder id:' (with a red asterisk), and 'Project:' (with a red asterisk). A note at the bottom states: 'Once the submission is created, a data staging area will be assigned. The path to the data staging area will be shown for the submission after you click on the "Save" button below.' A note at the very bottom says 'Required fields are marked with an *.' with buttons for 'Cancel' and 'Save'.

Hands on Exercise 1

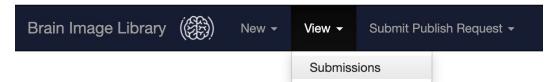
1. Navigate to submit.brainimagelibrary.org in your browser
2. Click Login and use the password you've set up when you created a PSC user account.
3. On the top left-hand corner of your screen – navigate to “New” and on the drop down select “New Submission”
4. Complete the New Submission form to create your first Data Submission, the more descriptive the better.

Submitting Data: Data Submission Created

Your new data submission has now been created.

Note under data path, a new directory has been created with a unique identifier that will always point to your data submission.

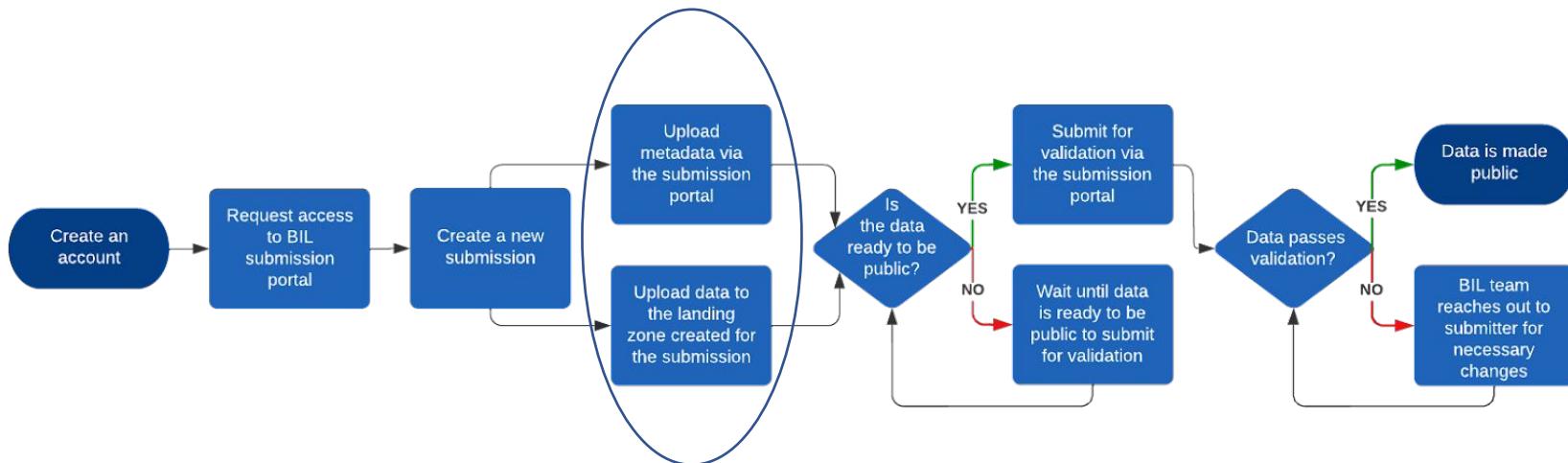
This data path is where you will be directing your upload to.

A screenshot of the submission details page for "Mouse Transcriptomics". The page shows the following fields:

- Submission Name: Mouse Transcriptomics
- Submission Description: 120 Mouse Transcriptomics Images
- Organization: PSC
- Lab: PSC
- Project funder: NIH
- Project funder ID: R24 Hoylewski
- Data staging area: /bio/biuh9e9a7c2419c283dc
- Validation Status: Not validated
- Submission Status: Not submitted

A note at the bottom states: "This submission doesn't have any metadata yet." At the bottom right are buttons for "Edit", "Delete", and "Cancel".

Submission Overview



Uploading Files to BIL

Several options are available transferring files depending on your operating system

Windows?

1. SFTP/SCP
 - o [WinSCP](#), [FileZilla](#) (User Interface applications)
2. [Globus](#)

MacOS X?

1. SFTP/SCP
 - o [FileZilla](#), [Fetch](#)
2. Rsync (Command Line)
3. [Globus](#)

Linux?

1. SFTP/SCP (Interfaces available depending on Linux Distribution)
2. Rsync (Command Line)
3. [Globus](#)

Uploading Data to data submission Directory

Methods in
uploading data:
rsync, sftp, scp,
globus

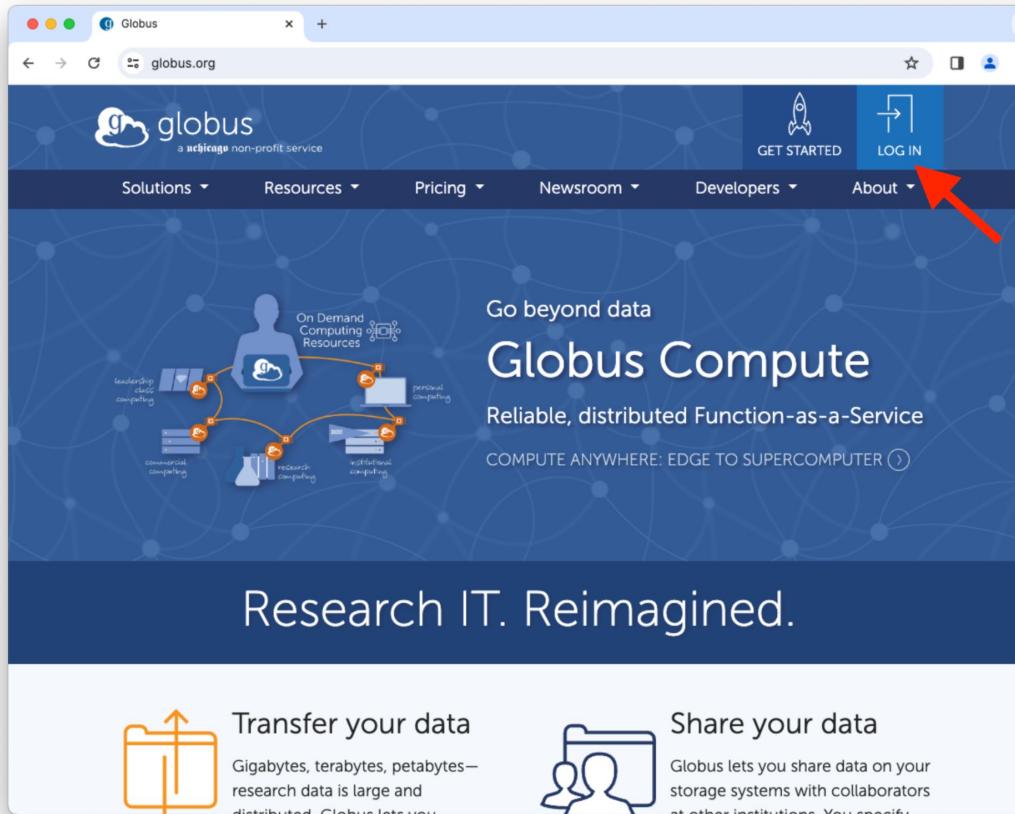
```
$ rsync -lrtpDvP mouse1 testuser@upload.brainimagelibrary.org:/bil/lz/testuser/abcdef0123456789
sending incremental file list
mouse1/
mouse1/data1.tiff
    1356122 100% 126.20MB/s      0:00:00 (xfer#1, to-check=0/2)

sent 1356392 bytes received 35 bytes  2712854.00 bytes/sec
total size is 1356122 speedup is 1.00
```

Using Globus for File Transfer to BIL

1. Go to <https://cilogon.org/>.
2. Select your institution from the 'Select an Identity Provider' list. If your Institution is not among those listed, but you have an ACCESS ID (formerly XSEDE username), then use ACCESS ID as your CILogon Identity Provider.
3. Click the 'Log On' button. You will be taken to your institutional login page.
4. Log in with your username and password for your institution.
 - If your institution has an additional login requirement (e.g., Duo), authenticate to that as well.
5. You will be returned to the CILogon webpage after successfully authenticating your institution's credentials.
6. Click on the User Attributes drop-down link to find the 'ePPN'. Select and copy the ePPN string (which typically looks like an e-mail address) to include in your e-mail to bil-support@psc.edu. If your CILogon ePPN string is blank, please let us know that, and also which CILogon Identity Provider you selected.
7. Send an email to bil-support@psc.edu with your CILogon Certificate Subject and ePPN fields, asking that they be mapped to your BIL username for Globus GridFTP data transfers.

Using Globus for File Transfer to BIL



The screenshot shows the official Globus website at globus.org. The top navigation bar includes links for "Solutions", "Resources", "Pricing", "Newsroom", "Developers", and "About". A prominent blue "LOG IN" button is highlighted with a red arrow pointing towards it. The main content area features a diagram illustrating a network of computing resources: leadership class computing, commercial computing, research computing, institutional computing, personal computing, and On Demand Computing Resources. The text "Go beyond data" and "Globus Compute" is displayed, along with the tagline "Reliable, distributed Function-as-a-Service" and the phrase "COMPUTE ANYWHERE: EDGE TO SUPERCOMPUTER". Below this, a dark blue banner contains the slogan "Research IT. Reimagined." At the bottom, there are two sections: "Transfer your data" (with an icon of a folder and arrow) and "Share your data" (with an icon of a folder and two user profiles).

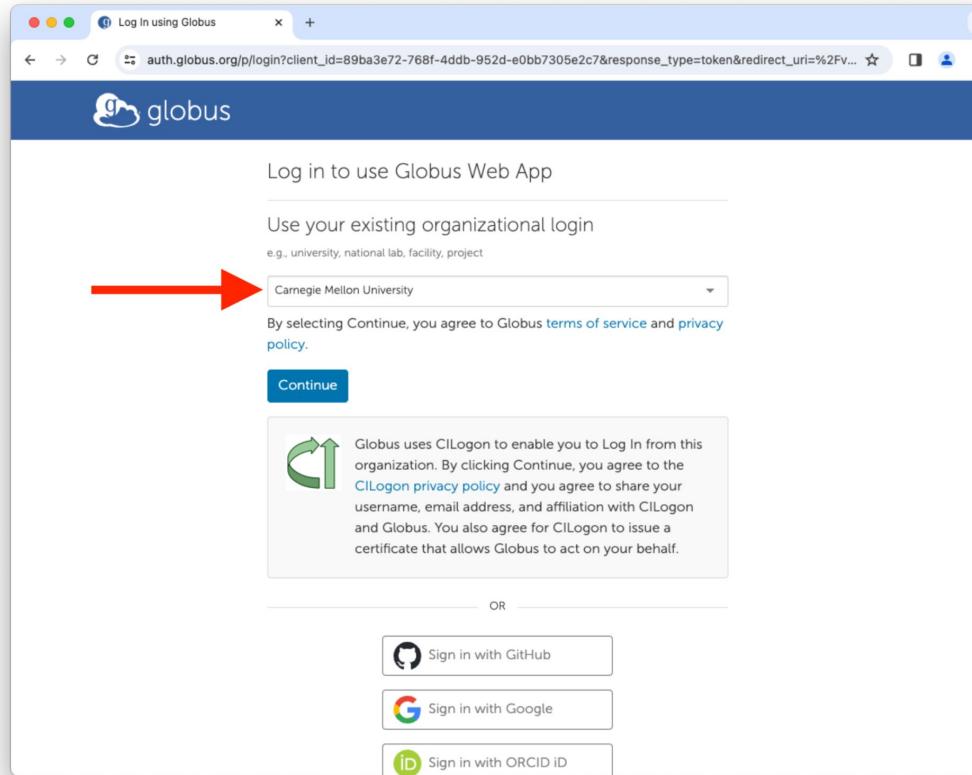
Transfer your data

Gigabytes, terabytes, petabytes—research data is large and distributed. Globus lets you

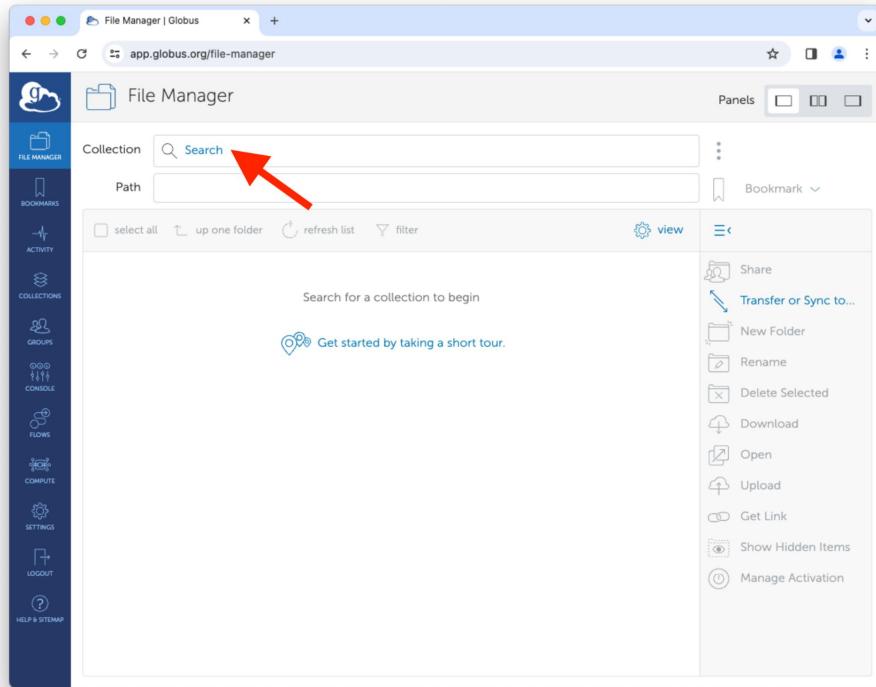
Share your data

Globus lets you share data on your storage systems with collaborators at other institutions. You specify

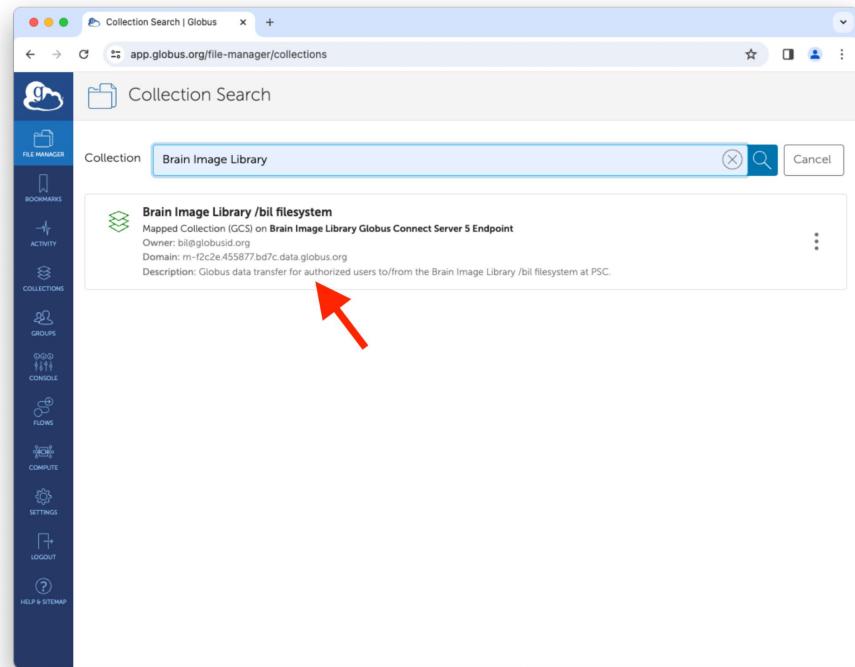
Using Globus for File Transfer to BIL



Using Globus for File Transfer to BIL



The screenshot shows the Globus File Manager interface. On the left is a sidebar with various icons for FILE MANAGER, BOOKMARKS, ACTIVITY, COLLECTIONS, GROUPS, CONSOLE, FLOWS, COMPUTE, SETTINGS, LOGOUT, and HELP & SITEMAP. The main area has tabs for 'FILE MANAGER' and 'COLLECTIONS'. A search bar at the top has the placeholder 'Search' and a magnifying glass icon. Below it is a 'Path' input field. A context menu is open on the right, listing options like Share, Transfer or Sync to..., New Folder, Rename, Delete Selected, Download, Open, Upload, Get Link, Show Hidden Items, and Manage Activation. At the bottom, there's a message: 'Search for a collection to begin' and a 'Get started by taking a short tour.' button.



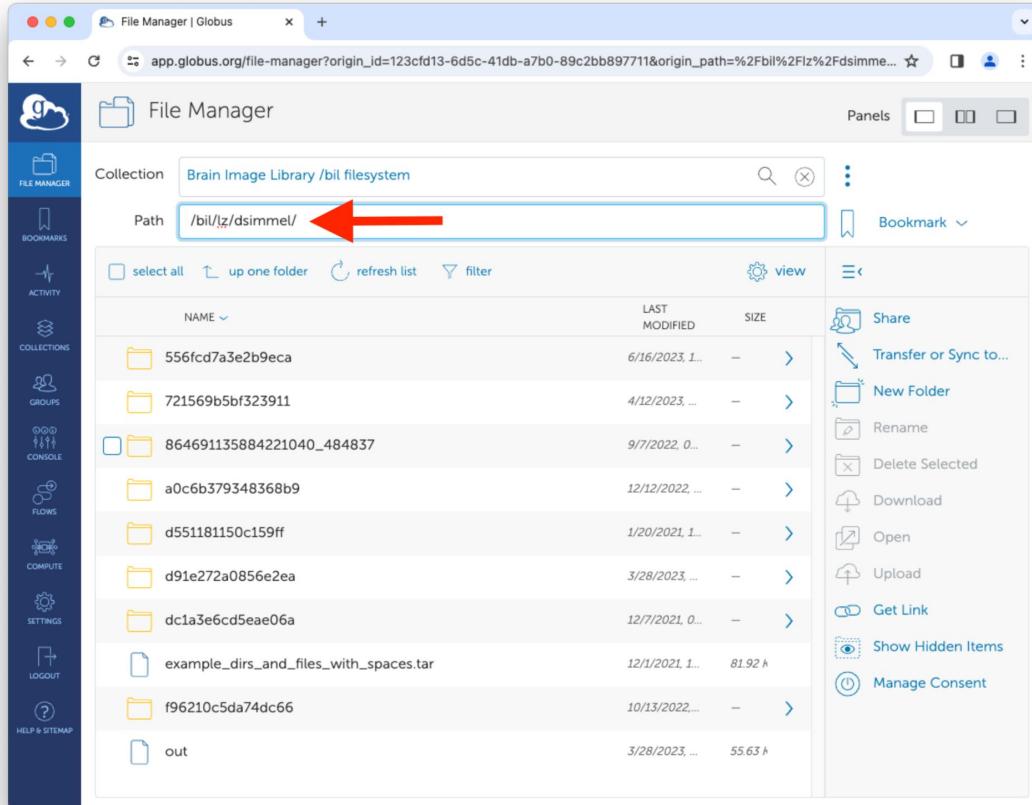
The screenshot shows the Collection Search interface. The sidebar on the left is identical to the one in the first screenshot. The main area shows a search result for 'Collection' with the text 'Brain Image Library'. Below it, a card displays the details for 'Brain Image Library /bil filesystem'. The card includes the following information: 'Mapped Collection (GCS) on Brain Image Library Globus Connect Server 5 Endpoint', 'Owner: bil@globusid.org', 'Domain: m-f2c2e455877bd7c.data.globus.org', and 'Description: Globus data transfer for authorized users to/from the Brain Image Library /bil filesystem at PSC.'. A red arrow points to this card.

Using Globus for File Transfer to BIL

A screenshot of a web browser showing the Globus authentication interface. The title bar says "Authentication Required using Globus". The main content area has a blue header "globus" with a cloud icon. Below it, the text "Identity Required" is displayed. A message states: "An identity from one of the following identity providers is required to continue. Reason: Brain Image Library /bil filesystem requires your consent to allow data access". It asks the user to "Please select the identity or identity provider to continue:" and lists two options: "dsimmel@access-ci.org" and "dsimmel@andrew.cmu.edu". A red arrow points to the second option. At the bottom, there is a "Continue" button with a red arrow pointing to it. A search bar at the bottom says "Look-up your organization..".

A screenshot of a web browser showing the Globus consent request interface. The title bar says "Consent Request using Globus". The main content area has a blue header "globus" with a cloud icon. It asks "Globus Web App would like to:" and lists three items with checked boxes: "Manage data using Globus Transfer", "Manage collections on Brain Image Library Globus Connect Server 5 Endpoint", and "Access your data on Brain Image Library /bil filesystem via HTTPS". Below this, it says "To work, the above will need to: ▾". A message at the bottom explains: "By clicking "Allow", you allow **Globus Web App**, in accordance with its [terms of service](#) and [privacy policy](#), to use the above listed information and services. You can rescind this and other [consents](#) at any time." A red arrow points to the "Allow" button.

Using Globus for File Transfer to BIL



The screenshot shows the Globus File Manager interface. The left sidebar contains icons for FILE MANAGER, BOOKMARKS, ACTIVITY, COLLECTIONS, GROUPS, CONSOLE, FLOWS, COMPUTE, SETTINGS, LOGOUT, and HELP & SITEMAP. The main area is titled "File Manager" and shows a "Collection" named "Brain Image Library /bil filesystem". The "Path" field contains the value "/bil/lz/dsimmel/" with a red arrow pointing to it. Below the path is a search bar and a refresh list button. The main table lists files and folders in the directory:

NAME	LAST MODIFIED	SIZE	Actions
556fc7a3e2b9eca	6/16/2023, 1...	-	>
721569b5bf323911	4/12/2023, ...	-	>
864691135884221040_484837	9/7/2022, 0...	-	>
a0c6b379348368b9	12/12/2022 ...	-	>
d551181150c159ff	1/20/2021, 1...	-	>
d91e272a0856e2ea	3/28/2023, ...	-	>
dc1a3e6cd5eae06a	12/7/2021, 0...	-	>
example_dirs_and_files_with_spaces.tar	12/1/2021, 1...	81.92 k	
f96210c5da74dc66	10/13/2022, ...	-	>
out	3/28/2023, ...	55.63 k	

The right side of the interface shows a context menu with options: Share, Transfer or Sync to..., New Folder, Rename, Delete Selected, Download, Open, Upload, Get Link, Show Hidden Items, and Manage Consent.

Using SCP/SFTP for File Transfer to BIL

Using scp to upload.brainimagelibrary.org: An example uploading data (data1.tiff) to the landing zone as testuser is shown below:

```
$ scp data1.tiff testuser@upload.brainimagelibrary.org:/bil/lz/abcdef0123456789/mouse1/data1.tiff
```

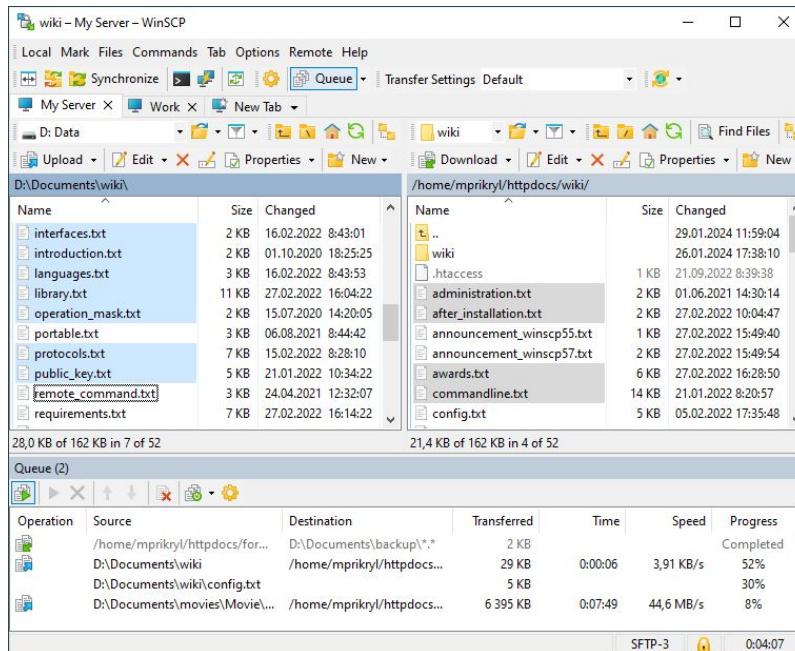
```
testuser@upload.brainimagelibrary.org's password:
```

```
data1.tiff                                0%   0      0.0KB/s
---:-- ETA
data1.tiff                                100% 1324KB  1.3MB/s
00:00
```

Using SCP/SFTP for File Transfer to BIL

WinSCP is a free software for SCP and SFTP file transfer with a user interface.

<https://winscp.net/eng/docs/start>



Hands on Exercise 2

Open terminal on local machine

- On Mac: Spotlight Search > “**Terminal**” > Enter
- On Windows: Search > “**cmd**” > Enter
- On Linux: **Ctrl + Alt + T**

Create a dummy file by running

1. Navigate to your home directory. Run: **cd**
2. Create a text file. Run **Cat > myfile.txt**
 - a. Type some text “**This is my new file**”
 - b. Type **Control + D** to finish writing to the file.

Alternatively, just create a text file with your preferred editor

Hands on Exercise 2

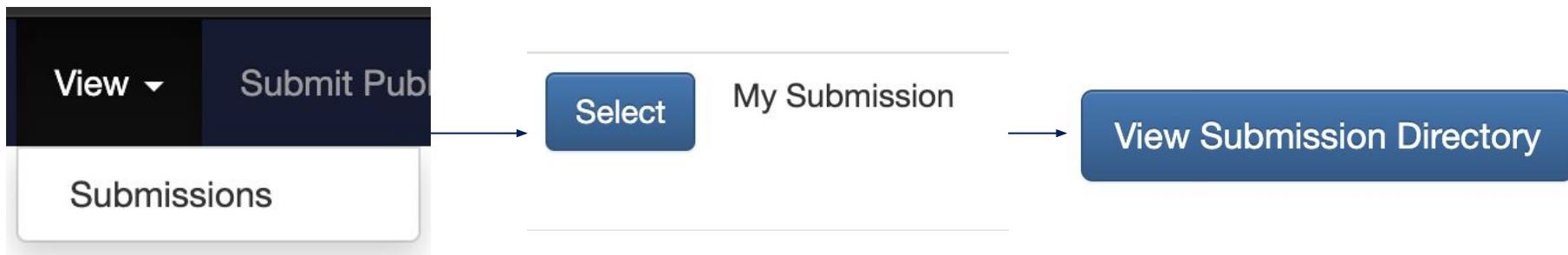
With our file created, let's upload it to your new BIL Submission

Run: `scp myfile.txt *username*@upload.brainimagelibrary.org:/bil/lz/*username*/UUID*/`

File Cleanup/Organization using OnDemand

Files can be edited/removed/restructured within Open OnDemand as well as your command line interface

Viewing and Editing your data submission is easily accessible inside the BIL Submission Portal by choosing a submission and clicking "View Submission Directory"



File Cleanup/Organization using OnDemand

Landing Zone directory is easily accessible through OnDemand for organization

Will link to your Landing Zone during Staging and later to a Public Directory after publication

The screenshot shows a file browser interface with the following elements:

- Path bar: / bil / lz / ltuite96 / a2ebd6758d8cc705 /
- Action buttons: Change directory (with a pencil icon) and Copy path (with a clipboard icon).
- Filtering options: Show Owner/Mode, Show Dotfiles, and a Filter input field.
- Status message: Showing 0 rows - 0 rows selected.
- Table header: Type, Name, Size, Modified at.
- Table body: No data available in table.



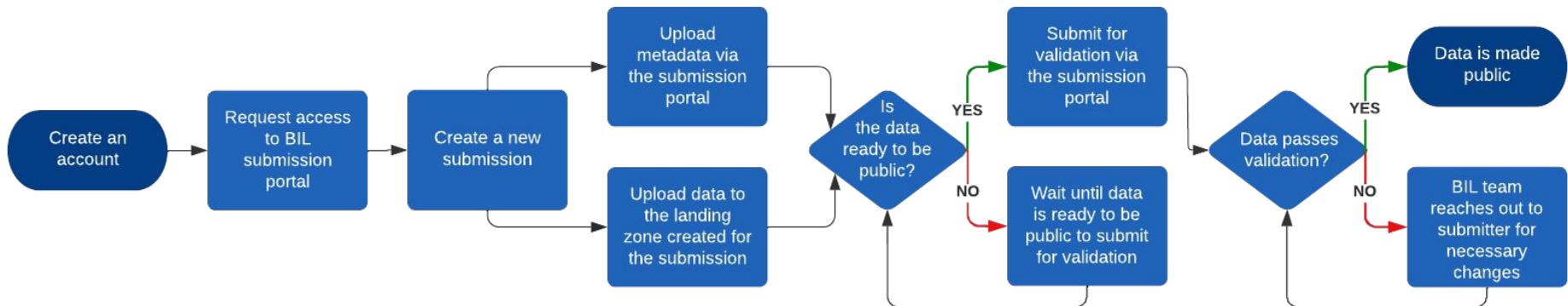
Carnegie
Mellon
University



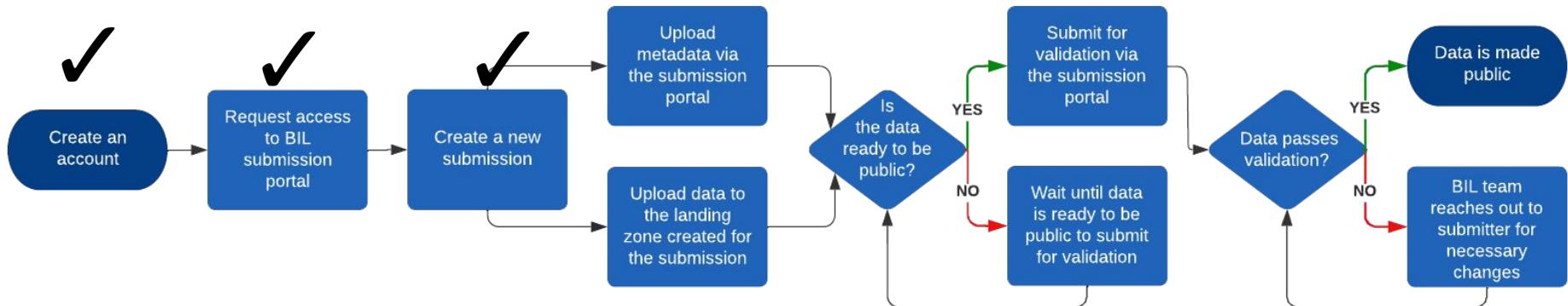
File Organization

Mariah Kenney

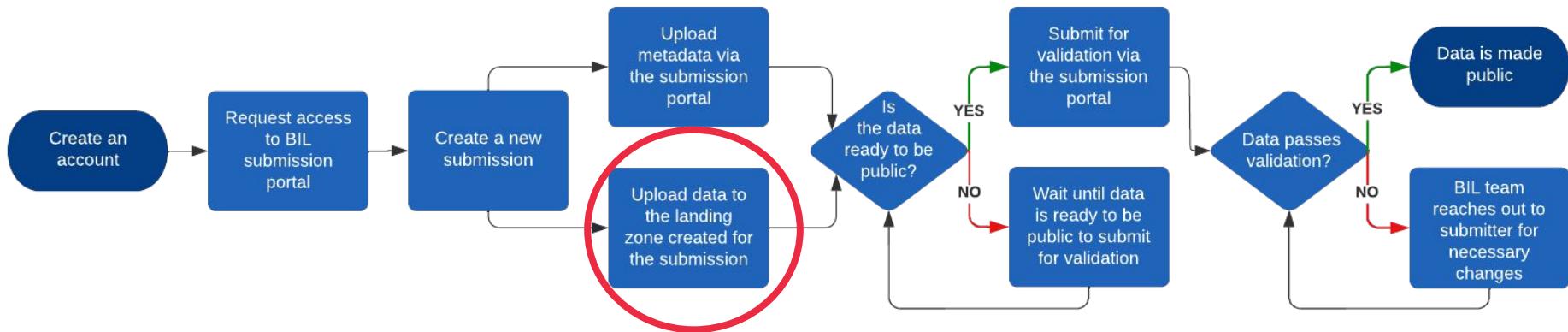
Submission Overview



Submission Overview



Submission Overview



Dataset vs. Submission

Dataset

- A stand-alone entry of an image-volume or image-set associated with a single subject or experimental unit with unique metadata.
- A single dataset is usually associated with a single donor or subject when submitting multiple subjects in a submission, or a single part of the brain when imaging many parts of the brain in a submission.
- Many datasets make up a submission.
- A dataset typically contains many 2d image files that are assembled to form a more complex two or three dimensional volume.

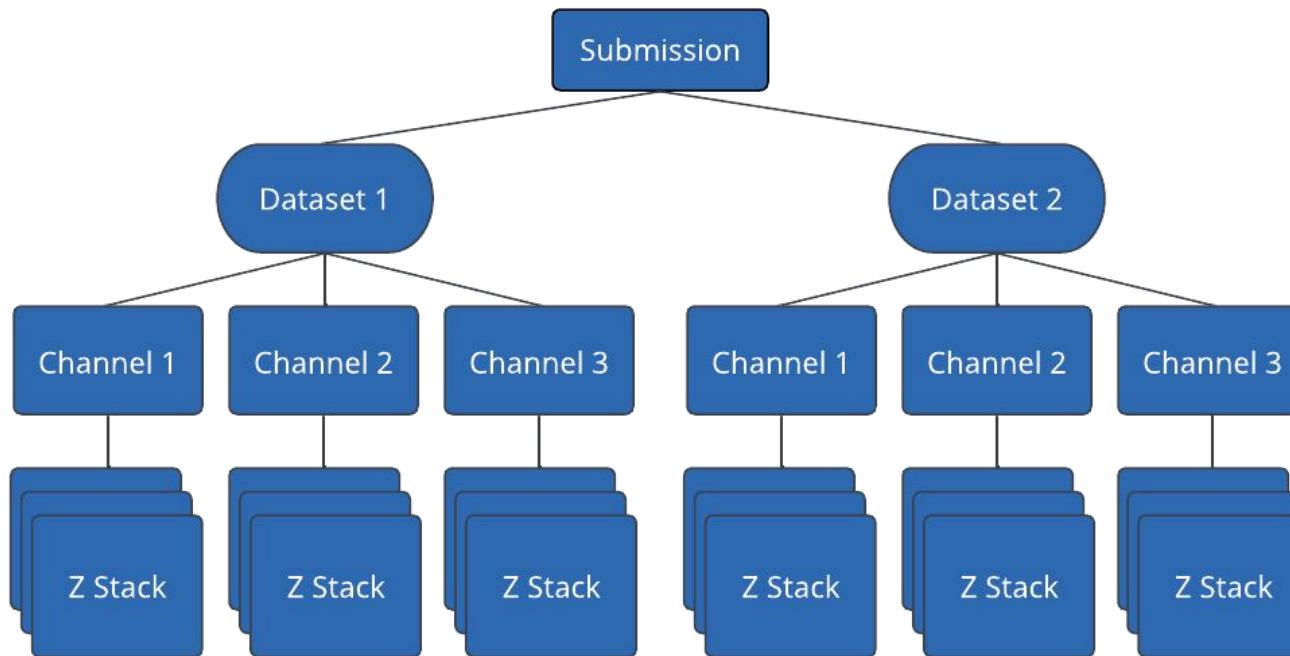
Submission

- Contains one or more related datasets and the associated metadata.
- Submissions will inherit project metadata (such as the NIH project, grant number, laboratory name, etc.), thus all datasets within a submission must belong to the same project.
- In general, smaller submissions are recommended because all datasets within a submission must pass the validation process for the datasets within the submission to be published.
- Each “level” of data should be uploaded in separate submissions (e.g. The set of raw data and the same data aligned to a reference are considered two separate submissions).

Dataset vs. Submission

- 1. Many datasets make up a single submission**
- 2. Each dataset represent a single donor/individual or experimental unit**

How to structure your data?



Best practices

Perform general file cleanup after data transfer

We suggest reviewing the files that have uploaded before you submit for validation. You can do this in your terminal or use tools such as OnDemand.

Consider file format

Avoid proprietary file formats

Avoid problematic special characters

Characters that are generally safe to include in filenames other than numbers and letters:

Underscore _

Dash -

Period .

Create unique directories for each dataset

Each dataset must have their own unique path that does not lead to an individual file, but rather a directory.

Additional Files

If you have additional metadata or files that you want to include with each dataset, place it in a subdirectory called extras inside each subdirectory:

/bil/lz/testuser/abcdef0123456789/mouse1/extras



Carnegie
Mellon
University



Break

15 Mins



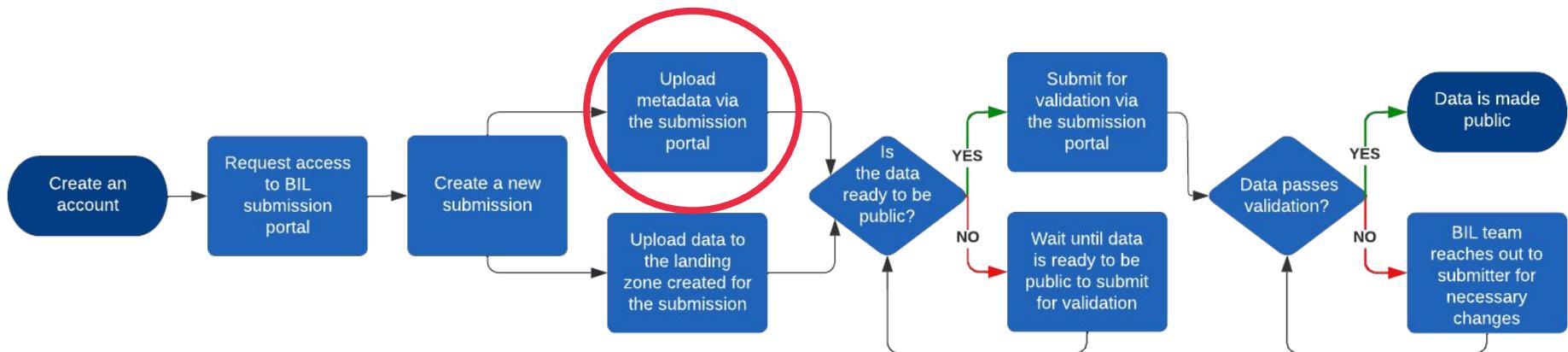
Brain Image Library Metadata



*How to share data and metadata that
serves the scientific community*

Mariah Kenney
Data Curator and Metadata Librarian

Submission Overview



What makes the BIL metadata practical?

- Goal: Increase FAIRness of microscopy data and encourage reuse and ensure data is useful to the scientific community (**F**indability, **A**ccessibility, **I**nteroperability, and **R**euse)
- Metadata schema is designed with the ability to issue DOIs in mind – BIL will be issuing dataset DOI's automatically (instead of by-request) in the near future
- BIL's initial implementation is spreadsheet-oriented
- Metadata models allow redundant fields to be removed for ease of data entry
- New metadata schema developed by BRAIN Initiative: Standard metadata for 3D microscopy.
DOI: [10.1038/s41597-022-01562-5](https://doi.org/10.1038/s41597-022-01562-5)

Overview

Tab	Description
Contributors	The Contributors record is used to identify and give credit for those contributing to the project. There can be multiple contributors – in this case the contributor record should be listed in order of importance. Contributors can be individuals or institutional groups.
Funders	Funders provide financial support for the project.
Publication	Related Pre-Prints, Publications, and Protocols associated with the dataset that have globally unique identifiers
Instrument	These fields define the capture instrument used
Dataset	Dataset title, Dataset Abstract, Directory (landing zone) linkage at BIL
Specimen	Information about the specimen (Donor, Organ, Samples)
Image	General information describing the dataset's images, including capture orientation, landmarks, channels and sizes.
SWC (optional)	Information on neuron tracing datasets in standardized swc format

Metadata Models

- You will choose a metadata model when you upload your metadata.
- Metadata models describe the structure of the submission and how the datasets are related to each other.
- Metadata models are used to ensure that your data is properly ingested and indexed in the database.
- Metadata will not be able to be uploaded if the metadata model selected and the structure of the metadata file is not correct.
- There are 5 metadata models - described in the coming slides

Metadata Models

Brain Image Library  New ▾ View ▾ Submit Publish Request ▾

kenneyml  Log out

Upload Metadata Spreadsheet

Please note that BIL has adopted a new metadata schema. Please use the template provided below.

Step 2 of 3: What does your data look like?

- 1. Each dataset was generated from a single sample and donor (e.g. whole brain imaging) on the same instrument. Multiple datasets can be accommodated in a single submission. The lines of the specimen tab in the metadata spreadsheet must match exactly to the lines in the dataset tab (i.e. the specimen listed in specimen tab row 5 is the specimen for dataset row 5). Each dataset tab should have a corresponding image tab. There should be exactly one line in the instrument tab.
- 2. The dataset was generated from multiple samples, specimens, or donors on the same instrument. Only a single dataset can be accommodated per submission. Multiple specimens should be listed in the metadata spreadsheet, but only a single dataset. There should be one line in the image tab. There should be exactly one line in the instrument tab.
- 3. All datasets were generated from one specimen, but are from unique regions of interest on the specimen. Multiple datasets can be accommodated in a single submission. List the specimen multiple times in the specimen tab. Identify the specific region of interest in the locations column. Multiple donor lines must match exactly in the metadata spreadsheet (i.e. the specimen listed in specimen tab row 5 is the specimen for dataset row 5). Each dataset tab should have a corresponding image tab. There should be exactly one line in the instrument tab.
- 4. All datasets were generated from one specimen – but use different experimental (machinery) parameters. Multiple datasets can be accommodated as a single submission. There should be exactly one entry in the specimen tab. The lines listed in the Instrument tab in the metadata spreadsheet must match exactly to the lines in the dataset tab (i.e. the instrument listed in instrument tab row 5 is the instrument for dataset row 5). Each dataset tab should have a corresponding image tab.
- 5. Reconstructed neurons in standardized swc format. All reconstructions were generated from one specimen. There should be one line listed in the specimen tab. There should be one line listed in the dataset tab. No information should be listed in the image tab. File format specification: <http://www.neuronland.org/NLMorphologyConverter/MorphologyFormats/SWC/Spec.html>

Step 3 of 3: Upload metadata for associated submission

Download and fill out either the [Excel](#) or [LibreOffice Calc](#) template.

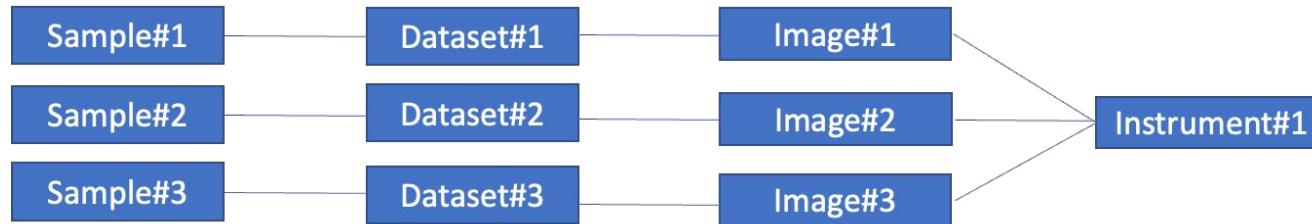
Choose a collection, then upload your metadata.

Associated collection:

[Upload Metadata](#)

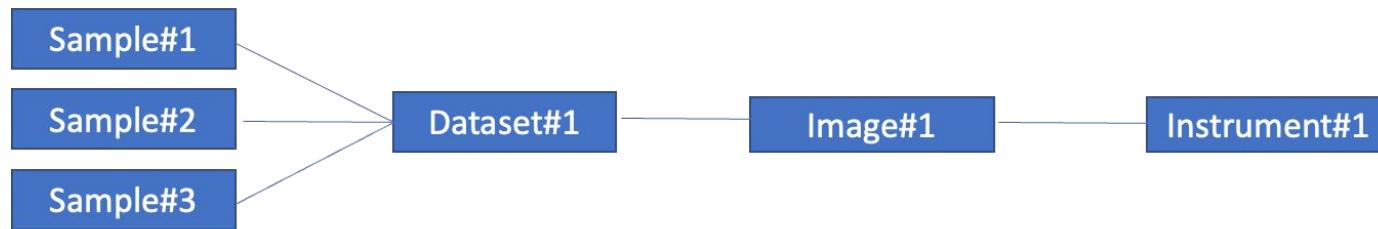
Metadata Model 1

1. Each dataset was generated from a single sample and donor (e.g. whole brain imaging) on the same instrument. Multiple datasets can be accommodated in a single submission. The lines of the specimen tab in the metadata spreadsheet must match exactly to the lines in the dataset tab (i.e. the specimen listed in specimen tab row 5 is the specimen for dataset row 5). Each dataset tab should have a corresponding image tab. There should be exactly one line in the instrument tab.



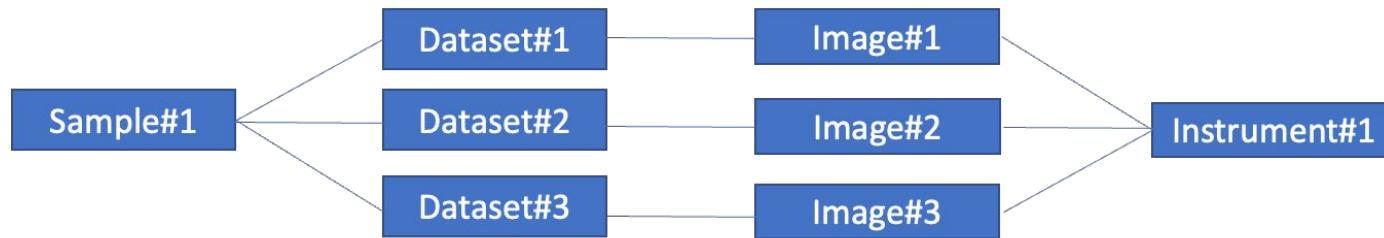
Metadata Model 2

2. The dataset was generated from multiple samples, specimens, or donors on the same instrument. Only a single dataset can be accommodated per submission. Multiple specimens should be listed in the metadata spreadsheet, but only a single dataset. There should be one line in the image tab. There should be exactly one line in the instrument tab.



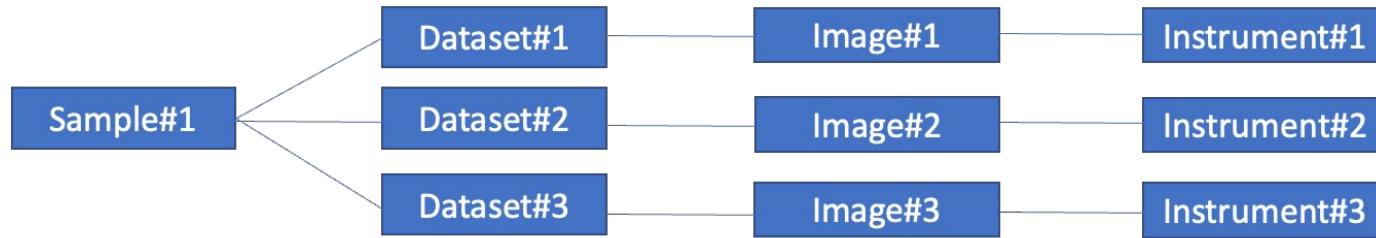
Metadata Model 3

3. All datasets were generated from one specimen, but are from unique regions of interest on the specimen. Multiple datasets can be accommodated in a single submission. List the specimen multiple times in the specimen tab. Identify the specific region of interest in the locations column. Multiple donor lines must match exactly in the metadata spreadsheet (i.e. the specimen listed in specimen tab row 5 is the specimen for dataset row 5). Each dataset tab should have a corresponding image tab. There should be exactly one line in the instrument tab.



Metadata Model 4

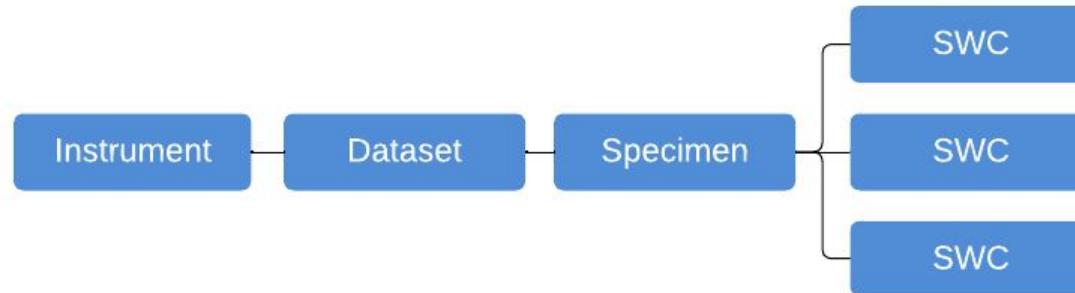
4. All datasets were generated from one specimen – but use different experimental (machinery) parameters. Multiple datasets can be accommodated as a single submission. There should be exactly one entry in the specimen tab. The lines listed in the Instrument tab in the metadata spreadsheet must match exactly to the lines in the dataset tab (i.e. the instrument listed in instrument tab row 5 is the instrument for dataset row 5). Each dataset tab should have a corresponding image tab.



Metadata Model 5

5. Reconstructed neurons in standardized swc format. All reconstructions were generated from one specimen. There should be one line listed in the specimen tab. There should be one line listed in the dataset tab. No information should be listed in the image tab. File format specification:

<http://www.neuronland.org/NLMorphologyConverter/MorphologyFormats/SWC/Spec.html>



Example 1

8 image files from 4 different samples (i.e. different mice) of the same part of the brain using the same microscope imaging settings.

Metadata Models

1. Each dataset was generated from a single sample and donor (e.g. whole brain imaging) on the same instrument. Multiple datasets can be accommodated in a single submission. The lines of the specimen tab in the metadata spreadsheet must match exactly to the lines in the dataset tab (i.e. the specimen listed in specimen tab row 5 is the specimen for dataset row 5). Each dataset tab should have a corresponding image tab. There should be exactly one line in the instrument tab.
2. The dataset was generated from multiple samples, specimens, or donors on the same instrument. Only a single dataset can be accommodated per submission. Multiple specimens should be listed in the metadata spreadsheet, but only a single dataset. There should be one line in the image tab. There should be exactly one line in the instrument tab.
3. All datasets were generated from one specimen, but are from unique regions of interest on the specimen. Multiple datasets can be accommodated in a single submission. List the specimen multiple times in the specimen tab. Identify the specific region of interest in the locations column. Multiple donor lines must match exactly in the metadata spreadsheet (i.e. the specimen listed in specimen tab row 5 is the specimen for dataset row 5). Each dataset tab should have a corresponding image tab. There should be exactly one line in the instrument tab.
4. All datasets were generated from one specimen – but use different experimental (machinery) parameters. Multiple datasets can be accommodated as a single submission. There should be exactly one entry in the specimen tab. The lines listed in the Instrument tab in the metadata spreadsheet must match exactly to the lines in the dataset tab (i.e. the instrument listed in instrument tab row 5 is the instrument for dataset row 5). Each dataset tab should have a corresponding image tab.
5. Reconstructed neurons in standardized swc format. All reconstructions were generated from one specimen. There should be one line listed in the specimen tab. There should be one line listed in the dataset tab. No information should be listed in the image tab.

Metadata Models

1. Each dataset was generated from a single sample and donor (e.g. whole brain imaging) on the same instrument. Multiple datasets can be accommodated in a single submission. The lines of the specimen tab in the metadata spreadsheet must match exactly to the lines in the dataset tab (i.e. the specimen listed in specimen tab row 5 is the specimen for dataset row 5). Each dataset tab should have a corresponding image tab. There should be exactly one line in the instrument tab.
2. The dataset was generated from multiple samples, specimens, or donors on the same instrument. Only a single dataset can be accommodated per submission. Multiple specimens should be listed in the metadata spreadsheet, but only a single dataset. There should be one line in the image tab. There should be exactly one line in the instrument tab.
3. All datasets were generated from one specimen, but are from unique regions of interest on the specimen. Multiple datasets can be accommodated in a single submission. List the specimen multiple times in the specimen tab. Identify the specific region of interest in the locations column. Multiple donor lines must match exactly in the metadata spreadsheet (i.e. the specimen listed in specimen tab row 5 is the specimen for dataset row 5). Each dataset tab should have a corresponding image tab. There should be exactly one line in the instrument tab.
4. All datasets were generated from one specimen – but use different experimental (machinery) parameters. Multiple datasets can be accommodated as a single submission. There should be exactly one entry in the specimen tab. The lines listed in the Instrument tab in the metadata spreadsheet must match exactly to the lines in the dataset tab (i.e. the instrument listed in instrument tab row 5 is the instrument for dataset row 5). Each dataset tab should have a corresponding image tab.
5. Reconstructed neurons in standardized swc format. All reconstructions were generated from one specimen. There should be one line listed in the specimen tab. There should be one line listed in the dataset tab. No information should be listed in the image tab.

Example 1

8 image files from 4 different samples (i.e. different mice) of the same part of the brain using the same microscope imaging settings.

Data files to upload:

Mouse1_brain_1.tif

Mouse1_brain_2.tif

Mouse2_brain_1.tif

Mouse2_brain_2.tif

Mouse3_brain_1.tif

Mouse3_brain_2.tif

Mouse4_brain_1.tif

Mouse4_brain_2.tif

Example 1

8 image files from 4 different samples (i.e. different mice) of the same part of the brain using the same microscope imaging settings.

Data files to upload:

Mouse1_brain_1.tif

Mouse1_brain_2.tif

Mouse2_brain_1.tif

Mouse2_brain_2.tif

Mouse3_brain_1.tif

Mouse3_brain_2.tif

Mouse4_brain_1.tif

Mouse4_brain_2.tif



Directory structure:

/bil/lz/userid/submissionid/Mouse1

/bil/lz/userid/submissionid/Mouse2

/bil/lz/userid/submissionid/Mouse3

/bil/lz/userid/submissionid/Mouse4

Example 1

8 image files from 4 different samples (i.e. different mice) of the same part of the brain using the same microscope imaging settings.

Let's look at how this would be organized in the spreadsheet and be submitted.

Example 2

Image sets taken at 2 different magnifications from the same mouse

Metadata Models

1. Each dataset was generated from a single sample and donor (e.g. whole brain imaging) on the same instrument. Multiple datasets can be accommodated in a single submission. The lines of the specimen tab in the metadata spreadsheet must match exactly to the lines in the dataset tab (i.e. the specimen listed in specimen tab row 5 is the specimen for dataset row 5). Each dataset tab should have a corresponding image tab. There should be exactly one line in the instrument tab.
2. The dataset was generated from multiple samples, specimens, or donors on the same instrument. Only a single dataset can be accommodated per submission. Multiple specimens should be listed in the metadata spreadsheet, but only a single dataset. There should be one line in the image tab. There should be exactly one line in the instrument tab.
3. All datasets were generated from one specimen, but are from unique regions of interest on the specimen. Multiple datasets can be accommodated in a single submission. List the specimen multiple times in the specimen tab. Identify the specific region of interest in the locations column. Multiple donor lines must match exactly in the metadata spreadsheet (i.e. the specimen listed in specimen tab row 5 is the specimen for dataset row 5). Each dataset tab should have a corresponding image tab. There should be exactly one line in the instrument tab.
4. All datasets were generated from one specimen – but use different experimental (machinery) parameters. Multiple datasets can be accommodated as a single submission. There should be exactly one entry in the specimen tab. The lines listed in the Instrument tab in the metadata spreadsheet must match exactly to the lines in the dataset tab (i.e. the instrument listed in instrument tab row 5 is the instrument for dataset row 5). Each dataset tab should have a corresponding image tab.
5. Reconstructed neurons in standardized swc format. All reconstructions were generated from one specimen. There should be one line listed in the specimen tab. There should be one line listed in the dataset tab. No information should be listed in the image tab.

Metadata Models

1. Each dataset was generated from a single sample and donor (e.g. whole brain imaging) on the same instrument. Multiple datasets can be accommodated in a single submission. The lines of the specimen tab in the metadata spreadsheet must match exactly to the lines in the dataset tab (i.e. the specimen listed in specimen tab row 5 is the specimen for dataset row 5). Each dataset tab should have a corresponding image tab. There should be exactly one line in the instrument tab.
2. The dataset was generated from multiple samples, specimens, or donors on the same instrument. Only a single dataset can be accommodated per submission. Multiple specimens should be listed in the metadata spreadsheet, but only a single dataset. There should be one line in the image tab. There should be exactly one line in the instrument tab.
3. All datasets were generated from one specimen, but are from unique regions of interest on the specimen. Multiple datasets can be accommodated in a single submission. List the specimen multiple times in the specimen tab. Identify the specific region of interest in the locations column. Multiple donor lines must match exactly in the metadata spreadsheet (i.e. the specimen listed in specimen tab row 5 is the specimen for dataset row 5). Each dataset tab should have a corresponding image tab. There should be exactly one line in the instrument tab.
4. All datasets were generated from one specimen – but use different experimental (machinery) parameters. Multiple datasets can be accommodated as a single submission. There should be exactly one entry in the specimen tab. The lines listed in the Instrument tab in the metadata spreadsheet must match exactly to the lines in the dataset tab (i.e. the instrument listed in instrument tab row 5 is the instrument for dataset row 5). Each dataset tab should have a corresponding image tab.
5. Reconstructed neurons in standardized swc format. All reconstructions were generated from one specimen. There should be one line listed in the specimen tab. There should be one line listed in the dataset tab. No information should be listed in the image tab.

Example 2

Image sets taken at 2 different magnifications from the same mouse

Data files:

Mouse10x_1.tif

Mouse10x_2.tif

Mouse10x_3.tif

Mouse30x_1.tif

Mouse30x_2.tif

Mouse30x_3.tif

Example 2

Image sets taken at 2 different magnifications from the same mouse

Data files:

Mouse10x_1.tif

Mouse10x_2.tif

Mouse10x_3.tif

Mouse30x_1.tif

Mouse30x_2.tif

Mouse30x_3.tif



Directory structure:

/bil/lz/userid/submissionid/Mouse10x

/bil/lz/userid/submissionid/Mouse30x

Example 2

Image sets taken at 2 different magnifications from the same mouse

Let's look at how this would be organized in the spreadsheet and be submitted.



Carnegie
Mellon
University



Uploading Metadata and Submission/Validation

Luke Tuite

Upload Metadata Spreadsheet

1. Download the metadata spreadsheet on this page and provide information requested
2. Provide a response for the appropriate Metadata Model matching the information in your spreadsheet
3. Upload your completed spreadsheet with the associated data submission selected

The screenshot shows the Brain Image Library (BIL) website interface. At the top, there is a navigation bar with links for 'New', 'View', and 'Submit Publish Request'. Below the navigation bar, a dropdown menu is open with options 'New Submission' and 'Add Metadata'. The main content area is titled 'Upload Metadata Spreadsheet'. A note below the title states: 'Please note that BIL has adopted a new metadata schema. Please use the template provided below.' Below this note, there is a section titled 'Step 2 of 3: What does your data look like?' containing five numbered options describing different dataset generation scenarios. Further down, another section titled 'Step 3 of 3: Upload metadata for associated submission' provides instructions for uploading the completed spreadsheet, including a link to a template and a dropdown menu for selecting an associated collection.

Brain Image Library  New ▾ View ▾ Submit Publish Request ▾

New Submission
Add Metadata

Brain Image Library  New ▾ View ▾ Submit Publish Request ▾

Upload Metadata Spreadsheet

Please note that BIL has adopted a new metadata schema. Please use the template provided below.

Step 2 of 3: What does your data look like?

- 1. Each dataset was generated from a single sample and donor (e.g. whole brain imaging) on the same instrument. Multiple datasets can be accommodated in a single submission. The lines of the specimen tab in the metadata spreadsheet must match exactly to the lines in the dataset tab (i.e. the specimen listed in specimen tab row 5 is the specimen for dataset row 5). Each dataset tab should have a corresponding image tab. There should be exactly one line in the instrument tab.
- 2. The dataset was generated from multiple samples, specimens, or donors on the same instrument. Only a single dataset can be accommodated per submission. Multiple specimens should be listed in the metadata spreadsheet, but only a single dataset. There should be one line in the image tab. There should be exactly one line in the instrument tab.
- 3. All datasets were generated from one specimen, but are from unique regions of interest on the specimen. Multiple datasets can be accommodated in a single submission. List the specimen multiple times in the specimen tab, identify the specific region of interest in the locations column. Multiple donor lines must match exactly in the metadata spreadsheet (i.e. the specimen listed in specimen tab row 5 is the specimen for dataset row 5). Each dataset tab should have a corresponding image tab. There should be exactly one line in the instrument tab.
- 4. All datasets were generated from one specimen – but use different experimental (machinery) parameters. Multiple datasets can be accommodated as a single submission. There should be exactly one entry in the specimen tab. The lines listed in the instrument tab in the metadata spreadsheet must match exactly to the lines in the dataset tab (i.e. the instrument listed in instrument tab row 5 is the instrument for dataset row 5). Each dataset tab should have a corresponding image tab.
- 5. Reconstructed neurons in standardized swc format. All reconstructions were generated from one specimen. There should be one line listed in the specimen tab. There should be one line listed in the dataset tab. No information should be listed in the image tab. File format specification: <http://www.neuronland.org/NLMorphologyConverter/MorphologyFormats/SWC/Spec.html>

Step 3 of 3: Upload metadata for associated submission

Download and fill out either the [Excel](#) or [LibreOffice Calc](#) template.

Choose a collection, then upload your metadata.

Associated collection:

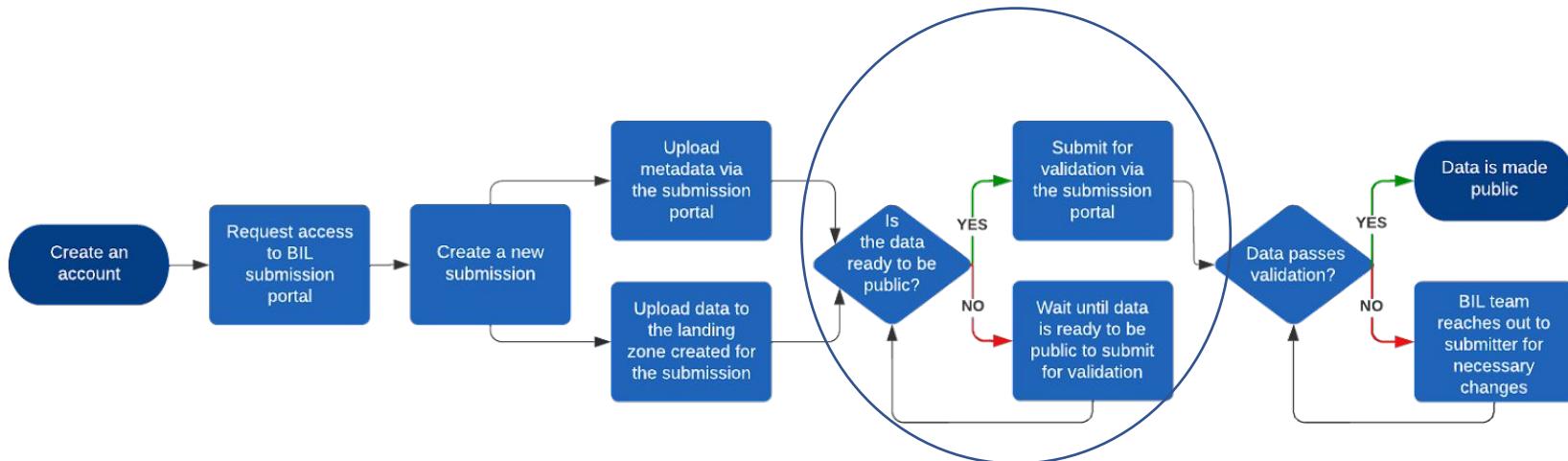
Upload Metadata

Cancel

Hands on exercise 3

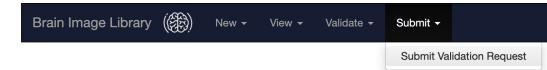
1. On the top left-hand corner of your screen – navigate to “New” and on the drop down select “Descriptive Metadata”
2. On the drop-down menu – select your data submission you’ve just created
3. Using the file downloaded from here
[https://github.com/brain-image-library/workshops/blob/master/2024/january/data submission/metadata model examples/BIL Metadata TemplateV3 Model 1.xlsx](https://github.com/brain-image-library/workshops/blob/master/2024/january/data%20submission/metadata%20model%20examples/BIL%20Metadata%20TemplateV3%20Model%201.xlsx)

Submission Overview



Final Step: Submit a Validation Request/Make Data Public

1. Complete this step when data is ready to be made public
2. Navigate to the Submit Validation Request tab
3. Select the data submission you've uploaded data and metadata for
4. Click "Submit Validation Request"
5. You will be notified when the validation is finished



Submit Validation Request

Submit Request to Publicize Collection

Select checkboxes for each collection to request validation and publication. This will submit a ticket to BIL support and be processed.

Name	Description	Organization	Lab Name	Project ID	Funder	BIL UUID	Data Path	Locked	Submission Status	Validation Status
<input checked="" type="checkbox"/> Whole Brain Cell Distribution March 2021	Whole Brain Cell Distribution Experiment on Agouti/Swiss mice imaged in March 2021	University of Pittsburgh	Watkins Lab	1-xxx-xxxx	NH	32f5403995986bc32	/af/2/tu/HtR0/32f5403995986bc32	False	NOT_SUBMITTED	NOT_VALIDATED

What happens during validation?

Several checks for data sanitization

- Looks for empty files
- Checks and fixes directory name structure i.e. spaces, illegal characters
- Checks and fixes filename structure i.e. spaces, illegal characters
- Images are put through BioFormats to make sure images are useable
- Bioformats reports back if any corrupted data or data with errors is found.
- Metadata is curated and confirmed to match data submitted.

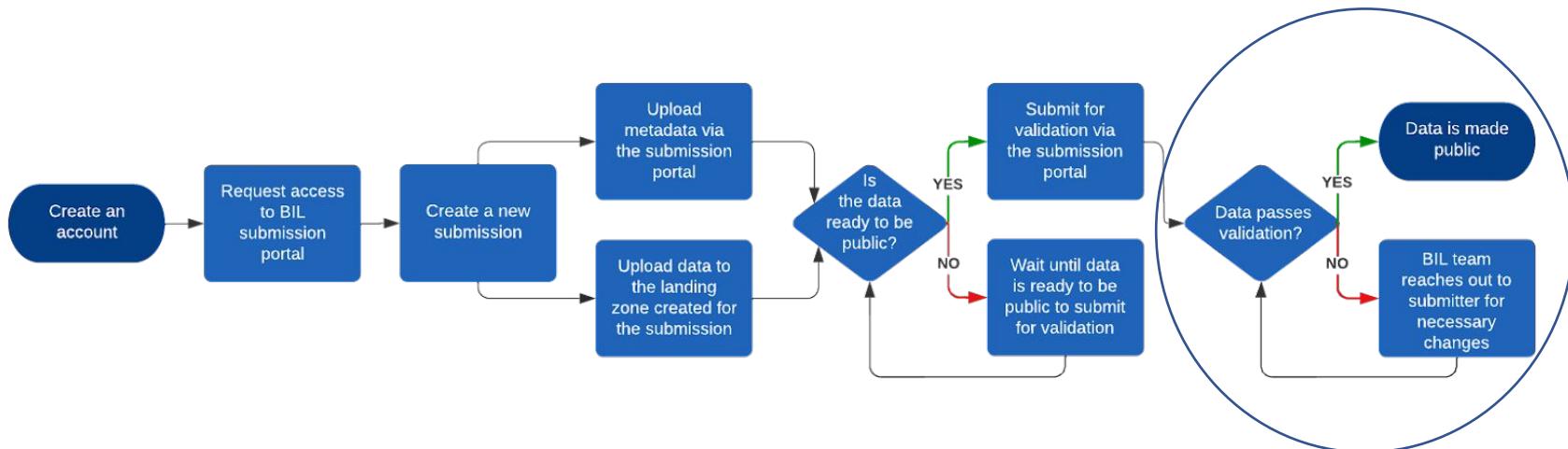
Our data validation does not reject based on quality of data, only on usability.

Validation process duration depends on amount of files and size.

Hands on exercise 4

1. On the top left-hand corner of your screen – navigate to “Submit” and on the drop down select “Submit Validation Request”
2. Click the checkbox on your data submission you’ve created and then click “Submit Validation Request”
3. You’re done!

Submission Overview



Validation timeline

- Begin the submission process in-advance of any deadlines that you may have for making the data public
- Plan for data transfer time
- 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

BIL PI Dashboard

Organize your data by Project and Sub-Project:

- Project creation is mandatory for BICAN projects
- Request PI Dashboard access through the ticket system at bil-support@psc.edu

Create a new project

Project Name:*

Funded By:

Consortia Affiliation:

Select your project's affiliation(s). Choose all that apply.

To choose multiple: Windows users: control-click / Mac users: command-click

Brain Initiative Cell Atlas Network (BICAN)
 Brain Initiative Cell Census Network (BICCN)

Required fields are marked with an *****.

Submit New Project

Cancel

PI's Detailed View

- Provides snapshot of BIL data by project
- The project view will require all Pi's to have a BIL submit portal account.
 - Previously only data submitters required to have a portal account
- Allows management of data submissions submitted for projects on BIL
- Views for statuses on all submissions on projects you oversee, i.e. submission status
- Manage users associated with your project – add users to give them access to submit data under your project name
- Future Tool: downloadable reports for information on all of your project's submitted data

PI's Detailed Project View

Project Details

Manage Projects

Select a project to view details including collections and users associated with it

Name	Funded By	Is BICCN?	Personnel	Collections
Ituite96 Project 2	Not Applicable	True	View Personnel	View Collections
Tuite fMOST	R24_12345_ABCD	True	View Personnel	View Collections

[Create a New Project](#)[Return to PI Dashboard](#)

Project Contributors

Personnel on Tuite fMOST

Name
Ituite96
dsimmel
ghood
ropelews
ecp

[Add Users to Project](#)[Return to Manage Projects](#)[Return to PI Dashboard](#)

Project Submission S

Collections for Project: Tuite fMOST

Name	Description	Submitter	Organization Lab	Lab Name	Project Funder ID	Project Funder	BILUuid	Locked	Submission Status	Validation Status	Modality	View Details
Raw Images 1	Raw mouse brain images	Ituite96	PSC	PSC	R24 Ropelewski	NIH	361bf4a66169cf82	False	NOT_SUBMITTED	NOT_VALIDATED	NIH	View
Raw Images 2	Raw Mouse Images	Ituite96	PSC	PSC	R24 Ropelewski	NIH	d9e8ba60481dcaa1	False	NOT_SUBMITTED	NOT_VALIDATED	NIH	View

Future Features: Self-Service Pre-Validation

Steps of BIL Validation will be available to run yourself via the Brain Image Library Submission portal

Keep an eye out for updates via email and on the BIL submission site for updates on Self Service features.



Carnegie
Mellon
University



How to cite your datasets

Mariah Kenney

How to cite your datasets

BIL is in the process of transitioning to a DOI system for datasets.

DOIs utilize the dataset ID. That is issued once dataset become public.

You can reach out to the BIL help desk at any time to request a DOI for datasets that do not yet have a DOI.

We **do not** recommend citing BIL datasets by:

- including the download link in your publication (i.e. <https://download.brainimagelibrary.org/00/9c/009c1e6fcc03ebac>)
- Including a query in your publication (i.e. "All datasets in BIL for grant X")



Dataset DOI

The screenshot shows a dataset page from the Brain Image Library. At the top, there's a navigation bar with links for "About", "Data Submission", "Data Access", and "Contact". Below the header, the dataset title is displayed: "Cell type distribution in female and male mouse brains / Nos1_GFP_M6_200525". A DOI link "[DOI: <https://doi.org/10.35077/gap>]" is provided. A large brain icon is centered on the page. The main content area includes sections for "Dataset Citation", "Abstract", "Methods", and "TechnicalInfo". Under "Dataset Citation", a citation is given: "Osten, Pavel. (2022). Cell type distribution in female and male mouse brains / Nos1_GFP_M6_200525. [Dataset / Microscopy]. Brain Image Library. <https://doi.org/10.35077/gap>". The "Abstract" section describes the project's goal to establish an Anatomical Collaboratory for Systematic Atlasing of Cell-Type Distribution and Morphology. The "Methods" section notes "No description found." The "TechnicalInfo" section includes a "DATASET METADATA" table and a "SPECIMEN METADATA" table. The "DATASET METADATA" table has four rows: "title" (Cell type distribution in female and male mouse brains / Nos1_GFP_M6_200525), "GeneralModality" (population imaging), "Technique" (other), and "Other" (cell distribution). The "SPECIMEN METADATA" table has three rows: "LocalID" (Nos1_GFP_M6_200525), "Species" (Mouse), and "NCBITaxonomy" (NCBITaxID10090).

DATASET METADATA:	
title	Cell type distribution in female and male mouse brains / Nos1_GFP_M6_200525
GeneralModality	population imaging
Technique	other
Other	cell distribution

SPECIMEN METADATA:	
LocalID	Nos1_GFP_M6_200525
Species	Mouse
NCBITaxonomy	NCBITaxID10090

DOIs are issued for each dataset.

<https://doi.brainimagerlibrary.org/doi/10.35077/gap>

Group DOI

The screenshot shows a Group DOI page for a dataset titled "Detection and Skeletonization of tracer injections using topological methods." The page includes a dataset citation, abstract, methods, technical info, funding, contributors, and related identifiers. A brain icon is present on the right.

Detection and Skeletonization of tracer injections using topological methods.
[DOI: <https://doi.org/10.35077/g.9>]

Dataset Citation:
Mitra, Partha. (2020). Detection and Skeletonization of tracer injections using topological methods.. [Collection / Dataset]. Brain Image Library. <https://doi.org/10.35077/g.9>

Abstract:
The Serial Two-Photon (STP) dataset presented in this paper was collected as a part of Brain Initiative Cell Census Network (BICCN) and downloaded from Brain Image Library (BIL). One STP dataset was involved in the development and demonstration of methods in this paper. The dual-color fluorescent micro-optical sectioning tomography(fMOST) data presented in this paper , both raw images and single neuron reconstruction data ("ground truth"), were collected as a part of BICCN and downloaded from BIL. One fMOST dataset was involved in the development and demonstration of methods in this paper.

Methods:
For STP data, a full description of the protocol is available at protocols.io: dx.doi.org/10.17504/protocols.io.sqcedsw For fMOST data, a full description of the data collection protocol is available at protocols.io: dx.doi.org/10.17504/protocols.io.ssgeebw

Technical Info:
Full description of the data usage and algorithm development is available from <https://arxiv.org/pdf/2004.02755v1.pdf>.

Funding:
NATIONAL INSTITUTE OF BIOMEDICAL IMAGING AND BIOENGINEERING R01-EB022899 METHODS FROM COMPUTATIONAL TOPOLOGY AND GEOMETRY FOR ANALYSING NEURONAL TREE AND GRAPH DATA;
National Institutes of Mental Health U19-MH114821 A COMPREHENSIVE CENTER FOR MOUSE BRAIN CELL ATLAS;
National Institutes of Mental Health U01-MH114824 COLLABORATORY FOR ATLASING CELL TYPE ANATOMY IN THE FEMALE AND MALE MOUSE BRAIN

Contributors:
Mitra, Partha (ProjectLeader) [ORCID: <https://orcid.org/0000-0001-8818-6804>] Cold Spring Harbor Laboratory

Related Identifiers:
Requires [URL : https://download.brainimagelibrary.org/22/06/2206db1eee169a9b/180830_JH_WG_Fezf2LSLflp_CFA_female_processed/];
Requires [URL : <https://download.brainimagelibrary.org/94/77/94775d6a2ddab320/339951-17781/>];
Requires [URL : <https://download.brainimagelibrary.org/b4/d4/b4d4211078a67217/17781/>];

DOIs can also be issued for groups of datasets

Should be a single Group DOI for citing in your manuscript/article

Allows for data that is submitted in different submissions and in different quarters to be easily cited at the same time.

Example:

<https://doi.brainimagelibrary.org/doi/10.35077/g.9>

Request a DOI

- **Group DOIs**
 - Group DOIs should be requested at bil-support@psc.edu
 - We will ask you to fill out and return a form that includes the datasets you would like to group in the DOI and information about the publication citing the data.
- **Dataset DOIs**
 - In the interim period before we are at the point of issuing DOIs automatically upon publication for each dataset, dataset DOIs should also be requested by reaching out to the help desk if needed immediately.





Carnegie
Mellon
University



Data Access

Where to find BIL data?

Overview: Where to find BIL data?

1. **BIL file system**
2. **BIL search web portal**
 - Web portal: <https://api.brainimagelibrary.org/web/>
3. **BIL metadata search API**
 - <https://www.brainimagelibrary.org/metadataapi.html>
4. **DOI (digital object identifier)**
 - Dataset DOI
 - Group DOI

1. BIL File System

Data is organized by submission in BIL

- The path to the public data utilizes the submission id number
- There are unique directories for each dataset
- The path includes sub-directories using the first 2 and then the third and fourth characters in the submission id
- Ex: collection id: abcdef0123456789

/bil/lz/user/abcdef0123456789/example_dataset_01



/bil/data/ab/cd/abcdef0123456789/example_dataset_01

1. BIL File System

Data is organized by submission in BIL

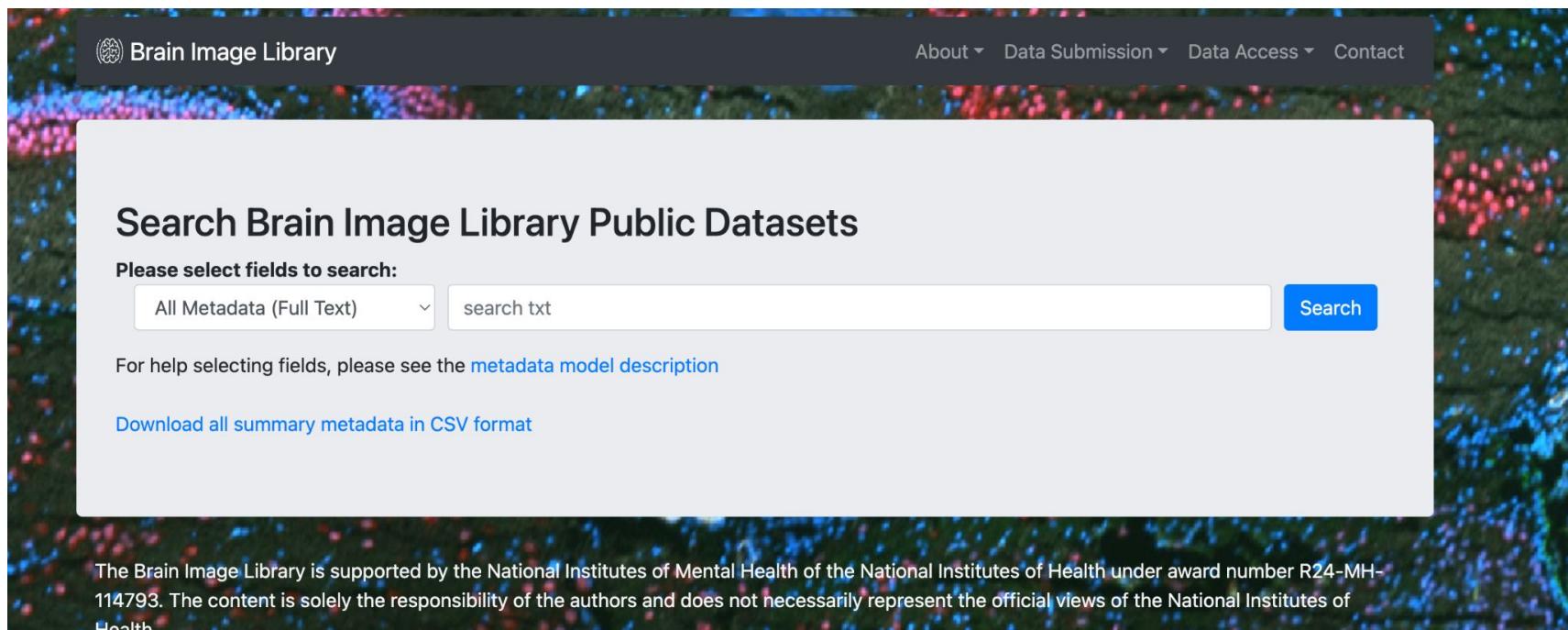
- The path to the public data utilizes the submission id number
- The path includes sub-directories using the first 2 and then the third and fourth characters in the submission id
- Ex: collection id: 1234abcd

/bil/data/ab/cd/abcdef0123456789/example_dataset_01

This path can be reformatted to a URL to access the data:

https://download.brainimagelibrary.org/ab/cd/abcdef0123456789/example_dataset_01/

2. Search portal - <https://api.brainimagerlibrary.org/web/>



The image shows the Brain Image Library search portal. At the top, there is a dark header bar with the "Brain Image Library" logo and navigation links for "About", "Data Submission", "Data Access", and "Contact". Below the header is a large, semi-transparent background image of a brain scan with blue and red highlights. In the center, there is a white search form. The title "Search Brain Image Library Public Datasets" is displayed in bold black font. Below it, a sub-header "Please select fields to search:" is followed by a dropdown menu set to "All Metadata (Full Text)" and a search input field containing "search txt". To the right of the search input is a blue "Search" button. Below the search form, there is a link to "metadata model description" and another link to "Download all summary metadata in CSV format". At the bottom of the page, a footer note states: "The Brain Image Library is supported by the National Institutes of Mental Health of the National Institutes of Health under award number R24-MH-114793. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health."

3. Search API

The metadata API is extremely flexible and capable of full-text searching or individual metadata fields.

The metadata API has two primary user-facing endpoints:

- query - the query endpoint will query the system and return a json document with matching entry. Queries use GET formatted URLs, which can be pasted in a web browser, utilized at the command-line through tools such as curl, or called programmatically. Query URLs take the form of:

`/query/metadata/division?metadataelement=querystring`

- retrieve - the retrieve endpoint will return the metadata as a json document for the given ids. Retrieve queries can be either GET formatted urls (for retrieving metadata for a single entry) or a POST request with a json payload. The json POST payload may be the results of an earlier query. The basic forms of the retrieve GET url are:

`/retrieve?bilid=the_desired_id_to_retrieve_metadata_for`

For example:

`https://api.brainimagelibrary.org/retrieve?bilid=ace-cot-pad`

3. Search API

Query: <https://api.brainimagelibrary.org/query/contributors?affiliation=Broad>

Result:

```
{  
    "success": "true",  
    "endpoint": "Contributors/affiliation",  
    "message": "GET success",  
    "affiliation": "Broad",  
    "bildids": [  
        "ace-big-pup",  
        "ace-big-pry",  
        "ace-big-pot",  
        "ace-big-pox",  
        "ace-big-pun"  
    ]  
}
```

3. Search API

Retrieve: <https://api.brainimagelibrary.org/retrieve?doi=https://doi.org/10.35077/rut>

Result:

```
{  
    "success": "true",  
    "endpoint": "Retrieve/doi",  
    "message": "GET success",  
    "doi": "https://doi.org/10.35077/rut",  
    "retjson": [  
        {  
            "Metadata": "2.0",  
            "Submission": {  
                "sheet": "57",  
                "collection": "1394",  
                "submission_uuid": "96ba8210cceceeb7",  
                "method": "ingest_1",  
                "doi": "https://doi.org/10.35077/rut"  
            }  
        }  
    ]  
}
```

(Continues...)

4. DOI

The screenshot shows a dataset landing page from the Brain Image Library. At the top, there's a navigation bar with links for 'About', 'Data Submission', 'Data Access', and 'Contact'. Below the header, the dataset title is displayed: 'Cell type distribution in female and male mouse brains / Nos1_GFP_M6_200525'. A DOI link is provided: [DOI: <https://doi.org/10.35077/gab>]. A large brain icon is centered on the page. The main content area includes sections for 'Dataset Citation', 'Abstract', 'Methods', and 'TechnicalInfo'. Under 'Dataset Citation', it lists the author, year, title, and source. The 'Abstract' section describes the project's goal of establishing an anatomical collaboratory for systematic atlasing of cell-type distribution and morphology. The 'Methods' section notes 'No description found.' The 'TechnicalInfo' section contains two tables: 'DATASET METADATA' and 'SPECIMEN METADATA'. The 'DATASET METADATA' table includes rows for title, GeneralModality, Technique, and Other. The 'SPECIMEN METADATA' table includes rows for LocalID, Species, and NCBITaxonomy.

DATASET METADATA:	
title	Cell type distribution in female and male mouse brains / Nos1_GFP_M6_200525
GeneralModality	population imaging
Technique	other
Other	cell distribution

SPECIMEN METADATA:	
LocalID	Nos1_GFP_M6_200525
Species	Mouse
NCBITaxonomy	NCBITaxid10090

Datasets with DOIs issued will have a landing page that displays more in-depth metadata for the associated dataset:

<https://doi.org/10.35077/gab>

DOIs can also be issued for groups of related datasets:

<https://doi.org/10.35077/g.9>

4. DOI

The screenshot shows a dataset landing page from the Brain Image Library. The title is "Detection and Skeletonization of tracer injections using topological methods." Below it is the DOI: <https://doi.org/10.35077/g.9>. The page includes a dataset citation, abstract, methods, technical info, funding, contributors, and related identifiers. A brain icon is on the right.

Detection and Skeletonization of tracer injections using topological methods.

[DOI: <https://doi.org/10.35077/g.9>]

Dataset Citation:
Mitra, Partha. (2020). Detection and Skeletonization of tracer injections using topological methods.. [Collection / Dataset]. Brain Image Library. <https://doi.org/10.35077/g.9>

Abstract:
The Serial Two-Photon (STP) dataset presented in this paper was collected as a part of Brain Initiative Cell Census Network (BICCN) and downloaded from Brain Image Library (BIL). One STP dataset was involved in the development and demonstration of methods in this paper. The dual-color fluorescent micro-optical sectioning tomography(fMOST) data presented in this paper , both raw images and single neuron reconstruction data ("ground truth"), were collected as a part of BICCN and downloaded from BIL. One fMOST dataset was involved in the development and demonstration of methods in this paper.

Methods:
For STP data, a full description of the protocol is available at protocols.io: dx.doi.org/10.17504/protocols.io.sqcedsw For fMOST data, a full description of the data collection protocol is available at protocols.io: dx.doi.org/10.17504/protocols.io.ssgeebw

Technical Info:
Full description of the data usage and algorithm development is available from <https://arxiv.org/pdf/2004.02755v1.pdf>.

Funding:
NATIONAL INSTITUTE OF BIOMEDICAL IMAGING AND BIOENGINEERING R01-EB022899 METHODS FROM COMPUTATIONAL TOPOLOGY AND GEOMETRY FOR ANALYSING NEURONAL TREE AND GRAPH DATA;
National Institutes of Mental Health U19-MH114821 A COMPREHENSIVE CENTER FOR MOUSE BRAIN CELL ATLAS;
National Institutes of Mental Health U01-MH114824 COLLABORATORY FOR ATLASING CELL TYPE ANATOMY IN THE FEMALE AND MALE MOUSE BRAIN

Contributors:
Mitra, Partha (ProjectLeader) [ORCID: <https://orcid.org/0000-0001-8818-6804>] Cold Spring Harbor Laboratory

Related Identifiers:
Requires [URL : https://download.brainimagelibrary.org/22/06/2206db1eee169a9b/180830_JH_WG_Fezf2LSLflp_CFA_female_processed/];
Requires [URL : <https://download.brainimagelibrary.org/94/77/94775d6a2ddab320/339951-17781/>];
Requires [URL : <https://download.brainimagelibrary.org/b4/d4/b4d4211078a67217/17781/>];

Datasets with DOIs issued will have a landing page that displays more in-depth metadata for the associated dataset:

<https://doi.org/10.35077/gab>

DOIs can also be issued for groups of related datasets:

<https://doi.org/10.35077/g.9>

Metadata version updates

Version 1

07504513d424f013	Image and morphological reconstruction of a cortical interneuron (3b) from marmoset, Yanma, after systemic viral IV injection with AAV2/9-hDlx-GFP-GFP, imaged in December 2020 - positive for DLX5/6-GFP, positive for Parvalbumin and negative for Thrombospondin Type 1 Domain Containing 7A.
Sample:	
Cortex 20201201_	
CJYanma_Ex10_	
PVALB+THSD7a_-	
Hippo-Region	
004	
 BICCN	
Investigator	Experiment
Guoping Feng	<i>Modality:</i> spatial transcriptomics
Feng Lab	<i>Method:</i> smFISH
Massachusetts Institute of Technology	<i>Technique:</i> smFISH
Funding	<i>Structure:</i> Cortex
1-U01-MH114819-01	<i>Organism:</i> marmoset
	<i>TransLine:</i> CJYanma
	<i>Cells:</i> 1
BIL: /bil/data/07/50/07504513d424f013/	
HTTPS: https://download.brainimagerlibrary.org/07/50/07504513d424f013/	

- The latest BIL metadata version was implemented at the beginning of this year
- The version 2 includes more fields and descriptive metadata
- All BIL data is in the process of being updated to the newest version where possible
- BRAIN Standards
(<https://doi.org/10.1038/s41597-022-01562-5>)

Version 2

TechnicalInfo:	
DATASET METADATA:	
title	Cell type distribution in female and male mouse brains / Nos1_GFP_M6_200525
GeneralModality	population imaging
Technique	other
Other	cell distribution
SPECIMEN METADATA:	
LocalID	Nos1_GFP_M6_200525
Species	Mouse
NCBI/Taxonomy	NCBI:txid10090
Age	unknown
Ageunit	Days
Sex	Male
Genotype	R26-CAG-LSL-H2B-GFP/R26-CAG-LSL-H2B-GFP; nNos1/nNos1
OrganName	Brain
SampleLocalID	Nos1_GFP_M6_200525
INSTRUMENT METADATA:	
MicroscopeType	Two Photon
MicroscopeManufacturerAndModel	TissueVision
IMAGE METADATA:	
xAxis	superior-to-inferior
yAxis	right-to-left
zAxis	anterior-to-posterior
Number	1,2,red,green
displayColor	(1.0, 0.0, 0.0),(0.0, 1.0, 0.0)
stepSizeX	4 micron/pixel
stepSizeY	4 micron/pixel
stepSizeZ	53 micron/pixel

Metadata version updates

- Metadata updates are made as needed for new data modalities
- SWC metadata was added for neuron tracing datasets in 2023 (Method 5)
 - New metadata fields were added for neuron tracing datasets that include:
 - sourceDataSample
 - sourceDataSubmission
 - coordinates
 - coordinatesRegistration
 - brainRegion
 - brainRegionAtlas
 - brainRegionAtlasName
 - brainRegionAxonalProjectio
 - brainRegionDendriticProjection
 - neuronType
 - segmentTags
 - proofreadingLevel
- UPCOMING: New metadata fields will be added for spatial transcriptomics datasets



Carnegie
Mellon
University



Tools available at BIL

How to interact with data pre or post submission



Tools Available at BIL

Computational Resources (at no cost to users)

- [BIL Analysis Ecosystem](#)
 - Flexible resource that is made up of several large memory machines equipped with modern GPUs
- [Bridges-2](#)
 - NSF funded supercomputer located at the Pittsburgh Supercomputing Center.
- [Neocortex](#)
 - Neocortex is a resource that targets AI-powered scientific discovery and provides hardware for the development of efficient algorithms for artificial intelligence and graph analytics.

Tools available to access some of these resources:

- [Open OnDemand](#)
 - Open OnDemand is an open-source portal that enables web-based access to HPC services
 - Jupyter Lab
 - RStudio
 - File System Interface
 - Terminal

BIL Analysis Ecosystem

Access to the BIL Analysis Ecosystem is granted to all BIL users upon account creation.

Large Memory Compute Nodes (L nodes)

Number of nodes: 8 ([I001 - I008](#))

- Processors: HPE ProLiant DL580 Gen9
- Cores per node: 80 - 20 per CPU
- RAM: 3TB
- CPU per node: 4
- CPU model: Xeon E7 8870

Software Available

Fiji, Bio-Formats, R, Python, MatLab, and more.



Bridges-2

<https://www.psc.edu/resources/bridges-2/>

Bridges-2 is available at no cost for research and education, and at cost-recovery rates for other purposes.

Regular memory nodes: 256GB - 512GB RAM (488 RM nodes have 256GB of RAM, and 16 have 512GB of RAM)

Extreme memory nodes: 4TB RAM (4 nodes)

GPU nodes: 512GB of RAM (Eight NVIDIA Tesla V100-32GB SXM2 GPUs) (24 GPU nodes)

The BIL file system is mounted on Bridges 2 so all data is directly accessible on these resources.

Bridges-2 is allocated through ACCESS. More information on ACCESS allocations and information on how to get started is available here: <https://allocations.access-ci.org/>



Visualization

The Brain Image Library provides following ways to visualize the data remotely, without signing in:

- Neuroglancer: Visualize BIL data in a web browser. Explore curated datasets at our brain explorer
- Napari-bil-data-viewer: A plugin for the popular napari image viewer that allows single-click visualization of thousands of datasets at BIL.

Interact with these tools and learn more at:

<https://www.brainimagedatabase.org/visual.html>



Questions?

BIL Office Hours

Join the BIL team virtually each month for hands-on assistance.

Check for occurrences on the BIL website here:

<https://www.brainimagelibrary.org/contact.html>



For questions at any time contact us at bil-support@psc.edu

If the data models do not accurately reflect your data, reach out and we can work with you to find the best fit for how to represent your data.



Carnegie
Mellon
University



Thank you for your time.

Reach out to bil-support@psc.edu with any questions