



Carnegie  
Mellon  
University



# Using the BIL Portal

*Luke Tuite*

# Brain Image Library Data Submission Process

Workflow for  
depositing data  
at BIL

Providing  
metadata for  
your submissions

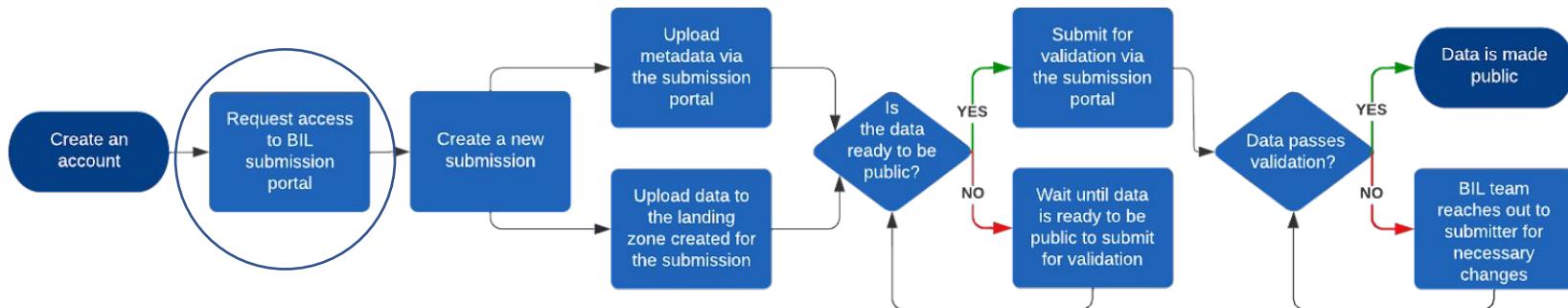
BIL PI Dashboard

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

- Structuring your metadata
- Proper inclusions

- 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](mailto: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](http://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.

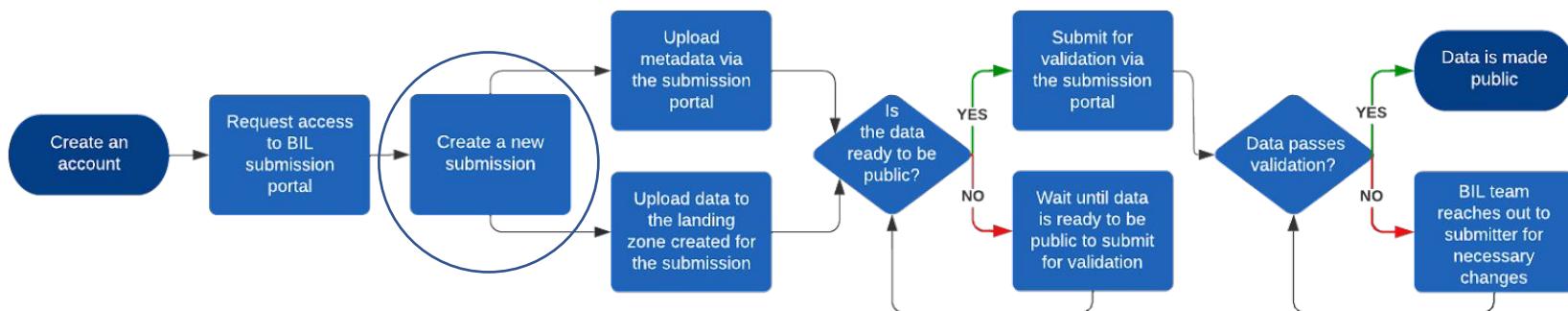
Username:

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

# 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

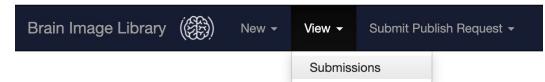
1. Navigate to submit.brainimagelibrary.org in your browser
2. Click Login and use the password you've set up when you created a PSC user account.
3. On the top left-hand corner of your screen – navigate to “New” and on the drop down select “New Submission”
4. Complete the New Submission form to create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

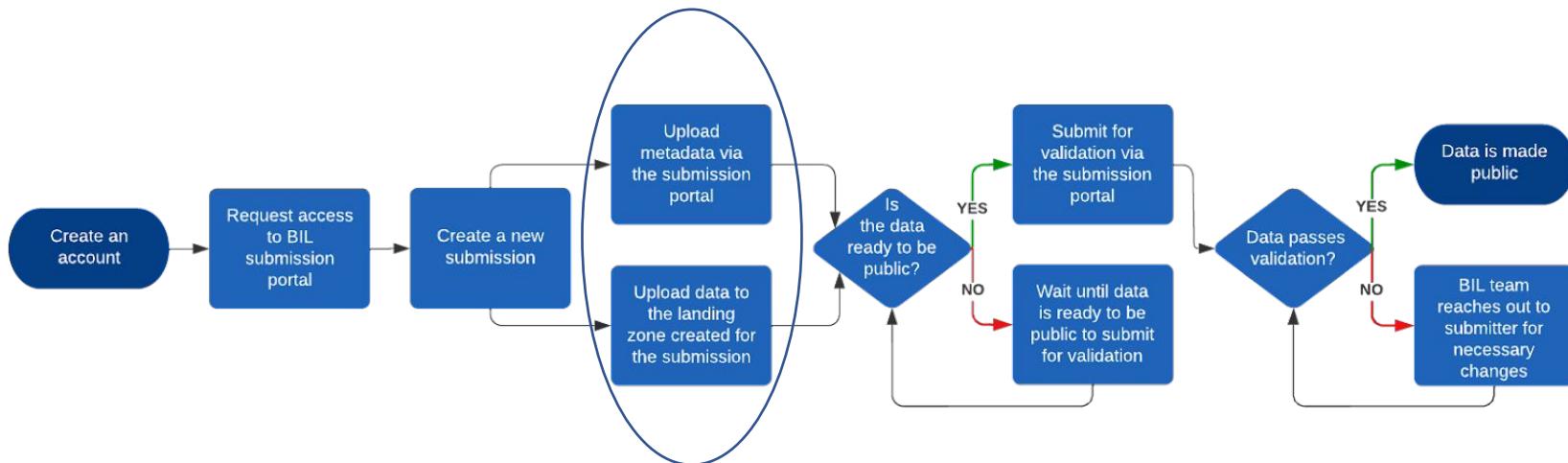
This data path is where you will be directing your upload to.

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)

# 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 'Certificate Information' drop-down link to find the 'Certificate Subject'. Select and copy the entire certificate subject string to include in your e-mail to [bil-support@psc.edu](mailto:bil-support@psc.edu).
7. 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](mailto:bil-support@psc.edu). If your CILogon ePPN string is blank, please let us know that, and also which CILogon Identity Provider you selected.
8. Send an email to [bil-support@psc.edu](mailto: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

Collection Search

Collection Brain Image Library

**Brain Image Library Download**  
GCSv4 Host  
Owner: bil@globusid.org  
Description: Download endpoint for authorized users of the Brain Image Library.

**Brain Image Library Upload with CILogon Authentication**  
GCSv4 Host  
Owner: bil@globusid.org  
Description: Endpoint for authorized Brain Image Library users to upload data to directories in which they have write access

g Cloud

FILE MANAGER

BOOKMARKS

ACTIVITY

COLLECTIONS

GROUPS

CONSOLE

FLOWS

ACCOUNT

LOGOUT

HELP & SITEMAP

# Using Globus for File Transfer to BIL

The screenshot shows the Globus File Manager interface. On the left is a sidebar with various icons: Cloud, File Manager (selected), Bookmarks, Activity, Collections, Groups, Console, Flows, Account, Logout, Help & Sitemap. The main area has tabs for 'File Manager' and 'Brain Image Library Upload with CILogon Authentication'. The 'Collection' is set to 'Lukes\_Macbook' and the 'Path' is '/~/'. A 'Transfer & Timer Options' dropdown is open. Below is a table of files:

NAME	LAST MODIFIED	SIZE
Documents	2/15/2023, 11...	—
Downloads	4/3/2023, 09...	—
Movies	2/15/2023, 11...	—
Music	2/15/2023, 11...	—
<input checked="" type="checkbox"/> myfile.txt	4/2/2023, 09...	16 B
opt	2/16/2023, 11...	—
Pictures	2/15/2023, 12...	—
Public	2/15/2023, 11...	—

A context menu is open over 'myfile.txt', listing options: Share, Transfer or Sync to..., New Folder, Rename, Delete Selected, Download, Open, Upload, Get Link, Show Hidden Items, and Manage Activation. The URL in the address bar is circled: /bil/lz/luite96/dc584feacd6a0382/.

# Using SCP/SFTP for File Transfer to BIL

Using scp to [upload.brainimagelibrary.org](http://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
```

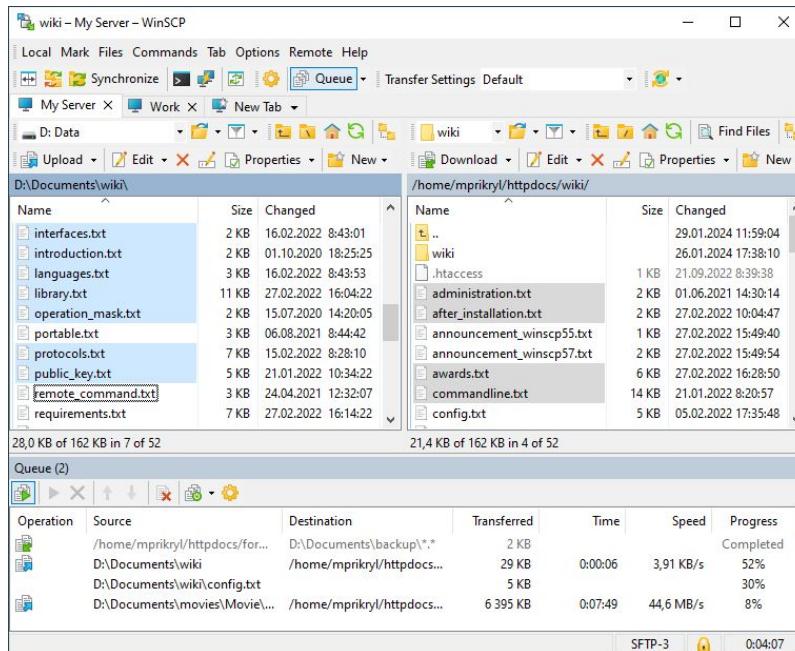
```
testuser@upload.brainimagelibrary.org's password:
```

```
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.

# Hands on Exercise 2

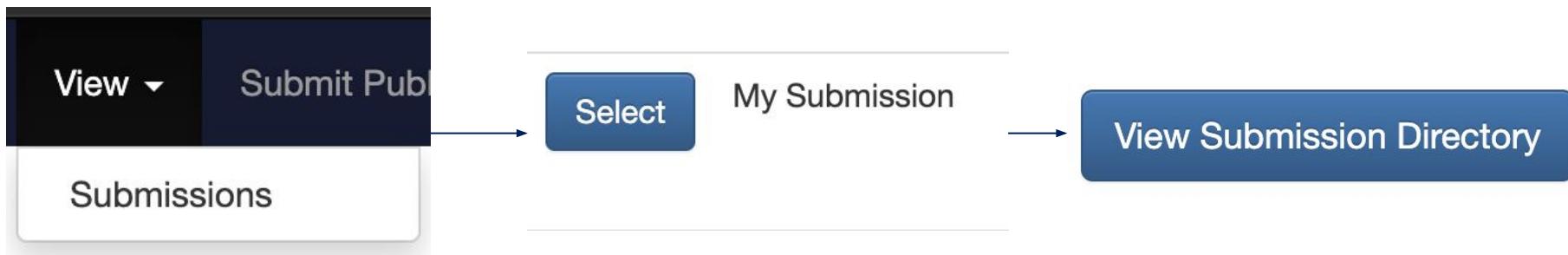
With our file created, let's upload it to your new BIL Submission

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.



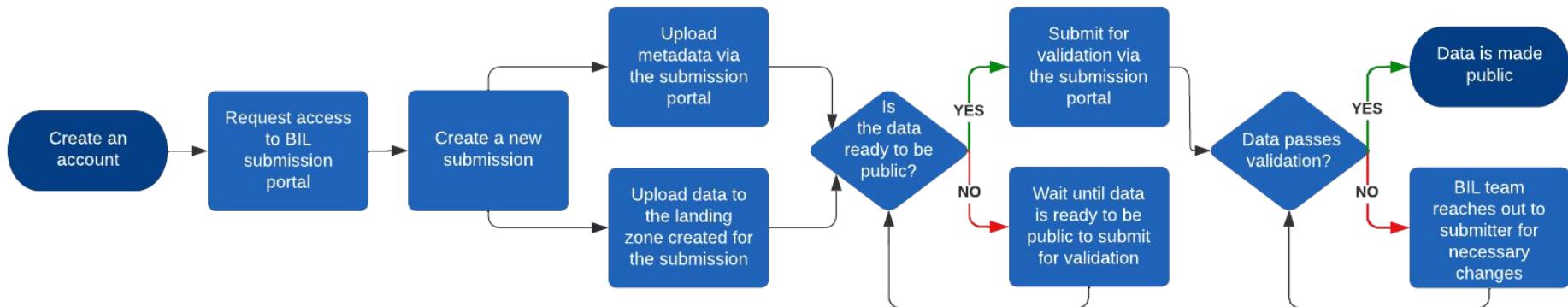
Carnegie  
Mellon  
University



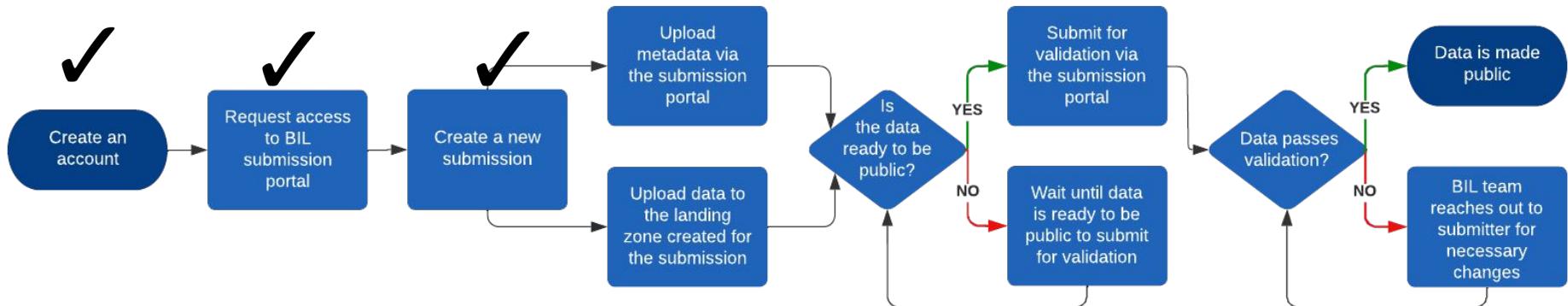
# File Organization

*Mariah Kenney*

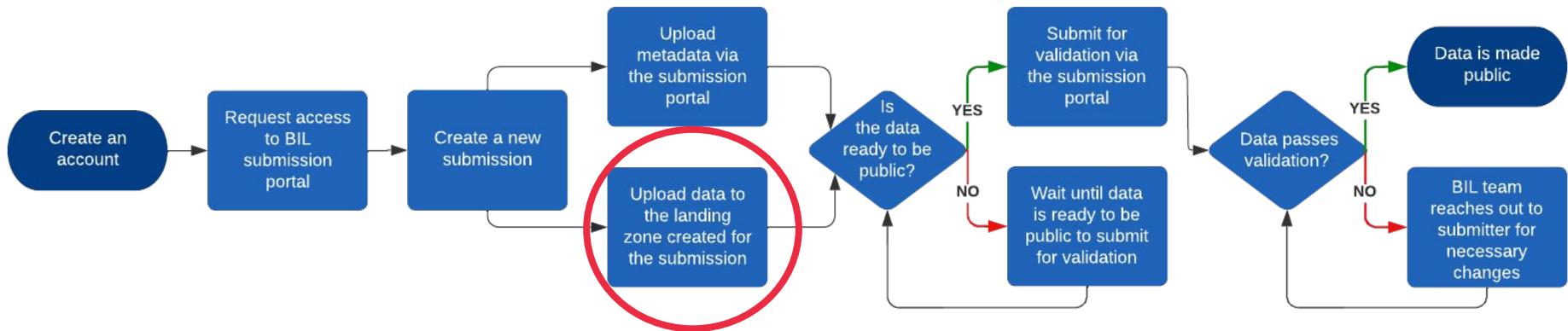
# Submission Overview



# 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.

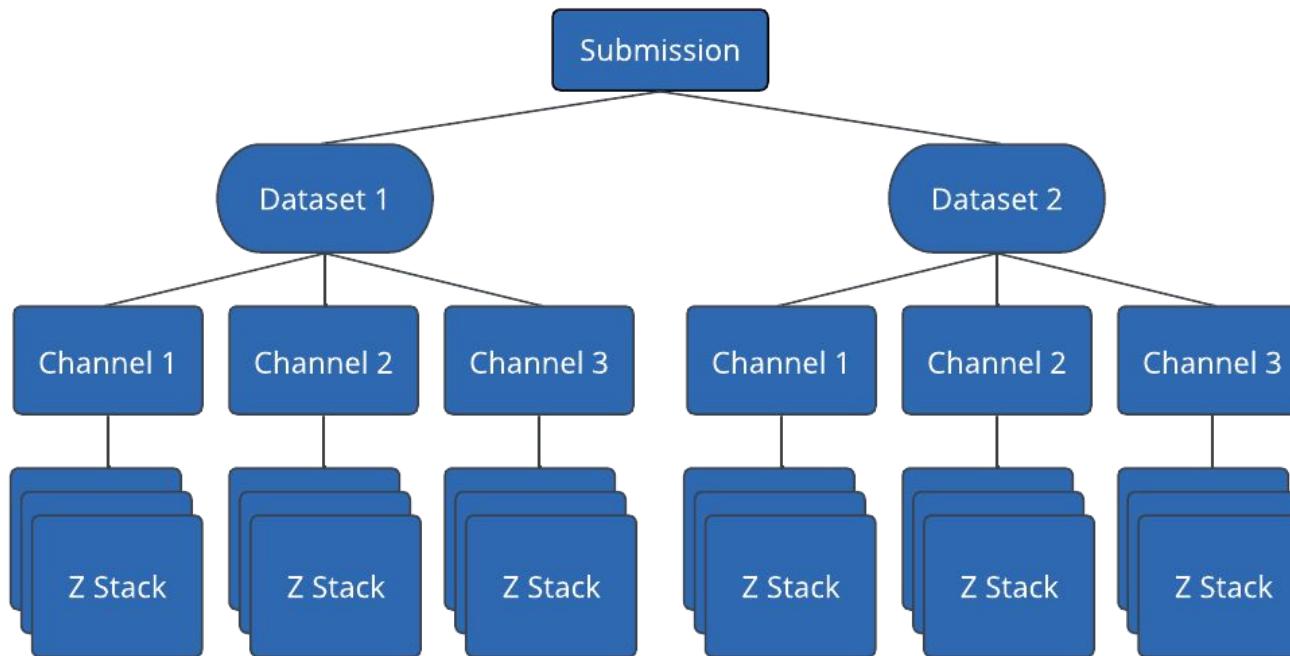
## Submission

- 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*

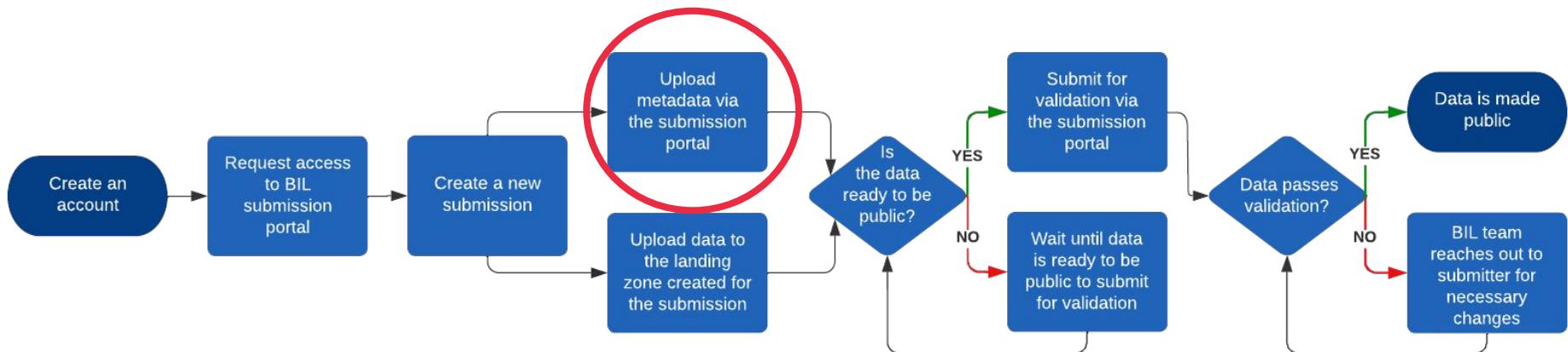
# Brain Image Library Metadata



*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: Collaborative Standards for BRAIN 3D Microscopy Hamilton (PI) arXiv:2105.09158v1 <https://arxiv.org/abs/2105.09158>

# 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.

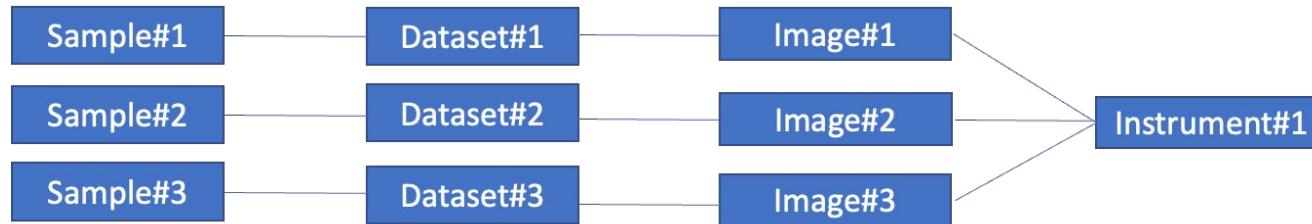
Choose a collection, then upload your metadata.

Associated collection:

[Upload Metadata](#)

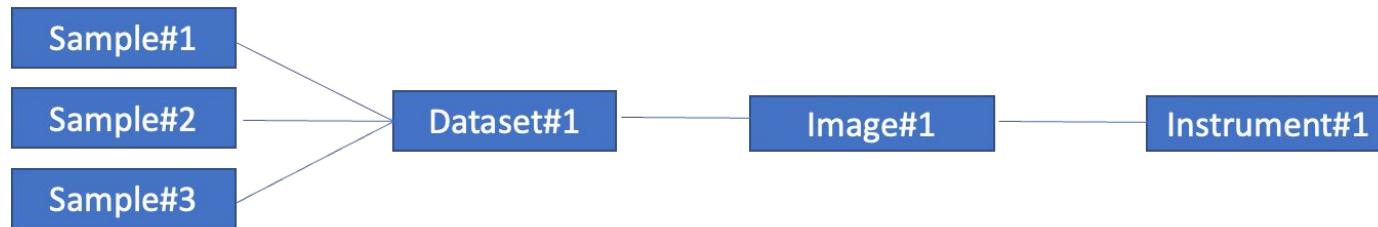
# Metadata Models

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.



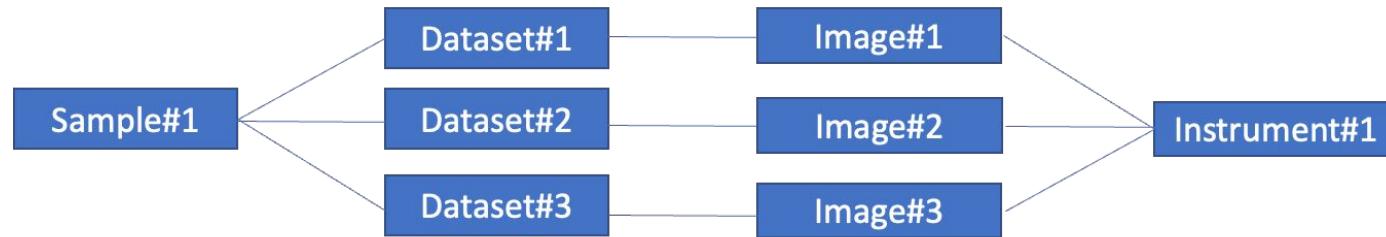
# Metadata Models

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.



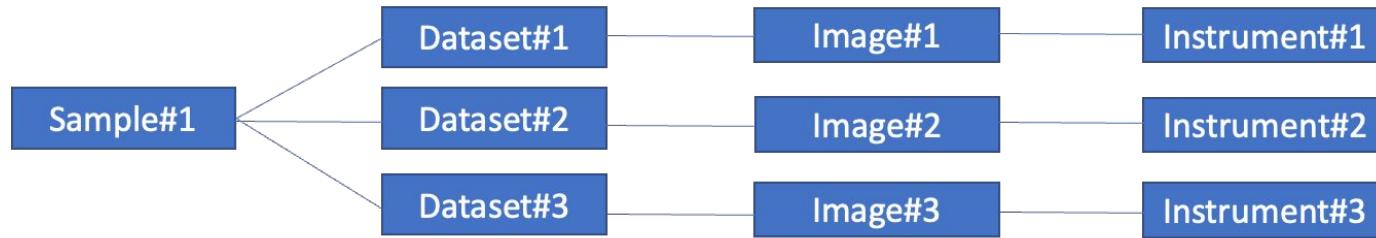
# Metadata Models

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 Models

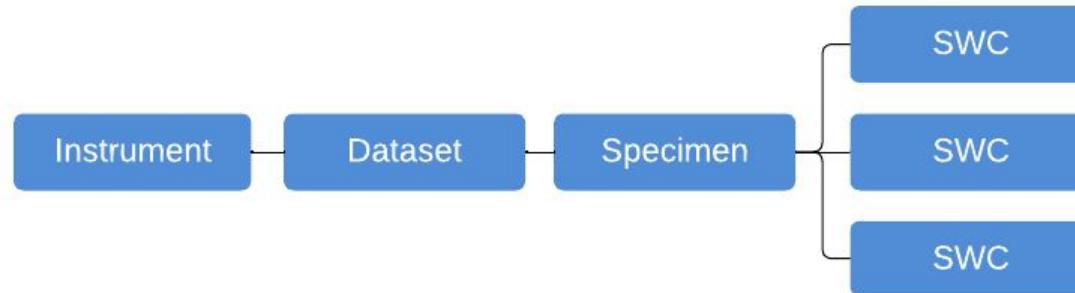
4. All datasets were generated from one specimen – but use different experimental (machinery) parameters. Multiple datasets can be accommodated as a single submission. There should be exactly one entry in the specimen tab. The lines listed in the Instrument tab in the metadata spreadsheet must match 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 Models

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.**

# 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 3

**4 image sets for 4 different ROIs from the same brain**

Directory structure:

/bil/lz/userid/submissionid/ROI1

/bil/lz/userid/submissionid/ROI2

/bil/lz/userid/submissionid/ROI3

/bil/lz/userid/submissionid/ROI4

# 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.

## Example 4

Used in cases where multiple samples are included with unique metadata and the samples cannot be separated into different files/directories - more common with higher level data.

Directory structure:

/bil/lz/userid/submissionid/datasetid



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, a section titled 'Step 2 of 3: What does your data look like?' contains five numbered options describing different dataset generation scenarios. At the bottom of the page, another section titled 'Step 3 of 3: Upload metadata for associated submission' includes instructions to download a template, choose a collection, and upload the metadata. There are also 'Cancel' and 'Upload Metadata' buttons.

Brain Image Library  New ▾ View ▾ Submit Publish Request ▾

New Submission  
Add Metadata

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:

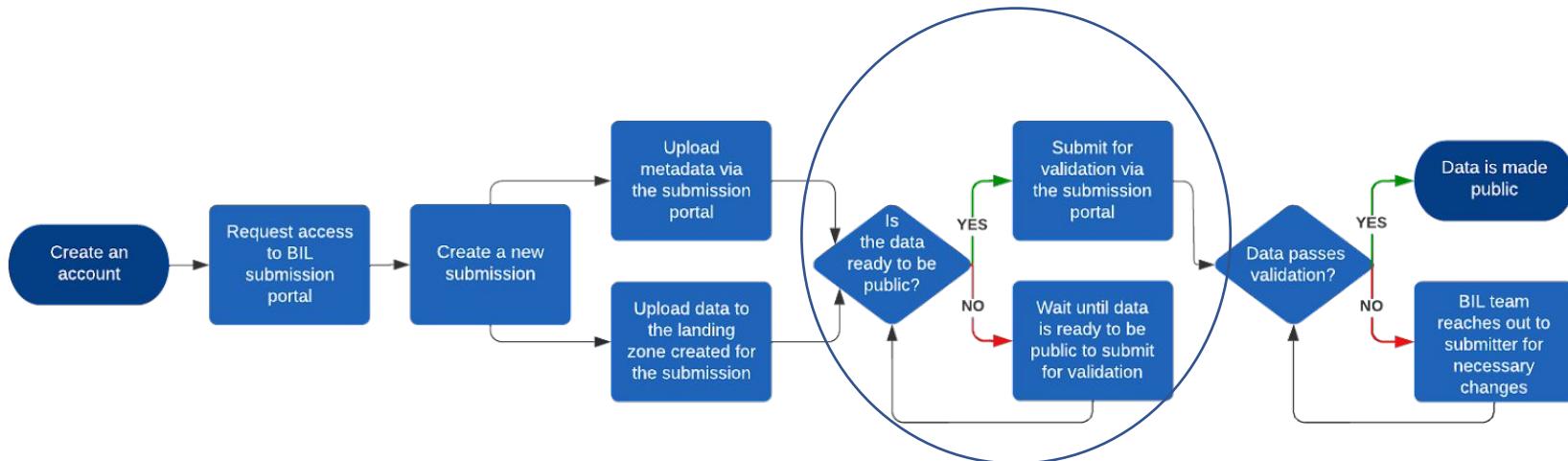
Upload Metadata

Cancel

# Hands on exercise 3

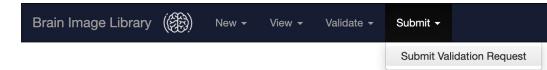
1. On the top left-hand corner of your screen – navigate to “New” and on the drop down select “Descriptive Metadata”
2. On the drop-down menu – select your data submission you’ve just created
3. Using the file you downloaded during the metadata presentation - open this file on the submission interface and you will be redirected to a detailed view of your data submission

# 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.

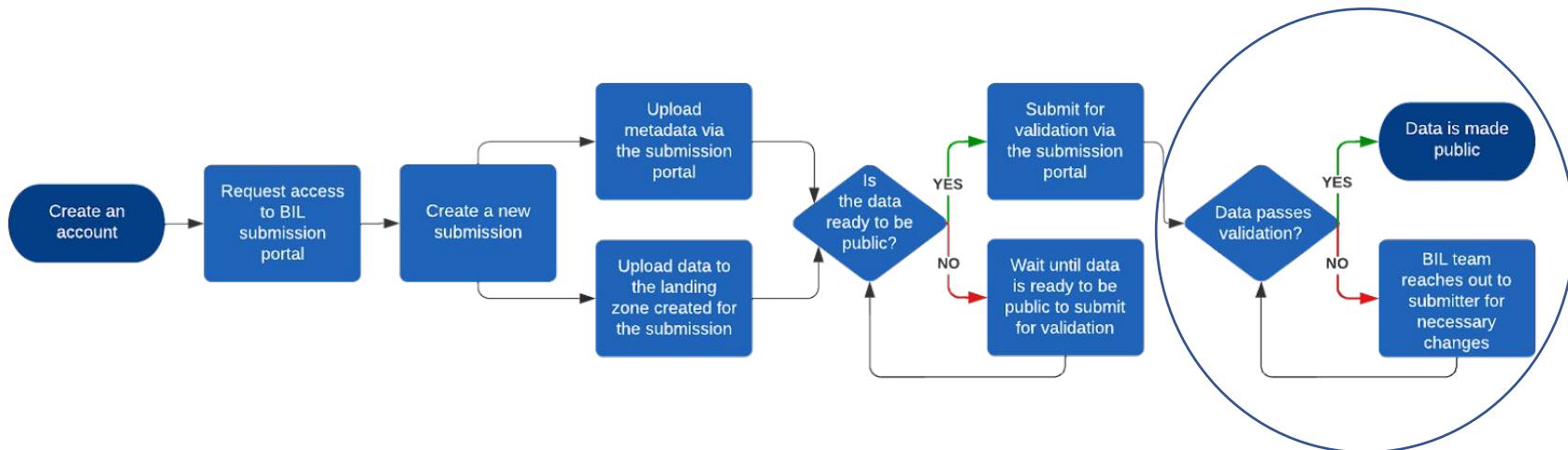
Our data validation does not reject based on quality of data, only on usability.

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:

- Project creation is mandatory for BICAN projects
- Request PI Dashboard access through the ticket system at [bil-support@psc.edu](mailto: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	<a href="#">View Personnel</a>	<a href="#">View Collections</a>
Tuite fMOST	R24_12345_ABCD	True	<a href="#">View Personnel</a>	<a href="#">View Collections</a>

[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	BIL Uuid	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	<a href="#">View</a>
Raw Images 2	Raw Mouse Images	Ituite96	PSC	PSC	R24 Ropelewski	NIH	d9e8ba60481dcaa1	False	NOT_SUBMITTED	NOT_VALIDATED	NIH	<a href="#">View</a>

# Future Features: Self-Service Pre-Validation

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

PIs will be able to define project sub-projects through the PI Dashboard

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.

For datasets that have been issued a DOI, the DOI landing page will contain the citation and DOI for each dataset.

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/](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](mailto: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 portal**
  - <https://submit.brainimagelibrary.org/search>
3. **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/data/ab/cd/abcdef0123456789/example\_dataset\_01**

# 1. BIL File System

Data is organized by submission in BIL

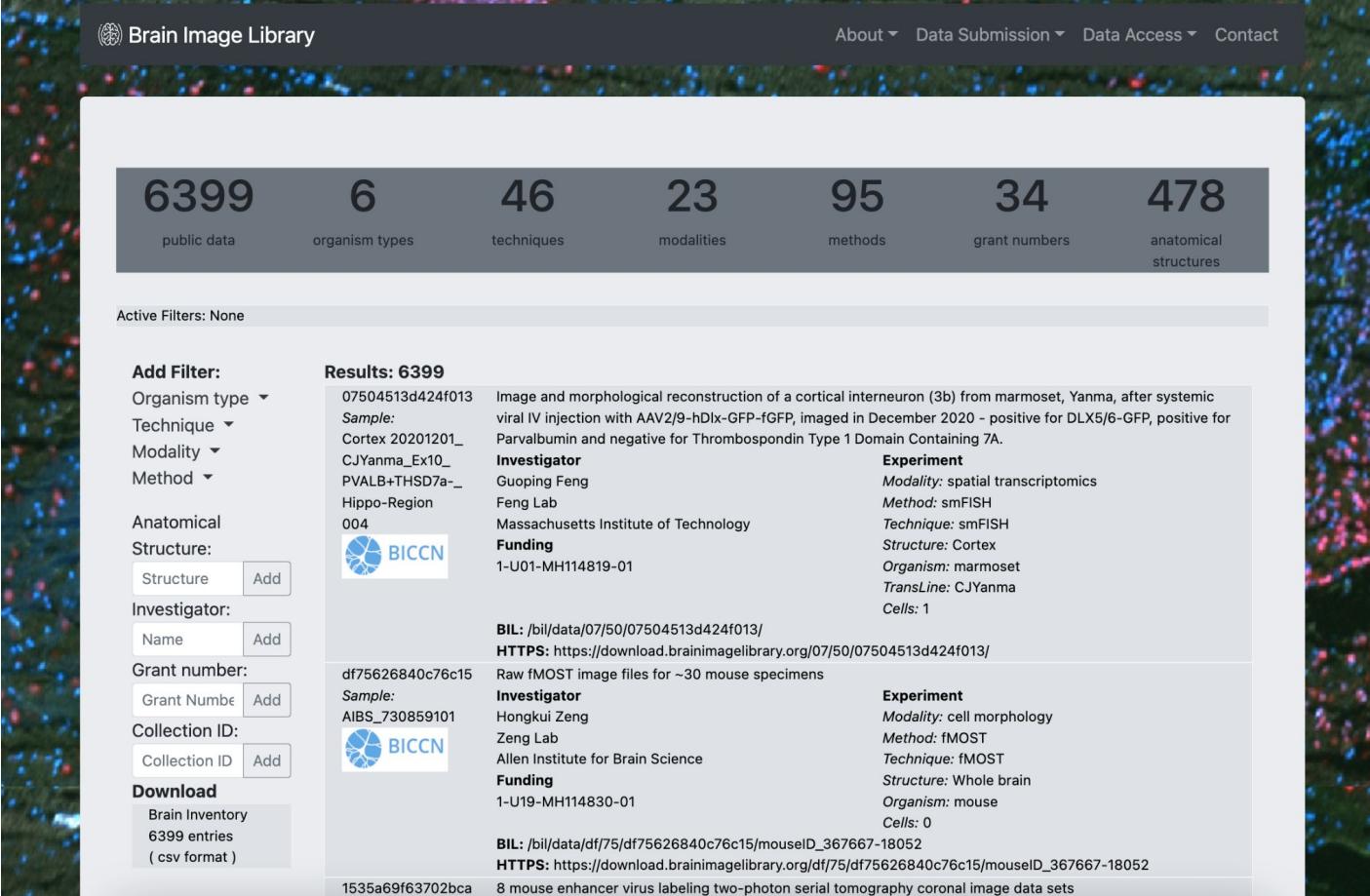
- The path to the public data utilizes the submission id number
- The path includes sub-directories using the first 2 and then the third and fourth characters in the submission id
- Ex: collection id: 1234abcd

**/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/](https://download.brainimagelibrary.org/ab/cd/abcdef0123456789/example_dataset_01/)**

## 2. Search portal - submit.brainimagelibrary.org/search



Brain Image Library

About ▾ Data Submission ▾ Data Access ▾ Contact

6399 public data    6 organism types    46 techniques    23 modalities    95 methods    34 grant numbers    478 anatomical structures

Active Filters: None

**Add Filter:**

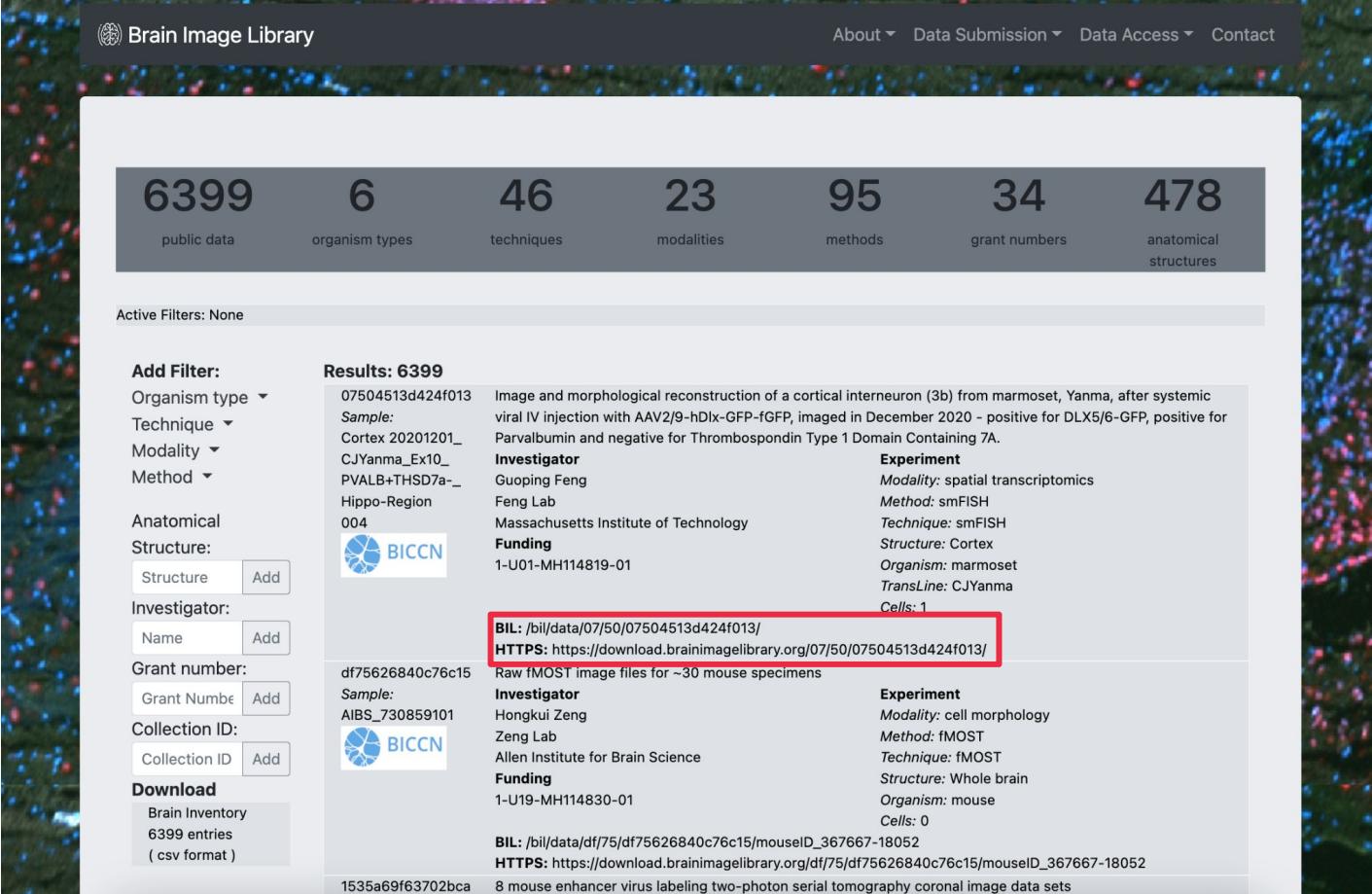
- Organism type ▾
- Technique ▾
- Modality ▾
- Method ▾
- Anatomical Structure:  
Structure
- Investigator:  
Name
- Grant number:  
Grant Number
- Collection ID:  
Collection ID
- Download  
Brain Inventory  
6399 entries  
( csv format )

**Results: 6399**

07504513d424f013	Image and morphological reconstruction of a cortical interneuron (3b) from marmoset, Yanma, after systemic viral IV injection with AAV2/9-hDlx-GFP-fGFP, imaged in December 2020 - positive for DLX5/6-GFP, positive for Parvalbumin and negative for Thrombospondin Type 1 Domain Containing 7A.	<b>Experiment</b> <i>Modality:</i> spatial transcriptomics <i>Method:</i> smFISH <i>Technique:</i> smFISH <i>Structure:</i> Cortex <i>Organism:</i> marmoset <i>TransLine:</i> CJYanma <i>Cells:</i> 1
Sample: Cortex 20201201_ CJYanma_Ex10_ PVALB+THSD7a- Hippo-Region 004	Guoping Feng Feng Lab Massachusetts Institute of Technology <b>Funding</b> 1-U01-MH114819-01	BIL: /bil/data/07/50/07504513d424f013/ <a href="https://download.brainimagelibrary.org/07/50/07504513d424f013/">HTTPS: https://download.brainimagelibrary.org/07/50/07504513d424f013/</a>
df75626840c76c15	Raw fMOST image files for ~30 mouse specimens	<b>Experiment</b> <i>Modality:</i> cell morphology <i>Method:</i> fMOST <i>Technique:</i> fMOST <i>Structure:</i> Whole brain <i>Organism:</i> mouse <i>Cells:</i> 0
Sample: AIBS_730859101	Hongkui Zeng Zeng Lab Allen Institute for Brain Science <b>Funding</b> 1-U19-MH114830-01	BIL: /bil/data/df/75/df75626840c76c15/mouselD_367667-18052 <a href="https://download.brainimagelibrary.org/df/75/df75626840c76c15/mouselD_367667-18052">HTTPS: https://download.brainimagelibrary.org/df/75/df75626840c76c15/mouselD_367667-18052</a>
1535a69f63702bca	8 mouse enhancer virus labeling two-photon serial tomography coronal image data sets	

© 2023 Brain Image Library, All Rights Reserved

## 2. Search portal - submit.brainimagelibrary.org/search



Brain Image Library

About ▾ Data Submission ▾ Data Access ▾ Contact

6399 public data    6 organism types    46 techniques    23 modalities    95 methods    34 grant numbers    478 anatomical structures

Active Filters: None

**Add Filter:**

- Organism type ▾
- Technique ▾
- Modality ▾
- Method ▾
- Anatomical Structure:  
Structure
- Investigator:  
Name
- Grant number:  
Grant Number
- Collection ID:  
Collection ID
- Download  
Brain Inventory  
6399 entries  
( csv format )

**Results: 6399**

07504513d424f013	Image and morphological reconstruction of a cortical interneuron (3b) from marmoset, Yanma, after systemic viral IV injection with AAV2/9-hDlx-GFP-fGFP, imaged in December 2020 - positive for DLX5/6-GFP, positive for Parvalbumin and negative for Thrombospondin Type 1 Domain Containing 7A.	<b>Experiment</b> <i>Modality:</i> spatial transcriptomics <i>Method:</i> smFISH <i>Technique:</i> smFISH <i>Structure:</i> Cortex <i>Organism:</i> marmoset <i>TransLine:</i> CJYanma <i>Cells:</i> 1
df75626840c76c15	Raw fMOST image files for ~30 mouse specimens	<b>Investigator</b> Hongkui Zeng Zeng Lab Allen Institute for Brain Science <b>Funding</b> 1-U19-MH114830-01
1535a69f63702bca	8 mouse enhancer virus labeling two-photon serial tomography coronal image data sets	<b>Experiment</b> <i>Modality:</i> cell morphology <i>Method:</i> fMOST <i>Technique:</i> fMOST <i>Structure:</i> Whole brain <i>Organism:</i> mouse <i>Cells:</i> 0

BIL: /bil/data/07/50/07504513d424f013/  
[HTTPS: https://download.brainimagelibrary.org/07/50/07504513d424f013/](https://download.brainimagelibrary.org/07/50/07504513d424f013/)

BIL: /bil/data/df/75/df75626840c76c15/mouseID\_367667-18052  
[HTTPS: https://download.brainimagelibrary.org/df/75/df75626840c76c15/mouseID\\_367667-18052](https://download.brainimagelibrary.org/df/75/df75626840c76c15/mouseID_367667-18052)

© 2023 Allen Institute, All Rights Reserved

## 2. Search portal - submit.brainimagelibrary.org/search

The image shows the Brain Image Library search portal. At the top, there is a navigation bar with links for "About", "Data Submission", "Data Access", and "Contact". Below the navigation bar, a banner displays various statistics: 6399 public data entries, 6 organism types, 46 techniques, 23 modalities, 95 methods, 34 grant numbers, and 478 anatomical structures. A message "Active Filters: None" is displayed below the statistics. On the left side, there is a sidebar with "Add Filter" dropdowns for Organism type, Technique, Modality, Method, Anatomical Structure, Investigator, Grant number, and Collection ID. There is also a "Download" section with a link to "Brain Inventory" and a "csv format" option. A red circle highlights the "Download" section. The main content area shows search results for "Results: 6399". The first result is for a cortical interneuron from a marmoset, imaged in December 2020. It includes details about the sample, investigator (Guoping Feng, Feng Lab, Massachusetts Institute of Technology), funding (1-U01-MH114819-01), experiment (smFISH, spatial transcriptomics), and organism (marmoset). The second result is for raw fMOTST image files from the Hongkui Zeng lab at the Allen Institute for Brain Science, using cell morphology and fMOTST techniques on mouse brain tissue.

Brain Image Library

About ▾ Data Submission ▾ Data Access ▾ Contact

6399 public data 6 organism types 46 techniques 23 modalities 95 methods 34 grant numbers 478 anatomical structures

Active Filters: None

Add Filter:

- Organism type ▾
- Technique ▾
- Modality ▾
- Method ▾
- Anatomical Structure:
- Structure
- Investigator:
- Name
- Grant number:
- Grant Number
- Collection ID:
- Collection ID

**Download**

Brain Inventory  
6399 entries  
( csv format )

**Results: 6399**

**07504513d424f013** Image and morphological reconstruction of a cortical interneuron (3b) from marmoset, Yanma, after systemic viral IV injection with AAV2/9-hDlx-GFP-fGFP, 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**  
Guoping Feng  
Feng Lab  
Massachusetts Institute of Technology

**Funding**  
1-U01-MH114819-01

**Experiment**  
Modality: spatial transcriptomics  
Method: smFISH  
Technique: smFISH  
Structure: Cortex  
Organism: marmoset  
TransLine: CJYanma  
Cells: 1

BIL: /bil/data/07/50/07504513d424f013/  
[HTTPS: https://download.brainimagelibrary.org/07/50/07504513d424f013/](https://download.brainimagelibrary.org/07/50/07504513d424f013/)

**df75626840c76c15** Raw fMOTST image files for ~30 mouse specimens

**Sample:** AIBS\_730859101

 **BICCN**

**Investigator**  
Hongkui Zeng  
Zeng Lab  
Allen Institute for Brain Science

**Funding**  
1-U19-MH114830-01

**Experiment**  
Modality: cell morphology  
Method: fMOTST  
Technique: fMOTST  
Structure: Whole brain  
Organism: mouse  
Cells: 0

BIL: /bil/data/df/75/df75626840c76c15/mouselID\_367667-18052  
[HTTPS: https://download.brainimagelibrary.org/df/75/df75626840c76c15/mouselID\\_367667-18052](https://download.brainimagelibrary.org/df/75/df75626840c76c15/mouselID_367667-18052)

1535a69f63702bca 8 mouse enhancer virus labeling two-photon serial tomography coronal image data sets

© 2023 Brain Image Library, All Rights Reserved

### 3. DOI - doi.brainimagelibrary.org/doi/

The screenshot shows a dataset landing page. At the top, there's a navigation bar with links for 'About', 'Data Submission', 'Data Access', and 'Contact'. Below the navigation is a large image of a brain. The main content area has a white background and contains the following information:

**Cell type distribution in female and male mouse brains /  
Nos1\_GFP\_M6\_200525**

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

**Dataset Citation:**  
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>

**Abstract:**  
This project proposes to establish an Anatomical Collaboratory for Systematic Atlasing of Cell-Type Distribution and Morphology in female and male brain. This Collaboratory will apply standardized, unbiased and largely automated methods developed in the group's laboratories to atlas the distribution and morphology of >80 molecularly defined cell classes and cell type across all regions of the female and male brain. This work will yield the most comprehensive characterization of cell type anatomy in the mammalian brain to date, establishing a structural basis for building an integrated Cell Type Brain Atlas.

**Methods:**  
No description found.

**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
NCBITaxonomy	NCBITaxid10090

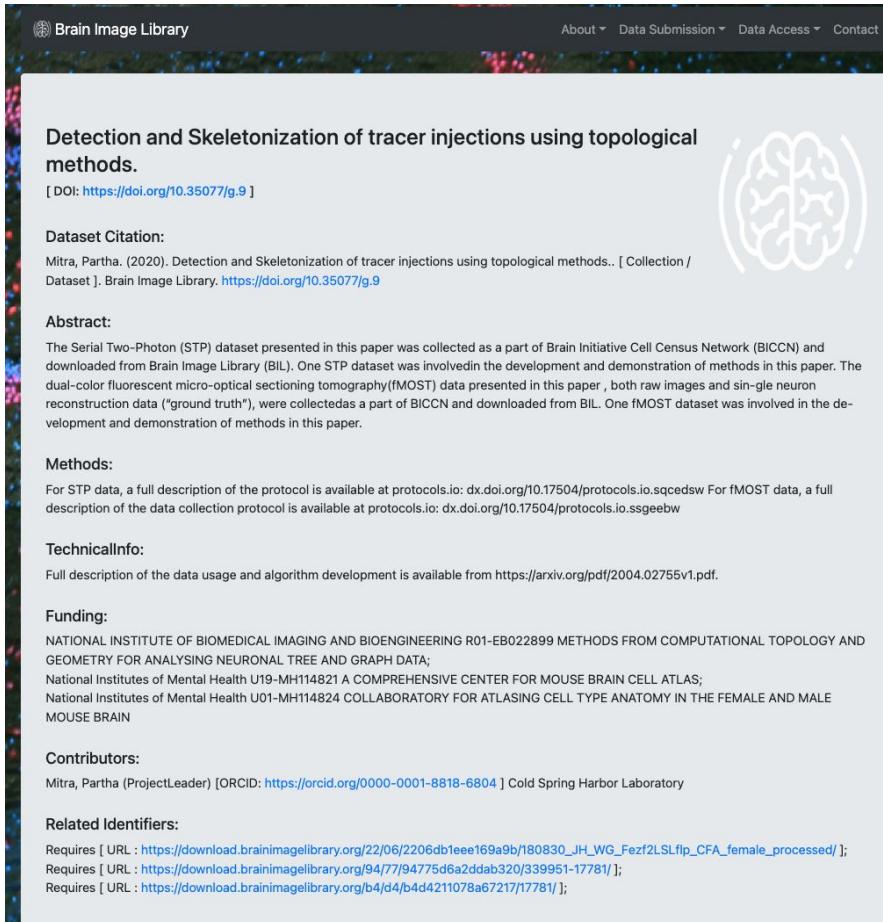
Datasets with DOIs issued will have a landing page that displays more in-depth metadata for the associated dataset:

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

DOIs can also be issued for groups of related datasets:

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

### 3. DOI - doi.brainimagelibrary.org/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 navigation is a large image of a brain. The main content area has the following sections:

- 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
- TechnicalInfo:**

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/](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.brainimagelibrary.org/doi/10.35077/g.9>

DOIs can also be issued for groups of related datasets:

<https://doi.brainimagelibrary.org/doi/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.
<b>Sample:</b>	
Cortex 20201201_	
CJYanma_Ex10_	
PVALB+THSD7a_-	
Hippo-Region	
004	
 BICCN	
<b>Investigator</b>	<b>Experiment</b>
Guoping Feng	<i>Modality:</i> spatial transcriptomics
Feng Lab	<i>Method:</i> smFISH
Massachusetts Institute of Technology	<i>Technique:</i> smFISH
<b>Funding</b>	<i>Structure:</i> Cortex
1-U01-MH114819-01	<i>Organism:</i> marmoset
	<i>TransLine:</i> CJYanma
	<i>Cells:</i> 1
<b>BIL:</b> /bil/data/07/50/07504513d424f013/	
<b>HTTPS:</b> <a href="https://download.brainimagerlibrary.org/07/50/07504513d424f013/">https://download.brainimagerlibrary.org/07/50/07504513d424f013/</a>	

- 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

<b>TechnicalInfo:</b>	
<b>DATASET METADATA:</b>	
title	Cell type distribution in female and male mouse brains / Nos1_GFP_M6_200525
GeneralModality	population imaging
Technique	other
Other	cell distribution
<b>SPECIMEN METADATA:</b>	
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
<b>INSTRUMENT METADATA:</b>	
MicroscopeType	Two Photon
MicroscopeManufacturerAndModel	TissueVision
<b>IMAGE METADATA:</b>	
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



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
- [X2Go](#)
  - X2Go is open source remote desktop software and gives remote access to a Linux system graphical user interface.
- [TGX](#)
  - TGX is a remote desktop for low-latency remote desktop workstation for intensive graphics applications up to 4K resolution.

# 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/>



# Questions?

## BIL Office Hours

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

Schedule: Second Wednesday of the month

Time: 2-3PM Eastern

Register for this meeting:

<https://cmu.zoom.us/meeting/register/tJElD-6vrTIsH9xdwa6uCDXIrBO-pIQicsLF>



***For questions at any time contact us at [bil-support@psc.edu](mailto: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](mailto:bil-support@psc.edu) with any questions**