Data Discovery / Metadata Subgroup
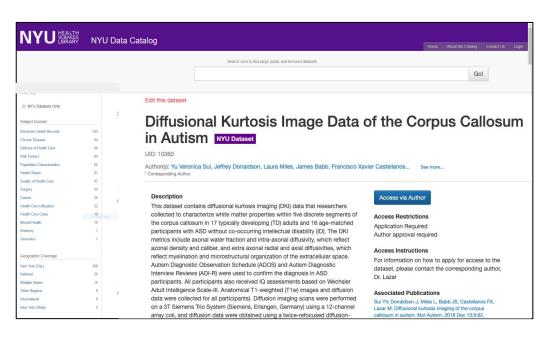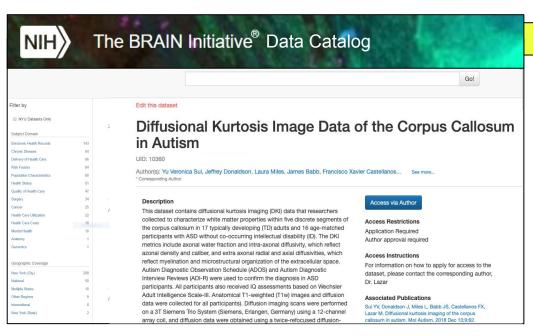
Improving Data Harmonization and Discovery of BRAIN Initiative Datasets

Making BRAIN Initiative Datasets Discoverable

Making datasets discoverable regardless of where they are stored

Open Source
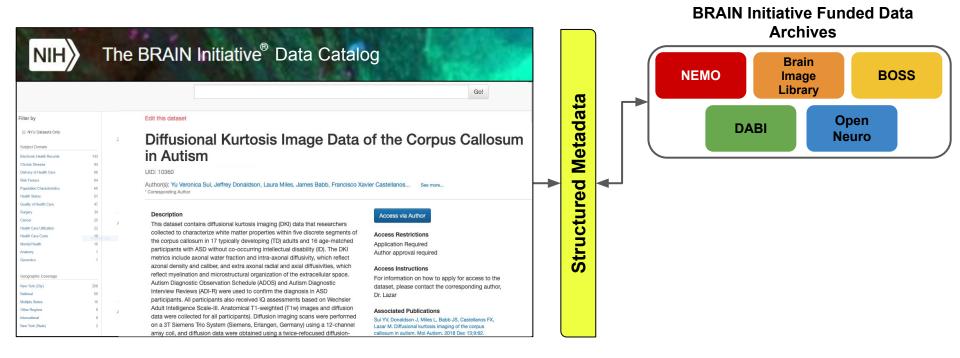
Implemented at 8 institutions

Making BRAIN Initiative Datasets Discoverable
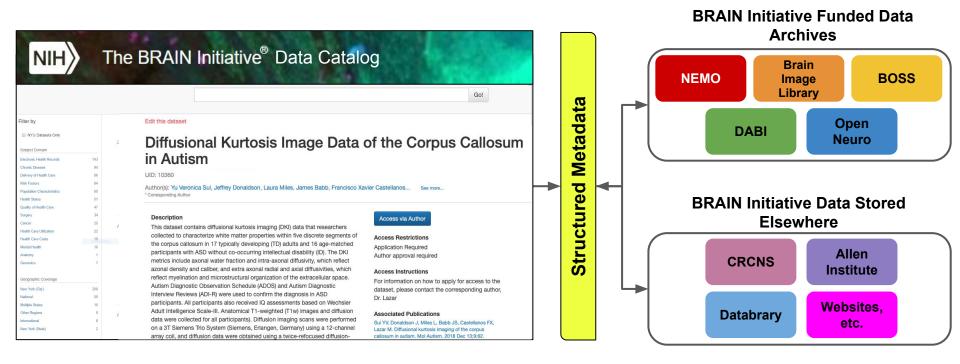

Structured Metadata

- Dataset Descriptions

- Experimental parameters

- Data standard used (i.e., NWB)

- Access instructions

- Data use agreements

- Associated publications

Making BRAIN Initiative Datasets Discoverable

BRAIN Initiative Funded Data Archives

Making BRAIN Initiative Datasets Discoverable

Making BRAIN Initiative Datasets Discoverable

DEMO in breakout session

BRAIN Initiative Funded Data Archives

NEMO · Brain Image Library · BOSS · DABI · Open Neuro

Structured Metadata

BRAIN Initiative Data Stored Elsewhere

CRCNS · Allen Institute · Databrary · Websites, etc.

Meeting Summary (2/1/2019)

Takeaways:

- No single storage/sharing/analysis platform will contain all data types

- Data Catalog points to all data regardless of where it is stored

- Customized metadata can be informed by NIH funded data archives and/or U19 subgroups (for datasets with no existing standards in place)

- Ideal for NIH to host Data Catalog

- **U19 Data Science Consortium best for initial scope of Data Catalog**

Support Needed:

Startup:
- Installation
- Branding
- Metadata customization

Ongoing:
- Curation of datasets
- Outreach efforts
- Software updates
- Metadata modifications
- Governance
- Maintenance

Neuroscience metadata database

- Electronic notebook for lab members to organize metadata from experiments and link to raw data

- Allows other investigators to discover and understand data both within and across laboratories

- A REST API - Coordinates analysis tools with that same notebook

Database navigation

PROJECTS · ANIMAL SUBJECTS · SESSIONS · PERSONAL MAZES · GENERAL ATTRIBUTES · ANIMAL METADATA · BEHAVIORAL PARADIGMS · PERSONAL EQUIPMENT · DATA REPOSITORIES

Working towards a centralized neuroscience metadata database

- User friendly web interface
- Data model to support wide range of experimental data types
- Minimal barrier between internal use, collaborations and public sharing
- Centralized, standardized metadata makes experimental data
  - More discoverable
  - More interpretable
  - More accessible

Effective data sharing

Direct sharing using globus.org or via a data archive like CRCNS

Effective data sharing
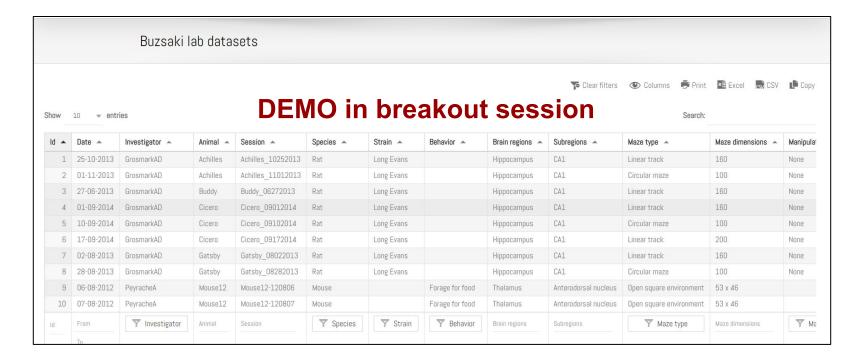

Buzsaki lab datasets

DEMO in breakout session

Show 10 ▾ entries   Search:

Id ▲	Date ▲	Investigator ▲	Animal ▲	Session ▲	Species ▲	Strain ▲	Behavior ▲	Brain regions ▲	Subregions ▲	Maze type ▲	Maze dimensions ▲	Manipula
1	25-10-2013	GrosmarkAD	Achilles	Achilles_10252013	Rat	Long Evans		Hippocampus	CA1	Linear track	160	None
2	01-11-2013	GrosmarkAD	Achilles	Achilles_11012013	Rat	Long Evans		Hippocampus	CA1	Circular maze	100	None
3	27-06-2013	GrosmarkAD	Buddy	Buddy_06272013	Rat	Long Evans		Hippocampus	CA1	Linear track	160	None
4	01-09-2014	GrosmarkAD	Cicero	Cicero_09012014	Rat	Long Evans		Hippocampus	CA1	Linear track	160	None
5	10-09-2014	GrosmarkAD	Cicero	Cicero_09102014	Rat	Long Evans		Hippocampus	CA1	Circular maze	100	None
6	17-09-2014	GrosmarkAD	Cicero	Cicero_09172014	Rat	Long Evans		Hippocampus	CA1	Linear track	200	None
7	02-08-2013	GrosmarkAD	Gatsby	Gatsby_08022013	Rat	Long Evans		Hippocampus	CA1	Linear track	160	None
8	28-08-2013	GrosmarkAD	Gatsby	Gatsby_08282013	Rat	Long Evans		Hippocampus	CA1	Circular maze	100	None
9	06-08-2012	PeyracheA	Mouse12	Mouse12-120806	Mouse		Forage for food	Thalamus	Anterodorsal nucleus	Open square environment	53 x 46	
10	07-08-2012	PeyracheA	Mouse12	Mouse12-120807	Mouse		Forage for food	Thalamus	Anterodorsal nucleus	Open square environment	53 x 46	

Direct sharing using globus.org or via a data archive like CRCNS

Alignment with FAIR Principles

	BRAIN INITIATIVE DATA CATALOG	METADATA TOOL
Findable	All BRAIN Initiative datasets discoverable through one tool	Discovery of data elements at granular level
Accessible		
Interoperable		Standardized vocabulary allows data to be linked
Reusable		Detailed metadata allows meaningful reuse

Alignment with FAIR Principles

	BRAIN INITIATIVE DATA CATALOG	METADATA TOOL
Findable	All BRAIN Initiative datasets discoverable through one tool	Discovery of data elements at granular level
Accessible		
Interoperable		Standardized vocabulary allows data to be linked
Reusable		Detailed metadata allows meaningful reuse

Alignment with FAIR Principles

	BRAIN INITIATIVE DATA CATALOG	METADATA TOOL
Findable	All BRAIN Initiative datasets discoverable through one tool	Discovery of data elements at granular level
Accessible		
Interoperable		Standardized vocabulary allows data to be linked
Reusable		Detailed metadata allows meaningful reuse

Subgroup Discussion Points

1. Where does your data live now?

    a. What metadata does it have?
    b. How do you determine what metadata is captured (e.g. standardized across a lab or collaboration?)

2. How much have you shared your own data? How did you do it?

    a. Did you use a repository (why/why not)?
    b. Did you encounter security or privacy issues?
    c. Other barriers you faced when sharing?

3. Have you searched for data? How did you do it?