

# **Data Discovery / Metadata Subgroup**

Improving Data Harmonization and Discovery of BRAIN  
Initiative Datasets

# Making BRAIN Initiative Datasets Discoverable

The screenshot shows the NYU Health Sciences Library Data Catalog interface. The header includes the NYU logo and 'NYU Data Catalog' with navigation links for Home, About the Catalog, Contact Us, and Login. A search bar is present with the placeholder text 'Search here to find large public and licensed datasets' and a 'Go!' button. The main content area displays the dataset 'Diffusional Kurtosis Image Data of the Corpus Callosum in Autism' with a 'NYU Dataset' tag. It includes the UID: 10360, author information (Yu Veronica Sul, Jeffrey Donaldson, Laura Miles, James Babb, Francisco Xavier Castellanos...), and a 'Description' section. A 'Description' section provides details about the dataset, mentioning it contains Diffusional Kurtosis Imaging (DKI) data for 17 typically developing (TD) adults and 16 age-matched participants with ASD. It also lists 'Access Restrictions' (Application Required, Author approval required), 'Access Instructions' (For information on how to apply for access to the dataset, please contact the corresponding author, Dr. Lazar), and 'Associated Publications' (Sul YV, Donaldson J, Miles L, Babb JS, Castellanos FX, Lazar M. Diffusional kurtosis imaging of the corpus callosum in autism. Mol Autism. 2018 Dec 13;9:62).

**NYU HEALTH SCIENCES LIBRARY** NYU Data Catalog

Home About the Catalog Contact Us Login

Search here to find large public and licensed datasets

Go!

[Edit this dataset](#)

## Diffusional Kurtosis Image Data of the Corpus Callosum in Autism NYU Dataset

UID: 10360

Author(s): Yu Veronica Sul, Jeffrey Donaldson, Laura Miles, James Babb, Francisco Xavier Castellanos... [See more...](#)

\* Corresponding Author

### Description

This dataset contains diffusional kurtosis imaging (DKI) data that researchers collected to characterize white matter properties within five discrete segments of the corpus callosum in 17 typically developing (TD) adults and 16 age-matched participants with ASD without co-occurring intellectual disability (ID). The DKI metrics include axonal water fraction and intra-axonal diffusivity, which reflect axonal density and caliber, and extra axonal radial and axial diffusivities, which reflect myelination and microstructural organization of the extracellular space. Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview Reviews (ADI-R) were used to confirm the diagnosis in ASD participants. All participants also received IQ assessments based on Wechsler Adult Intelligence Scale-III. Anatomical T1-weighted (T1w) images and diffusion data were collected for all participants. Diffusion imaging scans were performed on a 3T Siemens Trio System (Siemens, Erlangen, Germany) using a 12-channel array coil, and diffusion data were obtained using a twice-refocused diffusion-

[Access via Author](#)

### Access Restrictions

Application Required  
Author approval required

### Access Instructions

For information on how to apply for access to the dataset, please contact the corresponding author, Dr. Lazar

### Associated Publications

Sul YV, Donaldson J, Miles L, Babb JS, Castellanos FX, Lazar M. Diffusional kurtosis imaging of the corpus callosum in autism. Mol Autism. 2018 Dec 13;9:62.

Subject Domain	Count
Electronic Health Records	143
Chronic Disease	94
Delivery of Health Care	86
Risk Factors	84
Population Characteristics	60
Health Status	51
Quality of Health Care	47
Surgery	34
Cancer	25
Health Care Utilization	22
Health Care Costs	18
Mental Health	10
Anatomy	1
Genomics	1

Geographic Coverage	Count
New York (City)	208
National	50
Multiple States	10
Other Regions	9
International	6
New York (State)	2

**Making datasets discoverable regardless of where they are stored**

**Open Source**

**Implemented at 8 institutions**

# Making BRAIN Initiative Datasets Discoverable

The screenshot shows the NIH The BRAIN Initiative Data Catalog interface. At the top left is the NIH logo and the text 'The BRAIN Initiative® Data Catalog'. A search bar with a 'Go!' button is located at the top right. On the left side, there is a 'Filter by' section with a sidebar menu listing various categories and their counts. The main content area displays the dataset title 'Diffusional Kurtosis Image Data of the Corpus Callosum in Autism' with a UID of 10360. Below the title, the authors are listed: Yu Veronica Sul, Jeffrey Donaldson, Laura Miles, James Babb, Francisco Xavier Castellanos. A 'Description' section provides details about the dataset, including the number of participants and the imaging techniques used. There are also sections for 'Access Restrictions', 'Access Instructions', and 'Associated Publications'.

**NIH** The BRAIN Initiative® Data Catalog

Go!

Filter by

NYU Datasets Only

Subject Domain

Electronic Health Records	143
Chronic Disease	94
Delivery of Health Care	86
Risk Factors	84
Population Characteristics	60
Health Status	51
Quality of Health Care	47
Surgery	34
Cancer	25
Health Care Utilization	22
Health Care Costs	18
Mental Health	10
Anatomy	1
Genetics	1

Geographic Coverage

New York (City)	206
National	50
Multiple States	10
Other Regions	9
International	6
New York (State)	2

Edit this dataset

## Diffusional Kurtosis Image Data of the Corpus Callosum in Autism

UID: 10360

Author(s): Yu Veronica Sul, Jeffrey Donaldson, Laura Miles, James Babb, Francisco Xavier Castellanos... See more...  
\* Corresponding Author

**Description**

This dataset contains diffusional kurtosis imaging (DKI) data that researchers collected to characterize white matter properties within five discrete segments of the corpus callosum in 17 typically developing (TD) adults and 16 age-matched participants with ASD without co-occurring intellectual disability (ID). The DKI metrics include axonal water fraction and intra-axonal diffusivity, which reflect axonal density and caliber, and extra axonal radial and axial diffusivities, which reflect myelination and microstructural organization of the extracellular space. Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview Reviews (ADI-R) were used to confirm the diagnosis in ASD participants. All participants also received IQ assessments based on Wechsler Adult Intelligence Scale-III. Anatomical T1-weighted (T1w) images and diffusion data were collected for all participants. Diffusion imaging scans were performed on a 3T Siemens Trio System (Siemens, Erlangen, Germany) using a 12-channel array coil, and diffusion data were obtained using a twice-refocused diffusion-

**Access via Author**

**Access Restrictions**  
Application Required  
Author approval required

**Access Instructions**  
For information on how to apply for access to the dataset, please contact the corresponding author, Dr. Lazar

**Associated Publications**  
Sul YV, Donaldson J, Miles L, Babb JS, Castellanos FX, Lazar M. Diffusional kurtosis imaging of the corpus callosum in autism. *Mol Autism*. 2018 Dec 13;9:62.

## Structured Metadata

- Dataset Descriptions
- Experimental parameters
- Data standard used (i.e., NWB)
- Access instructions
- Data use agreements
- Associated publications

# Making BRAIN Initiative Datasets Discoverable

**NIH** The BRAIN Initiative® Data Catalog

Go!

Filter by

- NYU Datasets Only

Subject Domain

Electronic Health Records	143
Chronic Disease	94
Delivery of Health Care	86
Risk Factors	84
Population Characteristics	60
Health Status	51
Quality of Health Care	47
Surgery	34
Cancer	25
Health Care Utilization	22
Health Care Costs	22
Mental Health	18
Anatomy	1
Genetics	1

Geographic Coverage

New York (City)	208
National	50
Multiple States	10
Other Regions	9
International	8
New York (State)	2

[Edit this dataset](#)

## Diffusional Kurtosis Image Data of the Corpus Callosum in Autism

UID: 10360

Author(s): Yu Veronica Sul, Jeffrey Donaldson, Laura Miles, James Babb, Francisco Xavier Castellanos... [See more...](#)

\* Corresponding Author

[Access via Author](#)

**Description**

This dataset contains diffusional kurtosis imaging (DKI) data that researchers collected to characterize white matter properties within five discrete segments of the corpus callosum in 17 typically developing (TD) adults and 16 age-matched participants with ASD without co-occurring intellectual disability (ID). The DKI metrics include axonal water fraction and intra-axonal diffusivity, which reflect axonal density and caliber, and extra axonal radial and axial diffusivities, which reflect myelination and microstructural organization of the extracellular space. Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview Reviews (ADI-R) were used to confirm the diagnosis in ASD participants. All participants also received IQ assessments based on Wechsler Adult Intelligence Scale-III. Anatomical T1-weighted (T1w) images and diffusion data were collected for all participants. Diffusion imaging scans were performed on a 3T Siemens Trio System (Siemens, Erlangen, Germany) using a 12-channel array coil, and diffusion data were obtained using a twice-refocused diffusion-

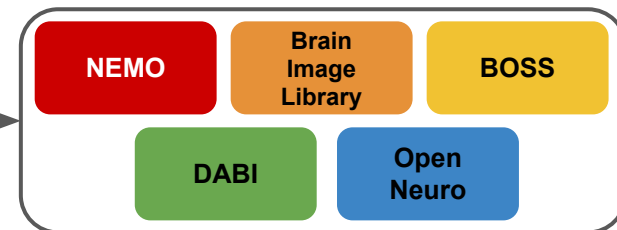
**Access Restrictions**  
Application Required  
Author approval required

**Access Instructions**  
For information on how to apply for access to the dataset, please contact the corresponding author, Dr. Lazar

**Associated Publications**  
Sui YV, Donaldson J, Miles L, Babb JS, Castellanos FX, Lazar M. Diffusional kurtosis imaging of the corpus callosum in autism. *Mol Autism*. 2018 Dec 13;9:62.

Structured Metadata

## BRAIN Initiative Funded Data Archives



# Making BRAIN Initiative Datasets Discoverable

**NIH** The BRAIN Initiative® Data Catalog

Go!

Filter by

- NYU Datasets Only

Subject Domain

Electronic Health Records	143
Chronic Disease	94
Delivery of Health Care	86
Risk Factors	84
Population Characteristics	60
Health Status	51
Quality of Health Care	47
Surgery	34
Cancer	25
Health Care Utilization	22
Health Care Costs	18
Mental Health	10
Anatomy	1
Genetics	1

Geographic Coverage

New York (City)	206
National	10
Multiple States	10
Other Regions	9
International	6
New York (State)	2

**Diffusional Kurtosis Image Data of the Corpus Callosum in Autism**

UID: 10360

Author(s): Yu Veronica Sul, Jeffrey Donaldson, Laura Miles, James Babb, Francisco Xavier Castellanos... See more...  
\* Corresponding Author

**Description**

This dataset contains diffusional kurtosis imaging (DKI) data that researchers collected to characterize white matter properties within five discrete segments of the corpus callosum in 17 typically developing (TD) adults and 16 age-matched participants with ASD without co-occurring intellectual disability (ID). The DKI metrics include axonal water fraction and intra-axonal diffusivity, which reflect axonal density and caliber, and extra axonal radial and axial diffusivities, which reflect myelination and microstructural organization of the extracellular space. Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview Reviews (ADI-R) were used to confirm the diagnosis in ASD participants. All participants also received IQ assessments based on Wechsler Adult Intelligence Scale-III. Anatomical T1-weighted (T1w) images and diffusion data were collected for all participants. Diffusion imaging scans were performed on a 3T Siemens Trio System (Siemens, Erlangen, Germany) using a 12-channel array coil, and diffusion data were obtained using a twice-refocused diffusion-

**Access via Author**

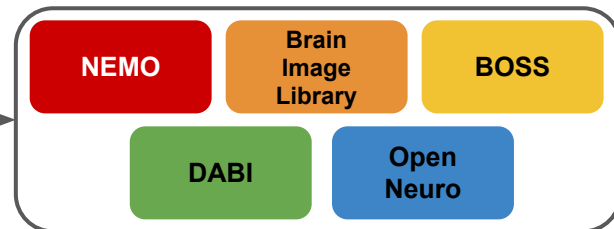
**Access Restrictions**  
Application Required  
Author approval required

**Access Instructions**  
For information on how to apply for access to the dataset, please contact the corresponding author, Dr. Lazar

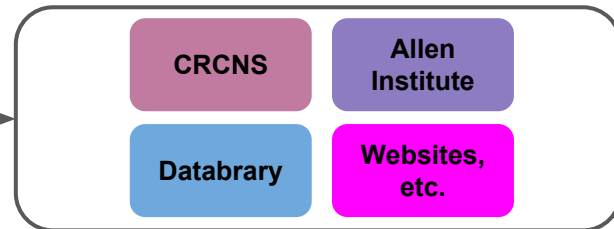
**Associated Publications**  
Sui YV, Donaldson J, Miles L, Babb JS, Castellanos FX, Lazar M. Diffusional kurtosis imaging of the corpus callosum in autism. *Mol Autism*. 2018 Dec 13;9:62.

Structured Metadata

## BRAIN Initiative Funded Data Archives



## BRAIN Initiative Data Stored Elsewhere



# Making BRAIN Initiative Datasets Discoverable

DEMO in breakout session

**NIH** The BRAIN Initiative® Data Catalog

Go!

Filter by

- NYU Datasets Only

Subject Domain

- Electronic Health Records 143
- Chronic Disease 94
- Delivery of Health Care 86
- Risk Factors 84
- Population Characteristics 60
- Health Status 51
- Quality of Health Care 47
- Surgery 34
- Cancer 25
- Health Care Utilization 22
- Health Care Costs 18
- Mental Health 10
- Anatomy 1
- Genetics 1

Geographic Coverage

- New York (City) 206
- National 10
- Multiple States 10
- Other Regions 9
- International 6
- New York (State) 2

**Diffusional Kurtosis Image Data of the Corpus Callosum in Autism**

UID: 10360

Author(s): Yu Veronica Sul, Jeffrey Donaldson, Laura Miles, James Babb, Francisco Xavier Castellanos... See more...  
\* Corresponding Author

**Description**

This dataset contains diffusional kurtosis imaging (DKI) data that researchers collected to characterize white matter properties within five discrete segments of the corpus callosum in 17 typically developing (TD) adults and 16 age-matched participants with ASD without co-occurring intellectual disability (ID). The DKI metrics include axonal water fraction and intra-axonal diffusivity, which reflect axonal density and caliber, and extra axonal radial and axial diffusivities, which reflect myelination and microstructural organization of the extracellular space. Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview Reviews (ADI-R) were used to confirm the diagnosis in ASD participants. All participants also received IQ assessments based on Wechsler Adult Intelligence Scale-III. Anatomical T1-weighted (T1w) images and diffusion data were collected for all participants. Diffusion imaging scans were performed on a 3T Siemens Trio System (Siemens, Erlangen, Germany) using a 12-channel array coil, and diffusion data were obtained using a twice-refocused diffusion-

**Access via Author**

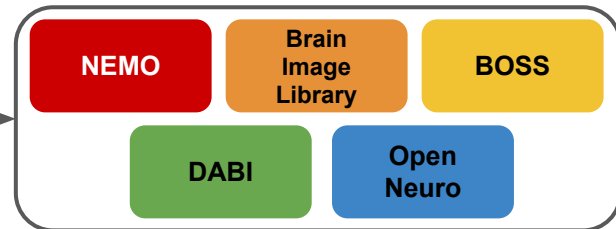
**Access Restrictions**  
Application Required  
Author approval required

**Access Instructions**  
For information on how to apply for access to the dataset, please contact the corresponding author, Dr. Lazar

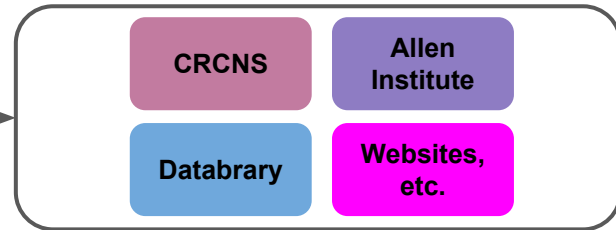
**Associated Publications**  
Sui YV, Donaldson J, Miles L, Babb JS, Castellanos FX, Lazar M. Diffusional kurtosis imaging of the corpus callosum in autism. *Mol Autism*. 2018 Dec 13;9:62.

Structured Metadata

## BRAIN Initiative Funded Data Archives



## BRAIN Initiative Data Stored Elsewhere



# 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

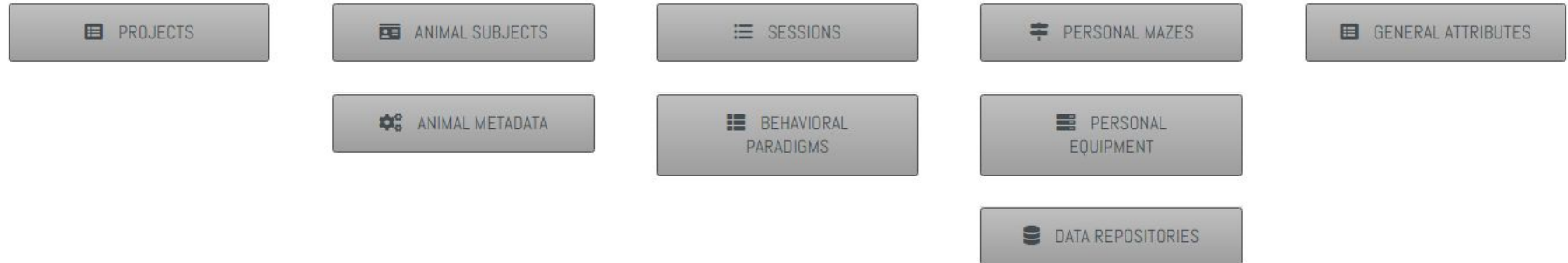
### Ongoing:

- 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





# 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

The screenshot shows a web interface titled "Sessions". It features a search bar labeled "Search all fields" and two dropdown menus for "Project" and "Animal". Below these are two rows of filters: "Behavioral paradigm", "Manipulation", "Brain region", "Date from", and "Date to". At the bottom, there are two toggle switches for "Spike sorted" and "Cell metrics". A "Search" button and a "Reset" button are located to the right of the search bar.

Sessions

Search all fields  Project  Animal

Behavioral paradigm  Manipulation  Brain region  Date from  Date to

Spike sorted  Cell metrics

Search Reset

# Effective data sharing

Buzsaki lab datasets

Clear filters Columns Print Excel CSV Copy

Show 10 entries Search:

Id	Date	Investigator	Animal	Session	Species	Strain	Behavior	Brain regions	Subregions	Maze type	Maze dimensions	Manipular
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	

Id From To Investigator Animal Session Species Strain Behavior Brain regions Subregions Maze type Maze dimensions Me

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

# Effective data sharing

Buzsaki lab datasets

Clear filters Columns Print Excel CSV Copy

DEMO in breakout session

Show 10 entries Search:

Id	Date	Investigator	Animal	Session	Species	Strain	Behavior	Brain regions	Subregions	Maze type	Maze dimensions	Manipular
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	

Id From To Investigator Animal Session Species Strain Behavior Brain regions Subregions Maze type Maze dimensions Me

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

# Alignment with FAIR Principles

## BRAIN INITIATIVE DATA CATALOG

## METADATA TOOL

<b>F</b> indable	All BRAIN Initiative datasets discoverable through one tool	Discovery of data elements at granular level
<b>A</b> ccessible		
<b>I</b> nteroperable		Standardized vocabulary allows data to be linked
<b>R</b> eusable		Detailed metadata allows meaningful reuse

# Alignment with FAIR Principles

	BRAIN INITIATIVE DATA CATALOG	METADATA TOOL
<b>F</b> indable	All BRAIN Initiative datasets discoverable through one tool	Discovery of data elements at granular level
<b>A</b> ccessible		
<b>I</b> nteroperable		Standardized vocabulary allows data to be linked
<b>R</b> eusable		Detailed metadata allows meaningful reuse

# Alignment with FAIR Principles

	BRAIN INITIATIVE DATA CATALOG	METADATA TOOL
<b>F</b> indable	All BRAIN Initiative datasets discoverable through one tool	Discovery of data elements at granular level
<b>A</b> ccessible		
<b>I</b> nteroperable		Standardized vocabulary allows data to be linked
<b>R</b> eusable		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?