Best practices for managing data sharing in the lab

Maryann E. Martone, Ph. D.*
Adam R. Ferguson, Ph. D.

PRECISE-TBI

*Founder and equity interest in SciCrunch Inc

ODC -TBI.org

What constitutes good data management

- An intentional lab-centric strategy designed to maximize potential for effective sharing:
  - You
  - Future you
  - Your lab
  - Your colleagues
  - Publishing data in a trusted repository

	Ad Hoc	One-Time	Active and Informative	Optimized for Re-Use
Planning your project	When it comes to my data, I have a "way of doing things" but no standard or documented plans.	I create some formal plans about how I will manage my data at the start of a project, but I generally don't refer back to them.	I develop detailed plans about how I will manage my data that I actively revisit and revise over the course of a project.	I have created plans for managing my data that are designed to streamline its future use by myself or others.
Organizing your data	I don't follow a consistent approach for keeping my data organized, so it often takes time to find things.	I have an approach for organizing my data, but I only put it into action after my project is complete.	I have an approach for organizing my data that I implement prospectively, but it not necessarily standardized.	I organize my data so that others can navigate, understand, and use it without me being present.
Saving and backing up your data	I decide what data is important while I am working on it and typically save it in a single location.	I know what data needs to be saved and I back it up after I'm done working on it to reduce the risk of loss.	I have a system for regularly saving important data while I am working on it. I have multiple backups.	I save my data in a manner and location designed maximize opportunities for re-use by myself and others.
Getting your data ready for analysis	I don't have a standardized or well documented process for preparing my data for analysis.	I have thought about how I will need to prepare my data, but I handle each case in a different manner.	My process for preparing data is standardized and well documented.	I prepare my data in such a way as to facilitate use by both myself and others in the future.

Borghi J, Abrams S, Lowenberg D, Simms S, Chodacki J (2018) **Support Your Data: A Research Data Management Guide for Researchers.** Research Ideas and Outcomes 4: e26439. https://doi.org/10.3897/rio.4.e26439

How? FAIR throughout data lifecycle

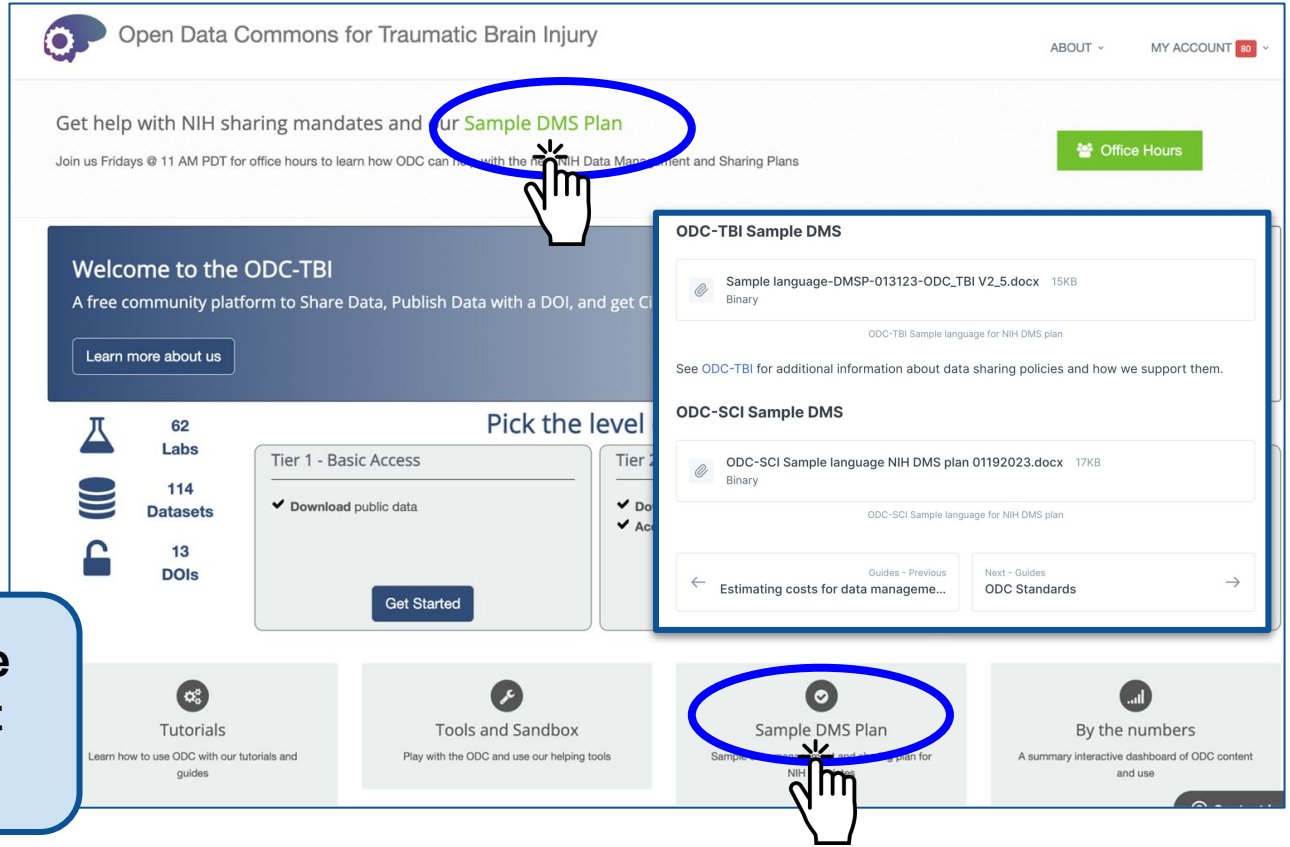
https://iussp.org/sites/default/files/FAIR_Data.png

- Practices that are designed to increase the utility of biomedical data
- Emphasizes both **human** and **computational** utility
- NIH data management and sharing policy designed to promote FAIR data stewardship:
  - Identifiers
  - Metadata
  - Documentation
  - Standards
  - Provenance
  - Licenses/Access rights

ODC-TBI: Trusted repository for FAIR preclinical data

- NINDS recommended repository for pre-clinical TBI
- Community governed: ODC-TBI implements PRECISE recommendations

Knowing where you are going can help you get there more quickly!



F: Meaningful title that aids in search

A: How do I access the files

R: What if I have a question?

F: Persistent identifier

I, R: Common, open formats, community standards and documentation

R: Who is allowed to use this data?

F: Description of data set

R: Provenance

I, R: Links to tools, protocols etc

F,R: Who performed the work?

ZIP CSV

Full Data Package | Data Dictionary

ODC-TBI Public

Home-cage monitoring spontaneous activity of C57BL/6J male mi… months after open-field low-intensity blast exposure

DOI:10.34945/F5FK5C

DATASET CITATION

Zuckerman A., Siedhoff H. R., Balderra…
male mice 3 months after open-fiel…
TBI:871 http://doi.org/10.34945/…

ABSTRACT

STUDY PURPOSE: Evaluate the…
home-cage-like environment.

DATA COLLECTED: A total of 52 male C… of two groups: Blast (n=29) or Sham (n=23). Mice … re (46.6 kPa, maximum impulse of 60.0 kPa*ms), under anesthesia… but were not exposed to the blast wave. 3 months post-exposure, the spontaneous activity of the mice was measured using the PhenoTyper® home-cages (L = 30 × W = 30 × H = 35 cm; Model 3000, Noldus Information Technology, The Netherlands). Each mouse was housed individually, and its activity was continuously measured for 72 hours at a sample rate of 15 fps. Program-acquired data were uploaded to the web-based AHCODA-DB (Sylics, Bilthoven, The Netherlands) for meta-analysis. Twenty behavioral parameters were analyzed and included in this dataset. See protocols and other related data in the relevant links section below.

CONCLUSIONS: No significant differences were found between the Blast and Sham mice in different parameters of general daily performance behaviors, such as activity, arrests, and feeding zone visits.
Although no significant difference in long shelter visits between the Blast and Sham mice, was found, significant differences were found in multiple parameters of short shelter visits such as "shelter visit threshold" and "short shelter visit duration" (relevant to anxiety-like behaviors). Blast mice visited their shelters more frequently and for shorter periods of time than Sham mice in both dark and light phases. These results suggest that LIB-exposed mice may hold stable perceptions of environmental stimuli as a threat during activity bouts, whereas sham controls experienced such responses to a lesser degree. This type of performance is consistent as trait anxiety in humans, defined as a tendency to respond with concerns, troubles, and worries to non-threatening situations.

KEYWORDS

primary open-field blast; home-cage monitoring; PhenoTyper; spontaneous activity; anxiety-like behaviors

PROVENANCE / ORIGINATING PUBLICATIONS

- Siedhoff HR, Chen S, Balderrama A, Sun GY, Koopmans B, DePalma RG, Cui J, Gu Z. Long-Term Effects of Low-Intensity Blast Non-Inertial Brain Injury on Anxiety-Like Behaviors in Mice: Home-Cage Monitoring Assessments. Neurotrauma Reports. Jan 2022. 3.1:27-38.. doi:10.1089/neur.2021.0063.

RELEVANT LINKS

Home-cage monitoring general behavior of C57BL/6J male mice during the CognitionWall test 3 months after open-field LIB exposure
https://dx.doi.org/10.34945/F59W23
Releted dataset in ODC-TBI

Open-field blast (OFB) model in mice protocol
https://dx.doi.org/10.17504/protocols.io.yxmvm2kwog3p/v1
Protocol for the Open-field blast (OFB) model in mice in protocols.io

Open-field Blast parameters dataset
https://dx.doi.org/10.34945/F5630G

DATASET INFO

Contact: Gu Zezong (guze@health.missouri.edu)

Lab: PRECISE-TBI Lab: Truman Memorial VA
ODC-TBI Accession:871

Records in Dataset: 3744
Fields per Record: 26

Last updated: 2023-06-09
Date published: 2023-06-09
Downloads: 5

Files: 2

LICENSE

Creative Commons Attribution License (CC-BY 4.0)

FUNDING AND ACKNOWLEDGEMENTS

Department of Veterans Affairs Offices of Research & Development (VA ORD) LAMb/ShEEP programs, BLR&D Director Service program UFR-002-18F, Open-Field Blast (OFB) Core, and the Collaborative Merit Review for TBI Research Program I01 BX004313-01A1 (ZG), DoD Congressionally Directed Medical Research Programs (CDMRP) for the Peer Reviewed Alzheimer's Research Program Convergence Science Research Award PRARPCSRA; AZ180043 (ZG), Research funds of the University of Missouri to ZG.

CONTRIBUTORS

Zuckerman, Amitai [ORCID:0000-0002-7075-1913]
Harry S. Truman Memorial Veterans' Hospital Research Service, Columbia, Missouri, USA; Department of Pathology and Anatomical Sciences, University of Missouri School of Medicine, Columbia, Missouri, USA
Siedhoff, Heather R.
Harry S. Truman Memorial Veterans' Hospital Research Service, Columbia, Missouri, USA; Department of Pathology and Anatomical Sciences, University of Missouri School of Medicine, Columbia, Missouri, USA
Balderrama, Ashley
Harry S. Truman Memorial Veterans' Hospital Research Service, Columbia, Missouri, USA; Department of Pathology and Anatomical Sciences, University of Missouri School of Medicine, Columbia, Missouri, USA
Cui, Jiankun
Harry S. Truman Memorial Veterans' Hospital Research Service, Columbia, Missouri, USA; Department of Pathology and Anatomical Sciences, University of Missouri School of Medicine, Columbia, Missouri, USA
Gu, Zezong [ORCID:0000-0002-2411-7460]
Harry S. Truman Memorial Veterans' Hospital Research Service, Columbia, Missouri, USA; Department of Pathology and Anatomical Sciences, University of Missouri School of Medicine, Columbia, Missouri, USA

http://doi.org/10.34945/F59W23

FAIR and data management

- Anecdotal evidence from repositories suggests that the PI and lab members are heavy users of their own data
- Stewardship
  - Identifiers
  - Metadata
  - Documentation
  - Standards
  - Provenance
  - Licenses/Access

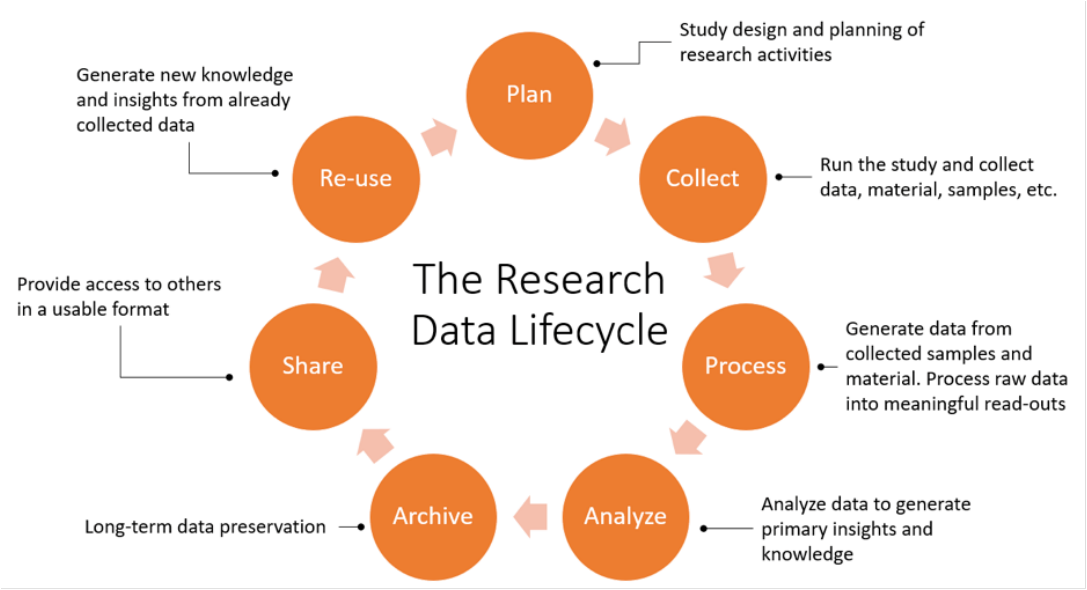


Figure 2. The Research data lifecycle

Fouad et al. (2023) **A practical guide to data management and sharing for biomedical laboratory researchers,** BioarXiv, in process.

Developing a data management workflow for your lab

1. **Requirements analysis:**
   a. What standards are available in my field?
   b. If I am required to publish in a particular repository, what does the repository require?
   c. What license do they require/allow: What data sharing agreements need to be put into place?
   d. Where will data be stored so it can be accessed by the entire lab?
   e. What metadata will be routinely collected to describe it?
2. **Create a standard data dictionary for routine data elements:**
   a. Use community CDE's if available (for preclinical TBI, they are!)
   b. Make sure the required metadata for community standards is acquired
   c. Consider a lab-wide data dictionary that is regularly updated (and versioned)
3. **Consider data formats:**
   a. Does the repository have a particular data format that is required?
   b. Are you storing your data in a proprietary format?
   c. Are you familiar with good data formatting practices? e.g., tidy spreadsheet format
4. **Generate a system for the unique identification of subjects and encourage single-subject data tracking:**
   a. Creating a central registry of subjects will help the lab learn about FAIR practices
   b. Greatly helps keep track of data collection, management, and analysis at the individual subject level.
   c. Provides clear provenance for all subjects within a given experiment
5. **Create documentation and SOPs for data workflow, including data management and sharing.**
   a. Serve as instructions, training material for newcomers, and documentation for grant applications (DMS plan).
   b. Consider storage and access, experiment registration, folder organization and file naming

Fouad et al. (2023) **A practical guide to data management and sharing for biomedical laboratory researchers,** BioarXiv, in process.

Developing a data management workflow for your lab

Create a standard data dictionary for routine data elements:

a. Use community CDE's if available (for preclinical TBI, they are!)
b. Make sure the required metadata for community standards is acquired
c. Consider a lab-wide data dictionary that is regularly updated (and versioned)

Consider data formats:

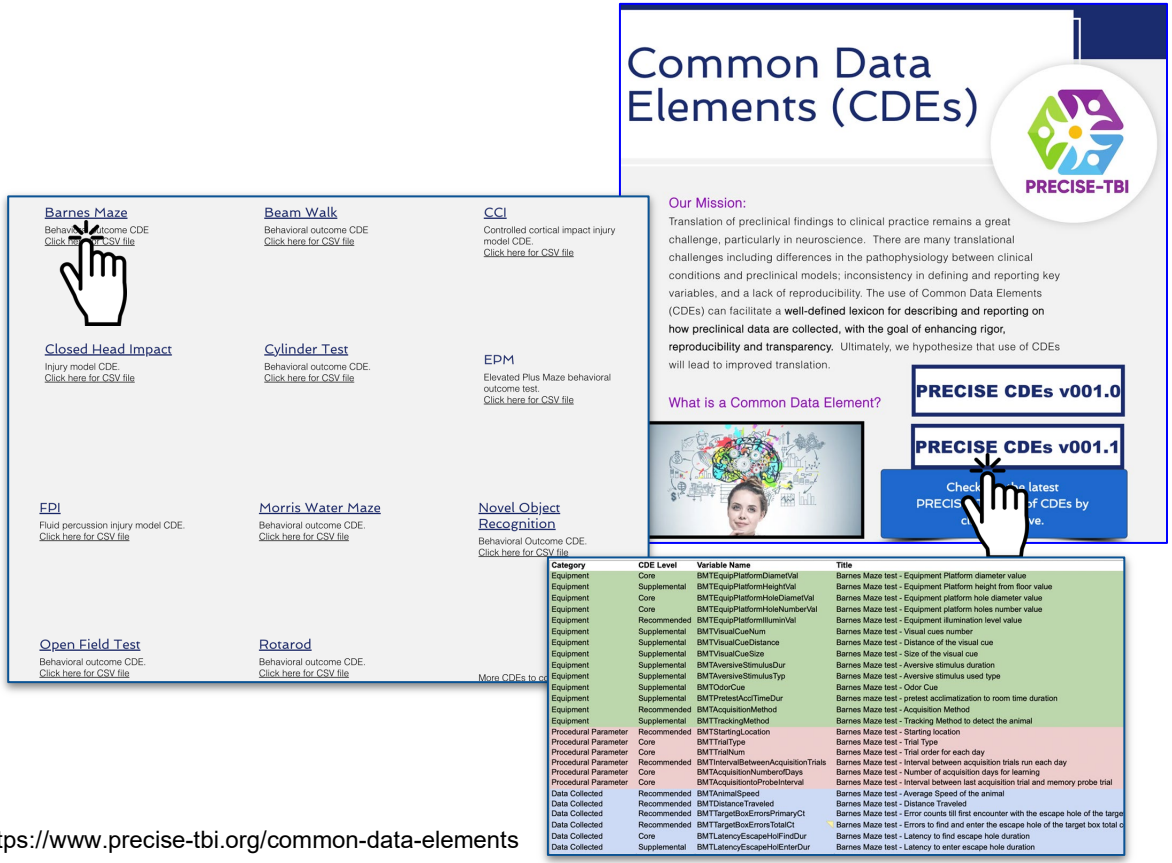
a. Does the repository have a particular data format that is required?
b. Are you storing your data in a proprietary format?
c. Are you familiar with good data formatting practices? e.g., tidy spreadsheet format

Generate a system for the unique identification of subjects and encourage single-subject data tracking:

a. Creating a central registry of subjects will help the lab learn about FAIR practices
b. Greatly helps keep track of data collection, management, and analysis at the individual subject level.
c. Provides clear provenance for all subjects within a given experiment

Community standards: Common Data Elements

- A **community standard for**
  a. What should be collected
  b. What it should be named
  c. How it should be structured
- Enable investigators to **systematically collect, analyze, and share data** across the community
- **Facilitates research** by improving data collection, analysis, harmonization and data sharing
- **CDEs for preclinical TBI** can be downloaded from the PRECISE website
- **ODC-TBI supports** CDEs for preclinical TBI

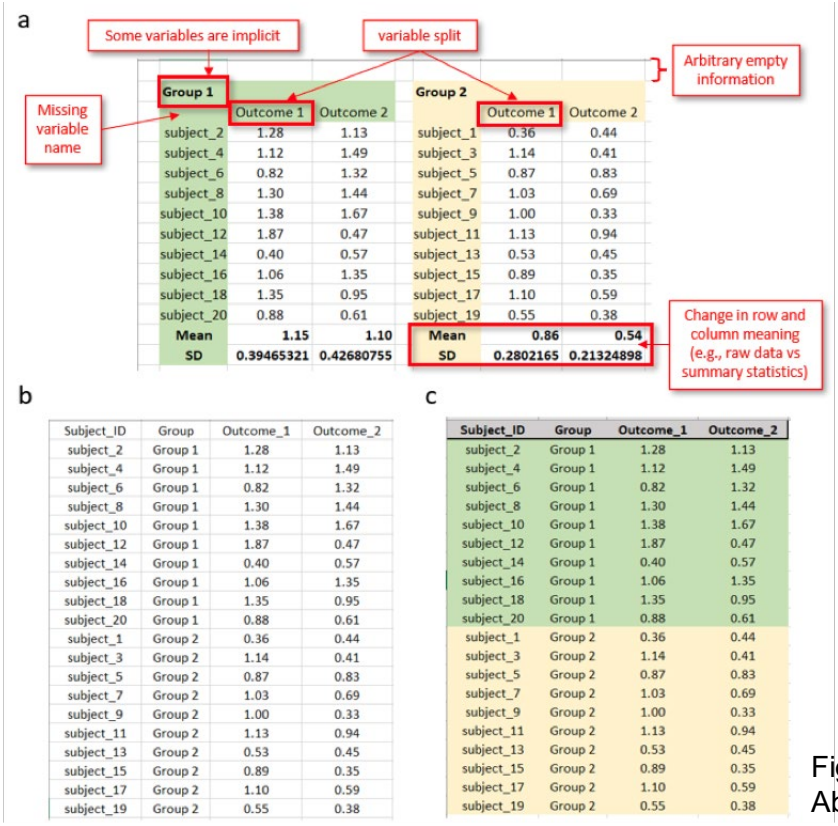


Category	CDE Level	Variable Name	Title
Equipment	Core	BMTEquipPlatformDiametVal	Barnes Maze test - Equipment Platform diameter value
Equipment	Supplemental	BMTEquipPlatformHeightVal	Barnes Maze test - Equipment Platform height from floor value
Equipment	Core	BMTEquipPlatformHoleDiametVal	Barnes Maze test - Equipment platform hole diameter value
Equipment	Core	BMTEquipPlatformHoleNumberVal	Barnes Maze test - Equipment platform holes number value
Equipment	Recommended	BMTEquipPlatformIlluminVal	Barnes Maze test - Equipment illumination level value
Equipment	Supplemental	BMTVisualCueNum	Barnes Maze test - Visual cues number
Equipment	Supplemental	BMTVisualCueDistance	Barnes Maze test - Distance of the visual cue
Equipment	Supplemental	BMTVisualCueSize	Barnes Maze test - Size of the visual cue
Equipment	Supplemental	BMTAversiveStimulusDur	Barnes Maze test - Aversive stimulus duration
Equipment	Supplemental	BMTAversiveStimulusTyp	Barnes Maze test - Aversive stimulus used type
Equipment	Supplemental	BMTOdorCue	Barnes Maze test - Odor Cue
Equipment	Supplemental	BMTPretestAcclTimeDur	Barnes maze test - pretest acclimatization to room time duration
Equipment	Recommended	BMTAcquisitionMethod	Barnes Maze test - Acquisition Method
Equipment	Supplemental	BMTTrackingMethod	Barnes Maze test - Tracking Method to detect the animal
Procedural Parameter	Recommended	BMTStartingLocation	Barnes Maze test - Starting location
Procedural Parameter	Core	BMTTrialType	Barnes Maze test - Trial Type
Procedural Parameter	Core	BMTTrialNum	Barnes Maze test - Trial order for each day
Procedural Parameter	Recommended	BMTIntervalBetweenAcquisitionTrials	Barnes Maze test - Interval between acquisition trials run each day
Procedural Parameter	Core	BMTAcquisitionNumberofDays	Barnes Maze test - Number of acquisition days for learning
Procedural Parameter	Core	BMTAcquisitiontoProbeInterval	Barnes Maze test - Interval between last acquisition trial and memory probe trial
Data Collected	Recommended	BMTAnimalSpeed	Barnes Maze test - Average Speed of the animal
Data Collected	Recommended	BMTDistanceTraveled	Barnes Maze test - Distance Traveled
Data Collected	Recommended	BMTTargetBoxErrorsPrimaryCt	Barnes Maze test - Error counts till first encounter with the escape hole of the targe
Data Collected	Recommended	BMTTargetBoxErrorsTotalCt	Barnes Maze test - Errors to find and enter the escape hole of the target box total c
Data Collected	Core	BMTLatencyEscapeHolFindDur	Barnes Maze test - Latency to find escape hole duration
Data Collected	Supplemental	BMTLatencyEscapeHolEnterDur	Barnes Maze test - Latency to enter escape hole duration

https://www.precise-tbi.org/common-data-elements

Formats: the importance of machine readability

- ODC-TBI enforces **a *tidy data format***
- FAIR requires both human and machine readability
- Some common practices make it hard for computers to read your data



	Group 1			Group 2	
	Outcome 1	Outcome 2		Outcome 1	Outcome 2
subject_2	1.28	1.13	subject_1	0.36	0.44
subject_4	1.12	1.49	subject_3	1.14	0.41
subject_6	0.82	1.32	subject_5	0.87	0.83
subject_8	1.30	1.44	subject_7	1.03	0.69
subject_10	1.38	1.67	subject_9	1.00	0.33
subject_12	1.87	0.47	subject_11	1.13	0.94
subject_14	0.40	0.57	subject_13	0.53	0.45
subject_16	1.06	1.35	subject_15	0.89	0.35
subject_18	1.35	0.95	subject_17	1.10	0.59
subject_20	0.88	0.61	subject_19	0.55	0.38
Mean	1.15	1.10	Mean	0.86	0.54
SD	0.39465321	0.42680755	SD	0.2802165	0.21324898

b

Subject_ID	Group	Outcome_1	Outcome_2
subject_2	Group 1	1.28	1.13
subject_4	Group 1	1.12	1.49
subject_6	Group 1	0.82	1.32
subject_8	Group 1	1.30	1.44
subject_10	Group 1	1.38	1.67
subject_12	Group 1	1.87	0.47
subject_14	Group 1	0.40	0.57
subject_16	Group 1	1.06	1.35
subject_18	Group 1	1.35	0.95
subject_20	Group 1	0.88	0.61
subject_1	Group 2	0.36	0.44
subject_3	Group 2	1.14	0.41
subject_5	Group 2	0.87	0.83
subject_7	Group 2	1.03	0.69
subject_9	Group 2	1.00	0.33
subject_11	Group 2	1.13	0.94
subject_13	Group 2	0.53	0.45
subject_15	Group 2	0.89	0.35
subject_17	Group 2	1.10	0.59
subject_19	Group 2	0.55	0.38

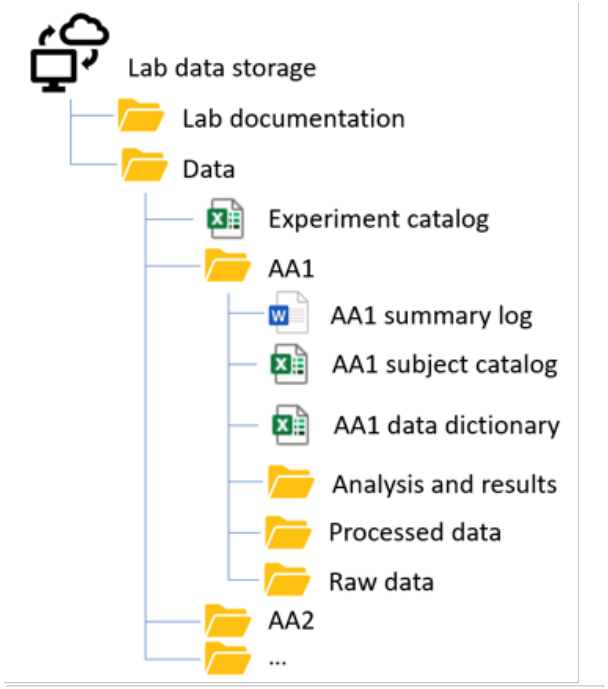
c

Subject_ID	Group	Outcome_1	Outcome_2
subject_2	Group 1	1.28	1.13
subject_4	Group 1	1.12	1.49
subject_6	Group 1	0.82	1.32
subject_8	Group 1	1.30	1.44
subject_10	Group 1	1.38	1.67
subject_12	Group 1	1.87	0.47
subject_14	Group 1	0.40	0.57
subject_16	Group 1	1.06	1.35
subject_18	Group 1	1.35	0.95
subject_20	Group 1	0.88	0.61
subject_1	Group 2	0.36	0.44
subject_3	Group 2	1.14	0.41
subject_5	Group 2	0.87	0.83
subject_7	Group 2	1.03	0.69
subject_9	Group 2	1.00	0.33
subject_11	Group 2	1.13	0.94
subject_13	Group 2	0.53	0.45
subject_15	Group 2	0.89	0.35
subject_17	Group 2	1.10	0.59
subject_19	Group 2	0.55	0.38

Figure courtesy of Dr. Abel Torres Espin

Fouad et al. (2023) **A practical guide to data management and sharing for biomedical laboratory researchers,** BioarXiv

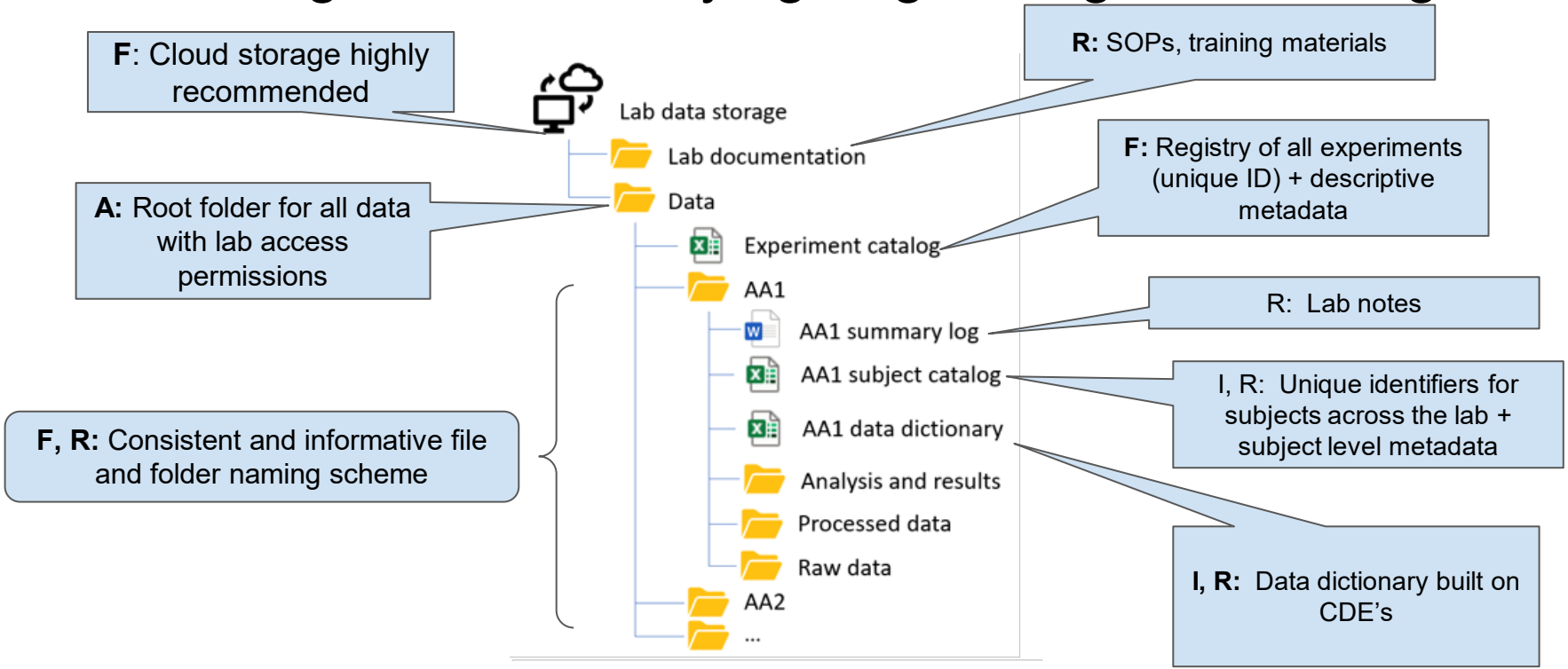
Lab management: Identifying, organizing and naming



Fouad et al. (2023) **A practical guide to data management and sharing for biomedical laboratory researchers,** BioarXiv, in process

Supplementary materials: https://zenodo.org/record/8071997

Lab management: Identifying, organizing and naming



Fouad et al. (2023) **A practical guide to data management and sharing for biomedical laboratory researchers,** , in process

Supplementary materials: https://zenodo.org/record/8071997

Questions?

Extra slides

ODC-TBI supports the FAIR data principles

- Findable:
  - Unique identifiers
  - Rich metadata describing the dataset
  - Published in a trusted repository
- Accessible:
  - Data can be accessed and is machine readable
  - Authorization and authentication as necessary
- Interoperable
  - Common vocabularies
  - Open formats
- Reusable
  - Metadata, metadata, metadata
  - Data dictionary
  - Access rights specified
  - Data collected according to community standards