DATA MANAGEMENT AND SHARING PLANS

PRESENTED BY ALLISON KAEFRING AND JAMIE BLOSS

LEARNING OBJECTIVES

•Attendees will have a basic understanding of the elements of a data management and sharing plan

•Attendees will be able to log into DMPTool and access different types of plans

•Attendees will be able to list the repository options hosted by or subscribed to by ECU

NELSON MEMO

•Scientific data underlying peer-reviewed scholarly publications resulting from federally funded research should be made freely available and publicly accessible by default at the time of publication.

https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-access-Memo.pdf

WHO NEEDS TO INCLUDE A DMSP?

All federal research funding agencies, including the National Endowment for the Humanities and National Endowment for the Arts.

All grant proposals submitted to the NIH for submission deadlines on or after January 25th, 2023, where the proposal produces scientific data.

WHY IS DATA SHARING IMPORTANT?

Replicability of studies/increased transparency

Multiple uses for one data set

- Reduces the number of projects that need funding

- Reduces strain on community of participants

- Supports data education/literacy

Helps track data preservation

MAKING DATA LEGIBLE/REPRODUCIBLE

Data should be understandable by people not a part of your research team

Example: naming conventions

DE-IDENTIFYING DATA

Legal Requirements

1. FERPA (Federal Educational Rights and Privacy Act)

2. HIPPA (Health Insurance Portability and Accountability Act)

3. North Carolina Identity Theft Act

Resources

1. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

2. https://www.ipc.on.ca/sites/default/files/legacy/2016/08/Deidentification-Guidelines-for-Structured-Data.pdf

3. https://qdr.syr.edu/guidance/human-participants/deidentification

GENERAL INFORMATION TO DE-IDENTIFY

Personally identifiable information

- Name

- SSN

- Address/ZIP

- phone number

- Picture

- voice

Indirect Identifiers

Varies by research and community being researched

Examples:

- Occupation

- Region

- Outliers

ETHICAL CONSIDERATIONS

- Is the community especially vulnerable or stigmatized and might need extra diligence in deidentifying?

- Even if data individually cannot identify individuals, can they be identified if taken in the context of the study/other data?

- If data can never be responsibly de-identified and how can you write that into your DMP

- Is the data still usable to other researchers?

VULNERABLE POPULATIONS/STIGMATIZED SUBJECTS

Insider v outsider

Is the community so small even vague information could identify someone?

Is the data de-identified to outsiders AND insiders

Quasi-identifiers

Identity revealing traits that allow linking to other (often external) data sets

ex. Voter registration is public

ex. Course ID unique combinations of courses could be linked to identifying information

HOW TO ANONYMIZE

Generalization

Create broader categories by combining categories together

- Can be less useful/reusable to other researchers

- Decrease in overall quality

- Distort correlations

Suppression

Delete/censor data that violates anonymity

- Can skew integrity of data and distort it

- Decrease in quality of specific censored sections

- Distort mean values of de-identified columns

Document what has been changed/anonymized and keep file in secure location separate from data

DMP TOOL

EXAMPLES

DMP tool Dashboard

LibGuide

https://libguides.ecu.edu/DataManagement/datamgtplans

https://library.ecu.edu/ >Research Guides> Faculty Resources> Data Management> Data Management Plans

BONES OF THE NIH DMSP

1.**Data Type:** Briefly describe the scientific data to be managed, preserved, and shared, including the data to be collected or used in the study; data to be preserved and shared; and documentation to be collected and shared with the data.

2.**Related Tools, Software and/or Code:** An indication of whether specialized tools will be needed to access or manipulate the shared scientific data to support replication or reuse, and name(s) of the needed tool(s) and software.

3.**Standards:** An indication of what standards will be applied to the scientific data and associated metadata (i.e., data formats, data dictionaries, data identifiers, definitions, unique identifiers, and other data documentation). If the discipline of the research does not have a shared approach to data structures, the Plan may indicate that no consensus data standards exist for this scientific data and metadata.

BONES OF THE NIH DMSP CONT.

4. **Data Preservation, Access, and Associated Timelines:** Plans and timelines for data preservation and access, including (a) specific repository or repositories where data will be shared (see tab on the left on repositories); what persistent identifier or other indexing tool will be provided to find and access the data; and when data will become available and for how long it will stay available.
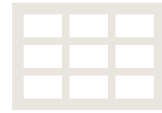
5. **Access, Distribution, or Reuse Considerations:** Any limits that will be placed on access, and why the limits are expected to be placed.

6. **Oversight of Data Management and Sharing:** Indicate how and on what schedule the DMS Plan will be monitored and managed, and by whom.

WHICH ELEMENT IS THIS?

Anyone wishing to access, display, process, or reuse the data would need access to a computer and a web browser. They would need to be able to open CSV, TXT, and PDF files. If any survey codebooks are created in a Microsoft word document (.docx), they will be converted to PDF files.

Data type?

Related tools, software and/or code?

Standards?

Data preservation, access, and associated timelines?

Access, distribution, or reuse considerations?

Oversight of data management and sharing?

WHICH ELEMENT IS THIS?

The only limitations that could restrict access to data is for privacy and confidentiality protections for patient data recorded by the FHASES system.

Data type?

Related tools, software and/or code?

Standards?

Data preservation, access, and associated timelines?

Access, distribution, or reuse considerations?

Oversight of data management and sharing?

WHICH ELEMENT IS THIS?

Following our IRB and best practices, we will protect participant identities who participate in surveys and qualitative research. We will implement protections to limit any ability to link data with individuals. We will strip identifying information from survey or interview responses as necessary.

Data type?

Related tools, software and/or code?

Standards?

Data preservation, access, and associated timelines?
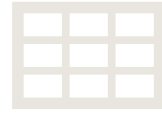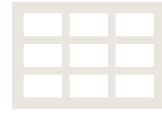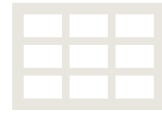
Access, distribution, or reuse considerations?

Oversight of data management and sharing?

WHICH ELEMENT IS THIS?

Co-Investigator Joseph Lee, ORCID: 0000-0001-9698-649X, will be responsible for the day-to-day oversight of team data management activities and data sharing with input and support from co-investigator Jamie E. Bloss, ORCID: 0000-0002-0706-7684. We plan to review this document every 12 months to update it and make sure that implementation is happening as the research progresses through all years of the award period.

Data type?

Related tools, software and/or code?

Standards?

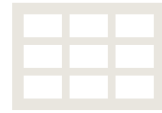Data preservation, access, and associated timelines?

Access, distribution, or reuse considerations?

Oversight of data management and sharing?

WHICH ELEMENT IS THIS?

Data will be stored in common and open formats, such as CSV and PDF files for our survey and quantitative data. We will ensure data and other products are stored in a standardized format (e.g., CSV, TXT, HTML, XML, PDF) that are expected to remain compatible with newer versions of software.

Data type?

Related tools, software and/or code?

Standards?

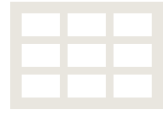Data preservation, access, and associated timelines?

Access, distribution, or reuse considerations?

Oversight of data management and sharing?

WHICH ELEMENT IS THIS?

We will keep all generated data files in two different locations, one of which is ECU's Microsoft OneDrive which the entire grant team has access to, which is approved for storage of research data. Copies will additionally be held locally at ECU on a secured server that is administered by ECU ITCS.

Data type?

Related tools, software and/or code?

Standards?

Data preservation, access, and associated timelines?

Access, distribution, or reuse considerations?

Oversight of data management and sharing?

DATA REPOSITORY OPTIONS

4. Data Preservation, Access, and Associated Timelines: Plans and timelines for data preservation and access, including (a) specific repository or repositories where data will be shared (see tab on the left on repositories); what persistent identifier or other indexing tool will be provided to find and access the data; and when data will become available and for how long it will stay available.

[Open Science Framework](https://osf.io) OSF allows you to have both public and private versions of your study. Once you have de-identified data in an OSF-hosted study, you can register the study and share it for your data sharing compliance. OSF offers DOIs and unique persistent URLs.

Dataverse*: https://dataverse.unc.edu/dataverse/ECU Open access datasets repository based on DataVerse. Creates a DOI.

ECU ScholarShip: https://thescholarship.ecu.edu/page/deposit Makes a permanent link but not a DOI.

https://libguides.ecu.edu/DataManagement/repositories

RESOURCES LIST

DMPTool: https://dmptool.org/

Open Science Framework: https://osf.io/

Dataverse: https://dataverse.unc.edu/dataverse/ECU

Updates: https://open.science.gov/

LibGuide: https://libguides.ecu.edu/DataManagement

PLEASE FILL OUT THIS EVALUATION AND ASSESSMENT

https://ecu.az1.qualtrics.com/jfe/form/SV_ePOyzXsDp6rtn9k

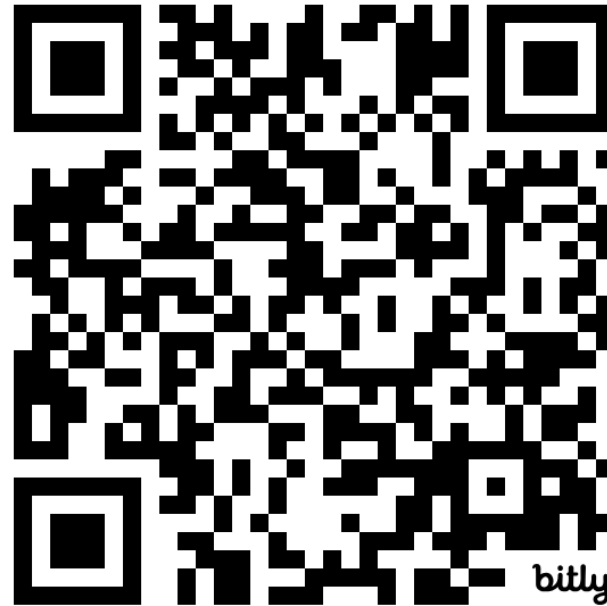**Contact us for more info and assistance:**

Jamie Bloss - health sciences campus

Blossj19@ecu.edu

Allison Kaefring - main campus

kaefringa22@ecu.edu

QUESTIONS?