# Managing Data From Your Research

**David Durden**

Data Services Librarian

👍

**Grab the Slides and Follow Along**

**https://go.umd.edu/rc181016**

🔍 **The Data Problem**

# 🔍 Where Does All the Data Go?

> ...for every yearly increase in article age, the odds of the data set being extant decreased by 17%.

# 🏛️ Why Data Management?

1. Money

2. Ethics & transparency

3. Reproducibility & accountability

4. Re-use

5. Good organizational practice

6. For the public good

# 🔀 Levels of Data Management

1. Administrative (policy)

  ○ Data Use Agreements

  ○ Data Management Plans

  ○ Data Safety and Monitoring Plans

2. Applied (procedural)

  ○ Workflows

  ○ data management plans

  ○ Documentation

# 📊 What Qualifies as "Research Data"?

# 📊 Research Data Are:

> Data that are used as *primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity*, and that are used as *evidence in the research process* and/or are commonly accepted in the research community as necessary to validate research findings and results. All other digital and non-digital content have the potential of becoming research data. *Research data may be experimental data, observational data, operational data, third party data, public sector data, monitoring data, processed data, or repurposed data.*
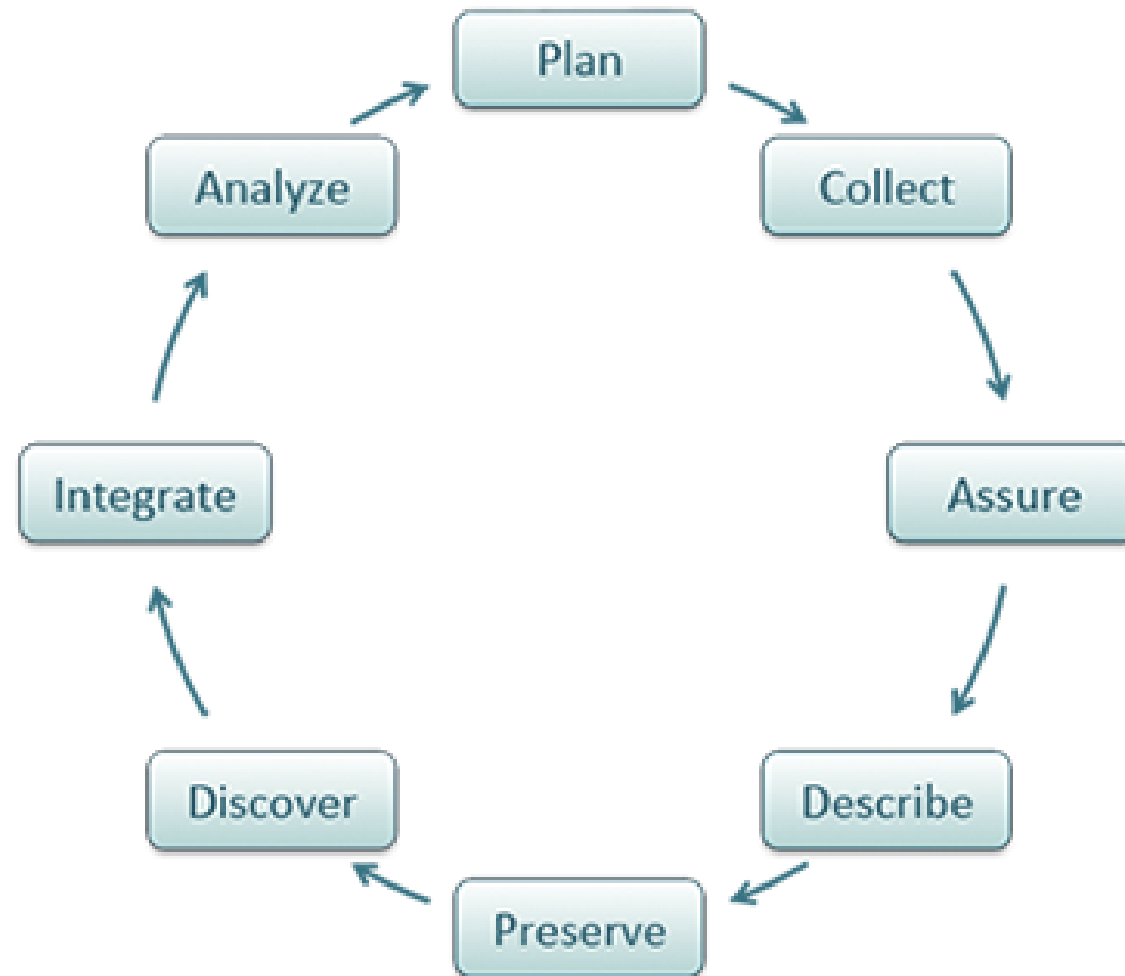
# 📊 What is "Research Data"?

- drafts

- bibliographies

- oral histories & A/V materials

- code/scripts

- survey instruments

- "data"

- description of methodologies

- etc.

# ♻ DataONE Lifecycle Model

# ♻️ DataONE Lifecycle Model

1. *Plan:* description of the data that will be compiled, and how the data will be managed and made accessible throughout its lifetime

2. *Collect:* observations are made either by hand or with sensors or other instruments and the data are placed a into digital form

3. *Assure:* the quality of the data are assured through checks and inspections

4. *Describe:* data are accurately and thoroughly described using the appropriate metadata standards

# ♻️ DataONE Lifecycle Model

5. *Preserve:* data are submitted to an appropriate long-term archive (i.e. data center)

6. *Discover:* potentially useful data are located and obtained, along with the relevant information about the data (metadata)

7. *Integrate:* data from disparate sources are combined to form one homogeneous set of data that can be readily analyzed

8. *Analyze:* data are analyzed

# 📄 The Data Management Plan

"*When things go wrong, they do so in the manner that yields the most difficulty.*"

   – The Principle of Maximum Inconvenience

## 💰 NSF DMP Requirements:

> Proposals submitted... must include a supplementary document of no more than two pages labeled "Data Managament Plan." This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results.

*Note:* **NSF DMP requirements vary by directorate. In fact, DMP requirements are hardly standardized across institutions and agencies.**

# 💡 Institutions That Often Require DMPs

- The Sloan Foundation

- Institute of Museum and Library Services (IMLS)

- Institute of Education Sciences (IES)

- DoD, DoE

- U.S. Geological Survey

- NASA, NOAA

- USDA

- NSF, NIH

# ⚙ The Basics of a DMP

A good DMP answers the questions:

- who,
- what,
- where,
- when,
- why,
- and how

# 📝 Write your DMP with DMPTool



Login at https://dmptool.org

# 📄 A General Description of Your Data

- how big will the data be?

- how fast will the data grow?

- what are the likely file formats for the data?

- how unique is the data?

- what is the source of the data?

- who owns the data?

# 📄 Storage and Backup

- how and where are the data stored?

- what is your backup schedule?

- who manages storage and backups?

- cost?

# 💿 Backup and Storage Strategies

*The level of redundancy in your storage media should reflect the value of your data.*

*Lots of copies keep stuff safe.*

*The cloud is just someone else's computer...* (Always read the terms of service for any cloud storage solution.)

# 💿 Backup and Storage Strategies

**The 3-2-1 backup rule:**

1. At least **three** copies,

2. in **two** different formats,

3. with **one** of those copies off-site.

# 📄 Security

*What security policies apply to your data?*

- UMD (e.g., data retention)

- IRB (human subjects and ethics)

- HIPAA (digital health records)

- FERPA (student and educational records)

- ITAR (DoD and other federal organizations)

- other ethical/legal concerns (at-risk populations, protected species, trade secrets, etc.)

# 📄 Post-project Data Management

- how will you store and backup your older data?
- how long will you maintain data after the project is completed (e.g., institutional data retention policies, state or federal policies)?
- will you migrate your storage media over time?
- how will you prepare data for long-term storage/preservation?
- will you use a third party to preserve your data? e.g., a data repository
- what happens to data if you leave your current institution?

# 📄 Data Sharing and Access

*One of the most important sections for publicly funded research.*

- what data will be shared and in what forms?

- who is the audience for your data?

- when and where will you share?

- how will the data be prepared for sharing?

- who is responsible for making the data available and/or answering question about access?

- how much will this cost?

# 📁 Example Folder Structure

Project_1

    Data

        AnalyzedData

        RawData

            20171201

            20171215

                Mterrapin-observ-1

    DataManagementPlan

    Documentation

    Draft

    LitReview

    Scripts

# 📄 File Naming Conventions

*The most granular level of metadata.*

**Use a combination of:**

- experiment type/number
- researcher name
- sample type/number
- analysis type
- date
- site/lab/location
- version number

# 📄 File Naming Conventions

*There is **no** absolute way to name your files.*

**Names should be:**

- descriptive

- unique

- consistent

- relatively short

# 📄 Some File Naming Conventions

**Avoid:**

- spaces – use dashes or underscores (use **CamelCase** or **pot_hole_case**)

- special characters / \:( )*?'< >[ ]&$@!"%^

- localized date formats – use the ISO 8601 date convention (YYYYMMDD or YYYY-MM-DD)

- using commas or periods in file names

# 💡 Data Management Wrap-up

# 💡 Data Management Wrap-up

- analyze your workflows

- create systems that work for you, and stick to them

- organize and describe your data and processes in way that is understandable and repeatable

- don't be afraid to change or make improvements

# 💡 Data Management Wrap-up

- use a data management plan as part of project planning

- omit DMP sections that won't apply

- don't reinvent the wheel – use existing policies and/or DMPs

- look out for 'future you'

31

# 🔧 Resources

- Local resources
    - https://lib.umd.edu/data

- RDM Education
    - https://dataone.org/education

    - https://mantra.edina.ac.uk/libtraining.html

- Data Management Plan Templates
    - https://dmptool.org/

# 💾 Thanks!

💻 https://lib.umd.edu/data

✉ lib-research-data@umd.edu

✉ durden@umd.edu