Creating Data Management Plans with DMPTool

David Durden

Data Services Librarian

University of Maryland Libraries

September 11, 2018

📌 Introduction

1. what is data management?

2. what is a data management plan?

3. why is a data management plan necessary?

4. how to make a data management plan in DMPTool

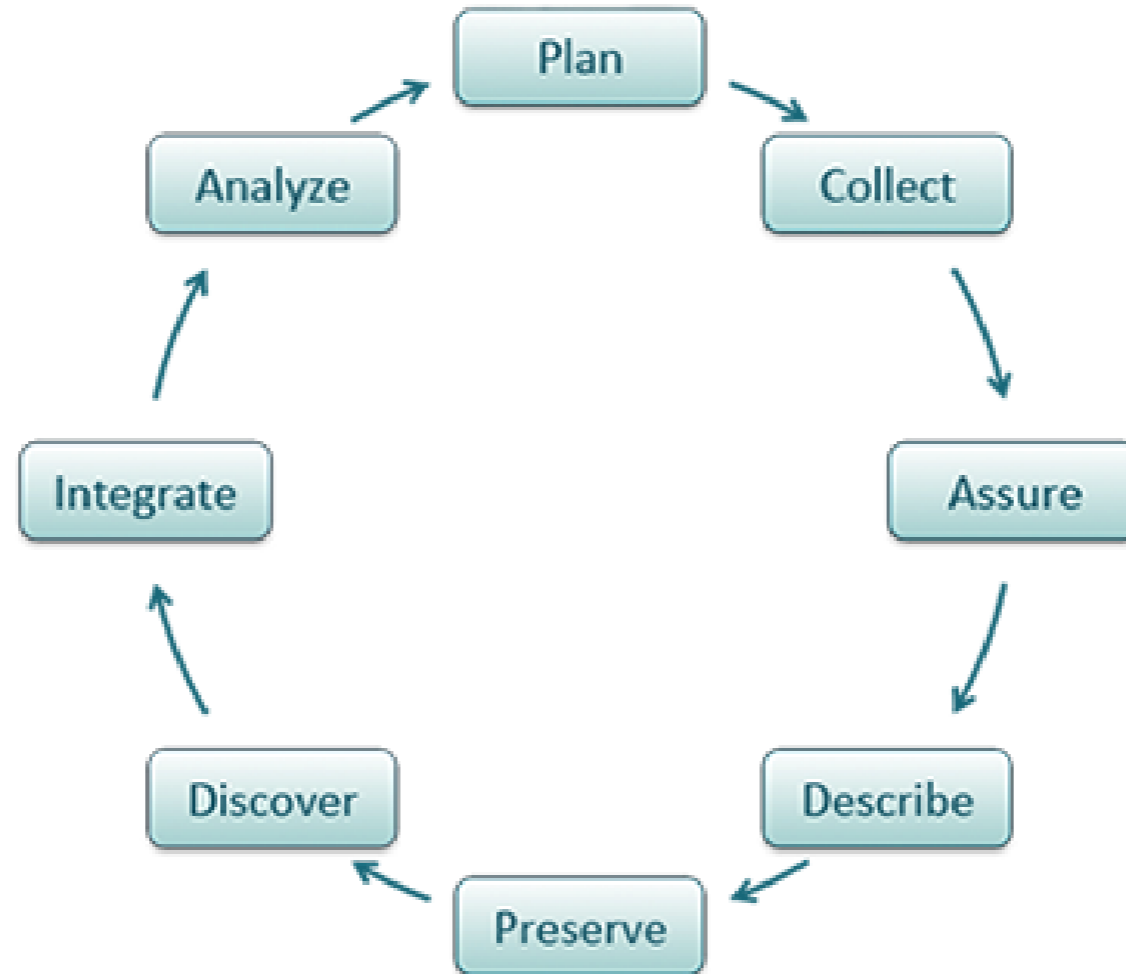🏛️ **What is data management?**

🏛️ What is data management?

1. *ensuring* physical integrity of data files

2. *ensuring* safety of content

3. *describing* the data (metadata) and recording its history (provenance)

4. *providing/enabling* appropriate access controls

5. *transferring* custody of data; *destroying* data when necessary

📊 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.*

♻️ A data lifecycle

6

♻ A data management lifecycle
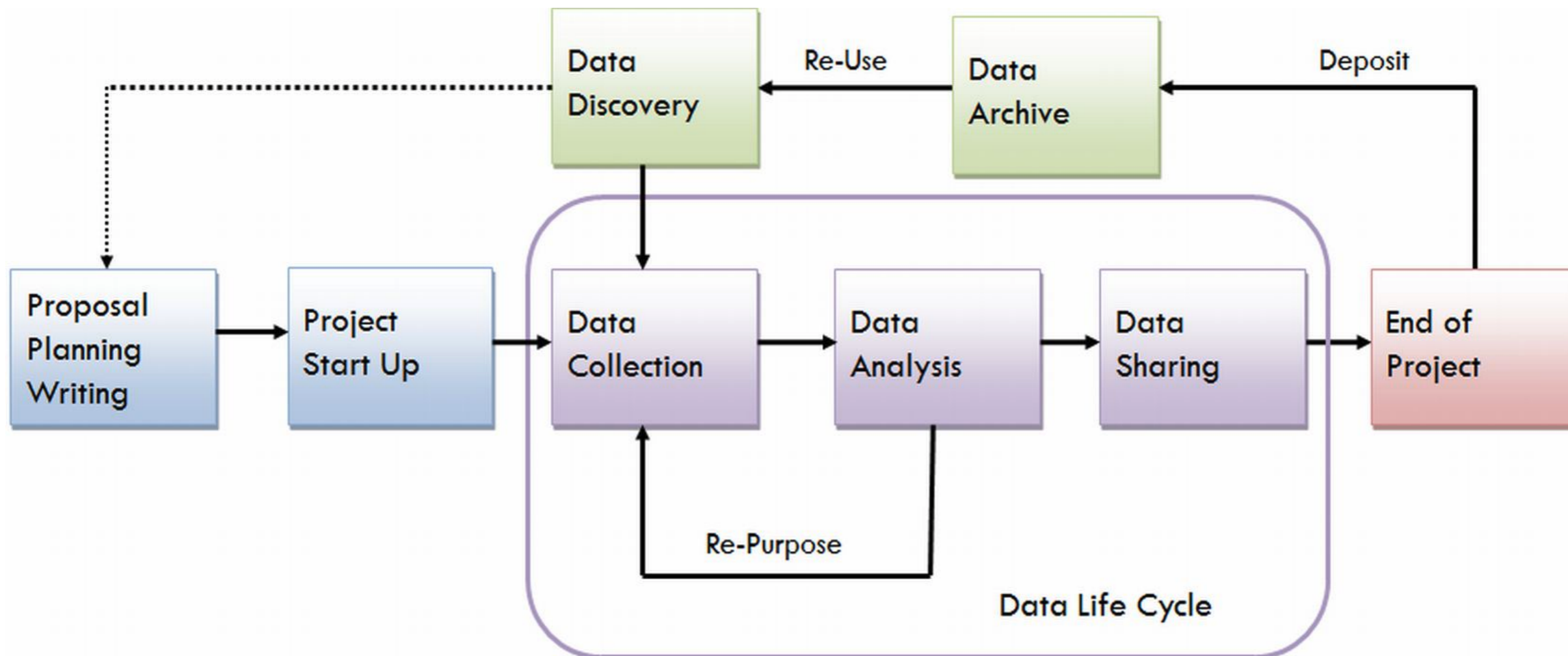
Some data management topics

1. choosing formats

2. organizing files and naming conventions

3. version control

4. access control & security

5. backup & storage

6. format conversions

7. project documentation

8. sharing and preservation plans

📄 What is a data management plan?

❓ What's a data management plan?

- a description of how data are managed throughout the lifecycle of a research project
- description of how data will comply with sharing and access mandates

🔀 Eamples of data management documents

1. Administrative -- high-level

- data use agreements

- data management plans (for grant applications)

- data safety and monitoring plans

2. Applied -- procedural

- workflows

- data management plans (for project operations)

- documentation

💀 The DMP skeleton

Describe:

1. types of data, samples, physical collections, software, curriculum materials produced by the research or project

2. standards and formats to be used for data, metadata, and content

3. access and sharing policies -- provisions for appropriate privacy, confidentiality, security, and intellectual property

4. policies and provisions for re-use, redistribution, and the production of derrivatives

5. plans for archiving, preservation, and access to data, samples, and other research products

📄 Data management plans: Asking the right questions

📄 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?

- scope and content?

📄 Organization and documentation

- how will you document your data?

- what metadata standards/formats will you use?

- how are your data organized

📄 Storage and backup

- how and where are the data stored?

- what is your backup schedule?

- who manages data storage?

- what will data storage cost?

📄 Security

- what security measures will you use?

- who has access to data and data stores?

- who manages the security systems?

- where is PII stored?

📄 Security: *What 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: *The important section 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?

💰 Why is a data management plan necessary?

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

    – The principle of maximum inconvenience

💡 Institutions that require DMPs*

- The Sloan Foundation

- Mellon Foundation

- Institute of Museum and Library Services (IMLS)

- Institute of Education Sciences (IES)

- Dept. of Defense, Dept. of Energy

- U.S. Geological Survey

- NASA, NOAA

- USDA

- NSF, NIH

*Not a comprehensive list

💻 **Overview of DMPTool**

Log in at https://dmptool.org

⏳ **DMPTool exercise**

📝 Intructions

Working in teams or pairs:

1. identify a funding agency

2. use the funder's template to create a DMP

3. read through guidance provided by the template

4. take note of issues/questions for discussion

💬 **Discussion**

🔧 Resources

Library resources

https://lib.umd.edu/data

DIY Research Data Management Tutorials

https://dataone.org/education

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

Texts

Briney, K. (2015). *Data management for researchers*. Exeter: Pelagic Publishing, UK.

💾 Thanks!

💻 https://lib.umd.edu/data

✉ lib-research-data@umd.edu

✉ durden@umd.edu

👏 Credits

Parts of this slidedeck are remixed from a similar slidedeck created by Sherry Lake, University of Viriginia Libraries and from the useful book, *Data management for researchers*, by Kristin Briney.

The End.