Research Data Management

Jo Klein

GIS & Data Visualization Librarian

Objectives

1. What is data management?

2. Why is data management important?

3. How does one put data management into practice?

What is data management?

Types of research data

- Primary vs. secondary

- Observational

- Experimental

- Simulation

- Derived/Compiled

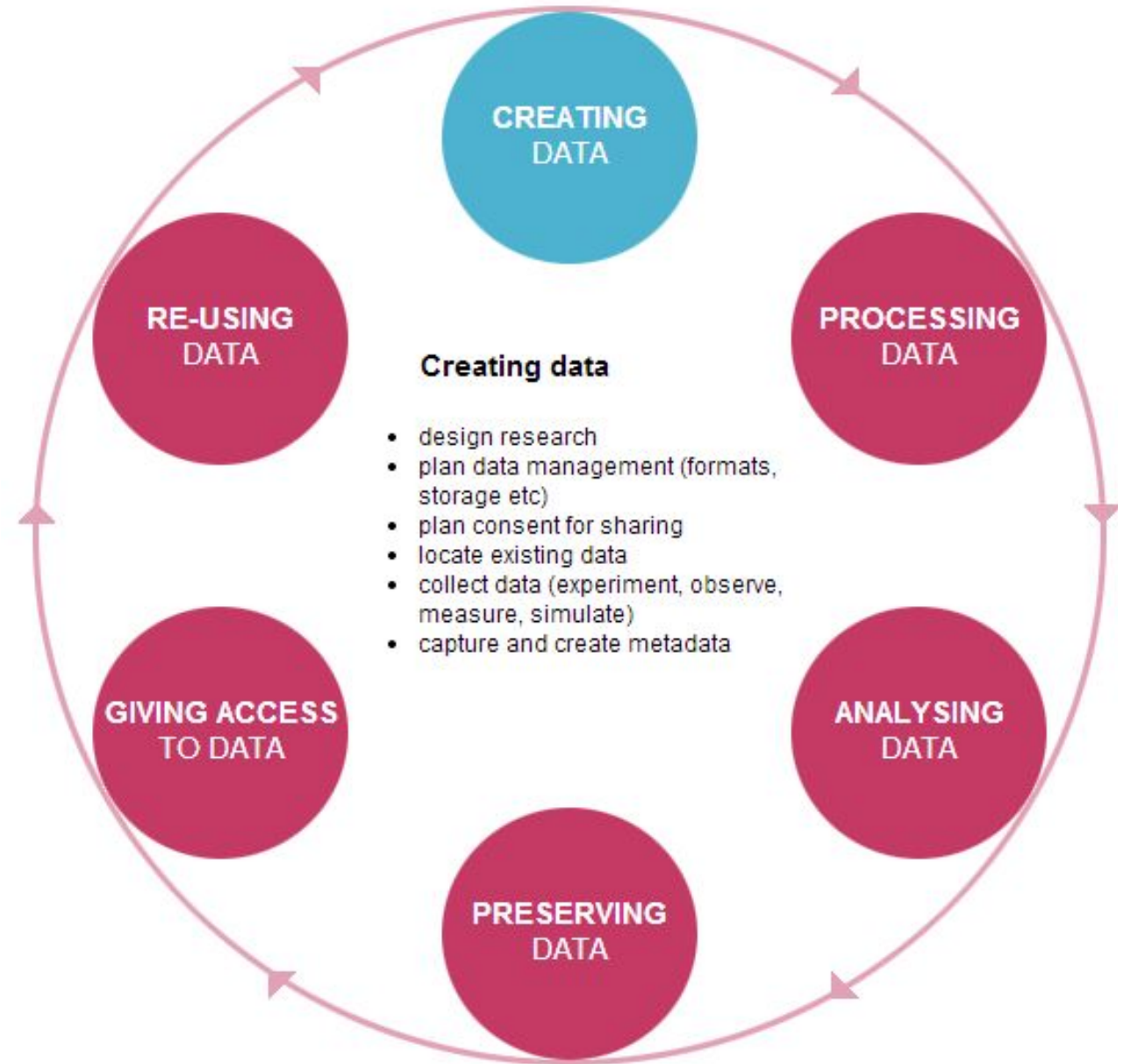**What do you work with?**
go.uncg.edu/rdm21-1

Research data includes...

- Text or Word documents
- Laboratory notebooks, field notebooks, diaries
- Questionnaires, transcripts, codebooks
- Audiotapes, videotapes
- Digital objects
- Data files and spreadsheets
- Application input, output, and log files for analysis or simulation software
- Methodologies and workflows
- Standard operating procedures and protocols
- Quality assurance/control records

Research data management is the practice of ensuring the **integrity** of data through careful consideration of:

- Ownership
- Collection
- Storage
- Sharing

(Steneck 2007, 87)

CREATING DATA

RE-USING DATA

PROCESSING DATA

GIVING ACCESS TO DATA

ANALYSING DATA

PRESERVING DATA

Creating data

- design research
- plan data management (formats, storage etc)
- plan consent for sharing
- locate existing data
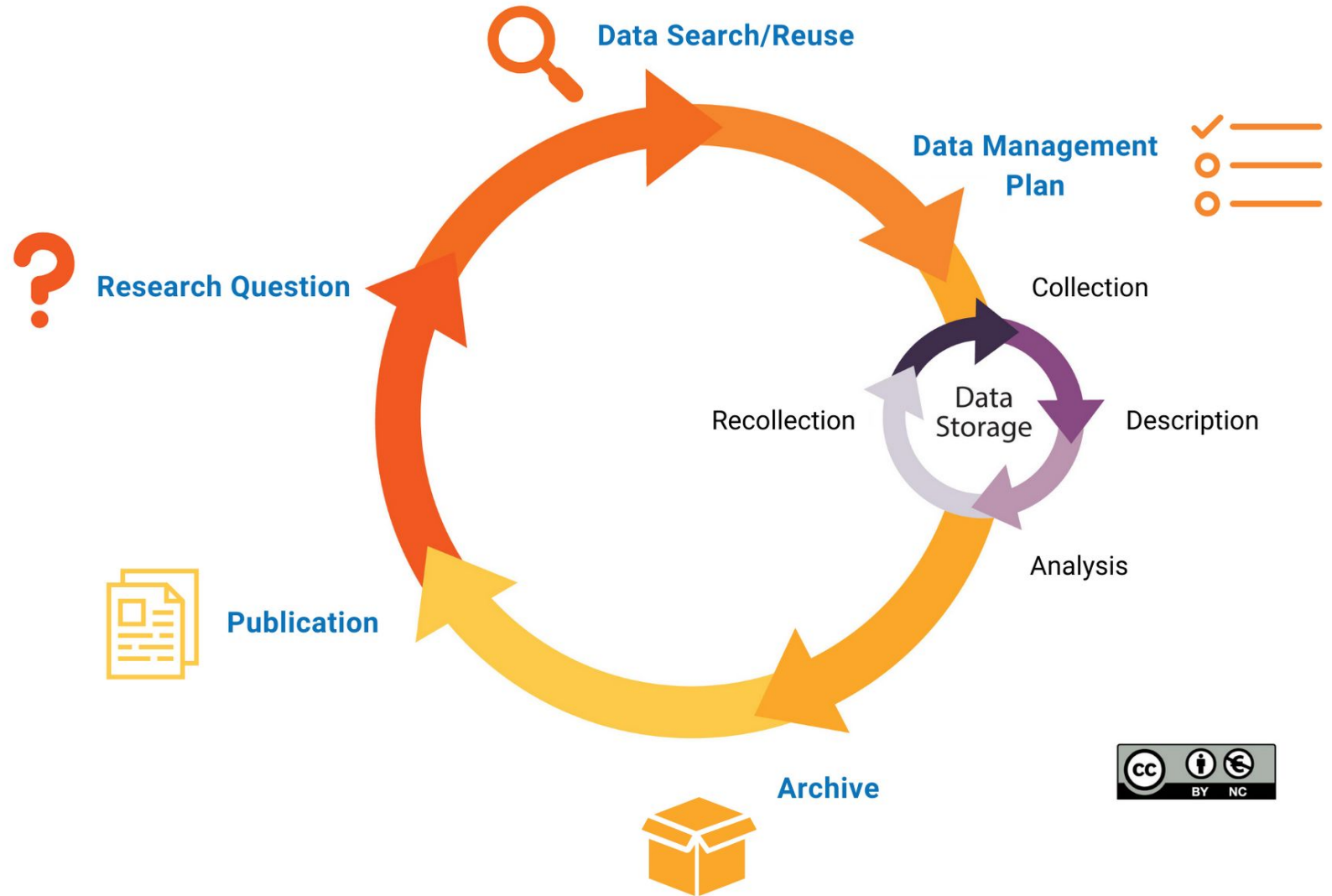- collect data (experiment, observe, measure, simulate)
- capture and create metadata

What do you do with your data at each of these steps?

What is missing?

What is extra?

What do you want to "shore up"?

The Research Data Management Lifecycle

Data Search/Reuse

Data Management Plan

Collection

Research Question

Recollection

Data Storage

Description

Analysis

Publication

Archive

Adapted from the Research Data Management Lifecycle (diagram), Data Management LibGuide, University of California, Santa Cruz, viewed 26th July 2019 <http://guides.library.ucsc.edu/datamanagement>

e Cycle of Scholarly Research

Research lifecycle

DATA SHARING & ARCHIVE · DATA MANAGEMENT PLANNING · DATABACKUP & SECURITY

Data Management Planning Checklist · University and Funder policy · Content of Data Management Plan · Ethical and Legal Considerations · Organisation and Documentation · Data Backup · Data Security · To share or not to share? · Deposit your data

Training, Advice and Support

Research Life Cycle

Re-use · Planning · Implementation · Publishing · Discovery & Impact · Preservation

Data Creation · Data Reuse · Data Access · Data Preservation

ous-Wide Research Lifecycle ...
rg

Library, The University of York
york.ac.uk

Print Version of the Data Model fo...
databaseanswers.org

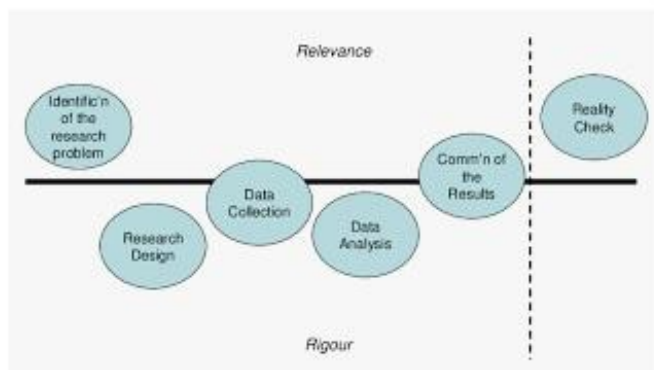Research Life Cycle" image from UC ...
pinterest.com

Research data lifecy...
blogs.ntu.edu.sg

RESEARCH LIFECYCLE — CONCEPT & PLANNING · EXPLORING & ORGANISING · COLLECTION & ANALYSIS · OUTPUTS & PUBLICATIONS · IMPACT & MANAGEMENT

Relevance / Rigour — Identific'n of the research problem · Research Design · Data Collection · Data Analysis · Comm'n of the Results · Reality Check

THE RESEARCH LIFECYCLE

Ideas · Project proposal · Results · Research · Research Data Lifecycle

Research Lifecycle — Create Data · Process Data · Analyse Data · Preserve Data · Publish Data · Reuse Data

Data man... done at al... the resear... *each step in t... own best pract...

searchers - Library
h.edu

b: Rigour and Relevance in the Extended ...
researchgate.net

Notre Dame Research FAQ...
research.nd.edu

Research Life Cycle | Göttingen ...
eresearch.uni-goettingen.de

Home - Data Management Planning ...
guides.nyu.edu

he Research Data Management Lifecycle — Data Search / Reuse · Data Management Plan · Collection · Description · Data Storage · Re-collection · Analysis · Archive · Publication

CREATING DATA · PROCESSING DATA · ANALYSING DATA · PRESERVING DATA · GIVING ACCESS TO DATA · RE-USING DATA · Creating data

intermediate research · final research · proposal · conception · post project

Completion · Conceptualization · Literature/Material Review · Research · Write Up · Assessment · Project Review · Publication/Dissemination

Data Discovery · Re-Use · Data Archive · Proposal Planning Writing · Project Start Up · Data Collection · Data Analysis · Re-Purpose · Data Life C...
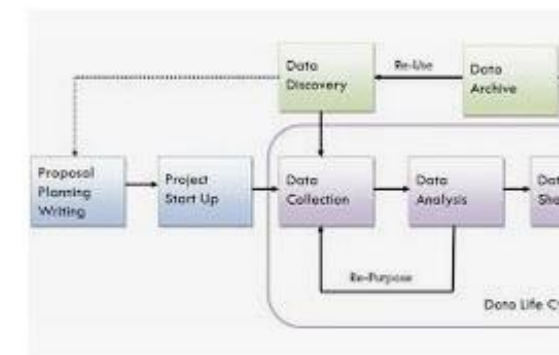
Life Cycles | NNLM

Research data lifecycle - NU ...

When to Engage | What is Public ...
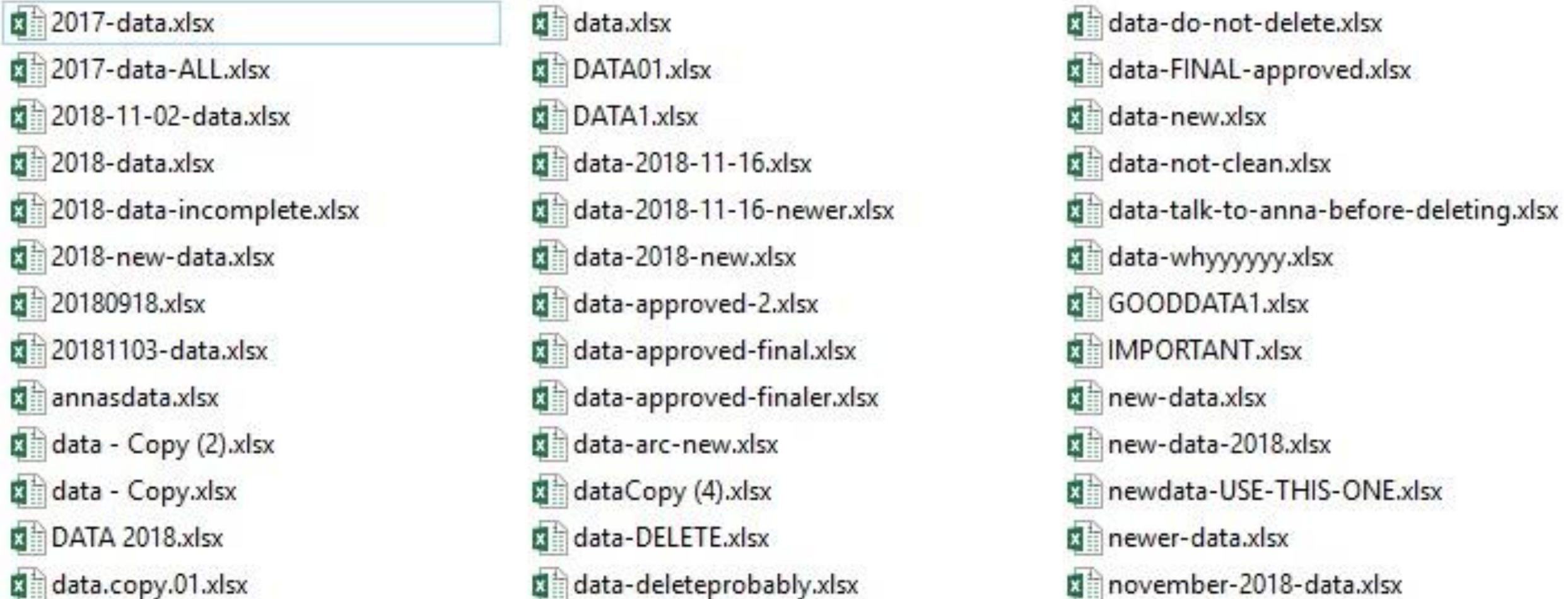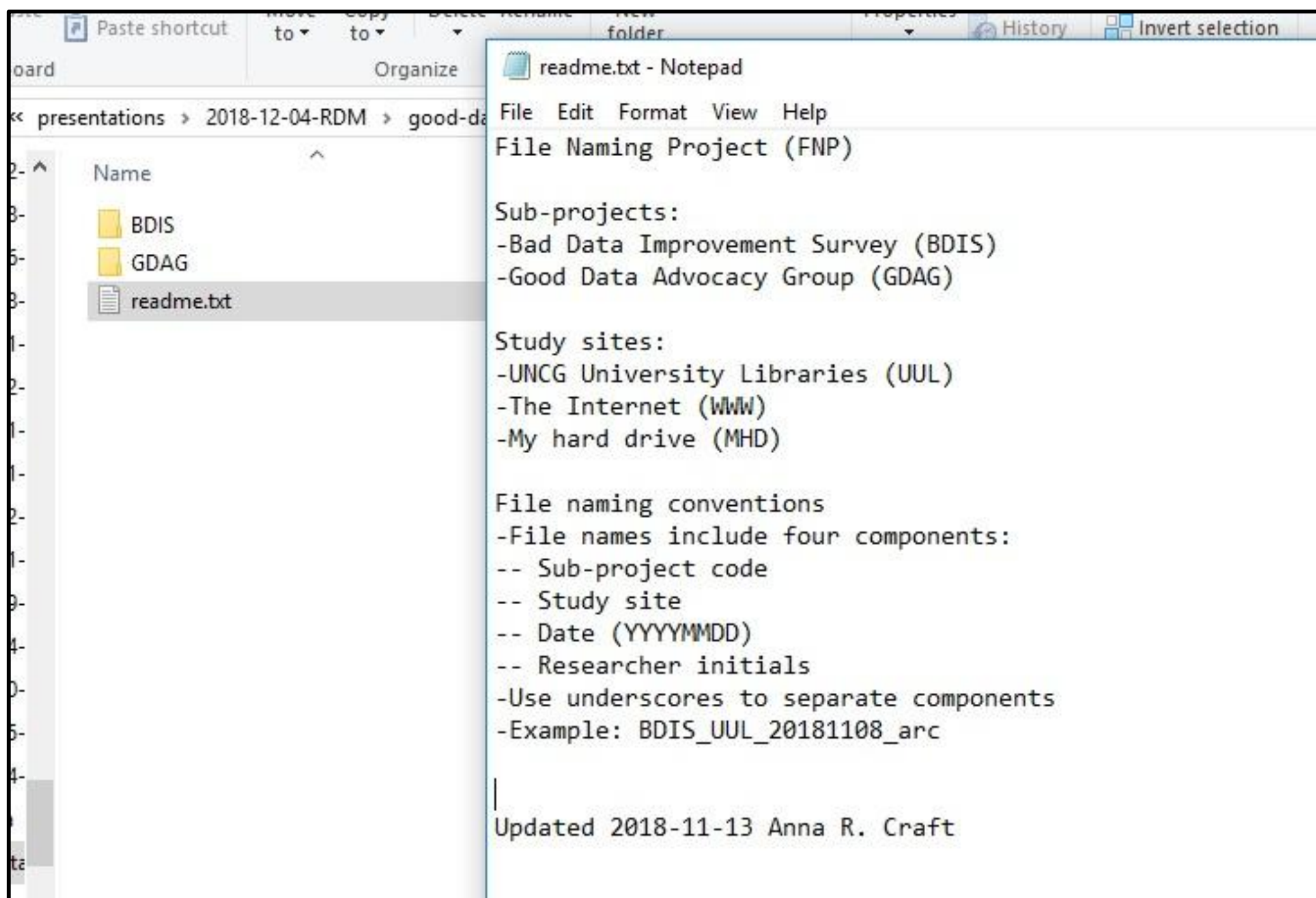
History Data Management Lifecycle (HDML ...

Steps in the Data Life Cycle ...

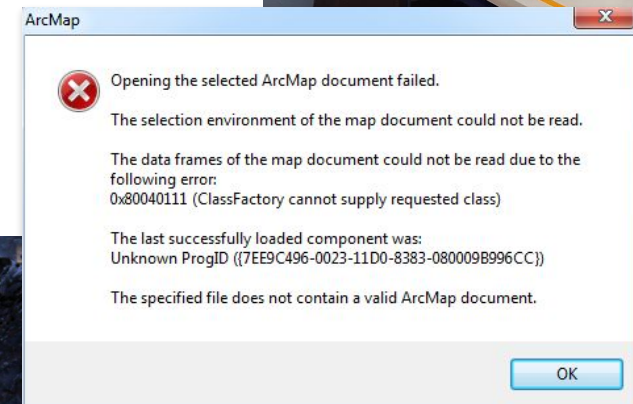# Why is data management important?

Which file do you need:

2017-data.xlsx	data.xlsx	data-do-not-delete.xlsx
2017-data-ALL.xlsx	DATA01.xlsx	data-FINAL-approved.xlsx
2018-11-02-data.xlsx	DATA1.xlsx	data-new.xlsx
2018-data.xlsx	data-2018-11-16.xlsx	data-not-clean.xlsx
2018-data-incomplete.xlsx	data-2018-11-16-newer.xlsx	data-talk-to-anna-before-deleting.xlsx
2018-new-data.xlsx	data-2018-new.xlsx	data-whyyyyyy.xlsx
20180918.xlsx	data-approved-2.xlsx	GOODDATA1.xlsx
20181103-data.xlsx	data-approved-final.xlsx	IMPORTANT.xlsx
annasdata.xlsx	data-approved-finaler.xlsx	new-data.xlsx
data - Copy (2).xlsx	data-arc-new.xlsx	new-data-2018.xlsx
data - Copy.xlsx	dataCopy (4).xlsx	newdata-USE-THIS-ONE.xlsx
DATA 2018.xlsx	data-DELETE.xlsx	newer-data.xlsx
data.copy.01.xlsx	data-deleteprobably.xlsx	november-2018-data.xlsx

Window chrome / toolbar:

Paste shortcut | Move to ▾ Copy to ▾ Delete ▾ Rename New folder | Properties ▾ History Invert selection

oard | Organize

« presentations › 2018-12-04-RDM › good-da

Name

- BDIS
- GDAG
- readme.txt

---

readme.txt - Notepad

File Edit Format View Help

File Naming Project (FNP)

Sub-projects:
-Bad Data Improvement Survey (BDIS)
-Good Data Advocacy Group (GDAG)

Study sites:
-UNCG University Libraries (UUL)
-The Internet (WWW)
-My hard drive (MHD)

File naming conventions
-File names include four components:
-- Sub-project code
-- Study site
-- Date (YYYYMMDD)
-- Researcher initials
-Use underscores to separate components
-Example: BDIS_UUL_20181108_arc

Updated 2018-11-13 Anna R. Craft

Data management...

Prevents data loss through:

- Loss of staffing competencies
- Loss of institutional commitment
- Changes in user expectations
- Natural disasters
- Storage failure
- Format obsolescence
- Human error
- Malicious attack

ArcMap

Opening the selected ArcMap document failed.

The selection environment of the map document could not be read.

The data frames of the map document could not be read due to the following error:
0x80040111 (ClassFactory cannot supply requested class)

The last successfully loaded component was:
Unknown ProgID ({7EE9C496-0023-11D0-8383-080009B996CC})

The specified file does not contain a valid ArcMap document.

OK

CC image by Sharyn Morrow on Flickr

Data management...

Facilitates publication, sharing, and preservation, which:

- Increases the impact and visibility of research
- Promotes innovation and potential new data uses
- Leads to new collaborations between data users and creators
- Maximizes transparency and accountability
- Encourages improvement and validation of research methods
- Reduces cost of duplicating or re-collecting data
- Provides important resources for education and training

How do I "do" data management?

The Research Data Management Lifecycle
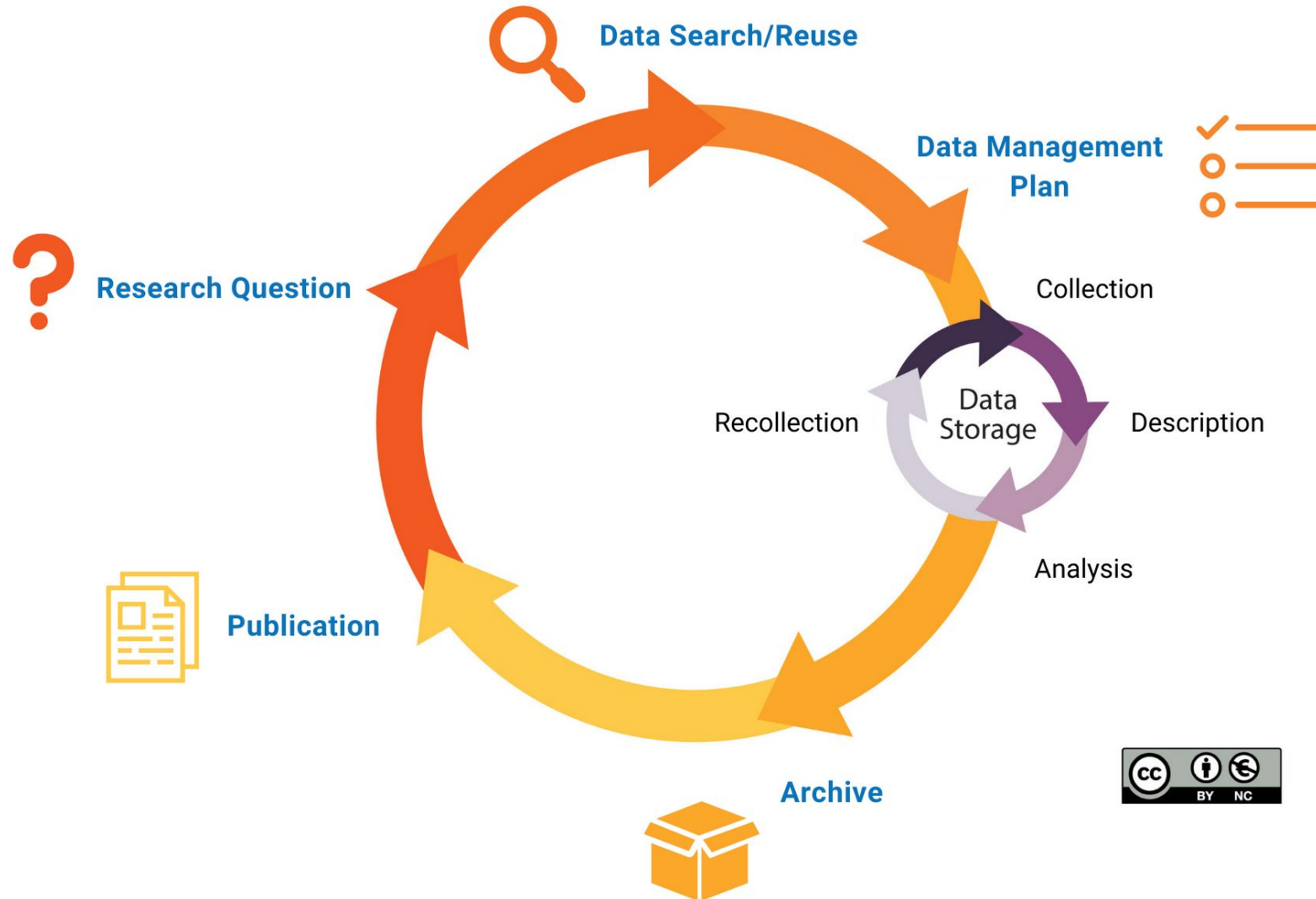
Data Search/Reuse

Data Management Plan

Collection

Description

Data Storage

Recollection

Analysis

Research Question

Publication

Archive

Adapted from the Research Data Management Lifecycle (diagram), Data Management LibGuide, University of California, Santa Cruz, viewed 26th July 2019 <http://guides.library.ucsc.edu/datamanagement>

Data management plan (DMP)

- Formal, written document

- Outlines what you will do with your data during and after research

- Helps to ensure your data are secure and available for immediate use and future re-use

- Required by most funding agencies before or after research

Components of a DMP

- Information about all data and data formats ("Data Inventory")

- Metadata content and format

- Policies for access, sharing, and re-use

- Long-term preservation or archiving

- Budget considerations

Data Inventory

- Description of data to be produced

- How will data be acquired?

- How will data be processed?

- What are the file formats?

- How do you ensure quality assurance and control during collection?

- What existing data will you use?

- How will data be managed in the short term?

Metadata content and format

- What metadata are needed to make your data understandable in the long-term?

- How will the metadata be created?

- What format or standards will be used for the metadata?
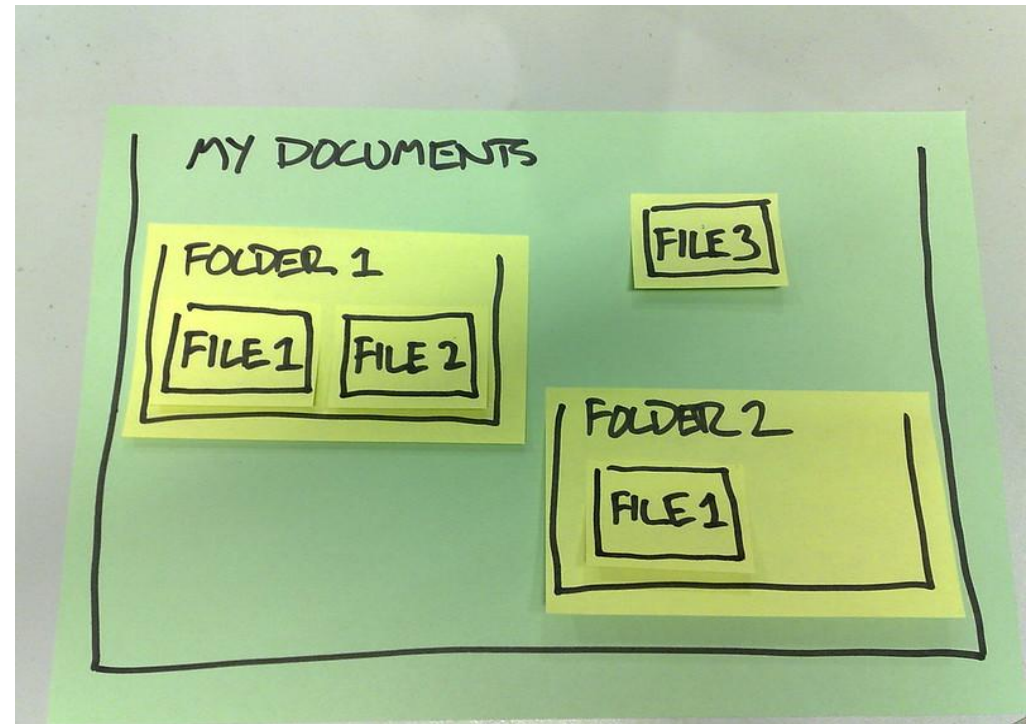
Photo: Flickr user Pat Joyce: (CC BY-NC 2.0)

Policies for access, sharing, and re-use

- How will the data be shared?

- Are there ethical or privacy issues with sharing the data?

- Are there existing or potential intellectual property & copyright issues?

- What are potential future uses for the data?

- Who will be able to access and use the data?

- Will users need to submit a use agreement?

- What is a preferred citation for the data?

Long-term preservation or archiving

- What data will be preserved for the long term?

- How long will the data need to be kept and made available?

- What data formats (& transformations) are needed to archive?

- Who will be responsible for archiving & maintaining the data?

Budget considerations

- What are the anticipated costs for short-term storage, maintenance, and long term archiving?

- Will you need to pay for training, equipment, and/or staff?

- Will you need to pay for data preparation or cleaning?

- Who will pay for those costs?

Photo: Flickr user Tristan Martin

go.uncg.edu/rdm21-2

What would a DMP help you with the most?

What do you want to "shore up"?
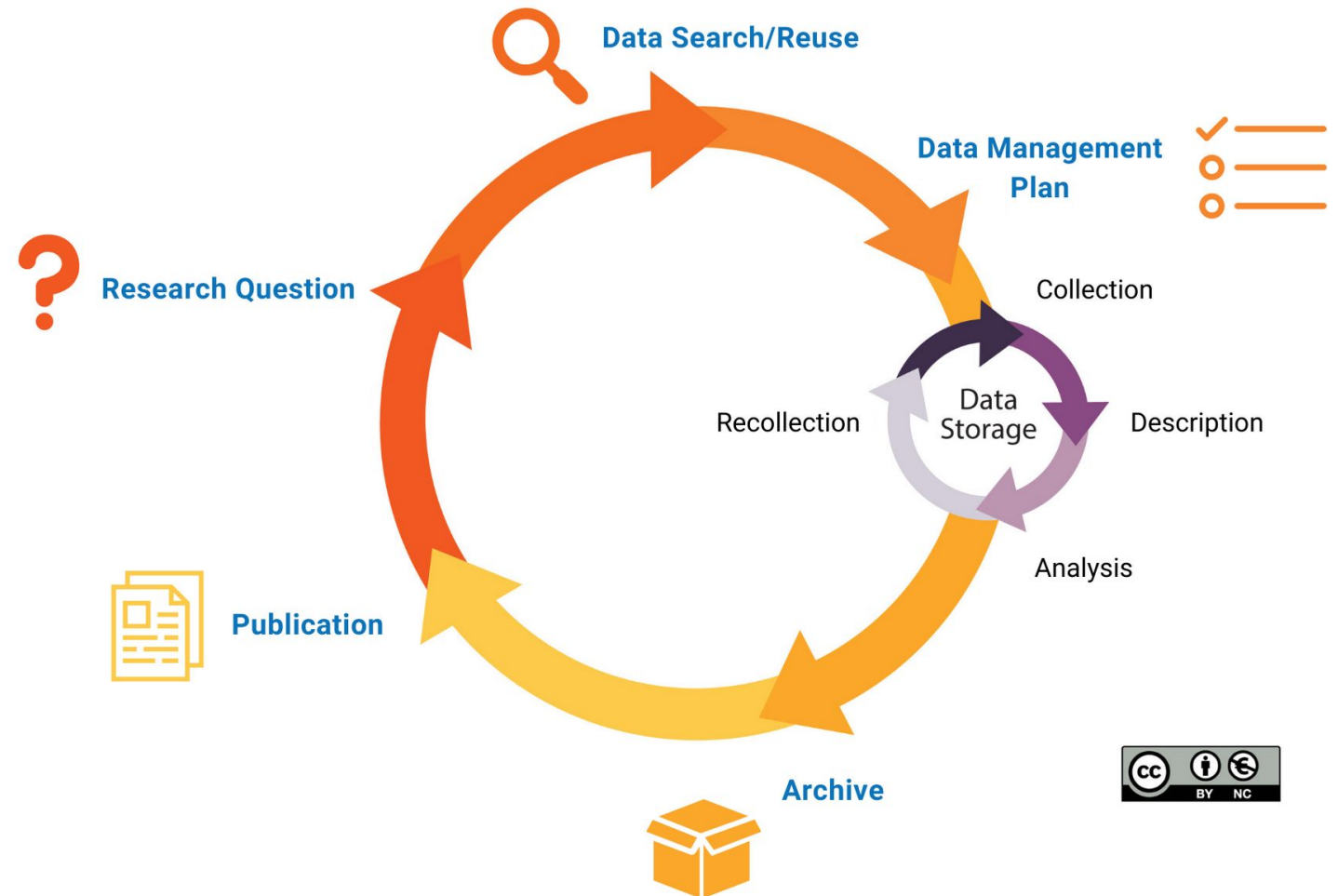
The Research Data Management Lifecycle

Data Search/Reuse

Data Management Plan

Research Question

Collection

Recollection

Data Storage

Description

Analysis

Publication

Archive

DMP requirements

Funding Agency	Types of Data	Contextual Details/Metadata	Storage, Backup and Security	Provisions for Protection/Privacy	Policies for Re-Use	Policies for Access and Sharing	Plans for Archiving and Preservation of Access
NSF	X	X	X	X	X	X	X
NIH		Partial		X	X	X	
CDC		Partial		X		X	
NEH	X	X	X	X	X	X	X
IMLS	X	X	X	X	X	X	X

DMP (and RDM) best practices

- Think in terms of the lifecycle of your data.

- Check the UNCG RDM guide or other guides.

- Ask for help!

- Use DMPTool (https://dmptool.org).

Guides and resources: RDM

- [Best practices for file naming | Stanford Libraries](#)
- [Data Management for Researchers](#) (ebook)
- [Further Reading on Data Management Issues | ICPSR](#)
- [Managing Scientific Information and Research Data (ch. 8 & 9)](#) (ebook)
- [NNLM RD3: Resources for Data-Driven Discovery](#)
- [ORI Introduction to the Responsible Conduct of Research](#) (report)
- [Research Data | University of Michigan](#)
- [Research Data Management | UNCG](#)
- [Research Data Management Toolkit | Jisc](#)
- [Understanding Research Data Management | University of Pittsburgh](#)

Questions?

Jo Klein

ejklein@uncg.edu

Schedule a consultation or chat with me:
https://uncg.libguides.com/RDM