Research Data Management 101

https://libraries.mit.edu/data-management/services/workshops/

Data Management: Why is it important?

Money

Time

Sharing

Integrity

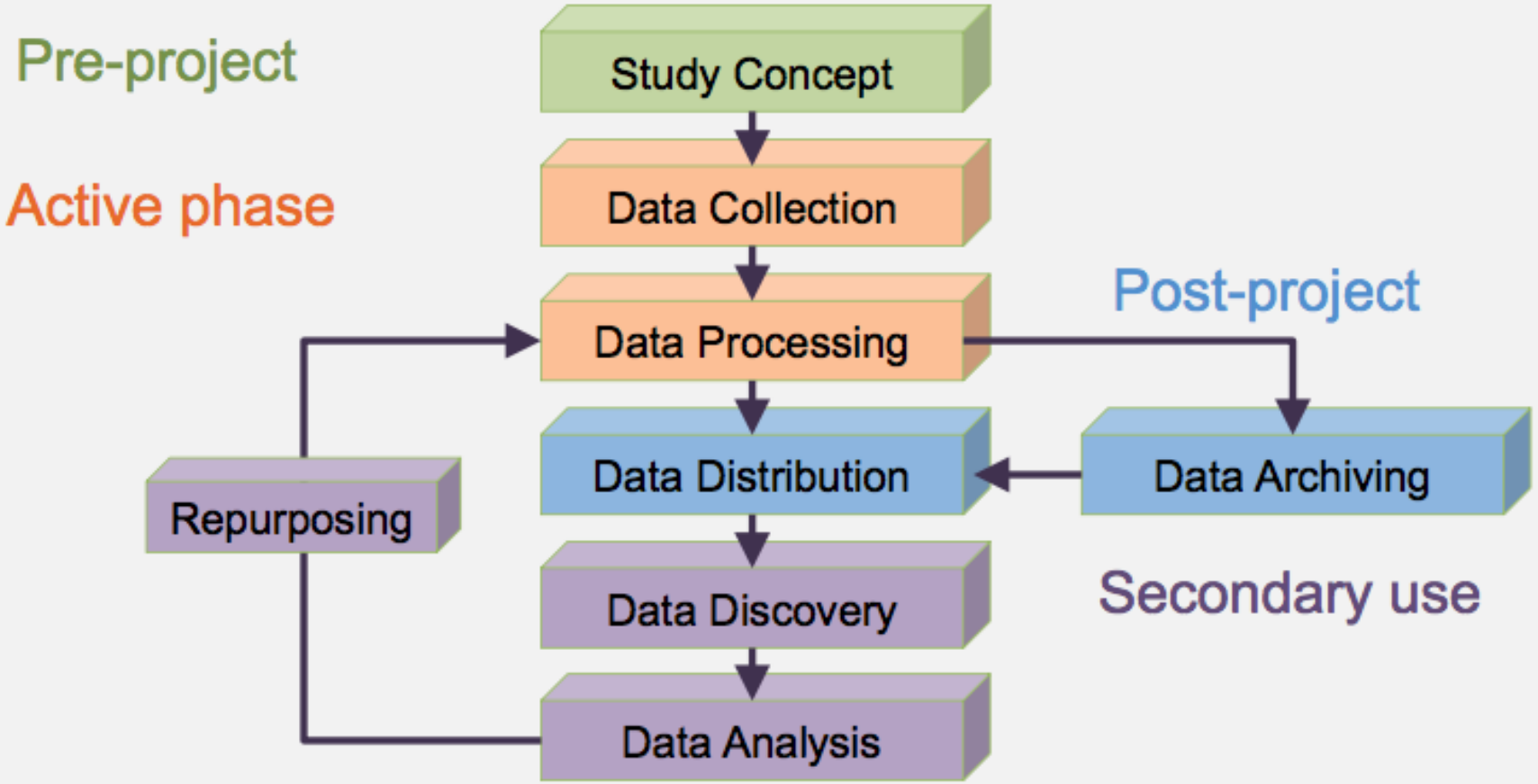
Improve your analysis

Writing a DMP

What do we mean by *data*?

General	Social Sciences	Natural/Physical Sciences
• images • video • mapping/GIS data • numerical measurements	• survey responses • focus group and individual interview transcripts • economic indicators • demographics • opinion polling	• measurements generated by sensors/laboratory instruments • computer modeling • simulations • observations and/or field studies • specimen

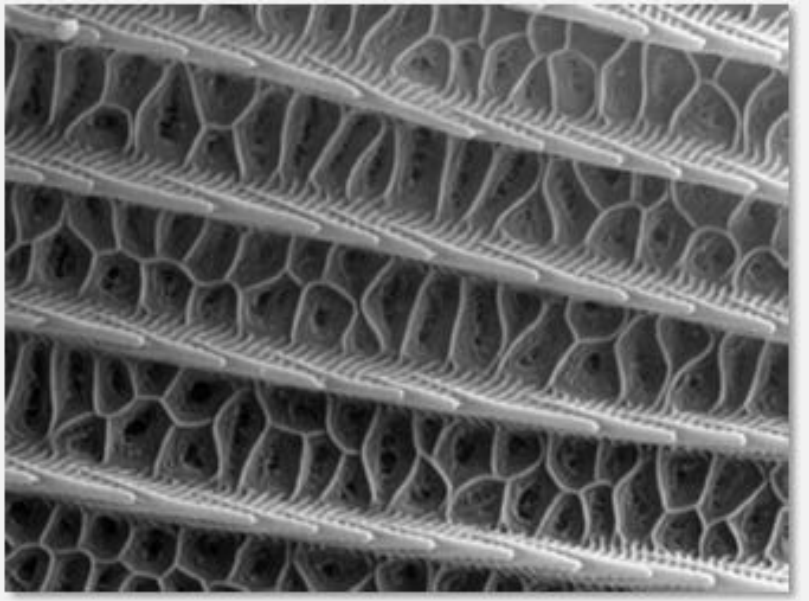
Research data lifecycle phases



Lifecycle adapted from DDI version 3.0 Combined Life Cycle Model

MIT Libraries

A case study



Data Documentation

AKA Metadata

Metadata: why does it matter?

Data is not self-describing.

Metadata, or **"data about data"** explains your dataset and allows you to document important information for:

- **Finding** the data later
- **Understanding** what the data is later
- **Sharing** the data (both with collaborators and future secondary data users)
- Consider it an **investment** of time that will save you trouble later several-fold

MIT Libraries

Documentation

Project name

Project summary

Funding info

Primary contact info

Other people working on the project

Location of data and supporting info (lab notebooks, procedures, etc.) for project

Organization and naming conventions used for data

Capturing your Metadata

In a filename

In a README.txt

In a spreadsheet

In an XML file

Into a database

README

```
Creator: Katherine McNeill
Subject: monarch butterfly wing
Description: this directory contains Sashimi ESEM images of a butterfly wing
I took after finding a butterfly floating by the Charles River.
Contributor: Mark Clemens helped me with these images
Date: 20151015
Original Format: Sashimi Microscope format (.sam)
Relation: this is a directory that will contain multiple files
Type: image
Coverage: By the Charles River in Cambridge, MA.
Rights: National Science Foundation (funder) owns the data (Grant number:
00213)
```

Data Organization

Why file organization is important

@AksharPathak
YASH BHARDWAJ & JUGAAD POSTERS

Once your research gets underway, there may be multiple files in various formats, multiple versions, methodologies, etc., all relating to your research.

MITLibraries

File organization: naming conventions

Best Practice	Example
Limit the file name to 32 characters (preferably less!)	32CharactersLooksExactlyLikeThis.csv
When using sequential numbering, **use leading zeros** to allow for multi-digit versions For a sequence of 1-10: 01-10 For a sequence of 1-100: 001-010-100	**NO** ProjID_1.csv ProjID_12.csv **YES** ProjID_01.csv ProjID_12.csv
Don't use special characters & , * % # ; * ( ) ! @$ ^ ~ ' { } [ ] ? < > -	**NO** name&date@location.doc
Use only one period and use it before the file extension	**NO** name.date.doc **NO** name_date..doc **YES** name_date.doc
Avoid using generic data file names that may conflict when moved from one location to another	**NO** MyData.csv **YES** ProjID_date.csv

MITLibraries

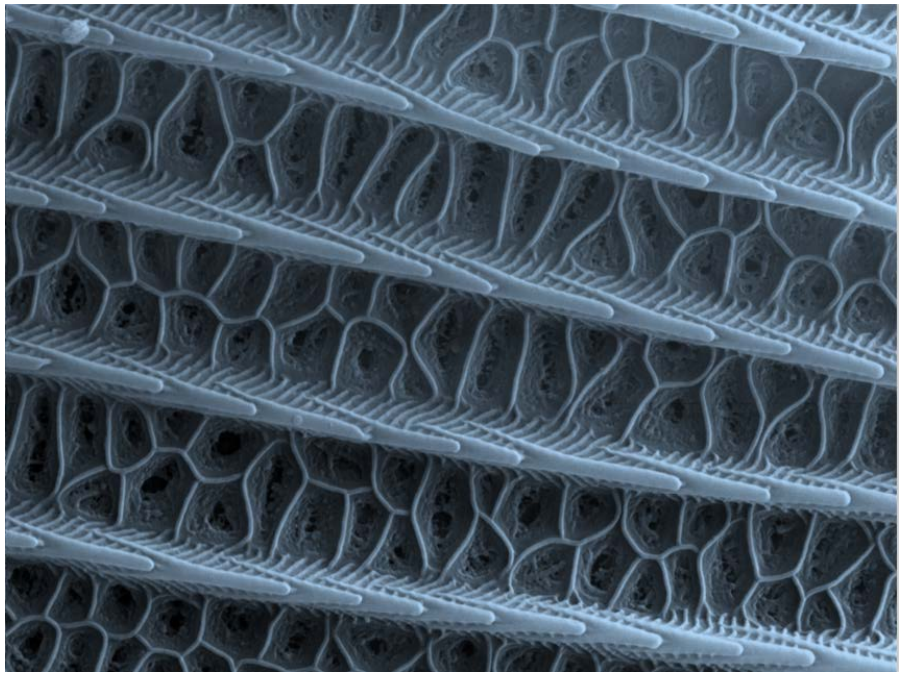
File Organization: naming conventions

Example for our case study:

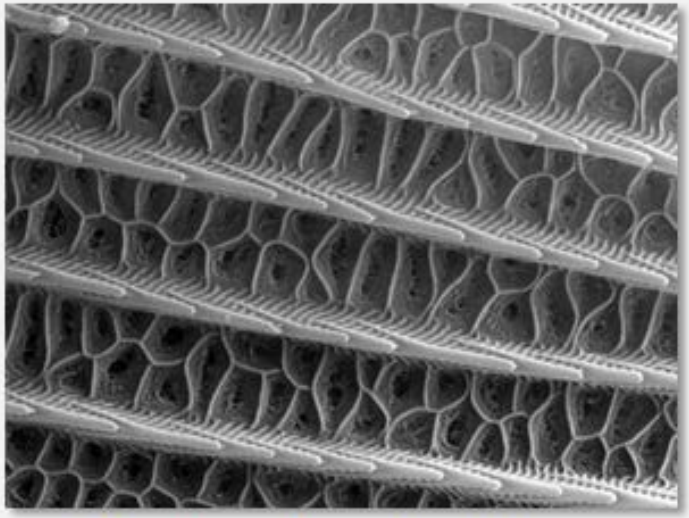
Picture_a2Xc38_butterfly.sam

How could this be better named?

Identifiers?



File organization: naming conventions



Maybe started with:

abcdefghijklmnopqrstuvwxyz.sam

Sashimi Microscope format

Date as YYYYMMDD

Ascension # because part of a series

sam_monarch_wing_20150415_CM_001.tif

File format

Descriptive element

Initials because working in a group

MITLibraries

Data Storage

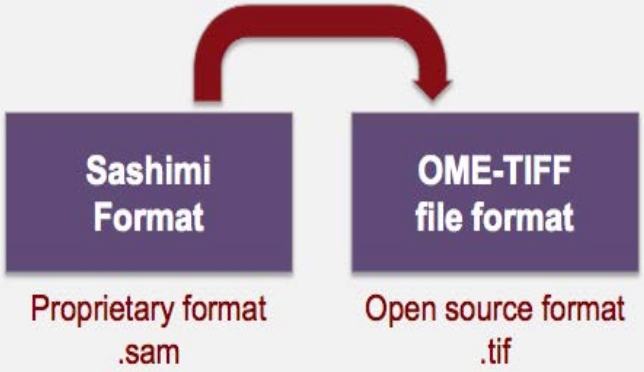
Data Storage: during active phase

Ideally keep 3 copies of your data!

- Local/working copy
- Local external copy (e.g. external hard drive)
- Remote copy offsite (e.g. cloud storage)

File formats: preferred examples

Proprietary Format	Alternative/Preferred Format
Excel (.xls, .xlsx)	Comma Separated Values (.csv) ASCII
Word (.doc, .docx)	plain text (.txt), XML, PDF/A, HTML, ODF or if formatting is needed, PDF/A (.pdf)
PowerPoint (.ppt, .pptx)	PDF/A (.pdf), ODP, JPEG 2000, PDF, PNG
Photoshop (.psd)	TIFF (.tif, .tiff),
Quicktime (.mov)	MPEG-4 (.mp4), MOV, AVI, MXF
Sounds	WAVE, AIFF
Containers	TAR, GZIP, ZIP
Databases	XML, CSV



Data Security

Software

Passwords

Safe Storage Environment

Encryption

Preservation and Sharing

Preserving and Sharing

- Preservation
  - Repositories
  - Individual
- Sharing
  - Website
  - Email upon request
  - Repository

Preserving and Sharing

Our Case Study chose…

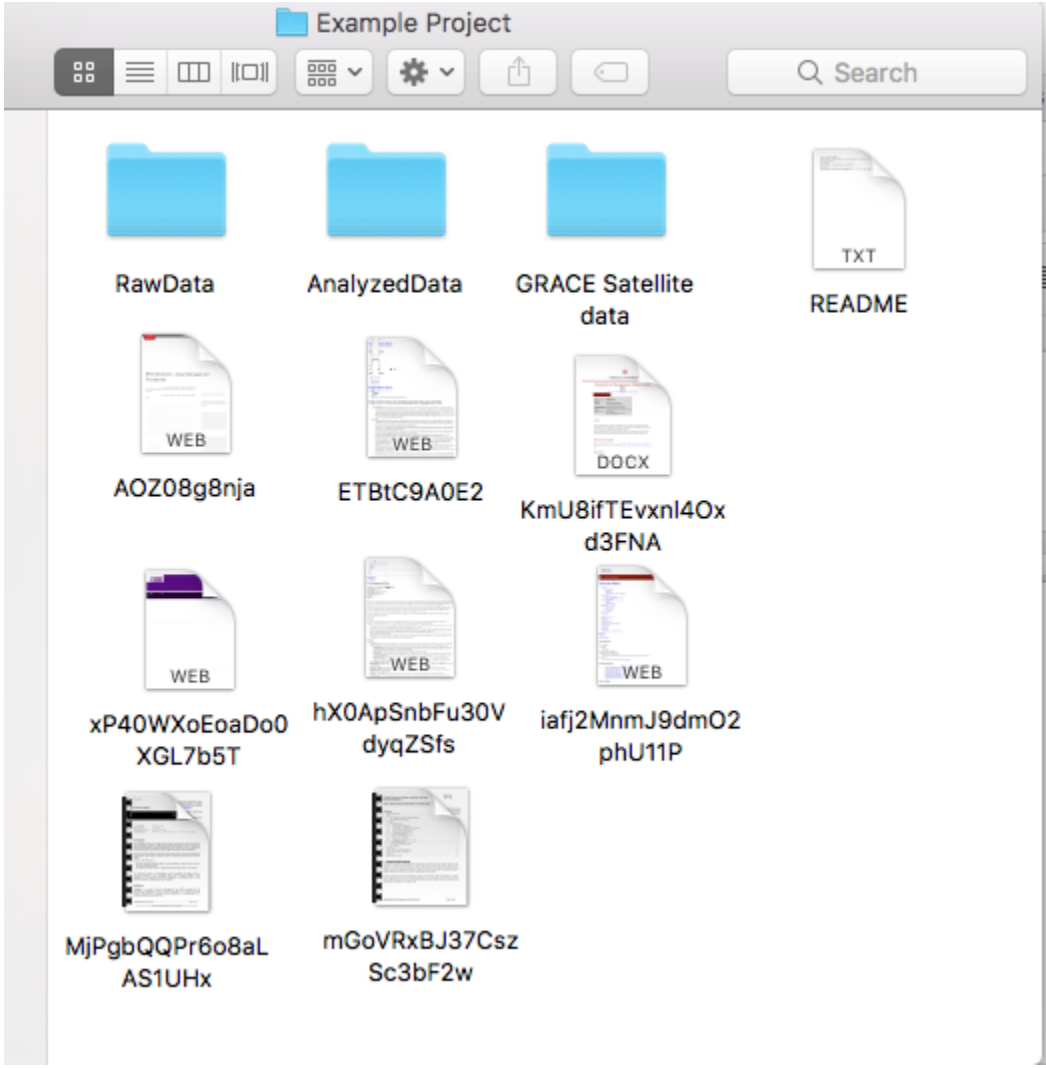
Email upon request

Why?

- informal/easy
- Keep control over who sees it
- Control over managing and preserving

Scenario: What's good and bad?

A project on ground water changes using GRACE satellite data.



Quick Practices

Check backups

Organization system

Take better notes

Review security plan

Check to see if you can access old files