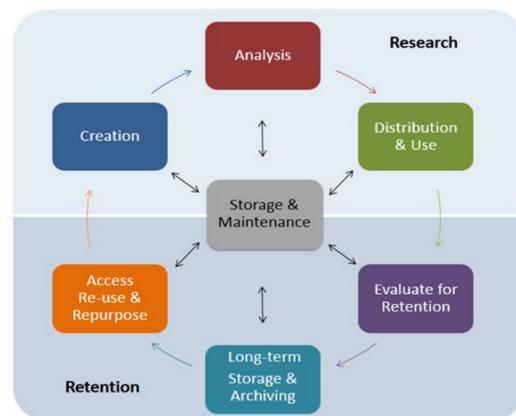


# Effective data management to promote reproducibility

Authors: Diep Ganguly, ANU, AUS ([0000-0001-6746-0181](mailto:0000-0001-6746-0181)); Peter Crisp, UMN, USA ([0000-0002-3655-0130](mailto:0000-0002-3655-0130)); Ben Schwessinger, ANU, AUS ([0000-0002-7194-2922](mailto:0000-0002-7194-2922))

## What is data management?

→ The organization, curation, and storage of data or information in a manner that promotes accessibility, and effective and reproducible processing.



Source: <https://datamanagement.hms.harvard.edu/>

- ↑ Findable
- ↑ Accessible
- ↑ Interoperable
- ↑ Reusable

## Benefits

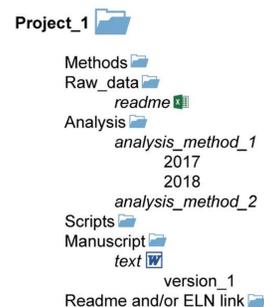
- It's good for (open) science!
  - ◆ Improved research quality and reproducibility
  - ◆ Increased accessibility and visibility of your research
  - ◆ Better for collaboration
- Better data archival and retention
- Easier data submission
- **START EARLY!** A data management plan should be incorporated during research development



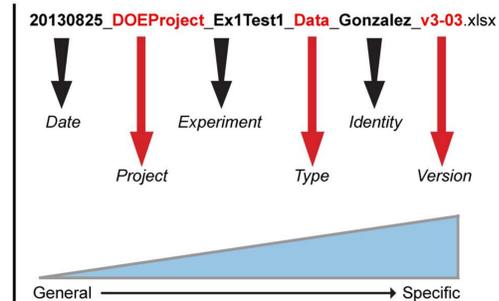
## What to incorporate into a data management plan?

- Scales of data management = folder structure, file naming conventions, file types, README files etc.
- **What** data will be produced as a part of the project
- **How** each type of data will be organized, documented, standardized, stored, protected, shared and archived
- **Who** will take responsibility for carrying out the activities listed above
- **When** these activities will take place over the course of the project (and beyond)
- **Metadata**

### PROJECT DIRECTORY STRUCTURE



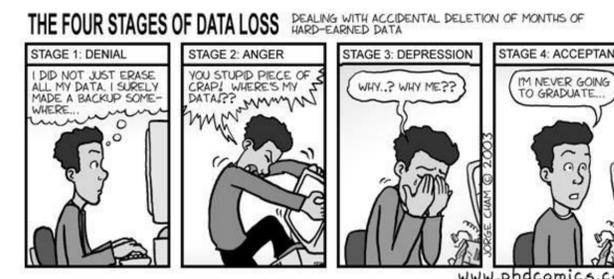
### FILE NAMING CONVENTION



<https://plantae.org/experimental-reproducibility-101-part-1/>

## What do I need for a data management plan?

- Data standards (including some level of data screening)
- Practical, organised file naming conventions and folder structure
  - ◆ Include date in yyyy-mm-dd format
  - ◆ Use meaningful abbreviations
  - ◆ Have group identifiers
  - ◆ Document your decisions
  - ◆ Be **consistent**
- Reliable storage options (including ample space)
- Have effective means of **backing your data up**
- Utilize available purpose-built software
  - ◆ LabArchives (<https://www.labarchives.com/>)
  - ◆ Open Science Framework (<https://osf.io/>)
  - ◆ PrometheusWiki ([http://prometheuswiki.org/tiki-custom\\_home.php](http://prometheuswiki.org/tiki-custom_home.php))



## Resources

- Reproducibility in plant science
  - <https://community.plantae.org/organization/reproducibility-in-plant-science/dashboard>
- File naming conventions <http://www.exadox.com/en/articles/file-naming-convention-ten-rules-best-practice>
- Folder organization
  - <https://zapier.com/blog/organize-files-folders/>
- Harvard University Data Management page
  - <https://datamanagement.hms.harvard.edu/>
- Kbroman Lab (Short primer on data storage and handling form Kbroman)
  - <http://kbroman.org/dataorg/>
- Purdue Library (Short primer on data management and file naming conventions)
  - <http://guides.lib.purdue.edu/c.php?g=353013&p=2378292>
- Data One Best Practices (Detailed resource on how to handle data throughout its life-cycle)
  - <https://www.dataone.org/best-practices>
- Mantra (Free online course for handling digital data)
  - <https://mantra.edina.ac.uk/>

