

# Introduction to Research Data Management

File organization

<http://hdl.handle.net/1969.1/164383>



# Workshops

1. Build an overview
2. Collect and document data
- 3. Store digital data**
4. Work with data
5. Share and preserve data
6. Plan ahead

# Introduction

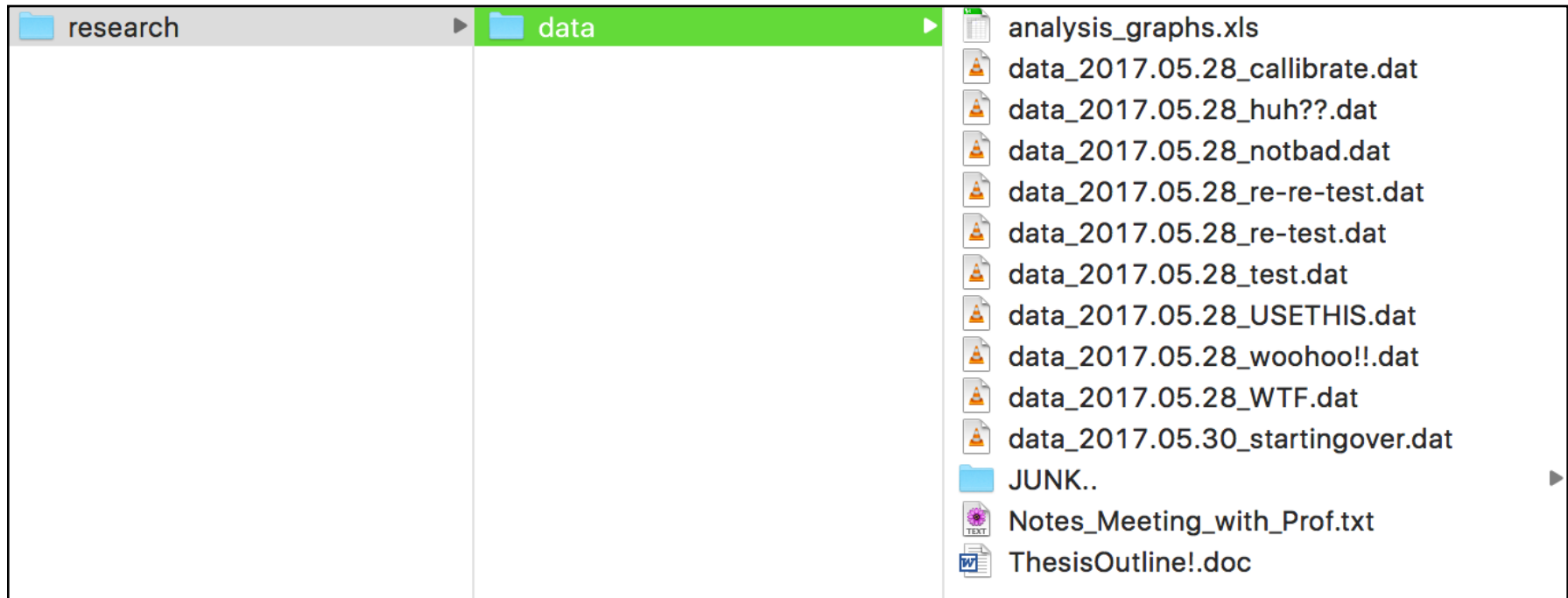
Focus: Organizing digital files with accessibility in mind.

Goal: Be aware of naming strategies that can help you focus on working with your data, rather than trying to find files.

# Discussion

Do you or your research group already have a particular way of naming files?

Can you describe what goes into a file name?



# Naming conventions

Have conventions for your:

1. File names
2. Folder names
3. Directory structure

# What's in a name?

Component	Possible use case	Tip
Name or acronym	Creator, project; team, named data	Relevant and simple
Sequential #	Run of experiment, version number	Include leading zeros
Date and time	Date of creation, date range of experiment	YYYY-MM-DD (ISO 8601) format
Identifier	Subject, project, grant	Relevant and simple
Research condition	Instrument, temperature, model	Relevant and simple
Type or keyword	Denote type of content in a file	Use a standard list
Extension	Denote file format	Use them

# Putting components together

1. Less than 32 characters, enough to uniquely identify.
2. Same components, in same order.
3. No special characters ‘( & \* % \$ £ ] { ! @)’
4. ‘-’ or ‘\_’ instead of ‘ ‘
5. Include file extensions, after ‘.’



# Arranging components

**Date** - chronological order

20170809-jansen-sample-data.csv  
20170823-zheng-sample-data.csv  
20170902-arora-sample-data.csv

**Type** - categorical order

data-v09.csv  
graph-v01.csv  
graph-v02.csv

**Name** - order by creator

arora-20170902-sample-data.csv  
jansen-20170809-sample-data.csv  
zheng-20170823-sample-data.csv

**Sequential number** - order by run

012-cesm-model-output.csv  
013-cesm-model-output.csv  
014-cesm-model-output.csv

# Example: Generic file naming convention

Data and figures

**Working version**

[period\*]-[creator\*]-[document\_title]-[type].[ext]

**Fixed version**

[period\*]-[creator\*]-[document\_title]-[type]-v[##].[ext]

\* = Optional component.

**[period]** = Date or dates covered, at least [YYYY].

**[creator]** = Name of primary creator or maintainer, a person, organization, or system.

**[title]** = project or experiment title.

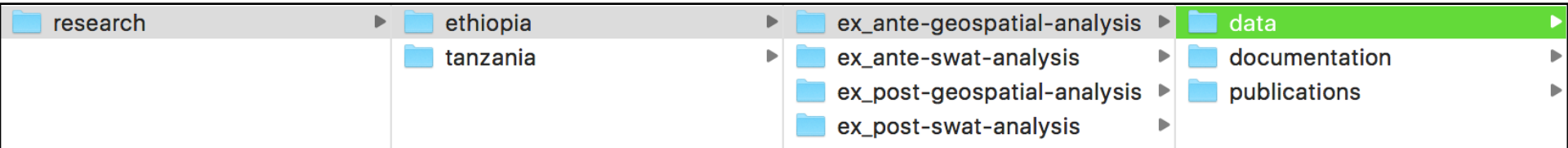
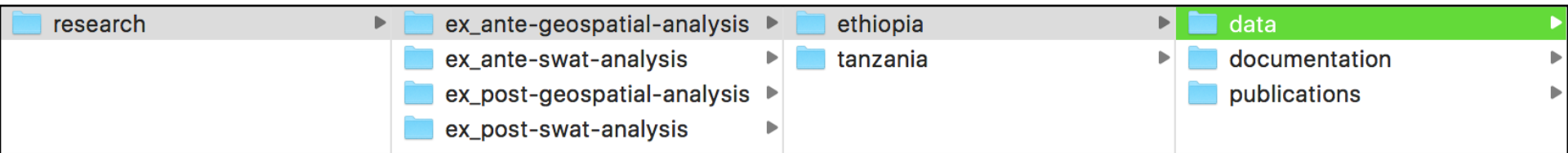
**[##]** = number with leading zeros, e.g. 01.

**[type]** = type of document within file:


















metadata, data, code, list, table, figure, plot, graph, diagram, schema

# Directory structure

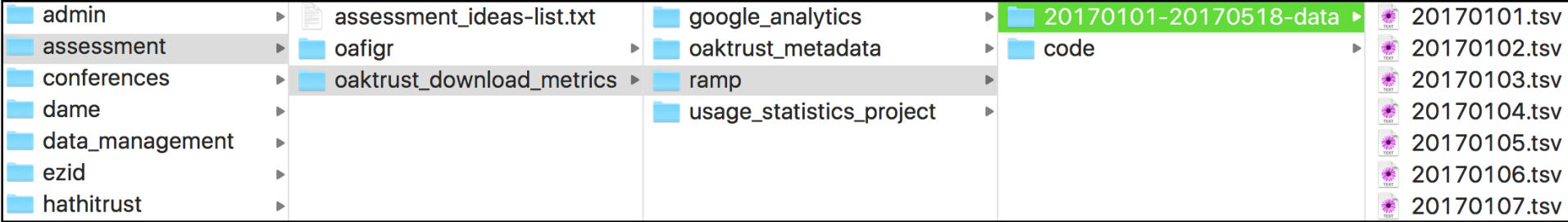
Based on the projects, subject matter, and types of documents you produce.



# Exercise: Critique me!

 admin	▶	 consulting	▶	 201702-lyon-data_and_copyright-slides.pptx
 assessment	▶	 <b>copyright</b>	▶	 20140312-enimil-copyright_data-slides.pdf
 conferences	▶	 data_science_institute	▶	 20170323-algenio-data_legal_issues-faq.docx
 dame	▶	 environmental_scan	▶	
 data_management	▶	 ideas	▶	
 ezid	▶	 outreach	▶	
 hathitrust	▶	 rdm_workshops	▶	

# Exercise: Critique me!



# Batch file renaming

**Finder:** Built-in Mac OS renaming utility.

**Advanced Renamer:** Windows software for renaming multiple files and folders at once.

**Bulk Rename Utility:** Windows software to rename files and folders based with flexible criteria.

# Conclusion

- Reviewed the purpose of file naming conventions.
- Discussed the components of sensible file names.
- Compared conventions.

# References and resources

- Cornell. "File management" [Website](<https://data.research.cornell.edu/content/file-management>)
- DMPTool "Data Management General Guidance" [Website]([https://dmptool.org/dm\\_guidance](https://dmptool.org/dm_guidance))
- Library Carpentry. [Lesson](<https://data-lessons.github.io/library-data-intro/03-foundations/>)
- MANTRA "Organizaing data" [Module](<http://mantra.edina.ac.uk/organisingdata/>)
- MIT Libraries [Website](<http://libraries.mit.edu/data-management/store/organize/>)
- Frazer, Meghan. 2013. "An eleveator pitch for file naming conventions" [Article](<http://acrl.ala.org/techconnect/post/an-elevator-pitch-for-file-naming-conventions>)