

DATA'S NAMING AND ORGANIZING

Midi de l'info scientifique

Audrey.Bellier@unige.ch

6 November 2017

LIBRARY



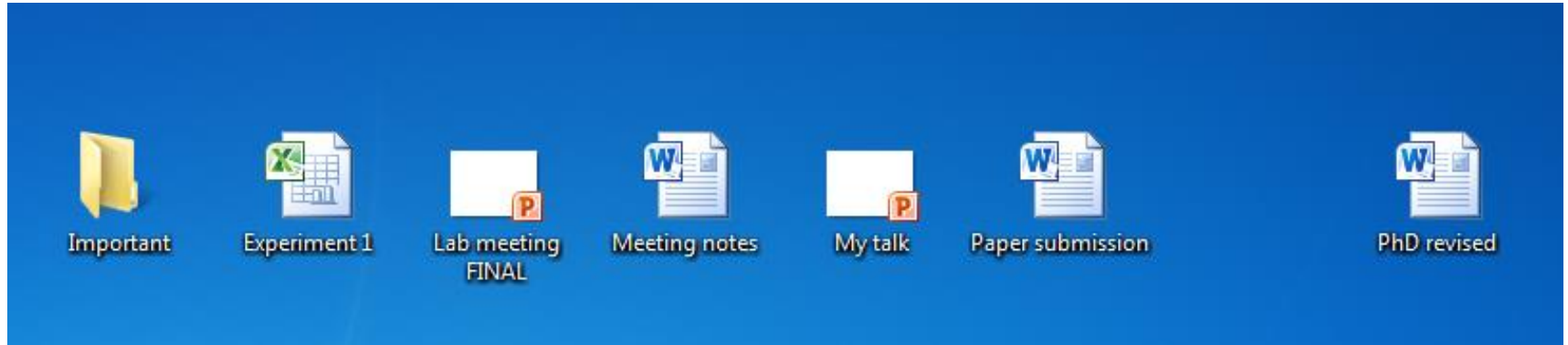
UNIVERSITÉ
DE GENÈVE

WHY ORGANIZING YOUR RESEARCH DATA ?

- Save **time** and **resources**
- For yourself, your colleagues and whoever might be interested in your data, even after you have left the lab



HOW DO YOU ORGANIZE YOUR DATA ?



- File structures
- File naming
- File versioning

DATA ORGANIZATION

- File structures
- File naming
- File versioning

KEY PRINCIPLES OF FILE ORGANIZATION

1. Spending a little time **upfront**, can save a lot of time later on
2. Be **realistic**: strike a balance between doing too much and too little
3. There's no single right way to do it; establish a **system that works for you**
4. Think about **who your system needs to work for**:
Just you? You and your lab group? Collaborators?

FILES ORGANIZATION: THE 5 C's

- ✓ Clear
- ✓ Concise
- ✓ Consistent
- ✓ Correct
- ✓ Conformant

WHICH EXAMPLE IS BEST? WHY?

Example A

Documents library

PhD data

- 25July
- Documents
- Experiments
- Experiments2
- From desktop
- Important
- Other
- PhD
- Talks
- Experiment 1
- Lab meeting FINAL
- Meeting notes
- My talk
- Paper submission
- PhD revised

Example B

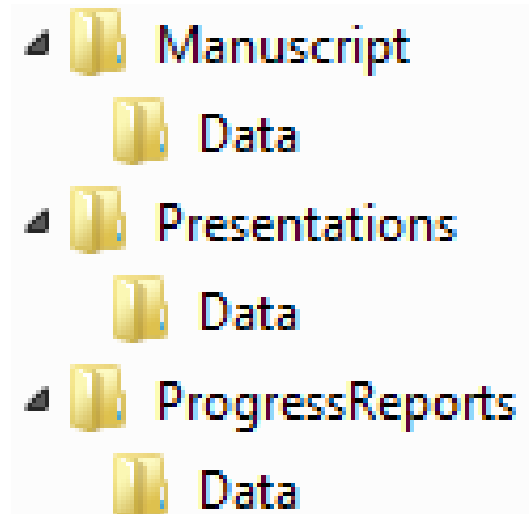
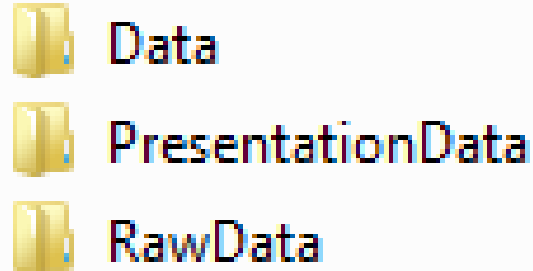
Documents library

PhD data2

- Conferences
 - Downloaded publications
 - Experimental data
 - Financial documents
 - PhD thesis
 - Presentations
 - Protocols
 - Reagents
 - Reports
 - Training

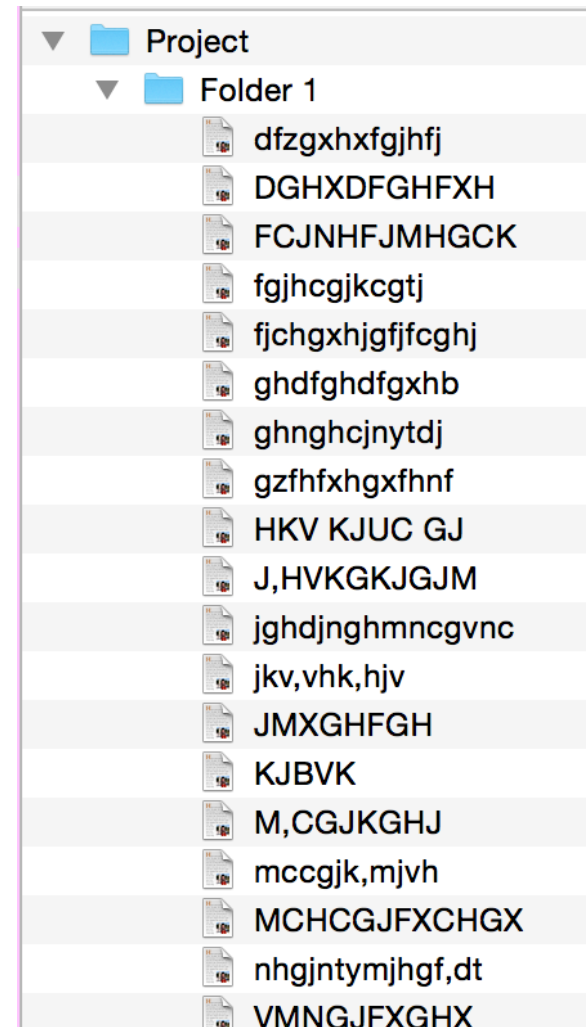
FILE STRUCTURE : HIERARCHICAL METHOD

1. avoid overlapping categories



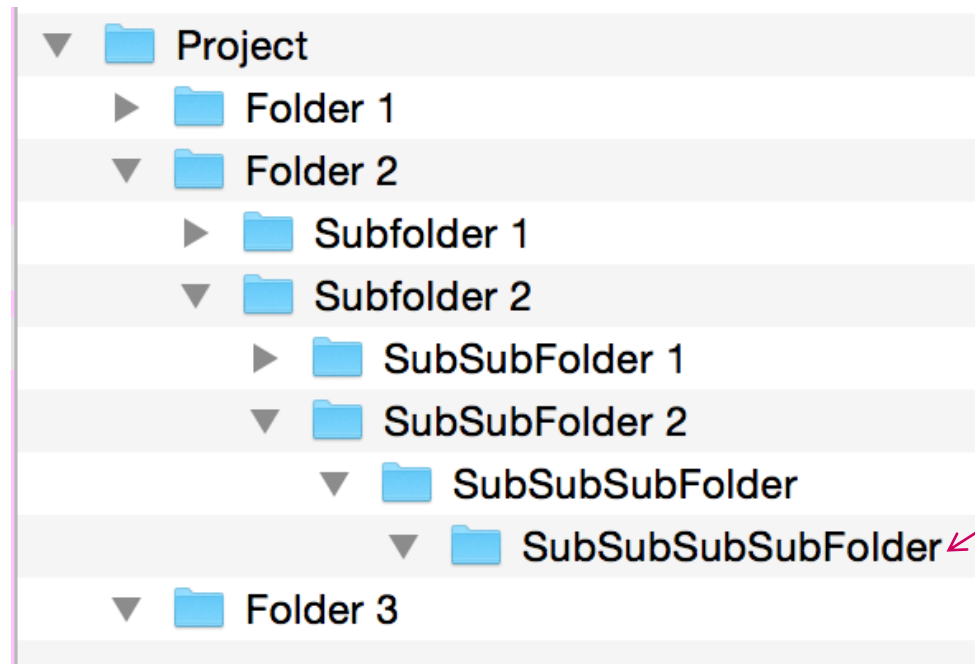
FILE STRUCTURE : HIERARCHICAL METHOD

1. Avoid overlapping categories
2. Don't let your folders get too big



FILE STRUCTURE : HIERARCHICAL METHOD

1. Avoid overlapping categories
2. Don't let your folders get too big
3. Don't let your structures go too deep



How many
clicks does it
take to get
there?

DATA ORGANIZATION

- File structures
- **File naming**
- File versioning

EXERCICE: FILE NAMING

My passwords.doc	My data.xls
IMPORTANT.doc	My study.doc
My Thesis final final.doc	Doc.1.doc
My Thesis version 12.doc	New doc.doc
Data 01/08/2016.xls	Int 1 (2).doc
Data 10 jan. 2016.xls	Interview 1.doc



1. Returning to your data in 1 year, will you recognize what these files contain?
2. What information needs to be in a file name to identify the content?
3. What would you change in these names?

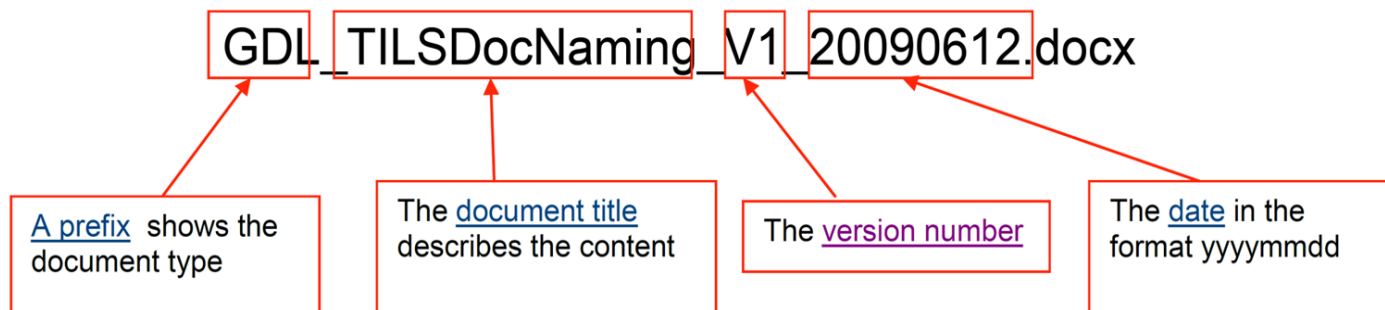
FILE NAMING CONVENTIONS

Best Practice	Example
Order dates beginning with the year to enable sorting by date	YYYYMMDD
Limit the file name to 32 characters	32CharactersLooksExactlyLikeThis.csv
For 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_2.csv ProjID_12.csv YES : ProjID_02.csv ProjID_12.csv
Don't use special characters & , * % # ; * () ! @\$ ^ ~ ' { } [] ? < > -	NO : name&date@location.doc YES : name_date_location.doc
Use only one period (for the file extension)	NO : name.date.doc YES : name_date.doc
Avoid using generic data file names that may be ambiguous when moved	NO : MyData.csv YES : ProjID_date.csv

FILE NAMING CONVENTIONS

TILS Document Naming Convention

Document naming for the TILS Division should follow this convention:



README.TXT

Readme.txt : File with explanations of your conventions, e.g. a table with the meaning of your different abbreviations

- For you, to stick with your standards
- For you, to understand your file names in 15 yrs (or 15 days!)
- For other people, to use your files
- For other people, to save files properly in your folders

DATA ORGANIZATION

- File structures
- File naming
- **File versioning**

"FINAL".doc



FINAL.doc!



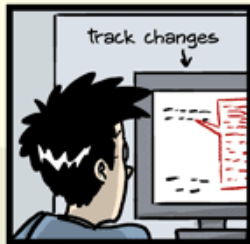
FINAL_rev.2.doc



FINAL_rev.6.COMMENTS.doc



FINAL_rev.8.comments5.
CORRECTIONS.doc



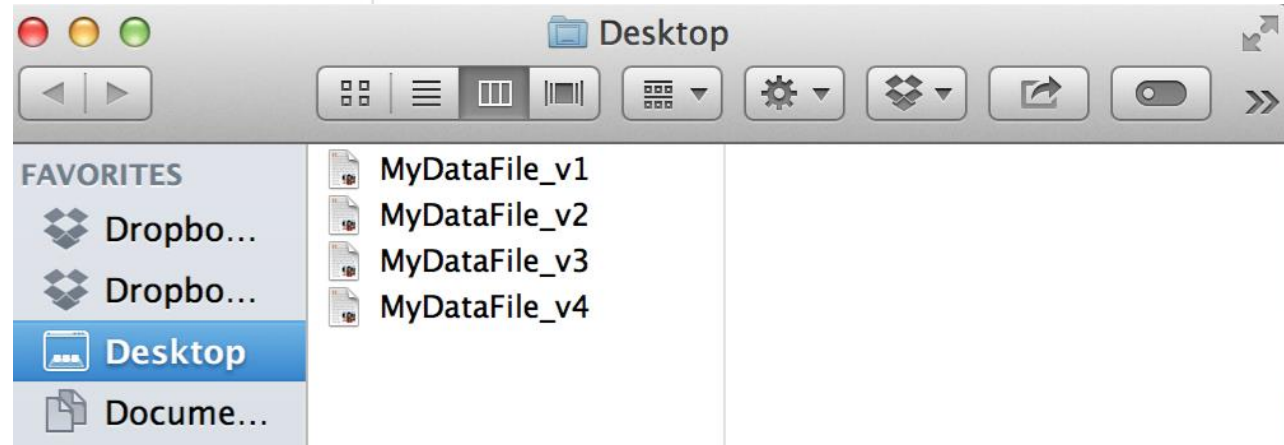
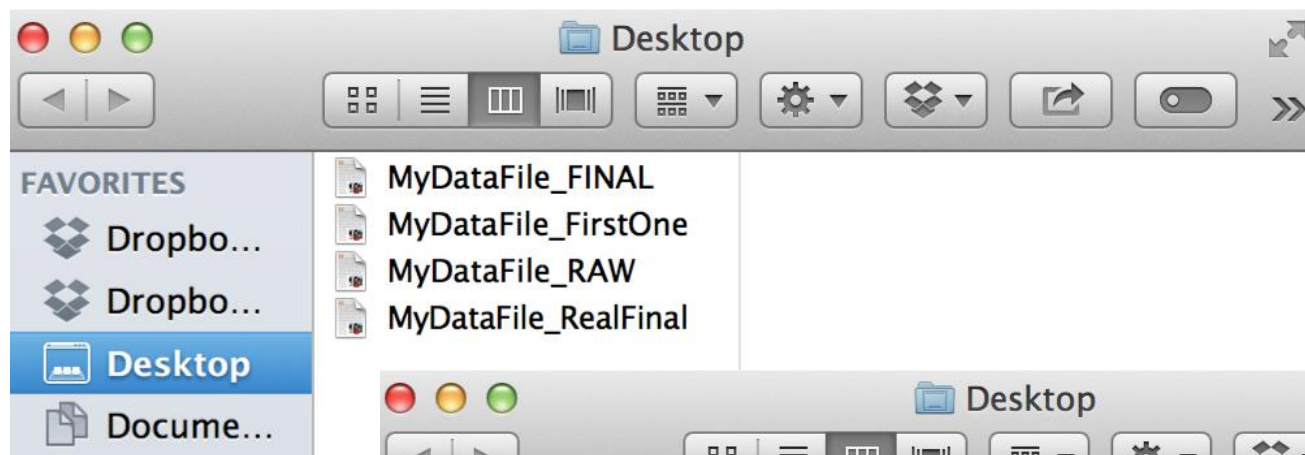
FINAL_rev.18.comments7.
corrections9.MORE.30.doc



FINAL_rev.22.comments49.
corrections.10.#@\$%WHYDID
ICOMETOGRADSCHOOL????.doc

FILE VERSIONING

Avoid imprecise “final” labels



FILE VERSIONING

1. Save new versions
2. Establish a consistent convention
3. Document your convention
4. Consider your version control needs

OUTILS DE RENOMMAGE

Ant Renamer (Windows):

<http://www.antp.be/software/renamer/fr>

Massive File Renamer (Windows, Linux, Mac):

<https://ivanrf.com/en/portfolio-posts/massive-file-renamer/>

Renamer 5 (Mac): <https://renamer.com/>

TAKE HOME MESSAGE

- Take time **upfront** to think what works best for you
- Organize your files in a logical **hierarchy**
- Name your files using **conventions**
- Create a **readme.txt** file
- Keep track of the different **versions**

NEXT «MIDIS DE L'INFO»

Tuesday 21.11	CMU	Customized publications monitoring Keep yourself up to date on last publications with PubMed, WoS and SciFinder	
Tuesday 05.12	CMU	Plagiarism detection with Compilatio Understand how the software works and be aware of its limits	
Tuesday 12.12	Sciences 2	Emerging trends in peer review Discover innovative forms of peer review for scientific communication	



Bibliothèque de l'UNIGE, 2017

Ce document est sous licence Creative Commons Attribution - Partage dans les mêmes conditions 4.0 International : <http://creativecommons.org/licenses/by-sa/4.0/deed.fr>.