

Data Organization

Dalal Rahme

Data Services Librarian

AUB Libraries



Outline

- File Structuring
- Naming Conventions
- Version Control
- File Formats
- Data Documentation
 - Dictionaries
 - Readme files

Why Organizing Data?

Data Sharing and Management Snafu in 3 Short Acts
by Karen Hanson, Alisa Surkis & Karen Yacobucci
NYU Health Sciences Libraries
August 3, 2012 (Last Update: December 12, 2012)

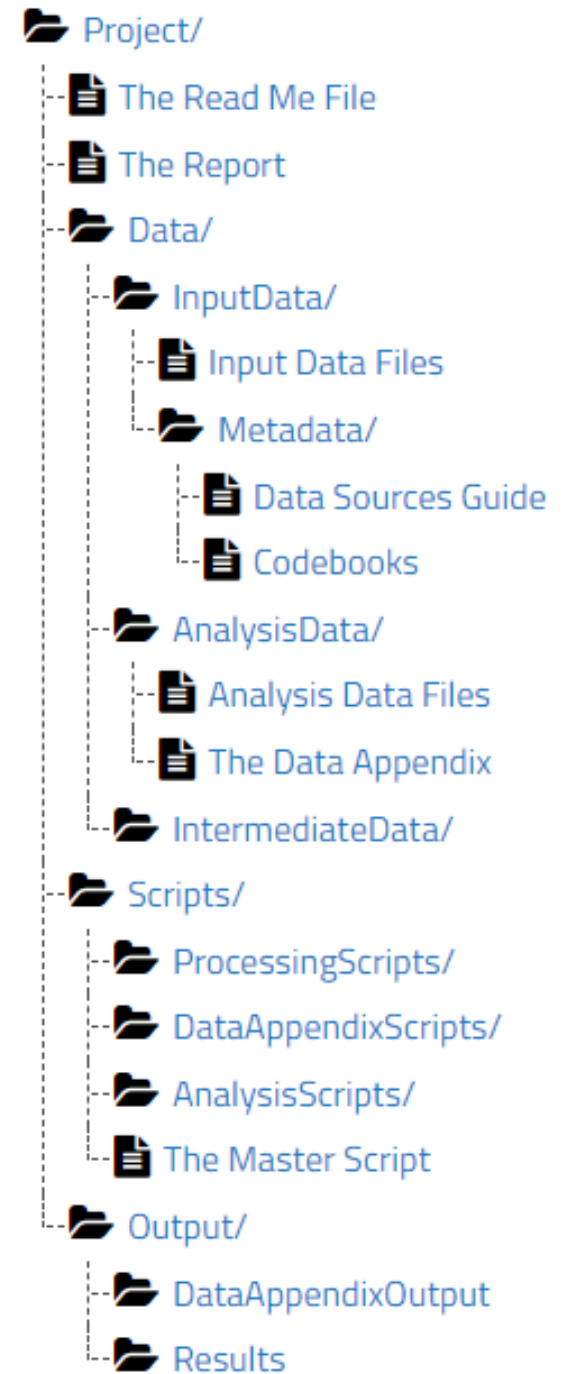


File Structuring

- Adhere to existing procedures
- Use folders
- Name folders appropriately
- Be consistent
- Separate ongoing and completed work
- Backup

Project Tier Protocol

<https://www.projecttier.org/tier-protocol/protocol-4-0/>



Naming Conventions

Files with no Naming Conventions

- Test data 2016.xlsx
- Meeting notes Jan 17.doc
- Notes Eric.txt
- Final FINAL last version.docx



Naming Conventions Dos

- Keep names short (≤ 25 characters).
- Choose names using 3-4 key unchanging pieces of information
- Use YYYY-MM-DD format for better sorting even over the span of many years.
- All numeric fields should be zero-padded for equivalent width.
- For better visibility, give preference to dashes over underscores.
- Create a README.txt describing the file naming convention.
- Be consistent!

Naming Conventions Don'ts

- Do not add spaces! They are often interpreted as delimiters and may cause problems.
- Do not include special characters such as: " / \ [] : ; | = , < ? > & \$ # ! ' { } () * .
- Do not rely on case to distinguish filenames. Not all systems are case sensitive.
- Avoid using words such as 'draft' or 'letter' at the start of file and folder names.

Naming Conventions Dos

Consider including:

- Unique Identifiers such as project name or grant number
- Conditions such as lab instrument, solvent..etc
- Date
- Version number

Example 1:

aub_aco000001_000001_m.tif

aub_aco000001_000001_d.tif

(organization_projectnamebooknumber_pagenumber_master/derivativecopy. format

Example 2:

micro_hhl_20170620_dr_001.tif

(instrument_insectname_date_initials_sequential
number.format)



Version Control

How many times have you tried to go back one step and you couldn't because you did not practice versioning?

Version Control

Version control is method used to track file or set file changes over time so that you can recall older versions at a later time.

Version Control

Revised on	Version	Description	Approved by
01-Oct-2015	0.1	Initial Document Creation	IT Manager
10-Jan-2016	1.0	Publish Document	IT Manager
15-Jan-2016	1.1	Review document, modify policy section reflecting organizational legal-contractual service level agreements for the protection of information, including ISO, PCI DSS	IT Head
20-Jan-2016	1.2	Formal Editorial work	PMO
27-Jan-2016	1.3	Formal review and minor edits	PMO
15-Feb-2016	1.3a	Include missing protection of information assets policy clause as per Network Security Policy Design stage	PMO
22-Feb-2016	1.3b	Update mandatory requirements of the Information Security Policy with respect to Policy requirement and Network Design	PMO
10-Mar-2016	1.4	Formal review and change to use new policy template	PMO
17-Mar-2016	1.5	Formal review to reflect jobs descriptions of IT Security roles defined by HR for the organization	HR Head
25-Mar-2016	2.0	Formal approval for Senior Management	Senior executives

A stack of several folders in various colors (beige, grey, blue) is shown in the upper left corner of the image. The folders are slightly offset, creating a sense of depth. The background is a solid, textured blue.

File Formats

Proprietary Vs Non-Proprietary

 Text	Preferred	.pdf/a, .odt, .odf
	Acceptable	.csv, .txt, .xml, .rtf, .doc, .xls, .pdf, .ps
 Image	Preferred	.tiff, jp2
	Acceptable	.tiff, .jp2, .gif, .png, .jpeg
 Audio	Preferred	.bwf, .flac
	Acceptable	.aif, .lla, .ogg, .wav, .aac, .mp3, .ram, .wmv
 Video	Preferred	.avi, .mj2, .mpj2
	Acceptable	.avi, .mj2, .mpj2, .mp2, .mp4, .wmv, .mov

Lossy vs Lossless Compressions



PNG

jpeg



Data Documentation



Data Documentati on

- **Detailing your methods helps people understand what you did**
- **And helps make your work reproducible**
- **Conclusions can be verified**

Ways to Document your Data

- ReadMe Files :
 - Template from Cornell Univ.
<https://data.research.cornell.edu/content/readme>
- Data Dictionaries
- Deposit in Repositories Supporting Standardized Metadata



Questions

- For more info get in touch:
- da31@aub.edu.lb
- Ext. 2958