Breakfast with Data Stewards II

Marie Šafner
Matúš Drexler

Program

1. **Where are we at with ELNs.** (Matúš)
2. **ReQuest system and its development.** (Matúš)
3. **Live demonstration: uploading a dataset to ASEP data repository.** (Marie)
4. **RDM assignments.** (Marie)

   Reflection and feedback on the first assignment (folder file organization), initiating the next assignment on metadata.

If you have any questions there won't be time to address, please drop in to **A2.46** that day anytime between **2-3 PM**.

ELNs and ReQuest System

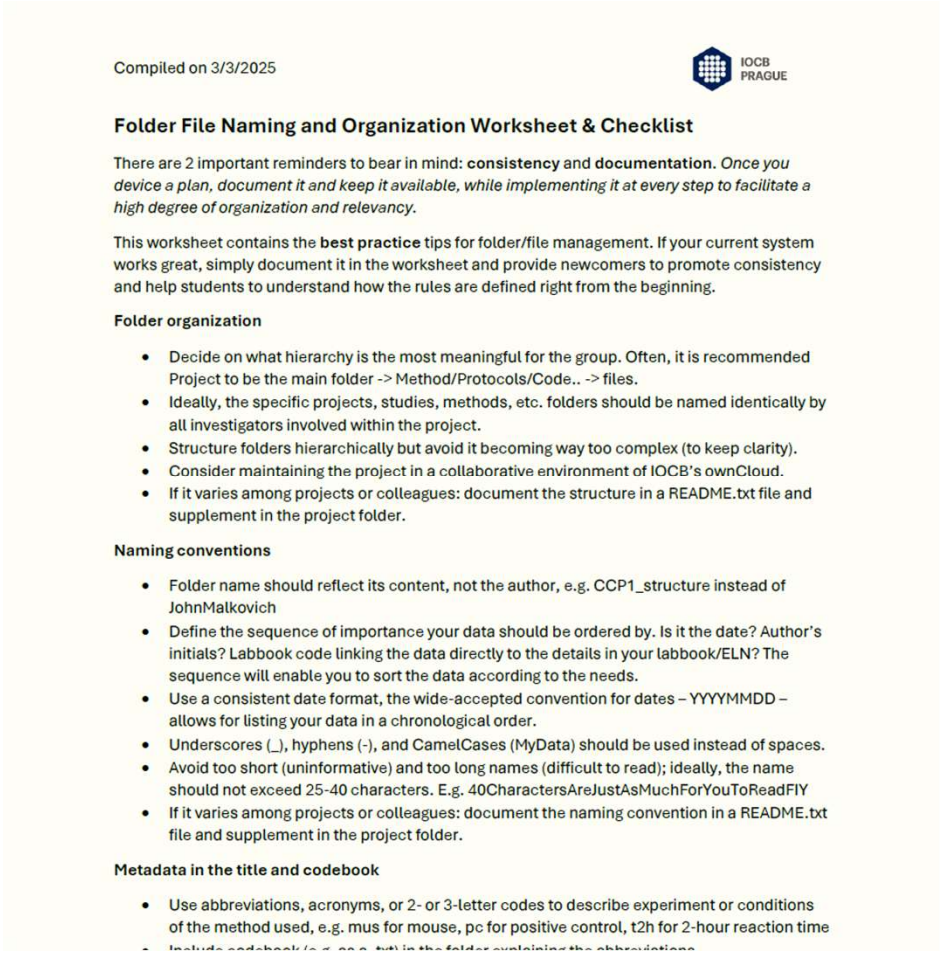
30th April – IOCB ELN Strategic meeting

ASEP: depositing a dataset.

- https://asep.lib.cas.cz/arl-cav/cs/prihlaseni/?opal=myasep&
- Datasets deposit agreement and self-test questions -> in 3. Agreement and Licensing

RDM assignments

File-Folder Organization



Compiled on 3/3/2025

IOCB PRAGUE

Folder File Naming and Organization Worksheet & Checklist

There are 2 important reminders to bear in mind: **consistency** and **documentation**. *Once you device a plan, document it and keep it available, while implementing it at every step to facilitate a high degree of organization and relevancy.*

This worksheet contains the **best practice** tips for folder/file management. If your current system works great, simply document it in the worksheet and provide newcomers to promote consistency and help students to understand how the rules are defined right from the beginning.

Folder organization

- Decide on what hierarchy is the most meaningful for the group. Often, it is recommended Project to be the main folder -> Method/Protocols/Code.. -> files.
- Ideally, the specific projects, studies, methods, etc. folders should be named identically by all investigators involved within the project.
- Structure folders hierarchically but avoid it becoming way too complex (to keep clarity).
- Consider maintaining the project in a collaborative environment of IOCB's ownCloud.
- If it varies among projects or colleagues: document the structure in a README.txt file and supplement in the project folder.

Naming conventions

- Folder name should reflect its content, not the author, e.g. CCP1_structure instead of JohnMalkovich
- Define the sequence of importance your data should be ordered by. Is it the date? Author's initials? Labbook code linking the data directly to the details in your labbook/ELN? The sequence will enable you to sort the data according to the needs.
- Use a consistent date format, the wide-accepted convention for dates – YYYYMMDD – allows for listing your data in a chronological order.
- Underscores (_), hyphens (-), and CamelCases (MyData) should be used instead of spaces.
- Avoid too short (uninformative) and too long names (difficult to read); ideally, the name should not exceed 25-40 characters. E.g. 40CharactersAreJustAsMuchForYouToReadFIY
- If it varies among projects or colleagues: document the naming convention in a README.txt file and supplement in the project folder.

Metadata in the title and codebook

- Use abbreviations, acronyms, or 2- or 3-letter codes to describe experiment or conditions of the method used, e.g. mus for mouse, pc for positive control, t2h for 2-hour reaction time
- Include codebook (e.g. ..co.e..txt) in the folder explaining the abbreviations

Have you all seen the document?

Did you find it helpful in the practical way?

Did you get support in the implementation process? What was the feedback in your group?

Any feedback is helpful as to know whether the time going to the work is worth investing.

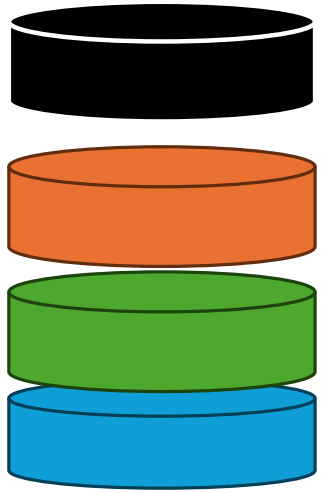
Metadata and documentation

Mandatory metadata sections in repositories

Readme documentation

Identifiers linked to an ELN record

File names-embedded metadata



✓

Codebook

Naming conventions

General information for a project/study

Information required to repeat an experiment

Dataset description

Readme files

- Readme.txt documentation
- Create readme files for logical “clusters” of related files / data. *Is it a project? Study within a project? Dataset cluster? Single dataset?*
- Name the readme in the way that it is easily associated with the data file(s) it describes.
- Format multiple readme files of the same type identically. Present the information in the same order, using the same terminology.
- Use standardized vocabulary when possible. Chemical Terminology, Ontology for Biomedical Investigations
- Good way is to explore community standards (e.g. domain-specific repository, experiment-specific metadata standard MIBBI)

Readme files

- Worksheet on readme files
- Feedback for your template available

Metadata tools

- RO-crate – data collection with automatically updated JSON format metadata with basic description (author, date, sequence of data entry)

Thank you for listening.

- IOCB strategy for sharing data via repositories.
- If you are unsure about the topic of repositories – please let me know.
- We already have a copy of an onsite installed DSW – “customization” phase.
- Drop-in session in **A2.46** today anytime between **2-3 PM**.