Lesson plan on data organisation

Justine Vandendorpe (ORCID: 0000-0002-9421-8582),[1*] Sophie Boße (ORCID: 0009-0002-6461-8291)[2]

[1]Data Science and Services, ZB MED - Information Centre for Life Sciences, Cologne, NRW, Germany
[2]Open Science, ZB MED - Information Centre for Life Sciences, Cologne, NRW, Germany

*Address correspondence to Justine Vandendorpe, vandendorpe@zbmed.de

Present address: Justine Vandendorpe, Data Science and Services, ZB MED – Information Centre for Life Sciences, Gleueler Str. 60, 50931 Cologne, NRW, Germany.

About this lesson plan

This lesson plan on data organisation was written in preparation for an interactive workshop on data organisation (see related slides here). The intended learning outcomes are largely based on Engelhardt *et al.* 2022 and Petersen *et al.* 2023 and you can choose them according to the level of your audience. The methods of engagement are presented in tables (with additional information below the table where necessary) and are largely based on Biernacka *et al.* 2020. The content of the lesson plan is based on Engelhardt *et al.* 2020, Biernacka *et al*. 2020, previous training and other sources (see Sources at the end of the document).

Intended Learning Outcomes (ILOs)

1. Explain the importance of a structured approach (basic).
2. Describe what a file naming convention is (basic).
3. List criteria for a good naming convention (basic).
4. Critically evaluate file naming and naming conventions (intermediate).
5. Develop naming conventions independently and apply them to own and shared research data (intermediate).
6. Apply supporting tools (e.g. renaming tools) to own files (intermediate).
7. Explain the purpose and use of versioning (basic).
8. Identify version control methods and tools (intermediate).
9. Apply simple versioning methods to your own research data (advanced).
10. Describe the general characteristics of an efficient folder structure (basic).
11. Explain why raw data should be stored separately (basic).
12. Critically evaluate different folder structures for files (intermediate).
13. Independently design a folder structure for own and collaborative research data (intermediate).
14. Explain the benefits and drawbacks of a clear and understandable folder structure (basic).
15. Explain tag-based file organisation (intermediate).

16. Explain benefits and drawbacks of a tag-based file organisation (intermediate).
17. Independently document your naming convention and folder structure (intermediate).
18. Organise data in a spreadsheet (basic).

Content

Motivation

Structured approach (ILO 1, Presentation slides 5-6)

Engagement method: ideas out loud [Biernacka *et al.* 2020]	
Learning target: ● Activation ● Query knowledge	**Description:** the teacher asks why we need a structured organisation of data and the participants call out answers without being asked directly. There is no particular order for the answers.
Duration: 5 min.	**Required materials:** Slide 5

Keeping the data and code layer of our project clean and organised saves us a lot of time and frustration [de Kok 2018]. It also ensures machine readability and helps to give context to our research: what, how and why things were done remains understandable even years later [Biernacka *et al.* 2020]. A structured approach also promotes reproducibility [Jacob *et al.* 2022] and avoids duplication of effort [Biernacka *et al.* 2020]. Of course, it also helps with all aspects of data organisation: it helps to easily identify the current state, to make naming conventions known to other researchers (including yourself), and to prevent data loss through overwriting or accidental deletion [Biernacka *et al.* 2020]. A structured approach also facilitates collaboration [Biernacka *et al.* 2020] and enables a traceable storage structure [Voigt *et al.* 2022]. It also supports data discovery and reuse by making it easier to search for data, finding it faster, and allowing other researchers to work with it. Overall, a structured approach is the foundation of your data-related work [Library Carpentry n.d.] and leads to more efficient work [Biernacka *et al.* 2020]

5S methodology (Presentation slide 7)

"5S" (2024) is a workplace organisation method that uses a list of five Japanese words translated into English as: sort, set in order, shine, standardise and sustain. In the context of organising research data (see Glossary), 'sort' would refer to deleting unnecessary files. 'Set in order' would refer to developing and documenting naming conventions and folder structures. 'Shine' would refer to following conventions and developing routines. 'Standardise' would refer to documenting rules and responsibilities and developing best practices and standard operating procedures (SOPs). And 'sustain' would refer to regularly checking that rules are being followed and making improvements where necessary (Fig. 1) [Assmann *et al.* 2022, November].

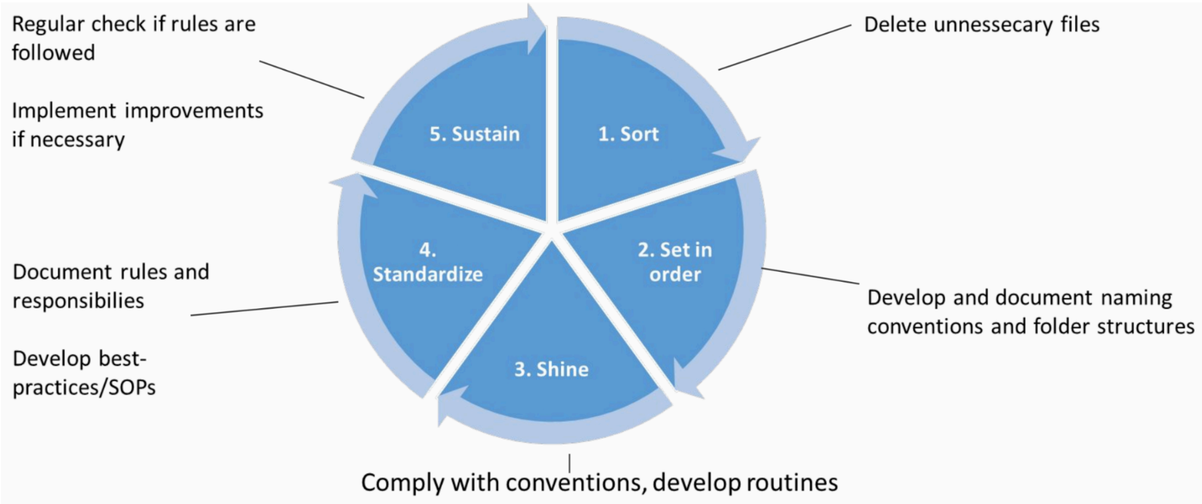

Fig. 1. 5S methodology [Assmann *et al.* 2022, November].

Further resources

- The 5S Methodology in Research Data Management - Article here [Lang *et al.* 2021].
- Coffee Lecture about 5S Data Organisation (german) here [Forschungsdatenmanagement Thüringen (TKFDM) 2021]

File naming

File naming convention (ILO 2, Presentation slide 9)

In order to maximise access to your data, to stay organised and to identify your files quickly, files and folders should be named in a meaningful and systematic way [LMA RDMWG 2024a, Rehwald *et al.* 2022]. A file naming convention provides a framework for naming your files and/or folders in a way that describes what they contain and how they relate to other files. This framework will help you, your future self, and others in a shared or collaborative group file-sharing environment to navigate your work more easily [LMA RDMWG 2024a].

Thus, within your research group, we recommend [Biernacka *et al.* 2020, Bobrov *et al.* 2021, Bres *et al.* 2022] that you:

1. Adopt a naming convention for files and folders (see ILO 5).
2. Document your file and folder naming convention (see ILO 17).
3. Stay consistent: the naming convention should be chosen in advance to ensure that it can be systematically followed and contains the same information (such as date and time) in the same order (e.g. yyyy-mm-dd) [Biernacka *et al.* 2020].

Criteria for a good naming convention (ILO 3, Presentation slide 10)

Avoid automatically generated names (e.g. from digital cameras) as they can lead to conflicting names due to repetition [Biernacka *et al.* 2020]. A good naming convention produces file names that are human-readable, machine-readable, and play well with your system's default ordering [Goldman 2020a].

Human-readability

File names should be descriptive and provide just enough contextual information to establish a link to a particular experiment or data collection [Bobrov *et al.* 2021, LMA RDMWG 2024a]. To achieve this, you should choose names that reflect the content and are unique [Lindstädt *et al.* 2019].

Machine-readability

In some operating systems, long file paths can cause technical problems. Therefore, file names should be as long as necessary and as short as possible to keep them concise and readable on any operating system. It is recommended to limit file names to ≤ 32 characters (32CharactersLooksExactlyLikeThis.txt). Avoid using special characters (e.g. {}[]<>()* % # ' ; " , : ? ! & @ $ ~), umlauts (ä, ö, ü, ß,...) or spaces [Biernacka *et al.* 2020]. Periods should only be used before version numbers and file extensions, which should be preserved from the system (e.g. .ERL, .CSV, .TIF) [Lindstädt *et al.* 2019]. You can use underscores (_), hyphens (-), or CamelCase instead to make file names both human- and machine-readable [LMA RDMWG 2024a].

Play well with default ordering

The computer organises files by name, character by character. To browse your files easily, you should choose names that can be sorted alphabetically, numerically or chronologically to ensure that the files appear in a logical order. If you want chronological order, start with a date in ISO 8601 format (YYYY-MM-DD or YYYYMMDD) [Briney 2020, Lindstädt *et al.* 2019]. When using sequential numbering, make sure to use leading zeros. For a sequence of 1-10: 01-10 and for a sequence of 1-100: 001-010-100. Scalability should be taken into account (e.g. if a two-digit file number is chosen, the number of files is limited to 99) [Biernacka *et al.* 2020].

Name components that are already part of the folder name do not have to be repeated in the file names [Biernacka *et al.* 2020]. Also consider the system under which the file is stored for later access and retrieval of the data.

Evaluate file naming and naming conventions (ILO 4, Presentation slides 11-14)

Below are some examples of file names that are human-readable (if you know the code/abbreviations), machine-readable, and properly sortable [Bres *et al.* 2022]:

- 2016-01-04_ProjectA_Ex1Test1_SmithE_v1.0.xlsx
- 2000_USNM_379221_01.tiff
- USNM_379221_01.tiff

Now look at these file names [Bres *et al.* 2022], they need to be improved.

- Test data 2016.xlsx
    - Replace whitespaces
    - Give more information: what is the test data?
    - Is this the only test data from 2016? If not, add month and/or day
    - Put the date first if you want to sort the files chronologically

- Meeting notes Jan 17.doc
    - Replace whitespaces
    - Give more information: what meeting was it?
    - Put the date first, in the format 2017-01, if you want to sort the files chronologically
- Notes Eric.txt
    - Replace whitespaces
    - Give more information: notes about what?

You can see some more examples of bad and good file names in Figure 2. The left part (i.e. bad names) shows how difficult it can be to find files that belong together if you don't think about the default ordering when naming your files: different files of the same experiment (July 2019) are separated by files of another experiment.

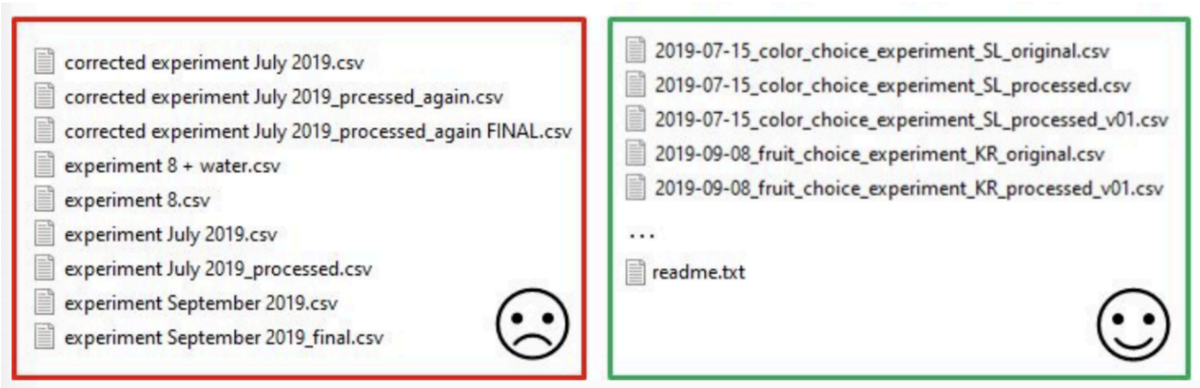

Fig. 2. Examples of bad (left) and good (right) file names [Rehwald *et al.* 2022].

Engagement method: exercise [Biernacka et al. 2020, Bres *et al.* 2022]	
Learning target: • Activation • Apply knowledge	**Description:** the teacher shows a list of file names and asks the participants to mark which of these examples follow a good naming convention.
Duration: 7 min.	**Required materials:** • List of file names (see below). • Technical solution: zoom poll (online), zoom annotation function (online), mentimeter (online or in person) or just microphones to discuss the names verbally (online, in person without microphone). • Slides 13 and 14.

List of file names:

- 4hr3nofnf3w49389utz304.mp3 (A: no)
- Projekt001_Probe045_MassSpec_20200824.csv (A: yes)
- Workshop_RDM.pdf (A: no/yes)
- Probe1 3004 Britta+Olga new edited!corrected.csv (A: no)
- 2021-05-18_US_NoDirtyDishesDay.pptx (A: yes)
- Olga_170413_probe17k.tiff (A: no/yes)
- Naturepaper karl britta james finished!.doc (A: no)
- Vm4520132Schmidt.pdf (A: no)
- 647749157.pdf (A: no)

- 170413_sample17k_olga.tiff (A: yes?)
- Naturepaper+karl+britta+james &nal.doc (A: no)
- Olga170413sample17k.tiff (A: no/yes)
- Krst_765_spct_1203.csv (A: no/yes)
- Naturepaper+karl+britta+james finished! reworked.doc (A: no)
- Cristal_765_spectr_20161203.tif (A: yes?)
- Nature_karlbrittajames_endendversion.doc (A: no)
- 28q8QGIHKwrRw.pdf (A: no)
- Conference_Digital_Science.pdf (A: no/yes)

Choose/create a file and folder naming convention (ILO 5, Presentation slides 15 to 22)

The following seven steps can help you create a naming convention for your files and/or folders [Briney 2020].

Step 1: Identify the group of files to which the naming convention will apply

First, you need to think about what related files you are working with, and what set of files your naming convention will cover [Briney 2020, LMA RDMWG 2024a]. Sometimes it is necessary to use different naming conventions for different sets of files [Briney 2020]. Now check if there are already established file naming conventions in your discipline or group that you can use or adapt for your file set [Goldman 2020a] For example, a file naming convention for FASTQ files can be found here.

- **Example:** this convention will apply to all of my microscopy files, including the raw and processed image data from both microscope A and microscope B [Adelaide Microscopy n.d.].

Step 2: Identify Metadata

Now decide what information you need in the file name to make it easy to find a particular file [LMA RDMWG 2024a]. Ideally, choose three pieces of metadata (see Glossary) and avoid using more than five. This metadata should be enough for you to visually scan the file names and easily understand what is in each one [Briney 2020].

If you find that you are encoding a large amount of metadata in the file names, you should consider storing this metadata in a master spreadsheet with your data for future reference [LMA RDMWG 2024a].

Below are some examples of metadata you might consider including in your file names [Biernacka *et al.* 2020, Lindstädt *et al.* 2019, LMA RDMWG 2024a]:

1. Content
2. Creator/person, researcher name/initials
3. Name of the team/research group
4. Funding body
5. Data type
6. Project number, ID, or part
7. Experiment name, acronym, or ID

8. Experimental conditions
9. Date or date range of the experiment
    a. Date of creation
    b. Date of processing
    c. Date of publication
10. Sample ID
11. Lab name, location
12. Software name, data collection method
13. Version number (see Versioning section), status

For other possible metadata or attributes, including examples, see Figure 3 [Adelaide Microscopy n.d.].

Examples of Specimen Attributes

Attribute	Example
Time	age (weeks, years); geologic period/epoch/era
Intervention or Procedure	drug vs. placebo, quenching vs. tempering
Location	anatomical (brain, heart); climatic (temperate, tropical); ecosystem (freshwater, marine)
Species (inc. subspecies or strains)	rodent vs. human; white- vs. black-backed magpie; C57BL/6 vs. BALB/c mice
Experimental	microscope settings (e.g. 10x vs. 20x magnification); staining media (safranin vs. iodine)
Function	load vs. non-load bearing; reproductive vs. vegetative
Phenotypic traits	Sex (male vs. female); mass, height or volume

Fig. 3: Examples of specimen attributes [Adelaide Microscopy n.d.].

- **Example 1:** For my files, I want to know the date, the sample ID and the image number for this sample on this date [Briney 2020].
- **Example 2:** For my images I want to know the genetic strain of the mouse, the treatment group, the sample number and the age [Adelaide Microscopy n.d.].

Step 3: Abbreviate or encode metadata

If any of the metadata from step 2 is described by a lot of text, decide what abbreviated information to keep to avoid too long file names. If any of the metadata you chose has regular categories, you can standardise the categories and/or replace them with 2- or 3-letter codes. Make sure you document these codes [Briney 2020, LMA RDMWG 2024a].

- **Example 1:** the sample ID uses a code consisting of: a 2-letter project abbreviation (project 1 = P1, project 2 = P2); a 3-letter species abbreviation (mouse = "MUS", fruit fly = "DRS"); and a 3-digit sample ID (assigned in a laboratory notebook) [LMA RDMWG 2024a].
- **Example 2:** genetic strain will use a 3-letter abbreviation (C57BL/6: "BL6", BALB/c: "BaC"), age in weeks will use a 3-letter code of the form "w##" (14 weeks old: "w14"; 16 weeks old: "w16") [Adelaide Microscopy n.d.].

Step 4: Select an order for the metadata in the file name

The computer arranges files by name, character by character [LMA RDMWG 2024a]. So think about how you want to sort and search your files to decide which metadata should appear at the beginning of the filename [Briney 2020]. Make sure you use the default order (alphabetical, numerical or chronological), so for example, if the date is most important, put it at the beginning so that dates are sorted chronologically [Briney 2020].

- **Example 1:** my sample ID is most important, so I will list it first, followed by the date, then the image number [Briney 2020].
- **Example 2:** the genetic strain is the most important attribute, followed by the treatment, the number of samples and the age of the mouse [Adelaide Microscopy n.d.].

Step 5: Deliberately separate metadata elements

Many computer systems cannot handle spaces in file names [Briney 2020, LMA RDMWG 2024a]. Think about what characters you will use to separate each piece of metadata and, if necessary, different words within a metadata element. To make file names both machine- and human-readable, use hyphens (e.g. file-name.xxx), underscores (e.g. file_name.xxx), or capitalise the first letter of each word, which is called CamelCase (e.g. FileName.xxx) instead [Briney 2020,LMA RDMWG 2024a]. Make sure you avoid special characters such as ~ ! @ # $ % ^ & * ( ) ` ; : < > ? . , [] { } ' " | LMA RDMWG 2024a] and umlauts [ä,ö,ü,ß,...].

- **Example 1:** I will use underscores to separate metadata and hyphens between parts of my sample ID [Briney 2020].
- **Example 2:** a unique sample ID will use a code consisting of strain, treatment and specimen number separated by hyphens ("-"). Underscores ("_") separate the remaining attributes [Adelaide Microscopy n.d.].

Step 6: Use versioning

Do you need to track different versions of each file? You can track versions of a file by adding version information to the end of the file name. Consider using a version number (e.g. "v1.0"), a version date (e.g. YYYY-MM-DD), or the status of your workflow (e.g. "raw", "processed", "composite") [Briney 2020]. For more information on versioning, see the Versioning section below.

- **Example 1:** as each image passes through my analysis workflow, I will append the version type to the end of the file name (e.g. "_raw", "_processed", and "_composite") [Briney 2020].
- **Example 2:** when images are processed in the analysis workflow, the version type (e.g. raw, processed, binary) is inserted within the file name (e.g. "_raw_", "_proc_", "_bin_") and immediately before the image number [Adelaide Microscopy n.d.].

Step 7: Document your naming convention (ILO 17)

Naming conventions should be documented so that others in your laboratory or department can follow this standard. Document the naming conventions in a Data Management Plan (DMP) (see Glossary), Standard Operating Procedures (SOPs) or a README file and keep it with your files. If the file is moved or shared, users will be able to identify the file by its filename [Biernacka *et al.* 2020, Briney 2020, LMA RDMWG 2024a]. Think about what information others need to understand and follow your naming conventions. Remember that choices, such as abbreviations, that make sense now may not be so obvious in 6 months' time. Consider including the following in your documentation (see Glossary): information about the author and the project, data types to which the conventions apply, naming structure, data format, explanation of metadata elements included, list and explanation of abbreviations and codes, examples of file names, location of files [Adelaide Microscopy n.d., Biernacka *et al.* 2020, Bres *et al.* 2023]. It may be useful to document the file naming convention together with the documentation of the folder structure (see "Document your folder structure").

- **Example 1:** my file naming convention is "SA-MPL-EID_YYYYYMMDD_###_status.tif". Examples are "P1-MUS-023_20200229_051_raw.tif" and "P2-DRS-285_20191031_062_composite.tif" [Briney 2020].
- **Example 2:** my file naming convention is strain-treatment-specimenNumber_age_version_imageNumber.fileFormat (with age in weeks). Examples are BL6-OVX-03_w18_raw_0001.bmp and BaC-CTL-03_w22_raw_0001.bmp [Adelaide Microscopy n.d.]. → Find the complete documentation here.
- **Example 3:** YYYY-MM-DD_JV_ProjectID_ExperimentID where IDs are linked to a table of data documentation such as metadata [Bobrov *et al.* 2021].
- **Example 4:** [date]_[title]_[creator]_[status]_[version] [Rehwald *et al.* 2022].

Examples of detailed READMEs describing file naming conventions for biological, geological and plant science microscopy data can be found here. Figure 4 shows an example of documenting multiple naming conventions for different data types. A template README file for documentation of naming conventions can be found here.

Data type	Data format	Unique properties* (& their abbreviations)	Version	Name draft**
Microscope Image	.tiff	• Date (YYYYMMDD) • author (Name Initials, eg. Max Exampleton = ME) • Image description • Microscope settings (x10, x20, x40 magnification)	No	YYYYMMDD_NameInitials_Animal_Tissue_Protein_Antibody_Magnification.tiff eg. 20211116_ME_LrpKO_Brain_GFAP_Cy3_x40.tiff
Protocol	.txt	• Date (YYYY-MM-DD) • Author (Name initials, eg. Max Exampleton = ME) • Method description (IHC, WB, ELISA)	Yes	YYYY-MM-DD_NameInitials_Method_Version.txt eg. 2021-11-16_ME_IHC_v1.txt

* Choose properties that are useful when searching for a file

** Put the properties in the order you want your files to be sorted in. For example, if you want to sort via date, the date should apear first in the file name: YYYYMMDD_exp001.xlsx

Based on: Data Management: File Organization by Data Management Services. Copyright © 2021-10-19 MIT, CC-BY 4.0 31

Fig. 4. Naming conventions - Do it yourself [Bres *et al.* 2023].

Develop a naming convention (ILO 5, Presentation slide 23)

Engagement method: exercise	
Learning target: ● Activation ● Apply knowledge	**Description:** the teacher asks the participants to come up with a naming convention for their own files and give some examples.
Duration: 10 min.	**Required materials:** ● Pen and paper or PC with a text editor. ● This worksheet (adapted worksheet for microscope data here) and/or this checklist may help. ● Slide 23.

Tools for simultaneous renaming (ILO 6, Presentation slide 24)

Sometimes it is necessary to rename multiple files at once, for example to change the automatically generated names from your digital camera or other software in one step, or to remove or replace spaces or other special characters from multiple file names in one operation [Biernacka *et al.* 2020]. There are various tools for such simultaneous file renaming for different operating systems, which you can find in Table 1.

Table 1: Tools for renaming files simultaneously.

Operating system	**Renaming tool**
Multiple OS	Adobe Bridge jExifToolGUI
Linux	Gnome Commander GPRename
Mac	ExifRenamer NameChanger Renamer 6
Unix	mv command
Windows	Advanced Renamer Altap Salamander Ant Renamer Bulk Rename Utility ExifToolGUI Rename-It Total Commander WildRename

Further resources

- File naming examples in Table 1 of the article here [Briney *et al.* 2020].
- Information and steps for creating naming conventions here [LMA RDMWG 2024a].

- Information about file naming in RDM Guide here [RDM Guide n.d.].
- Information on File Naming and Folder Hierarchy here [MIT Libraries n.d.a].
- Information and examples for microscopy data here [Adelaide Microscopy 2024].
- Worksheet in File Naming Conventions here [Briney 2020].
- Worksheet for Naming and Organizing Files here [MIT Libraries 2020].
- Checklist for FIle Naming Conventions here [Goldman 2020c].
- A detailed documentation of a File Naming Convention here [TILS 2009].

Versioning

Versioning or version control is the practice of tracking and managing changes to a file or set of files over time so that you can later retrieve specific versions.

We recommend that you meet with project partners to decide how versioning will be carried out, how version changes will be documented, and how a version change will be defined [Bres *et al.* 2022].

Purpose and use of versioning (ILO 7, Presentation slide 27)

Versioning helps you to keep a complete long-term change history of each file by tracking, tracing and annotating your steps (i.e. changes made to the file(s)) and also allows you to go back one step. Versioning also allows you to keep multiple versions of each file, and to create new versions of the same file - or even new results - by incorporating new data and/or changes to a file's structure; this is particularly important in the case of software. Versioning also supports debugging in software. Overall, versioning makes your research easier to understand [Biernacka *et al.* 2020, Bres *et al.* 2022, Di Russo 2020].

Version control methods and tools (ILO 8, Presentation slide 28)

Versioning can be done in the file name (see semantic versioning below), in the data (e.g. in the header or a column for comments), in a text file (e.g. in a README file), or using a version control system (VCS). A VCS is a software tool that helps to manage changes to one or more files over time. Examples of VCSs include Git (e.g. Bitbucket, GitHub, GitLab) and Apache Subversion [Git n.d.]. For collaborative document and storage locations (e.g. wiki, Google Docs, cloud), versioning is available in situ [Biernacka *et al.* 2020] (i.e. within the document/storage location and in real-time).

Apply versioning methods (ILO 9, Presentations slides 29 to 31)

Manual file versioning can be done using semantic versioning. You can do this by adding a "v" to the end of each file name, followed by a maximum of three numbers separated by a period (note that these are the only periods allowed in a file name other than the one before the extension). The first number is called MAJOR and indicates important changes. The second number is called MINOR and indicates less drastic changes. The third number is called PATCH, and is mainly used by software developers to indicate bug fixes. Examples of semantic versioning would look like this [Bobrov *et al.* 2021, Bres *et al.* 2022]: Filename_vMAJOR.MINOR.PATCH.FileExtension

- Ex1Test1_SmithE_v1.0.0.xlsx

- Ex1Test1_SmithE_v1.2.5.xlsx
- Ex1Test1_SmithE_v2.1.1.xlsx

If you decide to use manual file versioning, it is recommended that you use a version control table (a version control table template from the University of Sydney Library can be downloaded here). It is also recommended that you assign responsibilities for completing files, store milestone versions, and store obsolete versions separately after backup. How many versions of a file will be kept, which versions (e.g. major versions instead of minor versions (version 2.0 but not 2.1)), for how long, and how the versions will be organised need to be decided in advance [Biernacka *et al.* 2020], ideally with project partners.

<table>
<tr><td colspan="2">Engagement method: exercise [Bres et al. 2022]</td></tr>
<tr><td>Learning target:
Activation
Apply knowledge</td><td>Description: the teacher shows a file name and the participants try to improve it.</td></tr>
<tr><td>Duration: 3 min.</td><td>Required materials:
File name: Final FINAL last version.docx.
Technical solution: zoom poll (online), zoom annotation function (online), mentimeter (online or in person) or just microphones to discuss the names verbally (online, in person without microphone).
Slide 31.</td></tr>
</table>

Further resources

- Comparison of GitHub and GitLab here [Vaughan-Nichols 2022]
- Blogpost on Why and how to give up GitHub here [Software Freedom Conservancy]
- Git [Haenel & Plenz 2012]
- Recommendations on Data Versioning [Klump *et al.* 2024]
- Versionskontrolle Mit Subversion [Pilato *et al.* 2009]

Folder structure

To make it easier to find files, especially if you have a lot of data, you should avoid a chaotic or alphabetical approach to storing data. Instead, a proper folder structure is a hierarchical arrangement in which folders are created to make it easier to find data [Biernacka *et al.* 2020]. A typical hierarchical folder structure has a root folder and several levels of subfolders, as shown in Figure 5. A carefully planned folder structure, with understandable folder names and an intuitive design, is the foundation of good data organisation. The folder structure provides an overview of what information can be found where, enabling both current and future contributors to understand what files have been produced in the project [Mičetić *et al.* n.d.].

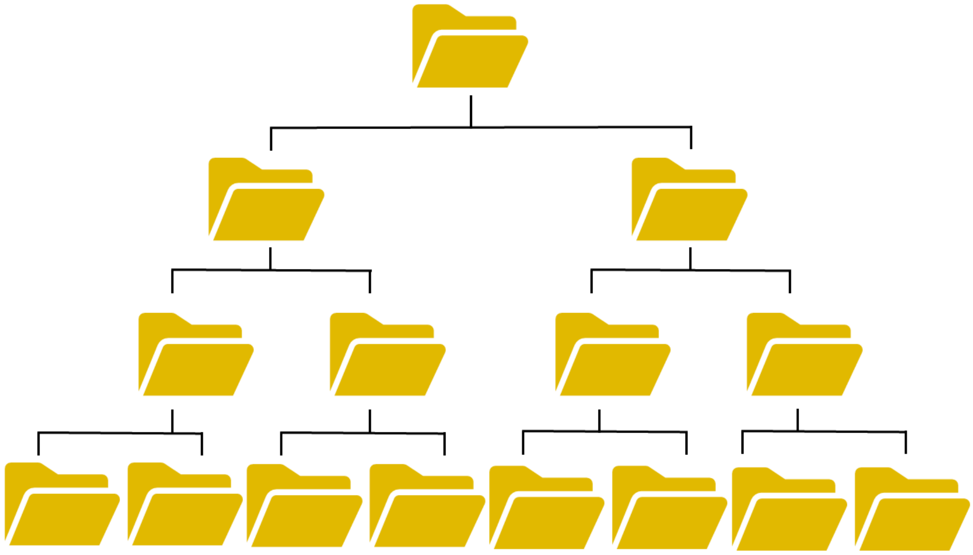

Fig. 5: hierarchical folder structure [Boße *et al.* 2024].

General characteristics of an efficient folder structure (ILO 10 and 11, Presentation slides 34 to 37)

An efficient folder structure allows "someone", perhaps your future self, to look at your files and immediately understand in detail what you have done and why [Goldman 2020b]. Therefore you should choose a folder structure that is hierarchical, clear, comprehensive, efficient and conclusive [Bres *et al.* 2022, Bres *et al.* 2023]. To make it clear and comprehensive for other team members, make sure the structure is self-explanatory and has intuitive navigation [Biernacka *et al.* 2020, Bobrov *et al.* 2021, Bres *et al.* 2022].
Short, meaningful folder names that follow a comprehensive naming convention make browsing a folder structure more efficient [Assmann *et al.* 2022, November; RDM Guide n.d.]. Sometimes it is a good idea to number the folders to ensure that they work well with the system's default order [Assmann *et al.* 2022, November]. For clarity, the folder structure should be identical on servers and local devices [Biernacka *et al.* 2020].

There is no one-size-fits-all solution: the optimal folder structure depends on the specific project requirements [Bres *et al.* 2023]. To make the structure easy to browse, do not make it too deep: use a maximum of 3 to 4 levels [Bres *et al.* 2022, Voigt *et al.* 2022]. In addition, if the folders are too large, it is difficult to find the right file in the folder: so limit your folders to a maximum of 10 items per folder [Bres *et al.* 2022].

There are several approaches to building your hierarchy. For some projects, it may be helpful to use a folder structure that follows the different parts or workflow of the project. This can support the step-by-step creation, analysis and publication of data [Biernacka *et al.* 2020, Bres *et al.* 2022]. You can also consider basing your hierarchy on functionalities, people involved, date or time period, data types, creation methods or processing steps [Bres *et al.* 2023]. Be careful to distinguish between [Schmid 2021, von der Dunk 2021]:

- Work vs. private material
- Own work vs. work of others (papers vs. literature)

- Research vs. administrative content
- Raw data (see Glossary) vs. processed data (see Glossary) vs. final data (see Glossary)
- Experiment vs. analysis
- Experimental runs/replicates (where appropriate)

You should avoid using generic "current stuff" folders. Also, be careful about creating researcher-specific folders within a project: folders are about the content, not the authors [Pasquier 2024]. If you use researcher-specific folders, external contributors will not be able to understand what data is stored in these folders. Use one folder per dataset, containing data and its description. If you have multiple datasets, the project information can be described in the parent folder [Rehwald *et al.* 2022].

Make sure you don't have overlapping categories, as you shouldn't have copies of files in different folders, since this can lead to confusion and make it difficult to keep track of different versions of the file [Goldman 2020b]. If you need to see a file in more than one folder, you can use shortcuts to the file instead. This allows you to keep a single reference file [RDM Guide n.d.]. In particular, make sure you have a 'raw data' folder for each type of data or experiment [RDM Guide n.d.]. It is important to store your raw data separately so that the original versions of the files or their documentation are preserved and the original files can be reconstructed [Biernacka *et al.* 2020].

Evaluate different folder structures for files (ILO 12, Presentation slides 38 to 41)

The requirements for a folder structure differ depending on whether it is used by a research team or an individual researcher.

Example of a folder structure for a researcher team

Figure 6 shows an example of a folder structure designed for a whole team of researchers. This structure distinguishes between the team's shared space and the project folders. The shared space contains all files related to the project team and its working structure. In addition, each research project has its own folder with its own subfolders, which can be customised to meet the needs of each individual project [Briney *et al.* 2020].

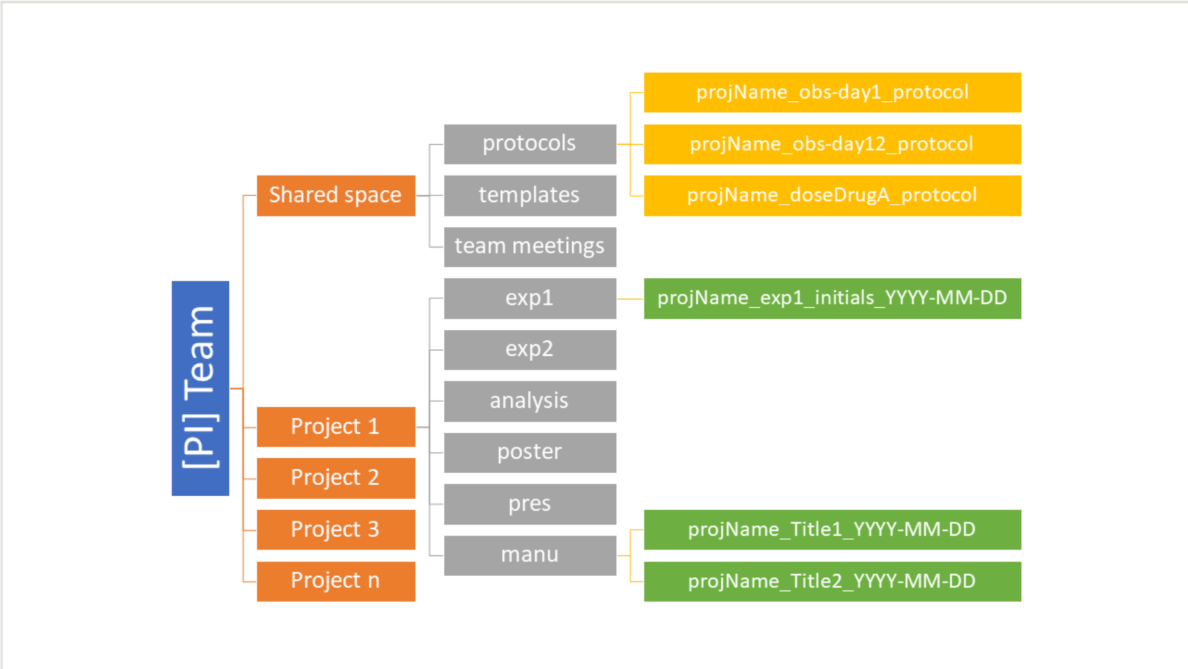

Fig. 6: Folder structure for a team of researchers [Briney *et al.* 2020].

Example of a folder structure for a single researcher

Figure 7 shows an example of a folder structure that might be suitable for an individual researcher who has a strict separation between their single research project and their private data. The research data folder is structured along the project workflow from raw data through data analysis to results. This structure can be extended to include additional folders for additional research projects if required [Von der Dunk 2021].

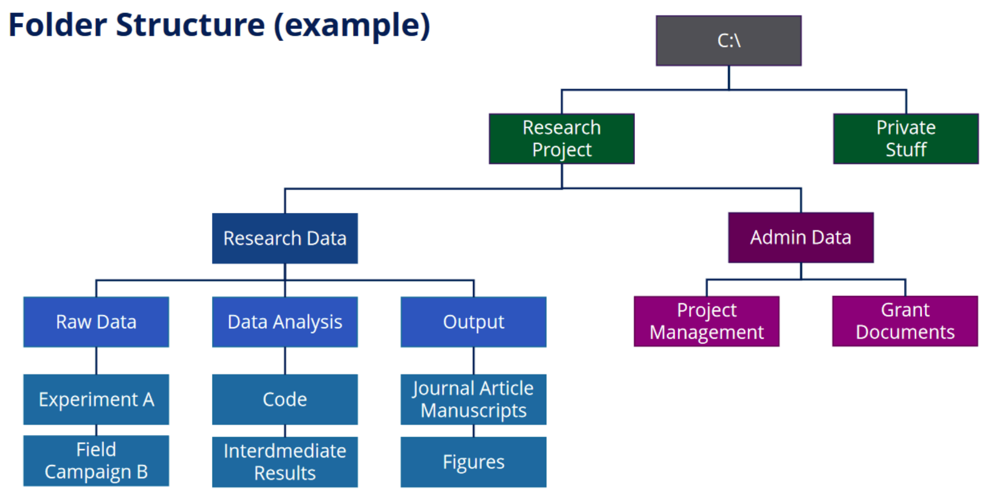

Fig. 7: Folder structure for a single researcher [Von der Dunk 2021].

Different approaches to store code and data

In contrast to the first example, the following two (Figures 8 and 9) differentiate between data and code at the same level, with raw data and processed data being separate subfolders of the root "data" folder.

- Project
    - Deliverable
        - 01_Data
            - Raw_data
            - Processed_data
        - 02_Code
        - 03_Protocols
        - 04_Manuscript
        - 05_Poster
    - Administrative information

Fig. 8: Folder structure.

- Project
    - Data
        - Raw_data
        - Processed_data
        - Documentation
    - Code
        - Src
        - Output
            - Plots
        - Documentation
    - Protocols
- Manuscripts
- Conference_reports
- Administrative_information

Fig.9: Folder structure [Bobrov *et al.* 2021].

The next example (Figure 10) introduces an approach that helps to keep track of the code-based analysis workflow by providing an additional pipeline folder for each code file. This makes it easy to see what the original data is, how the code should be executed, and how the data flows between the different code files. This folder structure can be downloaded here, and can be supplemented with other parent folders such as 'administrative' and 'paper' [de Kok 2018].

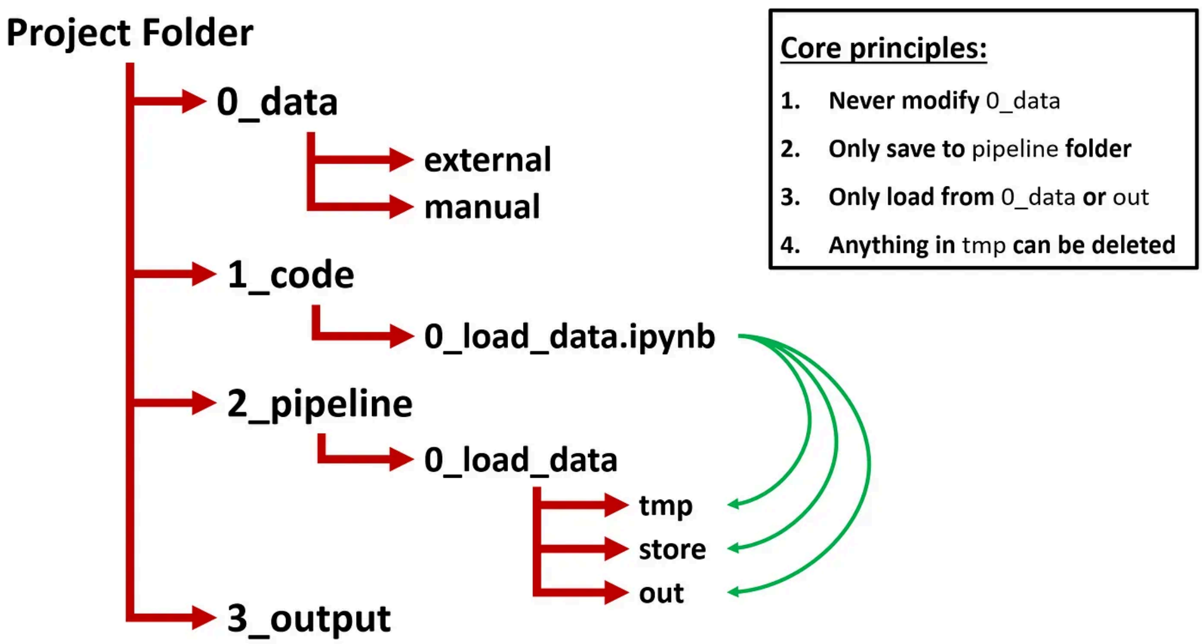

Fig. 10. Folder structure. 0_data: input data, either retrieved from external sources or created manually. 1_code: all your code files, starting with a number indicating the order of

execution. 2_pipeline: a separate subfolder for each code file contained in the 1_code folder, as all output generated by a code file ends up in its corresponding pipeline folder. Tmp: files you save for testing or other temporary reasons. Store: files that you save with the intention of loading them into the current code file. Out: files you save with the intention of loading them into a future code file. 3_output: all final output files to be included in the paper [de Kok 2018].

Engagement method: ideas out loud	
Learning target: ● Activation ● Apply knowledge	**Description:** the teacher asks the participants to think about things they could improve in their current folder structure and the participants call out answers without being asked directly. There is no particular order for the answers.
Duration: 3 min.	**Required materials:** Slide 42.

Design a folder structure (ILO 13)

Engagement method: exercise [Biernacka *et al.* 2020]	
Learning target: ● Activation ● Apply knowledge	**Description:** the teacher asks the participants to design a structure for their storage in the form of a directory tree. There are two options to choose from: 1. Create a structure for your own research data 2. Create a structure for collaborative research data.
Duration: 10 min.	**Required materials:** ● Pen and paper or PC with a text editor. ● This worksheet can help. ● Slide 43.

Benefits and drawbacks of a clear and understandable folder structure (ILO 14, Presentation slides 33, 34 and 44)

Benefits

Save time and nerves when searching for files

Organising files into systematically structured folders increases your productivity by making it easier to find files, group similar files together, and move files or groups of files [Goldman 2020b].

Improve workflow during the project

Standardised folder structures help maintain long-term organisation and record keeping.

A clear folder structure promotes good data management and makes it easy to identify which files need to be backed up [Goldman 2020b, MIT Libraries 2016]. You can distinguish between files to be finalised and final data or results by creating different subfolders for them. This allows subfolders to be treated as to-do lists to facilitate project and lab management [MIT Libraries 2016].

Collaboration made easy

Hierarchical folder structures are familiar and widely used and therefore support collaborative work. They are good at representing the nested structure of information [MIT Libraries 2016]. Once a folder structure is set up, it is easy to use and projects are easily understood by others [Boße *et al.* 2024, Goldman 2020b, MIT Libraries 2016]. In this way, a clear folder structure promotes the FAIR data principles and makes it easy to share your data [Boße *et al.* 2024].

Drawbacks

Difficulties in creating a standardised folder structure

Setting up a new folder structure can be time consuming [Boße *et al.* 2024, MIT Libraries 2016]. Especially in ongoing projects, there can be high transition costs to reorganise all the files. It can be difficult to find a good balance between breadth and depth when creating a new folder structure [MIT Libraries 2016].

Difficulties in using a standardised folder structure

The full effect of the folder structure is only achieved if the whole research group uses it consistently. This is sometimes difficult because different people may have a different 'understanding' of the data and therefore sort it differently [Boße *et al.* 2024]. In addition, too detailed a folder structure seems to be rather inconvenient for saving files, which may lead to less acceptance by project teams [Colomb *et al.* 2021]. The hierarchical folder structure categories are inherently discrete and can lead to forced separation of files [UC Merced Library n.d.].

Tag-based file organisation (ILO 15, Presentation slide 45)

Another approach to organising your files is to use embedded metadata called tags, which you add to each file [Carrano n.d.]. Tagging files basically works like using hashtags (#) in social media to tag and find information [UC Merced Library n.d.]. You can add one or more tags to your file, which will be included in the general information about your file, so that you can search for a file by its tag(s) using your computer's search functions [Carrano n.d.]. To do this, you need to decide on a set of tags to use at the beginning of the project and then use them consistently [MIT Libraries 2016, UC Merced Library n.d.]. Make sure you do not use too many similar tags, as this will make it impossible to find your file. For example, if some files are #cereals and some are #wheat but not #cereals, you will not find the #wheat file when searching for #cereals. Also be careful not to use too few tags, as your file is not unique, so you won't find the right file if all of them are tagged with a simple tag [UC Merced Library n.d.]. It may be helpful to use controlled vocabularies (see Glossary) or thesauri (see Glossary) established in your discipline to create your tag set.

How to add tags

Probably the easiest way to add tags to your files is through your operating system, as you don't need any additional tools. A guide to tagging files in MacOS can be found here. Windows allows you to add tags to at least images and native Microsoft Office documents [Onyimadu 2023]. For more information on tagging files in Windows, click here. In addition to manually creating tags in your operating system, there are some useful programs available to help you tag files of various formats.

- Tagsistant (Unix, free)
- Tabbles (Windows, free w/ limited features)
- TMSU (Unix, Windows, free)
- Adobe Bridge
- Adobe Acrobat (Guide here)
- ExifTool (Windows)
- jExifToolGUI (Mac/Linux/Windows)

Benefits and drawbacks of a tag-based approach (ILO 16, Presentation slide 46)

Compared to a hierarchical folder structure, a tag-based file organisation can be easier to set up. Especially when it comes to collaborative work, it is easier to combine than to merge different hierarchical folder structures. In addition, files can easily be assigned to more than one category without the risk of duplication and space limitations [MIT Libraries 2016, UC Merced Library n.d.].

On the other hand, this approach is less good at representing the structure of information. It has an increased risk of inconsistency, because if an item is not properly tagged when it is first acquired, it may be difficult to find that file later [MIT Libraries 2016, UC Merced Library n.d.]. This approach is not feasible for all file formats, as some do not allow embedded metadata tags to be added. Some file organisation tools create non-embedded tags, so when you export the files from the tools, you get the tags in separate metadata files. These need to be transferred separately with the files they describe [Carrano n.d.]. Not all metadata transfers well into new software programs [Macdonald & Smith 2013]. For this reason, embedded metadata of unsupported file formats, as well as non-embedded tags, can lead to loss of information when moving to another organisational tool or online system [Carrano n.d.].

Hybrid approach

If you want to combine the benefits of both the hierarchical folder structure and the tag-based approach, you can also add tags to your hierarchical folder structure [MIT Libraries 2016, Shanshan Ma 2010]. This hybrid approach takes some time to set up, save and tag files, but could make finding files easy and time-saving [Shanshan Ma 2010].

Recommendations for establishing a system (Presentation slide 47)

To maximise the benefits and minimise the drawbacks of a standardised folder structure, you should set up the new system carefully. Invest time in planning the folder structure before you start collecting data/writing your first line of code [de Kok 2018]. The more carefully you plan it, the easier it will be to navigate later [Biernacka *et al.* 2020]. To get maximum acceptance from your team, consider creating a system as a group and prioritising it for implementation [LMA RDMWG 2024b]. This is important as everyone needs to be able and willing to adhere to the system. So provide a method for easy adoption. Prepare a shared network with the folder hierarchy in place [LMA RDMWG 2024b]. Provide documentation (see "Document your naming convention and folder structure" below) with explanations, abbreviations and instructions for new contributors [Assmann *et al.* 2022, November; LMA RDMWG 2024b].

One option is to start small: create just one project/experiment and then expand from there [LMA RDMWG 2024b].

Document your folder structure (ILO 17, Presentation slide 47)

In order for everyone (including your future self) to be able to adhere to your folder structure or tag-based system, you should produce comprehensive documentation of it and make it easily available to all members of the team. You might consider documenting the folder structure together with the file naming convention in a DMP, SOP or README file [Biernacka *et al.* 2020, Briney 2020]. Relationships between elements (e.g. between code and the data file needed to run it), between the data file and associated documentation or metadata, or between multiple files should be documented as needed. Without such a file, it may be difficult to reconstruct these conventions after some years [Biernacka *et al.* 2020].

Think about creating README files for project-level documentation (information relevant to all files in the folder), as well as for file/folder-level documentation (information specific to each file/folder, Fig. 11). Information to include in the README can be found below [KU Leuven 2022]:

- The context in which the data was collected
- Origin of the data
- Content
- Information about data confidentiality
- Licences (see Glossary) and (legal) conditions for reuse

```
project/
  code/                 code needed to go from input files to final results
  data/                 raw and primary data (never edit!)
     raw_external/
     raw_internal/
     meta/
  doc/                  documentation of the study
  intermediate/         output files from intermediate analysis steps
  logs/                 logs from the different analysis steps
  notebooks/            notebooks that document your day-to-day work
  results/              output from workflows and analyses
     figures/
     reports/
     tables/
  scratch/              temporary files that can safely be deleted or lost
  README.txt            file and folder description
```

Fig. 11: README file documenting a folder structure [Mičetić *et al.* n.d.].

Further resources

- Checklist Directory Form here [Goldman 2020a].
- Worksheet for Naming and Organizing Files and Folders here [MIT Libraries 2020].
- Information on File Naming and Folder Hierarchy here [MIT Libraries n.d.a].
- Guide for Tagging and Finding Files here [MIT Libraries n.d.b].
- Comparison of hierarchical, tag-based and hybrid data organisation here [Shanshan Ma 2010].

Reusable Folder structures:

- GIN-Tonic: a digital shelf for your research files.
- Basic Folder Structure [Vukovic n.d.].
- Folder Structure generator [de Kok 2018].
- Template for research repositories [Colomb *et al.* 2020].
- simple Open Data template [de Plaa 2021].

Tidy data (Presentation slide 50)

The whole section draws heavily on Library Carpentry n.d. and Data Carpentry for Biologists n.d.a.

Tidy data is data that is well designed for working with computers. If you create tidy data as you collect it, it will be much easier to analyse later. One way to create tidy data is to use spreadsheets.

Organise data in a spreadsheet (ILO 18)

Using spreadsheet programs for data organisation (Presentation slides 51 to 55)

Spreadsheets are very useful graphical interfaces for designing data tables, getting an overview of the data, and performing very basic data quality control (see Glossary) functions. We can use them to enter and organise data, for some subsetting and sorting of data, and for statistics and plotting. Spreadsheet tools include the open-source (see Glossary) Apache OpenOffice Calc, Gnumeric and LibreOffice Calc, and the proprietary Apple Numbers, Google Sheets and Microsoft Excel.

<table>
<tr><td colspan="2">Engagement method: instant poll</td></tr>
<tr><td>Learning target: learn about participants' experiences of organising data in spreadsheets</td><td>Description: the participants answer a poll (see questions below) about their own experiences with organising data in spreadsheets. The trainers then briefly discuss the results.</td></tr>
<tr><td>Duration
● Explanation: 1 min.
● Execution of the poll: 3 min.
● Discussion of the results: 1 min.</td><td>Required materials:
● Instant poll tool.
● Slide 52.</td></tr>
</table>

- **Have you ever used spreadsheets in your work?** [Yes / No]
- **What kind of operations do you perform in spreadsheets?** [Multiple choice]

    - Data entry
    - Data organisation
    - Subsetting of data
    - Sorting of data
    - Statistics
    - Plotting
    - Quality control
- **What do you think spreadsheets are good for?** [Multiple choice]
    - Data entry
    - Data organisation
    - Subsetting of data
    - Sorting of data
    - Statistics
    - Plotting
    - Quality control

Spreadsheets can be very useful, but they can also be frustrating and sometimes even give us incorrect results.

Engagement method: ideas out loud [Biernacka *et al.* 2020]	
Learning target: ● Activation ● Query knowledge	**Description:** the teacher asks what the participants have accidentally done in a spreadsheet and/or if they have been frustrated by not being able to do something easily in spreadsheets. Participants call out answers without being asked directly. There is no particular order for the answers.
Duration: 5 min.	**Required materials:** Slide 53.

There are several problems with spreadsheets. Firstly, creating tables for reports in a spreadsheet is not optimal. Special formatting should be done in a word processor. Also, it can be very difficult, if not impossible, to retrace your steps (let alone someone else's), especially if your statistics or plots require more complex calculations. Finally, when doing calculations in a spreadsheet, it is easy to accidentally apply a slightly different formula to several adjacent cells.

A lot of real-world data isn't very tidy, mostly because most scientists aren't taught how to structure their data in a way that makes it easy to analyse.

Engagement method: improving messy data [Data Carpentry for Biologists n.d.b]	
Learning target: consolidation	**Description:** the participants have to improve messy data (see details below).
Duration: 10 min.	**Required materials:** ● Messy data (see below). ● Slide 55.

1. [Download]() a messy version of some data from the Portal project, including site, date, species identification, weight and sampling plot (within the site) for some small mammals.
2. Describe five things about this data that are not tidy.
3. Write answers to the following questions:
   - How could you correct the problems identified in 2?
   - Could this data be easily imported into a programming language or database in its current form?
   - Do you think it's a good idea to enter the data in this way and clean it up later, or to have a good data structure for analysis at the time of data entry? Why?

Formatting data tables in spreadsheets (Presentation slides 56 to 58)

The best formats (as well as software and interfaces) for data entry and analysis may be different. It is important to take this into account, and ideally to automate the conversion from one to the other.

To be able to reproduce your analyses, you need to keep the original dataset and create a new file with your cleaned or analysed data (see Glossary). You will also need to keep track of the cleaning and analysis steps, for example in another text file.

There are fundamental rules for structuring data in a spreadsheet. First, you should make your spreadsheet a rectangle: you should only have rows and columns (NO extra structure), you should put all your variables (i.e. what you are measuring) in columns (one column for one variable), and each observation in its own row (one row for one observation). Then, there should be only one value per cell, NOT multiple pieces of information combined in one cell (Fig. 12 and 13).

<table>
<tr><th colspan="4">RDM training</th></tr>
<tr><th>Date</th><th>Length (hours)</th><th>PGR|PDRA|other</th><th>Delivered by</th></tr>
<tr><td>4 Feb</td><td>1.5</td><td></td><td>GQ</td></tr>
<tr><td>7/8 Feb</td><td></td><td></td><td>GQ</td></tr>
<tr><td>20 Feb</td><td></td><td></td><td>GQ & DF</td></tr>
<tr><td>03/03/17</td><td>2</td><td>15|03|00</td><td>DF</td></tr>
<tr><td>04/03/17</td><td>2</td><td>30|0|0</td><td>DF</td></tr>
<tr><td>08/04/17</td><td>2</td><td>30|0|1</td><td>DF</td></tr>
<tr><td>26/05/17</td><td>2</td><td>27|0|0</td><td>DF</td></tr>
<tr><td>2 June?</td><td>2</td><td>24|02|00</td><td>DF</td></tr>
<tr><td>3 June?</td><td>1.5</td><td>12|07|04</td><td>DF</td></tr>
</table>

Fig. 12. Messy data [Library Carpentry n.d.].

Engagement method: one cell one value [Data Carpentry for Biologists n.d.b]

Learning target: consolidation	**Description:** the participants have to improve a data table (see details below).
Duration: 5 min.	**Required materials:** ● A messy data table (see below). ● Slides 57 and 58.

Table 2. Messy data to be improved [Data Carpentry for Biologists n.d.b].

Mass
26g
0.2kg

Table 3. Solution [Data Carpentry for Biologists n.d.b].

Mass	Unit
26	g
0.2	kg

RDM training					
Date	**Length (hours)**	**PGR**	**PDRA**	**other**	**Delivered by**
4 Feb	1.5				GQ
7/8 Feb					GQ
20 Feb					GQ & DF
03/03/17	2	15	3	0	DF
04/03/17	2	30	0	0	DF
08/04/17	2	30	0	1	DF
26/05/17	2	27	0	0	DF
2 June?	2	24	2	0	DF
3 June?	1.5	12	7	4	DF

Fig. 13. Tidy data [Library Carpentry n.d.].

You should also avoid duplicate chunks of data (Fig. 14) as they are difficult to update, more prone to error and take up more space. Instead, use one table for each category of data (Fig. 15), as you only need to make changes in one place and there is less repetitive typing. You should also leave the original / raw data as it is, and always make a copy before making any changes. Finally, you should record all the steps you take to clean your data in a plain text file, and export the cleaned data in an easy-to-read text file format such as CSV.

Family	Genus	Species	Plot	Abundance
Heteromyidae	Dipodomys	Spectabilis	1	2
Heteromyidae	Dipodomys	Spectabilis	2	7
Heteromyidae	Dipodomys	Spectabilis	3	5
Heteromyidae	Dipodomys	Spectabilis	4	3
Heteromyidae	Dipodomys	Ordii	1	5
Heteromyidae	Dipodomys	Ordii	2	9
Heteromyidae	Dipodomys	Ordii	3	12
Heteromyidae	Dipodomys	Ordii	4	11

Fig. 14. Bad: duplicate chunks of data [Data Carpentry for Biologists n.d.a].

SpeciesID	Plot	Abundance
disp	1	2
disp	2	7
disp	3	5
disp	4	3
dior	1	5
dior	2	9
dior	3	12
dior	4	11

SpeciesID	Family	Genus	Species
disp	Heteromyidae	Dipodomys	Spectabilis
dior	Heteromyidae	Dipodomys	Ordii

Fig. 15. Good: one table for each category of data [Data Carpentry for Biologists n.d.a].

Formatting problems/Common spreadsheet errors (Presentation slides 59 and 60)

There are common spreadsheet errors or formatting problems that are best avoided. Firstly, it is recommended to only have one table per sheet. If you create multiple tables within a spreadsheet (Fig. 16), you are making false associations between things for the computer, which sees each row as a single observation. You may also be using the same field name in several places, making it harder to clean up your data into a usable form. Then it is advisable to only have one tab per file. If you create separate tabs for each year, for example, you are more likely to inadvertently introduce inconsistencies into your data. Even if you manage to prevent all inconsistencies from creeping in, you will have to add an extra step before you can analyse the data, because you will have to combine the data into a single datatable. When entering data, before you create another tab or table, ask yourself if you could add another column to your original spreadsheet instead.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2		RDM training					Open access						
3		Date	Length (hours)	PGR|PDRA|other	Delivered by		Date	Len	Attendees	Delivered by			
4		12 Jan	1.5	45|0|0	FG		8 Jan	1.5 hours	20	FG			
5		7 Feb	2	38|0|0	GH		13 Jan	1 hour	21	JM			
6		4 Mar	2	43|3|0	GH		22 Jan	1 hour	35	JM			
7		6 Mar	1	21|7|0	GH		2 Feb	1.5 hours	36	JM		cancelled	
8		17 Mar	1.5	34|1|0	FG		3 Feb	1.5 hours	22	JM			
9		21 Mar	1	25|2|0	DQ		3 Feb	1 hours	30	JM			
10		23 Mar	2	32|10|0	FG		20 Feb	1.5 hours	36	FG			
11		19 Apr	1	34|0|0	GH		28 Feb	1.5 hours	28	JM			
12		30 Apr	1.5	37|0|0	FG		19 Mar	1.5 hours	33	FG			
13		4 Jun	1	45|0|0	GH		19 Mar	1 hour	39	JM			
14		12 Jun	2	36|0|0	DQ		4 Apr	1.5 hours	21	JM			
15		22 Jun	1.5	38|0|0	DQ		5 May	1.5 hours	25	JM			
16		25 Jun	1	35|4|0	GH		18 May	1 hour	22	JM			
17		30 Jun	1.5	44|3|0	FG		19 May	1.5 hours	20	FG			
18		1 Jul	1.5	40|0|4	FG		21 May	1.5 hours	21	JM			
19		6 Jul	1.5	21|0|0	GH		14 Jun	1.5 hours	37	JM			
20		7 Jul	1	37|4|1	DQ		18 Jun	1.5 hours	25	JM			
21		9 Jul	1	29|7|0	GH		4 Jul	1.5 hours	39	JM			
22		30 Jul	2	22|3|0	FG		6 Jul	1.5 hours	39	JM			
23		29 Aug	1.5	22|4|0	GH		10 Jul	1.5 hours	34	JM			
24		10 Sep	1	38|0|0	FG		13 Jul	1.5 hours	23	FG			
25		21 Sep	1	31|0|0	GH		17 Jul	1.5 hours	30	JM			
26		1 Oct	2	26|9|5	DQ		3 Aug	1.5 hours	28	JM			
27		25 Oct	1.5	20|4|0	DQ		20 Aug	1.5 hours	32	JM			
28		4 Nov	1.5	38|5|5	FG		26 Aug	1.5 hours	25	JM			
29		5 Nov	2	40|0|0	GH		28 Aug	1.5 hours	33	FG			
30		8 Nov	2	22|7|0	FG		1 Oct	1.5 hours	38	JM			
31		1 Dec	2	41|6|0	DQ		21 Oct	1.5 hours	34	JM			
32		19 Dec	2	39|9|1	GH		9 Nov	1.5 hours	32	JM			
33							15 Nov	1.5 hours	35	JM			
34							15 Nov	1.5 hours	27	JM			
35							2 Dec	1.5 hours	35	FG			
36							7 Dec	1.5 hours	23	JM			
37							11 Dec	1.5 hours	38	FG			
38							19 Dec	1.5 hours	20	FG			
39													

Fig. 16. Multiple tables [Library Carpentry n.d.].

You should also enter a zero if you mean zero, because to the computer 0 means data (i.e. something measured or counted), whereas an empty cell means a null value (i.e. no measurement). If you leave an empty cell instead of filling it in with a zero, it is the same as leaving out the data.

Speaking of null values, there are good (e.g. empty cell, NA when working in R) and bad (e.g. -999, 999, other numeric values, zero, text) choices for representing missing data (Figure 14).

Null values	Problems	Compatibility	Recommendation
0	Indistinguishable from a true zero		Never use
Blank	Hard to distinguish values that are missing from those overlooked on entry. Hard to distinguish blanks from spaces, which behave differently.	R, Python, SQL	Best option
-999, 999	Not recognized as null by many programs without user input. Can be inadvertently entered into calculations.		Avoid
NA, na	Can also be an abbreviation (e.g., North America), can cause problems with data type (turn a numerical column into a text column). NA is more commonly recognized than na.	R	Good option
N/A	An alternate form of NA, but often not compatible with software		Avoid
NULL	Can cause problems with data type	SQL	Good option
None	Uncommon. Can cause problems with data type	Python	Avoid
No data	Uncommon. Can cause problems with data type, contains a space		Avoid
Missing	Uncommon. Can cause problems with data type		Avoid
-,+,.	Uncommon. Can cause problems with data type		Avoid

Fig. 17. Choices for representing null values [White *et al.* 2013].

You should NOT use formatting to convey information or to make the spreadsheet look pretty (e.g. merging cells, using colours, fonts, italics) because it is hard to tell the computer how to handle this information. Instead, restructure your data (e.g. by creating a new field) to encode what the formatting would mean.
You should also NOT put units or comments in cells, but in a new column.
You should also be careful with column names: you should NOT repeat headers, as repeated headers can easily get mixed up with the data, causing problems later on. Instead, freeze the column headers. You should also avoid using numbers in column names (as some programs do not like field names that are text strings starting with numbers), special characters such as {}[]<>()* % # ' ; " , : ? ! & @ $ ~ (which often mean special things to computers and make the data harder to work with), and spaces (which computers use to separate commands). Instead of spaces, use underscores or CamelCase to include multiple words. You should also be clear and use short, meaningful names and be consistent with names, abbreviations and capitalisation. Finally, you should write dates as YYYY-MM-DD or have separate columns for year, month and day.
Special characters (including newlines, tabs and vertical tabs) should also be excluded from cells. Use only text and spaces in cells.
Finally, you should store the metadata as a separate file in the same directory as your data file, preferably in plain text format, with a name that is clearly associated with your data file.

Further resources

- Article on Data Organization in Spreadsheets here [Broman *et al.* 2018]
- Six tips for better spreadsheets (Article) here [Perkel 2022]
- Article on Tidy Data here [Wickham 2014]

Tools

- Data Curation Tool (FAIR4Health)
- FAIRDOM: “Project space [...] used by the community to organize, share and publish data, documents, literature and computational models, as well as to list contributors”
- G-Node Infrastructure (GIN) = Modern Research Data Management for Neuroscience (see Notes for more details)

Glossary

- **Analysed data:** processed data that has been interpreted [Goldman & Martin 2023].
- **Authority file:** type of controlled vocabulary that provides a consistent list of terms to describe different types of resources. It includes cross-references from variant or alternate terms and sometimes other contextual or biographical information to help users disambiguate. Authority files are often used to identify the correct forms of names [University of Pittsburgh Library System 2023].
- **Controlled vocabulary:** organised arrangement of words and phrases (from lists of terms to complex machine-readable ontologies) that are used to index and/or retrieve content through browsing and searching. A controlled vocabulary contains predefined, preferred and variant terms and is therefore used to describe a particular domain of knowledge (i.e. it has a defined scope). The use of a controlled vocabulary during the creation of data or metadata supports consistency and accuracy [Getty n.d., NBIS 2020-2022, University of Pittsburgh Library System 2023]. Controlled vocabularies include authority files, taxonomies, term lists and thesauri [University of Pittsburgh Library System 2023].
- **Data Management Plan (DMP):** formal and living document that defines responsibilities and provides guidance. It describes data and data management during the project and measures for archiving and making data and research results available, usable and understandable after the project has ended [Assmann *et al.* 2022, March; Cozatl *et al.* 2021; Lindstädt *et al.* 2019; Vandendorpe & Lindstädt 2020].
- **Documentation:** document or documents included with your data that describe(s) the details of the data and how it was generated [KU Leuven 2022].
- **FAIR Data Principles:** FAIR (Findable, Accessible, Interoperable, Reusable) is a term coined by the FORCE11 community in 2016 for sustainable research data management (RDM). The FAIR Data Principles are a concise and measurable set of principles that can be used as a guide for those wishing to improve the reusability of their data assets. The FAIR data principles promote professional management of research data and have been adopted by the European Commission and integrated into the Horizon Europe funding guidelines [forschungsdaten.info 2024, Wilkinson *et al.* 2016].
- **Final data:** curated data that support your research question [Goldman & Martin 2023], such as gene banks, national statistical archives [Darby n.d.].
- **Licence:** official permission to use something, i.e. “promise not to sue” based on existing rights [Bobrov *et al.* 2021].
- **Metadata:** data about data. It is a standardised description of the data in a formal, human- and machine-readable structure. Metadata is considered a subset of

documentation as it describes, explains, locates, or otherwise facilitates the retrieval, use, or management of a resource such as a dataset [Bres *et al.* 2022, KU Leuven 2022, Lindstädt *et al.* 2019, RfII 2021, Voigt *et al.* 2022].

- **Open source:** source code of software that is made publicly and freely available and can be redistributed and modified [Open source 2024, Open source n.d.].
- **Processed data:** raw data made useful [Goldman & Martin 2023].
- **Raw data:** potential information generated by a researcher for the first time during a research project. It is unprocessed, possibly even untouched by human hands, unseen by human eyes, unthought by human minds. It needs to be contextualised to make it accessible to the human audience [Darby n.d., Goldman & Martin 2023, Pomerantz 2015].
- **Research data:** research data is the collection of digital and non-digital objects (excluding scientific publications) that are generated (e.g. through measurements, surveys, source work), studied and stored during or as a result of scientific research activities. These objects are commonly accepted in the scientific community as necessary for the production, validation and documentation of original research results. In the context of Research Data Management, research data also includes non-data objects such as software and simulations.
- **Research Data Management (RDM):** the care and maintenance required to (1) obtain high-quality data (whether produced or reused), (2) make the data available and usable in the long term, independent of the data producer and (3) make research results reproducible beyond the research project [Biernacka *et al.* 2020, Bres *et al.* 2022, Bres *et al.* 2023, Pauls *et al.* 2023, RfII, Voigt *et al.* 2022]. It complements research from planning to data reuse and deletion.
- **Quality Control (QC):** process of controlling the use of data for an application or a process [Data quality 2024].
- **Taxonomy:** hierarchical classification or categorisation system in which all terms belong to a single hierarchical structure. It has parent/child or broader/narrower relationships to other terms and allows classification according to a predetermined system [University of Pittsburgh Library System 2023].
- **Term list:** a term list (or pick list) is the simplest type of controlled vocabulary. It is a simple list of words and/or phrases used to identify a particular characteristic of a person, event, object or other "thing" [University of Pittsburgh Library System 2023].
- **Thesaurus:** A thesaurus is a type of controlled vocabulary. It is a kind of dictionary of subject headings or descriptors that represents all the concepts for a particular domain in a consistent way, and labels each concept with a preferred term. A thesaurus contains preferred terms, variant terms, broader and narrower terms, as well as related terms, which may or may not belong to the same hierarchical structure of the term. Thesauri are used to organise a collection of documents for reference and retrieval [Merriam-Webster n.d., University of Pittsburgh Library System 2023].

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project ID 460129525 (NFDI4Microbiota to JV) and project ID 501899475 (FAIRagro to SB).

Sources

1. 5S (methodology). (2024, June 11). In Wikipedia. https://en.wikipedia.org/wiki/5S_(methodology)
2. Adelaide Microscopy. (2024). Data Access - Data management and access. https://www.adelaide.edu.au/microscopy/facilities-services/data-access#how-do-i-organise-my-microscopy-files
3. Adelaide Microscopy. (n.d.). File Naming Convention in 7 Steps. https://www.adelaide.edu.au/microscopy/ua/media/789/filenamingworksheet_v2.pdf
4. Assmann, C., Gadelha, L., Lindstädt, B., Markus, K., Shutsko, A. & Vandendorpe, J. (2022, November). Workshop on Research Data Management [Slides]. Friedrich Schiller University Jena & ZB MED - Information Centre for Life Sciences.
5. Assmann, C., Gadelha, L., Lindstädt, B., Shutsko, A. & Vandendorpe, J. (2022, March). Workshop on Research Data [Slides]. Friedrich Schiller University Jena & ZB MED - Information Centre for Life Sciences.
6. Biernacka, Katarzyna, Maik Bierwirth, Petra Buchholz, Dominika Dolzycka, Kerstin Helbig, Janna Neumann, Carolin Odebrecht, Cord Wiljes and Ulrike Wuttke: Train-the-Trainer Concept on Research Data Management. Version 3.0. Berlin, 2020. https://doi.org/10.5281/zenodo.4071471
7. Bobrov, E., Adam, L.-S., Söring, S., Jäckel, D., Herwig, A., Lindstädt, B., Vandendorpe, J. & Shutsko, A. (2021, October 07). Workshop on Research Data [Slides]. Berlin University Alliance & ZB MED - Information Centre for Life Sciences.
8. Boße, S., Vedder, L., Kleppert, I., Bonerath, A. & Külbel, F. (2024, January 15). Handling and organizing plant phenotyping data - a workshop on Research Data Management [Slides]. PhenoRob & ZB MED - Information Centre for Life Sciences.
9. Bres, E., Rudolf, D., Lindstädt, B. & Shutsko, A. (2022, February 09). Research Data Management in Medical & Biomedical Sciences [Slides]. Rhenish Friedrich Wilhelm University of Bonn & ZB MED - Information Centre for Life Sciences.
10. Bres, E., Rudolf, D., Lindstädt, B., Markus, K., Vandendorpe, J. & Riedel, R. (2023, July). Workshop on Research Data Management [Slides]. Rhenish Friedrich Wilhelm University of Bonn & ZB MED - Information Centre for Life Sciences.
11. Briney, K. A. (2020). File Naming Convention Worksheet (Submitted). California Institute of Technology. https://doi.org/10.7907/894Q-ZR22
12. Briney, K., Coates, H., & Goben, A. (2020). Foundational Practices of Research Data Management. In Research Ideas and Outcomes (Vol. 6). Pensoft Publishers. https://doi.org/10.3897/rio.6.e56508
13. Broman, K. W., & Woo, K. H. (2018). Data Organization in Spreadsheets. In The American Statistician (Vol. 72, Issue 1, pp. 2–10). Informa UK Limited. https://doi.org/10.1080/00031305.2017.1375989
14. Carrano, J. (n.d.). Tagging and Finding Your Files: Home. MIT Libraries. https://libguides.mit.edu/metadataTools/
15. Colomb, Julien, Thorsten Arendt, Keisuke Sehara, and The Gin-Tonic team. “Towards a Standardized Research Folder Structure.” Generation Research, 2021. https://doi.org/10.25815/WCY6-M233
16. Cozatl, R., Grenzebach, G., Lindstädt, B., Shutsko, A. & Vandendorpe, J. (2021, November). Workshop on Research Data Management [Slides]. Martin Luther University of Halle-Wittenberg & ZB MED - Information Centre for Life Sciences.

17. Darby, R. (n.d.). Research data defined. University of Reading Research Services. https://www.reading.ac.uk/research-services/research-data-management/about-research-data-management/research-data-defined
18. Data Carpentry for Biologists. (n.d.a). Tidy Data. https://datacarpentry.org/semester-biology/materials/tidy-data/
19. Data Carpentry for Biologists. (n.d.b). Improving Messy Data (Tidy Data). https://datacarpentry.org/semester-biology/exercises/Tidy-data-improving-messy-data-SQL/
20. Data quality. (2024, July 8). In Wikipedia. https://en.wikipedia.org/wiki/Data_quality
21. De Kok T. How to keep your research projects organized: folder structure. Medium. 2018, Nov 23. https://towardsdatascience.com/how-to-keep-your-research-projects-organized-part-1-folder-structure-10bd56034d3a
22. Di Russo J. (2020, February 18). A Simple Story to Explain Version Control to Anyone.https://towardsdatascience.com/a-simple-story-to-explain-version-control-to-anyone-5ab4197cebbc
23. Engelhardt, C., Biernacka, K., Coffey, A., Cornet, R., Danciu, A., Demchenko, Y., Downes, S., Erdmann, C., Garbuglia, F., Germer, K., Helbig, K., Hellström, M., Hettne, K., Hibbert, D., Jetten, M., Karimova, Y., Kryger Hansen, K., Kuusniemi, M. E., Letizia, V., … Zhou, B. (2022). D7.4 How to be FAIR with your data. A teaching and training handbook for higher education institutions (V1.1 DRAFT). Zenodo. https://doi.org/10.5281/ZENODO.5837500
24. Forschungsdatenmanagement Thüringen (TKFDM). (2021, February 4). 5S Data: Setz dich auf deine 5 Buchstaben und organisiere deine Daten! (Coffee Lecture) [Video]. YouTube. https://www.youtube.com/watch?v=73XzLsLrwMk
25. forschungsdaten.info (2024, March 06). FAIRe Daten. https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten/
26. German Council for Scientific Information Infrastructures (RfII). (2021). Research Data Management. https://rfii.de/en/topics/#forschungsdaten
27. Getty. (n.d.). 2. What Are Controlled Vocabularies? https://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/what.html
28. Git. (n.d.). 1.1 Getting Started - About Version Control. https://git-scm.com/book/en/v2/Getting-Started-About-Version-Control
29. Goldman, J. (2020a, July 27). Organize Your Files [Checklist]. https://osf.io/fp9j5
30. Goldman, J. (2020b, July 27). Organize Your Files [Slides]. https://osf.io/yeqjv
31. Goldman, J. (2020c, August 04). How to name a file. https://osf.io/dpu45
32. Goldman, J. & Martin, E. (2023, July 19). Case Study. Retrieved from osf.io/qazrk
33. Haenel, V., & Plenz, J. (2014). Git [PDF]. Munich, Germany: Open Source Press.
34. Jacob, B., Kroehling, M., Mertzen, D., Straka, J., Lindstädt, B., Shutsko, A. & Vandendorpe, J. (2022, May). Workshop on Research Data [Slides]. Forschungsdatenmanagement in Brandenburg: Technologien, Kompetenzen, Rahmenbedingungen (FDM-BB) & ZB MED - Information Centre for Life Sciences
35. Klump, J., Pampel, H., Rothfritz, L., & Strecker, D. (2024). Recommendations on Data Versioning. Zenodo. https://doi.org/10.5281/zenodo.13743876

36. KU Leuven. (2022). FAQ RDM - Metadata and Documentation. https://www.kuleuven.be/rdm/en/faq/faq-metadata-documentation#What%20is%20metadata
37. Lang, K., Roman, G., Jessica, R., Annett, S., Nadine, N., & Lehmann, A. (2021). The 5S Methodology in Research Data Management. Zenodo. https://doi.org/10.5281/ZENODO.4494258
38. Library Carpentry. (n.d.) Tidy data for librarians. https://librarycarpentry.org/lc-spreadsheets/
39. Lindstädt, B., Vandendorpe, J, & von der Ropp, S. (2019, December 09). Research Data Management [Slides]. RWTH Aachen University & ZB MED - Information Centre for Life Sciences.
40. Longwood Medical Area Research Data Management Working Group (LMA RDMWG). (2024a). File Naming Conventions. https://datamanagement.hms.harvard.edu/plan-design/file-naming-conventions
41. Longwood Medical Area Research Data Management Working Group (LMA RDMWG). (2024b). Directory Structure. https://datamanagement.hms.harvard.edu/plan-design/directory-structure
42. Macdonald, M. & Smith, K. (2013, January 28). Tagging and finding your files. MIT Libraries. https://libguides.mit.edu/ld.php?content_id=6600405
43. Massachusetts Institute of Technology Libraries Data Management Service (MIT Libraries). (2016). Data Management: File Organization. https://opendata.eawag.ch/eawagrdm/static/MITRES_STR_002S16_eawmod.pdf
44. Massachusetts Institute of Technology Libraries Data Management Service (MIT Libraries). (2020). Worksheet Naming and Organizing your files and folders https://www.dropbox.com/scl/fi/1zd63iszw33rh4hjcu1dl/Worksheet_fileOrg.docx?rlkey=q0t25t1wttp4qx2p1ne39qfhd&e=1&dl=0
45. Massachusetts Institute of Technology Libraries Data Management Service. (n.d.a). File naming and folder hierarchy. https://libraries.mit.edu/data-management/store/organize/
46. Massachusetts Institute of Technology Libraries Data Management Service (MIT Libraries). (n.d.b) Tagging and Finding Your Files. https://libguides.mit.edu/metadataTools/
47. Merriam-Webster. (n.d.). Thesaurus. In Merriam-Webster.com dictionary. Retrieved August 8, 2024, from https://www.merriam-webster.com/dictionary/thesaurus
48. Mičetić, I., Popleteeva, M., Ahokas, M., Muhammad, N., Fuchs, S. & Kallberg, Y. (n.d.). Data organisation. RDMkit. https://rdmkit.elixir-europe.org/data_organisation
49. NBIS - National Bioinformatics Infrastructure Sweden. (2020-2022).Controlled vocabularies & ontologies. https://nbisweden.github.io/module-metadata-dm-practices/02-ontologies/index.html
50. Onyimadu, A. (202, September 8). All Possible Ways to Add Tags to Files on Windows 11. Next of Windows. https://www.nextofwindows.com/how-to-add-tags-to-files-in-windows-11.
51. Open source. (2024, August 2). In Wikipedia. https://en.wikipedia.org/wiki/Open_source
52. Open source. (n.d.). Oxford Languages.
53. Pasquier, G. (2024, July 11). Folder Structure. Kathryn and Shelby Cullom Davis Library. https://libguides.graduateinstitute.ch/rdm/folders

54. Pauls, C., Feeken, C., Steen, E.-E., Lindstädt, B., Markus, K. & Vandendorpe, J. (2023, June). Workshop on Research Data Management [Slides]. Carl von Ossietzky University of Oldenburg & ZB MED - Information Centre for Life.
55. Perkel, J. M. (2022). Six tips for better spreadsheets. In Nature (Vol. 608, Issue 7921, pp. 229–230). Springer Science and Business Media LLC. https://doi.org/10.1038/d41586-022-02076-1
56. Petersen, B., Engelhardt, C., Hörner, T., Jacob, J., Kvetnaya, T., Mühlichen, A., Schranzhofer, H., Schulz, S., Slowig, B., Trautwein-Bruns, U., Voigt, A., & Wiljes, C. (2023). Lernzielmatrix zum Themenbereich Forschungsdatenmanagement (FDM) für die Zielgruppen Studierende, PhDs und Data Stewards. https://doi.org/10.5281/ZENODO.8010617
57. Pilato, C. M., Collins-Sussman, B., & Fitzpatrick, B. W. (2009). Versionskontrolle Mit Subversion (3rd ed.). O'Reilly Verlag.
58. Pomerantz, J. (2015). Metadata. Cambridge, MA: MIT Press.
59. RDM Guide. (n.d.). Folder structure. ELIXIR Belgium. https://rdm.elixir-belgium.org/folder_structure
60. Rehwald, S., Leimer, S., Lindstädt, B., Shutsko, A. & Vandendorpe, J. (2022, June). Workshop on Research Data Management in Medical and Biomedical Sciences [Slides]. CRC/TR 296 LocoTact & ZB MED - Information Centre for Life Sciences.
61. Research Data Management Thuringia (TKFDM). 5S Data: Setz dich auf deine 5 Buchstaben und organisiere deine Daten! (Coffee Lecture) [Video]. 04 February 2021. YouTube: https://youtu.be/73XzLsLrwMk?si=rXFwV7CcyXGJptP_
62. Schmid, F. (2021). Research data documentation - Best practices in filenaming and folder structure. https://ethz.ch/content/dam/ethz/associates/ethlibrary-dam/documents/Aktuell/Kurse/CoffeeLectures/2021-06-23_Coffee_Lecture_File-naming_final.pdf
63. Shanshan Ma. Using hierarchical folders and tags for file management. 2010. https://doi.org/10.17918/etd-3271
64. Software Freedom Conservancy. (n.d.). Give Up GitHub! https://sfconservancy.org/GiveUpGitHub/
65. Technology, Information and Learning Support (TILS). TILS Document Naming Convention. Queensland University of Technology, 2009. Last accessed 28.07.2020. https://www.data.cam.ac.uk/files/gdl_tilsdocnaming_v1_20090612.pdf.
66. UC Merced Library. (n.d.). File and Folder Organization - long draft. https://library.ucmerced.edu/node/66751
67. University of Pittsburgh Library System. (2023, Sep 26). Taxonomies and Controlled Vocabularies. https://pitt.libguides.com/metadatadiscovery/controlledvocabularies
68. Vandendorpe, J. & Lindstädt, B. (2020, August). Research Data Management [Slides]. ZB MED - Information Centre for Life Sciences.
69. Vaughan-Nichols, S. (2022, August 2). GitHub vs GitLab: Which program should you go with? ZDNET. https://www.zdnet.com/article/github-vs-gitlab-the-key-differences/
70. Voigt, P., Frericks, S., Lindstädt, B., Shutsko, A. & Vandendorpe, J. (2022, March). Workshop on Research Data [Slides]. Leipzig University & ZB MED - Information Centre for Life Sciences.
71. Von der Dunk, A. (2021). Lessons in Open Science - Best Practices in Personal RDM.

https://www.slub-dresden.de/fileadmin/groups/slubsite/default_upload_automagic/2021-10-01_-__Lessons_in_Open_Science_-_Personal_RDM.pdf
72. Vukovic, N. (n.d.). Setting up an Organised Folder Structure for Research Projects. http://nikola.me/folder_structure.html
73. White, E., Baldridge, E., Brym, Z., Locey, K., McGlinn, D., & Supp, S. (2013). Nine simple ways to make it easier to (re)use your data. In Ideas in Ecology and Evolution (Vol. 6, Issue 2). Queen's University Library. https://doi.org/10.4033/iee.2013.6b.6.f
74. Wickham, H. (2014). Tidy Data. In Journal of Statistical Software (Vol. 59, Issue 10). Foundation for Open Access Statistic. https://doi.org/10.18637/jss.v059.i10
75. Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. In Scientific Data (Vol. 3, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/sdata.2016.18

Rights and permissions

Cite this lesson plan

Vandendorpe, J., & Boße, S. (2024). Lesson plan on data organisation. https://doi.org/10.4126/FRL01-006484175

Published November 27, 2024

DOI

https://doi.org/10.4126/FRL01-006484175