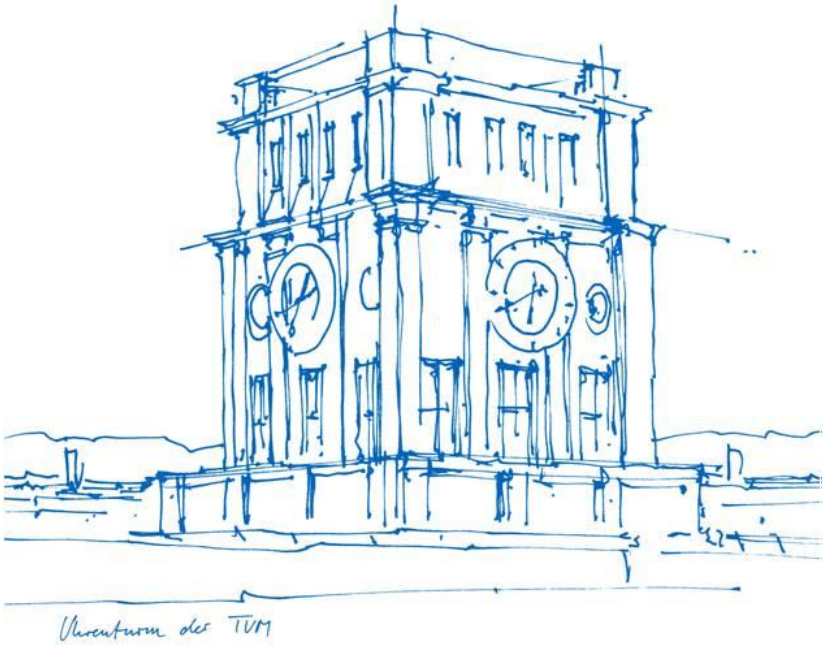
Research Data Management – Folder Structure

Technical University of Munich

University Library

Research Data Services

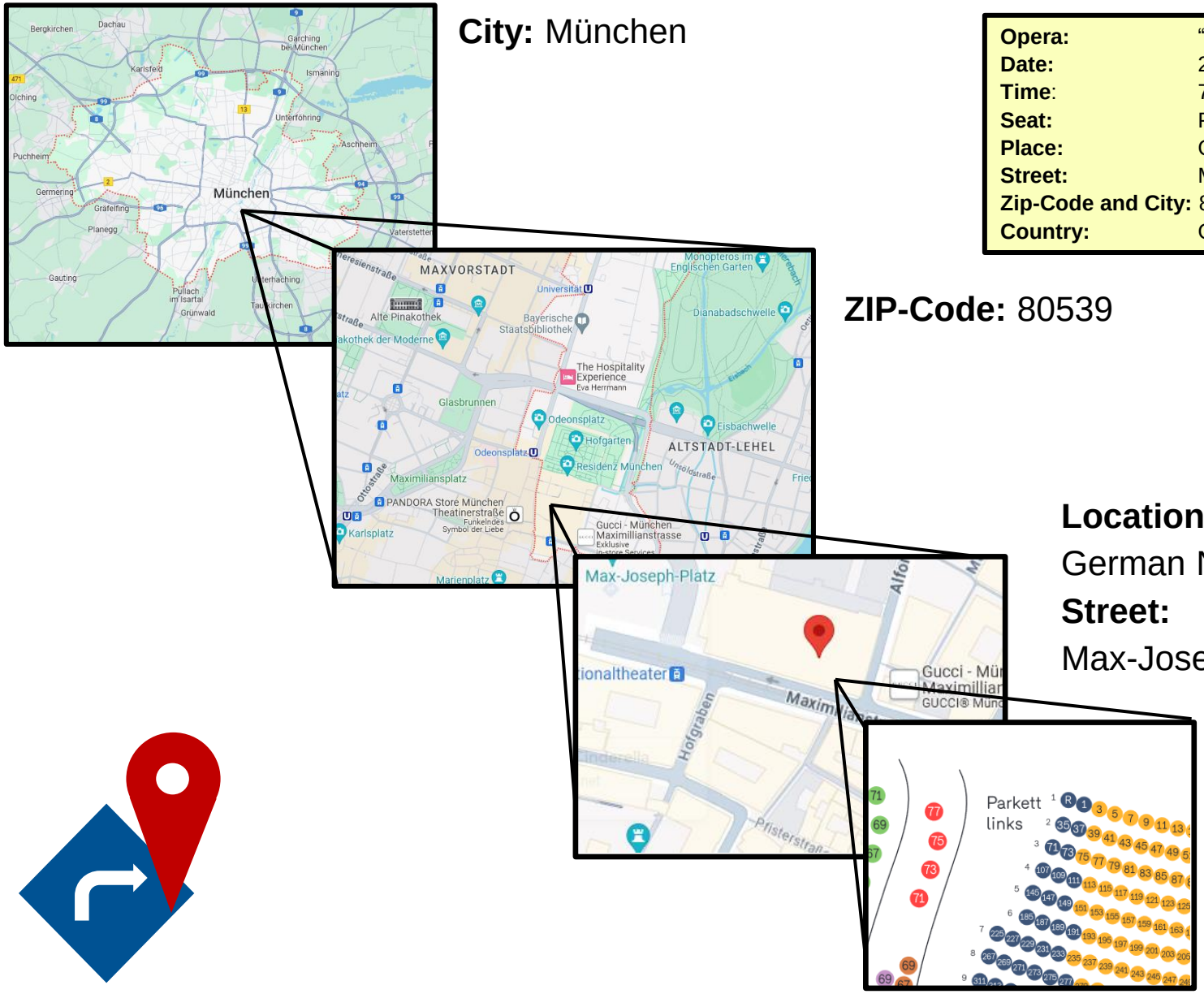
Munich



Our Agenda!

Time	Exercises
Day 1:	
09.00 a.m. – 09.30 a.m.	Introduction and Checklist
09.30 a.m. – 11.00 a.m.	Exercise 1: Reflection on the existing folder structure
10.30 a.m. – 10.45 a.m.	Break
10.45 a.m. – 11.30 a.m.	Research Data Life Cycle and
11.30 a.m. – 12.30 a.m.	Break
12.30 a.m. – 13.30 a.m.	Exercise 2a: Choosing metadata pairs
Homework:	Exercise 2b: Rearranging metadata pairs
Day 2:	
09.00 a.m. – 09.30 a.m.	Discussion Exercises 2a and b
09:30 a.m. – 10.30 a.m.	Exercise 3a and b: File and folder naming convention
10.00 a.m. – 10.15 a.m.	Break
10.15 a.m. – 10.45 a.m.	Introduction to clustering
10.45 a.m. – 12.00 a.m.	Exercise 4: Creation of folder structure
12.00 a.m. – 13.00 a.m.	Break
13.00 a.m. – 13.30 a.m.	Exercise 5: Documentation of folder structure

How did we find our way?



City: München

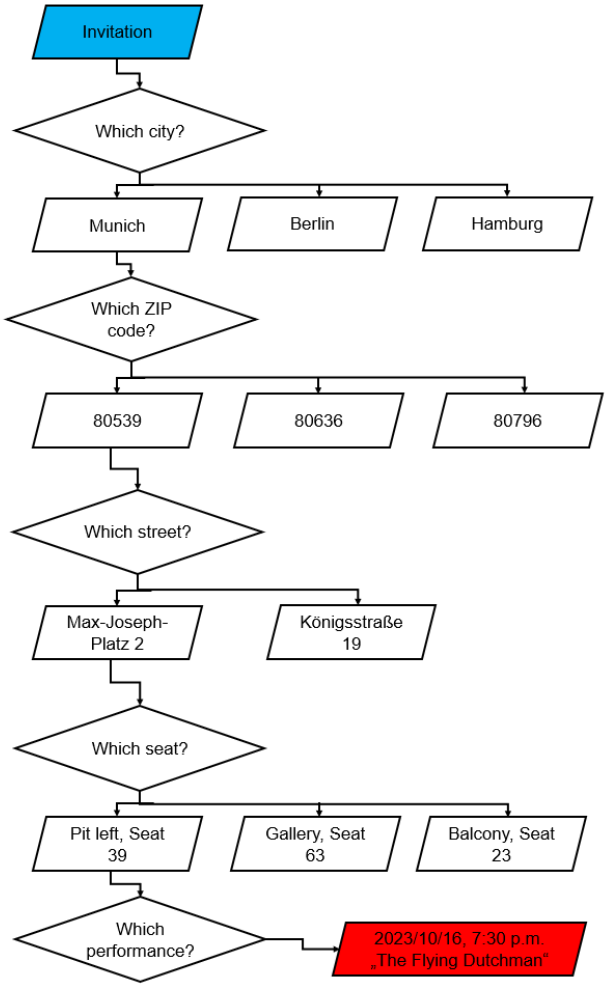
Opera:	“The Flying Dutchman”
Date:	2023/10/16
Time:	7:30 p.m.
Seat:	Pit left, Seat 39
Place:	German National Opera
Street:	Max-Joseph-Platz 2
Zip-Code and City:	80539 Munich
Country:	Germany

ZIP-Code: 80539

Location:
German National Opera
Street:
Max-Joseph-Platz 2

Seat:
Pit left, Seat 39

Which implicit procedure did we use?

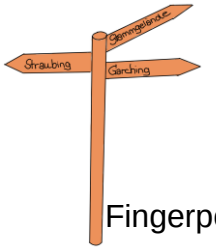
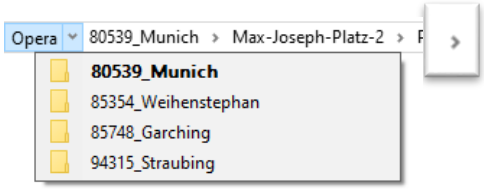


Opera:	“The Flying Dutchman”
Date:	2023/10/16
Time:	7:30 p.m.
Seat:	Pit left, Seat 39
Place:	German National Opera
Street:	Max-Joseph-Platz 2
Zip-Code and City:	80539 Munich
Country:	Germany

Opera › 80539_Munich › Max-Joseph-Platz-2 › Pit-Left_Seat-39

From navigation to folder structure!

Opera › 80539_Munich › Max-Joseph-Platz-2 › Pit-Left_Seat-39

Analogy		Folder		Flowchart Translation
 Area of interest	=	Opera › Root folder	=	Starting place
 Fingerpost	=	 Folder Path Sign	=	Decision point
München Road sign	=	› 80539_Munich › Folder	=	Transition place
 Destination	=	20231016_FlyingDutchman_7_30pm File	=	Target place

From navigation to folder structure!

The folder path is interpretable as:

Metadata!

But what are metadata? And how to use them?

Documentation

Metadata:

„Data about data“

- Data is findable
- Data is usable
- Data is manageable

First Name Surname	Katalin Karikó
Date: TT.MM.YYYY	17.01.1955
Location (ISO-3)	Szolnok (HUN)
Research Area	Biology mRNA

Schema (Metadata Key)

What information should be saved?

- General information
- Methods
- E.g., environmental influences, …

Advantages – Standards:

- Data is comparable
- Data can be linked
- Improved reproducibility

Vocabulary (Metadata Value)

How is the information phrased?

- Format for metadata is predefined
- Definition of which free text is permitted

Advantages – Standards:

- Data is easier to find
- Collaborations can work better together

How to systematically name files?

Metadata keys = Decision boxes = Arrows or Slashes in the file path!

What	Why	How	When	Where	Who
Dublin Core elements relevant for file naming, plus other useful elements (*)					
[Description] **[Identifier]** [Subject] [Type] [Citation] **[Version of]** [Table of contents] **[Title]**	[Is required by] [Is part of]	**[Format]** [Language] [Access Rights] [Instructional Method] [License] [Provenance] [Rights] [Source] **[Instrument]*** **[Method]*** **[Method step]***	[Date] **[Date Available]** **[Date Created]** [Date Accepted] [Date Copyrighted] [Date Submitted] [Date Issued] [Date Modified] [Temporal Coverage] [Date Valid]	**[Spatial Coverage]**	[Contributor] **[Creator]** [Publisher] [Audience] [Audience Education Level] [Mediator] [Rights Holder] [Compiler]

Example: Metadata keys (arrows) and metadata values (folders / files)

[Type (of events)] / [Spatial coverage (Munich)] / [Title (of building)] / [Identifier (of building)] / [Identifier (of seat)] / [Date Available (year-month-day-hours for event)]_[Subject (of event)]

Rules to structure your files!

1. **Does each file path go from general to specific?**
   - 1.1 Have metadata been put into the folder names if the metadata are redundant among a cluster of files?
   - 1.2. Does each file name have metadata going from broad to detailed?
   - 1.3. Does the file system have a navigable number of folder levels (1 to 7)?

2. **Does each file and folder name conform to technical requirements?**
   - 2.1. Is each file path without spaces, having underscores, hyphens or capital letters instead?
   - 2.2. If the order of files or folders matters: Does each file or folder name start with a number and/or letter for sorting?
   - 2.3. Is each file path shorter than the maximum length the computer allows?

3. **Does each child folder answer a single follow-up question[1] about its parent folder?**
   - 3.1. Does the follow-up question aim at static metadata[2] values?
   - 3.2. Are the answers without logical overlap?
   - 3.3. Do the answers leave no logical gaps?
   - 3.4. Is the folder naming convention documented?

4. **Does each file name answer a single follow-up question (with optional sub-questions) about its parent folder?**
   - 4.1. Does each file name have file-specific metadata values?
   - 4.2. Does each file name contain only static metadata[2] values?
   - 4.3. Is each file name short enough to be visible in the file explorer?
   - 4.4. Is the file naming convention documented?

[1] The follow-up question may be different for different parent folders.
[2] Static metadata: For each file, changing the value in the filename/path should not be necessary over time. For versioning, you will keep the previous file and create a new one – with each version file having a different but static metadata value v1, v2, etc.

Exercise 1: Check your folder structure against the checklist!

Follow the paths in your folder structure. Try explicitly naming the implicit question leading to each level's subfolder. You should ask **ONE** question only, to which the subfolders are the answer.

At each level, complete the checklist by answering it with “Yes” or “No”.

Wherever you have to answer with “No”, mark the section in your folder structure and write down which rules have not been followed.

To prepare for optimised file organisation, reflect on your results and try to answer the following questions:

- Which folders are most important?
- Which are the most universal or crucial categories you have?
- Which file types do you have?

Keep in mind that each arrow or slash corresponds to one decision point!

- Arrow or Slash = Metadata key
- Folder or File = Metadata value

Exercise: Research Data Life Cycle

Where do you have difficulties to find your data?

- Add red dots at the places!

Is the data used/collected relevant for other group members?

- Add a note with an explanation!

At which point of the life cycle do you spend most in your workflow?

- Add a note with your name!

Exercise 2: Thinking about degrees of freedom!

What	Why	How	When	Where	Who
Dublin Core elements relevant for file naming, plus other useful elements (*)					
[Description] **[Identifier]** [Subject] [Type] [Citation] **[Version of]** [Table of contents] **[Title]**	[Is required by] [Is part of]	**[Format]** [Language] [Access Rights] [Instructional Method] [License] [Provenance] [Rights] [Source] **[Instrument]*** **[Method]*** **[Method step]***	[Date] **[Date Available]** **[Date Created]** [Date Accepted] [Date Copyrighted] [Date Submitted] [Date Issued] [Date Modified] [Temporal Coverage] [Date Valid]	**[Spatial Coverage]**	[Contributor] **[Creator]** [Publisher] [Audience] [Audience Education Level] [Mediator] [Rights Holder] [Compiler]

Exercise 2: Thinking about degrees of freedom!

Exercise 2: Thinking about degrees of freedom

Information: File organisation is not trivial. This is because – although there are rules to follow – there are 3 degrees of freedom:

- Which metadata keys are necessary?
- What is a suitable order of metadata keys, from most universal and most crucial to least universal and least crucial?
- Where do you want to introduce folder levels?

To derive a project-specific folder structure, the group must devise its own conventions for dealing with these degrees of freedom.

Exercise 2: Thinking about degrees of freedom!

Exercise 2: Thinking about degrees of freedom

a) Choosing metadata pairs.

Take an exemplary file. For each metadata value, make the corresponding metadata key explicit, based on the Dublin Core elements. Write down the metadata pairs as a sequence of metadata keys and values. You should restrict the metadata pairs to a maximum of ten elements! Write down the sequence of metadata keys:

……/……/……/……/FileFormat

Reflection period: What to do next?

Where are we now?

- We discovered crunching points in your folder structure!
- We found a list of metadata keys and values for exemplary files!

What you can do until the second part?

- Rearrange the list of metadata keys!
- Continue the process of exercise two for all types of files!

What will we do next time?

- Documenting your metadata key/values pairs!
- Based on the metadata key/value pairs we discuss a new folder structure!
- We make your structure sustainable by starting a proper documentation routine!

Day 2 – From Files to Folders

Day 2:	
09.00 a.m. – 09.30 a.m.	Discussion Exercises 2a and b
09:30 a.m. – 10.30 a.m.	Exercise 3a and b: File and folder naming convention
10.00 a.m. – 10.15 a.m.	Break
10.15 a.m. – 10.45 a.m.	Introduction to clustering
10.45 a.m. – 12.00 a.m.	Exercise 4: Creation of folder structure
12.00 a.m. – 13.00 a.m.	Break
13.00 a.m. – 13.30 a.m.	Exercise 5: Documentation of folder structure

Reflection day 1 and preparation period

Where are we now?

- We discovered crunching points in your folder structure!
- We found a list of metadata keys and values for exemplary files!

What were your insights during the preparation period?

- You found a list of metadata keys and values for exemplary files?
- You rearranged the list of metadata keys?

What will we do today?

- Documenting your metadata key/values pairs!
- Based on the metadata key/value pairs we discuss a new folder structure!
- We make your structure sustainable by starting a proper documentation routine!

Exercise 2: Thinking about degrees of freedom!

Exercise 2: Thinking about degrees of freedom

If you compare your results to others, you might see that the order of your metadata differs.

b) Rearranging metadata pairs.

Rearrange the metadata pairs following the checklist from section 2. Go from general to specific and from most crucial to least crucial. Write down your metadata pairs in a hierarchical form:

……/……/……/……/FileFormat

Exercise 3a: (Single Person)

Documenting metadata keys and values for files and folders

a) Create file names from your list of metadata pairs in exercise 2b. You can either use bullet points or write down the characteristics in table format (see Table 4). The last three to four metadata keys/values correspond to your file name. The remaining key/value pairs will move to the folder structure. It might be necessary to add several bullet points for the folder-level characteristics for the same file type at different locations of your department structure.

Exercise 3a: (Single person)

Documenting metadata keys and values for files and folders

A file naming convention could contain the following information:

- File type: Microscope image
- Filename schema: [date]_[microscope]_[imageNumber]
- Schema key:
  - date: date of image capture YYYY-MM-DD format
  - Microscope: name/model of microscope
  - imageNumber: written in sequential form 00X
- Example filename: 2023-10-16_mic01_001.tif
- Folder level characteristics: you can name up to 7 metadata keys (and their values)

This procedure should be repeated for each file type.

File Type	Filename schema	Schema key	Example filename	Folder level characteristics
Image	Date_	Date: YYYY-MM-DD	2023-10-16	Experimental; Research
	microscope_	Microscope: name/model	_mic01_001.tif	
	imageNumber	imageNumber: 00X		

Exercise 3b: (Plenum)

Documenting metadata keys and values for files and folders

b) In the plenum you should discuss the folder-level metadata pairs and find a common standard. Compare the results and find common metadata keys for the different values. The next section will use the result to create a folder structure.

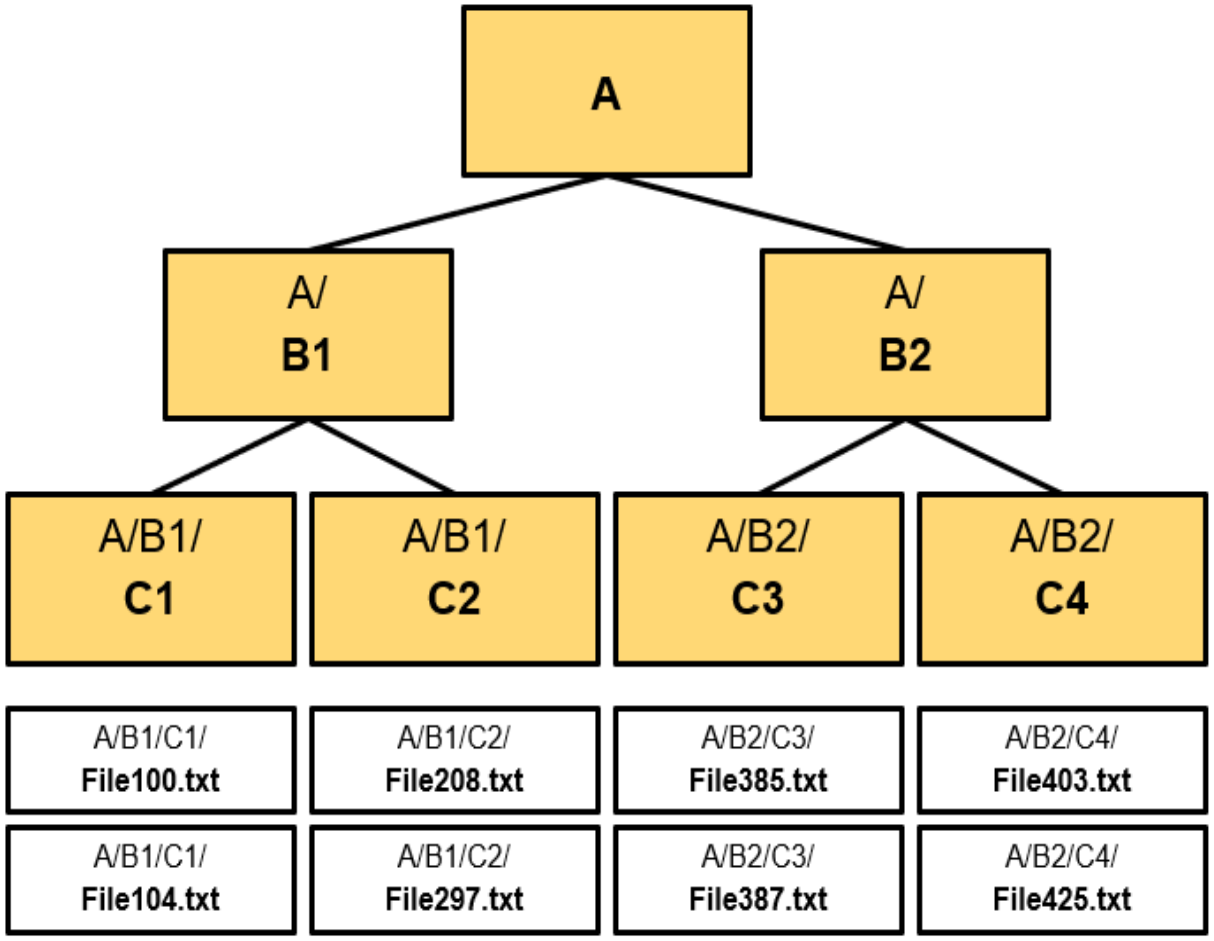
Clustering

STEP 4
Repeat step 3 until you can put all high-level folders into one top folder which represents the overall context.

STEP 3
Find folders with multiple identical metadata values and put each cluster of folders into a higher folder.

STEP 2
Find files with identical sets of contextual metadata values and put each cluster of files into a folder.

STEP 1
List all files names; plus contextual metadata values, sequenced from wider to narrower context.



Exercise 4: (Plenum)

Transferring metadata values from the file names to folder levels

In exercise 3, a file naming convention and folder-level metadata key/value pairs were generated. The folder-level characteristics are now used to generate a folder system. You can use the metadata value as a folder name and formulate a question by using the key. By doing this, you can continuously check if your system follows the checklist.

a) Use your list of generic filenames and folder metadata key/value pairs created in exercise 3.

b) Group together files with similar metadata values. Transfer matching values to the next higher level.

c) Repeat the procedure until you end up with a single folder.

Exercise 5: (Plenum/after Workshop)

Make your folder system and file naming sustainable

In this exercise, you will develop a documentation of the folder system and naming conventions. The file should be prominently placed within the first folder level and be updated in case of changes. The final document should be saved as plain text. This section builds on the well-described ReadMe instructions of MIT.

Profound documentation includes general information, an overview of the structure and the naming convention.

a) Overview

b) Folder structure

c) File naming schema

d) File naming abbreviations

MIT Library. (2023, 11 14). Data Management - Organize your files. Retrieved from https://libraries.mit.edu/data-management/store/organize/