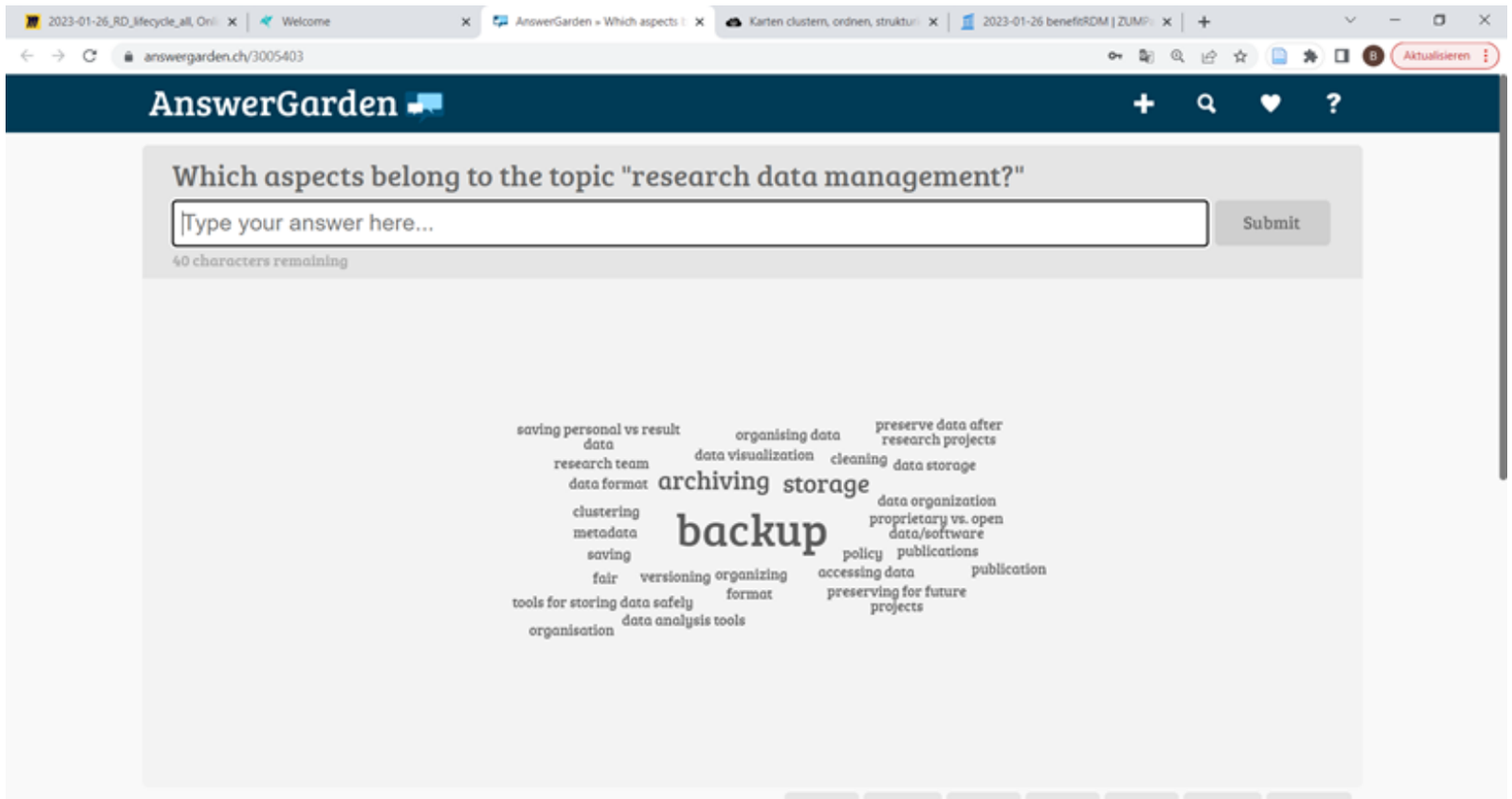
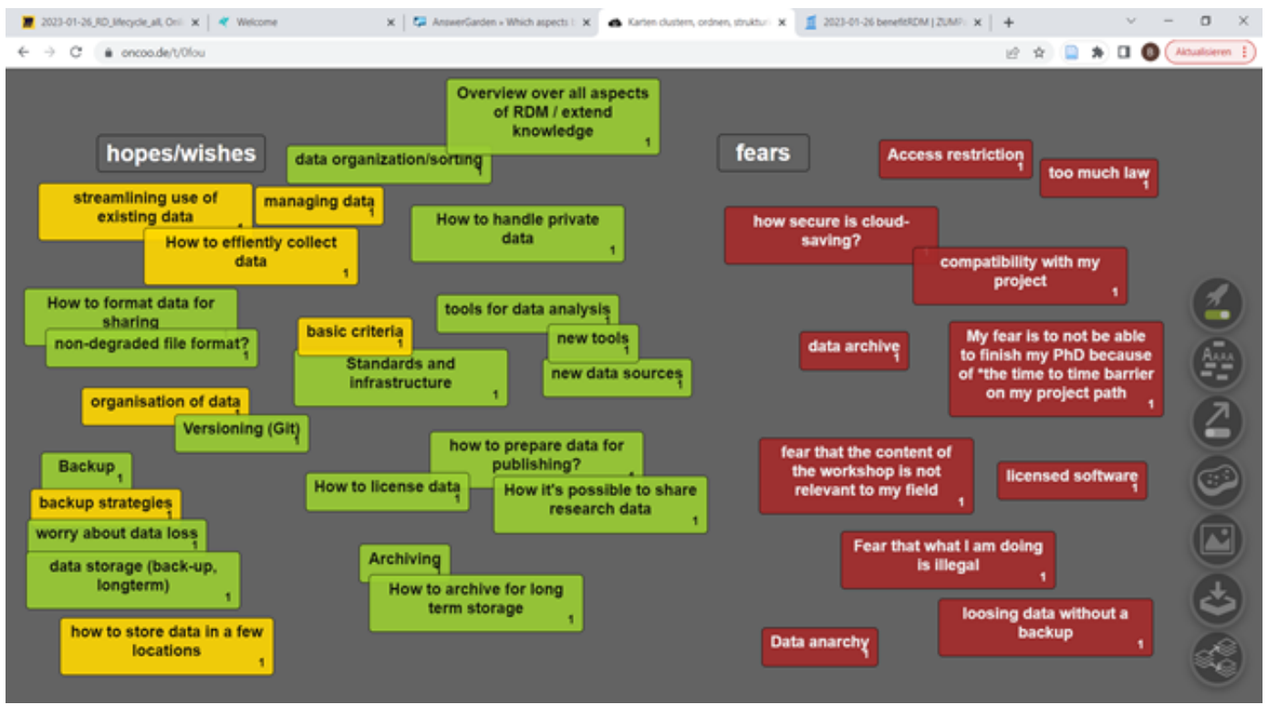
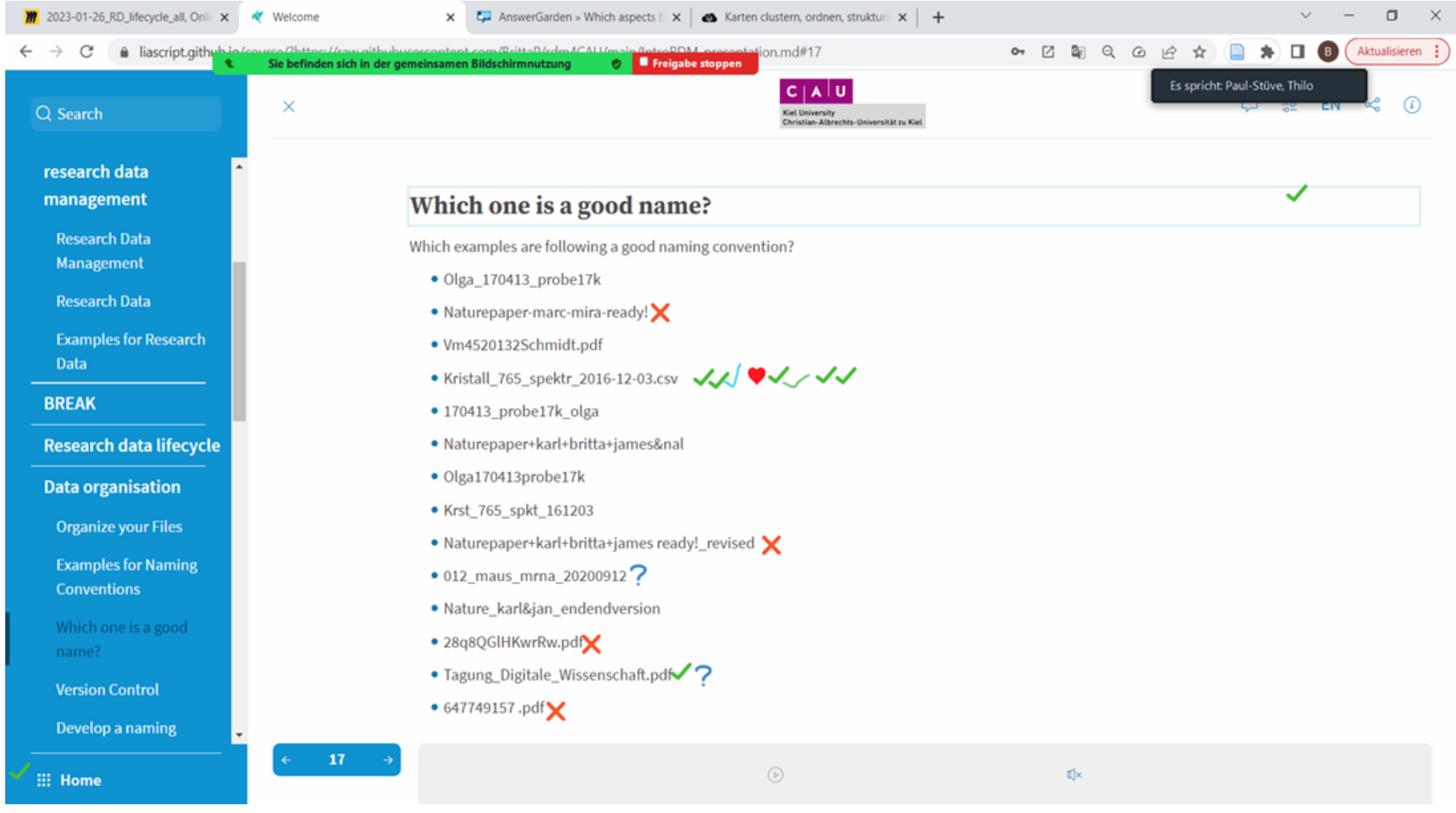

answergarden.ch/3005403
AnswerGarden
Which aspects belong to the topic "research data management?"
Type your answer here...
Submit
40 characters remaining
saving personal vs result data
organising data
preserve data after research projects
research team
data visualization
cleaning
data storage
data format
archiving
storage
data organization
clustering
backup
proprietary vs. open data/software
metadata
saving
policy
publications
fair
versioning
organizing
accessing data
publication
tools for storing data safely
format
preserving for future projects
organisation
data analysis tools


oncoo.de/t/0fou
Aktualisieren
hopes/wishes
Overview over all aspects of RDM / extend knowledge 1
data organization/sorting 1
streamlining use of existing data 1
managing data 1
How to handle private data 1
How to effiently collect data 1
How to format data for sharing 1
non-degraded file format? 1
basic criteria 1
tools for data analysis 1
new tools 1
Standards and infrastructure 1
new data sources 1
organisation of data 1
Versioning (Git) 1
how to prepare data for publishing? 1
Backup 1
How to license data 1
How it's possible to share research data 1
backup strategies 1
worry about data loss 1
Archiving 1
data storage (back-up, longterm) 1
How to archive for long term storage 1
how to store data in a few locations 1
fears
Access restriction 1
too much law 1
how secure is cloud-saving?
compatibility with my project 1
data archive 1
My fear is to not be able to finish my PhD because of *the time to time barrier on my project path 1
fear that the content of the workshop is not relevant to my field 1
licensed software 1
Fear that what I am doing is illegal 1
loosing data without a backup 1
Data anarchy 1

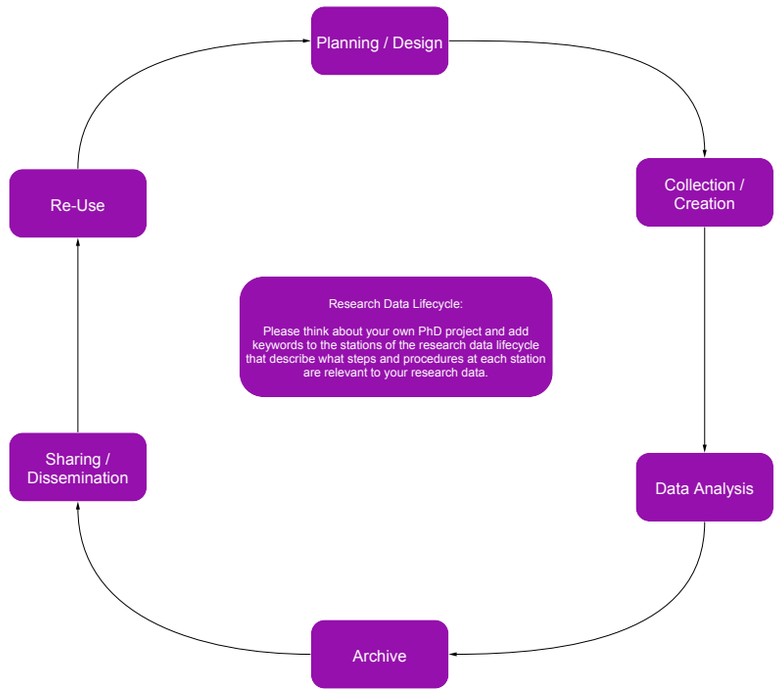

2023-01-26_RD_lifecycle_all, Onli
Welcome
AnswerGarden » Which aspects t
Karten clustern, ordnen, struktur
liascript.github.io/course/?https://raw.githubusercontent.com/BrittaP/rdm4CAU/main/IntroRDM_presentation.md#17
Aktualisieren
Sie befinden sich in der gemeinsamen Bildschirmnutzung
Freigabe stoppen
Es spricht: Paul-Stüve, Thilo
Search
research data management
Research Data Management
Research Data
Examples for Research Data
BREAK
Research data lifecycle
Data organisation
Organize your Files
Examples for Naming Conventions
Which one is a good name?
Version Control
Develop a naming
Home
C A U
Kiel University
Christian-Albrechts-Universität zu Kiel
EN
Which one is a good name?
Which examples are following a good naming convention?
Olga_170413_probe17k
Naturepaper-marc-mira-ready!
Vm4520132Schmidt.pdf
Kristall_765_spektr_2016-12-03.csv
170413_probe17k_olga
Naturepaper+karl+britta+james&nal
Olga170413probe17k
Krst_765_spkt_161203
Naturepaper+karl+britta+james ready!_revised
012_maus_mrna_20200912
Nature_karl&jan_endendversion
28q8QGlHKwrRw.pdf
Tagung_Digitale_Wissenschaft.pdf
647749157 .pdf
17

1management
2data
21geodata
22archdata
3code
4figures
5documents_pub
6praes
7lib

kernel-density_BA_per4.R

Laurenz

Git (gitLab)



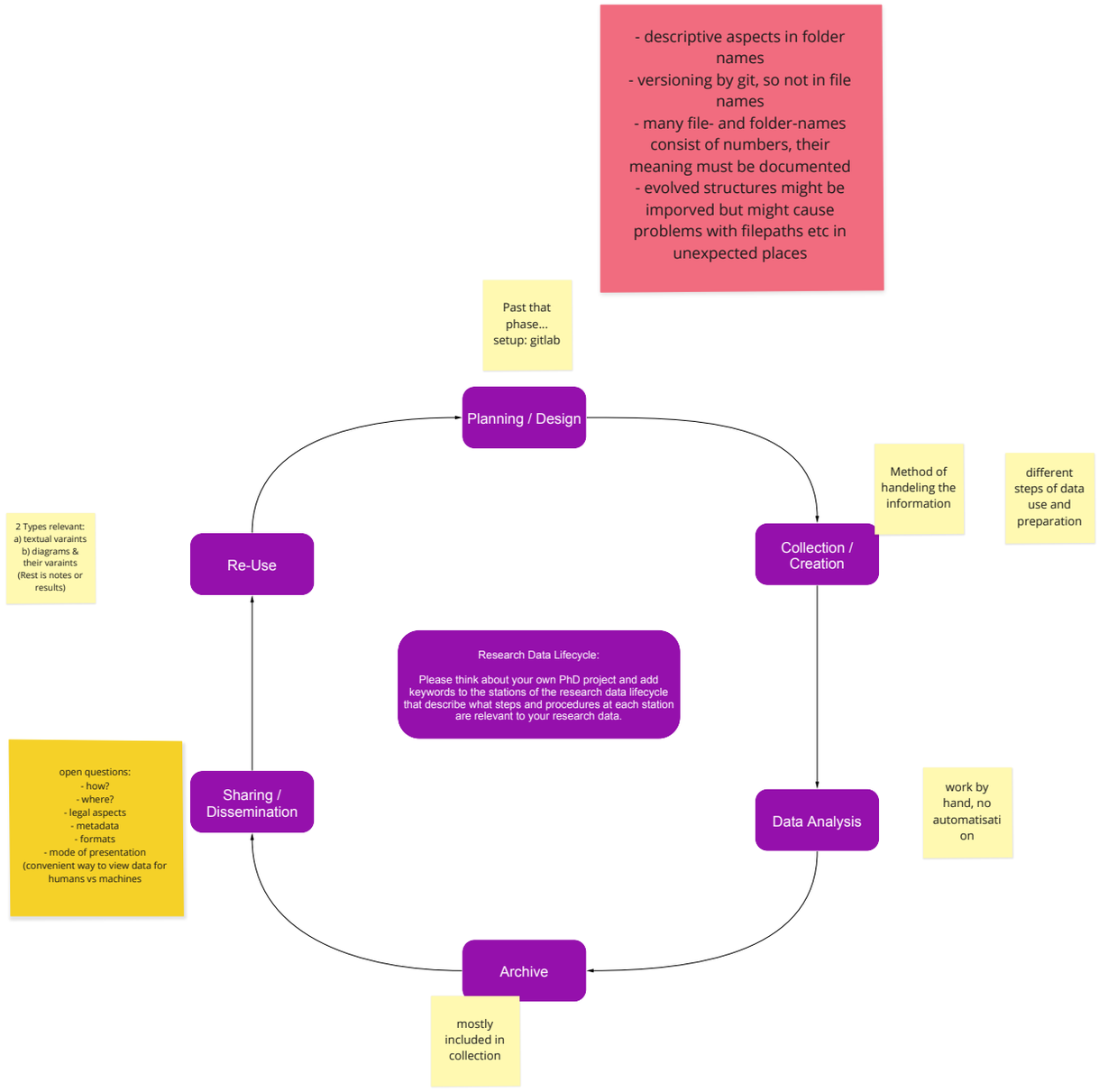
Open repositories (Spatial data, tables, models)

Git (gitLab)

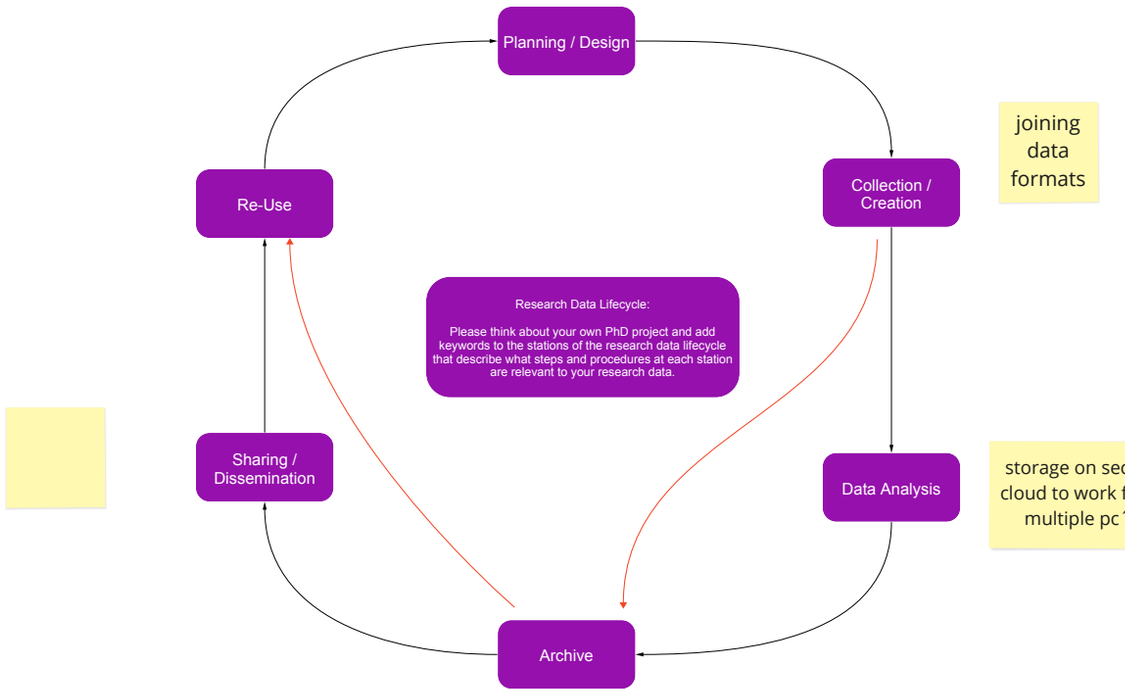
RStudio, Python & QGIS scripts

Git (gitLab) & CAU-Cloud

Lukas Wirth


- descriptive aspects in folder names
- versioning by git, so not in file names
- many file- and folder-names consist of numbers, their meaning must be documented
- evolved structures might be imporved but might cause problems with filepaths etc in unexpected places
Past that phase... setup: gitlab
Planning / Design
Method of handeling the information
different steps of data use and preparation
Collection / Creation
2 Types relevant:
a) textual varaints
b) diagrams & their varaints
(Rest is notes or results)
Re-Use
Research Data Lifecycle:
Please think about your own PhD project and add keywords to the stations of the research data lifecycle that describe what steps and procedures at each station are relevant to your research data.
open questions:
- how?
- where?
- legal aspects
- metadata
- formats
- mode of presentation
(convenient way to view data for humans vs machines
Sharing / Dissemination
Data Analysis
work by hand, no automatisation
Archive
mostly included in collection

Sebastian



joining data formats

storage on secure cloud to work from multiple pc´s

Samples:
[Chem. composition][wire diameter][strain][HT time + temp.]
Mg-2Ag_0.92_0.2_30s_425°C

Corrosion experiment:
[YYYY-MM-DD][Setup][Chem. composition].[suffix]

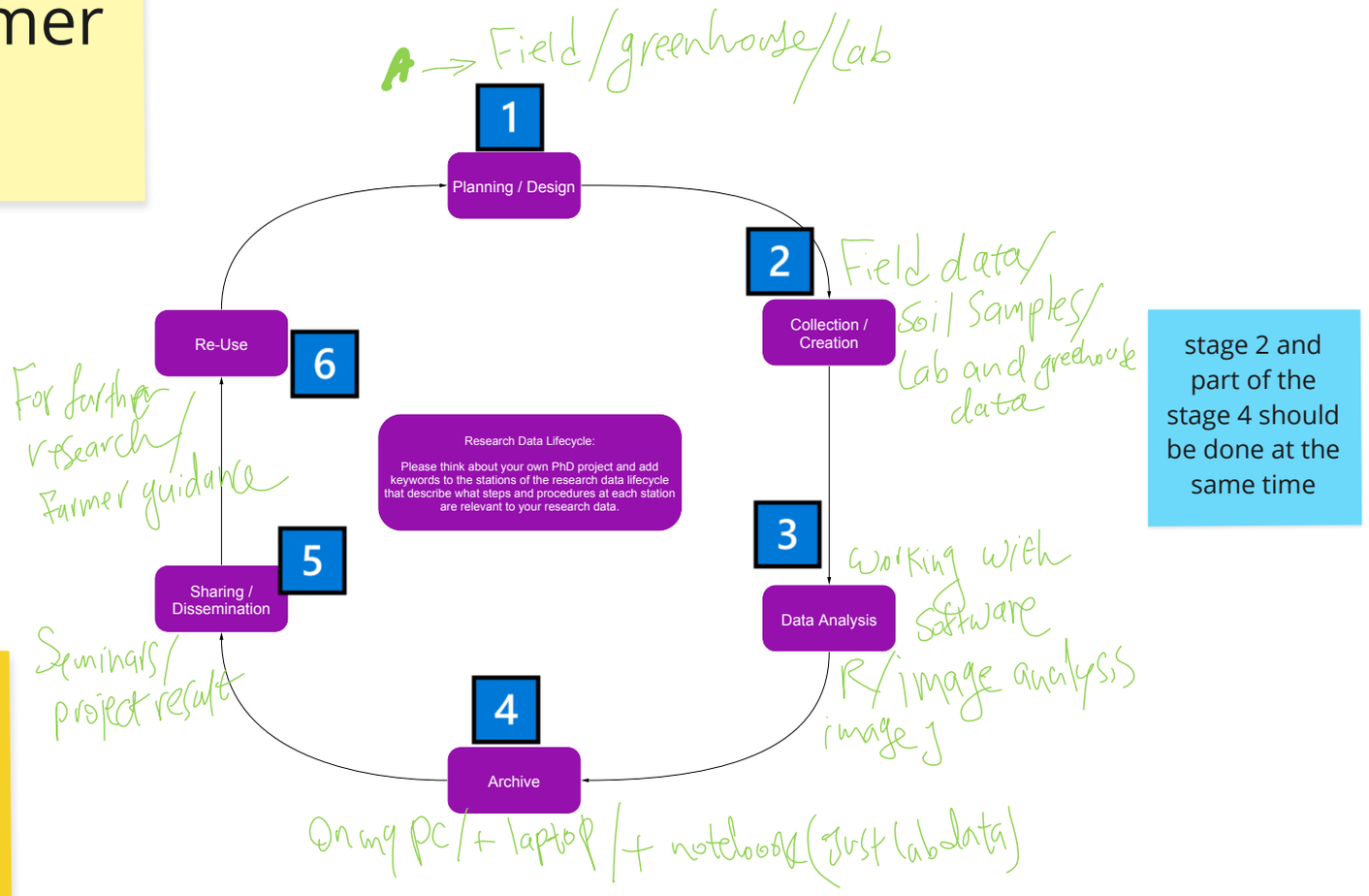
-PhD
  -Raw_data
    -Mechanical_test
    -Corrosion_test
    -Synchrotron
  -Publication
    -Conferences
    -Papers

Sahar Kretschmer

Project main name-(the data from greenhouse or fieldwork or I will add hyperspectral camera data and zymography data)- abbreviation of my name and family name- date (ddmmyyyy). file

Covercrop-greenhouse-SK-20012023.pdf

Covercrop-field-SK-01032023.pdf



stage 2 and part of the stage 4 should be done at the same time

Stephanie

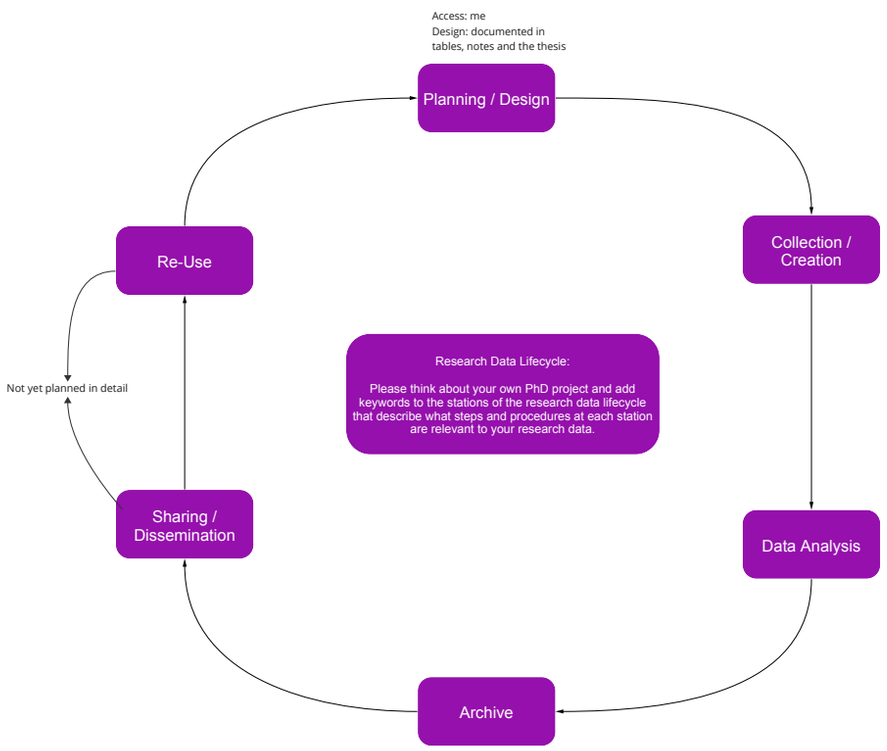
Naming convention

1. Folder structure: topics (perception, speech material/annotations, statistics, thesis) with subfolders
2. structure applies to audio files, annotations, data

[topic]_[speaker/participant/all]_[measurement(if applicable)]_[date(statistics, results, annotations, chapters)]_[version(statistics, results, annotations, chapters)].[suffix]

E.g. (for annotation):
ANN_MF_2023-01-26_3.2.5.TextGrid

In measurement folders:
MF.wav
MF.TextGrid

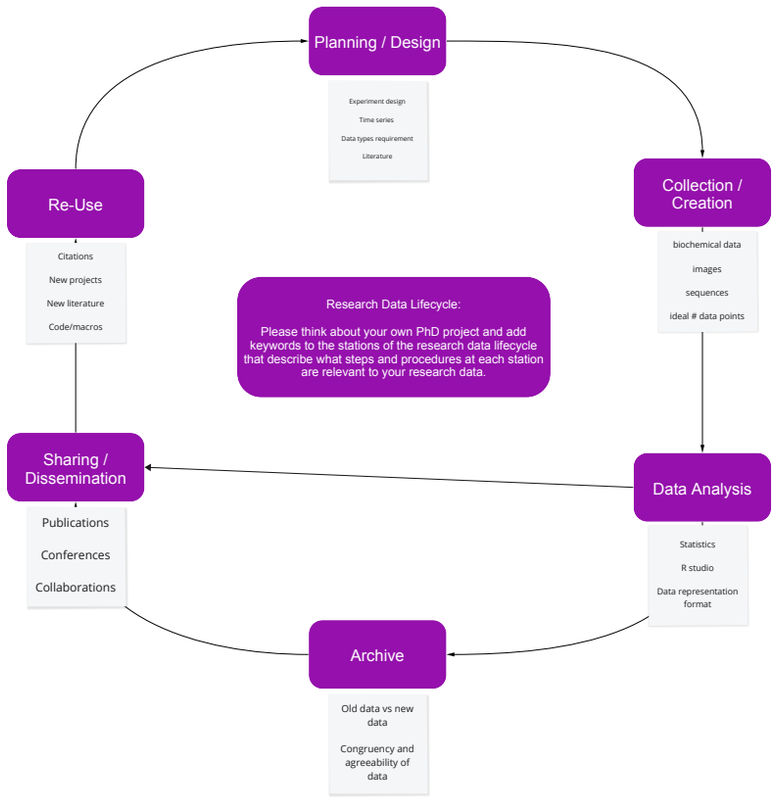


Collection:
Metadata -- LimeSurvey
Experiment data: Praat computer internal, Zoom interview, remote control and screen-sharing
--> data security and copyright issues
Annotations -- Praat
Measurements -- Praat, R scripts, Praat scripts (own and public)

R for statistics and visualizations --> codes and scripts

PhD folder --> folders per topic (perception, speech material, etc.)
Manual file names including versions and changes
Plan: Make master files when annotations, statistics etc. are done (for the time being) and keep only those

Kartikeyan

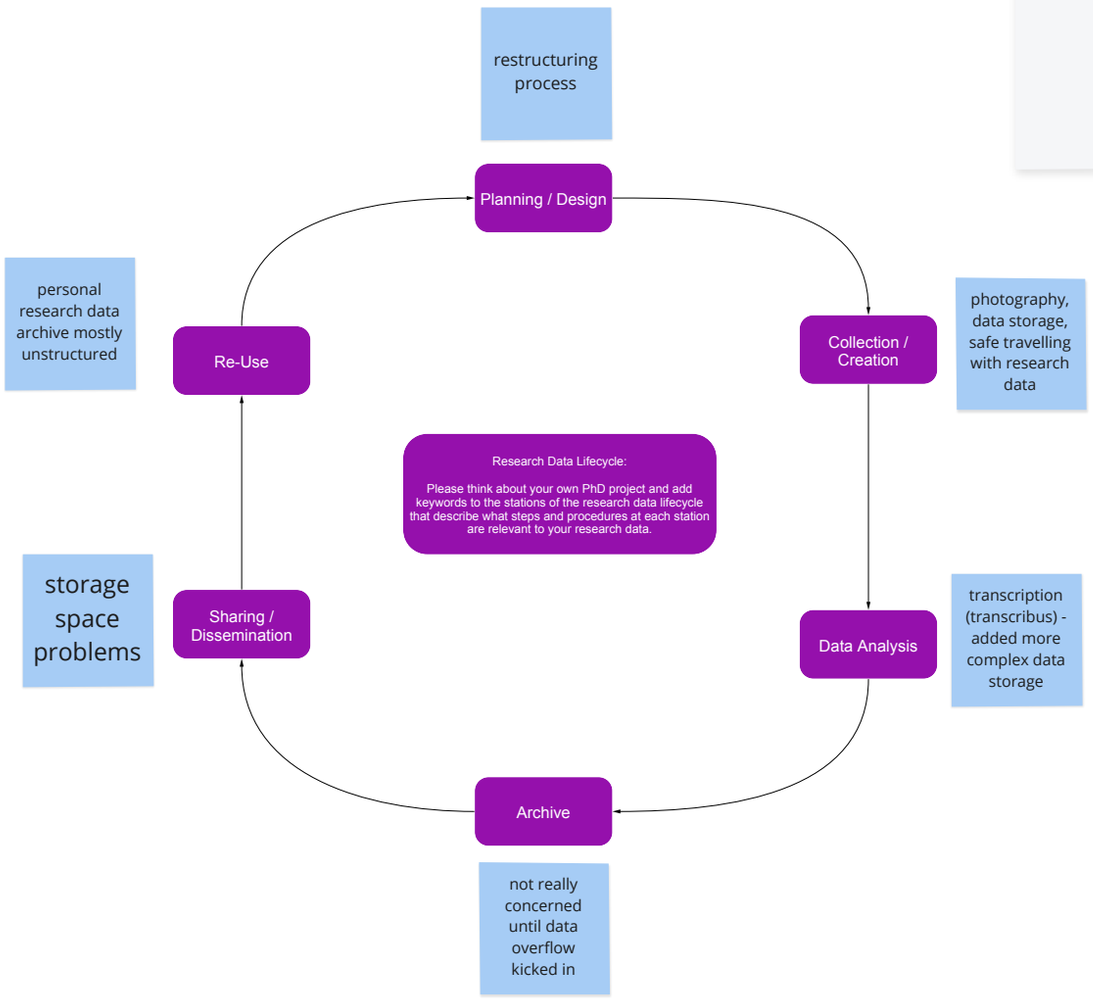


Naming convention

- Nested folders defining experiments
- Experiment with date, experiment conditions eg for images
- Statistics file, in the same folder with the same date and version
- Thesis
- Thesis_Version
- 0.0.0

Per-Ole

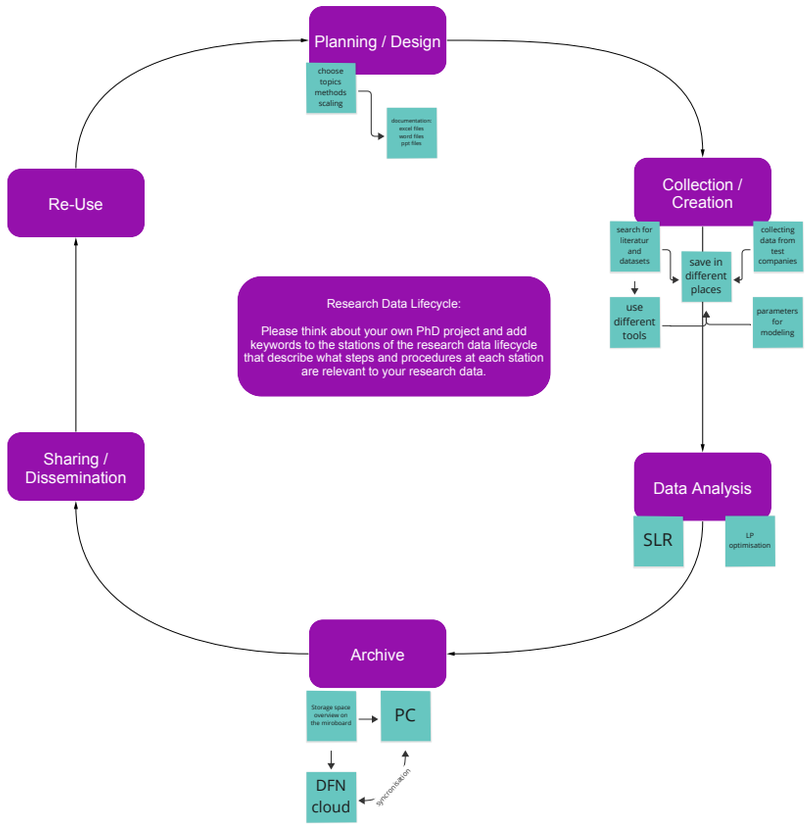
- Folder structure im preparation for the chapters
- Dates diffentiated between historical dates and research date)
- ...
-original file name (archive Numbers)
- abrevations needed (must be comprehensible and documented)
- versions scheme (1.0., 1.2, etc.)



Svea

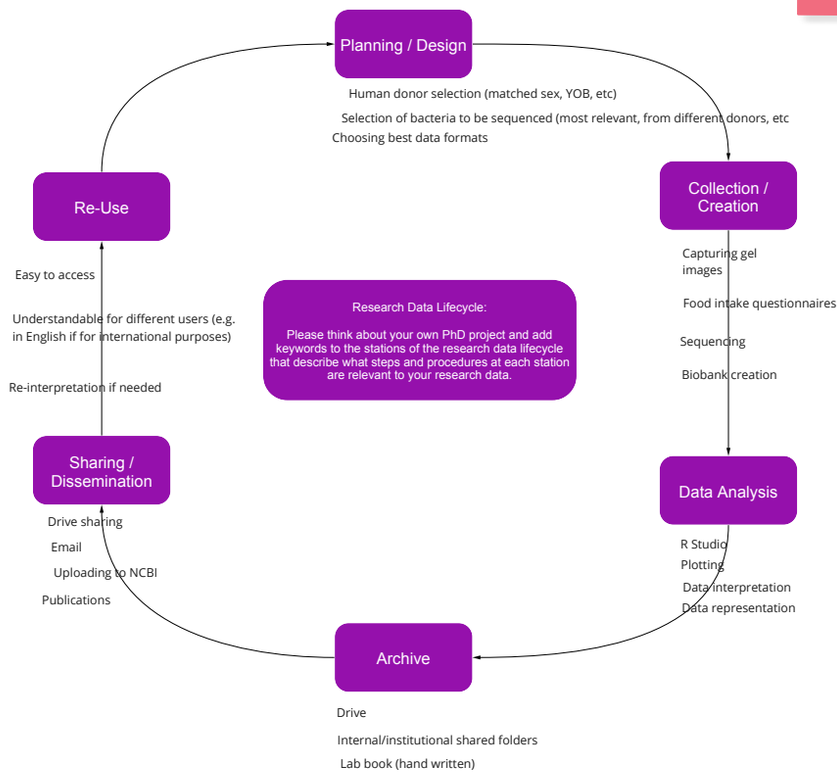
naming convention

Files	.docx .xls .ppt .csv .pdf .jpg .
sketch	
descriptive aspects	Project: Name of the Project in one word Kind: conference, essay, mindmap, koncept, exposee, paper, template, rawdata, documentation
abbrevation	Projects: Diss (PhD stuff in general) Anbauplanung Barrieren
versioning scheme	1.0 Draft 1.1 minor changes 2.0 major change 2.1.1 further major changes
convention	Projekt_Kind_Purpose_Version.sufix



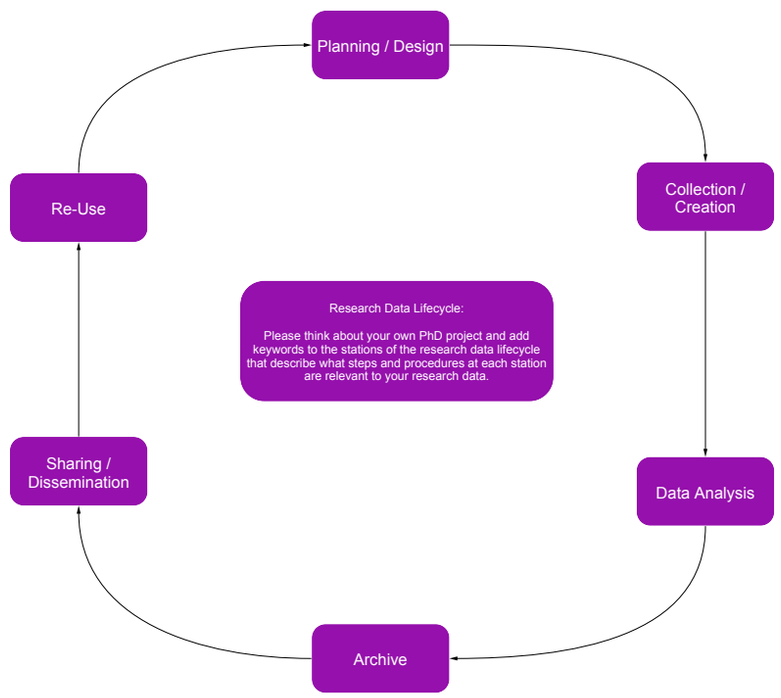
Sofia

Naming convention:
1) Sequencing requests
2) "2023_Sequencing"
3) Date, technique, worker
4) -
5) "2023-01-26_Illumina_Sofia"



Na

observation
constraints



GIT

simulations start

GIT
GEOMAR cloud

ferret., R, python, Matlab

nesh

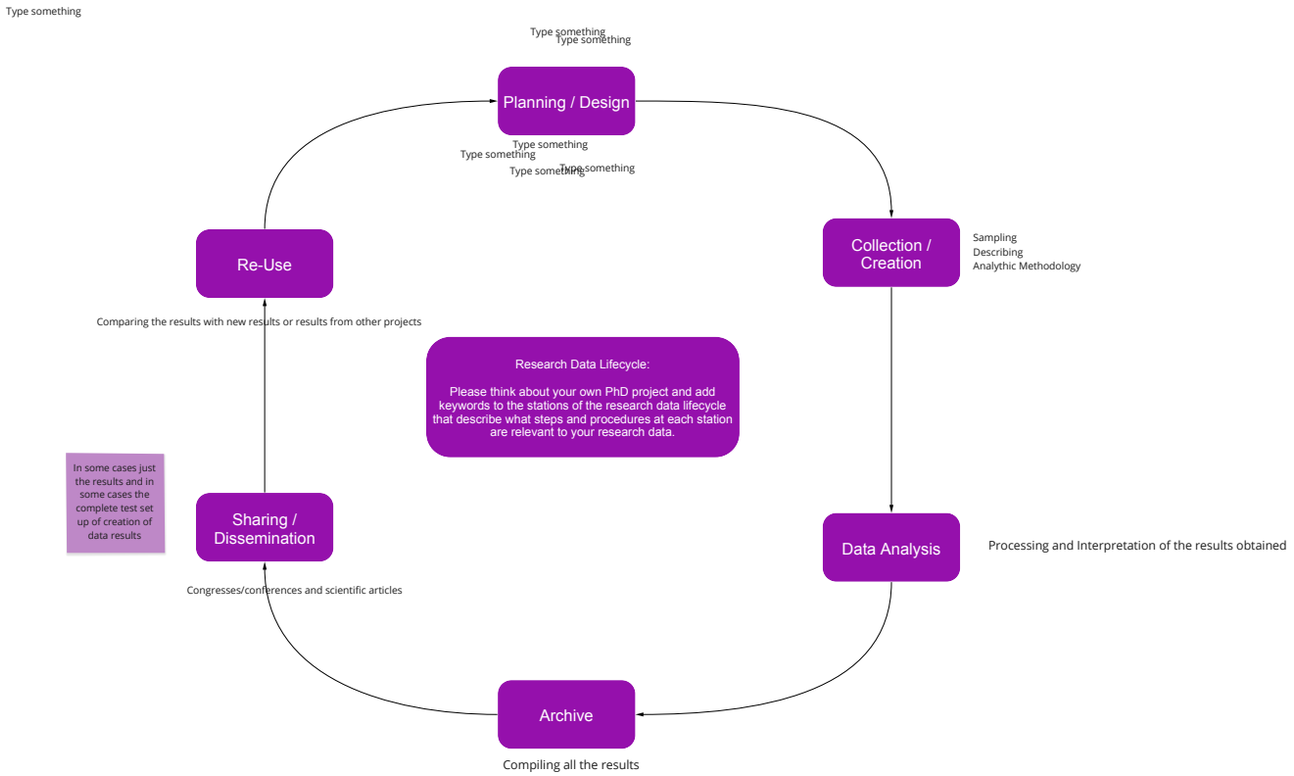
folder structure:
-> research
-> project
-> data;
 scripts;
 reference;
figs;
 template;
 meetings;
 process (for ongoing unsorted stuff)

naming convention:
[task]_[objective]_[approach]_[simulation-ID]_[date].[format-suffix]

Mafalda

Naming Convention:

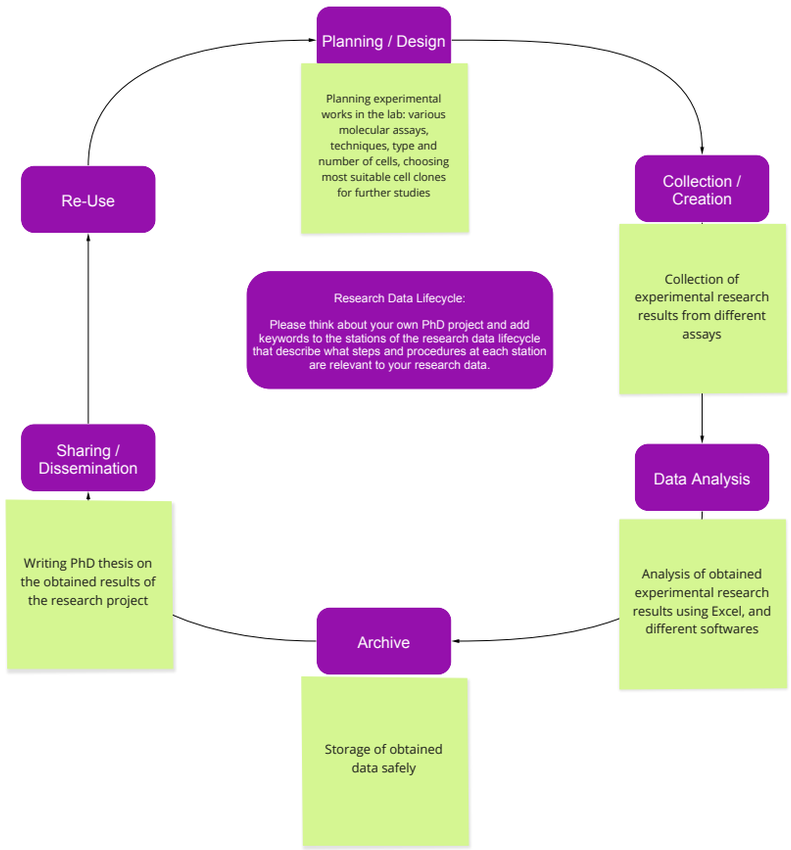
- Results files
  - Folder structure: Analythic Methodology/Tipe-of-Analyses/MudVolcanoName/
    1. Which methodology the result refers to
    2. Which analytic methodology was uses
    3.Which Mud volcano was analysed
  - Abbreviations:
    Ca=Calcium
    Yi= Yinazao mud volcano
    MV= Mud Volcano
  - Version excheme: v1, v2....
  - Date format: YYYY-MM-DD
  - Example of a file name: Isotopes_Ca_Yi-MV_2022-11_v1



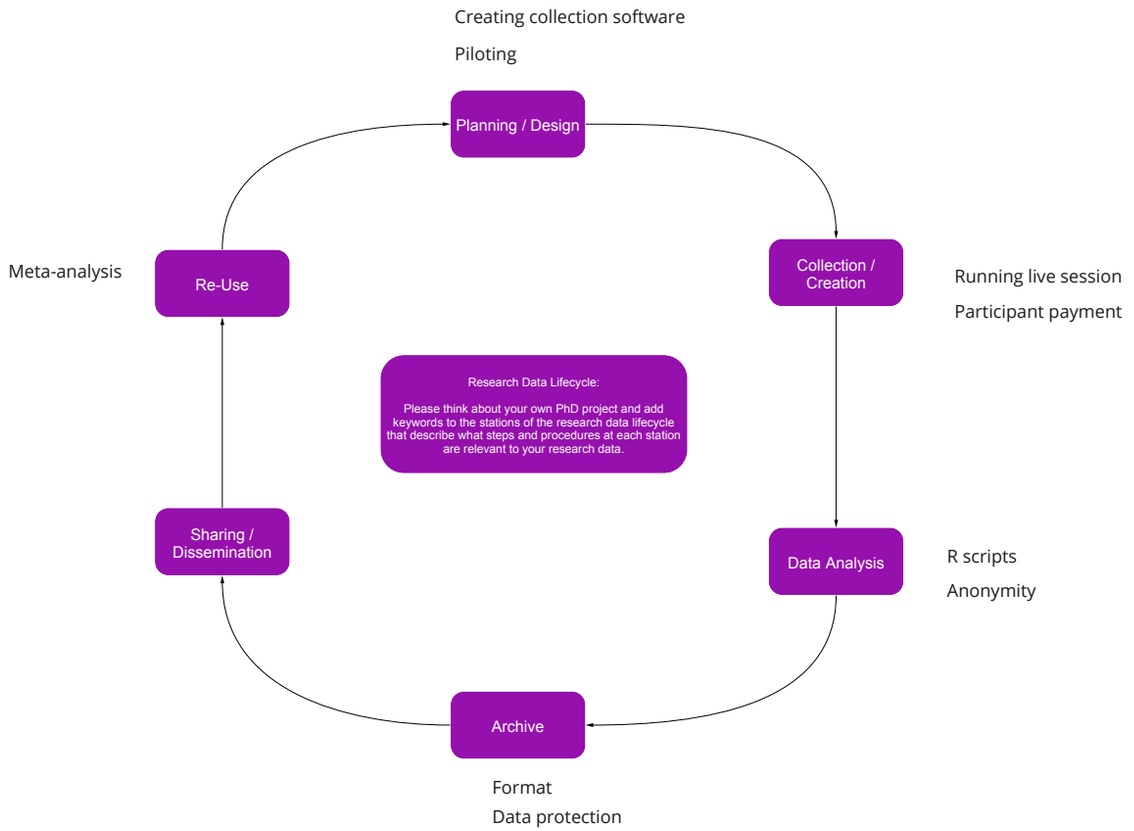
Dinara

Naming convention:
1. Date
2. Name of an experiment
3. Name of a cell clone, a gene or a protein examined / number of a tube

Example: 2023-01-26_ROS_ P4b9_1,5,11



Charlotte

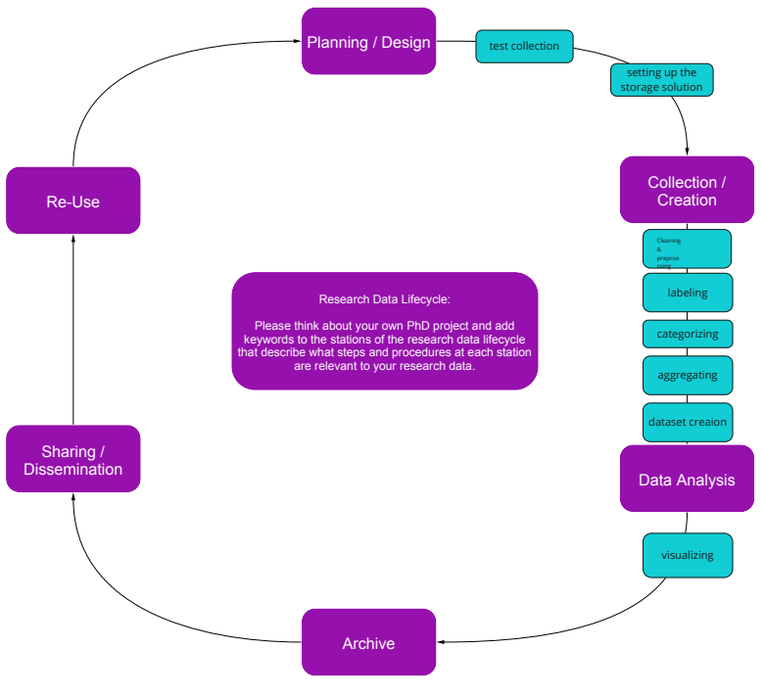


project/collectionstage/game/versions/session

gamename_version_treatment_sessioncode_date_datatype

I use Git

Anastasia

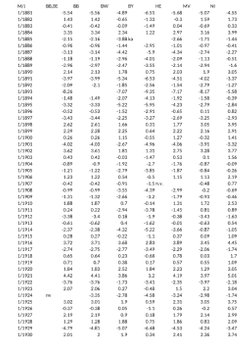

Planning / Design
test collection
setting up the storage solution
Collection / Creation
Cleaning & preprocessing
labeling
categorizing
aggregating
dataset creaion
Data Analysis
visualizing
Archive
Sharing / Dissemination
Re-Use
Research Data Lifecycle:
Please think about your own PhD project and add keywords to the stations of the research data lifecycle that describe what steps and procedures at each station are relevant to your research data.

You are working in a research group working on the ecology of forests and take over data from a previous project: average_d.xlsx

- Speculate what kind of data it could be.

Discuss and take notes

- Apart from the data itself, what information do you need to be able to work with a dataset?
- What do you notice in regard of data quality?



Kind of data:
measurement
per january each year
federal states
absolut? or percentages

further information needed:
units
clarification about missing data
measurement methods (east/west germany)
documentation as a textfile
why grouping federal states?

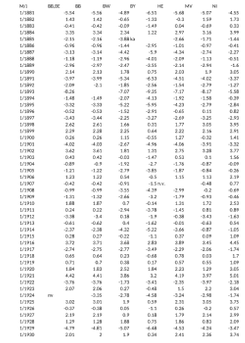
data quality:
unequal
decimals

You are working in a research group working on the ecology of forests and take over data from a previous project: average_d.xlsx

- Speculate what kind of data it could be.

Discuss and take notes

- Apart from the data itself, what information do you need to be able to work with a dataset?
- What do you notice in regard of data quality?



What is shown:

- probably German states in columns
- + an average across columns
- Rows probably January of each year

Data quality issues

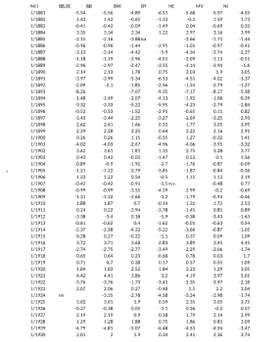
- differing notations for missing data (nothing, na, n.a., ka, k.a.)
- unit missing
- shortened numbers to two decimal spaces
- row labelling
- column labelling

You are working in a research group working on the ecology of forests and take over data from a previous project: average_d.xlsx

- Speculate what kind of data it could be.
    - Annual average of …

Discuss and take notes

- Apart from the data itself, what information do you need to be able to work with a dataset?
    - Abbreviations
    - Units of measurements
- What do you notice in regard of data quality?
    - Missing values named differently / blank
    - No data unit
    - Excel
    - Abbreviations are spaced differently (comma or slash)



Keywords:
- Firm level data
- Executive compensation
- Replication proc[illegible]

Partner work:

Collect all file formats you are working with and post them to the miro-Board.
Give some explanation on your file formats to your partner.
What kind of data hides behind the format?
What kind of software is needed to read it?
Anything special to consider?

.csv/.xlsx
.tif/.czi
.cm5
.docx/.pdf/.pptx
.py/ .R/
.fasta/ .nex /.treefile
etc (phylogenetics)

.txt / .tex / .docx
.csv /.xlsx
.img/.png/.tif
.pdf
.py /.bib
.opju
.pptx

.xlsx
.docx
.ppt
.txt

.nc, .csv, .rds, .R, .pdf, .tex, .m, .jnl, .key, .pptx

txt, LaTeX (etc.) + bib, png, eps, xml, as well as some pptx and xlsx

Group 2

.docx
.xlsx
.pptx
.csv
.pdf
.jpg

.csv
.r
.tif
.asc
.txt
.json
.py
.gz

.csv
.py
.r
.textgrid.
.wav
.table
.xcsl
.txt
.pdf
.tex