



FACULTY OF SCIENCE
Charles University



DATA MANAGMENT PLAN

Are we FAIR yet?

Marian Novotný

What (the hell) is Data Management Plan (DMP)?

- a new bureaucratic obstacle?
- document that describes a life cycle of data
- outlines how data are to be handled both during a research project, and after the project is completed

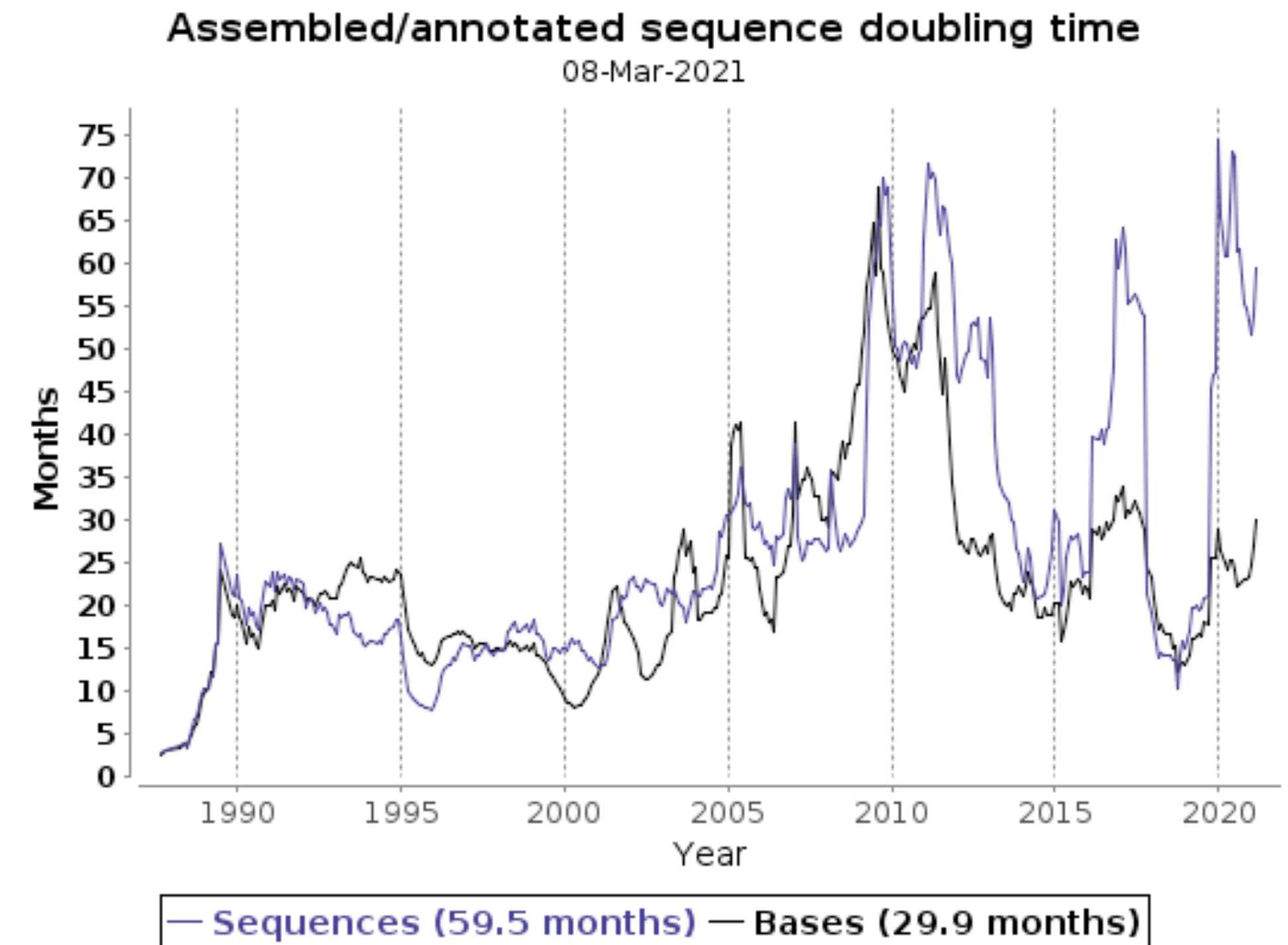
ELIXIR Norway:
Data life cycle management





How did we get to DMP?

- first used in aeronautical projects (1966)
- later used in big projects (e-commerce)
- many projects are really big now
- DMP's are required by 86% UK research councils and newly also in EU projects





What is DMP good for?

- the goal is to consider “all” aspects of data management BEFORE the projects starts
- if the data are organised, formatted and annotated well, it saves a lot of time, effort and money during the project
- it helps to present the data after a completion of the project
- it facilitates storage and archiving of the data
- winners shall be funders, research organisations and RESEARCHERS themselves!

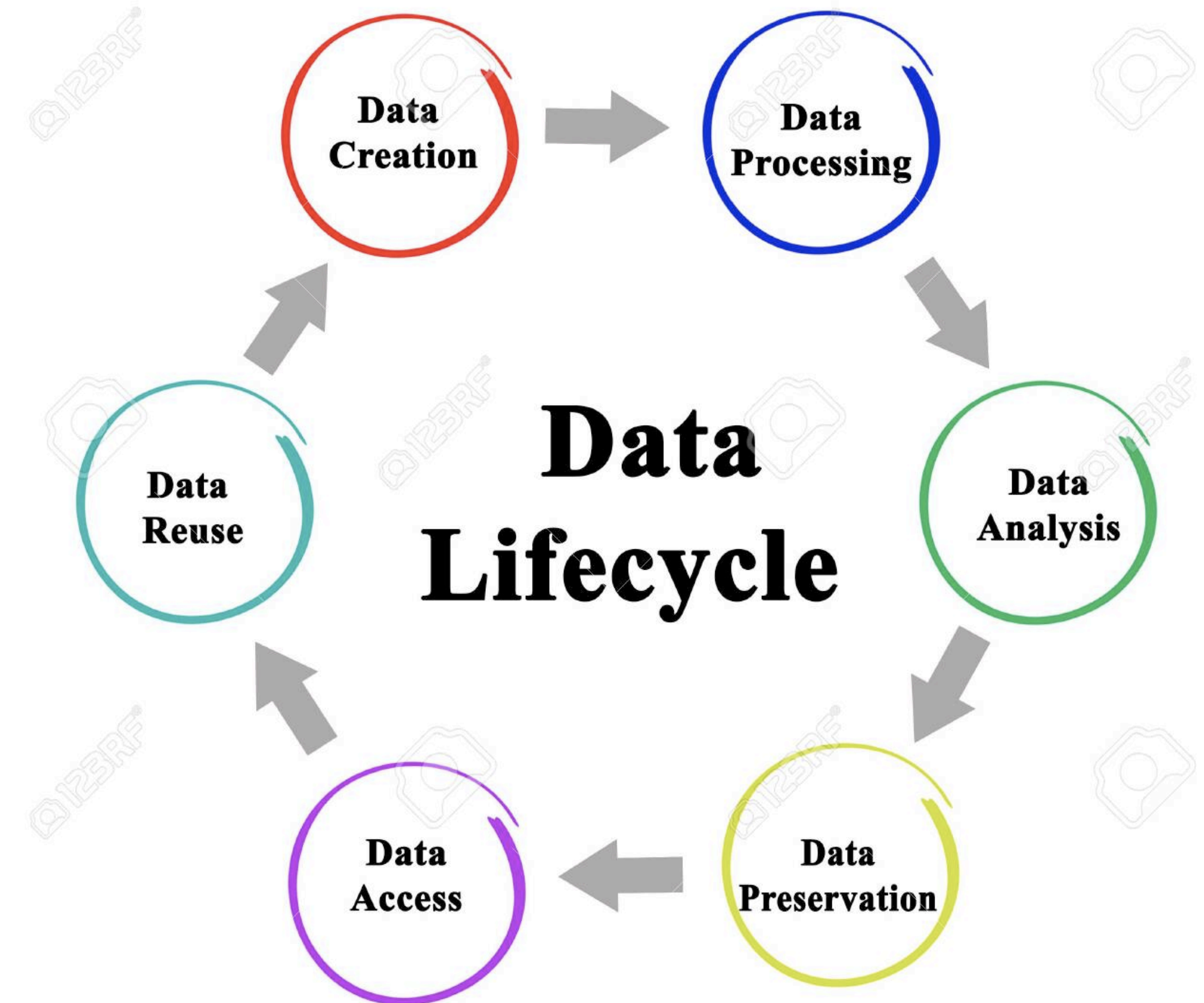


Why shall I bother with DMP?

- Professional benefits - time spared, anticipate problems, data loss or data breach, data continuity and consistency, make funders happy
- Economical benefits - realistic budget, data sharing
- Institutional benefits - better compliance and planning with institution

Data Management Plan - components

- Information about data and data format
- Metadata content and format
- Rules for data sharing and Reuse
- Long-term storage of data
- Budget





Information about data a data format

- Will I produce the data? Will I reuse the data?
- How much data will be produced? And where and how?
- How will be the data analysed? Do I have the tools and resources to do that?
- What will be the file formats?
- How will be responsible for dealing with data?
- How will be the data protected?
- Version control, used language?



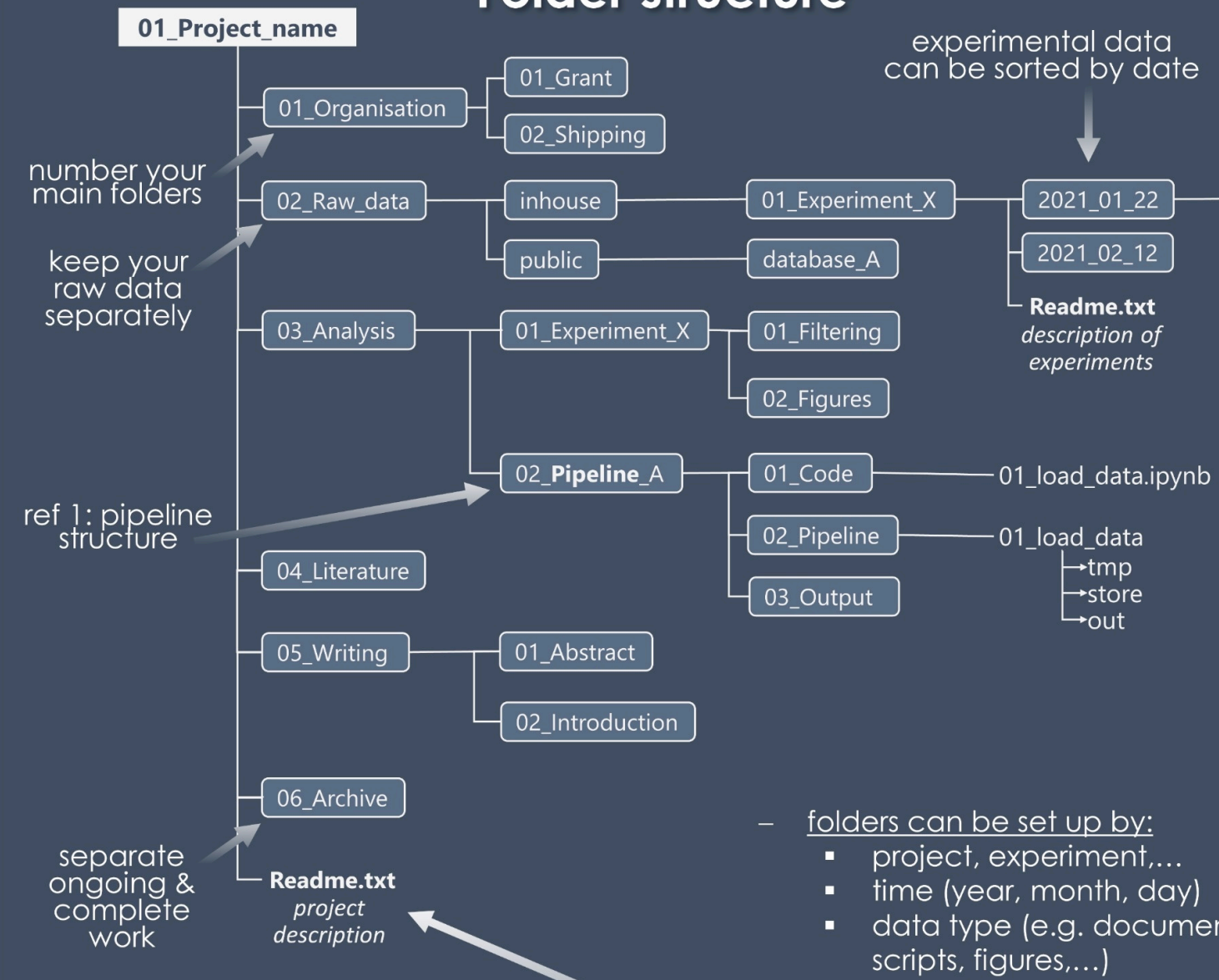
Data management tips

GOAL of good data management
→ optimise the discovery & reuse of data

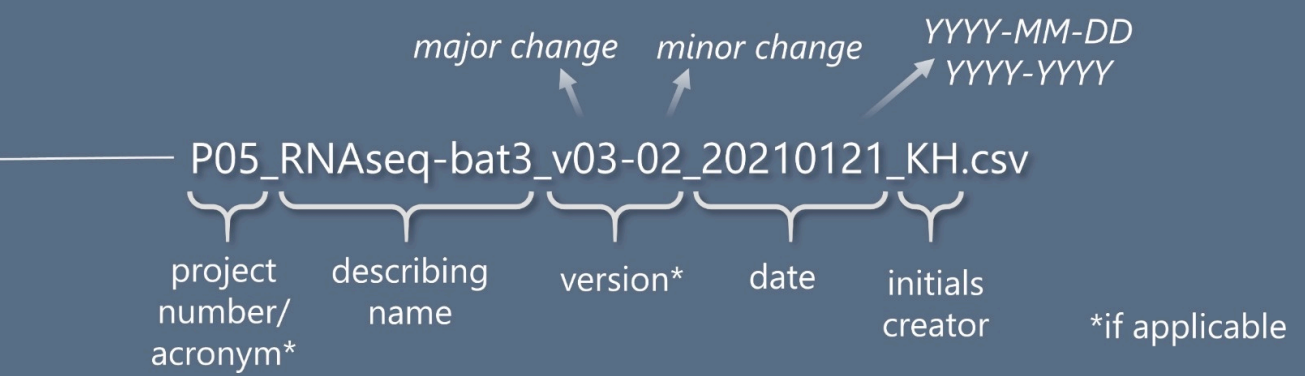
Questions to ask yourself

Are my files organised in a way that I can easily find what I am searching for?
What information would I need to understand and use my data in 20 years?
Could others understand and use my data?

Folder structure



File naming



- #### General naming tips for folders & files
- use unique, meaningful names
 - not too long (not >30-40 characters)
 - no spaces, dots, or special characters (\$"%!'&.*^()+=[:{}~@)
 - hyphens (-) & underscores (_) to separate elements

find the balance between a shallow & deep folder hierarchy

- too deep → too many clicks might be needed to get to the right file
- too shallow → too many files might end up in one folder (organise them in subfolders)



- folders can be set up by:
 - project, experiment,...
 - time (year, month, day)
 - data type (e.g. documents, scripts, figures,...)

Friendly Reminder
Comment your code! 😊

Metadata

Which information is necessary to interpret, understand, and use a given dataset?

readme.txt files
can be used to describe projects, folders, and files

References

¹ <https://towardsdatascience.com/how-to-keep-your-research-projects-organized-part-1-folder-structure-10bd56034d3a>
<https://www.wur.nl/en/Value-Creation-Cooperation/WDCC/Data-Management-WDCC/Doing/Organising-files-and-folders.htm>
<https://www.massey.ac.nz/massey/research/library/library-services/research-services/manage-data/organise.cfm>
<https://library.bath.ac.uk/research-data/working-with-data/organising-data>
<https://www.helsinki.fi/en/research/organizing-data-folders-with-5sdata-method>
<https://mantra.edina.ac.uk>
<https://old.dataone.org/education-modules>

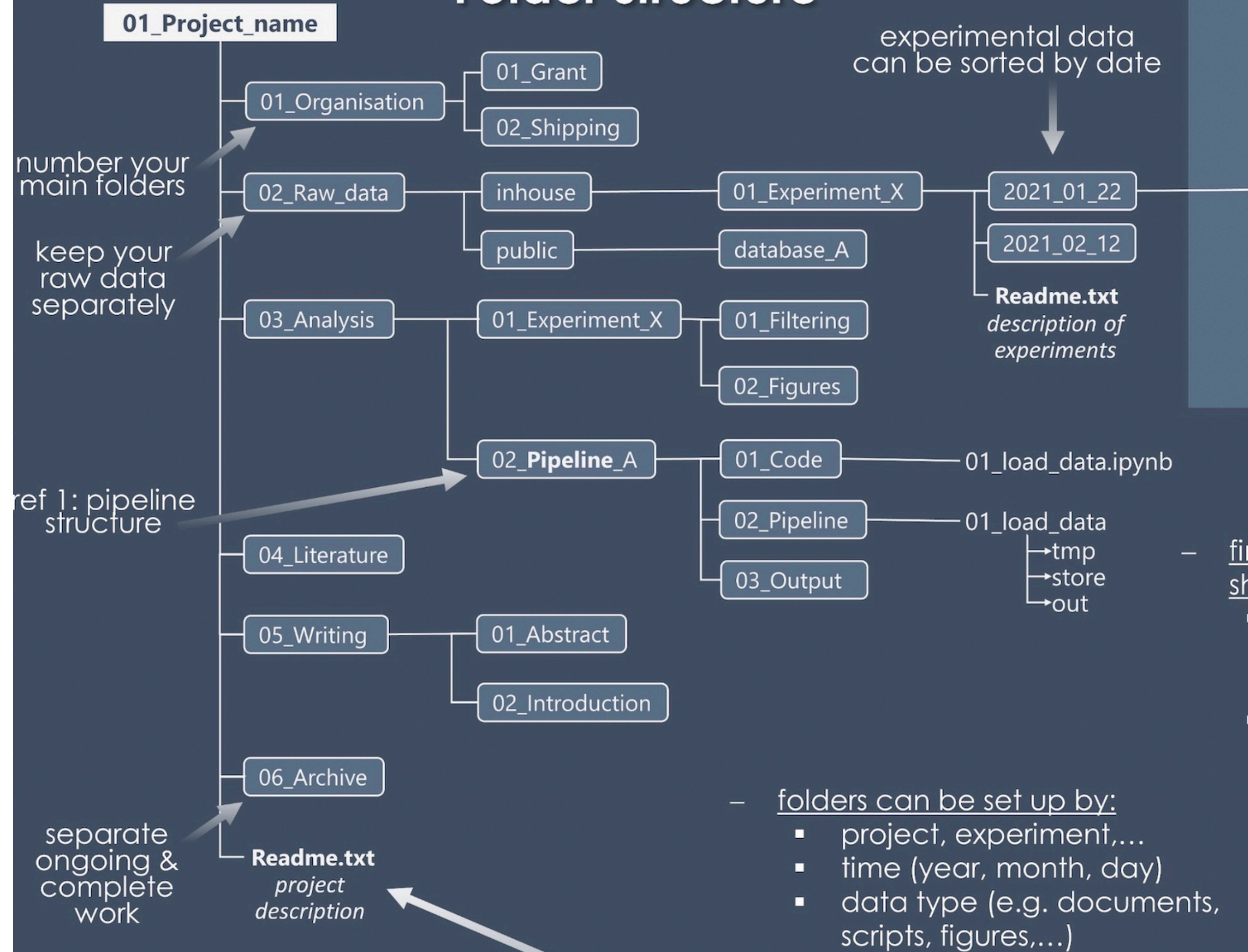


Important information

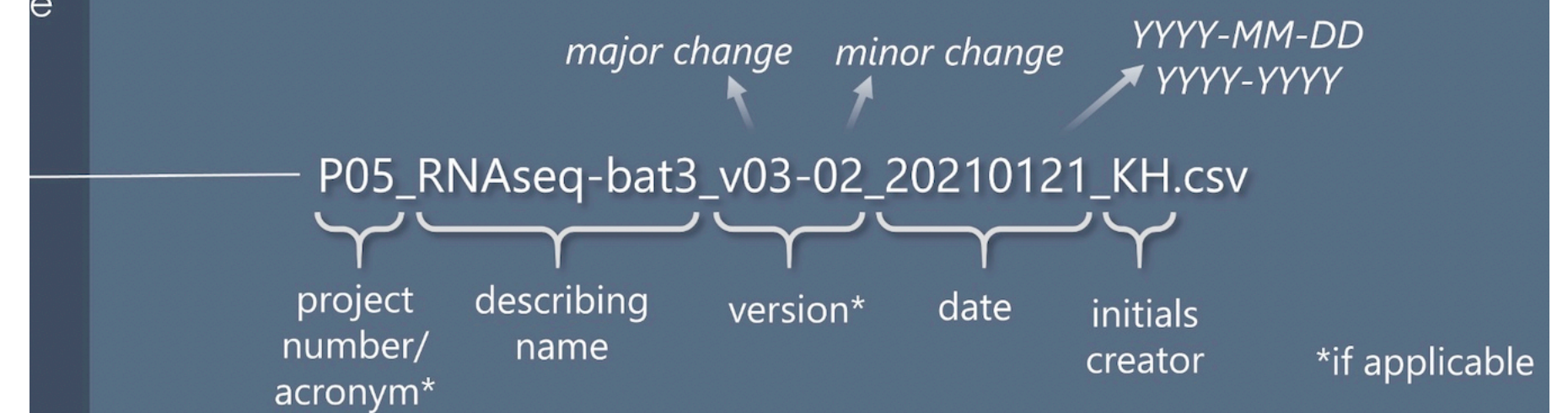
- Who** has created the data?
- What** is the content of the data?
- Which** questions have been answered?
- When** were the data created?
- How** were the data developed (methods)?
- Why** were the data developed?
- With whom** can the data be shared?



Folder structure



File naming



General naming tips for folders & files

- use unique, meaningful names
 - not too long (not >30-40 characters)
 - no spaces, dots, or special characters (€"\$%!"&*^()+=[:{}~@)
 - hyphens (-) & underscores (_) to separate elements
- find the balance between a shallow & deep folder hierarchy
- too deep → too many clicks might be needed to get to the right file
 - too shallow → too many files might end up in one folder



Metadata content and format

- Metadata are contextual “data about data”
- Descriptive - dates, organisms, keywords - precision medicine
- Structural, copyright rules, permissions, statistical analysis
- How will you generate metadata (lab notebooks, GPS units)?
- What metadata will you generate?
- Are there any standard formats for metadata in your field?



Access, sharing and reuse

- Define conditions for access, sharing and reuse
- Will be the data served in open access regime?
- Shall the data be protected?
- How shall they be cited?
- Will they be used after the completion of project? How?
- Any ethical issues?



Long term storage and archiving

- Identify how will be the data archived early in the project to have the data formats in agreement with archive
- What part of the projects needs to go into long-term storage?
- Who will be responsible for long-term storage?



Money, money, money...

- DMP helps to estimate realistic costs of data management
- Data management is not for free, having a plan makes it realistic to take care of data in future
- Personal costs - data preparation, documentation, archiving
- Software and hardware costs
- How will be these costs covered?



Do I need to and How could I ever do a DMP?

- When shall I consider preparing DMP?
- Do I need to employ a data scientist?
- Tools to help are available - DMP online, DSW (Data Stewardship Wizard)



FACULTY OF SCIENCE
Charles University

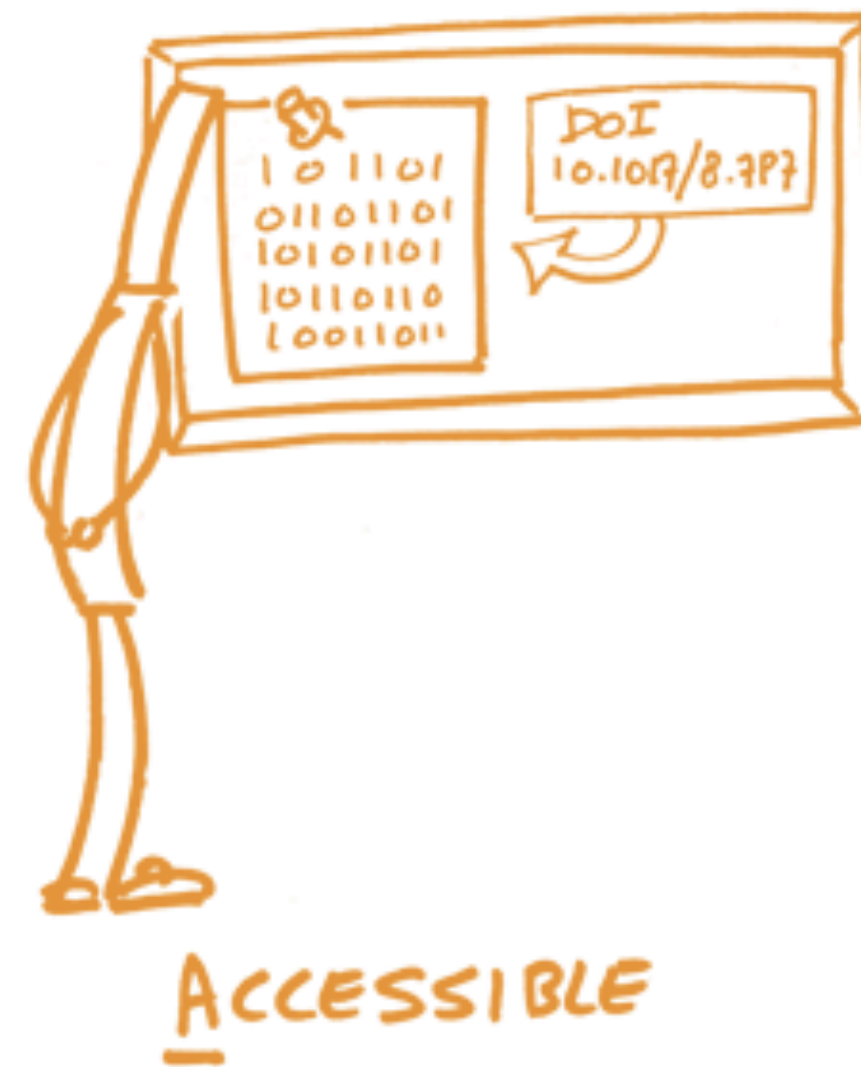


Are your data FAIR?

WHAT are FAIR data?

- EU funded projects require the data to be open access and follow FAIR principles

FAIR DATA PRINCIPLES





Findable

- How to find your data?
- Can computer find your data?
- Do you use persistent digital identifiers?
- Do you have naming conventions and version control?
- What about metadata and its format?



Accessible

- What will be accessible?
- Where will you make data accessible?
- What will be needed (software, protocols) to access the data?
- How will you regulate access and authenticate users?
- What metadata will be provided?
- Will you make metadata accessible even if the original data are not?



Interoperable

- Can you combine the data with other tools and datasets?
- Can you combine your metadata with other datasets?
- Do you use standard vocabularies for the data types you use?



Reusable

- How will be the data licensed?
- When will you make your data available?
- How long will they be reusable?
- Why do you put restrictions on reusability of data?



FAIR DMP

- There is an overlap between DMP and FAIR principles
- Some tools for DMP can tell you how FAIR will be your data
- Institution shall provide solid support for data management and DMP preparation



Summary

- DMP is scary, but also an opportunity to learn and profit more from your data
- Funders are likely to require DMP soon for majority of projects and hopefully will provide funds to treat the data properly
- Institutions will learn to provide support - it will be useful for them, too
- Open access data are vital for the progress of science!



FACULTY OF SCIENCE
Charles University





FACULTY OF SCIENCE
Charles University





FACULTY OF SCIENCE
Charles University





FACULTY OF SCIENCE
Charles University

