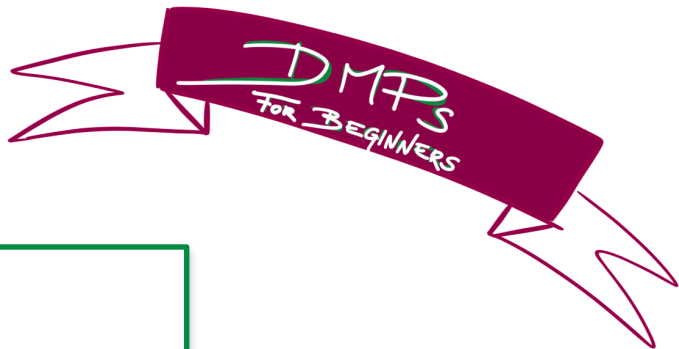


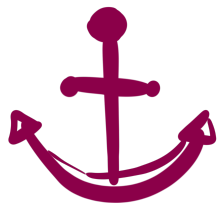
# BOARDING ...



*We will start at 10:00.*

- Please **turn on** your camera
- Please unmute yourself to **speak**
- The session will **not** be **recorded**
- Material & slides will be **provided**
- You might need pen & paper





# DATA MANAGEMENT PLANS FOR BEGINNERS

Jeanne Wilbrandt — Data Steward



**WELCOME!**



# WELCOME — OUR BACKGROUND

---

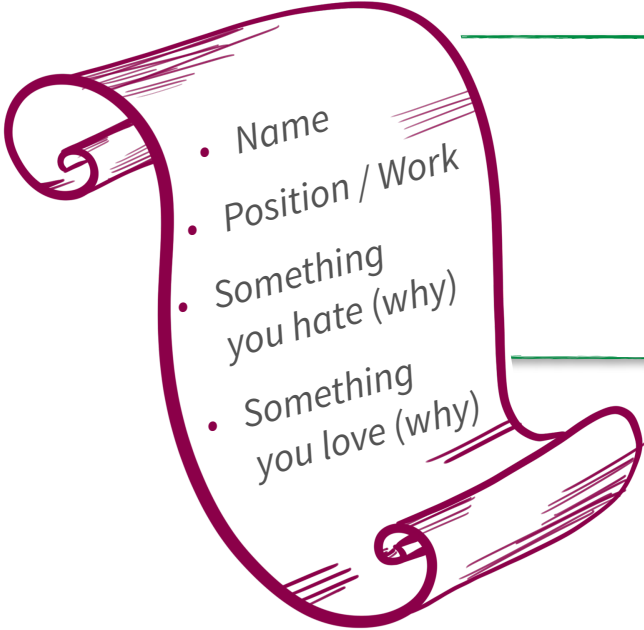
- National, academic and non-profit infrastructure
- Providing bioinformatics services and training to users in life sciences research and biomedicine
- Research focus on stem cell aging, regeneration, molecular damages and epigenetics of aging
- Research focus on principles of evolution, development and adaptation operating in crop plants





# WELCOME — WHO ARE WE?

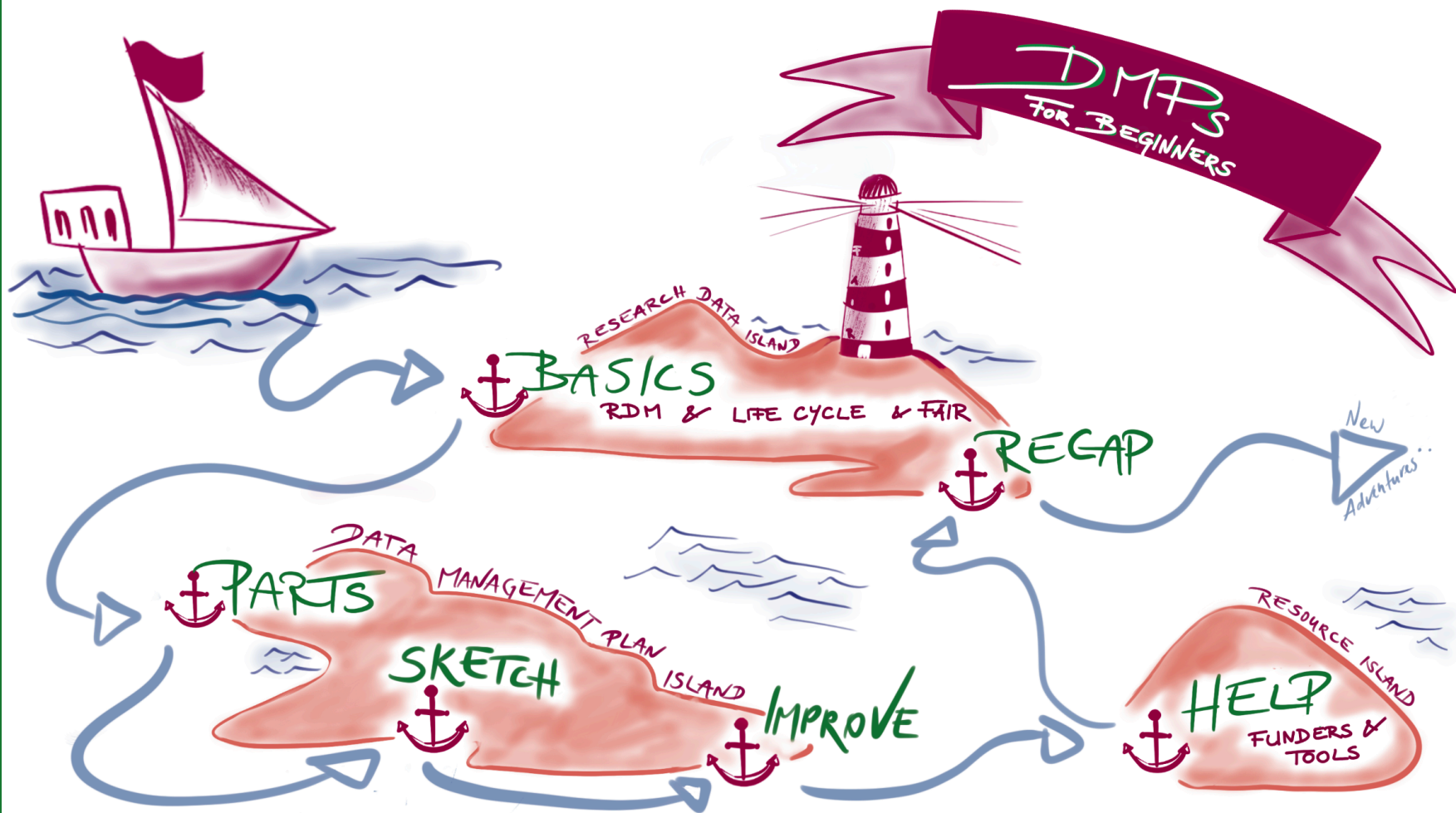
---

- 
- Name
  - Position / Work
  - Something you hate (why)
  - Something you love (why)

## *Link Chain*

Everyone





# WELCOME — AGENDA TODAY

---



Basics: RDM & FAIR

*Bio-Break*



Parts of a DMP

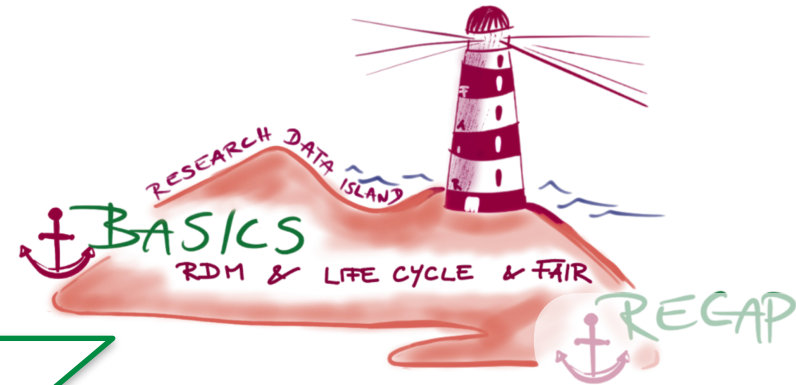


Writing a DMP Sketch





# RESEARCH DATA MANAGEMENT



# WHAT IS RESEARCH DATA?



Research Data?



+

Metadata?



# WHAT IS RESEARCH DATA?

## Research Data?

All digital data that are **used** or **produced** in the research process or are its **result**

- raw and processed data
- simulation data
- observational data
- code, algorithms, scripts
- audio-visual data



## Metadata?

**Data on data:**  
Provenance, genesis,  
background information

- who, when, why, how
- used resources
- used abbreviations, units, names
- licenses
- etc ...

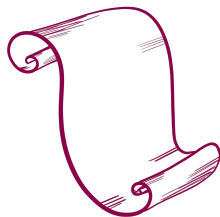


# WHAT IS DATA MANAGEMENT?

---



*RDM describes the entirety of all handling measures of research data.*

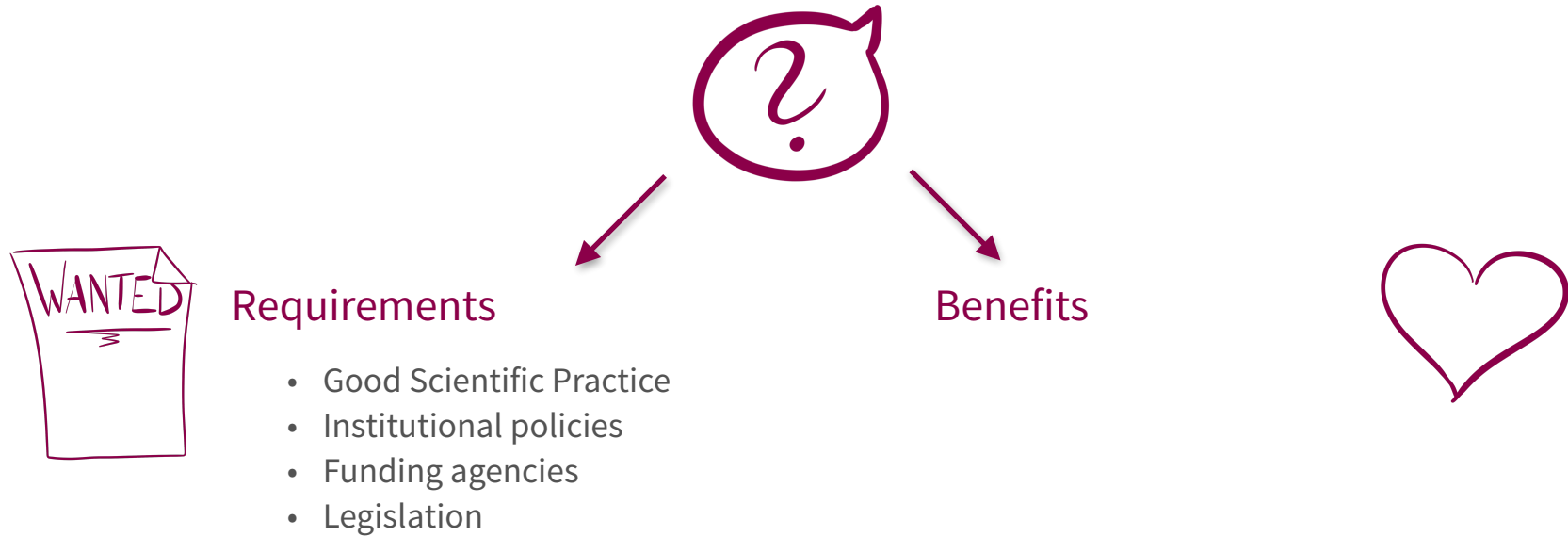


## Task fields (selection):

- Generation
- Processing
- Documentation
- Storage
- Publication
- Archiving

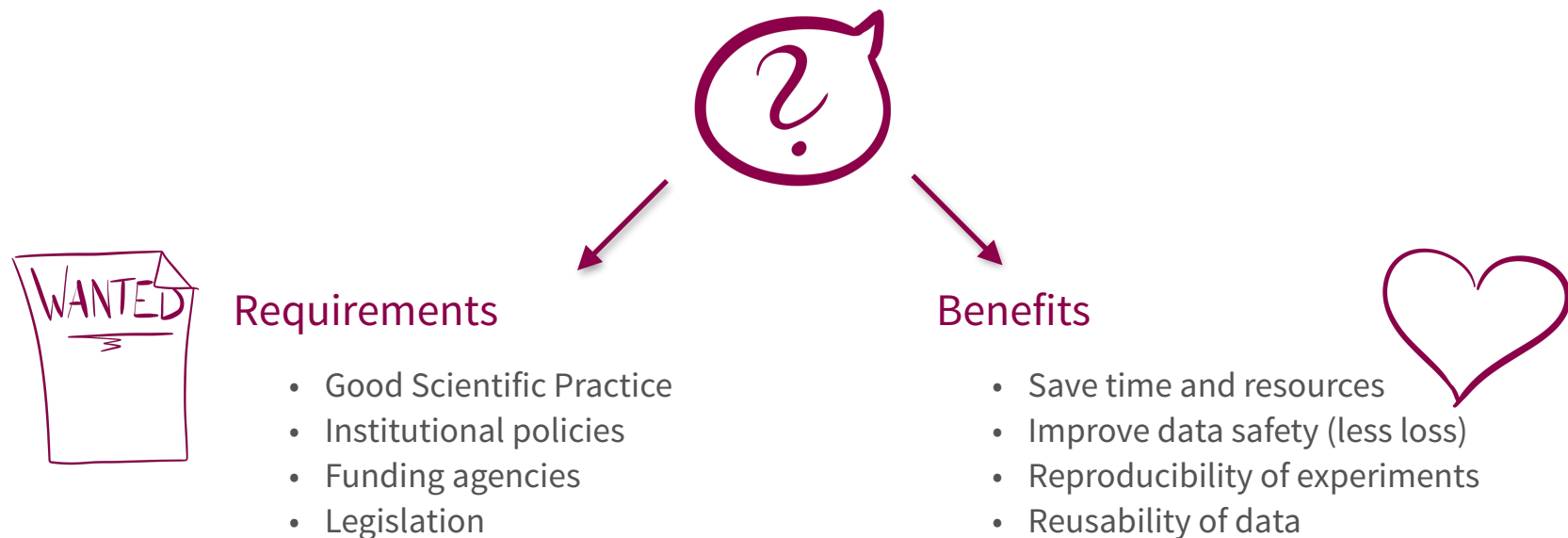


# WHY DATA MANAGEMENT?

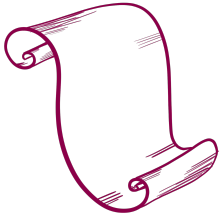




# WHY DATA MANAGEMENT?



# DATA MANAGEMENT: WHEN?



Task fields (selection):

- Generation
- Processing
- Documentation
- Storage
- Publication
- Archiving



*along data life cycle*



# RESEARCH DATA LIFE CYCLE



# RESEARCH DATA LIFE CYCLE



## Active project phase

- Identify existing data
- Storage and Backup
- Data organization
- File formats
- Data quality
- Documentation
- Safety and access

# RESEARCH DATA LIFE CYCLE

## Postproduction

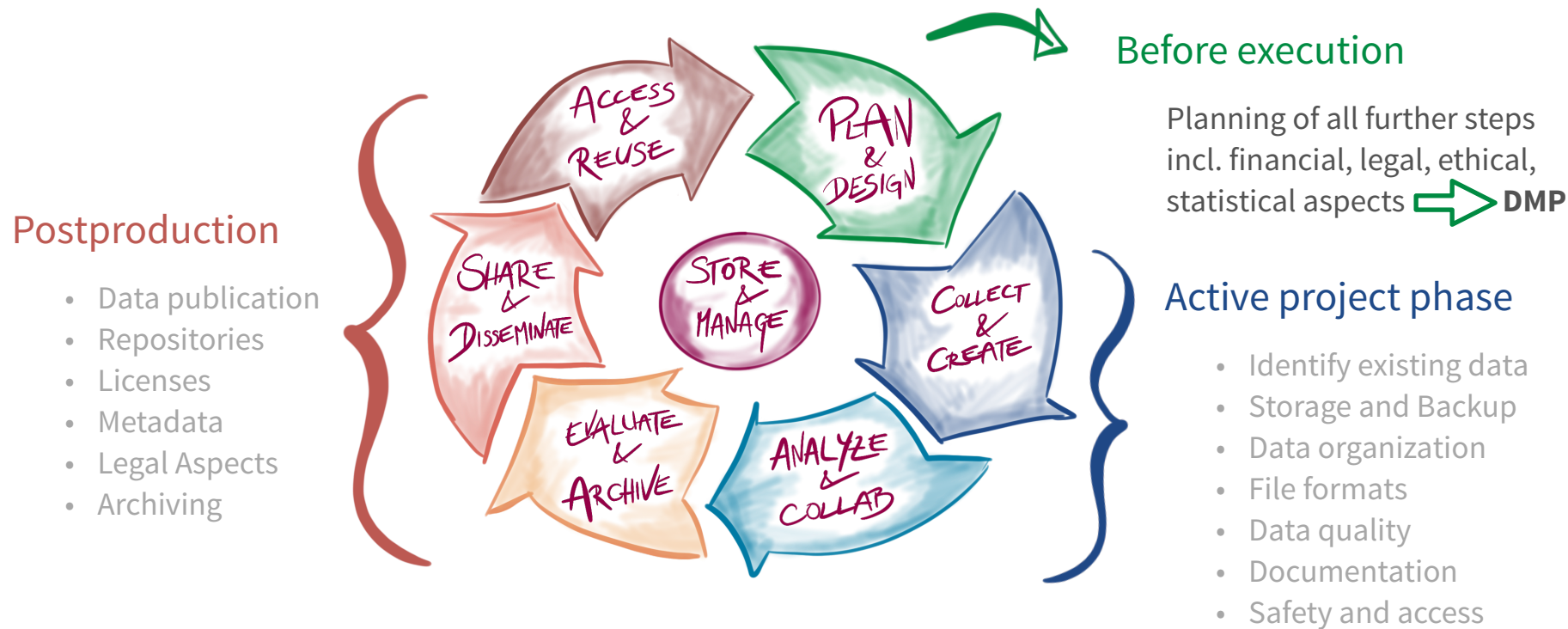
- Data publication
- Repositories
- Licenses
- Metadata
- Legal Aspects
- Archiving



## Active project phase

- Identify existing data
- Storage and Backup
- Data organization
- File formats
- Data quality
- Documentation
- Safety and access

# RESEARCH DATA LIFE CYCLE



# RESEARCH DATA MANAGEMENT

"Research data management (RDM)  
[has] the common goal of  
making the data **accessible**, **reusable** and  
**verifiable** in the **long term** and **independent**  
**of particular people**."

[forschungsdaten.info](https://forschungsdaten.info)



# FAIR (META)DATA PRINCIPLES



**FINDABLE**



+

**ACCESSIBLE**



+

**INTEROPERABLE**



=

**REUSABLE**



*... for machines and humans*





# FAIR (META)DATA PRINCIPLES

## FINDABLE



- Permanent ID
- Indexed in searchable resource

+

## ACCESSIBLE



+

## INTEROPERABLE



=

## REUSABLE



... for machines and humans

# FAIR (META)DATA PRINCIPLES

## FINDABLE



- Permanent ID
- Indexed in searchable resource

+

## ACCESSIBLE



- Defined retrieval protocol and license
- Authentication & authorization

+

## INTEROPERABLE



=

## REUSABLE



*... for machines and humans*

# FAIR (META)DATA PRINCIPLES

## FINDABLE



- Permanent ID
- Indexed in searchable resource

+

## ACCESSIBLE



- Defined retrieval protocol and license
- Authentication & authorization

+

## INTEROPERABLE



- Formats are open, free, and commonly standardized
- Easily harmonized

=

## REUSABLE



... for machines and humans

# FAIR (META)DATA PRINCIPLES

## FINDABLE



- Permanent ID
- Indexed in searchable resource

+

## ACCESSIBLE



- Defined retrieval protocol and license
- Authentication & authorization

+

## INTEROPERABLE



- Formats are open, free, and commonly standardized
- Easily harmonized

=

## REUSABLE



- License
- Information on genesis
- Conforming relevant standards

*... for machines and humans*



Where is ?!  
the page..



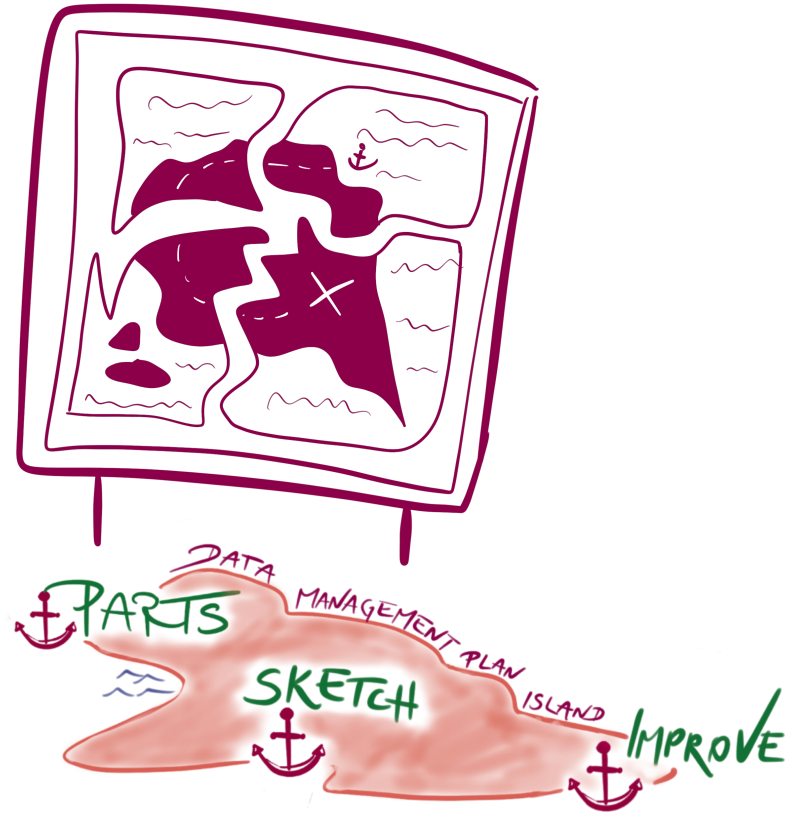
I don't know

No BACKUP?!!



## BIO-BREAK

# PARTS OF A DMP

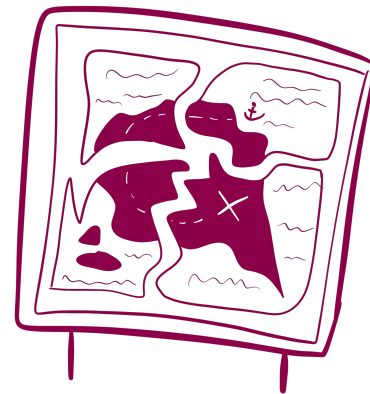


# WHAT ARE DATA MANAGEMENT PLANS?

---



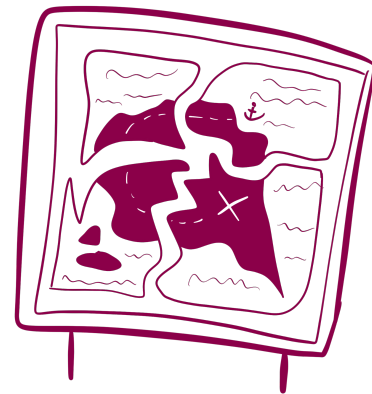
*A DMP is a formal, living document that outlines what you will do with your data during and after a research project.*



# WHAT ARE DATA MANAGEMENT PLANS?



*A DMP is a formal, living document that outlines what you will do with your data during and after a research project.*



- *What is covered by a DMP?*
- *What is the structure of a DMP?*
- *How does my work fit into a DMP?*
- *How do I phrase DMP texts?*



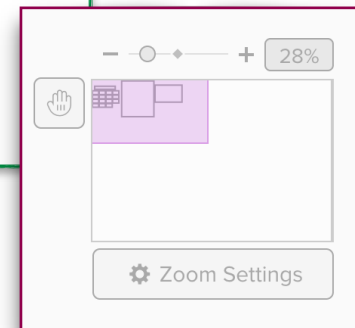


*What is covered by a DMP?*

*Sort DMP parts to life cycle phases*

Mural / Plenum

Mural



# DMP PARTS



Administrative information (Project name, data owner, further collaborators, contact, funding agency, etc.)

Finances

Resources

Project description

Roles and responsibilities

Ethics and legal compliance



Dataset description

Data collecting



Data types, formats, scale

Documentation

Backups

Metadata and standards

Licenses



Data sharing



Archiving and data storage





*What is the (general) structure of a DMP?*

*Sort pieces to DMP Template*

Mural / Plenum

Mural 2



# DMP SKETCHING



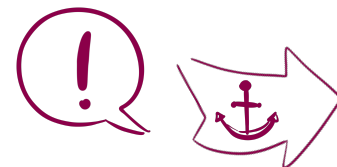
# WRITING A DMP SKETCH



*How does my work fit into a DMP?*

*With your project in mind: fill in the DMP  
template with bullet points*

Cloud + Word / Individual



# WRITING A DMP SKETCH



*How does my work fit into a DMP?*

*With your project in mind: fill in the DMP template with bullet points*

Pen & Paper / Word

*You have the opportunity to join the annual cruise of the R.V. Polarstern (cruise ANT X/6, part of the international JGOFS programme) from Punta Arenas to Capetown on 29 September - 29 November 1992. Your idea is to do ice and animal observations during this cruise (i.e., how thick is the ice, how many animals do you see and which).*



# WHAT SHOULD I WRITE?



*How do I phrase DMP texts?*

*Identify good DMP statements in 7 examples*

Discuss

Examples are either based on material of FDMentor (3.1) or slides by Julie Goldman (Love Data Week 21, DMPtool)



# DATA COLLECTION

---



A

The nature of the data or other materials produced under this sponsored project will include data characteristics such as observational, derived, referenced and/or other. The data types referenced could include images, reports, etc.

B

We will produce 2- and 3-dimensional microscopy image of mouse cells in TIFF format. We expect to produce 50 TB of image data (each image can range in low to high GB size) over the course of the project.





# DATA COLLECTION



A

The nature of the data or other materials produced under this sponsored project will include data characteristics such as observational, derived, referenced and/or other. The data types referenced could include images, reports, etc.

B

We will produce 2- and 3-dimensional microscopy image of mouse cells in TIFF format. We expect to produce 50 TB of image data (each image can range in low to high GB size) over the course of the project.



# DATA COLLECTION

---



A

The numeric data will be saved as TAB-delimited TXT files (ASCII) as well as ZIP-archives. The chosen formats and software are suitable for longterm archiving. The area has not yet been researched, no data for reuse exists.

B

The numeric data will be saved as an ASCII file with TAB delimitation in text file format as well as in archives of ZIP-format. The chosen formats and software are suitable for longterm archiving. The area has not yet been researched, no data for reuse exists.



# DATA COLLECTION



A

The numeric data will be saved as TAB-delimited TXT files (ASCII) as well as ZIP-archives. The chosen formats and software are suitable for longterm archiving. The area has not yet been researched, no data for reuse exists.

B

The numeric data will be saved as an ASCII file with TAB delimitation in text file format as well as in archives of ZIP-format. The chosen formats and software are suitable for longterm archiving. The area has not yet been researched, no data for reuse exists.



# DATA DOCUMENTATION



A

Metadata will be generated to describe the data generated in README format and will be stored alongside the data. Metadata standards (ISO 19115) will be applied during the creation of the metadata. File names will follow a standard as well.

B

Metadata will describe the data in README.txt format and will be stored alongside the data. Metadata standards (ISO 19115, Core) will be used and extended by coordinates in decimal degrees. File names will follow the standard ID\_YYYYMMDD\_R#.dat.



# DATA DOCUMENTATION



A

Metadata will be generated to describe the data generated in README format and will be stored alongside the data. Metadata standards (ISO 19115) will be applied during the creation of the metadata. File names will follow a standard as well.

B

Metadata will describe the data in README.txt format and will be stored alongside the data. Metadata standards (ISO 19115, Core) will be used and extended by coordinates in decimal degrees. File names will follow the standard ID\_YYYYMMDD\_R#.dat.



# DATA STORAGE & BACKUP



A

The data will be stored locally on the ship as well as directly backed up in the repository. It will be under the account of the person that generated and analyzed the data. These are standard procedures in our working group.

B

The code and data will be stored and backed up redundantly on the storage system managed by Central IT. This system includes daily backups on two physically separate sites. Data security is evaluated and assured according to data policy ABC.



# DATA STORAGE & BACKUP



A

The data will be stored locally on the ship as well as directly backed up in the repository. It will be under the account of the person that generated and analyzed the data. These are standard procedures in our working group.

B

The code and data will be stored and backed up redundantly on the storage system managed by Central IT. This system includes daily backups on two physically separate sites. Data security is evaluated and assured according to data policy ABC.



# DATA AVAILABILITY

---



A

The data will be made available immediately after collection via Pangaea and receive a Digital Object Identifier (DOI).

There will be no data usage restrictions. A data sharing agreement will not be required.

B

Data can be accessed by the general public through publications in conferences, journals, etc.





# DATA AVAILABILITY



A

The data will be made available immediately after collection via Pangaea and receive a Digital Object Identifier (DOI).

There will be no data usage restrictions. A data sharing agreement will not be required.

B

Data can be accessed by the general public through publications in conferences, journals, etc.



# DATA CONSERVATION

---



A

Data (see selection criteria above) will be long-term preserved for at least ten years in the discipline-specific repository Pangaea.

B

The data will be long-term preserved in the discipline-specific repository Pangaea. The fee for preservation will be 300,00 EUR per data upload. We will deposit there all data necessary to validate our findings.



# DATA CONSERVATION



A

Data (see selection criteria above) will be long-term preserved for at least ten years in the discipline-specific repository Pangaea.

B

The data will be long-term preserved in the discipline-specific repository Pangaea. The fee for preservation will be 300,00 EUR per data upload. We will deposit there all data necessary to validate our findings.



# DATA RESPONSIBILITIES

---



A

The data curator in close cooperation with the principal investigators will be responsible for implementing the DMP and data management. Data ownership will be contractually clarified.

B

Joe and Mary have completed training for the REDCap data capture and management system. The Open Science Framework (OSF) will be used in order to control permission level access.



# DATA RESPONSIBILITIES



A

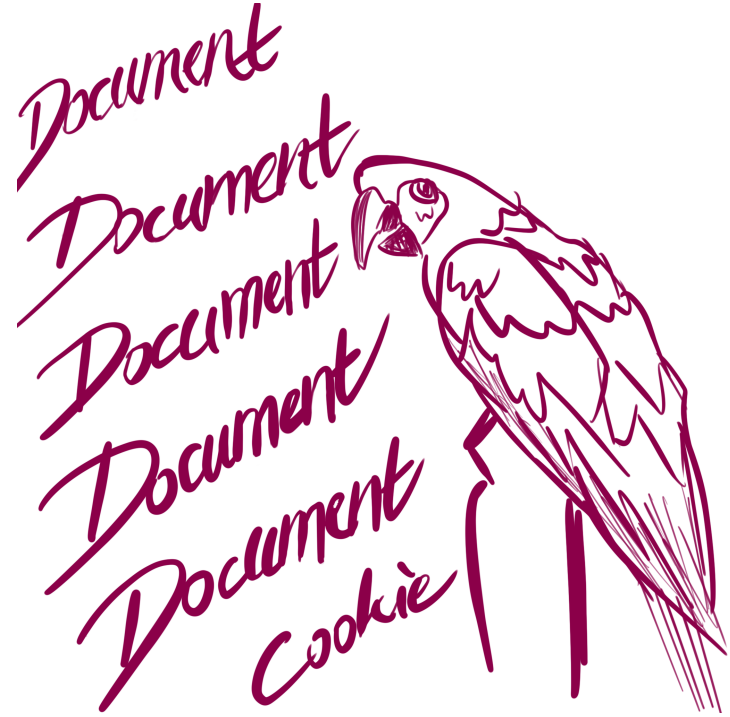
The data curator in close cooperation with the principal investigators will be responsible for implementing the DMP and data management. Data ownership will be contractually clarified.

B

Joe and Mary have completed training for the REDCap data capture and management system. The Open Science Framework (OSF) will be used in order to control permission level access.



# CLOSING



# TREASURE HUNT

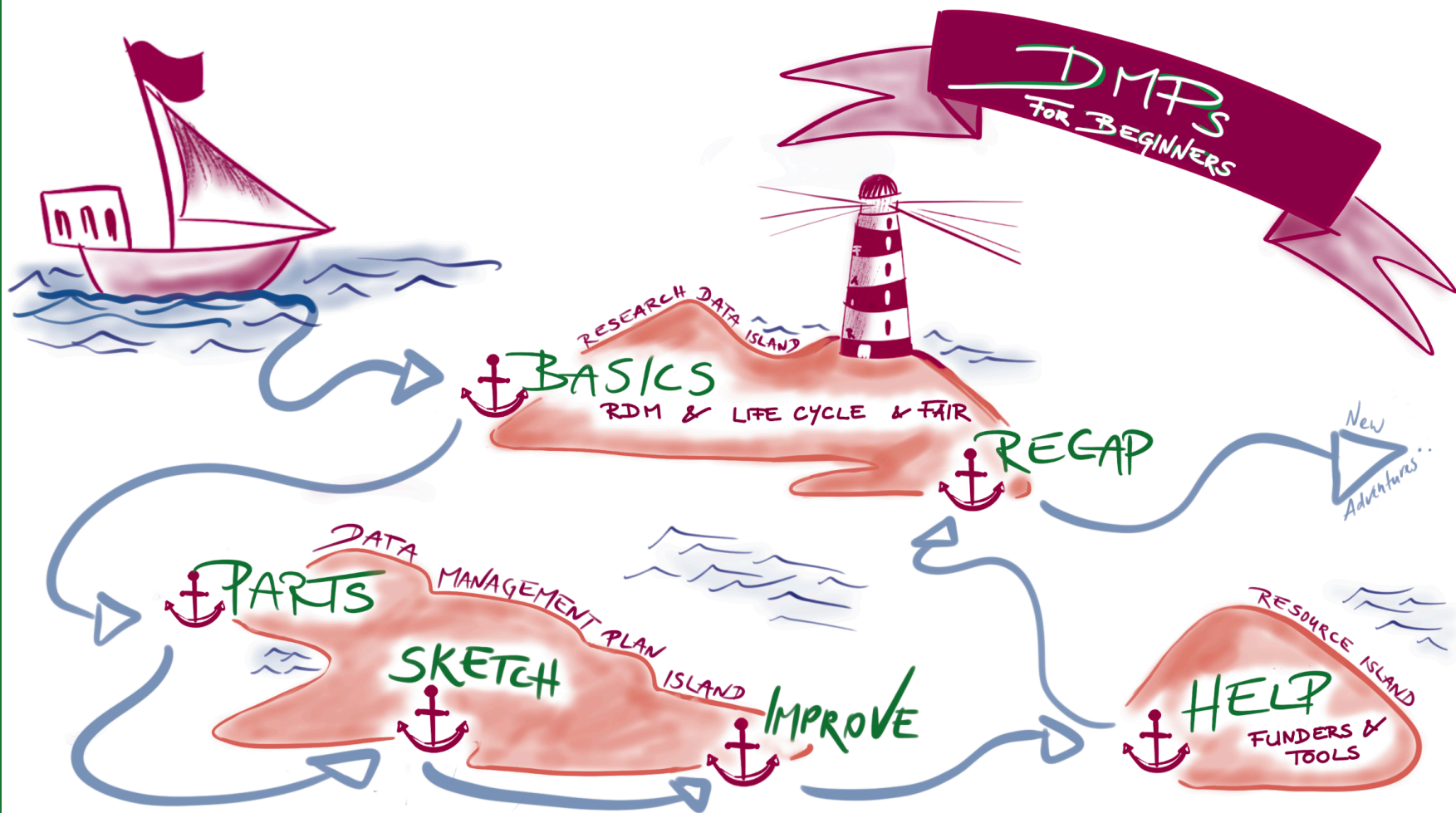
---



*Write down 13 things you learned today*

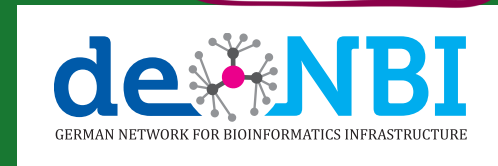
Pen and paper







Jeanne Wilbrandt  
Daniel Wibberg  
Amandine Nunes-Jorges  
Steve Hoffmann  
Maja Rey  
Sebastian Beier  
Matthias Lange  
Simon Parker  
Ulrike Wittig  
Martin Eisenacher  
Helena Schnitzer  
Christian Quast



# THANK YOU FOR PARTICIPATING!

