eResearch Lab: GRO.plan

- Survey: What is your language preference for today?
- Umfrage: Welche Sprache bevorzugen Sie für heute?

A: I would prefer the talk to be in English.

B: Ich würde den Vortrag lieber auf deutsch hören.

C: I don't care. / Ist mir egal.

Please type the respective letter into the chat.

Bitte geben Sie den entsprechenden Buchstaben in den Chat ein.

Data Management Planning and working with GRO.plan

10.02.2026

Timo Henne

henne@sub.uni-goettingen.de

eResearch Alliance

Survey: Who are you? (1)

- A: I am a bachelor student
- B: I am a master student
- C: I am a doctoral/PhD student or scientific assistant
- D: I am a PostDoc
- E: None of the above

Survey: Who are you? (2)

- I am from the …
- A: natural/structural sciences
- B: arts and humanities
- C: social sciences / jurisprudence
- D: life sciences
- E: None of the above

Survey: Who are you? (3)

- I am studying / doing a PhD / working

- A: in Göttingen
- B: not in Göttingen, but in Lower Saxony
- C: somewhere else

Survey: What do you know ?

- My experience with Data Management Plans (DMPs) is best described as…
- A: I have already worked on several DMPs for various projects.
- B: I was involved in writing DMPs occasionally.
- C: I have never worked on a DMP, but I have to provide one soon.
- D: I have no experience with DMPs, but am interested in the topic.
- E: None of the above

Agenda

- Introduction to Data Management Planning
- RDMO
- Introduction to GRO.plan
- GRO.plan Demo
- Discussion

Comments, questions and suggestions for using GRO.plan can be entered here:

https://pad.gwdg.de/eResearchLab_GROplan?both

INTRODUCTION TO DATA MANAGEMENT PLANNING

What is Research Data Management?

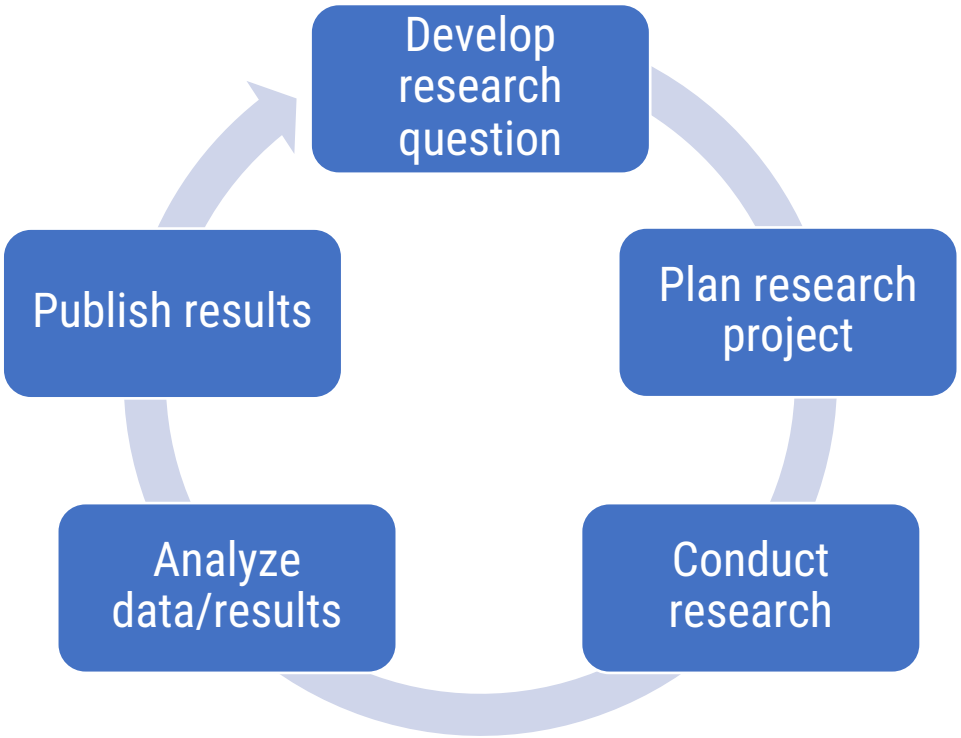
eResearch Alliance

Types of research data

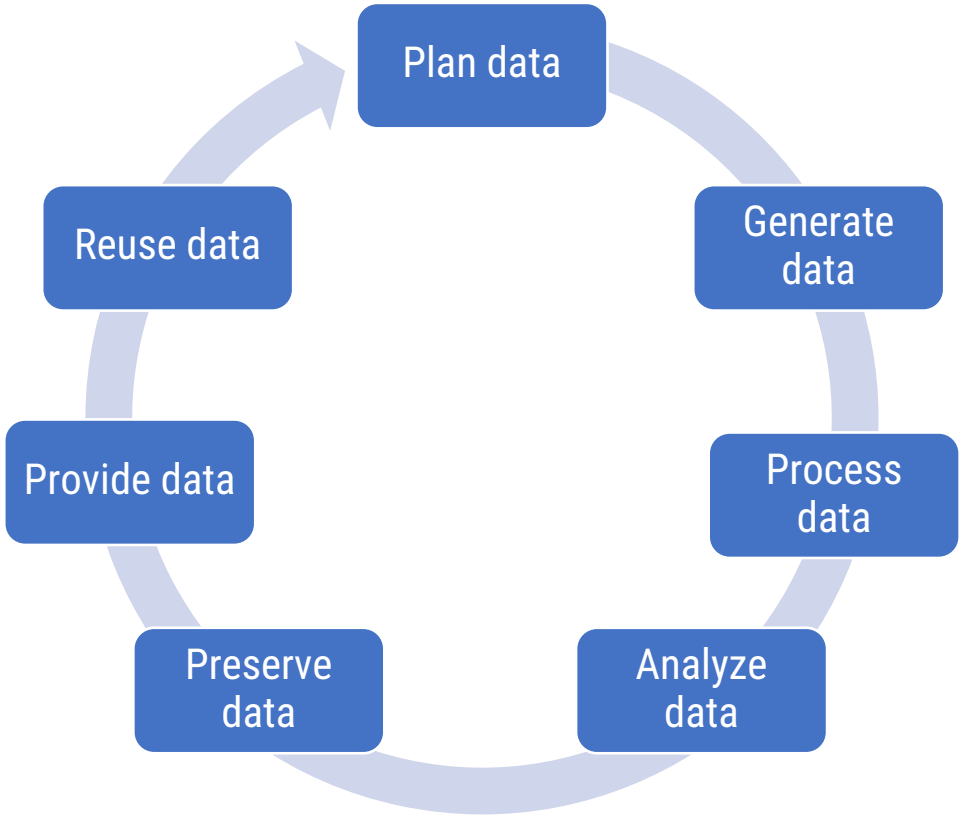
Type	Characteristics	Example
Observations	Data is collected in real time Mostly irreplaceable	Sensor data Survey data
Experiments	Mostly created in the laboratory Reproducible but expensive	Gene sequences Chromatogram
Simulations	Generated from test models Model and metadata more important than output	Climate models Economic models
Derived data	Derived or compiled from other data, reproducible	Text Mining 3D models
References	Collection of smaller data sets Mostly published	Gene sequence database Primary text sources
Digital copies	Digital version of an analog object, reproducible as long as the original exists	Manuscripts

Source: Dominique Ritze, Kai Eckert and Magnus Pfeffer. Research data.
In: PatrickDanowski, (Open) Linked Data in Libraries; 122-138, DeGruyter Saur, Berlin, 2013

Research cycles



Research life cycle



Research data life cycle

eResearch Alliance

Why Research Data Management?

- Improve your research
  - Prevent data loss & unnecessary work
  - Better data quality
  - Obviate retractions
- Increased visibility
  - Citations through data publications
  - Find future collaborations or employers
- Efficient use of resources
  - Data sharing
  - Productive collaboration
- External requirements
  - Good scientific practice
  - Funder or journal requirements
- Contribution to science & teaching
  - Feedback loops between empirical and modeling approaches
  - Test data for new software/algorithms or for teaching purposes
  - Initiate or support research questions in completely different fields

FAIR data Principles

- Set of guiding principles for research data
- Goal: make data **F**indable, **A**ccessible, **I**nteroperable and **R**eusable
- FAIR data principles
  - address data producers and data publishers to promote maximum use of research data
  - are aimed at both humans and machines
- Published in 2016:
  - Wilkinson, M., Dumontier, M., Aalbersberg, I. et al., The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

FAIR data Principles

Findable:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.

F2. data are described with rich metadata.

F3. (meta)data are registered or indexed in a searchable resource.

F4. metadata specify the data identifier.

Accessible:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol.

- A1.1. the protocol is open, free, and universally implementable.
- A1.2. the protocol allows for an authentication and authorization procedure, where necessary.

A2. metadata are accessible, even when the data are no longer available.

Interoperable:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles.

I3. (meta)data include qualified references to other (meta)data.

Re-usable:

R1. meta(data) have a plurality of accurate and relevant attributes.

- R1.1. (meta)data are released with a clear and accessible data usage license.
- R1.2. (meta)data are associated with their provenance.
- R1.3. (meta)data meet domain-relevant community standards.

eResearch Alliance

Source: https://www.force11.org/group/fairgroup/fairprinciples

Why plan your Research Data Management?

- Become aware of possible problems before they arise
  - Like planning your thesis or research project
  - Identify roles, responsibilities, resources and solutions before data are generated
- Prevent double work and time pressure
  - Keep data management problems to a minimum during hot research phases
  - Rely on knowing that your (intermediate) research results are well-managed
- Requirement by funders or institutions
  - EU: Horizon 2020 Open Research Data Pilot, mandatory in Horizon Europe
  - In many countries DMPs are mandatory for quite some time already
  - Research institutions often have their own research data policies, requiring data management planning in varying degrees

What is a Data Management Plan (DMP)?

- A data management plan (DMP) is a (living) document that outlines how data are to be handled during and after a research project.

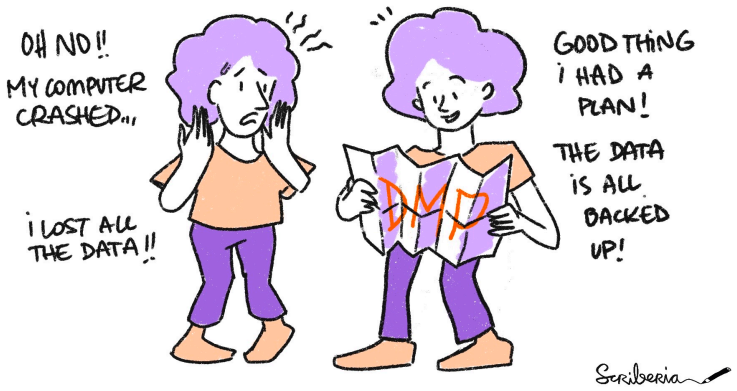


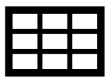
Image source: The Turing Way Community, & Scriberia. (2023). Illustrations from The Turing Way: Shared under CC-BY 4.0 for reuse. Zenodo. https://doi.org/10.5281/zenodo.8169292

- It describes:
  - What data will be created or reused
  - Who will own and have access to the data
  - What data management practices, tools and infrastructures will be used
  - Who will be responsible for each of these activities
  - What policies and regulations will apply to the data

Research data can mean a lot


Survey


Measurement

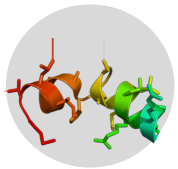

Observation

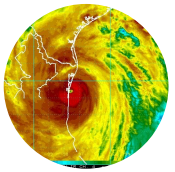

Code


Image


Document


Visualisation





Research data is any information that has been collected, observed, generated or created to validate original research findings.

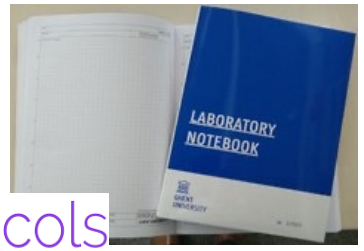
eResearch Alliance

Research output

Publications

Data

Other



Protocols

```
int[] arrayToSort = new int[] { 15, 4, 1, 29, 107, 56, 11, 22 };

int temporaryInt;

for(int passCount = 1; passCount < arrayToSort.Length;passCount++)
{
    for(int i = 0; i < arrayToSort.Length - 1;i++)
    {
        if(arrayToSort[i] > arrayToSort[i + 1])
        {
            temporaryInt = arrayToSort[i];
            arrayToSort[i] = arrayToSort[i + 1];
            arrayToSort[i + 1] = temporaryInt;
        }
    }
}
```

DMP

eResearch Alliance

DMPs in Horizon Europe

- „Under Horizon Europe (Work programmes 2021 and onwards), grantees of all ERC projects that generate research data **have to submit a DMP (at the latest six months after the start of the project)**, deposit such data in a 'trusted' repository and provide access to them, under the principle **"as open as possible, as closed as necessary"**. There are also a number of requirements concerning the **bibliographic and administrative metadata** of deposited data, which also have to be made openly accessible to enhance findability and facilitate reuse.

- Under Horizon Europe it is **not possible to opt out completely** from these obligations, but exceptions to the requirement to provide open access to data and metadata are possible. Grantees funded under Horizon Europe are advised to pay careful attention to the requirements detailed in the Horizon Europe Model Grant Agreement (MGA)6 and the explanations provided in the Horizon Europe Annotated Grant Agreement (AGA)7."

eResearch Alliance

Source:
https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf

Data Management at the DFG

"For this reason, the DFG expects research projects to include a description of how research data is handled. The description should be based on the checklist for handling research data

[...]

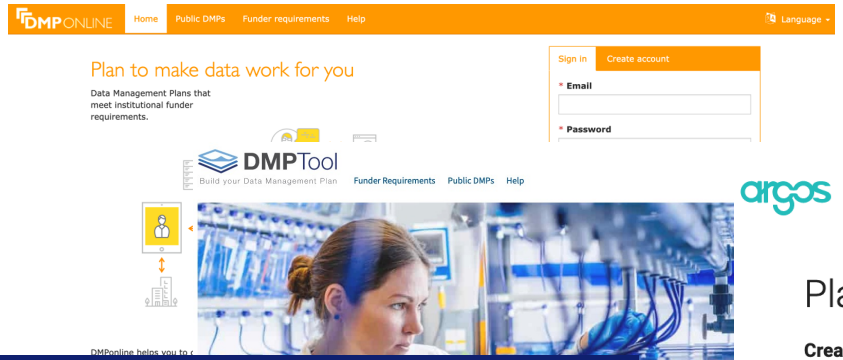
Making research data available, developing methods and standards and building data infrastructures are important contributions to the re-use of research findings as well as integrated part of the good research practice and should be listed as part of a researcher's preliminary work or academic profile."

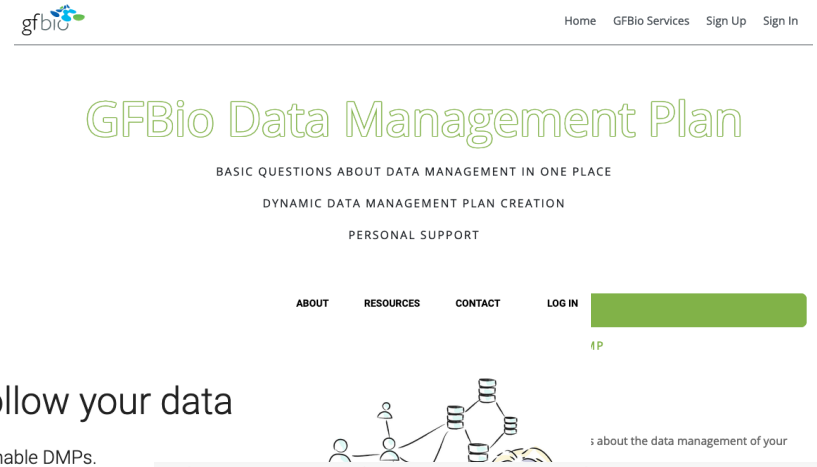
Source: https://www.dfg.de/en/research-funding/funding-initiative/research-data

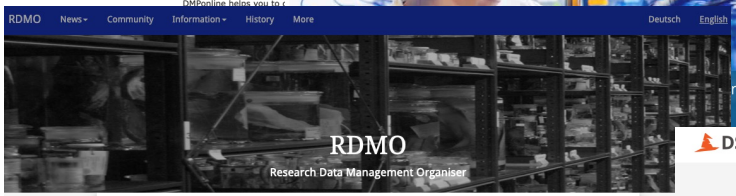
DMP as a requirement?

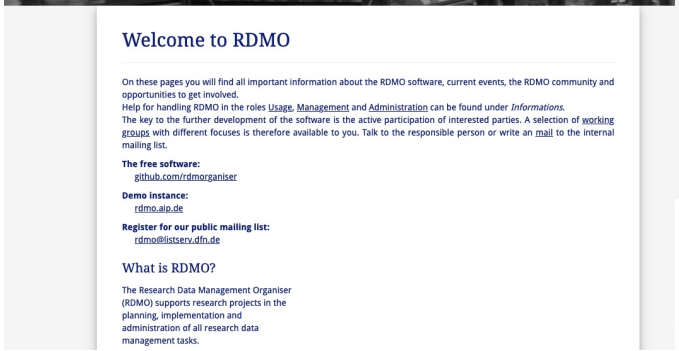
Funder/Program	DMP in proposal	Initial DMP	Final DMP	Specific DMP template
Horizon Europe	NO But include a RDM paragraph (exception: puplic emergency & if required in work programme)	YES (by month 6)	YES	YES (recommended)
European Research Council (ERC)	NO	YES (by month 6)	NO (but keep initial DMP up to date)	YES (recommended)
German Research Foundation (DFG)	NO But mandatory section on RDM in resesarch proposal Additional guidance/ recommendations for some disciplines available	NO (for some disciplines recommended)	NO	NO

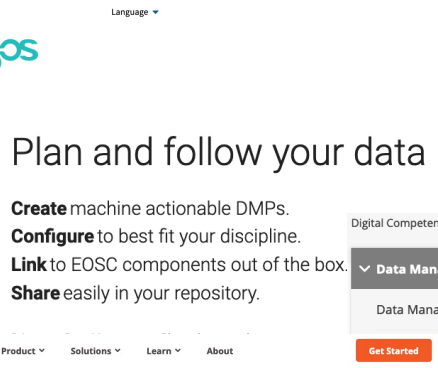
DMP tools: Examples

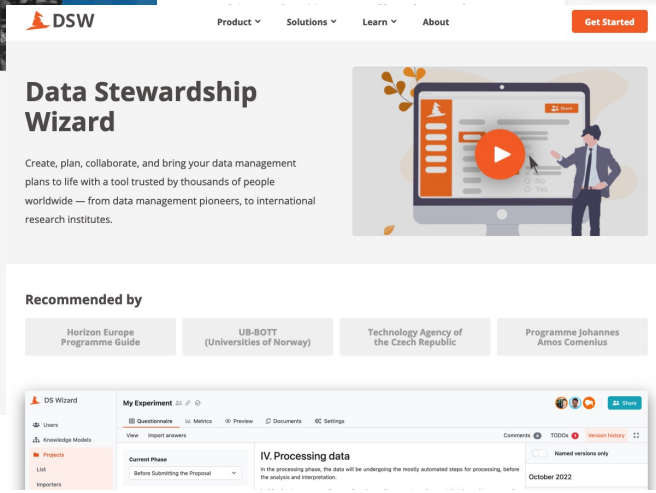


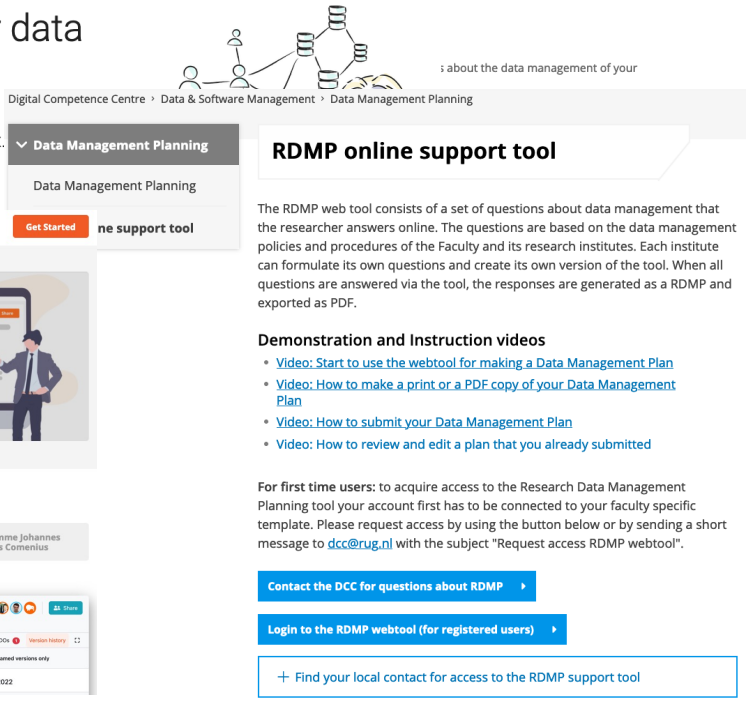












DMP tools: General structure

- Checklist or set of questions to be answered by users, based on
  - Funders‘ requirements
  - Institutional requirements
  - Community / discipline requirements
  - General research data management issues

- Varying degrees of customized support for answers
  - Preselected options
  - Conditional questions or sections
  - Suggestions for relevant further information
  - Connected or integrated support from institution or community

- Different options for storing and output of DMPs
  - Versioning or collaborative editing
  - Export to different formats

DMP sections

Data summary & description

Purpose, origin, type, sensitiveness, format & volume

Ethical & legal issues

Intellectual property, GDPR, etc.

Data storage & security

Security measures, back-up, data transfers & access control

Documentation & metadata

Information to keep data findable, understandable, reusable

Long-term preservation

What to preserve & where, for how long, access restrictions

Data sharing and reuse

Data repositories, access conditions, licenses or agreements

Roles & responsibilities

Costs & resources

eResearch Alliance

Source: Ghent University Data Stewards

DMPs assist in realising the FAIR Principles

Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

FINDABLE

Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.

ACCESSIBLE

Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.

INTEROPERABLE

Data and collections have a clear usage licenses and provide accurate information on provenance.

REUSABLE

eResearch Alliance

Source: LIBER, Implementing FAIR Data Principles: The Role of Libraries

How does a DMP address the FAIR principles? – I

Findable as the DMP provides information on:

- How data is stored and preserved during and after the project
- Persistent identifiers, e.g. DOIs, description of the data, metadata standards

Accessible because it includes information on:

- How the data can be accessed, what is required to access the data (authentication, authorization) and by what communication protocols (e.g. HTTP, HTTPS).

How does a DMP address the FAIR principles? – II

Interoperable through:

- Indicating which metadata standards, vocabularies, methodologies and tools were used.
- assisting different systems and services to exchange metadata and data (if machine-actionable)

Reusable because:

- Data can be described in some detail and accuracy, making it easier for others to understand, select and reuse.
- It indicates how the data is prepared for sharing and reuse, what licence and rules apply, how the data can be reused and for whom it might be valuable.

DMPs: General Recommendation

When certain information is not yet known (e.g. it is still too soon in the project):

➢ **indicate the planned action and timeline to resolve this.**

- e.g. Ethical approval for recruiting study participants in WP5 has been applied for, and reference to the ethical decision will be added here before this task is started.

- e.g. The long term preservation of the data will be managed by combining deposition in established generic repositories as well as making use of institutional storage platforms. Specific information regarding preservation will be added here at Milestone 3Y, three years after the start of the project.

- e.g. The exact volume of the datasets is not known yet, but we anticipate no issues with regard to storage or transfer because of the volume. In total 50 participants will be interviewed for 1 hour: video recording (1 GB / Hour) + transcript in Word doc. Online survey results for 250 participants will be collected: total less than 1 GB.

Data Description: approach 1

- **Straightforward project structure, limited number of datatypes**

⇨ **Write a description**

EXAMPLE

This project will produce qualitative observational data from interviews with pet owners conducted at the animal clinic in the faculty of Veterinary Science in Merelbeke, Belgium, between January and June 2020. These interviews will be done by myself.

Raw data will comprise digital audio recordings of interviews and hand-written notes. Audio files will be transcribed into digital text documents and notes will be digitized to prepare them for analysis.

The audio transcription and notes digitalization will be done in MS Word O365 (.docx extension). Around 40 one-hour long interviews will be done, with file size of 250 MB per interview, giving an estimate of 10 GB for all the interviews.

Data Description: approach 1

- **Straightforward project structure, limited number of datatypes**

⇨ **Write a description; 2nd example or 2nd part of the research project**

EXAMPLE 2

This project will measure biochemical characteristics of bovine urine samples collected at the animal clinic in the faculty of Veterinary Science in Merelbeke, Belgium, between January and June 2020.

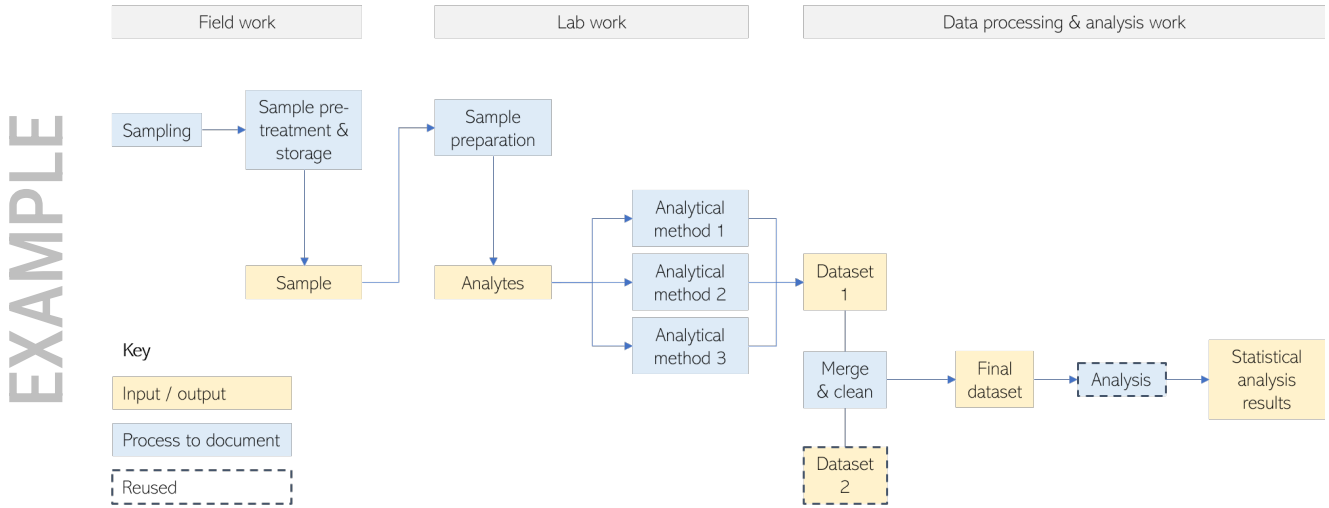
Raw data will comprise acidity and protein concentration measurement, collected in the bespoke database with case reports, and will be exported in spreadsheet format (csv) to enable processing and analysis by scripts generated in R using Rstudio.

40 urine samples will be analysed, constituting per sample 50kb in the record database. Combined, the spreadsheet, processed data tables and output datasets (all csv format) will be less than 2Gb.

Data Description: approach 2

- **Complex project structure, small to medium number of datatypes**

⇨ **Design a data flow**, add details in text or table



Data Description: approach 3

- **Complex project structure, medium to large number of datatypes**

⇨ **Create a data table**

EXAMPLE: NEXT SLIDE

Data Description: data table approach

EXAMPLE

Study phase	Data type	Origin	Format	Expected Volume
WP1	Observational qualitative	Questionnaire for farmers in Flanders	Paper and pencil + digitized transcript (.docx)	<1 GB, paper surveys of 300 participants
			Smartphone based questionnaire. Exported tables as .csv	<1 GB
WP1	Observational quantitative	Activity tracker for data subjects	Physical activity and sedentary time in .GT3X files	250 KB per subject x 300 subjects = 75 MB
WP1	Observational qualitative	Interview with a selection of farmers	.MP3 + transcript in .docx	About 5 GB
WP2	Experimental	Micro-randomized trial to determine the effect of isolated determinants on **physical activity**	Digital textual and numerical data combined in spreadsheets and obtained via tracker based experiment. CSV format and GT3X for raw activity tracker data	<5 MB for the CSV files. Raw accelerometer data (.GT3X files) are 250kB per subject x 40 subjects per experimental study = 10MB
WP3	Experimental	Micro-randomized trial to determine the effect of isolated determinants on **sedentary time**	Digital textual and numerical data combined in spreadsheets and obtained via a tracker based experiment. CSV format and GT3X for raw activity tracker data	<5 MB for the CSV files. Raw accelerometer data (.GT3X files) are 250kB per subject x 40 subjects per experimental study = 10MB

Data table (1)

- Bonus: Detect RDM needs
- There is not a "one size fits all" approach
  - How much granularity is needed and how do I break down the data into different rows?
    - It depends on your project and a good approach is to categorize the content in a way that proves useful for the management of your data
    - e.g. divide your data according to:
      - *content type (quantitative vs. qualitative)*
      - *data collection method (observational, experimental, simulation)*
      - *by purpose*
      - *by format*
      - *differentiate between own data and third-party data*

Data table (2)

- Bonus: Detect RDM needs
- There is not a "one size fits all" approach
  - How many columns or characteristics should I provide?
    - DMP templates will explicitly require some characteristics, e.g. description, format and expected size
    - Add other relevant characteristics that are specific to your data and that have implications for data management, e.g.:
      - *temporal and geographical scope of the data*
      - *data is sensitive or not*
      - *access rights*
    - If you are reusing data, it might be interesting to add the source and license of the dataset(s) in question

Data volume

- If relevant (i.e. high data volume), distinguish between the different data life cycle phases
  - how large (approximately) is the volume of generated data (per year, or over the entire project period)
  - how much data needs to be stored to ensure
    - accountability
    - verifiability
    - reusability
- Otherwise, briefly describe the expected general volume
  - e.g. “The exact volume of the datasets is not known yet, but we anticipate no issues with regard to storage or transfer because of the volume. In total 50 participants will be interviewed for 1 hour: video recording (1 GB / Hour) + transcript in Word doc. Online survey results for 250 participants will be collected: total less than 1 GB.”

Data format

- Important to reflect on file format choices at the start of the project.
  - File formats during project
  - File formats for preservation
  - Vulnerability to obsolescence: open vs. closed formats
    - e.g. Word (docx) vs Text (txt)
    - e.g. Excel (xlsx) vs CSV (csv)
- Keep in mind possible risks of file format migration
- Check if you need specific software to access the data

https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/recommended-formats/

Data Description – Tips

- Provide answers relating to research data, not publications!
- Remember that ‘research data’ come in many different forms
- Clear distinction between re-used and new data
- Break down and list your data types conveniently: e.g. by technique, by purpose, by research phase
- Provide enough details for outsiders to understand the sort of data involved
  - e.g. distinguish between digital/non-digital, quantitative/qualitative, raw & processed data
  - specify file formats (preferably no proprietary/unusual formats)
  - specify data collection methods (e.g. experimental, observational, simulation… data)
- Make a concise list of data(sets) and fill in at least the attributes requested by the template
  - e.g. not ‘personal data’ but ‘questionnaire for research institute policy personnel’

Public DMPs

- DMPonline, https://dmponline.dcc.ac.uk/public_plans
- ARGOS, https://argos.openaire.eu/explore-plans
- LIBER DMP Catalogue, https://zenodo.org/communities/liber-dmp-cat
- Gent University examples, https://www.ugent.be/en/research/datamanagement/before-research/datamanagementplan.htm

Data Management Planning: Further infos

Göttingen eResearch Alliance:

https://www.eresearch.uni-goettingen.de/knowledge-base/howto/how-to-data-management-planning/

Forschungsdaten.org (German):

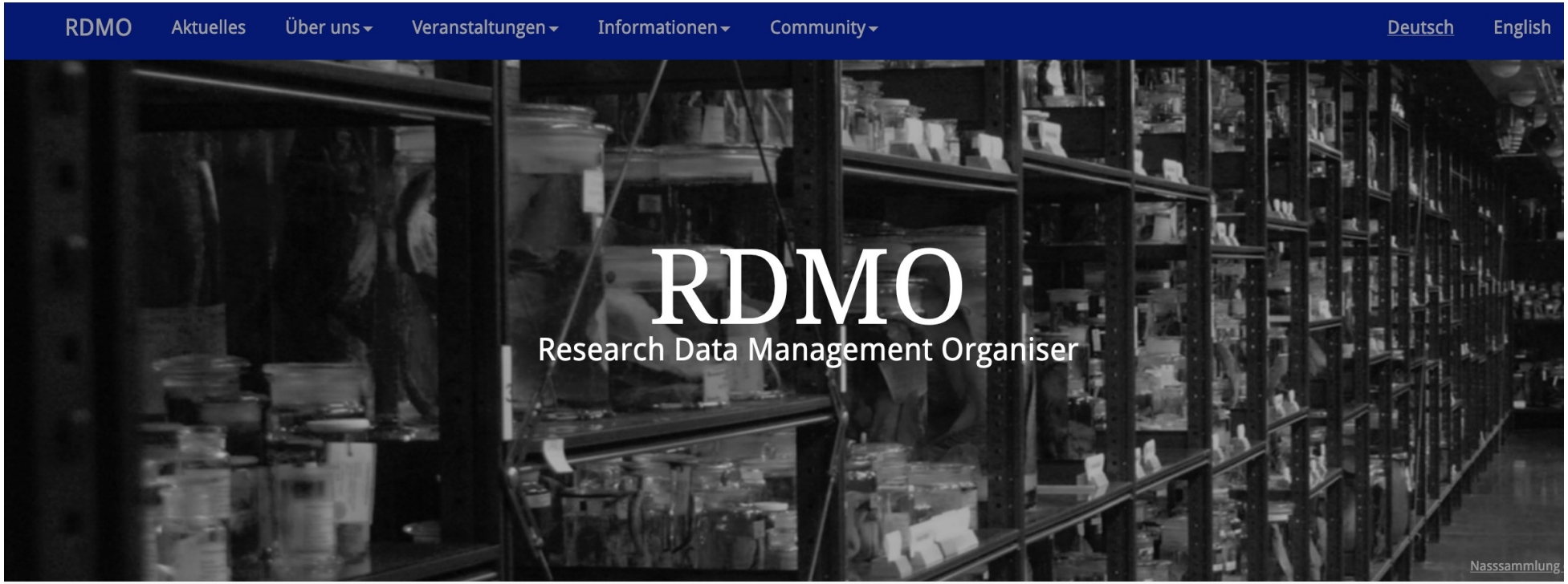
https://www.forschungsdaten.org/index.php/FAQs#Was_ist_ein_.28Forschungs-.29Datenmanagementplan.3F

Forschungsdaten.info (German):

https://www.forschungsdaten.info/themen/informieren-und-planen/datenmanagementplan/

RDMO: RESEARCH DATA MANAGEMENT ORGANIZER

Data Management Planning: RDMO

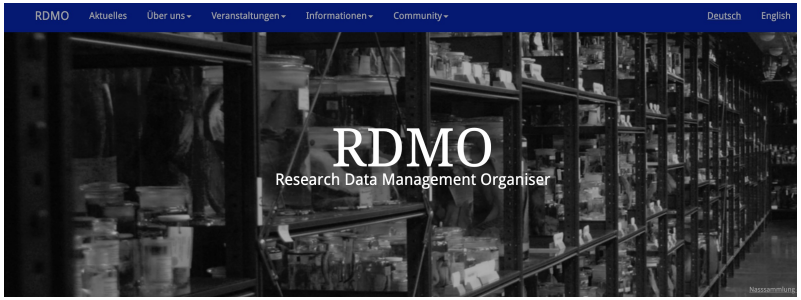


Project: https://rdmorganiser.github.io/
Demo instance: https://rdmo.aip.de/

eResearch Alliance

Data Management Planning: RDMO

- Research Data Management Organiser (RDMO) supports the systematic planning, organisation and implementation of research data management throughout the course of a project.



- DFG-funded project 2015-2020
  - Currently maintained and developed
  - through community-based consortium
- RDMO is being used productively at ca. 50 research institutions in Europe (mainly in Germany)
- Features:
  - versioning of DMPs to allow evolving DMPs over project lifecycle and beyond
  - output of DMPs as text documents according to funder requirements
  - built-in multilingual support
  - adaptable templates for various purposes, e.g. funder-required DMPs

INTRODUCTION TO GRO.PLAN

GÖTTINGENRESEARCHONLINE

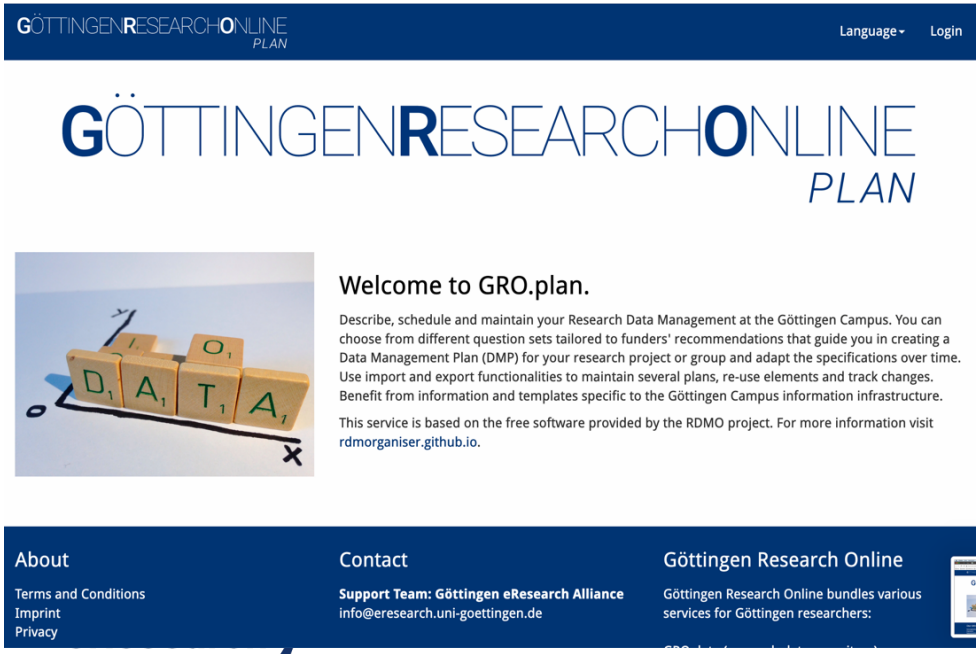
Publications	Data	Instruments	Plan	...
Publication data management	Research data repository	Large equipment portal	Data management planning	further services
Dspace-CRIS	*Dataverse*	*openIRIS*	*RDMO*	

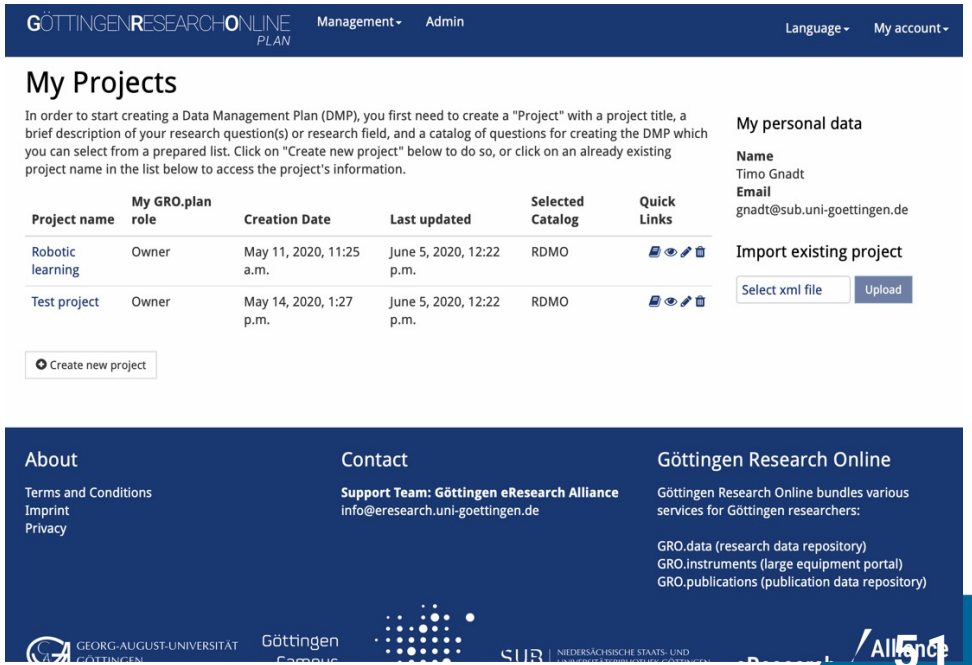
plan.goettingen-research-online.de

eResearch Alliance

GRO.plan – Data management planning tool

- based on RDMO software
- customizations according to Campus requirements
- integration into Portal Göttingen Research Online (GRO)
- allow connection with other GRO services
- changes in code and content are discussed with and fed back to RDMO community





GRO.PLAN DEMO

Thank you for your participation!

The comments, questions and suggestions from the participants of the GRO.plan eResearch Labs can be found here:

https://pad.gwdg.de/eResearchLab_GROplan#

CONTACT:

info@eresearch.uni-goettingen.de

www.eresearch.uni-goettingen.de