eResearch Alliance

eResearch Lab: GRO.plan

Data Management Planning and working with GRO.plan

Timo Henne

henne@sub.uni-goettingen.de

eResearch Alliance

eResearch Alliance

Structure

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

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

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

eResearch Alliance

Introduction to Data Management Planning

eResearch Alliance

Types of research data

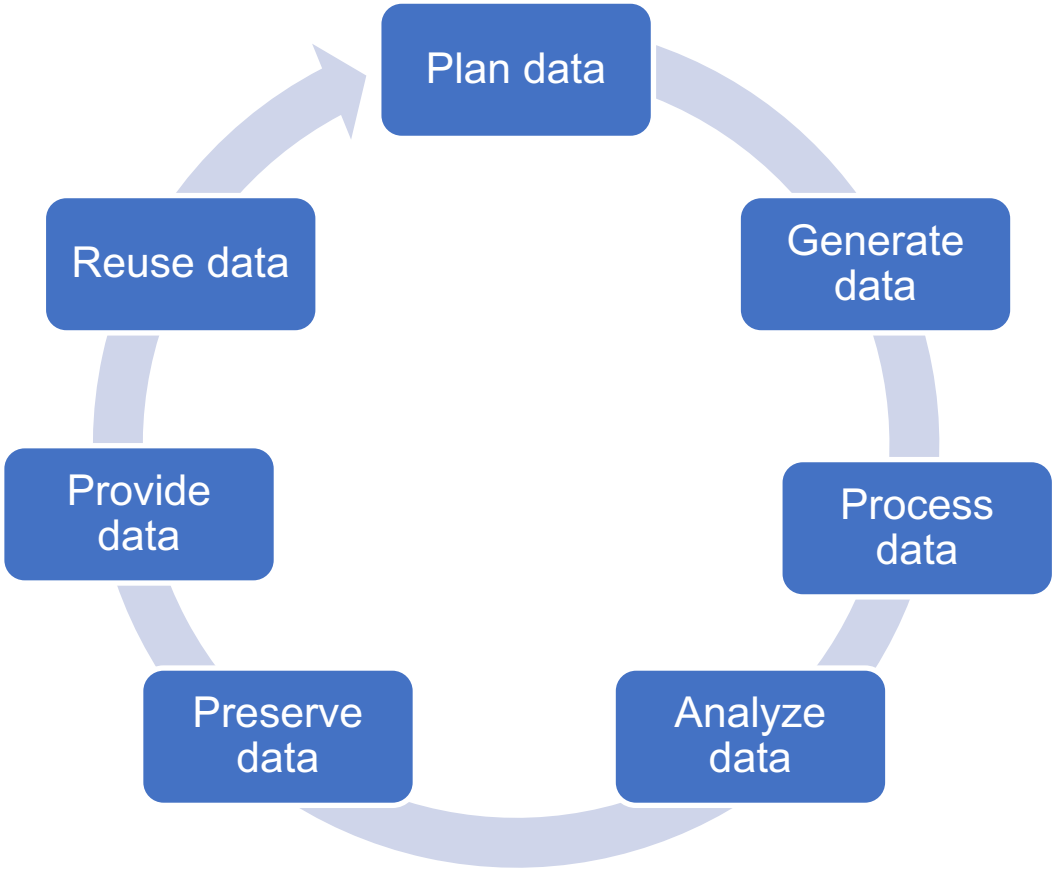
Type	Characteristics	Example
Observations	Data is collected in real time Mostly irreplaceable	Sensor data Survey data
Experiments	Mostly created in the laboratory Reproducible but expensive	Gene sequences Chromatogram
Simulations	Generated from test models Model and metadata more important than output	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 cycle

eResearch Alliance

Research data cycle



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

eResearch Alliance

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**.

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

Research data guideline of the Georg-August-University of Göttingen

- Officially published on August 28, 2014
- One of the first German universities with such a guideline

Göttingen eResearch Alliance

Focus areas

Current page: Homepage 28. August 2014: Res... Search | Deutsch

28. August 2014: Research data policy of the Georg-August University Goettingen (incl. UMG)

Preamble:

The Georg-August-University Goettingen is committed to diligently preserve results of scholarship, to produce novel results through research, and to make results accessible and reusable for academia and the wider society, now and for future generations. The management, protection, preservation and sustainable provision of research data must therefore be carried out in accordance with recognized standards, meet high expectations and fulfil legal and ethical obligations. The University acknowledges that the implementation of this guideline will depend on the settings and requirements of each subject area.

1. The University promotes and supports open access to research data.
2. Research data are those data collected, observed, simulated, derived, or generated during the course of research.
3. Management of research data includes their planning, collection, processing, and preservation. It ensures the access to, and the reuse, reproducibility, and quality assurance of all research data underpinning research results.
4. Research data management is generally the responsibility of the person leading a project and the researcher who is acting in an individual capacity. A particular responsibility is the adherence to good practices of research as well as standards in their subject area.
5. Research projects with research data require a data management plan that includes but is not restricted to the topics of access rights to research data and necessary precautions for handling them.
6. The University provides support and advice for research data management in the preparatory stages of research projects, during their conduct and after their completion, and provides appropriate training.
7. The University implements and maintains essential services for research data infrastructure that ensures adequate storage and technical availability of digital research data. Specific requirements have to be aligned among all stakeholders and may involve additional funding.
8. Storage and archiving of digital research data is carried out within the technological and informational infrastructure of the University or in acknowledged external or internal subject repositories.
9. The University and its researchers adhere in their research data management to given conditions of ethics, data protection, intellectual property, privacy and disclosure. This leaves regulations untouched that relate to an assessment of research data according to the German employee invention act and specific contractual agreements.
10. If exploitation or publication rights of data were transferred to third parties, it should be a precondition that research data remain openly and freely available for research purposes.

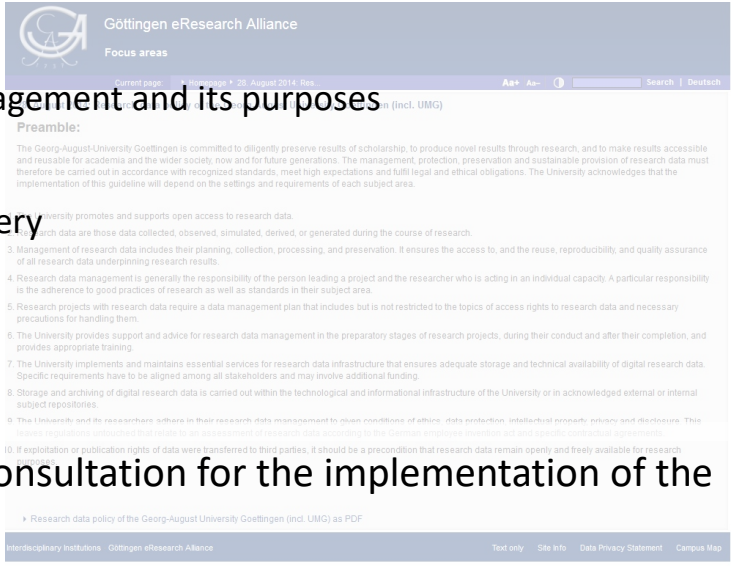
Research data policy of the Georg-August University Goettingen (incl. UMG) as PDF

Interdisciplinary Institutions Göttingen eResearch Alliance Text only Site Info Data Privacy Statement Campus Map

Source: http://www.uni-goettingen.de/en/488918.html

Research data policy of the Georg-August-University of Göttingen

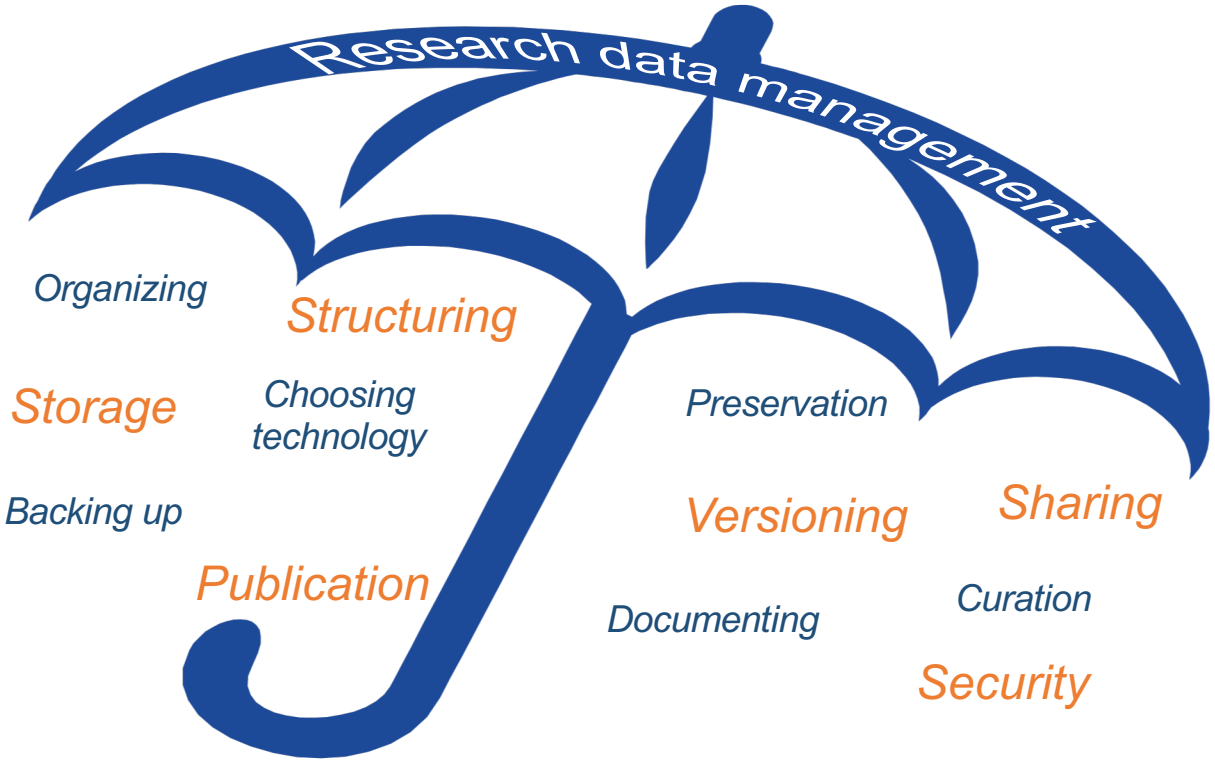
- Officially published on August 28, 2014
- One of the first German universities with such a guideline
- Topics covered:
  - Research data, research data management and its purposes
  - Data management plans
  - Support, training and service delivery
  - Storage solutions
  - Ethical and legal standards
  - Open Access
- **eResearch Alliance:** Support and consultation for the implementation of the policy for the Göttingen Campus



Source: http://www.uni-goettingen.de/en/488918.html

eResearch Alliance

What is Research Data Management?



Why research data management?

- **Improve your research**
- **Adhere to Good Scientific Practice**
- **Improve collaboration with colleagues**
- **Get credit through Data Publication**
- **Enable new research questions**

Why plan your Research Data Management?

eResearch Alliance

1. Become aware of problems before they arise

- Like planning your thesis or research project
- Identify roles, responsibilities, resources and solutions *before* data are generated

2. 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

3. Requirement by funders

- EU: Horizon 2020 Open Research Data Pilot, mandatory in Horizon Europe
- DFG: Guidelines on handling of research data
    - plus various discipline-specific recommendations and requirements
- BMBF: varying requirements depending on discipline, e.g.: https://www.bmbf.de/foerderungen/bekanntmachung.php?B=774
- In US and UK DMPs are mandatory for quite some time already

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."

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

DMPs in Horizon Europe

"A DMP should provide information on:

1. Dataset description (…)
2. Standards and metadata (…)
3. Name and persistent identifier for the datasets (…)
4. Curation and preservation methodology (…)
5. Data sharing methodology (…)"

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

Aspects of a data management plan

Amongst others, a data management plan consists of:

- Administrative information
- Description of project and datasets
- Metadata and standards
- Data exchange, sharing and publication
- Data archival and storage
- Responsibilities
- Costs

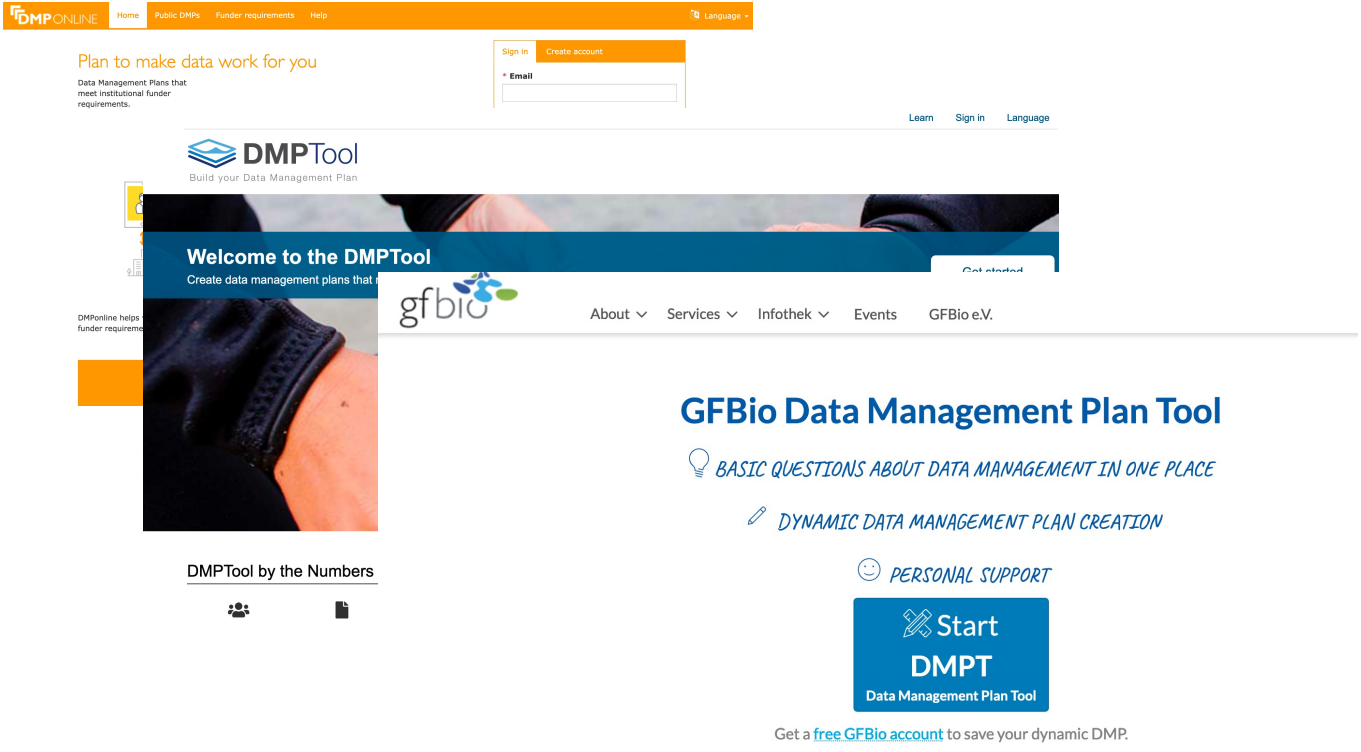
Example DMP tool questions

- *…*
- *How will data be generated or acquired?*
- *How will data be processed?*
- *How will data stored and archived?*
- *Will data be published? Which data, and how?*
- *Who will be responsible for research data management activities?*
- *What costs will arise for these activities?*
- *…*

DMP tools: General structure

- **Checklist or set of questions to be answered by users, based on**
  - funders requirements
  - instutional requirements
  - community / discipline requirements
  - general research data management issues
- **Varying degrees of customized support for answers**
  - preselected options
  - suggestions for relevant further information
- **Different options for storing and output of DMPs**

DMP tools: Examples



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

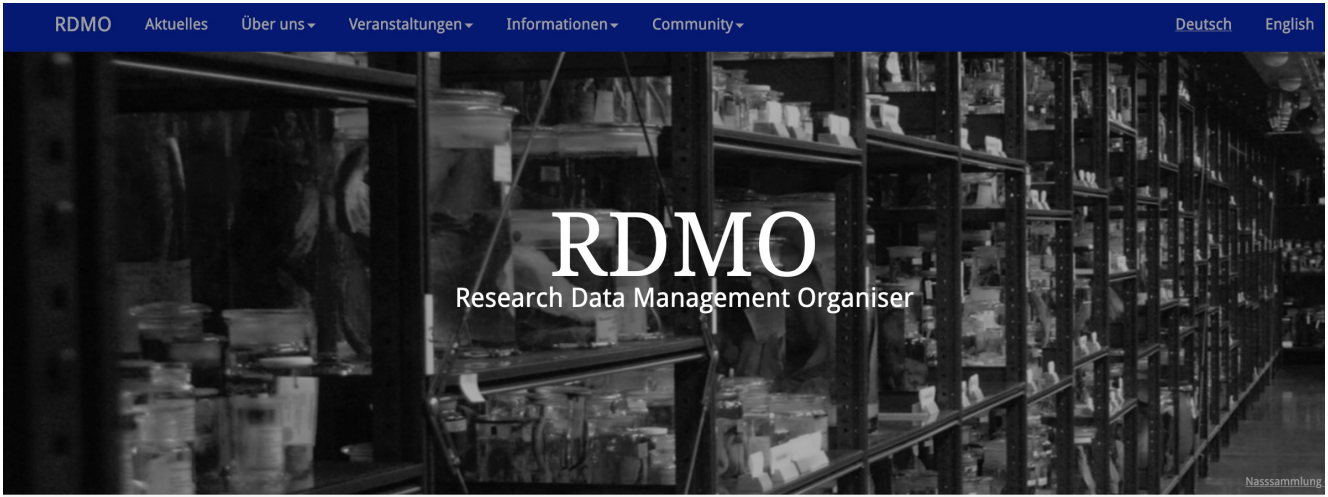
Forschungsdaten.info (German):

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

eResearch Alliance

RDMO: Research Data Management Organizer

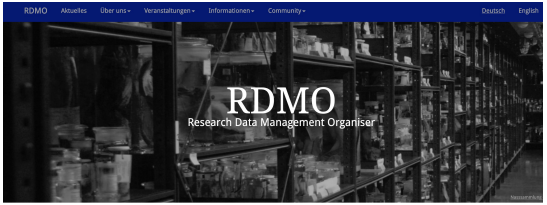
Data Management Planning: RDMO



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

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.



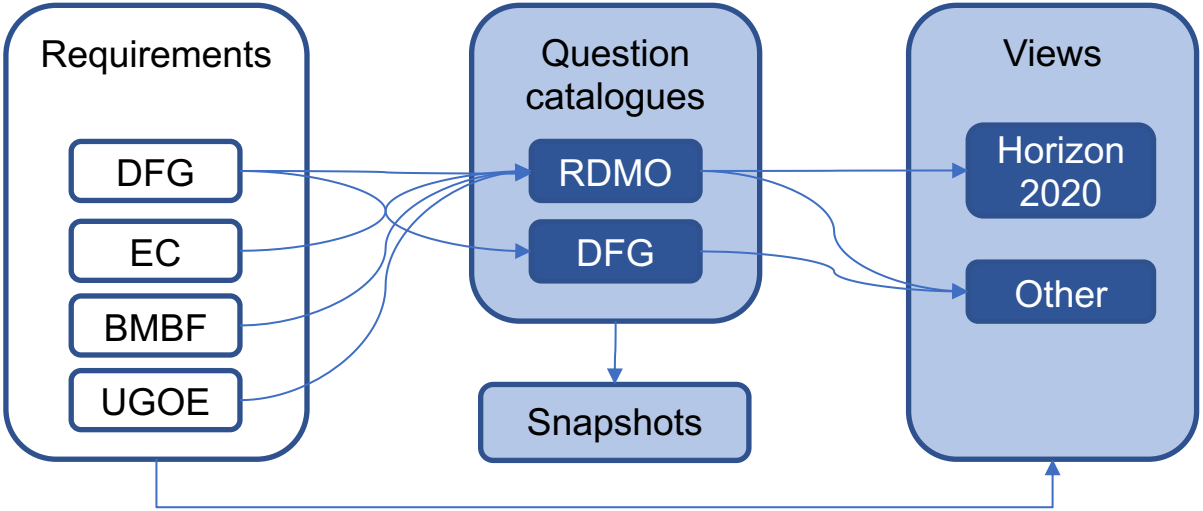
- DFG-funded project 2015-2020
- Currently maintained and developed through community-based consortium

- RDMO is being used productively at 22 research institutions in Germany, and being tested for productive use at over 20 more institutions.

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

Data Management Planning: RDMO



- Question catalogues
  - Used for defining the questions to be answered
  - Requirements from several funders or funder- or discipline specific
- Views
  - Used for defining the selection and display of given answers
  - Selection of answers according to one funder, discipline or call

eResearch Alliance

Introduction to GRO.plan

eResearch Alliance

GÖTTINGENRESEARCHONLINE

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

www.goettingen-research-online.de

eResearch Alliance

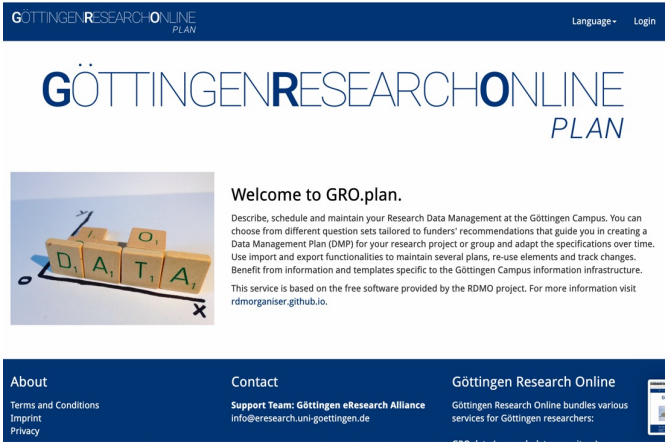
GÖTTINGENRESEARCHONLINE

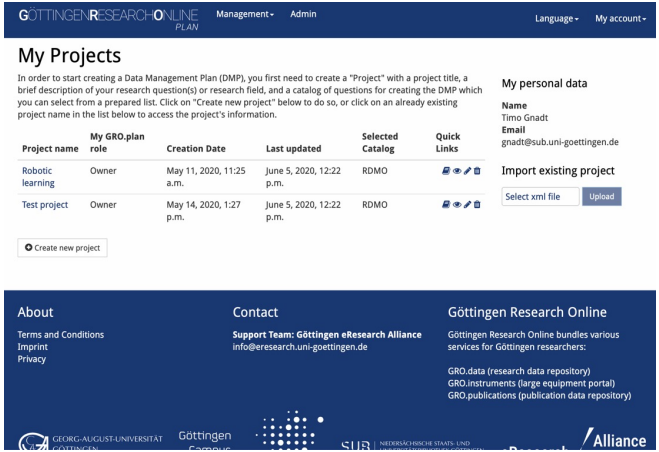
Publications	Data	Instruments	Plan	...
Publication data management	Research data repository	Large equipment portal	Data management planning	further services
Dspace-CRIS	*Dataverse*	*openIRIS*	*RDMO*	

plan.goettingen-research-online.de

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





RDMO: Upcoming features

- Indication of optional questions
- Default answers possible in all languages
- Tooltips & Overlays
- Export & Import to/from datacite metadata schema
- Recording of Medical Subject Headings
- Catalogue for Horizon Europe based on ScienceEurope checklist

eResearch Alliance

Thank you for your participation!

The comments, questions and suggestions of the participants for this eResearch Lab can be found here:

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

CONTACT:

info@eresearch.uni-goettingen.de

www.eresearch.uni-goettingen.de