FAIR Data Stewardship

Erik Schultes, Barbara Magagna, Alessa Gambardella
5 December 2024

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani

Data Management Plan

15:00-15:20
(Alessa)

Alessa Gambardella, PhD
Data Steward
Faculty of Science
Leiden University
&
GO FAIR Foundation Fellow
3PFF Facilitator Assistant

https://orcid.org/0000-0002-4930-2662
a.a.gambardella@science.leidenuniv.nl

What is a Data Management Plan (DMP)?

- A DMP documents the actions that take place at different stages of the research data lifecycle
  - How data is…
    - Collected
    - Organized
    - Documented
    - Stored
    - Secured
    - Shared
    - Archived
    - Preserved

Data lifecycle picture from UK Data Archive

what one will do in each stage depends largely on the type of data

General sections

0. Cover Information
1. Data Collection
2. Data Storage and Security
3. Data Documentation
4. Data Access and Sharing
5. Data Preservation and Publication
6. Costs

** Most recently, developments also exist to better incorporate software and models*

Key points to remember for a DMP

- **Plan early & often:** Involve data management planning from project onset, including data sharing and preserving considerations
- **Keep it simple:** Use clear and concise language; a DMP needs to be available for 20 years after research is finished
- **Be specific:** Provide detailed information about data handling decisions, both at the present time and those expected at a later time
- **Be flexible:** Allow for adjustments as the project progresses, e.g. as new tools become available, the situation changes, or research methodology changes
- **Collaborate:** Consult with data, IT, and other experts
- **Plan to be ‘as open as possible, as closed as necessary’**

Why draft a Data Management Plan (DMP)?

- Helps one to think about how to carry out research
- Provides information to others who may need to help
- Complies with funding agency or other requirements
- Ensures data quality and integrity
- Facilitates data sharing and reuse
- Protects intellectual property rights
- Enhances research reproducibility
- Clarifies roles and responsibilities

How is a DMP drafted?

- Methods:
  - Via paper and pen ☺
  - Via a word processor (e.g. in Microsoft Word)
  - Via an online tool (eg. DMPOnline, DMP Tool, DAMAP, DSW, OPIDOR, Argos)
  - Via an online tool for machine-actionability (next presentation)

Taking a closer look: *example template from Leiden University*

via word processor, e.g. Microsoft Word

Universiteit Leiden

Leiden University Data Management Plan template, v 4.3
Latest version of the template available on Zenodo:
http://doi.org/10.5281/zenodo.10210504
Tips and tricks for writing a Data Management Plan Leiden University:
https://doi.org/10.5281/zenodo.3903307

Administration details

0.1	Contact details	*Name and email address*
0.2	ORCID	*Add link to ORCID info*
0.3	Name of project to which this data management plan applies	*Name your project, group and supervisors*
0.4	Description of the research	*Briefly describe your research to help others understand the purposes for which the data are being collected or created. Max. 50 words.*
0.5	Project duration	Start: *DD-MM-YYYY* End: *DD-MM-YYYY*
0.6	Names of people and their responsibilities for data management	*List name, position, affiliation and ORCID (if known) plus the responsibility for data associated with this project, for example, collecting data, describing data, giving permissions for sharing and archiving the data.* *Naming anyone with specific roles and responsibilities for data management is especially important for collaborative projects that involve many researchers and/or partner organisations.*
0.7	Funding body(ies)	*If applicable. This is important because specific requirements may apply*
0.8	Grant number	*If applicable. A grant number provides unique identification for the grant.*
0.9	Partner organisations	*If applicable. These may be research partners that use your data, or that you use data from.*
0.10	Relevant agreements or protocols	*Mention any consortium agreement, institutional data protocol or data management plan at group level, etc. that complements this plan.*
0.11	Ethical review	*If applicable mention the registration number of your protocol and the name of the ethics committee.*
0.12	Personal data	☐ I do not collect personal data. ☐ I collect personal data and I will contact the privacy officer. ☐ I collect personal data and I have filled out a Data Processing Register for Research.
0.13	Name of data management support staff consulted during the preparation of this plan	*Mention name of individual and / or organisational unit*
0.14	Date of consultation with support staff	*DD-MM-YYYY*

Leiden University Template Data Management Plan v. 4.3 / 20231124

About this Data Management Plan

Date of creation	*DD-MM-YYYY*
Updated on (date of last update):	*DD-MM-YYYY* *A new version of the DMP should be created whenever important changes to the project occur due to inclusion of new data sets, changes in consortium policies or external factors.*
Changes in this version	*Indicate here what changes have been made to this plan since the last*

Data Collection

1.1	**Will the project use existing or third party data ?** ☐ No ☐ Own / group previous research ☐ Academic collaborators ☐ Commercial collaborators ☐ Publicly available database / archive ☐ Specialist commercial data provider ☐ Other (please specify) *If you will use existing or third party data describe briefly origin and type of existing data.*
	.....
1.2	**Is there an agreement for the use of existing data?** ☐ No ☐ Yes, I have a data transfer agreement (DTA) ☐ Yes, this is written down in a consortium agreement ☐ Yes, this is written down in a research agreement ☐ Yes, other (please specify) *Indicate if there any restrictions or requirements for use of this data, such as licensing conditions, informed consents, credits for data providers, permission to merge datasets?*
	.....
1.3	**How will you collect and/or create your data?** *Describe the research methodology.*
	.....
1.4	**What type(s) of data will you collect and create, and in what file format(s)?** *Consider also the data created by the processing of the raw data. See 'Tips and tricks' for types of data and sustainable formats.*
	.....
1.5	**What tools, instruments, equipment, hardware or software will you use to capture, produce, collect, create and process the data?** *Please give the names of the tools and state if they are already available. If not, state how you intend to acquire them. If applicable, describe whether you use a paper or electronic labjournal.*
	.....
1.6	**What is the estimated size of the data?** *Make an estimation (adopt stages if applicable).*

Leiden University Template Data Management Plan v. 4.3 / 20231124

Data stage	Specification of type of research data	Software choice and file format	Maximum data size during project	Final data size when project is finished
Raw data				
Processed data				
Published data				
Other…				

Data Storage & Security

2.1	**Are there any commercialisation, ethical or confidentiality restrictions about handling your data during your research?** ☐ No ☐ Yes *Please describe restrictions (e.g. contractual obligations, copyright, intellectual property, patentability, protection of personal data (privacy law), ethical restrictions, informed consent, institutional policies)*
	.....
2.2	**What are the main risks to data security?** ☐ Accidental deletion or file corruption ☐ Theft of, or damage to, equipment ☐ Overwriting or version loss ☐ Data leak, unauthorized access, or unauthorized use ☐ Other (explain): … *Please describe what would happen if the data get lost or become unusable.*
	.....
2.3	**What measures will you take to comply with the security requirements and to mitigate the risks described above?** *Describe how you can restore your data in the event of data loss and who is responsible.* *If applicable, please describe procedures to ensure personal data are handled confidentially and who is responsible.* ☐ Access restrictions (physical or digital) ☐ Encryption ☐ Reduce data sensitivity ☐ Regular and timely back-ups ☐ Master (locked) copy stored on university network storage ☐ Master (locked) copy stored elsewhere ☐ Data handling procedures and/or training for data handlers ☐ Other, namely: … *Please explain.*
	.....
2.4	**Where will you store your data during your research?** *Multipe answers possible.* ☐ On university departmental network storage / workgroups (J:) ☐ On university personal network storage (P:) ☐ In a virtual research environment (Sharepoint) ☐ Physical storage (e.g. USB, external hard drive) ☐ Cloud service (e.g. SURFdrive) ☐ Institutional service (e.g. Dataverse), namely:

Leiden University Template Data Management Plan v. 4.3 / 20231124

Data Documentation

	☐ Other, namely: …
	.....
2.5	**How will your data be backed up?** *Please specify briefly for each storage device frequency, location of backups and who is responsible.* ☐ I store my data on the university network storage which is backed-up by the ISSC. ☐ I have my own provision which I describe below.
	.....
2.6	**Are there any non-digital data or outputs that the project will generate? Where will these outputs be stored?** *Do you have a protocol for the storage and deletion of non-digital data? Please specify briefly and describe who is*
3.1	**Do you use a standard or convention for file naming and folder structures?** ☐ Yes, I use a standard that is common in my discipline, namely *(specify below)*: ☐ Yes, my group has a convention which I will describe below. ☐ No, I use my own method which I will describe below.
	.....
3.2	**How do you handle version control to maintain all changes that are made to the data?** *Please explain your choice briefly. Remember to also document any deletion of data, if applicable.* ☐ I lock raw data ☐ No version control (e.g. original files are overwritten) ☐ Version control software, namely: … ☐ Data/version number in filename/folder ☐ 'Track changes' feature in software ☐ By saving the script with which I process my data ☐ Other, namely: …
	.....
3.3	**What standard will you use to describe your data?** *Please refer to any metadata standards in your field if they exist.* ☐ I have a discipline-specific metadata standard, namely:… ☐ Archival metadata standard (e.g. Dublin Core), namely:… ☐ Other metadata standard, namely: … ☐ I have my own documentation which I will describe below.
	.....
3.4	**Where will metadata be registered?** ☐ In a separate README file ☐ Within the data file(s) ☐ As a separate formatted file ☐ In the data storage interface / platform ☐ Other, namely……:
	.....
3.5	**What supporting information / documentation will be needed to understand and reuse the data** *Please describe briefly how peers should be able to understand the data. Examples are lab journals, a codebook, survey questions, software documentation, readme.txt etc. Some institutes have mandatory publication packages.*

Leiden University Template Data Management Plan v. 4.3 / 20231124

Taking a closer look: *example template from Leiden University*

via word processor, e.g. Microsoft Word

Data Access & Sharing

4.1	**During your project, before publishing your results, with whom will you share your data?** *In most cases you will share raw or processed data with at least one person.*
	☐ Only my supervisor ☐ My immediate collaborators (including supervisor) ☐ Collaborators and (consortium) partners ☐ Any researcher in my field ☐ Anyone interested
	.....
4.2	**Are any restrictions placed on sharing your data?** *Please account for not sharing (parts of) your data.*
	☐ I have no restrictions ☐ I have restrictions on sharing (parts of) my data but I will share at least the metadata. Restrictions are due to: ☐ Protection of personal data ☐ Intellectual property ☐ Copyright ☐ Commercial reasons ☐ Security-related issues ☐ Ethical issues ☐ Other (explain):
	.....
4.3	**Do your participant consent forms include information about sharing, retention and deletion of data?**
	☐ Not applicable. ☐ Yes. *Please specify the relevant formula in the consent form.*
	.....
4.4	**What conditions do you set when you share your data?** *Describe the criteria that you apply to give access to your data, such as use limited to specific purposes, security measures, deletion after use, credits etc. Have these conditions been defined in a consortium or data sharing agreement (or equivalent)?*
	.....
4.5	**Who has authority to grant access to your data?** *Please describe briefly.*
	☐ You ☐ A colleague from the project, namely: … ☐ Supervisor ☐ Data Access Committee ☐ Funder ☐ Collaborator / research partner organisation ☐ Other, namely: …
	.....
4.6	**How will potential users discover your data?** *Interested users might find out about your data on your projectwebsite, via papers on conferences, data catalogue, University website, Dataverse, data repository etc.. Think of the different audiences.*

Data Preservation & Publication

5.1	**Are there any requirements on making (part of) your data public after your project?** *If your funder or publisher requires you to make your data FAIR (Findable, Accessible, Interoperable, Reusable) and / or open, please provide link to relevant policies or guidelines. Will your data be available freely or upon request? Mention embargo period if applicable.*
	☐ I will share (all or parts of) my data open access immediately upon publishing my results ☐ I will share my data upon request for the reasons I explain: ☐ I will share (all or parts of) my data open access after an embargo period of …. *(specify and explain)* ☐ I cannot share the following data for the reasons I explain:
	.....
5.2	**Which criteria do you use to decide which data has to be archived for the long term?** *Please briefly describe your choices.*
	☐ Type of data (raw, processed) and how easy it is to reproduce it ☐ Relevance of content for others ☐ Usability of format for others ☐ Data underlying publications ☐ Verification of research ☐ Costs ☐ Other, namely: …
	.....
5.3	**How long should your data be preserved?** *According to the University regulations on data management you must make your data findable, accessible and reusable for at least 10 years after the closure of a project. State any other obligations set by law, funder, etc. if any.*
	.....
5.4	**Are there any requirements regarding the destruction of data (digital and non-digital)?**
	.....
5.5	**Which of the following will you use for long-term findability and availability of your data?**
	☐ I will deposit data in a trusted data repository (e.g. DANS Easy, 4TU.ResearchData) as indicated below: ☐ According to the data protocol of my institute, I will archive data in the data repository indicated below (e.g. Dataverse): ☐ I will deposit data in a discipline-specific data repository as indicated below: ☐ I will use an archive specifically for my collaboration, namely: ☐ I will not use a data repository and will explain below how I will make my data findable and accessible for the long term. ☐ I will not make my data findable and accessible and I will explain why.
	.....
5.6	**If archiving in a data archive or repository, does it provide a persistent identifier (PID)?** *Please add additional information if needed and the PID when available.*
	☐ Yes, a DOI ☐ Yes, a different PID ☐ No
	.....
5.7	**What will you do to prepare your data for archiving? Will there be extra costs for this preparation?** *Describe how you intend to meet publisher or database / archive / repository requirements, e.g. converting the file formats, providing supplementary documentation. Mention (expected) costs in section 6.*
	.....

5.8	**Describe your strategy for publishing software that will be generated in this project.** *Indicate whether potential users need specific tools or software (e.g. specific scripts, codes or algorithms developed during the project) to access, interpret and (re-)use the data.*
	.....
5.9	**What license will you apply to your data?**
	☐ I will use the default license of the repository, namely: ☐ I will use a creative common license, namely: ☐ I will use an open source licence, namely: ☐ Other
	.....
5.10	**Who is responsible for the data after the project ends?**

Costs

6.1	Estimation and coverage of costs for data management

	Costs (in Euros)	Covered in the project budget yes / no	If not in budget, covered by …
Storage costs during the project			
Archive costs after the project			
Other costs (personnel for curation or datastewardship, etc., see guidance)			

Specify other costs

.....

Additional Information

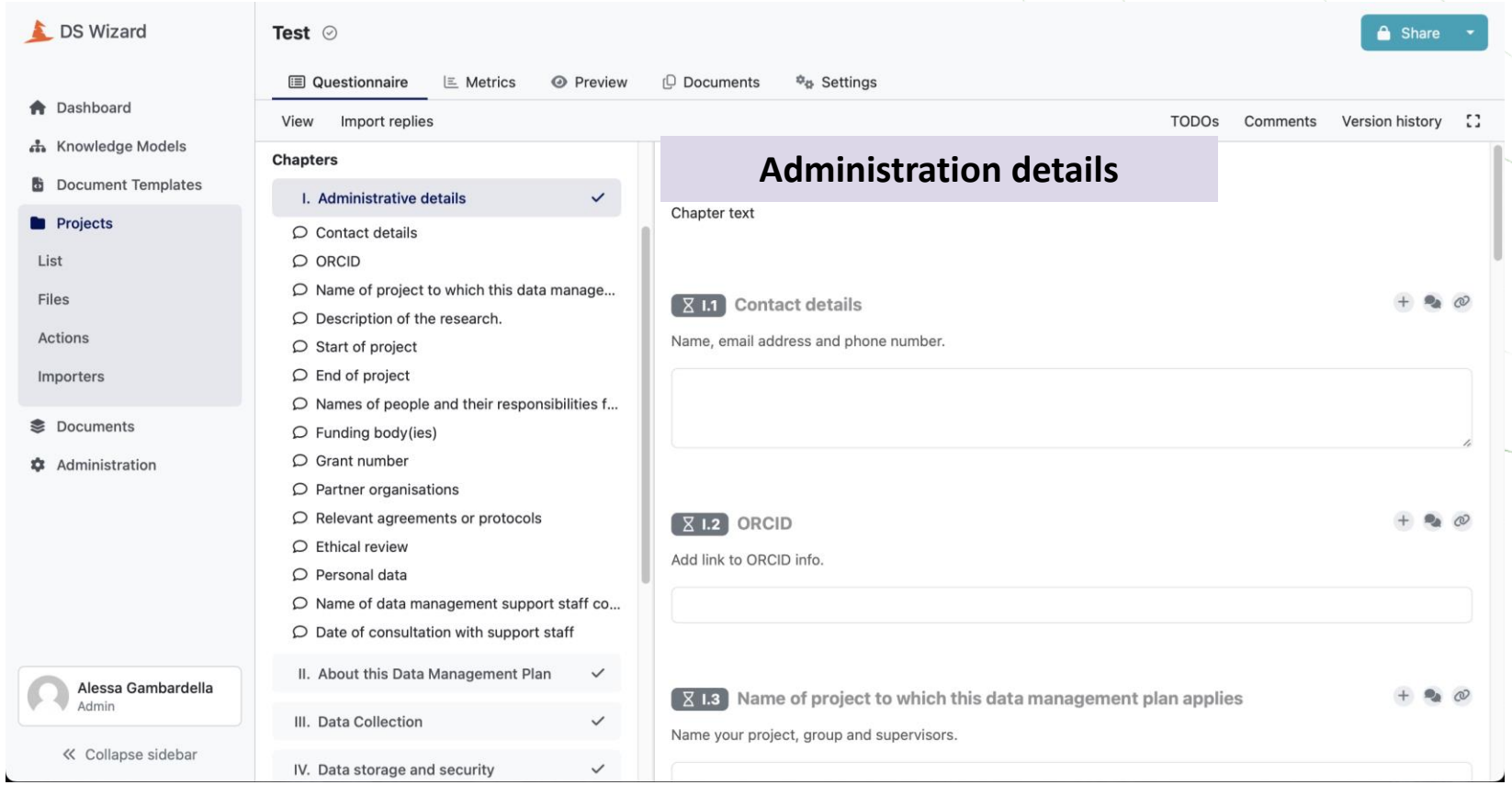
7.1	**Here you can put any additional information that you were not able to list in the boxes above.**
	.....

8. Review	
	If required by your institute or faculty, have your DMP reviewed.

	Submitted by _____ on DD/MM/YYYY (name PhD student) Reviewed by _____ on DD/MM/YYYY (name reviewer)

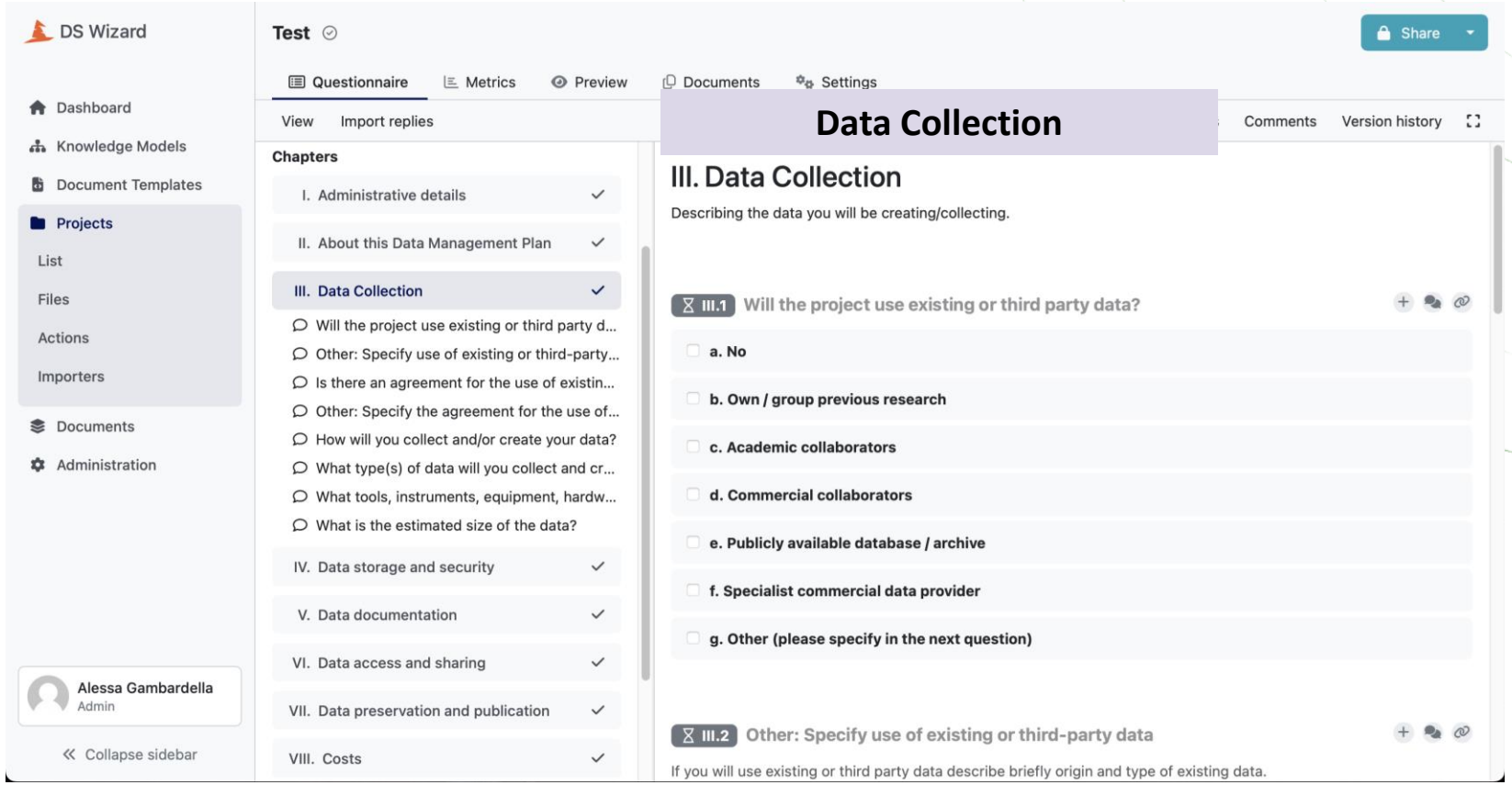
Taking a closer look: *example template from Leiden University*

- via online tool, e.g. Data Stewardship Wizard (DSW)



Taking a closer look: *example template from Leiden University*

- via online tool, e.g. Data Stewardship Wizard (DSW)



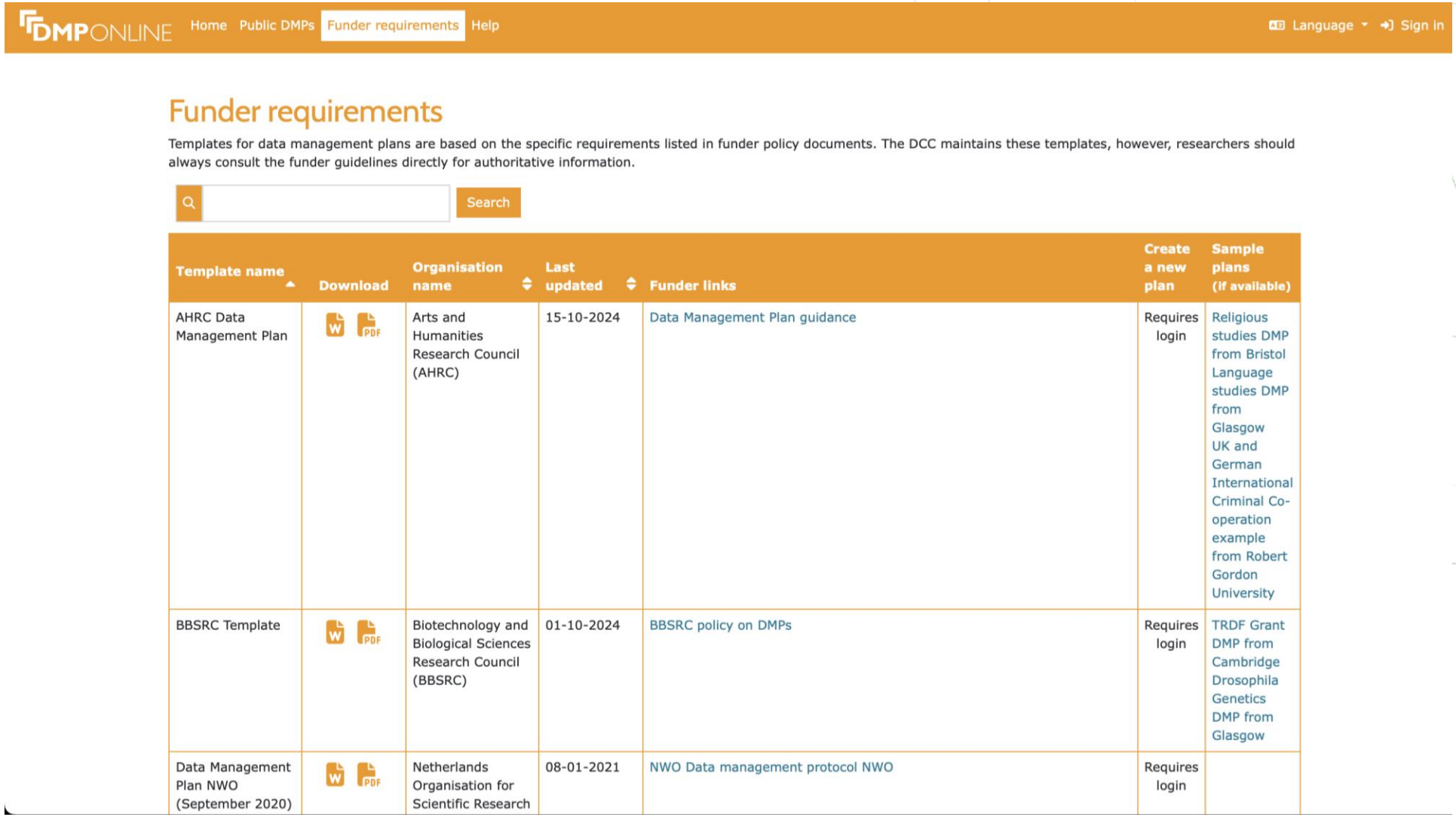
Taking a closer look: *comparing DMP templates*

- similar but different and partially inspired by FAIR

Leiden	NWO	ERC	H2020	ZonMW
1. Data collection	1.What data will be collected or produced, & what existing data will be re-used?	Summary	1. Data summary	1. General features of the project & data collection
2. Data storage and security	3. How will data and metadata be stored & backed up during the research?	5. Allocation of resources & data security	2.2 Making data openly accessible / 4. Data security	1.10 Storage / 2.3-4 legislation & protection
3. Data documentation	2. What metadata & documentation will accompany the data?	1. Making data findable / 3. Making data interoperable	2.1 Making data findable / 2.3. Making data interoperable	3 Making data findable / 5 Making data interoperable
4. Data access and sharing	4. How will you handle issues regarding the processing of personal information & intellectual property rights & ownership?	2. Making data openly accessible / 4. Increase data re-use	2.4 Increase data re-use / 5 Ethical aspects	2. Legislation / 4. Making data accessible
5. Data preservation and publication	5. How and when will data be shared & preserved for the long term?	2. Making data openly accessible / 4. Increase data re-use / 5. Allocation of resources & data security	2.2 Making data openly accessible / 2.4 Increase data re-use	4. Making data accessible / 6. Making data re-usable
6. Costs	6. Data management costs	5. Allocation of resources & data security	3. Allocation of resources	6. Making data re-usable

Table copied from: Leiden University Libraries, Centre for Digital Scholarship

Taking a closer look: *resource for comparing DMP templates*



DMPONLINE Home Public DMPs Funder requirements Help Language Sign in

Funder requirements

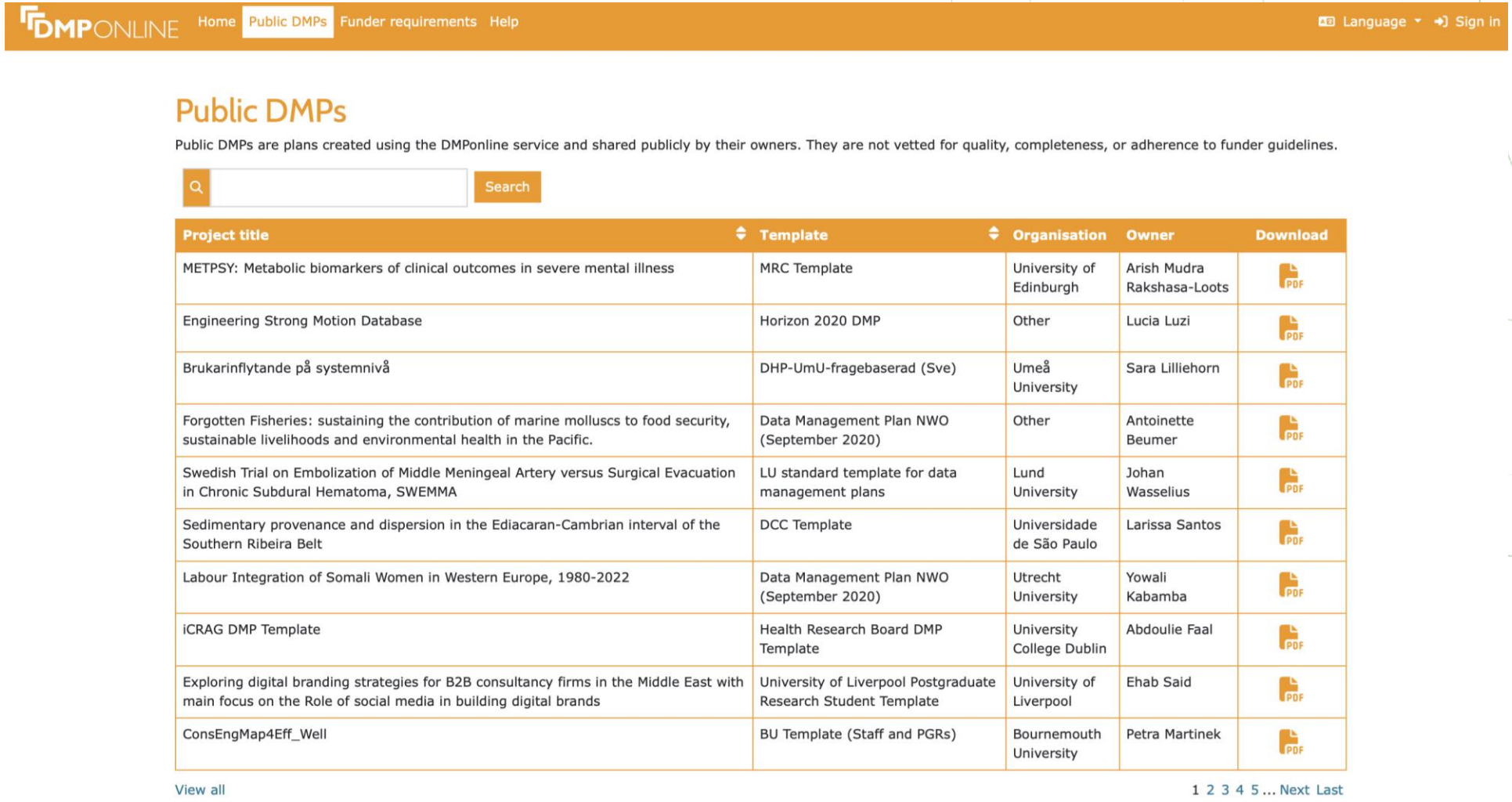
Templates for data management plans are based on the specific requirements listed in funder policy documents. The DCC maintains these templates, however, researchers should always consult the funder guidelines directly for authoritative information.

Search

Template name	Download	Organisation name	Last updated	Funder links	Create a new plan	Sample plans (if available)
AHRC Data Management Plan		Arts and Humanities Research Council (AHRC)	15-10-2024	Data Management Plan guidance	Requires login	Religious studies DMP from Bristol Language studies DMP from Glasgow UK and German International Criminal Co-operation example from Robert Gordon University
BBSRC Template		Biotechnology and Biological Sciences Research Council (BBSRC)	01-10-2024	BBSRC policy on DMPs	Requires login	TRDF Grant DMP from Cambridge Drosophila Genetics DMP from Glasgow
Data Management Plan NWO (September 2020)		Netherlands Organisation for Scientific Research	08-01-2021	NWO Data management protocol NWO	Requires login	

DMPOnline – Funder requirements: https://dmponline.dcc.ac.uk/public_templates

Taking a closer look: *resource for completed DMP examples*



DMPONLINE Home | Public DMPs | Funder requirements | Help | Language | Sign in

Public DMPs

Public DMPs are plans created using the DMPonline service and shared publicly by their owners. They are not vetted for quality, completeness, or adherence to funder guidelines.

Search

Project title	Template	Organisation	Owner	Download
METPSY: Metabolic biomarkers of clinical outcomes in severe mental illness	MRC Template	University of Edinburgh	Arish Mudra Rakshasa-Loots	PDF
Engineering Strong Motion Database	Horizon 2020 DMP	Other	Lucia Luzi	PDF
Brukarinflytande på systemnivå	DHP-UmU-fragebaserad (Sve)	Umeå University	Sara Lilliehorn	PDF
Forgotten Fisheries: sustaining the contribution of marine molluscs to food security, sustainable livelihoods and environmental health in the Pacific.	Data Management Plan NWO (September 2020)	Other	Antoinette Beumer	PDF
Swedish Trial on Embolization of Middle Meningeal Artery versus Surgical Evacuation in Chronic Subdural Hematoma, SWEMMA	LU standard template for data management plans	Lund University	Johan Wasselius	PDF
Sedimentary provenance and dispersion in the Ediacaran-Cambrian interval of the Southern Ribeira Belt	DCC Template	Universidade de São Paulo	Larissa Santos	PDF
Labour Integration of Somali Women in Western Europe, 1980-2022	Data Management Plan NWO (September 2020)	Utrecht University	Yowali Kabamba	PDF
iCRAG DMP Template	Health Research Board DMP Template	University College Dublin	Abdoulie Faal	PDF
Exploring digital branding strategies for B2B consultancy firms in the Middle East with main focus on the Role of social media in building digital brands	University of Liverpool Postgraduate Research Student Template	University of Liverpool	Ehab Said	PDF
ConsEngMap4Eff_Well	BU Template (Staff and PGRs)	Bournemouth University	Petra Martinek	PDF

View all

1 2 3 4 5 … Next Last

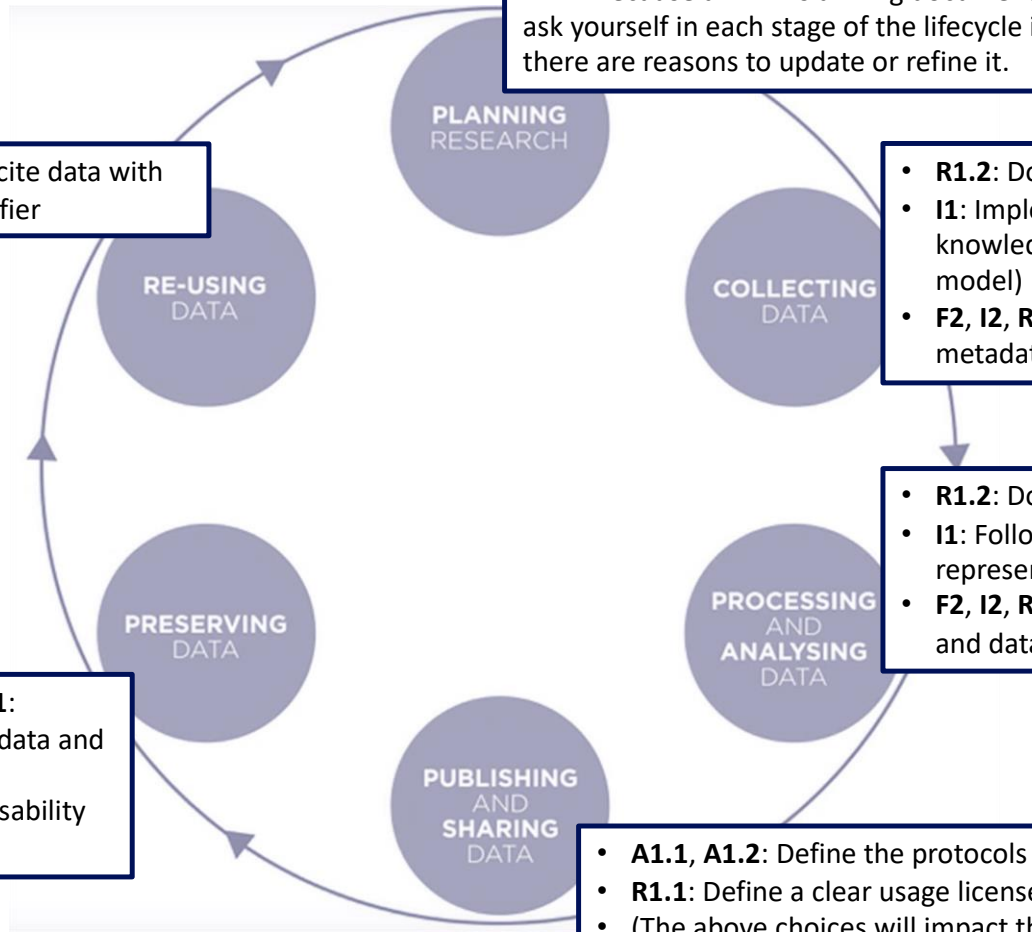
DMPOnline – Public DMPs: https://dmponline.dcc.ac.uk/public_plans

How does a DMP help in making data FAIR?

PRINCIPLES	Aspects of the DMP
FINDABLE	organization, documentation, naming conventions
ACCESSIBLE	storage, access permissions, preservation
INTEROPERABLE	formats, standards, vocabularies
REUSABLE	sharing, licenses, retention strategies

a well-crafted DMP is a roadmap for making research data FAIR

Revisiting the data lifecycle with the FAIR Principles



Plan for FAIR and well-managed data with a DMP. Because a DMP is a living document, ask yourself in each stage of the lifecycle if there are reasons to update or refine it.

- **F1**: Make sure you cite data with its persistent identifier

- **R1.2**: Document what you do
- **I1**: Implement procedure for knowledge representation (data model)
- **F2**, **I2**, **R1.3**: Implement choice of metadata and data fields

- **R1.2**: Document what you do
- **I1**: Follow procedure for knowledge representation (data model)
- **F2**, **I2**, **R1.3**: Follow choice of metadata and data fields

- **A1.1**, **A1.2**: Define the protocols for access
- **R1.1**: Define a clear usage license
- (The above choices will impact the choice of repository)

- **F1**, **F3**, **F4**, **A1**, **A2**, **I3**, **R1**: Deposit data with metadata and documentation for interoperability and reusability with repository

Aktau, A., Gambardella, A., Hettne, K., & Ulzen van, N. (2023). “Final outcomes of the taskforce FAIRification as as Service.” Zenodo. doi: 10.5281/zenodo.7546767.

Current DMP challenges are like those for managing data itself

- Unclear expectations
- Lack of common standards
- Seen as pushing paper
- Time consuming, for researchers and reviewers
- Not easily reusable, by oneself or others
- Lack of sharing
- Lack of recognized value

Current DMP challenges are like those for managing data itself

- Unclear expectations
- Lack of common standards
- Seen as pushing paper
- Time consuming, for researchers and reviewers
- Not easily reusable, by oneself or others
- Lack of sharing
- Lack of recognized value

Many current roads toward making DMPs FAIR themselves

Standardization

- E.g. RDA Common Standard

Machine ‘actionable’ DMPs (maDMPs)

- Many tools are looking into various export formats
- E.g. European Project funded by Horizon Europe ‘OS Trails’
- Readable, interpretable, actionable…

FAIR maDMPs

- Mapping the FAIR principles (e.g. FIP to DMPs)

Miksa, T., Walk, P. and Neish, P. (2020) “RDA DMP Common Standard for Machine-actionable Data Management Plans”. Zenodo. doi: 10.15497/rda00039.; OS Trails: https://ostrails.eu; Hettne, K., Magana, B., Gambardella, A., Suchánek, M., Schoots, F., & Schultes, E. (2023). FIP2DMP: Linking Data Management Plans with FAIR Implementation Profiles. Fair Connect. doi: 10.3233/FC-221515.

Summary

- DMPs are a work in progress
  - the what, the why, and the how
- DMPs suffer from challenges like data
- DMPs and FAIR map over the data lifecycle
- FAIRify data *and* FAIRify DMPs

THANKS!

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani