Data Management Plan

Milica Ševkušić, EIFL

Data Management Plan

A formal document explaining how research data will be handled throughout the data lifecycle.

- Mandatory in the Horizon Europe and ERC projects (funded by the European Commission)
- National and institutional mandates
- Increasingly often required as part of PhD requirements

OpenAIRE

Training challenges

- Discipline-specific parts of DMPs
- The complexity of IPR-related issues
- Achieving FAIRness in practice
- Finding best practice examples

Source: https://www.freepik.com/vectors/performance-icon

Horizon Europe Data Management Plan Template

- Guidance: issues to be covered
- Structured approach
- Living document

Initial DMP → Updates → Periodic reporting

https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

https://enspire.science/wp-content/uploads/2021/09/Horizon-Europe-Data-Management-Plan-Template.pdf

EU Grants: Data Management Template (HE):V1.0 – 05.05.2021

The Horizon Europe Model Grant Agreement requires that a data management plan ('DMP') is established and regularly updated. The use of this template is recommended for Horizon Europe beneficiaries. In completing the sections of the template the requirements for research data management of Horizon Europe as described in article 17 and analysed in the Annotated Grant Agreement, article 17, must be addressed.

1. Data Summary

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

What types and formats of data will the project generate or re-use?

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

What is the expected size of the data that you intend to generate or re-use?

What is the origin/provenance of the data, either generated or re-used?

To whom might your data be useful ('data utility'), outside your project?

2. FAIR data

2.1. Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Will metadata be offered in such a way that it can be harvested and indexed?

2.2. Making data accessible

Repository:

Will the data be deposited in a trusted repository?

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Will the data be accessible through a free and standardized access protocol?

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

How will the identity of the person accessing the data be ascertained?

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

Metadata:

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

OpenAIRE

Assistance: tools

- Wizards
  - Web-based tools
  - Usually free for individual researchers
  - Inbuilt templates
  - Customizable (for institutions)
  - Machine readable DMPs
  - Integration with repositories
- Checklists
  - Swedish National Data Service DMP checklist
  - Customized versions are often found on the websites of academic libraries

DMPTool

DSW

argos

DMPONLINE

DataWiz

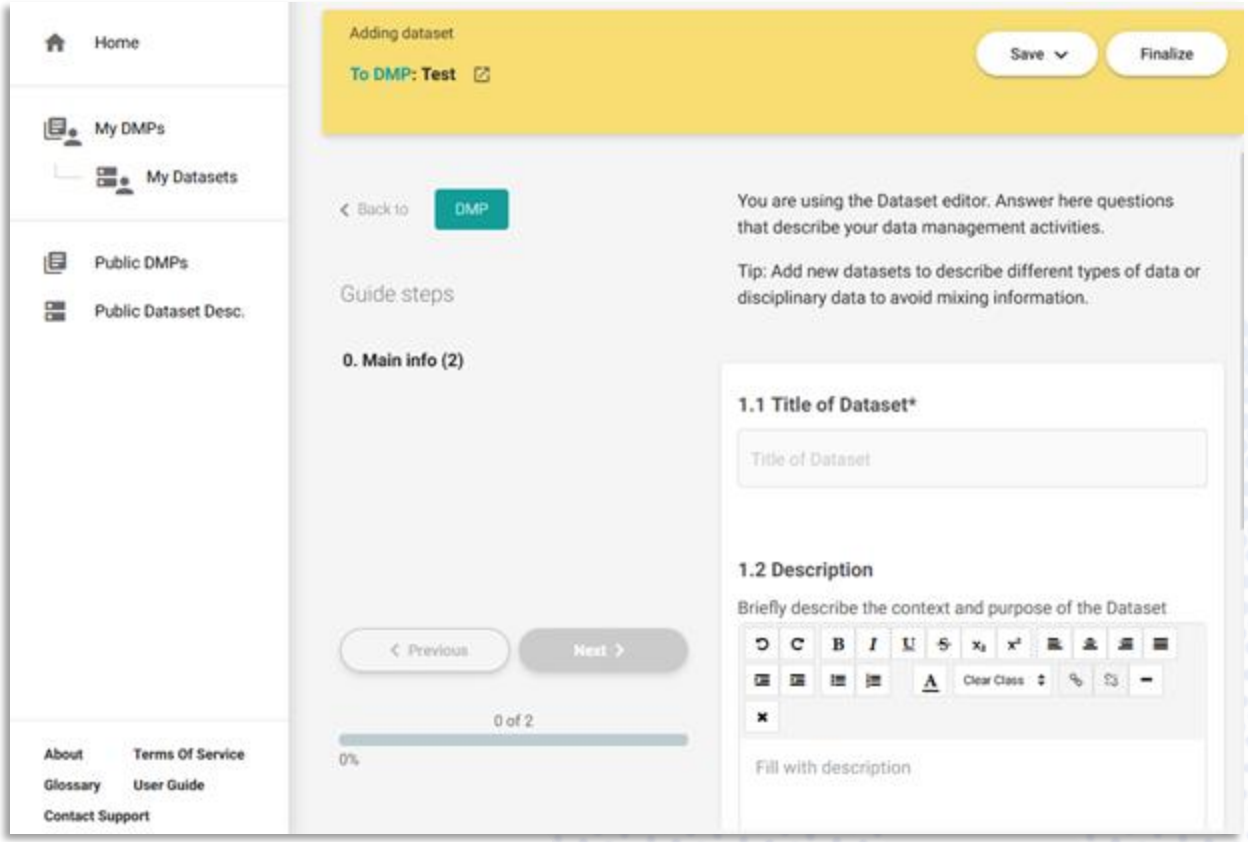
OpenAIRE

https://argos.openaire.eu

Argos

- Established by OpenAIRE
- Based on open-source software
- Free for individual researchers
- User interface in many languages
- Templates
- Public DMPs and datasets
- Machine actionable (based on the RDA Common Standard for machine-actionable Data Management Plans)
- Export: JSON, xml
- Publish on Zenodo

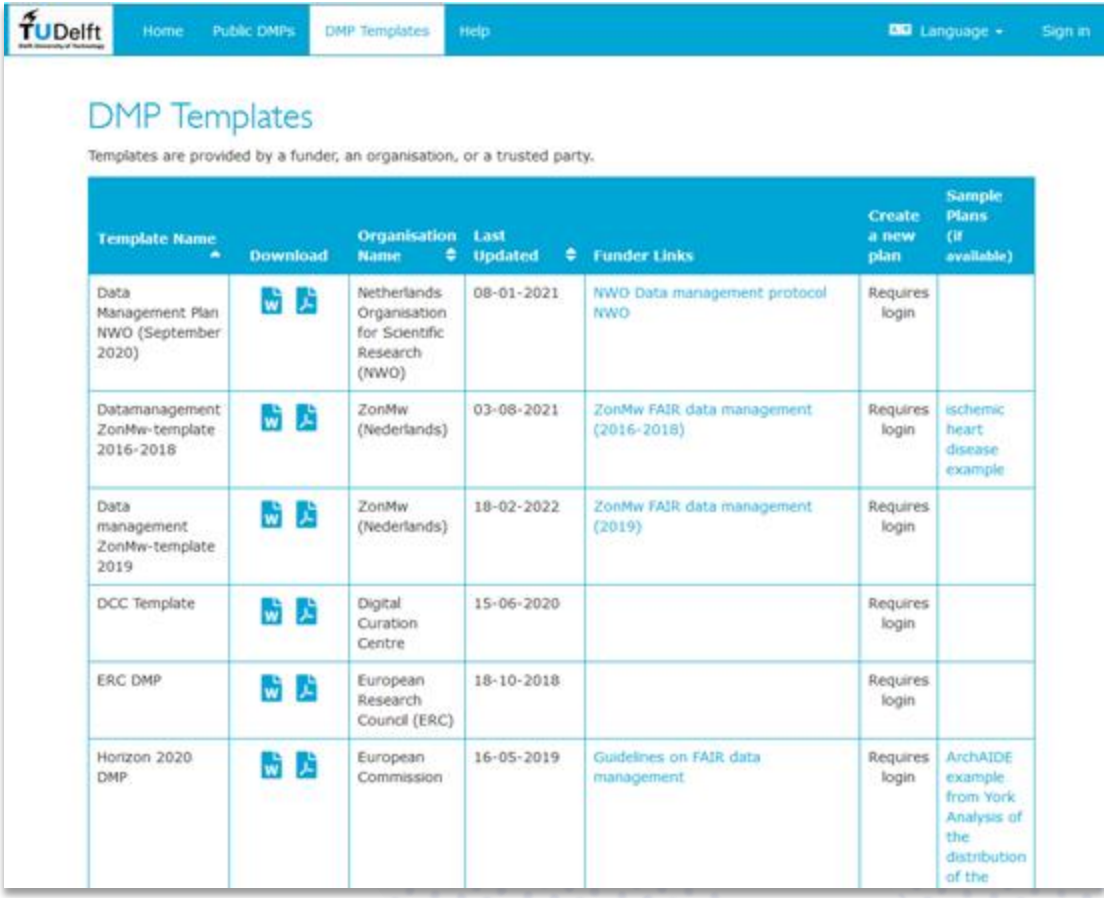
- Strong community support – monthly calls: https://www.openaire.eu/argos-community-calls



DMP Online

- Established by DCC
- Free for individual researchers
- Institutional instances
- Templates
- Public DMPs
- Export: xml, csv, txt, PDF, html

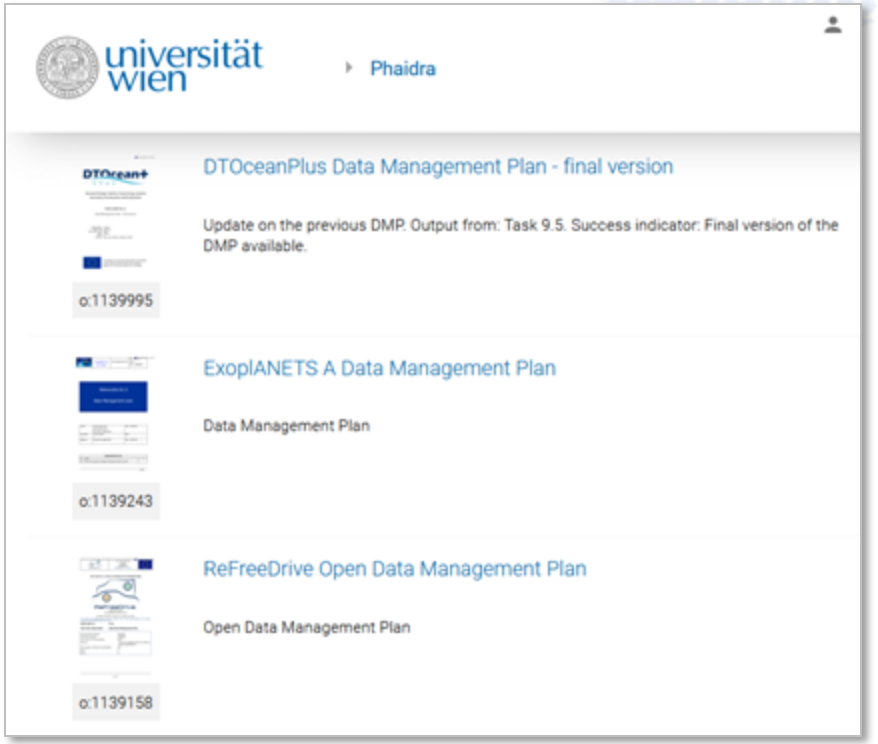
https://dmponline.dcc.ac.uk/



OpenAIRE

Best practice: DMP examples and use cases

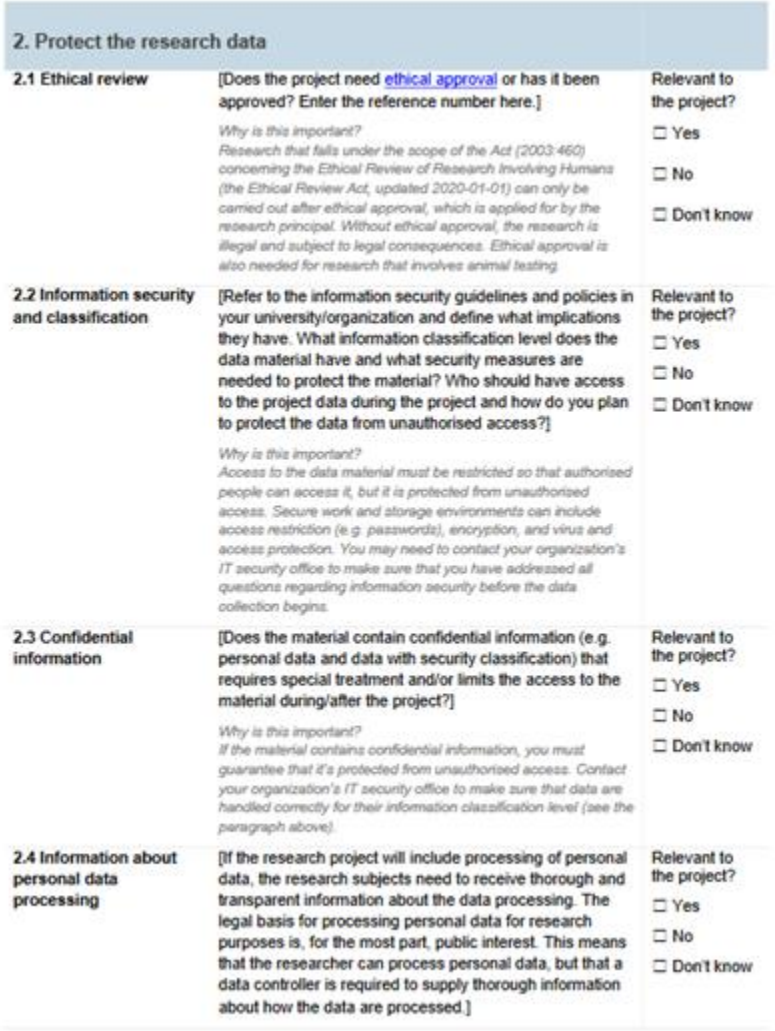
- DMP Use Case Project (OpenAIRE Austria): a collection of public DMPs of EC-funded projects: https://hdl.handle.net/11353/10.1140797
- Public DMPs in Argos, DMP Online etc.
- Search in Zenodo



Swedish National Data Service DMP checklist

https://snd.gu.se/en/manage-data/guides/dmp-checklist

- "The SND DMP checklist is designed so that it can be used as a complete and comprehensive plan for an entire research project. It can be used for different research areas, data types, phases of the research process, requirements from funding bodies, as well as various legal requirements."
- Very useful to those working with sensitive data



2. Protect the research data		
2.1 Ethical review	[Does the project need ethical approval or has it been approved? Enter the reference number here.] *Why is this important?* *Research that falls under the scope of the Act (2003:460) concerning the Ethical Review of Research Involving Humans (the Ethical Review Act, updated 2020-01-01) can only be carried out after ethical approval, which is applied for by the research principal. Without ethical approval, the research is illegal and subject to legal consequences. Ethical approval is also needed for research that involves animal testing.*	Relevant to the project? ☐ Yes ☐ No ☐ Don't know
2.2 Information security and classification	[Refer to the information security guidelines and policies in your university/organization and define what implications they have. What information classification level does the data material have and what security measures are needed to protect the material? Who should have access to the project data during the project and how do you plan to protect the data from unauthorised access?] *Why is this important?* *Access to the data material must be restricted so that authorised people can access it, but it is protected from unauthorised access. Secure work and storage environments can include access restriction (e.g. passwords), encryption, and virus and access protection. You may need to contact your organization's IT security office to make sure that you have addressed all questions regarding information security before the data collection begins.*	Relevant to the project? ☐ Yes ☐ No ☐ Don't know
2.3 Confidential information	[Does the material contain confidential information (e.g. personal data and data with security classification) that requires special treatment and/or limits the access to the material during/after the project?] *Why is this important?* *If the material contains confidential information, you must guarantee that it's protected from unauthorised access. Contact your organization's IT security office to make sure that data are handled correctly for their information classification level (see the paragraph above).*	Relevant to the project? ☐ Yes ☐ No ☐ Don't know
2.4 Information about personal data processing	[If the research project will include processing of personal data, the research subjects need to receive thorough and transparent information about the data processing. The legal basis for processing personal data for research purposes is, for the most part, public interest. This means that the researcher can process personal data, but that a data controller is required to supply thorough information about how the data are processed.]	Relevant to the project? ☐ Yes ☐ No ☐ Don't know

Discipline-specific DMPs

- Guidance Document Presenting a Framework for Discipline-specific Research Data Management (Science Europe, 2018), https://www.scienceeurope.org/our-resources/guidance-document-presenting-a-framework-for-discipline-specific-research-data-management
- Training materials in the SSH Open Marketplace: https://marketplace.sshopencloud.eu/search?order=score&q=data+management+plan&categories=training-material

Assessment frameworks for DMPs

- DMP Evaluation Rubric (part of Practical Guide to the International Alignment of Research Data Management)
- “Reverse engineering” ☺
- Structured approach
- Customized versions available on the websites of academic libraries

DMP question	DMP guidance	Performance level	
		Sufficiently addressed	Insufficiently addressed

Science Europe. (2021). Practical Guide to the International Alignment of Research Data Management - Extended Edition. https://doi.org/10.5281/zenodo.4915862

Science Europe. (2021). Practical Guide to the International Alignment of Research Data Management - Extended Edition. https://doi.org/10.5281/zenodo.4915862

2b **What data quality control measures will be used?**	• Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeated samples or measurements, standardised data capture, data entry validation, peer review of data, or representation with controlled vocabularies.	• Clearly describes the approach taken to ensure and document quality control in the collection of data during the lifetime of the project.	• Provides no information or only a vague mention on how data quality is controlled and documented during the lifetime of the project.
3 STORAGE AND BACKUP DURING THE RESEARCH PROCESS			
Guidance for Researchers		**Sufficiently Addressed The DMP…**	**Insufficiently Addressed The DMP…**
3a **How will data and metadata be stored and backed up during the research?**	• Describe where the data will be stored and backed up during research activities and how often the backup will be performed. It is recommended to store data in at least two separate locations. • Give preference to the use of robust, managed storage with automatic backup, such as provided by IT support services of the home institution. Storing data on laptops, stand-alone hard drives, or external storage devices such as USB sticks is not recommended.	• Clearly (even if briefly) describes: › The location where the data and backups will be stored during the research activities. › How often backups will be performed. › The use of robust, managed storage with automatic backup (for example storage provided by the home institution). or • Explains why institutional storage will not be used (and for what part of the data) and describes the (additional) locations, storage media, and procedures that will be used for storing and backing up data during the project.	• Provides no information or very vague reference to how data will be stored and backed up during the project.
3b **How will data security and protection of sensitive data be taken care of during the research?**	• Explain how the data will be recovered in the event of an incident. • Explain who will have access to the data during the research and how access to data is controlled, especially in collaborative partnerships.	• Clearly explains: › How the data will be recovered in the event of an incident. › Which institutional and/or national data protection policies are in place and provides a link to where they can be accessed. › Who will have access to the data during the research.	• Provides little or no details on how the data will be recovered in the event of an incident, which institutional data protection policies are in place, and who will have access to the data during the research.

Handling training challenges

- Make the best use of the structured approach
- Don’t obscure the complexity!
- Involve experts
- Use real-life examples

Source: https://www.freepik.com/vectors/computer-cartoon

Breakout rooms

- 6 groups
- 2 tasks
- But each group does only one task on Jamboard (find your group on a separate page): https://jamboard.google.com/d/1yS4TLT7hB3aY80s-9WyxPqsT8QIq341gz1EF5pRhEJc/edit?usp=sharing
- **Task 1**: Use colours to mark the sections/topics of the DMP that are the most difficult to present (left) and deliver (right).
- **Task 2**: Discuss the following questions: Where would you recommend Researcher A, B and C to deposit their data? Why? What are potential disadvantages of your choice? Can they share their data under an open licence? What tools and formats would you recommend?