

The data management plan

Stéphanie Cheviron



Part 1 | What is a DMP?

Part 2 | The funders' requirements

Part 3 | How to write a DMP?

Part 4 | More information

Part 1

What is a DMP?

What is a DMP?

To document the data life cycle of a research project, we use a **data management plan** (DMP)

It gives context to the data:

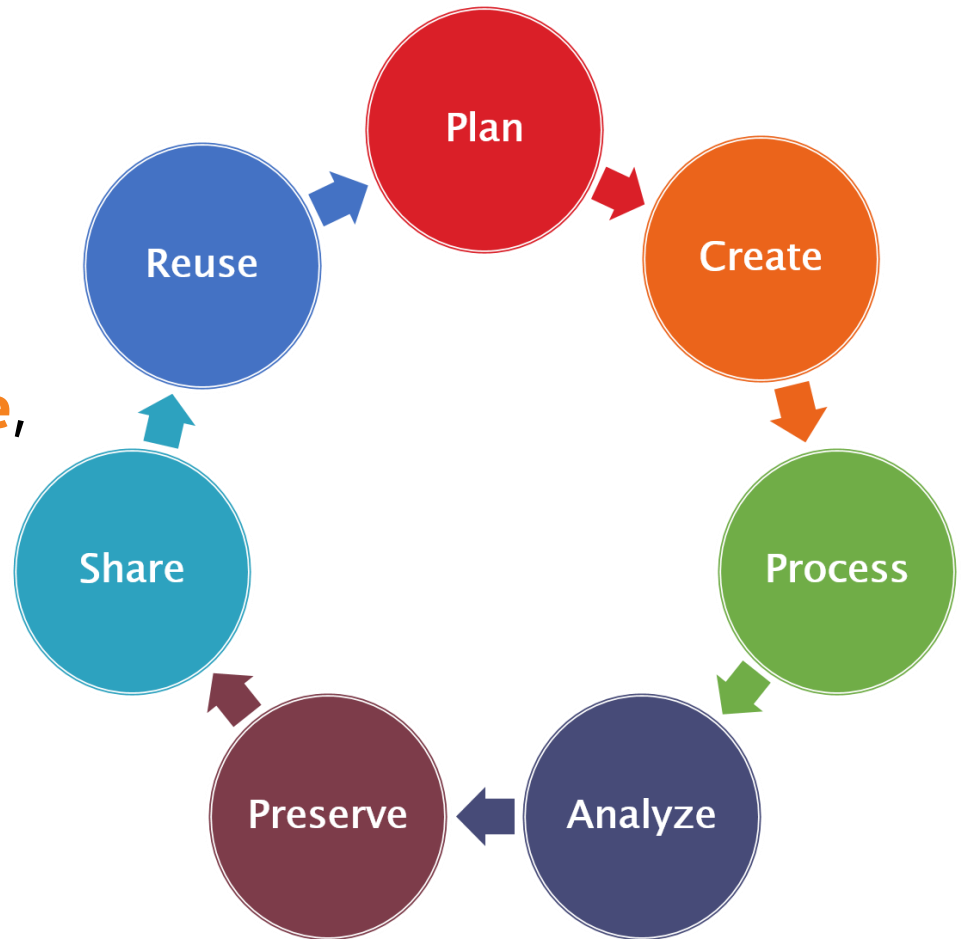
A DMP is a living document gathering **administrative, legal, technical and documentary information** about the data produced or collected during a research project.

It is increasingly requested by funders.

The data life cycle

It also makes it possible to **anticipate the different stages of the data life cycle**, right from the design stage of the research project.

What costs for managing my data?



Source : [Université d'Ottawa](#)

What do we find in a DMP?

You will always find these topics in a DMP:

- **Responsibilities:** who created the data? Who is responsible for their management and quality? Are contact details provided?
- **Data collected/generated:** provenance, files formats, volume, workflows,...
- **Metadata and documentation:** what are the metadata standards used to describe the data? Is there a data documentation? What are the file tree structure and naming convention?
- **Storage:** during the project, how are they stored and shared? What security measures are in place to protect access? What risk and confidentiality management?

What do we find in a DMP?

- **Ethics:** if the data is sensitive, has the project been submitted to the ethics committee?
- **Legal aspects:** Is the data subject to copyright or any other right that may restrict its dissemination?
- **Sharing and reusing:** Will the data be deposited in a data repository? Under which license? If not, why?
- **Archiving:** What is the planned strategy for long-term data preservation and access?
- **Resources :** what are the resources allocated to data management?

The FAIR principles

A set of 15 requirements to make your data:

→ **F**indable

ex : persistent identifiers (DOI, Handle,...), data repositories

→ **A**ccessible

ex : access to the metadata

→ **I**nteroperable

ex : controlled vocabulary

→ **R**eutilisable

ex : license



→ [DORANum](#)

→ [Force11](#)

Why a DMP?

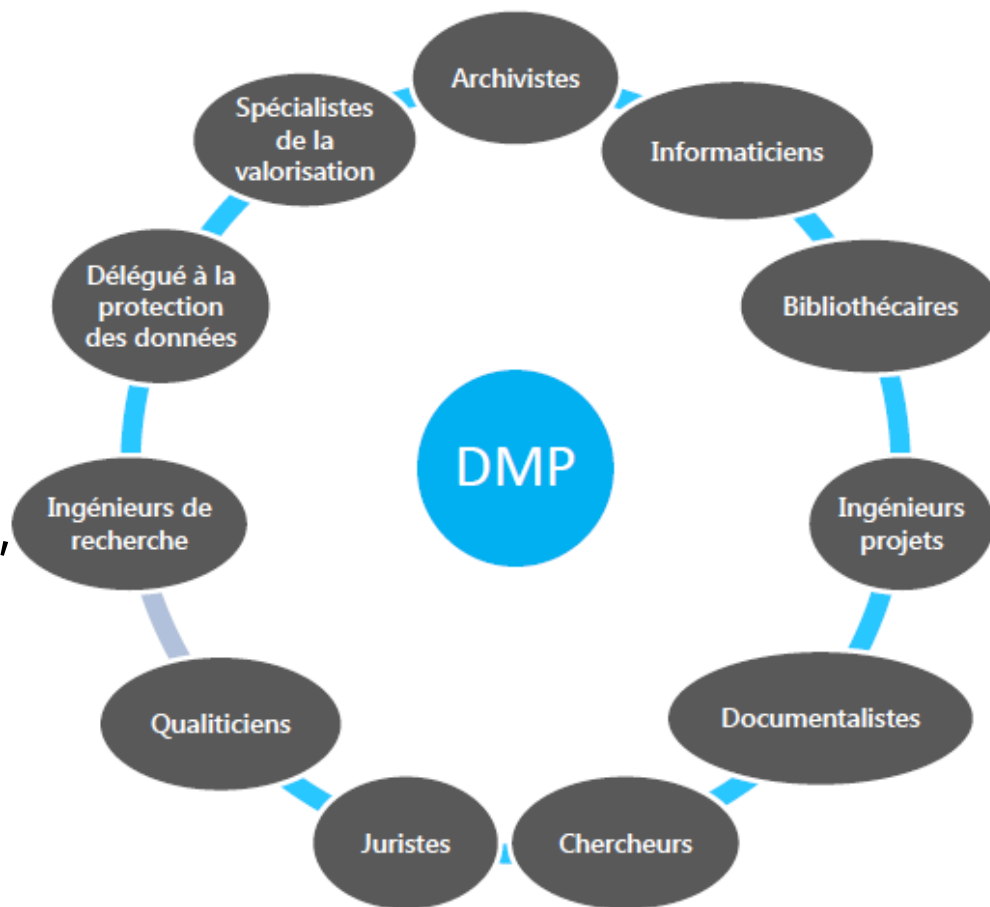
A DMP is more than a deliverable you submit to your funding agency:

- It contextualizes your data
→ Simplified **sharing** and **archiving**
- Increases the **reproducibility** and **integrity** of research
- It is useful when new members are joining your team.
→ **transmission of knowledge**

You are not alone!

Writing a DMP is a **collaborative work**

The document can be produced and updated by several members of the team, be reviewed by legal and IT experts, by data librarians, archivists, etc.



Nathalie Reymonet, Magalie Moysan, Aurore Cartier, Renaud Délémontez. Réaliser un plan de gestion de données « FAIR » : modèle . 2018. [{sic_01690547v2}](#)

Part 2

The funders' requirements



Horizon Europe & Euratom

The successor to the H2020 funding program requires:

- a **DMP as a deliverable regularly updated** during the research project
 - 1st version within the **first 6 months** of the project
 - A completed DMP at the end of the research project

Note: An updated DMP deliverable must also be produced mid-project (for projects longer than twelve months)

- FAIR data deposited in a **trusted data repository**:
 - As soon as possible
 - As indicated in the DMP
 - In open access



Horizon Europe & Euratom

- Beneficiaries must **explain in the DMP the exception(s)** under which they choose to or must restrict access to some or all of the research data
"As open as possible, as closed as necessary"
- **Use of CC-BY or CC-0** or equivalent licenses for open data
- Information via the repository about any other research products, instruments, tools needed to reuse or validate the data
- Metadata requirements are the same as for publications
- Costs associated with data management are **eligible for reimbursement**



More information on the [Annotated Model Grant Agreement \(AGA\)](#) [PDF], annex 5, Article 17, « HE COMMUNICATION, DISSEMINATION, OPEN SCIENCE AND VISIBILITY

Agence nationale de la recherche (ANR)

As of the 2019 AAPG, the ANR requires:

- To submit a **data management plan as a deliverable**:
 - **1st initial draft within the first 6 months of the project**
 - A completed DMP at the end of the research project

Please note: for the projects of more than 30 months, an updated DMP at mid-term

- To **deposit the data in a data repository** ("as open as possible, as closed as necessary").

Eligible costs

- All expenses related to data management: acquisition, collection, storage, dedicated staff...
- The cost of data storage in case of recourse to a third party is eligible **up to 5 years after the scientific end date of the project**, if contracted before the end of the project



V1 within the first **6 months**



V2 for projects of **more than 12 months**



V2 for projects of **more than 30 months**



V3 completed



Part 3

How to write a **DMP**?

DMP OPIDoR

(Bases on **DMP**

Roadmap by the

Digital Curation Centre

And the University of

California Curation Centre, and customized by the INIST for the French scientific community.



dmp.opidor.fr


Several templates are available: those of funders (ANR, H2020, ERC) and those of research organizations and universities.

A **Software Management Plan** (SMP) template is also available on OPIDoR.

Other DMP tools online: [Argos](#), [Data Stewardship Wizard](#), [DMPOnline](#), [DMPTools](#)

1. Data description and collection or re-use of existing data (2 questions)

1a. How will new data be collected or produced and/or how will existing data be re-used?

B *I*    

Enregistrer

Recommandations




Commentaires

ANR

- Explain which methodologies or software will be used if new data are collected or produced.
- State any constraints on re-use of existing data if there are any.
- Explain how data provenance will be documented.
- Briefly state the reasons if the re-use of any existing data sources has been considered but discarded

1. Data description and collection or re-use of existing data (2 questions)

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

B *I*    

Enregistrer

Recommandations

Commentaires

ANR

- Give details on the kind of data: for example numeric (databases, spreadsheets), textual (documents), image, audio, video, and/or mixed media.
- Give details on the data format: the way in which the data is encoded for storage, often reflected by the filename extension (for example pdf, xls, doc, txt, or rdf).
- Justify the use of certain formats. For example, decisions may be based on staff expertise within the host organisation, a preference for open formats, standards accepted by data repositories, widespread usage within the research community, or on the software or equipment that will be used.

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

B *I* ☰ ▾ ☰ ▾ 🔗 📄 ▾

Enregistrer

[Les métadonnées dans un DMP](#), Marie Puren, INRIA, 2017

[PDF]

[Comment décrire ses données avant de les publier ?](#) (DATAACC, physique/chimie)

[Understanding metadata. What is metadata, and what is it for?](#), Riley Jenn, NISO, 2017.

Recommandations

Commentaires

ANR

- Indicate which metadata will be provided to help others identify and discover the data.
- Indicate which metadata standards (for example DDI, TEI, EML, MARC, CMDI) will be used.
- Use community metadata standards where these are in place.
- Indicate how the data will be organised during the project, mentioning for example conventions, version control, and folder structures. Consistent, well-ordered research data will be easier to find, understand, and re-use.
- Consider what other documentation is needed to enable re-use. This may include information on the methodology used to collect the

2b. What data quality control measures will be used?

B *I*    

Enregistrer



[Guidelines for creating a README file](#), 4TU et université de Delft


Recommandations Commentaires

ANR

- 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.



3a. How will data and metadata be stored and backed up during the research?

B *I*    

Enregistrer

Recommandations


Commentaires

ANR

- 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 least at 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.

3. Storage and backup during the research process (2 questions)

3b. How will data security and protection of sensitive data be taken care during the research

B *I*    

Enregistrer

Recommandations

Commentaires

ANR

- 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.
- Consider data protection, particularly if your data is sensitive for example containing personal data, politically sensitive information, or trade secrets. Describe the main risks and how these will be managed.
- Explain which institutional data protection policies are in place.

4. Legal and ethical requirements, code of conduct (3 questions)

4a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

B *I*    

Enregistrer

Recommandations

Commentaires

ANR

Ensure that when dealing with personal data data protection laws (for example GDPR) are complied with:

- Gain informed consent for preservation and/or sharing of personal data.
- Consider anonymisation of personal data for preservation and/or sharing (truly anonymous data are no longer considered personal data).
- Consider pseudonymisation of personal data (the main difference with anonymisation is that pseudonymisation is reversible).
- Consider encryption which is seen as a special case of pseudonymisation (the encryption key must be stored separately from

4. Legal and ethical requirements, code of conduct (3 questions)

4b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

B *I*    

Enregistrer


Recommandations

Commentaires

ANR

- Explain who will be the owner of the data, meaning who will have the rights to control access:
 - Explain what access conditions will apply to the data? Will the data be openly accessible, or will there be access restrictions? In the latter case, which? Consider the use of data access and re-use licenses.
 - Make sure to cover these matters of rights to control access to data for multi-partner projects and multiple data owners, in the consortium agreement.
- Indicate whether intellectual property rights (for example Database Directive, *sui generis*)

4c. What ethical issues and codes of conduct are there, and how will they be taken into account?

B *I*    

Enregistrer


Recommandations

Commentaires

ANR

- Consider whether ethical issues can affect how data are stored and transferred, who can see or use them, and how long they are kept. Demonstrate awareness of these aspects and respective planning.
- Follow the national and international codes of conducts and institutional ethical guidelines, and check if ethical review (for example by an ethics committee) is required for data collection in the research project.

5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

B *I*    

Enregistrer



[Re3data](https://re3data.org/)

Recommandations

Commentaires

ANR

- Explain how the data will be discoverable and shared (for example by deposit in a trustworthy data repository, indexed in a catalogue, use of a secure data service, direct handling of data requests, or use of another mechanism).
- Outline the plan for data preservation and give information on how long the data will be retained.
- Explain when the data will be made available. Indicate the expected timely release. Explain whether exclusive use of the data will be claimed and if so, why and for how long. Indicate whether data sharing will be postponed or restricted for example to publish, protect

Re3data.org

Re3data.org is a registry of research data repositories.

The screenshot shows the Re3data.org search interface. The search bar contains the term 'quantum' and has a search button. The navigation menu includes 'Search', 'Browse', 'Suggest', 'Resources', and 'Contact'. The DataCite logo is in the top right. On the left, there is a 'Filter' sidebar with various categories like 'Subjects', 'Content Types', 'Countries', etc. The main content area shows search results for 'quantum'. The first result is 'Stark Broadening Parameters for Neutral and Singly Charged Ions' with a French subtitle 'Paramètres d'élargissement Stark des éléments neutres et une fois ionisés'. It lists subjects like 'Optics, Quantum Optics and Physics of Atoms, Molecules and Plasmas' and 'Natural Sciences'. The second result is 'NIST Energy Levels of Hydrogen and Deuterium' with the subtitle 'NIST Standard Reference Database 142'. It lists subjects like 'Optics, Quantum Optics and Physics of Atoms, Molecules and Plasmas' and 'Natural Sciences'. The page indicates 'Found 84 result(s)' and has a pagination control showing page 1 of 4.

[Nakala](#) (by Huma-Num), the data repository for humanities



En ▾

Login



Share, publish and enhance your scientific data

Upload my data in nakala

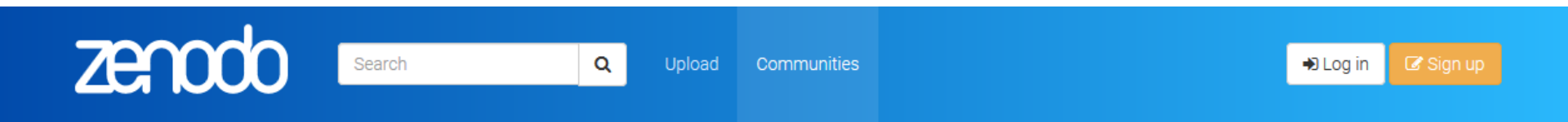
Search, cite and reuse scientific data

search in NAKALA...



Zenodo

The disciplin-agnostic european repository [Zenodo](https://zenodo.org).



Université de Strasbourg

Recent uploads

May 24, 2019 (1.0.0) **Dataset** **Open Access**

QSPR models for bioconcentration factor (BCF): Are they able to predict data of industrial interest?

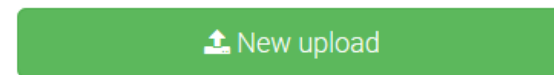
Lunghini, Filippo; Marcou, Gilles; Azam, Philippe; Patoux, Remi; Enrici, Marie-Hélène; Bonachera, Fanny; Horvath, Dragos; Varnek, Alexandre;

This dataset is described and studied in the article "QSPR models for bioconcentration factor (BCF): Are they able to predict data of industrial interest?" published in SAR and QSAR Environmental Research (Taylor&Francis). Files description: SI_BCFtrainset.xlsx: a collection

Uploaded on October 1, 2019

July 10, 2019 (1.0) **Dataset** **Open Access**

Consensus models to predict oral rat acute toxicity and validation on a dataset coming from the industrial context



Community


The logo of the University of Strasbourg, featuring the text 'Université de Strasbourg' in a stylized, blocky font within a white rectangular frame.

Université de Strasbourg

Curated by:
DataLibrarian

Curation policy:

5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?

B *I*    

Enregistrer



[FACILE - Service de validation de formats](#) (CINES)


Recommandations

Commentaires

ANR

- Indicate what data must be retained or destroyed for contractual, legal, or regulatory purposes.
- Indicate how it will be decided what data to keep. Describe the data to be preserved long-term.
- Explain the foreseeable research uses (and/or users) for the data.
- Indicate where the data will be deposited. If no established repository is proposed, demonstrate in the data management plan that the data can be curated effectively beyond the lifetime of the grant. It is recommended to demonstrate that the repositories policies and procedures (including any metadata standards, and costs involved) have been checked.

5c. What methods or software tools are needed to access and use data?

B *I*    


Enregistrer

Recommandations Commentaires

ANR

- Indicate whether potential users need specific tools to access and (re-)use the data. Consider the sustainability of software needed for accessing the data.
- Indicate whether data will be shared via a repository, requests handled directly, or whether another mechanism will be used?

5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

B *I*    

Enregistrer

3-2-1 rule :

- **3** copies
- **2** different media
- **1** backup off-site

<https://www.networkworld.com/article/3527303/for-secure-data-backup-here-s-how-to-do-the-3-2-1-rule-right.html>



Recommandations Commentaires

ANR

- Explain how the data might be re-used in other contexts. Persistent identifiers should be applied so that data can be reliably and efficiently located and referred to. Persistent identifiers also help to track citations and re-use.
- Indicate whether a persistent identifier for the data will be pursued. Typically, a trustworthy, long-term repository will provide a persistent identifier.



6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

B *I*    


Enregistrer

Recommandations Commentaires

ANR

- Outline the roles and responsibilities for data management/stewardship activities for example data capture, metadata production, data quality, storage and backup, data archiving, and data sharing. Name responsible individual(s) where possible.
- For collaborative projects, explain the co-ordination of data management responsibilities across partners.
- Indicate who is responsible for implementing the DMP, and for ensuring it is reviewed and, if necessary, revised.
- Consider regular updates of the DMP.

6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

B *I*    

Enregistrer



RDM costing tool :

→ [Costing data management](#)

→ [Data Management Costing Tool](#) (Delft)

Recommandations

Commentaires

ANR

- Explain how the necessary resources (for example time) to prepare the data for sharing/preservation (data curation) have been costed in. Carefully consider and justify any resources needed to deliver the data. These may include storage costs, hardware, staff time, costs of preparing data for deposit, and repository charges.
- Indicate whether additional resources will be needed to prepare data for deposit or to meet any charges from data repositories. If yes, explain how much is needed and how such costs will be covered

Part 4

More information

Guidance from funders



- The [ANR Open Science web page](#)
- The [Horizon Europe Programme Guide](#) [PDF] page 38.
- [The Annotated Model Grant Agreement \(AGA\)](#), annex 5, starting p. 152

Useful guidance



Guidance produced by the WG
Data from Couperin:

- [Improving your ANR project thanks to Open Science, 2020](#)
- [Manage research projects with Open Science: Develop specific guidance, training and support for project coordinators and consortia, 2020 \[SLIDES\]](#)

Legal guidance



- [guide “Ouverture des données de recherche. Guide d’analyse du cadre juridique en France” \(v.2, 2017\) \[PDF\]](#)
- [Guide de la Recherche « Les SHS et la protection des données à caractère personnel dans le contexte de la science ouverte » \[PDF\]](#)
- **Unistra DPO** : dpo@unistra.fr

Documentation for writing a

DMP



- [Science Europe Guidance. Presenting a Framework for Discipline-specific Research Data Management.](#) [PDF]
- [DCC Resources for digital curators](#)
- [Dorandum](#)
- Nathalie Reymonet, Magalie Moysan, Aurore Cartier, Renaud Délémontez. Réaliser un plan de gestion de données « FAIR » : modèle . 2018. [{sic_01690547v2}](#)
- [Datapartage](#) (INRA)
- [Metadata Standards Catalog](#)

Publics DMPs



- DMP Opidor : [public DMPs](#)
- [841 DMP H2020](#)
- Portage : [Exemples de PGD](#)
- DCC : [Example DMPs and guidance](#)
- DMPTools: [Public DMPs](#)
- [LIBER DMP Catalogue](#)

Assessment tools



Self-assessment tool:

→ [FAIR data self-assessment tool](#)
(ARDC)

DMP assessment tools:

→ [Grille de relecture du DMP ANR](#)

→ [Wellcome Trust](#)

→ [EPSRC](#)

RDM cost assessment tools:

→ [Costing data management](#) (UK Data Service)

→ [Costs of data management](#) (Utrecht)

Thank you!



Service des **bibliothèques**

Université de Strasbourg

[Scienceouverte.unistra.fr](https://scienceouverte.unistra.fr)

Stéphanie Cheviron

Research Data Librarian

+33 (0)3 68 85 07 93

scheviron@unistra.fr