DMP in European Research: from requirements to practical value. NAPMIX Workshop

Marjolaine Bodin
ISDD/Software Group/Data Automation Unit, ESRF

marjolaine.bodin@esrf.fr

OUTLINE

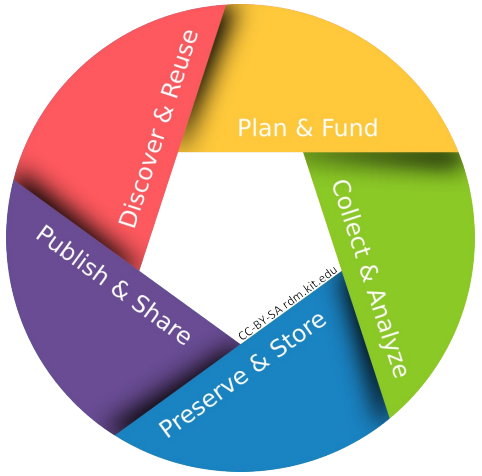
- Why Data Management Plans matter for science
- European Open Science & policy requirements
- What good data management looks like (principles & practices)
- How DMPs are created (templates, tools, support)
- From static DMPs to living / machine-actionable approaches

WHAT IS A DATA MANAGEMENT PLAN (DMP)?

A Data Management Plan (DMP) is a document describing how research data is managed across the project lifecycle.

It covers:

- data creation
- storage
- documentation
- sharing
- preservation



RDM kit
https://www.rdm.kit.edu/english/researchdata_rdm_cycle.php

A DMP turns research data into a well-managed, reusable long-term asset.

WHY DO DMPs MATTER?

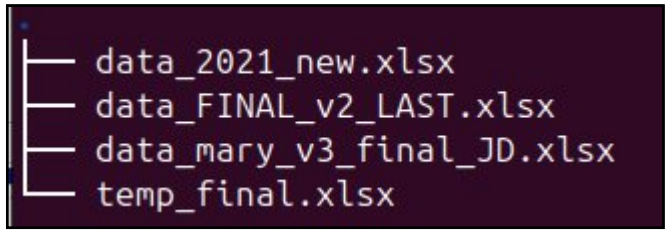
Without a DMP

- Scattered files across laptops, drives, emails
- Missing documentation: data hard to understand
- Risk of data loss or duplication
- Difficult collaboration
- Data becomes unusable after the project

With a DMP

- Organized and structured data
- Clear documentation and metadata
- Secure storage and backup
- Easier collaboration within teams
- Data remains reusable and valuable

Which one would you trust two years after the project?



The difference is not the data, it's how we manage it.

WHAT DOES A DMP COVER?

Data and documentation
- data description (formats, volume, types)
- metadata and standards

Storage & protection
- storage, backup, access control
- legal and ethical aspects

Sharing and reuse
- data sharing
- external access conditions
- long-term preservation

Responsibilities
- roles and responsabilities

It outlines data management practices to ensure alignment with the FAIR principles (Findable, Accessible, Interoperable, Reusable).

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

THE EVOLUTION OF DMPS

Traditional DMP (static)

- Written at proposal stage
- PDF / Word document
- Updated occasionally (if at all)
- Mainly for compliance (funder)

Problem: quickly becomes outdated and disconnected from practice

Active DMP

- Continuously updated during the project (e.g. add related publication)
- Integrated into research workflows
- Used as a living planning and operational tool, supporting continuously updated datasets in research infrastructures (e.g. marine sensors, astronomical observations)

Benefit: supports decision-making during the project, not just reporting

Machine Actionnable DMP (maDMP)

- Structured and machine-readable (e.g. JSON, metadata models)
- Can be integrated with tools and services
- Enables automation:
  - storage provisioning
  - metadata generation
  - repository deposition

Benefit: connects DMPs to research infrastructures and services

DMPs are no longer just document: they are becoming integrated into research processes.

Miksa et al. (2019), Ten principles for machine-actionable data management plans, PLOS Computational Biology. https://doi.org/10.1371/journal.pcbi.1006750

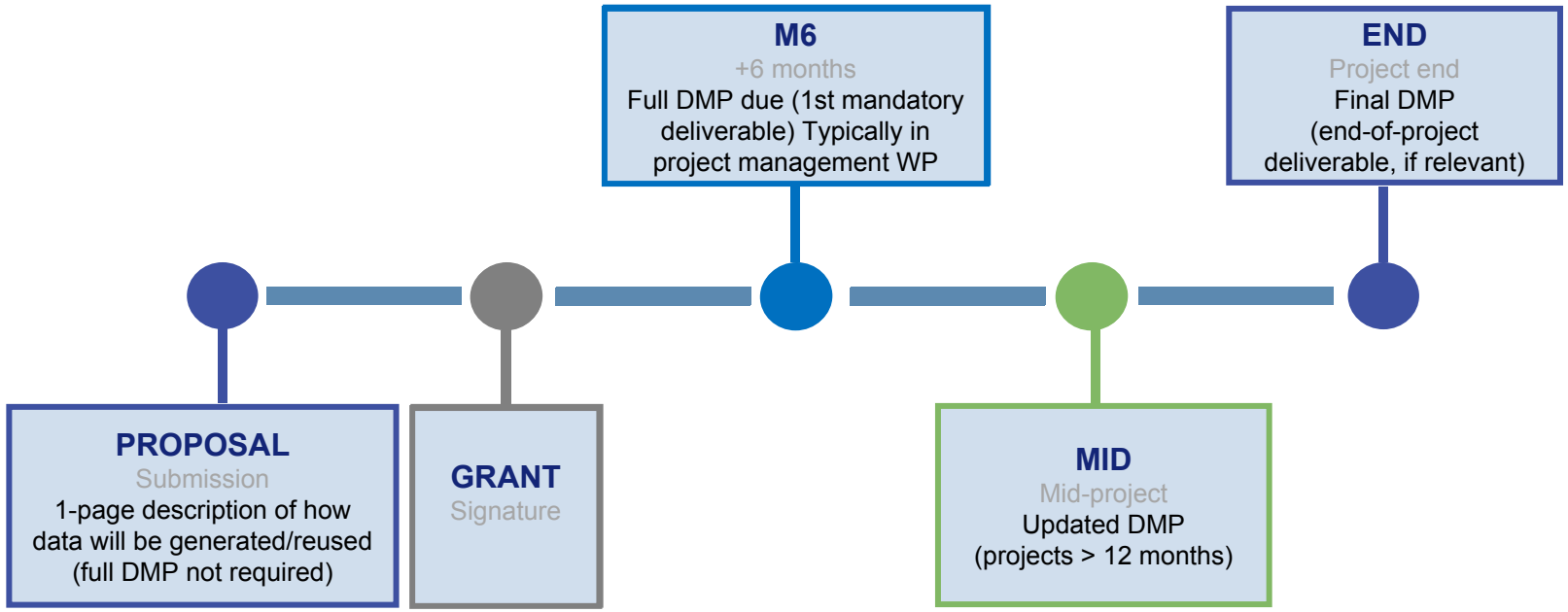
DMPS IN EU-FUNDED PROJECTS

DMPs are mandatory in most EU-funded projects (e.g. Horizon Europe)

The proposal must include a short (maximum one-page) description of how data and other research outputs will be generated and/or reused as part of the project methodology. A full DMP is not required at submission stage, unless explicitly required by the work programme or call.

A DMP is a mandatory project deliverable, typically due within 6 months of grant signature. For projects exceeding 12 months, an updated version is required at mid-project.
A final version is included among the end-of-project deliverables, where relevant.

It is typically included in the project management work package.



DMP IN OTHER RESEARCH SYSTEMS

- US: mandatory in NSF / NIH grants
- UK: integrated into national funding policies
- Global trend: increasing focus on data sharing and Open Science practices

In the US, agencies like NSF and NIH have required DMPs for many years. What is changing globally is not the requirement itself, but the nature of DMPs: moving toward structured and machine-actionable approaches.

https://upstream.force11.org/what-is-the-future-of-data-management-plans/

What about FP10, the next EU programme, 2028-2034?
maDMPs are likely to replace static documents by enabling automated, continuous and interoperable data management across systems.

DMPs are not specific to Europe, they are becoming a global standard in research funding. Within the broader open science movement, the sharing of publicly funded research data is becoming standard practice.

OPEN SCIENCE: PRINCIPLES AND CHALLENGES

The sharing of research data is required in line with the principle

“as open as possible, as closed as necessary.”

Some categories of data may be excluded from open sharing due to legal, ethical, or practical constraints, such as personal data, sensitive environmental or geospatial data, data still under validation.

The challenges are multiple:

Facilitating research work

- Ensure effective use of data throughout the project
- Enable medium-term data preservation
- Allow reproducibility of scientific results, even by the original research team
- Facilitate reuse of data for new research
- Increase visibility by valuing not only results, but also data and code
- Foster new collaborations and research approaches

Societal challenges

- Ensure sovereignty over produced research data
- Safeguard scientific integrity
- Guarantee transparency and strengthen public trust in research

DATA MANAGEMENT ENABLING OPEN SCIENCE

At different scales, data management enables:

At the team or laboratory level

- Enable appropriate archiving of research data
- Structure data to prevent loss of information (containers + metadata)
- Ensure effective transfer of knowledge within the lab (between temporary and permanent staff, and vice versa)

At the community level

- Enable data exchange within the scientific community
- Share methodologies and workflows
- Build reference datasets
- Support training of machine learning / neural networks

At a broader scale

- Meet legal and regulatory requirements
- Bridge different scientific communities

Example in the photon and neutron (PaN) community shows concrete adoption of FAIR data principles and Open Science workflows.

Murphy, B. M., Gotz, A., Gutt, C., McGuinness, C., Ronnow, H. M., Schneidewind, A., Deledda, S. & Pietsch, U. (2025). IUCrJ, 12, 8-15. https://doi.org/10.1107/S2052252524011941

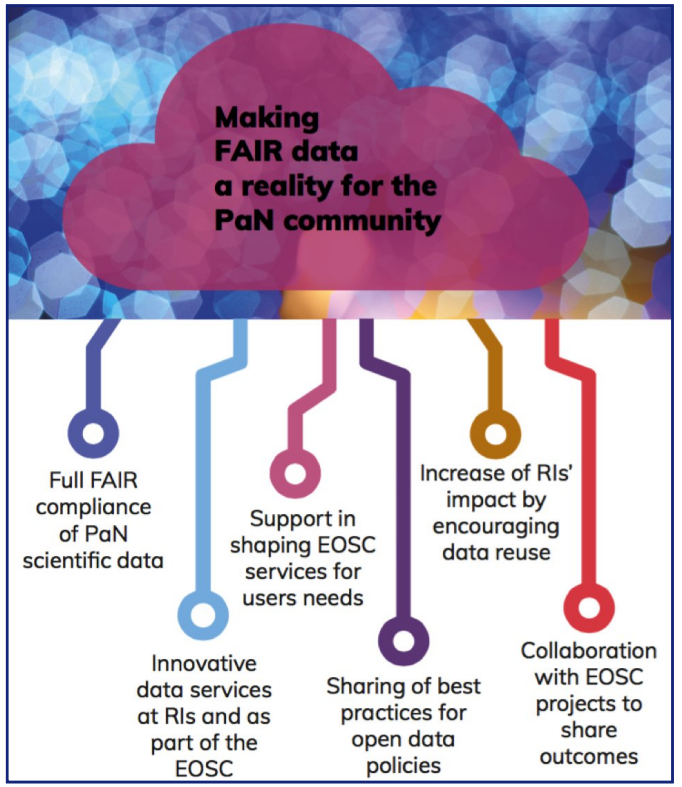
PANOSC: COMMUNITY IMPLEMENTATION

The Photon and Neutron Open Science Cluster (PaNOSC)

- PaNOSC supports Open Science in the photon and neutron community
- Development of FAIR data policies and services
- Harmonisation across European research infrastructures
- Support for data management practices and metadata standards
- Contribution to the European Open Science Cloud ecosystem

What PaN communities share

- Metadata standards e.g. NeXuS and domain based
- Facility and domain based data repositories e.g. SciCat + ICAT
- Ontologies – experimental techniques + domain e.g. PaNET
- Common AAI for SSO – UmbrellaID (moving to myAccessID)
- VRE platform for remote data analysis e.g. VISA
- Software packaging and distribution e.g. CVFMS
- Workflows for data processing e.g. Ewoks
- Visualisation of HDF5 files e.g. H5Web, silx view
- Experience & know-how on RDM, RSE, FAIR, …



Credits: **Andy Götz** (ESRF)

PaNOSC shows how Open Science is implemented within a specific scientific domain. At a broader level, these efforts are part of the European Open Science Cloud.

EOSC: THE EUROPEAN ECOSYSTEM

EOSC: European ecosystem for FAIR and interoperable research data

- European Open Science Cloud connects data, services, and research infrastructures across Europe
- Enables FAIR, interoperable, and reusable research data
- Builds a federated environment (not a single platform)
- Supports cross-domain data sharing and reuse
- Provides the context for initiatives like PaNOSC and OSTrails

OSTRAILS: TOWARDS ACTIONABLE AND INTEROPERABLE DMPs

OSTrails European project (2024-2027) Open Science Plan-Track-Assess Pathways

Involving 38 partners and 24 pilots across 15 national and 5 disciplinary research ecosystems
Coordinated by OpenAIRE, with 5 ESFRI Clusters and 22 research performing organisations

- **FAIRification of DMPs: moving beyond static documents toward machine-actionable, findable, and interoperable plans integrated into scientific knowledge graphs.**
- **Development of maDMPs to improve automation, reuse, and traceability of research outputs across the research lifecycle.**
- **Integration of FAIR assessment mechanisms for DMPs within tools and workflows (DMP FAIR evaluation in OSTrails)**

Supporting eosc

This project has received funding from the European Union's Horizon Europe framework programme under grant agreement No. 101130187. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the European Research Executive Agency can be held responsible for them.

ostrails

EXAMPLE MACHINE ACTIONABLE DMP


```
{
  "title": "DMP for Synchrotron X-ray Diffraction Study of Battery Materials",
  "created": "2026-01-10",
  "modified": "2026-03-05",
  "language": "eng",
  "contact": {
    "name": "Marjolaine Bodin",
    "mbox": "marjolaine.bodin@esrf.fr",
    "affiliation": {
      "name": "European Synchrotron Radiation Facility (ESRF)",
      "id": {
        "type": "ror",
        "value": "https://ror.org/02550n020"
      }
    }
  },
  "project": {
    "title": "Operando study of solid-state battery degradation mechanisms",
    "start": "2026-01-01",
    "end": "2026-01-08",
    "facility": {
      "name": "European Synchrotron Radiation Facility (ESRF)",
      "id": {
        "type": "ror",
        "value": "https://ror.org/02550n020"
      }
    },
    "beamtime": [
      {
        "beamline": {
          "name": "ID31",
          "id": "https://doi.esrf.fr/10.15151/ESRF-INST-AB12"
        },
        "experiment_type": "high-resolution powder diffraction",
        "proposal_id": "HC-1234"
      }
    ],
    "funding": [
      {
        "funder": {
          "name": "European Synchrotron Radiation Facility (ESRF)",
          "id": {
            "type": "ror",
            "value": "https://ror.org/02550n020"
          }
        },
        "grant_id": {
          "type": "proposal",
          "value": "ID31-2026-HC1234"
        }
      }
    ]
  },
  "datasets": [
    {
      "title": "Raw diffraction images",
      "type": "raw",
      "format": "HDF5",
      "personal_data": "no",
      "sensitive_data": "no",
      "distribution": [
        {
          "title": "ESRF Data Portal",
          "access_url": "https://data.esrf.fr/dataset/xxxx",
          "access": "embargoed",
          "host": {
            "name": "ESRF Data Portal",
            "id": {
              "type": "url",
              "value": "https://data.esrf.fr"
            }
          }
        }
      ]
    }
  ]
}
```


Export a DMP to machine-actionable format

OSTrails has published a formal mapping between the Science Europe DMP template and the RDA DMP Common Standard

https://docs.ostrails.eu/en/latest/commons/dmp/mappings/science-europe.html

"*Who is responsible for data management?*" maps to **contributor/role** in JSON

DMP TEMPLATES

A variety of DMP templates already exist, developed by different funding agencies and institutions. In parallel, initiatives are underway to create templates for machine-actionable DMPs.

- Each funding agency may require or recommend a specific DMP template.
- Similarly, institutions can define their own required or recommended templates.
- These templates may take the form of a text-based questionnaire or a machine-actionable format.

Templates

- **European Commission** / Horizon Europe template: Standard structure used in most EU-funded projects
- **Science Europe**: Defines a common structure for DMPs
- the **DMP Common Standard** model from the Research Data Alliance can be adopted (machine-actionable DMP template)

The European Commission recommends the use of this template, but its use is optional.

There is no single template, but a common structure is widely shared across Europe.

THINK ABOUT YOUR DMP EARLY

Don’t wait until the project starts to think about your DMP.
Plan for data management already during the proposal phase to anticipate key impacts on your project.

What to consider early?

Budget
- data management costs
- storage and computing resources

Legal and ethical
- personal or sensitive data
- GDPR and compliance

Data governance
- data ownership
- access and sharing rights
- consortium agreement

Early data management planning helps avoid technical, legal, and financial issues later.

COLLABORATIVE DMP

A DMP is a collaborative process

- Involve all project partners from the start
- Clarify roles and responsibilities early
- Define common data management rules
- Improve coordination across the consortium

Ensure the DMP is understandable for everyone. Keep the DMP clear and concise, avoiding long sentences and vague or generic phrasing.

Each project partner should ideally assign a data management contact, while one person should oversee the DMP to ensure it remains a coherent, unified document.

A DMP is most effective when built collaboratively.

DMP TOOLS

Collaborative / ecosystem platforms

- **Argos**: *(SaaS / hosted platform, community-driven deployment)* a collaborative initiative by OpenAIRE and EUDAT providing an open platform for Data Management Planning.
- **Opidor**: *(institutionally hosted service)* French institutional DMP support platform.

FAIR-oriented / questionnaire-based tools

- **Data Stewardship Wizard** (DSW): *(SaaS + on-premise possible, highly configurable deployments)* an open-source tool that enables collaborative creation of data management plans through intelligent, customizable questionnaires, including FAIRness assessment.

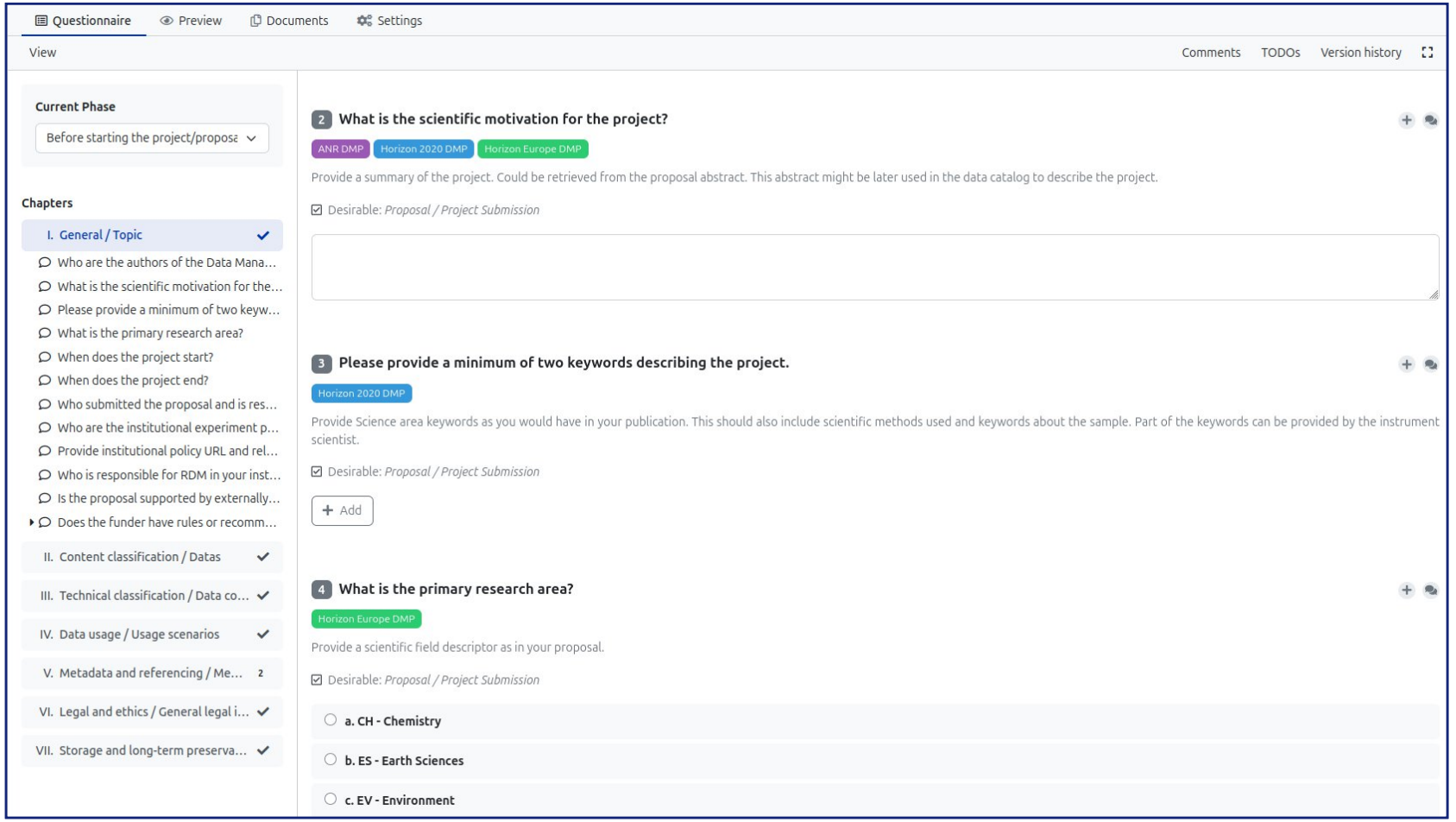
Machine-actionable / workflow-integrated tools

- **DAMAP**: *(open-source, typically on-premise / institutional deployment)* a tool designed to produce machine-actionable Data Management Plans.

Institutional / widely adopted platforms

- **DMPonline**: *(SaaS or institutional hosting)* a widely adopted tool across Europe, often provided by universities and research institutes for their researchers.
- **DMPTool**: *(SaaS / hosted by institutions)* broadly used in US academic environments.

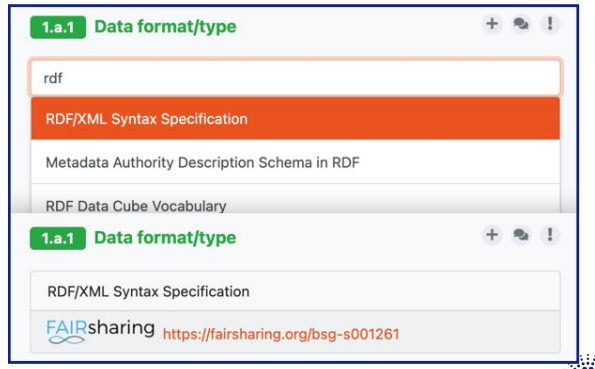
DATA STEWARDSHIP WIZARD



https://ds-wizard.org/

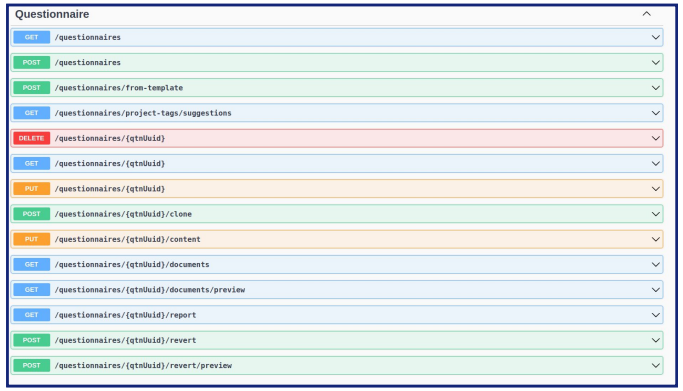
Smart questionnaire

- Several types of questions (options, list of items, open questions, followup questions)
- Specify guidance information
- Project phases
- Tags
- Can be linked to external sources enable fast completion and automated FAIRification.

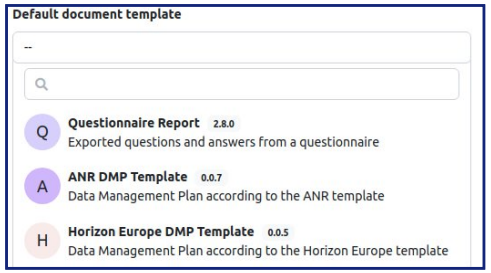


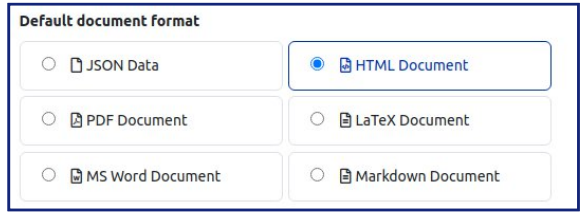
Pergl, R., Hooft, R., Suchánek, M., Knaisl, V., & Slifka, J. (2019). "Data Stewardship Wizard": A Tool Bringing Together Researchers, Data Stewards, and Data Experts around Data Management Planning. Data Science Journal, 18. https://doi.org/10.5334/dsj-2019-059

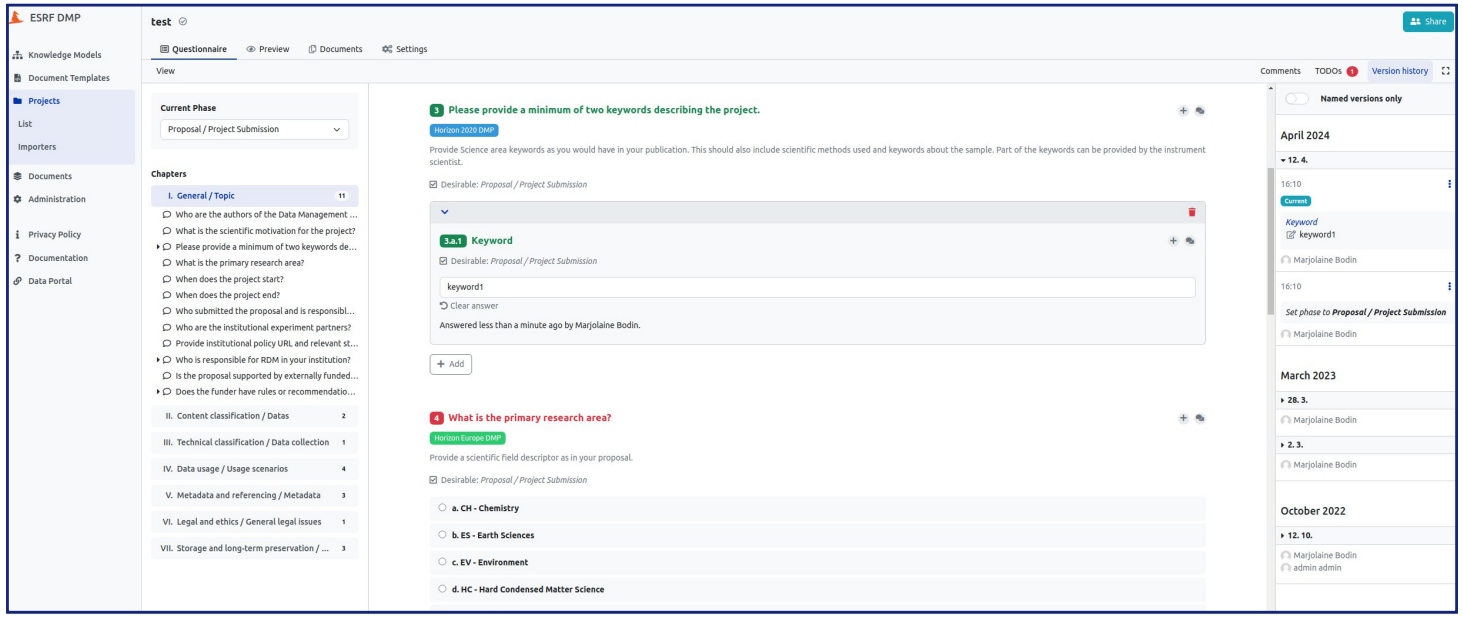
Rest API which can be used for integration purposes



Export using different templates (funder-specific), different formats (Word, PDF, JSON)



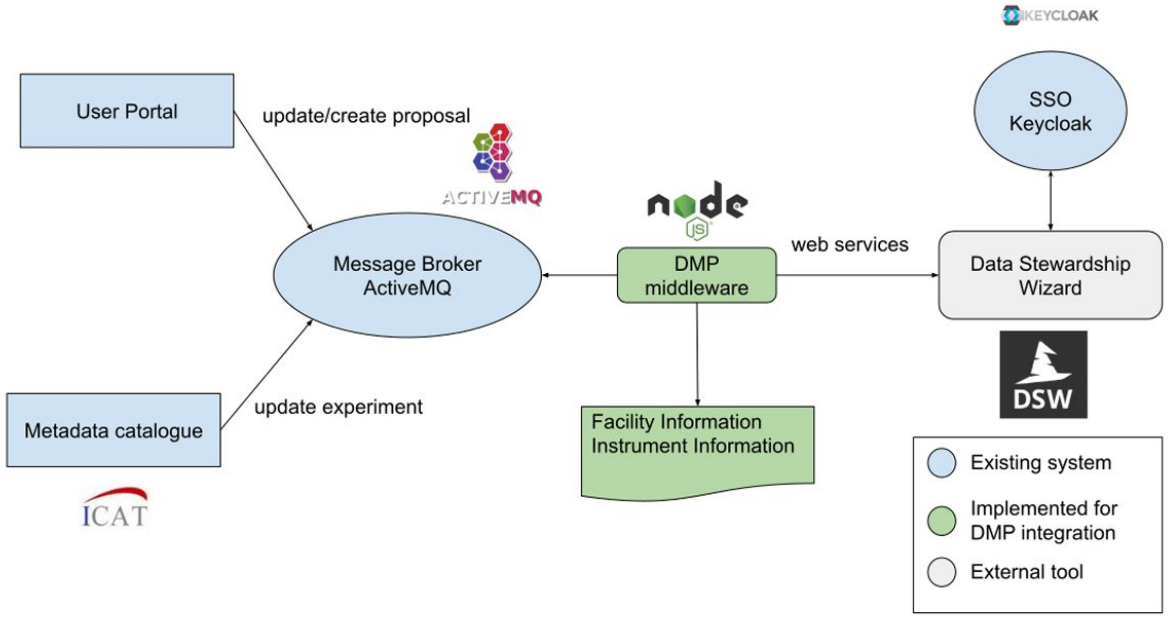




The questionnaire is filled by the user

- Access rights management and project sharing
- Online collaboration
- Version history

DSW@ESRF



Architecture and DMP data flow at the ESRF

https://dmp.esrf.fr/

- **DMP automatically created 6 weeks before the first experiment.**
- **Not mandatory**
- **Excluding industrial proposals**
- **Common PaN knowledge model (questionnaire)**
- **Based on information from the User Portal + static information (data policy, IT)**
- **Updated with information from our ICAT metadata catalog**

=> 60% questionnaire filled automatically

- **Develop 2 export templates** *(ANR, Horizon Europe)*

Bodin, M., Bolmsten, F., Aulin, P., Ivănoaica, T., Olivo, A., Malka, J., Wrona, K. and Götz, A., 2023. Data Management Plans for the Photon and Neutron Communities. *Data Science Journal*, 22(1), p.30.DOI: https://doi.org/10.5334/dsj-2023-030

EXAMPLE HORIZON EUROPE DMP

6 sections

1. Data Summary
2. FAIR Data
3. Other research outputs
4. Allocation of resources
5. Data Security
6. Ethics

Project overview

- Short description including:
  - a technical/specific section
  - a simpler section for non-specialists
- Project timeline
- Participants (and affiliations):
  - contacts: project lead, DMP author
  - project members
  - partners

There are no incorrect answers; simply describe what is carried out within the project.

EXAMPLE HORIZON EUROPE DMP: DATA SUMMARY

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?

- **Short description of each data type:**
  - **whether it already exists or not, and conditions of use**
  - **originating from or linked to specific equipment or software**
  - **constraints related to handling or reuse**
  - **brief description of the planned processing**

- **Estimated volume: an order of magnitude is sufficient**

- **Formats used: commonly used in the domain, preferably open, non-proprietary, and cross-platform**

- **Presence of sensitive data?**
  - **security requirements**
  - **involvement of the Data Protection Officer (DPO)**

EXAMPLE HORIZON EUROPE DMP: FAIR DATA

FAIR DATA

1. Making data findable, including provisions for metadata
2. Making data accessible
   - Repository
   - Data
   - Metadata
3. Making data interoperable
4. Increase data re-use

- **Findable**
  - **Unique and persistent identifier: Examples: DOI (data), ORCID (researchers), ROR (institutions), PIDINST (instruments)**
  - **described by rich metadata**
  - **stored or indexed in a searchable system/database**
  - **Use a clear, consistent folder structure** (project / experiment / sample-run)
  - **Separate raw and processed data**
  - **Use meaningful, stable file names**
  - **Ensure structure is understandable without personal knowledge**
- **Accessible**
  - **Trusted, certified, long-term repository. Example CoreTrustSeal certification**
  - **Accessible via standard protocols**
  - **In case of embargo or restricted access: metadata remain accessible**

EXAMPLE HORIZON EUROPE DMP: FAIR DATA

FAIR DATA

1. Making data findable, including provisions for metadata
2. Making data accessible
    - Repository
    - Data
    - Metadata
3. Making data interoperable
4. Increase data re-use

- **Interoperable**
    - **File format: open standard**
    - **Metadata: Use domain standards when available to improve interoperability: ontologies, controlled vocabularies**
    - **Availability of the source code where processing software exists**
- **Reusable: ready for reuse in future research, including through computational method**
    - **Provide clear and detailed documentation: experimental conditions, sample details, instruments, and software versions**
    - **Specify a clear usage license**

EXAMPLE HORIZON EUROPE DMP: OTHER RESEARCH OUTPUTS

OTHER RESEARCH OUTPUTS

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).
Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.?
To whom might your data be useful ('data utility'), outside your project?

Management of

- **Digital outputs: software, code, workflows, models, protocols**
- **Physical outputs: samples, materials, antibodies, etc.**

- **Apply FAIR principles where relevant (Findable, Accessible, Interoperable, Reusable)**
- **Document outputs with clear metadata and documentation**
- **Use versioning and persistent identifiers when possible**
- **Define access conditions and licenses**
- **Ensure outputs are well-described for future reuse (human + computational)**
- **Use community standards when available**

EXAMPLE HORIZON EUROPE DMP: ALLOCATION OF RESOURCES

ALLOCATION OF RESOURCES

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?
How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)
Who will be responsible for data management in your project?
How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

- **Costs include data storage, curation, metadata preparation, software/tools, security, and long-term archiving**
- **Data management costs are eligible under the Horizon Europe grant and integrated into project work packages**
- **A Data Manager / WP leader coordinates data management activities; each partner is responsible for their own data**
- **Data are preserved in trusted repositories, selected based on scientific value, reuse potential, and legal constraints**
- **Long-term preservation and access conditions are defined at consortium level in line with project and institutional policies**

EXAMPLE HORIZON EUROPE DMP: DATA SECURITY

DATA SECURITY

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?
Will the data be safely stored in trusted repositories for long term preservation and curation?

- Storage locations: encrypted hard drives, laboratory servers, institutional infrastructure, etc.
- Backup methods: manual and/or automatic backups
- Security measures: anonymisation processes, encryption, use of institutional infrastructures, etc.
- Access management: definition of user rights, controlled access, and secure data sharing and communication channels
- Key consideration: No use of commercial cloud services

EXAMPLE HORIZON EUROPE DMP: ETHICS

ETHICS

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).
Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

- **Compliance with ethical principles and legal frameworks (e.g. GDPR)**
- **Identification of any ethical or legal issues affecting data sharing and reuse**
- **Description of how ethics requirements from the project review are addressed**
- **Use of informed consent when personal data are involved (if applicable)**
- **Consideration of responsible and transparent use of AI methods (if relevant)**

CONCLUSION

A DMP is both a requirement and a scientific asset.

Beyond compliance

Improves data quality and reuse, not just a funder checkbox.

Start early

Early planning avoids rework and aligns storage, roles and responsibilities.

Living document

Evolves with the project, involves researchers, data stewards, and IT.

Consortium alignment

Clarifies data ownership and access rights, preventing conflicts between partners.

Better science

Improves reproducibility, citation of datasets and long-term data legacy.

Start now

- ✓ Open DSW or your DMP tool
- ✓ Identify your data steward
- ✓ Draft your data section before submission

Tools help, but scientific decisions remain yours.

"Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation."

Wilkinson et al., Scientific Data, 2016

ACKNOWLEDGEMENTS

We thank Andy Götz (Head of DAU) and Nils Blanc (Scientist BM02) for their input and discussions.

This project has received funding from the European Union's Horizon Europe framework programme under grant agreement No. 101130187.

ostrails

https://ostrails.eu/

ESRF
The European Synchrotron

THANKS FOR YOUR ATTENTION