

Data Management Plans in Horizon Europe Projects

Requirements and Solutions

Andre M. Pietsch (JLU Gießen)  <https://orcid.org/0000-0003-3757-9557>



A Short Introduction

Background: computer linguistic and text technology

Research data consultant since march 2020

Part of the Hessian research data infrastructure project *HeFDI*

Contact

Mail: forschungsdaten@uni-giessen.de

Phone: 0641/99-14013

Stay updated on research data management (RDM)

Local mailing list

<https://lists.uni-giessen.de/sympa/subscribe/fdm>

HeFDI newsletter

<https://www.uni-marburg.de/en/hefdi/data-information-consulting/hefdi-data-news>



Agenda

RDM Requirements in Horizon Europe Projects

What to consider in terms of
RDM within a Horizon Europe
project?

FAIR Principles and the FAIRification Framework

What are the FAIR Principles
specifically about?

What is the FAIRification
Framework?

Horizon Europe DMP Template

How to work with the Horizon
Europe DMP template?

How does a good practice
example look like?



Requirements on RDM in Horizon Europe Projects

What are the requirements?

Proper Research Data Management (RDM) is **mandatory** for any Horizon Europe project generating or reusing research data. It is a key part of Horizon Europe's open science requirements.

In Horizon Europe, *beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the **FAIR principles*** and should at least do the following:

- Prepare a **Data Management Plan (DMP)** and keep it updated throughout the course of the project
- Deposit data in a **trusted repository** and provide **open access** to it ('as open as possible, as closed as necessary')
- **Provide information** (via the same repository) about any research output or any other tools and instruments needed to re-use or validate the data

Keep in mind that 'research data' is a very broad concept and certainly not limited to numerical/tabular data.

<https://www.openaire.eu/how-to-comply-with-horizon-europe-mandate-for-rdm>



FAIR Principles

and the FAIRification Framework



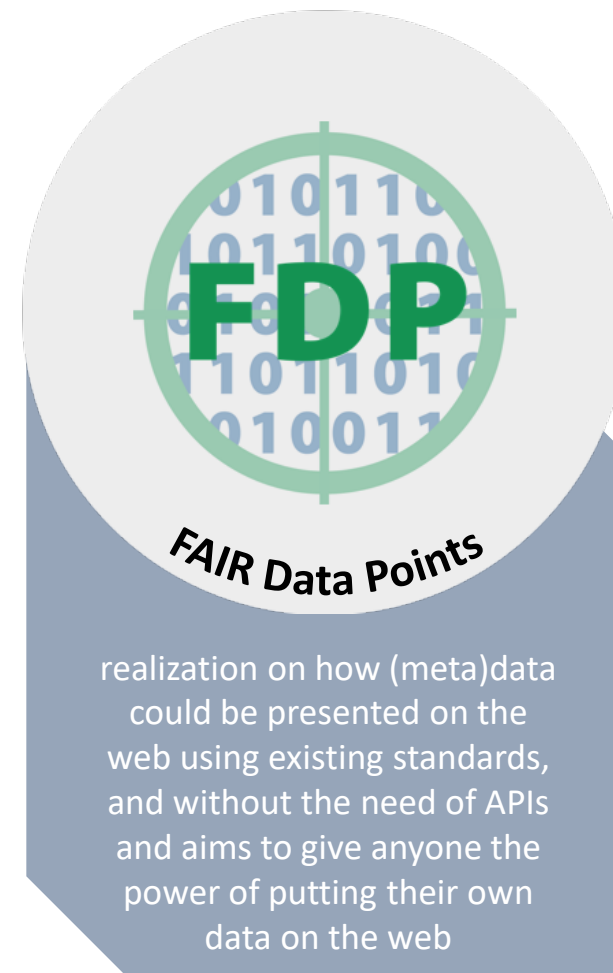
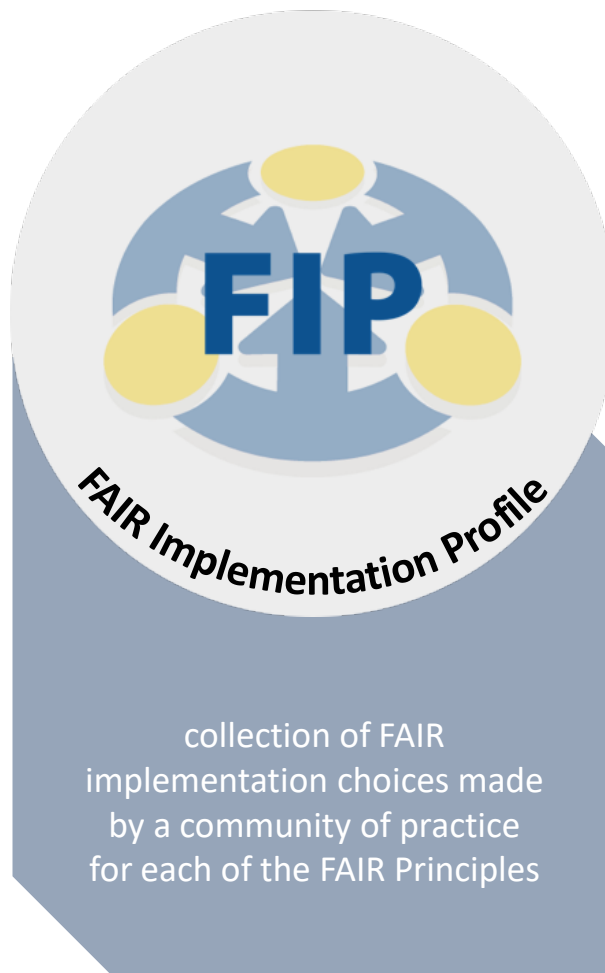
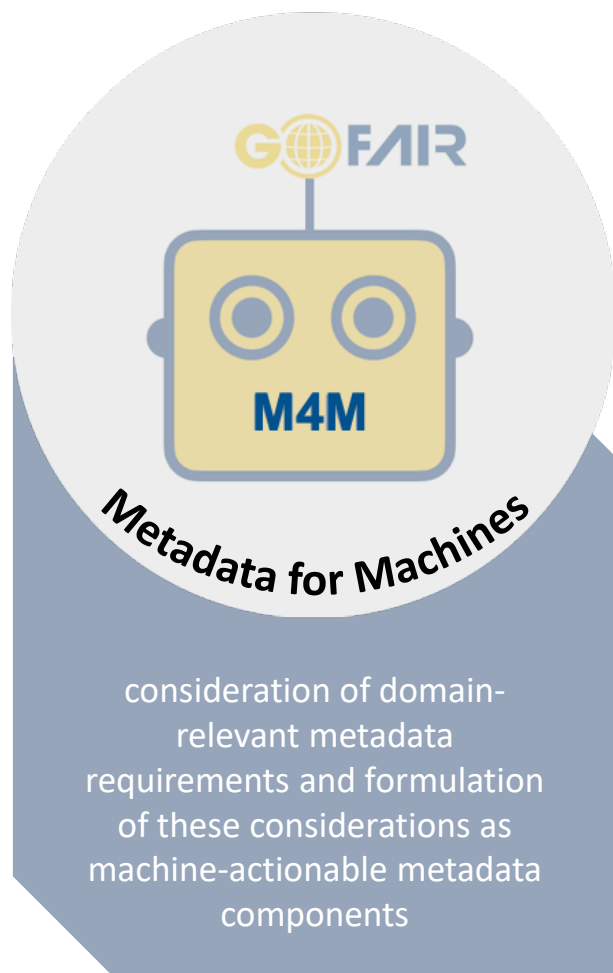
In 2016, the ‘**FAIR Guiding Principles for scientific data management and stewardship**’ were published in *Scientific Data*. The authors intended to provide guidelines to improve the **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse of digital assets. The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.

A practical “how to” guidance to go FAIR can be found in the **Three-point FAIRification Framework**.

<https://www.go-fair.org/fair-principles/>



Three-point FAIRification Framework



Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the **FAIRification process**.

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource

<https://www.go-fair.org/fair-principles/>



Accessible

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available



Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

12. (Meta)data use vocabularies that follow FAIR principles

13. (Meta)data include qualified references to other (meta)data



Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

R1. (Meta)data are richly described with 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 detailed provenance

R1.3. (Meta)data meet domain-relevant community standards



Data Management Plans (DMPs) Across the Project Life Cycle

First version

Once a project has had its funding approved and has started, you **must submit a first version of your DMP** (as a deliverable) within the first 6 months of the project. The Commission provides a DMP template in annex, the use of which is recommended but voluntary.

Updates

The DMP **needs to be updated over the course of the project** whenever significant changes arise, such as (but not limited to):

- new data
- changes in consortium policies (e.g. new innovation potential, decision to file for a patent)
- changes in consortium composition and external factors (e.g. new consortium members joining or old members leaving).

The DMP **should be updated as a minimum in time with the periodic evaluation/assessment of the project.**

- If there are no other periodic reviews foreseen within the grant agreement, then such an update needs to be made in time for the final review at the latest.
- Furthermore, the consortium can define a timetable for review in the DMP itself.

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



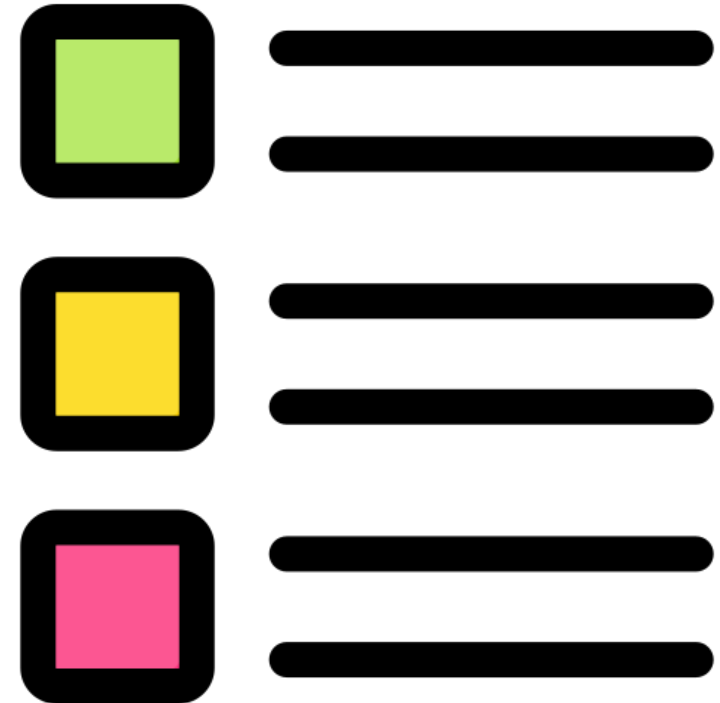
Horizon Europe DMP Template

and a [Good Practice-Example](https://doi.org/10.5281/zenodo.10670433) (<https://doi.org/10.5281/zenodo.10670433>)



Overview

1. Data Summary
2. FAIR Data
 - a. Making Data Findable, Including Provisions for Metadata
 - b. Making Data Accessible
 - c. Making Data Interoperable
 - d. Increase Data Re-use
3. Other Research Outputs
4. Allocation of Resources
5. Data Security
6. Ethics
7. Other Issues



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?



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**?



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?



Making Data Accessible: 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)?



Making Data Accessible: 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?

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?



Making Data Interoperable

What data and metadata **vocabularies, standards, formats or methodologies** will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you **provide mappings to more commonly used ontologies**? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

Will your data include **qualified references to other data** (e.g. other data from your project, or datasets from previous research)?



Increase Data Re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

Will the data produced in the project be useable by third parties, in particular after the end of the project?

Will the provenance of the data be thoroughly documented using the appropriate standards?

Describe all relevant data quality assurance processes.



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.



Allocations 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)?



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?



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?



Other Issues

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?



Summary



RDM
Requirements
in Horizon
Europe
Projects

- Manage data in line with the **FAIR Principles**
- Create a Data Management Plan (**DMP**) and keep it updated
- Deposit data in **trusted repositories as open as possible, as closed as necessary** and **as soon as possible**
- Provide **sufficient information/documentation** about everything that is needed to re-use or validate the data

FAIR Principles
and the
FAIRification
Framework

- Guidelines to improve the **Findability, Accessibility, Interoperability** and **Re-use** of digital assets
- Focus lies on **machine-actionability**
- Three-point **FAIRification Framework** gives a community of practice a „how to“ guidance for the FAIRification of data:
 - (1) Conducting **Metadata for Machines-Workshops** within a community of practice
 - (2) Creating a **FAIR Implementation Profile**
 - (3) Creating a **FAIR Data Point** that provides access to metadata in a manner that follows the FAIR Principles for data/metadata publishing

Horizon
Europe DMP
Template

- Focus on information to the **re-use of data** addressing the FAIR Principles
- Give information about all **related resources**
- Describe **data security measures**
- Describe any **ethical or legal issues** and how you solve them
- Name all **guidelines** you follow in terms of data management and the publication of data



Questions?

