H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020

European Commission

H2020 Programme

Guidelines on

FAIR Data Management in Horizon 2020

Version 3.0
26 July 2016

History of changes				
Version	**Date**	**Change**		**Page**
2.1	15.02.2016	▪	the guide was also published as part of the Online Manual with updated and simplified content	all
3.0	26.07.2016	▪	This version has been updated in the context of the extension of the Open Research Data Pilot and related data management issues	all
		▪	New DMP template included	6

1. Background – Extension of the Open Research Data Pilot in Horizon 2020

⚠️ Please note the distinction between open access to scientific peer-reviewed *publications* and open access to research *data*:

- **publications** – open access is an *obligation* in Horizon 2020.

- **data** – the Commission is running a flexible pilot which has been extended and is described below.

See also the guideline: Open access to publications and research data in Horizon 2020.

This document helps Horizon 2020 beneficiaries make their research data **findable, accessible, interoperable and reusable** (**FAIR**), to ensure it is soundly managed. Good research data management is not a goal in itself, but rather the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse.

Note that these guidelines do not apply to their full extent to actions funded by the ERC. For information and guidance concerning Open Access and the Open Research Data Pilot at the ERC, please read the Guidelines on the Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020.

The Commission is running a flexible pilot under Horizon 2020 called the Open Research Data Pilot (ORD pilot). The ORD pilot aims to improve and maximise access to and re-use of research data generated by Horizon 2020 projects and takes into account the need to balance openness and protection of scientific information, commercialisation and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservation questions.

In the 2014-16 work programmes, the ORD pilot included only selected areas of Horizon 2020. Under the revised version of the 2017 work programme, the Open Research Data pilot has been extended to cover all the thematic areas of Horizon 2020.

While open access to research data thereby becomes applicable by default in Horizon 2020, the Commission also recognises that there are good reasons to keep some or even all research data generated in a project closed.

The Commission therefore provides robust *opt-out* possibilities at any stage, that is

- during the application phase

- during the grant agreement preparation (GAP) phase and

- after the signature of the grant agreement.

The ORD pilot applies primarily to the data needed to validate the results presented in scientific publications. Other data can also be provided by the beneficiaries on a voluntary basis, as stated in their Data Management Plans. Costs associated with open access to research data, can be claimed as eligible costs of any Horizon 2020 grant.

Participation in the ORD pilot is not part of the evaluation of proposals. In other words, proposals are not evaluated more favourably because they are part of the pilot and are not penalised for opting out of the pilot.

For more on open access to research data, please also consult the H2020 [Annotated Model Grant Agreement](#).

> ⚠ Participating in the ORD Pilot does not necessarily mean opening up all your research data. Rather, the ORD pilot follows the principle "**as open as possible, as closed as necessary**" and focuses on encouraging sound data management as an essential part of research best practice.

2. Data Management Plan – general definition

Data Management Plans (DMPs) are a *key element* of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DMP should include information on:

- the handling of research data during and after the end of the project
- what data will be collected, processed and/or generated
- which methodology and standards will be applied
- whether data will be shared/made open access and
- how data will be curated and preserved (including after the end of the project).

A DMP is required for all projects participating in the extended ORD pilot, unless they opt out of the ORD pilot. However, projects that opt out are still encouraged to submit a DMP on a voluntary basis.

3. Proposal, submission & evaluation

> ⚠ Whether a proposed project participates in the ORD pilot or chooses to opt out does not affect the evaluation of that project. In other words, proposals will not be penalized for opting out of the extended ORD pilot.

Since participation in the ORD pilot is not an evaluation criterion, the proposal is **not** expected to contain a fully developed DMP. However, good research data management as such should be addressed under the impact criterion, as relevant to the project. Your application should address the following issues:

- What standards will be applied?
- How will data be exploited and/or shared/made accessible for verification and reuse? If data cannot be made available, why?
- How will data be curated and preserved?

Your policy should also:

- reflect the current state of consortium agreements on data management
- be consistent with exploitation and Intellectual Property Rights (IPR) requirements

You should also ensure resource and budgetary planning for data management and include a deliverable for an initial DMP at month 6 at the latest into your proposal.

4. Research data management plans during the project life cycle

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.

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.

Periodic reporting

For general information on periodic reporting please check the following sections of the online manual

- How to fill in reporting tables for publications, deliverables
- Process for continuous reporting in the grant management system.

5. Support

Reimbursement of costs

Costs related to open access to research data in Horizon 2020 are eligible for reimbursement during the duration of the project under the conditions defined in the H2020 Grant Agreement, in particular Article 6 and Article 6.2.D.3, but also other articles relevant for the cost category chosen.

Data Management Plan

A DMP **template** is provided in Annex I. While the Commission does not currently offer its own online tool for data management plans, beneficiaries can generate DMPs online, using tools that are compatible with the requirements set out in Annex 1 (see also section 7 of Annex I).

ANNEX 1

Horizon 2020 FAIR Data Management Plan (DMP) template

Version: 26 July 2016

Introduction

This Horizon 2020 FAIR DMP template has been designed to be applicable to any Horizon 2020 project that produces, collects or processes research data. You should develop a **single DMP for your project** to cover its overall approach. However, where there are specific issues for individual datasets (e.g. regarding openness), you should clearly spell this out.

FAIR data management

In general terms, your research data should be 'FAIR', that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation-solution.

This template is not intended as a strict technical implementation of the FAIR principles, it is rather inspired by FAIR as a general concept.

More information about FAIR:

[FAIR data principles (FORCE11 discussion forum)](#)

[FAIR principles (article in Nature)](#)

Structure of the template

The template is a **set of questions** that you should answer with a level of detail appropriate to the project.

It is not required to provide detailed answers to all the questions in the first version of the DMP that needs to be submitted by month 6 of the project. Rather, the DMP is intended to be *a living document* in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. Therefore, DMPs should have a clear version number and include a timetable for updates. As a minimum, the DMP should be updated in the context of the periodic evaluation/assessment of the project. If there are no other periodic reviews envisaged within the grant agreement, an update needs to be made in time for the final review at the latest.

In the following the main sections to be covered by the DMP are outlined. At the end of the document, Table 1 contains a summary of these elements in bullet form.

This template itself may be updated as the policy evolves.

1. Data Summary

What is the purpose of the data collection/generation and its relation to the objectives of the project?

What types and formats of data will the project generate/collect?

Will you re-use any existing data and how?

What is the origin of the data?

What is the expected size of the data?

To whom might it be useful ('data utility')?

2. FAIR data

2. 1. Making data findable, including provisions for metadata

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

What naming conventions do you follow?

Will search keywords be provided that optimize possibilities for re-use?

Do you provide clear version numbers?

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

2.2. Making data openly accessible

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for *opting out.*

How will the data be made accessible (e.g. by deposition in a repository)?

What methods or software tools are needed to access the data?

Is documentation about the software needed to access the data included?

Is it possible to include the relevant software (e.g. in open source code)?

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

Have you explored appropriate arrangements with the identified repository?

If there are restrictions on use, how will access be provided?

Is there a need for a data access committee?

Are there well described conditions for access (i.e. a machine readable license)?

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

2.3. Making data interoperable

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

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?

2.4. Increase data re-use (through clarifying licences)

How will the data be licensed to permit the widest re-use possible?

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

How long is it intended that the data remains re-usable?

Are data quality assurance processes described?

Further to the FAIR principles, DMPs should also address:

3. Allocation of resources

What are the costs for making data FAIR in your project?

How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

Who will be responsible for data management in your project?

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

4. Data security

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

Is the data safely stored in certified repositories for long term preservation and curation?

5. Ethical aspects

Are there any ethical 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).

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

6. Other issues

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

7. Further support in developing your DMP

The Research Data Alliance provides a [Metadata Standards Directory](#) that can be searched for discipline-specific standards and associated tools.

The [EUDAT B2SHARE](#) tool includes a built-in license wizard that facilitates the selection of an adequate license for research data.

Useful listings of repositories include:

- [Registry of Research Data Repositories](#)

- Some repositories like [Zenodo](#), an OpenAIRE and CERN collaboration), allow researchers to deposit both publications and data, while providing tools to link them.

- Other useful tools include [DMP online](#) and platforms for making individual scientific observations available such as [ScienceMatters](#).

SUMMARY TABLE 1

FAIR Data Management at a glance: issues to cover in your Horizon 2020 DMP

This table provides a summary of the Data Management Plan (DMP) issues to be addressed, as outlined in Annex I. You should refer to the annex and the main text of the guidelines for further guidance.

DMP component	Issues to be addressed
1. Data summary	- State the purpose of the data collection/generation - Explain the relation to the objectives of the project - Specify the types and formats of data generated/collected - Specify if existing data is being re-used (if any) - Specify the origin of the data - State the expected size of the data (if known) - Outline the data utility: to whom will it be useful
2. FAIR Data 2.1. Making data findable, including provisions for metadata	- Outline the discoverability of data (metadata provision) - Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? - Outline naming conventions used - Outline the approach towards search keyword - Outline the approach for clear versioning - Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how

2.2 Making data openly accessible	• Specify which data will be made openly available? If some data is kept closed provide rationale for doing so • Specify how the data will be made available • Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? • Specify where the data and associated metadata, documentation and code are deposited • Specify how access will be provided in case there are any restrictions
2.3. Making data interoperable	• Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. • Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
2.4. Increase data re-use (through clarifying licences)	• Specify how the data will be licenced to permit the widest reuse possible • Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed • Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why • Describe data quality assurance processes • Specify the length of time for which the data will remain re-usable
3. Allocation of resources	• Estimate the costs for making your data FAIR. Describe how you intend to cover these costs • Clearly identify responsibilities for data management in your project • Describe costs and potential value of long term preservation

4. Data security	• Address data recovery as well as secure storage and transfer of sensitive data
5. Ethical aspects	• To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
6. Other	• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)