EUROPEAN COMMISSION
Directorate-General for Research & Innovation

Guidelines on Data Management in Horizon 2020

Version 2.1

15 February 2016

HORIZON 2020

Introduction

> Horizon 2020 includes a limited and flexible **pilot action on open access to research data**. Participating projects must develop a **Data Management Plan** (DMP) specifying which data will be openly accessible.

This document is intended to help applicants and beneficiaries of projects under Horizon 2020 meet their responsibilities as regards **research data quality, sharing and security**.

DATA MANAGEMENT PLANS

A new element in Horizon 2020 is the **use of Data Management Plans** (DMPs). A DMP describes the data management life cycle for all datasets to be collected, processed or generated by a research project. It must cover:

- the handling of research data during & after the project
- what data will be collected, processed or generated
- what methodology & standards will be applied
- whether data will be shared /made open access & how
- how data will be curated & preserved

The use of a Data Management Plan is **required for projects participating in the Open Research Data Pilot**. Other projects are invited to submit a DMP if it is relevant to their planned research.

For a detailed description of the Open Research Data Pilot requirements and its scope visit the [Open Access to Scientific Publications and Research Data in Horizon 2020 Guidelines](#).

PROPOSAL SUBMISSION AND EVALUATION

All project proposals submitted as 'research and innovation actions' or 'innovation actions' include a section on research data management, which is assessed under the 'impact' criterion. As an applicant for these actions, you must **outline your data management policy**, where relevant. Include the following issues:

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

Your policy should

- reflect the current state of consortium agreements on data management
- be consistent with exploitation and IPR requirements

The data management section is also a checklist for the future and a reference for the resource and budget allocations associated with data management.

Projects participating in the Open Research Data Pilot and those intending to develop a DMP voluntarily should include into their proposal a deliverable for an initial DMP at month 6 at the latest. (See more details below on submitting DMP)

⚠ Important: the section on data management policy is **not** a Data Management Plan. Fully developed Data Management Plans are not expected at submission stage and are **not** taken into account as part of the evaluation.

RESEARCH DATA MANAGEMENT IN THE MODEL GRANT AGREEMENT

References to research data management as part of the Pilot on Open Research Data in the Horizon 2020 are included in Article 29.3 of the Model Grant Agreement, which applies to all projects participating in the Pilot:

> 'Regarding the digital research data generated in the action ('data'), the beneficiaries must:
>
> (a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:
>
> (i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;
>
> (ii) other data, including associated metadata, as specified and within the deadlines laid down in the data management plan (see Annex I);
>
> (b) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and —

These provisions are further explained in the Open Access Guidelines and in the Annotated Model Grant Agreement.

DATA MANAGEMENT PLANS DURING THE PROJECT CYCLE

First version: Projects taking part in the Open Research Data Pilot and those intending to develop a DMP voluntarily must provide a first version of the Data Management Plan **within the first six months** of the project. These projects should ensure that their proposal includes an initial DMP as a deliverable at month 6 at the latest.

Updates: Data Management Plan should be updated - if appropriate - during the project lifetime (*in the form of deliverables*). Please create and submit a new version of the DMP whenever significant changes arise in your project such as

- new data sets

- changes in consortium policies

- external factors.

Deadline of updates:

- as a minimum, update the DMP in time to be evaluated as part of the mid-term and final project reviews for projects that have planned reviews

- DMP updates needed at least at the end of projects, for projects where reviews are not planned

- and at other moments if it is defined by project consortia.

Periodic reporting: check further details in the Online Manual on how to fill in reporting tables for publications, deliverables and the process for continuous reporting in the grant management system of the Participant portal.

SUPPORT

Costs related to data management in Horizon 2020 are **eligible** for reimbursement during the duration of the project (see Article 6.2.D.3 of the Annotated Model Grant Agreement).

For now the Commission does not offer a DMP tool. Beneficiaries can generate their own data management plan or **use tools that are compatible with the requirements** set out in Annex 1 - Data Management Plan template - such as the Digital Curation Centre's DMP Online tool.

Support for research data management is available from OpenAIRE 2020. These services are primarily available to research projects funded under Horizon 2020, with preference given to those participating in the Open Research Data Pilot.

ANNEX 1: Data Management Plan (DMP) template

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the applicants with regard to all the datasets that will be generated by the project.

The DMP is not a fixed document, but evolves during the lifespan of the project.

The DMP should address the points below on a dataset by dataset basis and should reflect the current status of reflection within the consortium about the data that will be produced.

- **Data set reference and name**

  Identifier for the data set to be produced.

- **Data set description**

  Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.

- **Standards and metadata**

  Reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created.

- **Data sharing**

  Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).

  In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

- **Archiving and preservation (including storage and backup)**

  Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.

ANNEX 2: Additional guidance for Data Management Plans

This can be applied to any project that produces, collects or processes research data, and is included here as reference for elaborating DMPs in Horizon 2020 projects. This guide is structures as a series of questions that should be ideally clarified for all datasets produced in the project.

Scientific research data should be easily

1. Discoverable

a. DMP question: are the data and associated software produced and/or used in the project discoverable (and readily located), identifiable by means of a standard identification mechanism (e.g. Digital Object Identifier)?

2. Accessible

a. DMP question: are the data and associated software produced and/or used in the project accessible and in what modalities, scope, licenses (e.g. licencing framework for research and education, embargo periods, commercial exploitation, etc.)?

3. Assessable and intelligible

a. DMP question: are the data and associated software produced and/or used in the project assessable for and intelligible to third parties in contexts such as scientific scrutiny and peer review (e.g. are the minimal datasets handled together with scientific papers for the purpose of peer review, are data is provided in a way that judgments can be made about their reliability and the competence of those who created them)?

4. Useable beyond the original purpose for which it was collected

a. DMP question: are the data and associated software produced and/or used in the project useable by third parties even long time after the collection of the data (e.g. is the data safely stored in certified repositories for long term preservation and curation; is it stored together with the minimum software, metadata and documentation to make it useful; is the data useful for the wider public needs and usable for the likely purposes of non-specialists)?

5. Interoperable to specific quality standards

a. DMP question: are the data and associated software produced and/or used in the project interoperable allowing data exchange between researchers, institutions, organisations, countries, etc. (e.g. adhering to standards for data annotation, data exchange, compliant with available software applications, and allowing re-combinations with different datasets from different origins)?