



EUROPEAN COMMISSION
Directorate-General for Research & Innovation

Guidelines on Data Management in Horizon 2020

Version 2.0
30 October 2015



Introduction

In Horizon 2020 a limited and flexible **pilot action on open access to research data** will be implemented (see guidance on Open Access to Scientific Publications and Research Data in Horizon 2020). Participating projects will be required to develop a Data Management Plan (DMP), in which they will specify what data will be open.

This document provides **guidelines for data management in Horizon 2020** and is addressed to applicants and beneficiaries of projects under the Framework Programme. These guidelines are intended to give them indications on how they can comply with their responsibilities regarding research data quality, sharing and security.

Proposal submission and evaluation

All project proposals submitted to "Research and Innovation actions" as well as "Innovation actions" include a section on research data management which is evaluated under the criterion 'Impact'. Where relevant, applicants must provide a short, general outline of their policy for data management, including the following issues:

- What types of data will the project generate/collect?
- What standards will be used?
- How will this data be exploited and/or shared/made accessible for verification and re-use? If data cannot be made available, explain why.
- How will this data be curated and preserved?

The described policy should reflect the current state of consortium agreements regarding data management and be consistent with those referring to exploitation and protection of results. Please refer to the proposal Template B for structure and additional instructions.

The data management section can be considered also as a checklist for the future and as a reference for the resource and budget allocations related to data management.

Data Management Plans (DMPs) are introduced in the Horizon 2020 Work Programmes:

A further new element in Horizon 2020 is the use of Data Management Plans (DMPs) detailing what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and 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 Data Management Plan if relevant for their planned research.

A detailed description and scope of the Open Research Data Pilot requirements is provided on the Participants Portal (Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020).

Projects taking part in the Pilot on Open Research Data are required to provide a first version of the DMP as an early deliverable within the first six months of the project. Projects participating in the pilot as well as projects who submit a DMP on a voluntary basis because it is relevant to their research should ensure that this deliverable is mentioned in the proposal. Since DMPs are expected to mature during the project, more developed versions of the plan can be included as additional deliverables at later stages. The purpose of the DMP is to support the data management life cycle for all data that will be collected, processed or generated by the project.

Grant agreement / Pilot on Open Research Data in Horizon 2020

References to research data management are included in Article 29.3 of the Model Grant Agreement (article applied to all projects participating in the Pilot on Open Research Data in Horizon 2020):

'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 — where possible — provide the tools and instruments themselves).'*

A DMP as a document outlining how research data will be handled during a research project, and after it is completed, is very important in all aspects for projects participating in the Horizon 2020 Open Research Data Pilot as well as almost any other research project. Especially where the project participates in the Pilot it should always include clear descriptions and rationale for the access regimes that are foreseen for collected data sets. This principle is further clarified in the following paragraph of the Model Grant Agreement:

As an exception, the beneficiaries do not have to ensure open access to specific parts of their research data if the achievement of the action's main objective, as described in Annex I, would be jeopardised by making those specific parts of the research data openly accessible. In this case, the data management plan must contain the reasons for not giving access.

Project reviews

A DMP describes the data management life cycle for all data sets that will be collected, processed or generated by the research project. It is a document outlining how research data will be handled during a research project, and even after the project is completed, describing what data will be collected, processed or generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved. The DMP is not a fixed document; it evolves and gains more precision and substance during the lifespan of the project.

The first version of the DMP is expected to be delivered within the first 6 months of the project. This DMP deliverable should be in compliance with the template provided by the Commission (see Annex 1 to this document). More elaborated versions of the DMP can be delivered at later stages of the project. The DMP would need to be updated at least by the mid-term and final review to fine-tune it to the data generated and the uses identified by the consortium since not all data or potential uses are clear from the start. New versions of the DMP should be created whenever important changes to the project occur due to inclusion of new data sets, changes in consortium policies or external factors. Suggestions for additional information in these more elaborated versions are provided below in Annex 2.

Support on research data management for projects funded under Horizon 2020 is planned through projects funded under the Research Infrastructures Work Programme 2014-15 (call 3 e-Infrastructures). Exact plans for service delivery will be available in the Participants Portal by end of 2014. Full support services support is expected to be available latest in 2015 and coordinated with similar national initiatives. These services will be available only to research projects funded under Horizon 2020, with preference 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 structured 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)?