

# European Commission: Horizon 2020 - Initial DMP

## Data summary

Provide a summary of the data addressing the following issues:

- 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

## 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 metadata will be created and how.

*Guidance:*

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

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.

*Guidance:*

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.

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.

The [Registry of Research Data Repositories](#) provides a useful listing of repositories that you can search to find a place of deposit.

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?

*Guidance:*

Interoperability means 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.

2.4 Increase data re-use (through clarifying licenses):

- Specify how the data will be licensed 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 usable 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.

*Guidance:*

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

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.

## **Allocation of resources**

Explain the allocation of resources, addressing the following issues:

- 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

### *Guidance:*

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).

Costs 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 security**

Address data recovery as well as secure storage and transfer of sensitive data.

### *Guidance:*

Also consider whether the data is safely stored in certified repositories for long term preservation and curation.

## **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.

### *Guidance:*

Consider whether there are any ethical or legal issues that can have an impact on data sharing. For example, is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

## **Other**

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any).

# European Commission: Horizon 2020 - Updated DMP

## 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')?

## 2.1 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.

### *Guidance:*

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

## 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.

### *Guidance:*

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.

### *Guidance:*

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 platforms for making individual scientific observations available such as [ScienceMatters](#).

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?

*Guidance:*

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

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?

## **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?