TEMPLATE HORIZON 2020 DATA MANAGEMENT PLAN (DMP)

Annotated version for the use of participants under Societal Challenge 1

- **Instructions and footnotes in blue must not appear in the text.**
- **For options *[*in square brackets*]*: the option that applies must be chosen.**
- **For fields in [grey in square brackets] (even if they are part of an option as specified in the previous item): enter the appropriate data.**

Introduction

This Horizon 2020 DMP template has been designed to be applicable to any Horizon 2020 project that produces, collects or processes research data. In order to address specific issues under the different thematic priorities of the programme, though, updated versions can be prepared based on identified needs. In this template you will find specific annotations for projects funded from Societal Challenge 1 (Health, demographic change and wellbeing).

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.

Guidelines on FAIR Data Management in Horizon 2020 are available in the Online Manual.

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.

Project[1] Number: [insert project reference number]

Project Acronym: [insert acronym]

Project title: [insert project title]

DATA MANAGEMENT PLAN

[1] The term 'project' used in this template equates to an 'action' in certain other Horizon 2020 documentation

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')?

Specific for SC1 projects

Details on research methods (e.g. DNA sequencing) as such do not need to be described in the DMP. The DMP should consider, however, to what extent such methods might be an important element of the metadata needed to annotate certain data.

2. FAIR data

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

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.

Specific for SC1 projects

Apart from the exceptions provided by law, the processing of personal data can only occur with consent. The specifications of the consent need to be provided, so that it is clear which sharing and re-use of personal data can occur within the existing consent and for which sharing and re-use of personal data re-consenting will be required.

[2] *Metadata: the details about what, where, when, why, and how the data were collected, processed, and interpreted. Metadata include descriptions of how data and files are named, physically structured, and stored as well as details about the experiments, analytical methods, and research context. It is generally the case that the utility and longevity of data relate directly to how complete and comprehensive the metadata are. The amount of effort devoted to creating comprehensive metadata may vary substantially based on the complexity, types, and volume of data."*
(Source: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4619636/)

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

Specific for SC1 projects:

Please find here below a non-exhaustive list of repositories linked to health research:

https://www.ucl.ac.uk/library/research-support/research-data/learn-develop-teach/Subject-specific_resources/data_repositories

https://www.cancerresearchuk.org/sites/default/files/hands_on_data_sharing_advice_-_population.pdf

https://www.nature.com/sdata/policies/repositories

For personal data deposition in a repository that has established certain access requirements, might also be possible.

*If a repository for a specific type of health research data is not available, such data could be shared through a more generic repository such as OpenAIRE (*https://www.openaire.eu/*).*

Specific for SC1 projects:

***Clinical studies** must be registered and related summary results must be reported in WHO/ICMJE-approved registries. The related mandatory deliverables (see Annex II of the template for clinical studies) must be included in the DoA. We are currently not aware of a repository that allows OPEN sharing of individual patient data (IPD) underlying the results of a clinical study in a manner that is compliant with the protection of personal data. DMPs can refer to this exception from open data sharing and/or can refer to alternative suggestions on how IPDs might be shared.*

ICJME recommendations: http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html

WHO International Clinical Trials Registry Platform: http://www.who.int/ictrp/network/primary/en/

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?

Specific for SC1 projects:

Does the research undertaken by your action entail issues pertinent to Public Health Emergencies (PHERs)? If yes, has your project joined any national or international data sharing initiative to address PHERs? Could you please describe the data sharing policy that the project applies and the initiative it may have joined?

Does the research undertaken by your action involve monitoring of chemical substances in humans (human-biomonitoring), the environment, food and feed or in product and indoor air? If yes, it is strongly recommended to deposit your research data at the following database(s): IPCHEM platform: https://ipchem.jrc.ec.europa.eu

Example of a data sharing in the case of infectious diseases: Global Research Collaboration for Infectious Disease Preparedness (GloPID-R)

https://www.glopid-r.org/

https://www.glopid-r.org/find-out-about-our-work/data-sharing-working-group/

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.

Specific for SC1 projects:
What is the policy on how data will be owned/stored (cloud, repositories) once the project is over (sustainability of data)?

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).

Projects participating to the ORDP might present information relevant to the ethical aspects (data protection) in the DMP. In such a case, the ethics chapter of the DoA may simply refer to the DMP for more information on the details of the ethics aspects related to data.

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 above.

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)

HISTORY OF CHANGES		
Version	**Publication date**	**Change**
1.0	13.10.2016	▪ Initial version
2.0	15.02.2018	▪ Revised version introducing annotations specific to SC1 projects