How to create your Data Management Plan for H2020 and HE

A practical guide

Version 2/2021

Table of contents

Preface	p. 3
How to create your DMP	p. 18
1. The Administrative Plan details	p. 20
2. Data summary	p. 22
3. FAIR Data	p. 37
3.1. Making data findable, including provisions for metadata	p. 39
3.2. Making data openly accessible	p. 45
3.3. Making data interoperable	p. 53
3.4. Increase data re-use (through clarifying licenses)	p. 58
4. Allocation of resources	p. 64
5. Data security	p. 68
6. Ethical aspects	p. 71
7. Other issues	p. 79

Preface

Course aim

The aim of this course is to enable researchers to complete a FAIR Data Management Plan (DMP) on their own, consistent with the requests of Horizon2020 and Horizon Europe.

What do research funders want?

Funders expect the DMP to outline how data will be created, managed, shared, disseminated & preserved, justifying any restriction that needs to be applied.

When you complete your DMP, be *clear and concise*: funders typically expect a succinct (but exhaustive) DMP.

What are you expected to do?

Choose and demonstrate that the selections you made in the DMP about research data management are the most appropriate for your context, that of your discipline and future users. Similarly, you need to present a convincing case for any restrictions on data sharing.

Write a good plan: funders want to see that you understand their requirements and have realistic plans in place to meet there. The description of planned work should be clear and achievable, so markers can feel confident that you will be able to deliver what is proposed.

What is Research Data?

Research data is any information that has been collected, observed, generated or created to validate original research findings. Research data also includes non-digital formats such as laboratory notebooks, diaries and sketchbooks.

Research data can take many forms. It might be:

documents, spreadsheets, laboratory notebooks, field notebooks, diaries, questionnaires, transcripts, codebooks, audiotapes, videotapes, photographs, films, test responses, slides, artefacts, specimens, samples, collections of digital outputs, data files, database contents (video, audio, text, images), models, algorithms, scripts, contents of an application (input, output, logfiles for analysis software, simulation software, schemas), methodologies and workflows, standard operating procedures and protocols.

Sources of research data

- **Observational data** is captured in real-time, and is usually irreplaceable, for example sensor data, survey data, sample data, and neuro-images.
- **Experimental data** is captured from lab equipment. Examples of experimental data are gene sequences, chromatograms, and toroid magnetic field data.
- **Simulation data** is generated from test models where model and metadata are more important than output data. For example, climate models and economic models.
- **Derived or compiled data** has been transformed from pre-existing data points. Examples are data mining, compiled databases, and 3D models.
- **Reference or canonical data** is a static or organic conglomeration or collection of smaller (peer-reviewed) datasets, most probably published and curated. For example, gene sequence databanks, chemical structures, or spatial data portals.

7 important infos about DMP

1. About the Data Management Plan
2. Who & when to write a DMP
3. Data & metadata
4. Data: as open as possible, as closed as necessary
5. Sensitive data or data with copyright restrictions
6. Data minimisation
7. Why FAIR data management

About the Data Management Plan (DMP)

The DMP is a formal and structured document which depicts the entire lifecycle of research data. It assures that data are traceable, available, authentic, citable, properly stored, according to the FAIR principles.

Data adhere to clearly defined legal parameters and appropriate safety measures governing subsequent use.

The DMP provides specific information on:

- what data will be created and how;
- how data will be described, organised, stored and managed;
- how long they will be stored (accordingly to the policy for example);
- how will be responsible for each activities;
- how data will be shared, explaining any use restriction that could apply.

Who & when to write a DMP

Who: the principal investigator is the main responsible person for the compilation.

When: in the initial phase of data collection (> month 6). In any case, the sooner the better: take your time to think in advance about your data, will offer more chance to plan and allocate resources & calculate costs (eligible in the plan, if planned).

The DMP may evolve into versions (1st, updates, final) during your project: it needs to be updated whenever significant changes arise, e.g. new data production and, as a minimum, in time with periodic evaluation/assessment of the project.

Data & metadata (1)

When you produce some resource (document files, images, videos, etc.) the data is tagged with information that best describe it.

These tags form the resource metadata and they allow machine and user to find your data. Data is almost never separated from the metadata.

Data & metadata (2)

Metadata serve a variety of purposes

1. *facilitating interoperability and integrating resources by humans & machines* → in order to exchange info among different systems, platforms, data structures and interfaces;
2. *facilitating digital identification via standard numbers* → combining metadata as a set of identifying data (DOI, ISSN, ORCID ID, ISBN, etc.);
3. *protect resources and their future accessibility* → for archiving and preservation purposes, metadata track and describe the object, so it can be replicated on technologies in the future;
4. *the data & related metadata evaluated & included in the DMP* → produced during the research & needed for its validation in scientific publication (underlying data); other data (e.g. raw data).

This is the reason why metadata are *compliant with international standards*.

Data: as open as possible, as closed as necessary

If you plan to keep some datasets closed, justify this decision choosing which datasets should be made available and when, and possibly change the decision along the way.

The reason for keeping the datasets closed are

1. data restriction for possible economic exploitations;
2. data protection issues (e.g. security reasons, potential harm for individual);
3. incompatibility with the obligations to protect special categories of personal data ('sensitive data');
4. in the event that making data open is a risk for achieving the main goal of the project;
5. other legitimate reasons to describe and motivate.

Sensitive data and data with copyright restrictions

If you are working with *sensitive data* or data with *copyright restrictions*, you need a clear plan that addresses any potential privacy or legal issue.

The term «sensitive data» commonly refers to personal data that can reveal & identify univocally a natural person and data concerning racial and ethnic origin, religious or philosophical beliefs, political opinions, trade union membership as well as genetic data, health, sex life or sexual orientation (art. 9 GDPR). These data are related to the most intimate sphere of each individual.

They must be processed applying special safeguards and there must always be a specific condition that allow their processing (art. 9 GDPR).

The principle of purpose limitation and data minimisation (art. 5 GDPR)

Purpose limitation is a principle related to personal data processing that states that data collected and processed should not be held or further used unless for reasons that were clearly stated in advance.

Data minimisation means that personal data must be adequate, relevant, limited to what is necessary for the purposes for which data are processed

Data minimisation reduces the risk of unauthorised access.

Why FAIR data management

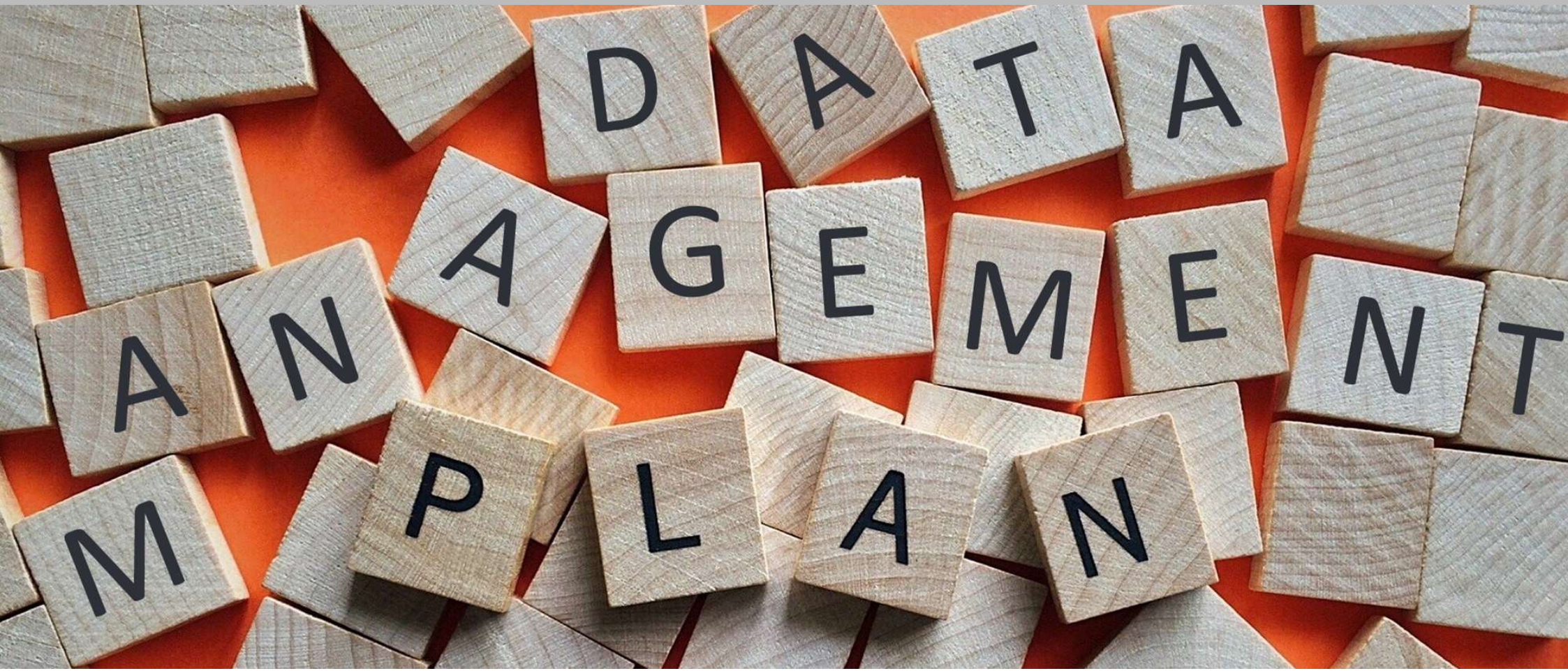
Funders (and publishers) require that research data be *FAIR*

FINDABLE – ACCESSIBLE – INTEROPERABLE – RE-USABLE

1. can be found on the Internet;
2. are accessible (not necessarily open) with clear rights and licenses;
3. are in usable format and reliable;
4. are identified in a unique & persistent way

so that data and metadata are easy to manage by *humans & computer*.

How to create your



Structure of the Data Management Plan

Structure your DMP in 7 parts, according to the Horizon2020 FAIR Data Management Plan template

1. Administrative plan details (p. 18)
2. Data summary (p. 20)
3. FAIR data (p. 35)
4. Allocation of resources (p. 62)
5. Data security (p. 66)
6. Ethical aspects (p. 69)
7. Other issues (p. 77)

Part 1
The Administrative plan details

The ID card of the project

Provide some administrative information

Project number	Enter the call ID of your programme *e.g. H2020 ID 882255*
Project Acronym	Enter the acronym of your project *e.g. THGA*
Project title	Enter the name of your project *e.g. Theory of Games – THGA*
Funder	Enter the funding institution name *e.g. European Commission H2020*
Principal Investigator/ Researcher	Enter the name of the researcher who writes the DMP *e.g. Laura Rossi*
Principal Researcher ORCID ID	Enter your ORCID ID *e.g. 0000-0003-4170-6345* If you do not have it, go to https://orcid.org/ and register your profile
Project data contact	Enter your phone number and your institutional @ *e.g. +39 041 2349999, laura.rossi@unive.it*
Version and date of the DMP	Specify version and date of this DMP *e.g. 1st Version, Update, Final xx.xx.20xx*

Part 2
Data summary

Describe the expected data to be gathered, the nature and scale of the data that will be generated or collected

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

In this section *describe the type of study* and justify your choices: make a *brief description* of your research and the study to be carried out, the *scope* of the investigation, the *duration*, and the *methodology* you intend to follow for data collection and management.

> *This four-year project will produce … To achieve this object, this study will develop new methods based on Artificial Intelligence for digitally identifying, collecting, integrating, managing and sharing diverse data …*

What types & formats of data will the project generate/collect? (1)

Describe the types of data to be managed during the research: write if they are quantitative, qualitative, generated from surveys, interviews, clinical trials, administrative records, images, etc.

This project will acquire & manipulate, collect & generate sets of different data:

- *acquired datasets (archival documents, topographic maps) from public archives;*
- *collected datasets gathered during the activities (GPS coordinates);*
- *generated datasets produced from processing, digitizing acquired datasets (images, map documents, …);*
- *Data collcted through personal interviews (audio-video recordings, questionaries, etc.)*
- *similar data: survey data available in thematic repositories.*

What types & formats of data will the project generate/collect? (2)

You can manage *personal* and *non-personal* data: they are handled in different ways due to data protection law.

Personal data are any information relating to an identified/ identifiable living individual.

The processing and the protection of personal data must be compliant with GDPR that defines the techniques and the measures to be used for processing personal data.

What types & formats of data will the project generate/collect? (3)

Examples of *personal data*

- personal details (name, surname, gender, age, date of birth);
- contact details (address, e-mail address, telephone no., Skype Id);
- identity documents and unique identifiers (passport, license plate no.);
- CV and related professional and academic experiences;
- recorded voice (not altered) or videos;
- images and comments from which it is possible to identify a person;
- online identifiers (IP address, geo-location points).

What types & formats of data will the project generate/collect? (4)

Common security measures can be seen in anonymisation, pseudonymisation, encryption.

Personal data that has been pseudonymised (personal data can no longer be attributed to a specific data subject without the use of additional information) can be used to re-identify a person, thus remains personal data and falls within the scope of the GDPR.

Personal data that has been rendered anonymous (e.g. information which does not relate to an identified or identifiable natural person, as well as if the individual is not or no longer identifiable) is no longer considered personal data.

What types & formats of data will the project generate/collect? (5)

Anonymous data in no way allows the identification of the individual because personal information is not *ab origine* received or is subsequently deleted, in order to make the data "anonymous".

What types & formats of data will the project generate/collect? (6)

Pseudonymisation is "the processing of personal data in such a manner that the personal data *can no longer be attributed* to a specific data subject without the use of additional information, provided that such additional information is *kept separately* and is *subject to technical and organisational measures* to ensure that the personal data are not attributed to an identified or identifiable natural person".

What types & formats of data will the project generate/collect? (7)

Data that do not involve personal information do not require special handling (except for confidantional information e.g. trade secrets, etc.).

Some examples

- information concerning companies, associations, committees;
- non-personal address such as info@company.com;
- anonymous data (e.g. aggregate data).

If you do not manage personal data, write in your DMP

I do not intend to produce personal data

What types & formats of data will the project generate/collect? (8)?

Continue describing the format and scale of your data (open or non-proprietary formats are preferable)

- text documents (DOC, ODF, PDF, TXT);
- images (JPG, GIF, SVG, PNG, TIFF);
- video/film (MPEG, AVI, WMV, MP4);
- audio (MP3, WAV, AIFF, OGG);
- structured data (HTML, JSON, TEX, XML, RDF);
- table (CSV, ODS, TSV, XLS, SAS, STATA, SPSS portable);
- source code (C, CSS, JAVASCRIPT, JAVA);
- configuration data (INI, CONF).

Will you re-use any existing data and how?

Describe the data reuse and integration in the research project and under which conditions: establish the *Intellectual Property Rights* (if any) and legal requirements, to avoid problems with ©-protected material and make data available.

Most of the data come from public and private repositories and archives. Some of them will be realized retaining their private data policy… All the data collected from thematic repositories are copyright free and in any case do not cause IPR infringement.

About possibilities for integration and reuse: collected and generated data could provide useful background information for similar projects.

What is the origin of the data? (1)

Describe the provenance of your data, how data will be gathered or generated and if you intend to conduct surveys, questionnaires, observe people behavior, carry out clinical trial, analyze data from web or from social networks, process information contained in archives or databases. In case of pre-existing data, explain if they require processing or data cleaning.

The project will reuse a variety of existing data from different sources: data on … managed by …; data from EU and Government databases (such as…); data from communication sources produced by … (e.g.: policy documents, reports, magazines, …); reports created by Local Authorities; tweets and blog posts; interviews, surways, etc. …

What is the origin of the data? (2)

List the devices used for data acquisition and if they already have the possibility of encryption or other appropriate security mechanisms by design (e.g. password, fingerprint).

Remember that an image, a video, a voice note, etc., *constitute personal data*, since they allow to identify a certain natural person. The collection, storage, use, communication and other operations carried out with such categories of data need appropriate precautions, adequate computer and organisational measures, appropriate guarantees. For example:

- immediate deletion of the audio/video file after the transcription of the anonymous conversation;
- preservation of the voice note with artificially modified tone, etc…

What is the expected size of the data?

Estimate the *amount of data* you'll manage: MB, GB, TB, PB.

If it is difficult to quantify, consider in case of

- interview → how many interviews and how long the video or audio last;
- report → how many reports;
- data or statistics → how many tables or datasets are collected/produced.

To whom might it be useful (data utility)?

State the *added value* that the new data would provide in relation to existing data and the scientific community.

Data collected and generated provide benefits to the advancement of archaeological research on local basis and important information for researchers, students, policy makers, stakeholders…

Part 3
FAIR data


Findable
Persistent Identifiers (PIDs)
Rich metadata
Indexed data repositories
PIDs in metadata
Accessible
Standard communications protocol
Open, free protocol
Authentication, where necessary
Metadata is always available
Interoperable
Vocabularies
Vocabularies are FAIR
Linked metadata
Reusable
Metadata have multiple attributes
Usage license
Provenance
Community standards

Part 3.1
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)?

Data and metadata should be easy to find by both humans & machines.

Data should be described with rich metadata that allow indexing in data repositories and discoverability (describing content, context and characteristics of the data will assure findability).

Data repositories selected, should use Persistent Identifiers (such as DOI for data, ORCID for researchers).

What naming conventions do you follow?

Describe your naming conventions. For example

- for data set file(s)
  [PROJECT ACRONYM]_WPnumber_Tnumber_coverage or other content specifications_date (YYYYMMDD)_version.file extention
- for readme file(s)*
  README_[PROJECT ACRONYM]_WPnumber_Tnumber_coverage or other content specifications_date (YYYYMMDD)_version.file extention

*A "README" file is a document containing relevant information about data set authorship, terms of reuse and responsibilities, explaining data set content and structure, collection procedures and analysis (such as file specifics, methodologies, codebooks of variables, data sources, and further necessary notes). (See Annex II to visualise the suggested README file template).

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

If there are controlled vocabularies or thesauri typical of the disciplinary area, please indicate them: they are very useful for indexing and discoverability.

Do you provide clear version numbers?

Describe how you will keep track of the versioning

- it refers to saving new copies of your files when you make changes so that you can go back and retrieve specific versions of your files later;
- when creating new versions of your files, record what changes are being made to the files and give the new files a unique name. Include a version number, e.g. “v1”, “v2“ (for major revisions) or "v2.1“ (for minor revisions);
- include information about the status of the file, e.g. “draft” or “final”.

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

Dublin Core metadata set (or Dublin Core enriched) could be used. A minimum set of metadata should include:

- Title
- Creator
- Subject
- Description
- Contributor
- Date
- Type
- Format
- Language
- Identifier
- Rights management

Part 3.2
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?

Try to make your data openly accessible whenever possible in order to allow dissemination, validation and re-use of research results.

Choose a data repository FAIR compliant and indexed (Re3data,Open AIRE).

For each deposited data set, make available all relevant documentation explaining data collection procedures and analysis (codebooks, methodologies, etc.) in order to guarantee intelligibility, reproducibility and the validation of the project findings. Specify, when necessary, the tools and software recommended for reproducing and reusing 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)?

In case specific software is needed to allow the reuse of the data you should list them. For example

- open spreadsheet and document editors, such as OpenOffice or LibreOffice
- free CSV file viewers, such as CSV viewer
- R, free software environment for statistical computing and graphics
- open or free image viewers
- etc…

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.

Indicate which repository you will choose and what its features are and describe if it is a certified repository or an indexed repository.

For each deposited data set, all relevant documentation explaining data collection procedures and analysis (such as codebooks, methodologies, etc.) will be made available along with the data, in order to guarantee intelligibility, reproducibility and the validation of the project findings.

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

It is possible that some datasets will be published after an embargo period.

If some datasets cannot be made accessible even in the future, it is necessary to indicate the reason for it (data protection).

Your DMP should identify the versions or parts of the data sets that cannot be freely shared providing the specific motivations.

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

The metadata must also include usage licenses (preferably Creative Commons licenses) or a copyright text that allows users to understand what uses are allowed.

In the DMP you should describe which CC license will be used (CC 0 or CC BY) or a text of copyright.

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

Describe who will have access to the datasets, with which permissions, who will authorise these permissions and what will be the authentication method.

Part 3.3
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)?

To allow interoperability, for indexing and discoverability, data sets should be described using standard descriptive metadata, such as *Dublin Core* and *DataCite Metadata Schema*.

All relevant documentation explaining codebooks, users' manuals, data collection procedures and analysis should be available along with the data in order to guarantee intelligibility, reproducibility and the validation of the project findings.

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

To allow data exchange and re-use among researchers, institutions, organisations, countries, etc., partners will convert all shareable data from proprietary formats and will made them available in well-known and documented open formats.

If metadata vocabularies and standard are available for your discipline you should use them.

Choose for your data format open standards (see slide 42) and describe them in your DMP.

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

It is important to use standard vocabularies or ontologies, well known and shared in the scientific communities.

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?

If there is a need to create specific ontologies or vocabularies for the project, it is recommended to provide mappings to more commonly used ontologies.

Part 3.4
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. (1)

Specify how the data will be licenced to permit the widest reuse possible and when the data will be made available for re-use. If applicable, why and for what period a data embargo is needed.

When possible, the dataset will be licensed under an open access license. This will depend on the level of privacy and the IPR involved in the data set.

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. (2)

An embargo period will be necessary only if the data set contains specific IPR or other exploitable results will justify an embargo. The embargo period should have a reasonable lenght that should be motivated.

If a text of copyright is used instead of an OA license, the rights of reuse should be described.

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.

Indicate whether and for how long the datasets will be made available to third parties. If there are parts that will remain secret (even if they are always accessible) they should be indicated.

Define how long after the end of the project the datasets will remain available for reuse. A sensible period could be 10 years but it also depends on the policies of the individual institution and the repository.

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

Specify how long the data will be stored (this often also depends on the policies of the archives).

Data will be stored for a period of … years after the end of the project.

Are data quality assurance processes described?

It is important to describe the methodologies of data production and aggregation, any software must be indicated.

Specific methods of data processing must be described with accuracy.

Part 4.
Allocation of resources

Making data fair requires a certain amount of human effort and investments: creating documentation and making your data understandable is very time consuming, so be realistic about how much effort is needed to prepare your data for sharing and preservation.

What are the costs for making data FAIR in your project? How will these be covered?

Outline and justify costs: storage, additional hardware, software and technical expertise, project website setting up, outsource services (back-up and preservation), etc.

Costs related to data management support and documentation, conversion of proprietary data files into open formats, processing of interviews (transcription, translation, anonymisation), services on open access issues, etc.

Who will be responsible for data management in your project?

Specify who collects, processes, structures, stores the data: Researchers, PHD students, external partners, collaborators.

Responsible for data management are the data sets creators who are directly involved in research data organisation and collection.

Researchers are identified with the unique persistent identifier ORCID. The ORCID ID allows for automated linkages between the researcher's identity + his/her research activity & output.

As for data protection, the data controller is normally the Ca' Foscari University with respect to the research projects conducted under the supervision of the University itself. The single researcher who acts for the University is the person appointed to personal data processing; notwithstanding this, he/she has to take any action on behalf of the University for respecting the obligations set out by the data protection law. Other subjects may be involved in the data collection, storage, dissemination, etc. (e.g. online platform for survey analysis): their specific role and responsibility have to be identified case by case (data controller, data processor, etc.)

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)?

Long-term access & maintenance data means having a plan that accounts for the next 5 – or more – years.

Describe

- where are you storing your data
- are you going to partner with Institutional resources or third party
- who is sponsoring it and what are the terms.

Part 5.
Data security

This section is strategic especially, but not only, if you manage 'sensitive data'. When managing data it is important to think about storage, backup and security.

Planning WHO will have access to WHAT is an important step also to prevent inappropriate uses. Indeed think that your personal device could become damaged and a good backup plan could save your research project.

Did you ever lost data??

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

During the research, raw data will be primarily loaded and maintained onto individual organisational computers for working purposes and regularly backed-up on a shared institutional server space with access restricted to only the team members. Data on individual computers will be removed following the end of the project …

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

Long term preservation for curated data is ensured by chosen, certified data repositories that observe specific preservation policies for long-term preservation (all data will be stored for at least 5 years).

Maybe your Institution has adopted one, otherwise we suggest Zenodo.

All data sets that are licensed with an open license will be preserved in Zenodo (with a policy on long-term preservation) and they will be available for use in the future. Sensitive data will be preserved in the internal project database, which has a long-term support of 5 years after the project ending.

Part 6.
Ethical aspects

Some researches involve human participant: you need to take care about safeguarding of any data that is being generated through and from them. The DMP will follow strict guidelines on ethical conduct involving human participants and this includes following standards set for Ca' Foscari researchers through «Codice etico di Comportamento». Outline in your DMP the steps you will take to protect research participants, e.g. data anonymisation; by negotiating informed consent for data collection. The University Ethics Committees provide sample consent forms.

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). (1)

Justify your decision to keep some (or all) datasets closed: report and complete this section with the relevant part of

https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-self-assess_en.pdf

To make data available it is essential to know who owns the IPR

See an example in the next slide

Are there any ethical or legal issues that can have an impact on data sharing? (2)

This research involves questionnaires, interviews and surveys with adults participants with key informants (selected practitioners and experts).

All aspects of collection of data involving personal data are covered by the Ethics Requirements document. All personal data collected from questionnaires, interviewers, surveys are carefully protected in compliance with relevant national data protection legislation of the EU member states implementing the GDPR and with the procedures defined by the Codice etico di Comportamento of the Ca'Foscari University of Venice that transposes the European Code of Conduct for Research Integrity.

Are there any ethical or legal issues that can have an impact on data sharing? (3)

As a general principle, personal data resulting from interviews and questionnaires will be separated from the research results, and will be handled by different members of the research team. In regards to the respondents in the survey, they will be selected at random and their name and address will not be recorded. The data will be stored in a away not to allow the identification of the subject, adopting measures for anonymisation (i.e. names replaced by initials or pseudonyms); results of questionnaires and interviews will be transmitted or made available to the other project partners as anonymous data or at least in a de-identified manner. However, personal data generated from interviews should be de-identified as soon as the purposes for which a researcher processes such personal data do not or do no longer require the identification of a data subject (for example, after the full transcriptions of the interviews).

A research may involve 'sensitive data' if strictly necessary for the purpose of the study (e.g. participants might disclose political opinions); this should be highlighted immediately at the time when information is provided to the researcher. Appropriate technical and organisational measures to ensure data protection of sensitive data must be implemented.

Are there any ethical or legal issues that can have an impact on data sharing? (4)

Files containing questionnaire data for statistical analysis, transcripts of interviews, photos, minutes, videos…, are stored in computers, laptops, intranets or hard-drives of the research institutions accessible through institutional password modified periodically (every 3 months in case of sensitive data), and protected by regularly updated antiviruses. Files containing "sensitive data" will be stored encrypted. Password-protected and encrypted files are accessible only to authorised members of the team. None of the project data will be left inadvertently available by being left on desks or in unlocked rooms.

All the research materials stored in computers are subjected to back up regularly (according with the institution's regulation). In order to safeguard them from accidental loss.

Are there any ethical or legal issues that can have an impact on data sharing? (5)

Data and information collected from questionnaires will be disseminated and published only in aggregate and/or anonymous form. Publications will only report aggregate data and shall not contain information that may permit the identification of individual participants. Data that are not shareable will be stored for the time required by … and will be subsequently destroyed. Where personal data are no more necessary for the research, they will be immediately destroyed. Qualitative data files can be accessible with public access as long as any information that can lead to identification of an individual participant is deleted.

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data? (1)

All those investigation (e.g. surveys, questionnaires, interviews) that involve human participants (internal and externals) will be handled according to the applicable data protection law.

This requires that everybody is being asked for their consent in relation to their involvement and all potential risks and benefits of their voluntary participation.

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data? (2)

- No individual will be identified by name in any published materials.
- Survey data will be stored securely in the server of the partner conducting the survey, following the national and international data privacy regulations.
- Selected material may be shared across the project team, by only by researchers for research purposes.
- The project will use pseudonymisation for sharing sensitive data internally.
- Any material shared externally or used for the dissemination of data will not utilise any information that could identify individuals and participants will be informed of this.
- Researchers involved in the project are aware about the importance of data protection and privacy issues in this kind of research.

Part 7.
Other issues

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

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

Credits

Data Monitoring Board of the Ca' Foscari University of Venice

IOSSG Italian Open Science Support Group