

# UEF Guidelines on Data Management for the Horizon 2020 funding applications

## INITIAL DMP

These guidelines can be found in [DMPTuuli](#) when choosing University of Eastern Finland as the primary research organisation and European Commission (Horizon 2020) as the funding organisation.

### 1. Data summary

Provide a summary of the data addressing the following issues:

- 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')?

## UEF Guidance

### Guidance on Existing Data

Briefly describe what kind of research data you will use, collect and produce in your project. Shortly outline how the data will be collected: e.g. literary survey, survey, interview study, case study, observation, measurement, collected by machine or instrument, laboratory work or programming.

Describe in short what types of data will be used and is expected to be produced: e.g. tables, texts, images, photographs, videos, statistics, diagrams, chemical or physical reactions, physical samples, sequences, codes, modeling or meta-analysis.

#### *Tips for best practices*

- Explain your methods, experimental arrangements and data content in more detail in the research plan.
- Clearly distinguish the data which is produced in this project from the study data that has been produced earlier.
- By reusing data produced by you or others, you will avoid duplicating work already done.

If you have any questions regarding your data management plan, please contact the UEF Library Research Support at [opendata@uef.fi](mailto:opendata@uef.fi).

### Guidance on Data Format

Your choice of file format is a primary factor in the accessibility and reusability of your data in the future. The format and software in which research data are created usually depend on how researchers collect and analyse data. Once data are prepared for storing, researchers should consider converting their research data to standard, open, non-proprietary and commonly used formats.

#### *Tips for best practices*

- List the file formats for every type of data that will be stored and shared in; e.g. excel tables as .csv, word documents as .txt and videos as .mp4.
- When listing the file formats you will be using, make sure to include any software necessary to view the data.

- Favour software and formats based on open standards to enable data reuse, interoperability and sharing.

### *Links to general guides*

- UEF guidance on open science, [Open UEF](#)
- [File formats and software](#), Finnish Social Science Data Archive

If you have any questions regarding your data management plan, please contact the UEF Library Research Support at [opendata@uef.fi](mailto:opendata@uef.fi).

## **Guidance on data management plan**

### *Why should you manage your research data and plan its management?*

Research data management and its planning (DMP) is an integral part of good research practices. By writing a DMP before your project starts you will help minimise unexpected problems.

The invaluable advantages of data management planning include e.g. the following:

- reducing the risk of losing data
- saving time and money
- meeting funder and policy requirements
- maintaining/ensuring data integrity.

A clearly outlined DMP will also help you to overcome complex ownership and user right issues in advance, and to support open access in order to promote new discoveries and productive future collaborations.

[Publishing and Data Policy of the University of Eastern Finland](#) states that a research data management plan must be made for the research at the planning phase. The research data management plan must specify the collection methods, processing, ownership, access rights, storage (including long-term storage), reuse, opening, publishing and planned destruction (if necessary) of the research data, as well as the required resources for these measures. Furthermore, the requirements set by the funder on the research data and the research data management plan are to be observed.

Your DMP should describe how you manage data during the whole research life cycle - and also cover what happens after the active phase of the project. The DMP is a living document which should be updated as the research project develops.

Good luck with your DMP!

If you have any questions regarding your DMP please contact the UEF Library Research Support at [opendata@uef.fi](mailto:opendata@uef.fi).

## **2. FAIR data**

### **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.**

## EC Guidance

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

## UEF Guidance

Your research data management practices should follow the [FAIR principles](#) which dictate how your data will be Findable, Accessible, Interoperable, and Re-usable. Data documentation enables data sets and files to be discovered, used, and properly cited. Documentation explains the terms, variable names and labels, codes or abbreviations used. Explain clearly and concisely how you will describe and document your data during the research project. Consider which documents need to be stored.

Plan how the data will be organized during the project:

- How will you name your files?
- How do you separate different versions of files?
- What is functional folder structure for your project?
- Who will have access to research data?

[Metadata](#) is standardized structured information about the data explaining the purpose, origin, time, location, creator of data collection as well as access conditions and terms of use. Metadata may also contain details about experiments, analytical methods and research context. Consider what information is needed to find, use, cite and interpret the data now and in the future.

Save metadata as a separate file that is readable also without the original data. [Metadata Standards Directory](#) (The Research Data Alliance) can be searched to find discipline-specific standards and associated tools. **If you will open your metadata or data, check the requirements for metadata of the data repository that you are planning to use and produce metadata according to those requirements.**

Consider relevant keywords to describe your data. Keywords are essential for finding your data in the future.

- Finnish subject headings ([YSA](#))
- Medical subject headings ([MeSH](#))

### *Links to general guides*

- [Data Management Guidelines: Data Description and Metadata](#), Finnish Social Science Data archive
- [Documentation and metadata](#), Helsinki University Library

If you have any questions regarding your data management plan, please contact the UEF Library Research Support at [opendata@uef.fi](mailto:opendata@uef.fi).

## **2.2 Making data openly accessible:**

- **Which data produced and/or used in the project will be made openly available as the default?**
- **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?**
- **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?**

### **EC 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.

The Commission recognises that there are good reasons to keep some or even all research data generated in a project closed. Where data 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.

Preference should be given to certified repositories which support open access where possible. The [Registry of Research Data Repositories](#) provides a useful listing of repositories that you can search to find a place of deposit.

### **UEF Guidance**

Ideally, eventually your data should be deposited to international or national data repositories where other researchers in your field can find, cite and reuse them. Find out if there are well-established data repositories that are used in your research field. There are widely recognized international general-purpose repositories, such as [Zenodo](#), [EUDAT](#) and [Dryad](#). The national service, [Etsin Research Data Finder](#), or similar service can be used for publishing metadata in order to increase the findability of your research data.

Search [data repositories](#).

Describe which data will be shared and how:

- Can all of your data be shared, or parts of it?
- Consider data sharing both during and shortly after the research project.
- Mention if specific software is needed to access the data.
- Linking to research products like publications and related research data creates a more complete understanding of the study.
- Describe how data and metadata can be found.

If your data or parts of it will not be shared, please explain why. Good reasons why data will not be shared might include confidentiality, trade secrets or ownership issues (license, copyright). Sometimes data can't be shared due to the unreasonable effort required for its sharing (e.g. big data volumes, legacy data). However, too often data can't be shared or reused in research, teaching and learning because of poor data management planning in the beginning of the research project.

According to [UEF data policy](#) research data relating to published findings obtained through public funding are, as a rule, open whenever this is possible in view of the agreements concluded (ownership, access rights, immaterial property rights and non-disclosure), legislation and the principles of research ethics, while not compromising the interests of the university. The discoverability and citability of research data must be ensured.

***Links to general guides***

- [Research Data Management: Sharing and reuse](#), Helsinki University Library

If you have any questions regarding your data management plan, please contact the UEF Library Research Support at [opendata@uef.fi](mailto:opendata@uef.fi).

### 2.3 Making data interoperable:

- **Are the data produced in the project interoperable?**
- **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?**

### EC Guidance

Inteoperability 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.

### UEF Guidance

Especially for data long-term storing and sharing, prefer file formats that are interoperable with several software applications.

#### *Links to general guides*

- [Recommended file formats for data sharing, reuse and preservation](#), UK Data Service
- [Metadata](#), UEF Library

If you have any questions regarding your data management plan, please contact the UEF Library Research Support at [opendata@uef.fi](mailto:opendata@uef.fi).

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

- **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 usable 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?**
- **Describe data quality assurance processes**

### EC Guidance

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

### UEF Guidance

#### *Tips for best practices*

- Consider when and for how long your data will be made available?

- Are your data available for all re-users or do you need to restrict it? If access to the data is restricted, explain how the identity of the person requesting the data can be verified (e.g. use of Haka identification, personal contacting).
- Who will own the data and who can issue permissions to reuse it?

Research collaboration and data protection may require transfer of rights of the results from the researchers to the university. For consortium projects, intellectual property rights (IPR) and ownership must be covered in the consortium agreement. If you use research material or data collected or produced by a third party, consider the copyright issues and potential licenses which may affect its distribution.

Any restrictions needed on data sharing must be outlined e.g. to protect proprietary or patentable data. These issues should be solved already at the planning stage of the research project. If ownership issues have not been considered early enough in the research life cycle, sharing and reusing the data may become impossible.

Principal investigator is responsible for concluding contracts on the ownership and user rights of research data at as early a stage as possible. Consider the relevant funder, institutional and departmental policy on copyrights or IPR.

It is recommended to use [Creative Commons](#) (CC), GNU, MIT or another relevant license.

The recommended CC licenses:

- Research data: CC-BY
- Metadata: CC0

Describe how you assure data remaining of high-quality and usable. Data quality control ensures that no data will be lost or accidentally changed during data collection, entry, digitization or checking. Quality control measures can include having clear instructions, using standardized methods, taking multiple measurements or samples and calibrating instruments.

#### *Links to general guides*

- [Licensing research data](#), Open UEF
- [Research Data Management: Legal aspects & research integrity](#), Helsinki University Library

If you have any questions regarding your data management plan, please contact the UEF Library Research Support at [opendata@uef.fi](mailto:opendata@uef.fi).

### **3. Allocation of resources**

**Explain the allocation of resources, addressing the following issues:**

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

#### **EC 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.

## **UEF Guidance**

Putting data into a usable format and making it meaningful to other researchers takes time and costs money in terms of software, hardware and personnel. Read your plan and make sure that there are rows in the budget to provide for the people who manage the data as well as paying for the required hardware, software and services.

### ***Tips for best practices***

- Consider the additional computational facilities and resources you need access to, and estimate the associated costs.
- Does your data need anonymisation? If anonymisation is needed, plan the work and related costs beforehand.
- How will the responsibilities for data management and costs be divided across partner sites in collaborative research projects?
- Take advantage of external free of charge services e.g. [cloud computing](#), CSC

If you have any questions regarding your data management plan, please contact the UEF Library Research Support at [opendata@uef.fi](mailto:opendata@uef.fi).

## **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?**

## **UEF Guidance**

### **Guidance on data security**

Describe how you access, handle, and store the data safely. Keep your research data safe and secure during the research project. Determine who has access to your data and what they are authorized to do with it. Providing unauthorized people with access to the data may be illegal. Access controls should always be proportionate to the kind of data and level of confidentiality involved.

The UEF IT Services produce most of the IT services for researchers, e.g. workstation services, network connections and server resources. Besides the IT services produced by the UEF, the researchers can use some external services, e.g. from CSC.

The UEF information security guides must be complied with in the management of research data. (See links below for more information).

### **Research data security**

The university's general [instructions on information security in information processing](#) also applies to research data.

Research data is divided into three different groups based on the confidentiality and publicity requirements:

- high protection level

- base protection level
- public information.

**High protection level** - information comparable to patient records or otherwise related to health conditions and other sensitive personal data (e.g. client relationship with social welfare), research subject's ex-directory phone number or contact information

**Base protection level**, for example:

- personal identification code and other similar personal data (e.g. contact information and family status)
- confidential information related to studies and study performance (e.g. test or examination results, student welfare, or exemptions granted from teaching)
- thesis or research plans, and technological or other development work
- research subject's annual income, net assets and other information related to a person's financial standing
- information security and other security and safety precautions undertaken in order to protect persons, facilities and IT systems
- procurement or tender documents until they enter the public domain
- information related to accounting or funding
- information released to the statistics authority for statistical purposes
- private business or professional secrets of third parties
- protected intellectual property rights (IPR) of the university
- business or professional secrets of the university
- recorded information and documents at the stage of internal preparation, drafts and memoranda. A higher level of protection can be applied to research data by one's own discretion or if the contract or commission requires it.

***Tips for best practices***

- The use of a personal or shared network drive enables you to control who can access and use your data. Furthermore, all data are backed up by the UEF IT Services.
- If you need larger volumes of storage space, please contact the UEF IT Services and describe your needs.

**Guidance on storage and backup**

As a researcher, you need to ensure that your data is stored properly both during and after the research process.

Consider who will be responsible for backup and recovery. If there are several researchers involved, create a plan with your collaborators and ensure safe transfer between participants.

**Research data storage services at UEF**

Personal storage space services

- For information at the base protection level and for public information: Office365 OneDrive for Business file space, 5 TB. Remote access with a browser or with mobile apps.
- For information at the high protection level: personal LocalDrive file space, 2 GB. Remote access only from a workstation that is administered by UEF.
- For research data at the high protection level and at the base protection level: researcher's disk space with backup, 50 GB, and researcher's disk space without backup, 2 TB. Remote access only from a workstation that is administered by UEF.

OneDrive for Business file space is automatically created for all users. Personal LocalDrive and researcher's disk space can be requested via e-Services.

#### Research project's or department's storage space services

- For information at the base protection level and for public information: Office365 workspaces for units/projects, no space limitation. Remote access with a browser or with mobile apps.
- For information at the high protection level: a LocalDrive file space for units/projects, 1–16 GB. Remote access only from a workstation that is administered by UEF.
- For research data at the high protection level and at the base protection level: research group's disk space without backup, disk quota negotiable. Remote access only from a workstation that is administered by UEF.

Office365 work spaces, LocalDrive file space and research group's disk space can be requested via e-Services.

Office365 OneDrive for Business and Sharepoint services can be used with a remote access connection through a web browser and mobile devices. The service providers' environments are fault-tolerant but the researchers themselves must take care of the backups and long-term storing of research data.

High protection level research data it-services are situated in the university's own server rooms and can be accessed from the workstations within the university network or from outside the university network through the VPN connection with workstations administered by the university. The UEF IT Services takes care of the daily backups of the material on the university's own servers.

Aforementioned services are sufficient for most researchers as the basic storage services of the research data but it is possible that they are not sufficient for all research purposes due to the features or capacity. If there are some special needs regarding the storage services, the UEF IT Services should be contacted in order to find a suitable solution.

If a researcher uses research data produced or stored by others, s/he must describe the key storing principles or refer to the data processing instructions of the organisation in question.

#### *Tips for best practices*

- The use of a safe and secure storage provided and maintained by your organisation's IT support is preferable during your research
- Do you have sufficient storage or will you need to include charges for additional services?

#### *Links to data storage guides and services*

- [UEF IT Services](#)
- [UEF Information Security](#)
- [Services for research](#), CSC
- [IDA Research data storage service](#), CSC

If you have any questions regarding your data management plan, please contact the UEF Library Research Support at [opendata@uef.fi](mailto:opendata@uef.fi).

## **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?

## UEF Guidance

### *Tips for best practices*

- Read the guidelines for [data protection and processing of personal data in UEF](#).
- Describe how you will maintain high ethical standards and comply with relevant legislation.
- Check whether an ethical review is required for your research project.
- Describe shortly how you will inform research participants about the research project and how you will obtain an informed consent from them. Does the informed consent used in your research project include provisions for data sharing and long-term preservation?
- If your data include personal or sensitive information and you plan to make anonymisation for it, describe your data anonymisation procedure in your data management plan.
- If your data or part of them cannot be shared, explain why. The reasons may include e.g. confidentiality issues, trade secrets or ownership issues (licence, copyright). In most cases, research involving patient data is confidential and cannot be made openly available.

### *Links to general guides*

- [Ethical reviews of research projects and studies at the University of Eastern Finland](#)
- [The Research Ethics Committee of the Northern Savo Hospital District](#) (in Finnish)
- [Toimitettavat asiakirjat, The Research Ethics Committee of the Northern Savo Hospital District](#) (in Finnish)
- [The Research Ethics Committee of the Northern Savo Hospital District](#), in English
- [Finnish National Board on Research Integrity](#)
- [National Committee on Medical Research Ethics, TUKIJA](#)
- [Informing Research Participants](#), Finnish Social Science Data Archive
- [Anonymisation and Personal Data](#), Finnish Social Science Data Archive

If you have any questions regarding your data management plan, please contact the UEF Library Research Support at [opendata@uef.fi](mailto:opendata@uef.fi).

## **6. Other**

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

If you have any questions regarding your data management plan, please contact the UEF Library Research Support at [openadata@uef.fi](mailto:openadata@uef.fi).