

FAIR data in the application for Horizon Europe

We at the CZU Library have prepared this guide to help you describe FAIR data in your application for Horizon Europe support. This guide contains information on what should be described in your application. The manner in which you ensure that your data is FAIR must be chosen on an individual basis.

What you will find in this guide:

What is FAIR data?.....	1
How do I create a description of FAIR data for my application?.....	1
Types of data/research outputs:	2
Findability of data/research outputs:.....	2
Accessibility of data/research outputs:	3
Interoperability of data/research outputs:	3
Reusability of data/research outputs:.....	4
Curation and storage/preservation costs:.....	4

What is FAIR data?

FAIR data was first introduced in 2016 in an article formulating the [FAIR guiding principles for scientific data management \(Wilkinson et al. 2016\)](#). The word FAIR is an acronym of four principles: Findability, Accessibility, Interoperability, Reusability. By observing these principles, the value of data can be maximized after its publication. Currently, many initiatives strive to make FAIR data the standard. **In your application for Horizon Europe, a brief description (max. 1 A4) must be provided as to how you will ensure that the data generated or used in the project is FAIR (see the Horizon programme guide, pp. 43-53).** As individual parts of FAIR data overlap and complement one another, there is no need to worry if any information is provided in more than one part.

How do I create a description of FAIR data for my application?

In the first step, we recommend reading this Guide for an overview of what should be remembered when describing the data. For inspiration, you should also consult any available Data Management Plans from other research projects - for instance, you can use [Public DMPs on DMPOnline](#) or description of [Zenodo FAIR principles](#) (please note that Data Management Plan and a description of FAIR data for the application are documents similar in content, but with some differences).

In the next step, you begin the actual creation of a description of FAIR data. You should try to describe data and metadata as best as you can. Give a thought to the selection of repository, licence, etc., in accordance with all the following points. If you are interested, we at the CZU Library can provide consultation in the area of FAIR data. However, we are not experts in all disciplines, but we can still provide a certain library insight and perhaps also pick up any maladies. **If you are interested in consultation, please contact us at openscience@lib.czu.cz.**

If you would like to learn more about FAIR data, you can visit the website of [CZU Library, web OpenAIRE](#), or look at the [Data Management Plan template for Horizon Europe](#).

Types of data/research outputs:

Provide a basic description of research data that will be generated in the project or used by you.

What should be described in the application?

- **What data will you use?**
 - Describe whether you will use data already generated or whether you will generate new data as a part of your project. If you use data that has already been generated, describe in this section how and why you want to reuse such data.
- **What is the format of the data?**
 - Describe the format of data that you will create or use in the project (e.g. you will work with a text, images, numbers - these must be assigned an appropriate format such as txt, csv, jpg, etc.)
- **Why do you need the data in the project?**
 - Explain why the given data is important to you, specify its anticipated size, origin as well as the potential for further use (i.e. data utility).

Findability of data/research outputs:

The key idea of FAIR data is the findability of data. The findability of data can be ensured by assigning [persistent identifiers \(PID\)](#) to data, which is also richly described with metadata. (Meta)data is available in an online searchable source (e.g. a catalogue or data repository). Importantly, metadata must be machine-readable.

What should be described in the application?

- **What PID will you assign to the data?**
 - Persistent identifiers ensure a permanent and unique identification of objects. A persistent identifier is, for example, ISBN, ISSN, URL or DOI or Handle. For data, a DOI or Handle format is the most commonly used PID, but there are specialized PIDs depending on the discipline.
 - Most repositories assign a PID to objects uploaded to the given repository. We therefore recommend that you check with the repository operator what PIDs they assign (see the section Accessibility of data). For instance, the repository [Zenodo assigns a DOI to uploaded data](#).
- **Authors must be clearly identified**
 - PIDs are also suitable for a unique identification of authors of data. It is advisable to mention in your application that data will be linked to [ORCID \(or another similar author identifier\)](#).
- **What metadata scheme will you use to describe the data?**
 - [Metadata](#) describes data and as such, help render data findable, easily citable and reusable. As a part of your application, you should describe what metadata schema will be used. There are not only general metadata schemas (e.g. [Dublin Core](#), [DDI](#)), but also discipline-specific schemas. If you are unsure as to what metadata scheme would be suitable for your discipline, visit the [DCC Metadata Directory](#), [RDA Metadata Directory](#) or [FAIRsharing](#).
 - Certain repositories inform on the websites what metadata scheme they use. For instance, [Zenodo](#) uses [DataCite's Metadata Schema](#), which complies with the communication protocol [OAI-PMH](#). Metadata is also available through the [public REST API](#).
 - Importantly, data must be richly described and optimized to ensure its potential reuse (both by people and machines - i.e. data must be machine-readable). In the project application, this fact is described by indicating what metadata schema will be used.
 - Metadata description is no spooky matter, it may look intricate at first, but in fact, you just need to follow the step-by-step instructions when uploading data to the repository (most often, no more is required than just to fill in the data meticulously in the pre-prepared form).

Accessibility of data/research outputs:

Data must be accessible under clear-cut conditions, using standardized communication protocols, both for people and machines. It should be noted that FAIR data does not mean that data must be open.

What should be described in the application?

- **Where and how will you store the data?**
 - It is fundamental to ensure security, accessibility and description of data. Definitely, you should not indicate here that you store your data on, e.g. Google Drive or OneDrive.
 - It is advisable to use repositories – either generic repositories, such as [Zenodo](#), [FAIRsharing](#), [Harvard Datavers](#), or discipline-specific repositories. Unsure which repository would be suitable for your discipline? Try searching the repository databases [Re3data](#) or OpenDOAR.
- **What licence will you assign to the data?**
 - Add brief information about the licences that you will use for data (see the section Reusability of data).
- **Will the data be open?**
 - According to the European Commission, FAIR data is driven by the motto “As Open as Possible, As Closed as Necessary”. This motto means that it is unnecessary to share all data, but at least a description of the data should be made available (metadata). If the data is not open, an explanation must be provided (e.g. the data is sensitive, etc.).
 - The openness of data is subject to the licence used by you – see the section Reusability. In the case of sensitive data, a note should be made of the approval of the [CZU Ethical Commission](#), or an establishment of a special committee that would evaluate any individual requests for access to the data.
 - If special software (e.g. newly developed software or unusual software) is required to access the data, specify the type of software or indicate that documentation will be created for the software. There is no need to state that you will use excel etc. for cvs files.

Interoperability of data/research outputs:

Interoperability of data will make sure that data can be integrated with other data, applications, work procedures, etc. Interoperability is achieved by ensuring that data and metadata correspond to recognized standards and formats typical of the given discipline.

What should be described in the application?

- **In what format/standard will you generate the data?**
 - Data should be described in understandable and open formats/standards. Ideally, a controlled vocabulary or thesaurus will be used: simply, the correct terminology of the given discipline (you should refer to an ontology or a specific thesaurus which you will follow).
- **What format/standard will you use to describe the metadata?**
 - The same rules described in the above paragraph apply here. Metadata should follow a certain, generally recognized standard (see the section Findability)

Reusability of data/research outputs:

FAIR data is data that is ready to be reused in future research (for example, for research review, loose follow-up to research, etc.).

What should be described in the application?

- **Documentation of data**
 - Describe how a systematic documentation of data will be ensured (description of codes and abbreviations used by you. An ideal procedure is to follow a globally recognized standard/format - see the section Interoperability).
 - A good idea is to create data documentation (e.g. README document, where everything essential for the work with data will be described).
- **Data integrity control**
 - Describe how data integrity and versioning will be checked – e.g. whether you will use [Github](#) or perhaps [Open Science Framework](#) or other tools and functions of data control.
- **Licensing**
 - Data must be assigned the correct licence to allow for the reuse of data. Since each data has different sharing methods and options, you need to find out under what licence you can share your data. [Creative Commons](#) licences are used most often. You should check requirements of finance providers – some of them specify what licences can be used.
 - Most repositories assign a certain licence automatically. When choosing a repository, you should therefore obtain information on the assigned licences.
 - Unsure about what licence to assign? Try [License Selector](#) or [Choose a CC licence](#).
- **Will the data be reused?**
 - If you already know that someone will reuse the data, you can mention this in this part.

Curation and storage/preservation costs:

The generation of FAIR data is a demanding activity in terms of time and finances. This part should contain a description of what costs of generation, storage and overall data management will be needed.

What should be described in the application?

- **Costs of generation and management of FAIR data**
 - When preparing a grant application, you should also consider the funds allocated to FAIR data management. Those funds are recognized as direct costs within the Horizon Europe programme.
 - Costs relating to the FAIR data generation and management include e.g. costs of data repositories, staff (Data Steward – data specialist, data librarians, lawyer – licensing specialist, discipline-oriented specialist), etc.
 - On determining the costs, you should provide both the amount and the period for which the costs will ensure data accessibility (e.g. if a fee for data storage in a certain repository needs to be paid, you should indicate the period for which the data will be accessible, or how the data accessibility will be solved after the project ends).
- **Responsibility for data management**
 - Provide the name of the person/persons who will be responsible for data management throughout the project. You can briefly describe their work positions and responsibilities.
 - If you create specialized positions for FAIR data generation and management within the project, those positions could be described in more detail here (e.g. a Data Steward position).
- **Data security and ethics**
 - Describe how data security will be ensured (see the section Accessibility). Where the data is ethically sensitive, provide information as to how this will be solved (see the section Accessibility).

Published by the Czech University of Life Sciences Prague within the project Office for the Support of International Projects
focused on Life Sciences within the European Research Area III.
(Ministry of Education and Science INTER-EXCELLENCE programme, sub-programme INTER-INFORM, project no.
LT120001)

© 2021 Czech University of Life Sciences Prague