Co-funded by the European Union

Ministry of Education, Youth and Sports of the Czech Republic

Guidelines for the Preparation of a Data Management Plan

for P JAC beneficiaries

version 1.0

"If you fail to plan, you are planning to fail." - Benjamin Franklin

Content

Introduction

The Ministry of Education, Youth and Sports of the Czech Republic (MEYS), as the provider of financial support/grants under the Johannes Amos Comenius Programme (P JAC), imposes, in the Specific Rules of certain calls, an obligation on beneficiaries to ensure the responsible management of research data in accordance with the FAIR principles. At the same time, beneficiaries are required to prepare and continuously update a **Data Management Plan** (DMP) that complies with the requirements of the respective call.

A DMP describes how research data will be handled during and after the project. It helps address all key aspects of research data management so that data is of high quality, secure, sustainable in a long term, and, where possible, accessible and reusable. It is a **living document** that should be regularly reviewed and updated. If certain information is not available at the beginning, it should be provided as the project progresses.

Within P JAC, the DMP should have a structured format, and its content must substantively correspond to the Horizon Europe DMP template. This template includes basic administrative information and seven key thematic areas accompanied by guiding questions that facilitate the preparation of the DMP. Beneficiaries are provided with a MEYS Data Management Plan template, which is aligned with the Horizon Europe template. However, the use of an appropriate digital tool for the preparation of the Data Management Plan is strongly recommended.

This **Guide** is intended for beneficiaries and for individuals who support researchers in the management of research data (e.g. research support staff such as data stewards). It contains practical recommendations and examples for the individual topics, in line with the requirements of OP JAK and with the Checklist for the Data Management Plan for P JAC beneficiaries, which—similarly to this Guide—has a purely methodological and non-binding character. Binding requirements are always set out in the legal act and in the relevant call, including the associated documentation forming annexes to the legal act.

How to read the guide

This Guide provides general information on the individual topics, as well as useful colour-coded information presented in boxes, the meaning of which is explained below. For the sake of clarity, the Guide distinguishes between mandatory requirements, recommended practices, and examples.

Guiding questions help identify what information should be included in the Data Management Plan. Do not answer the questions with a simple “yes” or “no”; instead, provide a brief explanation and relevant information for each topic. It is not required to answer every question. However, if a particular topic is not applicable to the project, this should be clearly stated in the Data Management Plan. As this is a living document, the level of detail may progressively increase and the information may be further refined over the course of the project.

Before answering the guiding questions, we recommend familiarising yourself with the policies, guidelines, and procedures related to research data management (RDM) that apply to your institution or discipline. If you need assistance, contact research data support services within your institution, such as Open Science staff, data stewards, or project support offices.

Why it is important
➢ Here you will find the reasons why a specific topic in the DMP is important.

Guiding questions
- *Here you will find guiding questions that are consistent with the Horizon Europe and MEYS DMP templates.*

Explanations of what to state in response to guiding questions
- Here you will find information on what is expected to be described for each topic.

Examples
Here you will find two types of examples: Type A in regular font, and Type B in *italic*.
- Type A: Examples presented as lists of items, for example types of data formats.
- *Type B: Illustrative examples of possible answers to the guiding questions.*

Attention! Data management must always comply with applicable legislation and with the research data management guidelines and procedures that apply at the research organisation(s) and in the jurisdiction(s) where the research is conducted.

In case of a consortium consisting of entities under different jurisdictions (i.e. international collaboration), each part of the project needs to comply with the legislation in the respective jurisdictions. Resolution of collision cases, and the governing law should be described in the consortium agreement to prevent confusion.

What is data

For the purpose of DMP, "**research data**" means information, other than scientific publications, in electronic form that is collected or generated in the course of research or development and is used as evidence in the research or development process, or that is generally accepted by the research community as necessary to validate the findings and results of research or development. Different disciplines may use different terms to describe both quantitative and qualitative data underlying a research project.

Examples of common data types across disciplines:
- measurements, observations, or instrument readings
- experimental data
- survey or questionnaire data
- interview transcripts
- field notes or ethnographic records
- clinical data or patient records
- imaging data (e.g. MRI, microscopy, satellite images)
- sequencing and genomic data
- simulation data or model outputs
- source code, scripts, or computational workflows

Depending on the discipline and the specifics of the research project, data may be archived or published in the form in which they were initially collected (raw data), or they may undergo substantial processing before becoming the final dataset (i.e. unprocessed vs. processed data). A **dataset** is a structured collection of related data, typically accompanied by metadata and documentation necessary for its interpretation, that can be archived or published as a unit.

Tools for data management planning

This section focuses on the tools available for preparing the Data Management Plan and their use in managing research data within the project.

Why it is important

- ➢ DMP tools facilitate collaboration among project members and research support staff.
- ➢ They allow users to provide answers easily, for example through structured questions with pre-filled options and recommendations, which simplify the completion of the DMP.

Beneficiaries of P JAC are recommended to use an appropriate digital tool for the preparation of the Data Management Plan. A wide range of such tools is currently available, for example Argos (OpenAIRE), DMPonline (Digital Curation Centre, UK), Data Stewardship Wizard (DSW), or FAIR Wizard (FW, commercial version of DSW). The **DSW** and **FW** tools are developed in the Czech Republic and can therefore provide a context aligned with the local research environment, for example by reflecting Czech funding bodies, the National Repository Platform (NRP), and applicable legislation. At the same time, they enable the creation of a Data Management Plan (DMP) using the MEYS model template and the Horizon Europe template, respectively. Due to their integration with the NRP, they are expected to allow for a high degree of automation in the future, thereby simplifying the preparation and management of DMPs[1]. However, the choice of a specific digital tool always depends on the requirements and needs of the respective institution.

Administrative information

This section covers general administrative information about the DMP and the research project to which the DMP relates.

Why it is important

- ➢ Contextual information about the DMP and research project is useful for anyone reading or evaluating the DMP.

Attention! For projects with particular information security requirements (e.g., dual-use research, protection of intellectual property rights, and commercial interests), the DMP itself may already contain sensitive information. If applicable, consult your support staff or IT department.

[1] For more information see https://dmp.eosc.cz

About the Data Management Plan

Information about the DMP itself must be clearly stated on the front page of the DMP document. This includes DMP title, version and date. The history of changes between versions should be documented (see the table of changes below).

DMP title [XXXX], e.g.: *Research Project Czech Context DMP*

The title of the DMP should clearly indicate the research project to which it relates.

DMP version and date
Version [X.X], e.g.: 1.0
Date [DD.MM.YYYY], e.g.: 01.01.2020

It is recommended to include both a date and a version number and to update them each time the DMP is revised.

DMP history of changes
It is recommended to include a table summarising the main changes between the different versions of the DMP. Describe substantial updates or modifications of broader topics compared to the previous version(s). Minor edits, such as clarifications of individual sentences or typographical corrections, do not need to be listed.

History of changes		
Version	**Date**	**Changes**
[1.0]	[DD.MM.YYYY]	[Original version]
[2.0]	[DD.MM.YYYY]	[Short description of the main changes, e.g.: added data formats, updated data size, clarified data access conditions and licensing]

Contributors

In this section, provide a list of all persons contributing to research data management within the project. It is recommended to clearly identify who is responsible for the DMP, and each DMP should include at least one contact person. If possible, provide the name of each contributor, their affiliation, role in data management, and contact details, together with relevant persistent identifiers.

Recommended information for each contributor:

- **Name and surname** (written as *Given name Family name*, e.g., *Jan Novák*)
- **Contact email** (institutional email addresses are preferred)
- **ORCID iD** (e.g., 0000-0000-0000-0000)
- **Role(s)** in the project (e.g., data steward, data collector, data curator)
- **Affiliation** (institution name and persistent identifier, e.g., ROR)

Example of a contributor:
Jan Novák
jan.novak@example.cz, ORCID: 0000-0000-0000-0000
Role: Data steward
Affiliation: Charles University, ROR: https://ror.org/024d6js02

To clearly identify contributors, it is recommended to use their **ORCID iD** (a free, persistent identifier for researchers and contributors). Registration and more information are available at: https://orcid.org/.

Affiliations should be provided at the institutional level. Using persistent identifiers from the Research Organization Registry (**ROR**) is also recommended (https://ror.org/).

When specifying **roles**, use the types defined by the DataCite Metadata Schema. Contributors may have multiple roles (e.g., data collector and data curator).

Attention! Some roles may have a different meaning in DataCite than they do in everyday usage. Before selecting a role, make sure you understand the definitions.

Examples of roles:
- Contact Person
- Data Collector
- Data Curator
- Data Manager
- Data Protection Officer
- Data Steward
- Distributor
- Editor
- Producer
- Project Leader
- Project Manager
- Project Member
- Researcher
- Rights Holder
- Sponsor
- Supervisor
- Work Package Leader
- Creator of DMP
- Other (if none of the roles above fit)

About the Project

In this section, provide information about the research project, such as the project title, start and end dates, project description, and project funding. Update this information whenever there are changes, for example, if the project end date is extended.

[Project name]

Project acronym: if applicable

Example: RPCC for a project named "Research Project Czech Context"

Project number: [XXXX]

Example: CZ.02.01.01/00/23_020/0008214

Project start date: [DD.MM.YYYY] – recommended

Project end date: [DD.MM.YYYY] – recommended

Funding: [XXXX] – Provide information about the research funding. It is recommended to select a funder from the ROR research funder registry.

Example: Ministry of Education, Youth and Sports, ROR: https://ror.org/04ydcpm48

Project abstract/description (recommended)**:** [Project abstract/description]

Providing a short project description helps to place the DMP in context. The abstract can be included in two ways:

- by linking to existing information in an external source (ensure the link points to openly accessible information), or
- by manually entering the textual project description in the text box.

1. Data summary

In this section, describe the research data that will be (or are already) generated or collected in the project. Specify the data types, formats, estimated size, their origin, purpose, and whether the data is newly generated or reused. This includes both data originating outside the project that will be reused for the purposes of the project and new data collected, captured, generated, or created by the project team.

Why it is important

- ➢ Defining what data will be collected or generated is essential for effective research data management.
- ➢ Estimating data size helps with planning storage, processing capacity, repository selection, and long-term preservation.
- ➢ The choice of data format affects accessibility, interoperability, and reusability.

Guiding questions

- *Will you **reuse** any **existing data** and for what purpose?*
- *What **types and formats of data** will the project generate or reuse?*
- *What is the **expected size** of the data that you intend to generate or reuse?*
- *What is the **purpose of the data generation or reuse** and how does it relate to the project objectives?*
- *What is the **origin/provenance** of the data, either generated or reused?*
- *Who outside the project could benefit from or reuse your data ("**data utility**")?*

Reuse of existing data

- Specify if **existing data** will be reused (if any). If yes, describe what data, including the source/repository, and provide a link to its persistent identifier (e.g. DOI) and licence, if possible.
- If the reuse of existing data has been considered but discarded, provide a brief explanation why.

Many research projects use existing datasets, digital records, or sources (such as material in public archives, media archives, legal resources, or large amounts of digital literature) to produce new outputs. Existing data can be used as a reference, combined with other data, or analysed with new research questions. They are often combined with newly generated project data. Existing data can come from other researchers or project teams, for example from previous projects.

There can be various reasons for not reusing existing data, such as lack of relevance, methodological differences, or insufficient quality. Identifying these reasons during project design helps justify the collection of new data.

Example of answers:
- *We will reuse publicly available survey data from the European Social Survey (ESS), accessed via the CESSDA repository. Data is available at [DOI:] under a CC BY licence.*
- *Existing genomic reference data will be reused from an established international database [Name, link]. Access is subject to specific conditions of use, which will be respected in accordance with the licence terms.*
- *Reuse of existing data was considered, however available data did not meet the required methodological criteria and therefore were not suitable for the objectives of the project.*

Data types and formats

- Specify the **types and formats of data** that will be generated or collected during the project. It is recommended that you indicate whether the format is **standard**[2] and **open format**[3], and suitable for long-term preservation.
- If you are using **proprietary**[4] **formats**, please provide a justification and a description of any strategies for converting the data to a more suitable open format.

Data formats refer to file formats, and their selection directly affects accessibility, interoperability, and reuse. For long-term preservation and sharing, it is advisable to deposit data preferably in standard, lossless[5], and open formats to ensure continued accessibility, interoperability, and sustainability over time (typically 5–20 years). Most repositories provide lists of preferred formats on their websites. Where standard formats are used, it is recommended to provide a link to a reference, e.g. via FAIRsharing.

Attention! Some standard formats are not necessarily suitable for long-term preservation. These include proprietary formats, lossy compression formats, and formats that require specific or expensive software. These formats are generally avoided because they can lead to data loss or become impossible to open as technology changes, limiting future access.

If **proprietary** or **discipline-specific formats**[6] are used, the DMP should justify their use and indicate whether open or preservation-friendly versions will be created for sharing and long-term preservation. It is not necessary to convert all files: original formats may be retained for raw data, while final datasets can be converted into more suitable formats for archiving.

Examples of data file formats suitable for long-term preservation:
- [Data type]: [Data format]
- File collections: TAR, GZIP, ZIP

[2] **Standard formats** are widely used and well-documented data format files commonly accepted within a discipline, supporting interoperability and long-term reuse.
[3] **Open formats** are non-proprietary formats that can be accessed and reused without requiring specific commercial software, making them suitable for long-term preservation.
[4] **Proprietary formats** are controlled by specific software vendors that may limit data access or reuse.
[5] **Lossless formats** preserve all original data during compression (ensuring that no data or quality loss will occur during file manipulation), e.g., JPEG 2000 is lossless compared to standard JPEG.
[6] **Discipline-specific formats** are data formats designed for specific scientific fields or specialised software and often require specialised tools for opening and analysis.

- Databases: XML, CSV, JSON
- Geospatial: SHP, DBF, GeoTIFF, NetCDF
- Video: MPEG, AVI, MXF, MKV
- Audio: WAVE, AIFF, MP3, MXF, FLAC
- Statistics: DTA, POR, SAS, SAV
- Images: TIFF, JPEG 2000, PDF, PNG, GIF, BMP, SVG
- Spreadsheet data: CSV, TXT
- Text: XML, PDF/A, HTML, JSON, TXT, RTF
- Web archive: WARC

- **Standard and open formats**, e.g.: CSV, TXT, XML, JSON, TIFF, PDF/A
- **Proprietary formats**, e.g.: PSD (Adobe Photoshop), SAV (SPSS), DWG (AutoCAD)

Sources and lists of long-term preservation formats:
- Confluence: File formats for archiving
- Digital Preservation Coalition: Digital Preservation Handbook
- MIT Digital Libraries File formats for long-term access
- Cornell Data Services: File formats for preservation
- ASEP Repository: Recommended file formats

Expected size of the data

- Estimate the **expected size of the data** to be produced in the project.

The **expected size of the data** can be expressed in terms of storage space required (kilo/giga/terabytes) and/or the numbers of objects or files, preferably for each file format separately. Providing only an overall estimate for the entire project is usually insufficient, as it does not clearly demonstrate that data management needs have been properly considered.

Example answers:
- *We expect that the overall data size generated during the project will not exceed 10 TB.*
- *Text document format (TXT): An open, standardised format suitable for long-term preservation. We expect to have approximately 0.5 GB of data to be stored in this format.*
- *Portable Document Format (PDF/A): A standardised format, suitable for long-term archiving. We expect to have 200 GB of data in this format.*
- *Microsoft spreadsheet file (XLSX). A standardised but proprietary format that is not suitable for long-term archiving. We plan to convert it to a suitable open CSV format by the end of the project. We expect to have 1 GB of data in this format.*
- Comma-separated Values (*CSV*): *An open, standardised format suitable for long-term preservation. Only a small amount of data is expected in this format.*

Purpose of the data collection/generation

- State the **purpose** of the data collection or generation and explain how it relates to the objectives of the project.

When creating a project, you should already have an idea about the types of data that will need to be generated or reused.

Example answers:
- *We generate new data on dissolution profiles of [drug name] to investigate its potential for the development of new formulations of an anticancer drug. Since it has anticancer abilities but has not been investigated in the open literature, this research aims to assess whether its dissolution profile complies with [country] pharmacopoeia requirements.*
- *We are reusing benchmark data for machine learning to evaluate a new image-classification model. Instead of collecting new images, we reuse the CIFAR-10 benchmark dataset, which already contains labelled images of animals and vehicles. We train our model on CIFAR-10 and compare its accuracy to other models that use the same dataset.*
- *We are generating new data through questionnaires to understand the effect of electricity bills on couples' decisions to have children. We combine these data with those previously gathered by our research team [add DOI/URL] on the cost of water consumption, in order to assess the overall impact of living costs on couples' planning decisions.*

Data origin/provenance

- Specify the **origin of the data** (**data source**), e.g., from your own or third-party devices, questionnaires, measurement equipment, etc. It is recommended to **list datasets that will be acquired.** If the datasets have already been created and published, it is recommended to provide a **list of these published datasets**, including dataset titles, assigned persistent identifiers (e.g., DOI), the repository, and the licence (if applicable).

The **origin of the data** (data source) refers to where the datasets come from and how they were or will be obtained. You should indicate whether the datasets are newly generated within the project (e.g., through experiments, surveys, measurements, or simulations) or reused from existing sources (e.g., repositories, databases, previous projects, or external providers). For reused data, the original source, access conditions, and licence should be indicated.

Data source/origin:
- a survey questionnaire (written or electronic),
- experiment or sensor,
- device/instrument (own or third-party),
- an existing dataset,
- a database or archive,
- a service center or research infrastructure,

- an organisation providing the data.

Example answers (datasets to be acquired):
- *RNA Sequencing Data from Human Lung Tissue Samples*
- *Microscopy Images of Cell Differentiation under Hypoxic Conditions*
- *Survey Data on Public Attitudes toward Climate Policy in Czechia (2024)*
- *Experimental Measurements of Thermal Conductivity in Composite Materials*

Data created/captured using equipment:
- *We will capture data using Fourier-transform infrared spectroscopy (FTIR) via a FTIR spectrometer Thermo Nicolet 6700. The measurements will be done by experts in the project using in-house equipment. The equipment is well documented and widely known. Calibration procedures and peer data review will be applied to ensure data quality.*

Non-equipment data/datasets/records:
- *We will use questionnaires and interviews. Personal data will be anonymised upon collection.*

Example answers (published datasets):
Dataset name: Clinical Trial Dataset on Treatment Response in Type 2 Diabetes
- *DOI: https://doi.org/10.1016/j.jmatsci.2025.04.012,*
- *General repository: Zenodo*
- *Licence: CC BY 4.0*

Data provenance (Fig. 1) describes not only the source of the data but also their origin, processing (including secondary data), transformations, and combinations. It represents a structured, metadata-based record capturing the complete history of a dataset, supporting traceability, accountability, reproducibility, and the assessment of data quality.

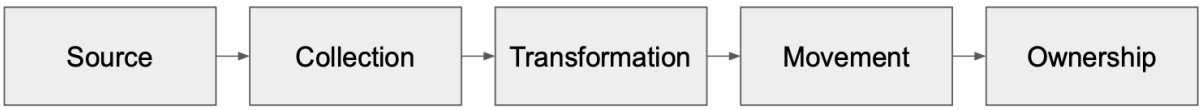

Fig. 1: Data provenance

Attention! **Data provenance** should be documented as comprehensively as possible, using appropriate standards where available. It is recommended to describe how data provenance information will be **documented** and how changes will be recorded when the data is modified or combined with other sources. Further details on how provenance will be documented should be provided in the Data documentation section.

Data utility

- Identify who may benefit from or reuse the data outside the project.

Estimating the **potential target audience** for data re-use supports the project's impact, promotes good data management practices, and helps clarify the purpose of data creation in relation to future users.

Example answers:
- Scientists in related disciplines can reuse the data to analyse greenhouse gas measurements in an urban landscape.
- The [specific] industry sector can benefit from the data created by the project, for example for product development or process optimisation.
- The [specific partner, e.g., hospital, public authority] can benefit from the data generated by the project for decision-making or operational purposes.
- Policymakers and other stakeholders can use the data to define, implement, monitor, and verify policies and climate-related actions.
- Citizens can create or use services based on a better understanding of the current state of the city or region in which they live.

2. FAIR Data

According to the rules of the respective calls, beneficiaries of P JAC research calls are required to manage research data in accordance with the FAIR principles, ensuring that the data is findable, accessible, interoperable, and reusable. The aim is to optimise data sharing and reuse by both humans and machines. **FAIR data do not necessarily have to be open to everyone**. Making FAIR data accessible follows the principle "as open as possible, as closed as necessary."

Attention! While the FAIR principles are domain-independent and can be applied to other research outputs (e.g., designs, software), the means of fulfilling the FAIR principles can be domain-specific. Despite this, some aspects are common across all fields, and these are described below.

2.1 Making data findable, including provisions for metadata (Findability)

In this section, describe how you will make your data and metadata findable. Specify which persistent identifiers (PIDs) will be assigned and which metadata standards will be applied. Findable (meta)data should be easy to discover for both humans and machines. The best way to achieve findability is to deposit datasets in a trusted data repository that assigns globally unique persistent identifiers (such as DOIs) and to provide as much contextual information (metadata) as possible when depositing a dataset into the repository.

Why is it important

- ➢ Without findability, data cannot be cited, reused, verified, or integrated into further research, thus limiting their impact.
- ➢ Persistent identifiers play a crucial role in ensuring the long-term findability and citability of datasets by providing stable, unchanging references.
- ➢ Using metadata standards improves discoverability by enabling search engines and repositories to accurately index datasets, making them easier to find and reuse.

Guiding questions

- *Will data be identified by a **persistent identifier**?*
- *Will rich **metadata** be provided to allow discovery? What metadata will be created? What disciplinary or general **standards** will be followed? If metadata standards do not exist in your discipline, outline what type of metadata will be created and how.*
- *Will search **keywords** be provided in the metadata to optimise the possibility for discovery and then potential reuse?*
- *Will **metadata** be offered in such a way that it **can be harvested** and **indexed**?*

Persistent identifiers

- Indicate whether your datasets will be assigned a persistent identifier (PID). If so, which one (based on the chosen repository).
- If persistent identifiers will not be assigned to the data, state why.

Persistent identifiers (PIDs) are permanent, unique references used to reliably identify and access people, organisations, and other objects (e.g., publications, research data) in a digital environment. PIDs support citation, discovery, linking, and long-term access to research outputs. When datasets are deposited in a repository, metadata is provided and a PID is assigned, usually in a form of DOI[7], Handle[8], or ARK[9].

Example answers:
- *All published datasets will be assigned a persistent identifier (DOI) by the [selected repository], ensuring long-term findability and citability.*
- *All datasets deposited in the Dataverse repository will be assigned a persistent identifier (DOI), ensuring long-term findability and citability.*
- *PIDs will not be assigned because the data is subject to contractual restrictions imposed by third-party partners, which prevent their deposit in a repository.*
- *Dataset [dataset name] will not be assigned a PID because the data contain highly sensitive personal information and cannot be deposited in a public repository.*

Metadata standards

- Specify **standards for metadata** creation (general or disciplinary) that you will follow. It is recommended to provide a name and link to the standard (e.g., from FAIRsharing). If there are no standards in your discipline, describe what metadata will be created and how.
- Outline the approach towards search **keywords** (if relevant).

Metadata is information about data or other research outputs. It can take the form of free text as well as structured, machine-readable information. Machine-readable metadata supports automation and discovery, while unstructured, human-readable text is used to understand the context of the data.

Metadata standards are agreed, structured frameworks for describing data consistently, enabling discovery, interoperability, and reuse across systems and disciplines. Consider which metadata standards or schemas are supported or required by the repository you have chosen for publishing your research data.

[7] **DOI** (Digital Object Identifier) is a persistent identifier used for digital objects (e.g. data, publications).
[8] **Handle** is a general system for persistent references to digital objects.
[9] **ARK** (Archival Resource Key) is used in digital archives for long-term data preservation.

General metadata standards describe basic information (such as title, creator, date, identifier, and format). They are universal and applicable across disciplines to support data discovery and citation.

Discipline-specific metadata standards are tailored to particular scientific fields and ensure the correct interpretation, validation, and re-use of data. It is recommended to use standards commonly adopted in the relevant discipline to ensure that data is described in a consistent and understandable way for other users.

The recommended requirements for describing research data in repositories (Tab. 1) are specified in the General Recommendation for Metadata Description (NTK, 2022; in Czech only). Ideally, the repository will have dedicated fields for this information. If not, these can be included in other appropriate fields, such as the abstract or notes section.

Tab. 1: Recommended minimum set of information (metadata) for describing datasets in a repository.

Required:	Recommended:
- **title** of the dataset - **creators** (i.e. authors and contributors, where possible with their affiliations) - **publisher** (institution providing data) - **date** of dataset deposit or publication date - **description** or abstract of dataset - **data access** (i.e. open, restricted, embargoed, closed; other access details) - **licence** of dataset (if applicable)	- **PIDs**: ➢ datasets, publications (e.g., DOI) ➢ organisations (e.g., ROR), etc. ➢ persons (e.g., ORCID), - **information about funding**: ➢ research funder (e.g., via ROR) ➢ project number - **classification** to scientific fields - **keywords** (preferably from a controlled vocabulary/ontology)

Examples of general and specific metadata standards
- Dublin Core – a widely used general metadata standard with basic descriptive elements
- DataCite Metadata Schema – a general standard for describing and citing research data
- DCAT – a specification for describing datasets in public/governmental data catalogues
- schema.org – a general web-based metadata vocabulary that improves data discovery
- BioSchemas – metadata profiles based on Schema.org for life sciences data
- DDI – metadata standard for documenting survey and social science data
- EAD – standard for describing archival and historical collections
- ISO 19115 – international standard for geospatial metadata
- EML – metadata schema for ecological and environmental datasets
- MIAME – minimum metadata requirements for microarray experiments
- CF Conventions – metadata conventions for climate and forecast data
- NetCDF – format and metadata standard for multidimensional scientific data

Examples of registries of metadata standards
- RDA (Research Data Alliance): Metadata Standards

- DCC (Data Curation Centre): Disciplinary Metadata
- FAIRSharing

Example answers:
- *The data will be stored in the Zenodo repository, which uses DataCite metadata schema.*
- *The metadata will follow the Dublin Core metadata schema for all datasets created with the use of instruments/equipment.*
- *All datasets will be described and organised according to the discipline-specific metadata standard Brain Imaging Data Structure (BIDS).*
- *Keywords are based on the ELSST thesaurus and topics created from the data repository will follow the CESSDA thematic classification.*
- *Since no appropriate metadata standards exist [for discipline], we will use the following metadata scheme to describe all datasets: Title; Description; Date; Creator; Rights/licence; Format; Volume; Experimental factors; Species; Observational Unit; Response variable; Technique; Experimental design [other].*

Metadata findability

- Indicate whether metadata in the repository will be searchable and findable, and whether their indexing and harvesting will be enabled[10].

Metadata harvesting and indexing are usually enabled by default in trusted repositories. It is recommended to first check if your selected repository supports metadata harvesting and indexing (e.g., through protocols such as OAI-PMH) by external services and discovery platforms. Most repositories provide this information on their websites.

Example answers:
- *The [Repository, link] supports metadata harvesting via standard protocols and assigns DOIs, ensuring that metadata can be indexed and discovered in search engines.*
- *Metadata will be searchable within the [Name of repository and link]; however, automated harvesting by external service*s is limited.

2.2 Making data accessible (Accessibility)

In this section, provide information on the accessibility of research (meta)data, which should be made available in a trusted repository as soon as possible, unless prevented by the principle of "**as open as possible, as closed as necessary**". Consider privacy, personal data protection, confidentiality, legitimate interests, third-party intellectual property rights, national security, or other justified interests. If some or all data cannot be made openly accessible, this must be clearly justified in the Data Management Plan, and the justification should be reviewed

[10] **Harvesting** is the process of systematically collecting/extracting (meta)data from one or more sources. This enables the integration of data from different repositories into central search platforms, e.g., OpenAIRE harvests metadata from European open access repositories. It is carried out using standardised protocols such as OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting).

on a regular basis.

Why it is important

➢ Accessibility ensures transparency, reproducibility, and compliance with funder policies.

➢ Accessible data can be retrieved under well-defined conditions, even if restrictions apply.

➢ Standardised protocols allow both automated systems and humans to access data.

Guiding questions

Repository:

- *Will the data be deposited in a **trusted repository**? Have you explored appropriate arrangements with the identified repository where your data will be deposited? Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?*

Data:

- ***Will all data be made openly available**? **If certain datasets cannot be shared** (or need to be shared under restricted access), **explain why**, clearly separating legal or contractual reasons from intentional restrictions.*
- ***If an embargo is applied** to give time to publish or protect intellectual property (e.g., patents), **specify why** and **how long this will apply**, keeping in mind that research data should be made available as soon as possible.*
- ***Will the data be accessible through a free, standardised access protocol**[11]? **If there are restrictions on use**, **how will access be provided to the data**, both during and after the end of the project? How will the identity of users accessing restricted data be verified? Is a **data access committee** needed (e.g., to evaluate/approve access requests to personal/sensitive data)?*

Metadata:

- *Will **metadata** be made **openly available** and **licenced**? If not, explain why.*
- *Will metadata contain **information enabling users to access the data**?*
- *Will metadata remain available even after the data itself is no longer accessible?*
- *Will **documentation** or **reference** about any **software needed** to access or read the data be included in metadata?*

Trusted repositories

- List repositories where the datasets will be stored. It is recommended to provide the repository name and link (e.g., from FAIRsharing, Re3data databases, if possible).

Trusted repositories to which you will deposit data may be disciplinary, institutional, national, or general-purpose. For the purposes of P JAC, a repository is considered trusted if it is listed in the above-mentioned registries, is certified (e.g. with the CoreTrustSeal), or is widely used

[11] **Free and standardised access protocol** is a publicly available, widely used technical method for accessing data (e.g. HTTP, HTTPS, FTP) to anyone without requiring proprietary software/payment.

by the research community. For the deposition of research data, the use of disciplinary repositories is recommended where they are available for the given field.

Information on repositories (e.g., certification, assignment of PIDs, access conditions, standards supported) can be found, for example, in the Re3data database. Additional practical information, such as accepted data types, formats, licensing conditions, access options, and any applicable fees, is usually available on the repository website.

Trusted repositories:
- Austrian NeuroCloud (ANC) – discipline-specific repository for neuroscience data
- ASEP Repository – institutional repository of the Czech Academy of Sciences
- Zenodo – a multipurpose repository maintained by CERN

Example answers:
- *All data will be stored in the institutional repository ASEP, which can assign DOIs to datasets, and allows the linking of other data and bibliographic records.*
- *Data will be stored in the [disciplinary repository, e.g., Hardware Dataverse], which provides certified long-term preservation, DOI assignment and adherence to FAIR principles.*

Data accessibility

- Specify which data will be made openly available. If there are any restrictions on access to the data, specify whether the restriction applies to all data or only to parts of it, and explain the reasons.
- Specify when the data will be made available for reuse. If applicable, explain why a data embargo is needed and for how long, particularly after the end of the project.
- Indicate whether any (documented) methods or software tools will be needed to access the data. It is recommended to specify which ones (including name and link).

Data accessibility describes how and under what conditions research datasets can be accessed and reused, including whether datasets are **openly available**[12], **embargoed**[13], shared with **restricted access**[14], or **closed**[15]. In P JAC, beneficiaries of research calls (see the specific call conditions and related documentation) are required to ensure open access to deposited research data in repositories, following the principle "as open as possible, as closed as necessary", preferably under a CC BY 4.0 or equivalent licence.

Attention! If (some) datasets cannot be shared openly, or can only be shared under restricted conditions, explain why. Specify whether the restrictions apply to all data or only

[12] **Open access** to datasets stored in a repository that are freely accessible to everyone via a free and standardised access protocol.
[13] **Embargoed access** may be applied to the publication of data. Datasets are not immediately openly available after deposition in a repository for a defined period of time.
[14] **Restricted access** to datasets is granted only under specific conditions by limiting data file access while enabling discovery.
[15] **Closed data** may be stored in the repository with metadata description, but data is not freely accessible to third parties ("closed access").

to parts of it (e.g., raw data, interview data) and whether it is due to legal/contractual reasons (e.g., privacy or intellectual property protection) or intentional restrictions (e.g., until the article is published).

There are many reasons why data may not be shared openly, including ethical and/or legal considerations such as trade secrets, intellectual property protection, security, or other considerations. In projects with multiple partners, specific partners may keep their data closed if making it public would conflict with their legitimate interests or other legal requirements. Restrictions should be supported, e.g., by a consortium agreement or agreements with project partners, especially for projects with multiple partners.

Example answers:
- *Datasets are openly available in the Harvard Dataverse repository under a CC BY licence. Users do not need to log in to access the data.*
- *Datasets are under embargo until the finalisation of our submitted publication that showcases the data. The embargo end-date is [DD.MM.YYYY]. After that date, the data will be publicly available under CC BY licence. The datasets are deposited in the Zenodo repository with DOI: https://doi.org/10.5281/zenodo.18991825.*
- *Datasets cannot become completely open. The dataset "MRI_Human_3" contains confidential information and will not be made openly available. It will be securely retained or deleted in accordance with the applicable ethical approval, legal requirements, and institutional retention policies.*
- *Dataset(s) [name] containing sensitive personal data will not be made publicly available due to the risk of harm to individuals. These data will be used solely for research purposes and subsequently destroyed or securely stored in an internal archive with restricted access until the end of the project.*
- *Data with personal information will be pseudonymised and shared under custom access terms. Access will be granted after communication with the contact person listed in the metadata.*
- *Dataset: [Dataset name], available in the Zenodo repository under CC BY 4.0, DOI: https://doi.org/10.5281/zenodo.15446419.*
- *The [name of dataset] contains BRML files that can be opened and read using the DIFFRAC.EVA software (DIFFRAC.EVA, Bruker).*

Metadata availability

- Specify if the metadata will be openly available.
- Indicate if the metadata will contain information on how to access the data (if applicable).

In accordance with the rules of the respective calls, **metadata of research data deposited in a repository** must be publicly accessible in compliance with the FAIR principles, to the extent permitted by legitimate interests or restrictions. Where access to the data is restricted,

the metadata should clearly indicate how a user may request access (e.g., if authentication[16] or authorisation[17] is required, and the procedure to request access).

Example answers:
- *Metadata of all records will be openly available under CC0 public domain.*
- *The metadata will remain accessible even if the deposited data is no longer available.*
- *Users requesting access to restricted data must justify the reason for their request.*
- *Requests for access to [personal or sensitive] data will be approved by a committee.*
- *The metadata will include clear information on access procedures, authentication, or authorisation requirements for datasets under restricted access (e.g., due to ethical or legal reasons).*

2.3 Making data interoperable (Interoperability)

In this section, describe how you will ensure that your data and metadata is interoperable. Interoperable datasets are data that can be compared, combined, and integrated with data from different sources. The best way to achieve interoperability is to use community standards and vocabularies relevant to your data type, choose a repository that allows linking or referencing to other related data, and store data preferably in standard, open (non-proprietary) formats suitable for long-term preservation and reuse.

Why is this important
- ➢ Data often need to be integrated with other datasets and must work with different applications or workflows for analysis, storage, and processing.
- ➢ Using standards increases (meta)data findability, interoperability, and reusability, allowing data exchange and reuse within and across disciplines.
- ➢ Different systems, tools, and researchers can interpret and integrate data correctly.

Guiding questions
- *What (meta)data **vocabularies, standards, formats or methodologies** will you follow?*
- *Will you follow **community**-endorsed interoperability **best practices**? Which ones? If you need to use uncommon or generate project-specific ontologies or vocabularies, will you provide mappings to widely used ontologies?*
- *Will your data include **qualified references**[18] to other datasets (e.g., other data from your project, or from previous research)?*

[16] **Authentication** is the process of verifying the identity of a user accessing the repository. It is all about telling a system who you - or your system - are.
[17] **Authorisation** is the process where the system is evaluating, if you are allowed to access a given resource. It can determine what actions or data the authenticated user is allowed to access or reuse.
[18] **Qualified reference** is a cross-reference that explains its intent, e.g., X is regulator of Y. The goal is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data.

Standards, formats, methodologies, ontologies, vocabularies

- Specify which data and metadata **standards, vocabularies, ontologies, or methodologies** you will follow to facilitate interoperability. It is recommended to provide a name and link to the standard (e.g., from FAIRsharing), where possible.
- If **community-accepted best practices** are used, specify which ones and **provide references**, if available.

Metadata standards and examples are described in the Metadata standards section.

Standard formats and examples are described in the Data types and formats section.

Controlled vocabularies[19]**, thesauri**[20]**, and ontologies**[21] are standardised collections of defined terms and relationships, usually specific to a research domain. Their use ensures consistent data description, supports data sharing and discovery, and improves interoperability. They also make it easier to understand and compare data across disciplines and language communities.

Community best practices: Many research communities have (meta)data domain-relevant community standards and/or "minimal information" standards (e.g., MIAME for microarray data) that define what information should be documented and shared to enable reuse and reproducibility.

Controlled vocabularies, thesauri and ontologies:
- CESSDA Vocabulary – a thematic classification for datasets in the social sciences
- ELSST – a thesaurus that supports consistent indexing of social science datasets
- MeSH – a hierarchical dictionary of medical terms
- AGROVOC – a vocabulary for describing concepts in agriculture and food sciences
- GO – gene ontology for describing biological functions, processes, and components
- UMLS – a biomedical ontology integrating multiple medical dictionaries
- SNOMED CT – a medical ontology covering diseases, procedures, and anatomy
- Library of biomedical ontologies – a platform with a wide range of biomedical ontologies

Example answers:
- *We will follow community-endorsed interoperability best practices by using established standards and vocabularies, including [metadata standard, e.g., DDI, ISO 19115, DataCite], [controlled vocabulary or thesaurus, e.g., ELSST, MeSH, AGROVOC], and standard file formats such as [CSV, NetCDF, TIFF]. No project-specific ontologies or vocabularies will be created.*
- *All datasets will be described using the following ontologies: Experimental Factor Ontology (EFO), Statistics Ontology (STATO) and Plant Ontology (PO).*

[19] **Controlled vocabulary** is a normative collection of terms used to describe and categorize information to ensure consistent labelling and minimize inconsistencies in terminology.
[20] **Thesaurus** is an extensive controlled dictionary that, in addition to a list of terms, contains hierarchical and semantic relationships between them (synonyms, superordinate and subordinate terms). It helps with searching and categorizing information.
[21] **Ontology** is a more sophisticated form of a structured model of domain knowledge, containing formal definitions of concepts and their relationships. It supports automated processing and linking of data between different systems.

- *For datasets [Name], suitable standard ontologies are not available. We will therefore create a project-specific controlled vocabulary and map it to the following standard ontologies: [Names].*
- *The following established vocabularies will be used: (GCMD) and (CF Conventions).*

To further support interoperability, it is important that data descriptions contain linked references using persistent identifiers to other related digital objects (e.g., publication, workflow, software, or source data) and to relevant entities such as authors, institutions, projects, and research outputs.

2.4 Increase data reuse (Reusability)

In this section, describe how you will make your data reusable. Data should be well-described and documented so that they can be understood, replicated, or combined in different contexts. Indicate what documentation will accompany your data and what is necessary for reuse. Specify the terms of use (licence) and the processes for ensuring data quality. The best way to achieve reusable data is to provide rich contextual information through comprehensive metadata when depositing datasets in a repository, apply an open licence (preferably CC BY 4.0 or an equivalent licence), and use community-accepted standards relevant to your data type and research field.

Why it is important

➢ Sharing data that can be reused by others is a main goal of the FAIR principles.

➢ It supports transparency and reproducibility and reduces unnecessary duplication of data collection.

➢ Reusable data can be applied in future research, innovation, and policy-making, increasing its value and impact.

Guiding questions

- *How will you provide **documentation** needed to validate data analysis and facilitate data reuse? Will the **provenance** of the data **be thoroughly documented** using appropriate standards?*
- *Will your data be openly available to permit the widest possible reuse? Will your data be licensed using standard **licences** for reuse? Will the data produced in the project be usable by third parties, in particular after the end of the project?*
- *Describe all relevant **data quality** assurance processes.*

Data documentation

- Provide information about the accompanying **data documentation** such as data provenance, data collection methodologies, data organisation, etc.

Data documentation includes various types of information (i.e. a detailed description of the data) that help to understand the context in which the data were generated, as well as the

structure and content. Appropriate documentation enables others to understand the data, reproduce results, and reuse the data in different contexts. Data documentation is typically provided through README[22] files, codebooks[23], data dictionaries, methodological descriptions, and analysis workflows, as well as other relevant materials. Data should be documented through all stages of the research data lifecycle.

What should be documented:
- How and why data have been collected, created, or modelled.
- How different data files and versions are organised.
- What changes have been made between different versions of data files.
- The meaning of codes, abbreviations, variable names, etc.
- Which software and software versions were used for data processing and analysis.
- Legal, ethical, or other restrictions that limit data reuse.
- Whether (and how) data have been reused in other research projects.

Example answers:
- *Documentation will be provided through structured metadata, README files, and data dictionaries deposited alongside the datasets in the repository. Variable definitions, codes, and abbreviations will be explained, and information on the software, tools, and versions used for analysis will be included. Data provenance will be documented using disciplinary and general metadata standards [which one].*
- *Data provenance will be documented in README files.*
- *All datasets will be clearly annotated with metadata that provides provenance information following best practice guides from the [e.g., FluxNet] community.*
- *Documentation describing data file structures, provenance, and workflows will be provided in [e.g., README file or codebook] in the repository.*
- *The README file will include references and links to the specific software and scripts needed to reuse or read the data.*

Data licencing

- Specify under which licence the data will be openly available to permit possible reuse, e.g., CC BY.
- Specify whether the data produced and/or used in the project is usable by third parties.

Licensing supports the lawful sharing and reuse of data in line with the “Reusability” principle of the FAIR principles. A licence does not transfer ownership of the data but grants permission for their use and sets the conditions for their further use. In most cases, a licence is assigned when data are deposited in a repository. Before applying a specific licence, its terms and conditions should be carefully reviewed.

[22] A **README** is a (text) file that provides basic information about the structure, components, data files, and associated metadata, including descriptions of folders and file organisation. It should preferably also contain extensive metadata and a clear description of the research methodology.
[23] A **Codebook** is a document that describes each variable in a dataset—its meaning, values, units, formats, and coding conventions. Similar to a README file, it serves as a reference guide that enables others to correctly understand, interpret, and use the data.

Within P JAC, research data should preferably be made available under the CC BY 4.0 or an equivalent licence. The CC BY licence allows anyone to share, adapt, and reuse the data, provided that the original author is properly acknowledged. Other licences may also be applied, such as Open Data Commons (ODC) that are specifically designed for databases and datasets.

Attention! Only **content protected by intellectual property rights** can be licensed, such as **copyrighted works** or **databases**[24] protected by the sui generis database right. Raw facts, measurements, transaction records, or descriptive values are generally not protected by copyright. Czech copyright law explicitly states that a "data as such" is not a work, nor are ideas, procedures, principles, or methods.

However, protection may apply to the creative structure of a database, the selection and arrangement of data, accompanying documentation, or visualisations. Protection may also extend to the database as a whole under the sui generis database right, which safeguards the investment made in creating the database. It is therefore advisable to assess, for each dataset, whether protection applies to the database structure, accompanying materials, or the database as an investment as a whole.

Further information on when research data are subject to copyright protection (creative content) or to the sui generis database right (protecting investment in database creation) is available in the overview provided here. Guidance on selecting an appropriate licence and defining conditions for the use of open data is available in the methodology Defining Conditions for the Use of Open Data (in Czech only) on the data.gov.cz portal.

Example answers:
- *All datasets will be made openly available under a CC BY 4.0 licence, allowing anyone to reuse, adapt, and share the data with attribution.*
- *Sensitive data will be under restricted access to protect intellectual property rights.*
- *Where possible, data will be openly available under a CC BY 4.0 licence. Some datasets containing personal or confidential information will be made available under controlled access with specific reuse agreements.*
- *Only fully anonymized data will be shared under a CC BY 4.0 licence.*

Data quality

- Describe the data quality assurance/control processes that will be applied to ensure that data is accurate, consistente, complete, and reliable throughout the project lifecycle.

Data quality assurance/control processes include planned and systematic actions. Such processes may be applied during data collection, processing, analysis, and storage. Typical quality assurance measures include instrument calibration, repeated sampling or measurements, standardised data collection and recording procedures, data entry validation,

[24] A **database** (for legal purposes) means a collection of independent works, data, or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means.

peer review or mutual data evaluation, and the use of controlled vocabularies or dictionaries to ensure consistency.

Example answers:
- *Experiments will be repeated, and results will be averaged, including the calculation of standard deviations to assess variability and reliability. Measurement instruments will be regularly calibrated to ensure consistency and comparability of data across the project.*
- *Standardised data collection protocols and templates will be used to ensure consistent data recording.*
- *Data entries will be validated through automated checks and manual review to identify inconsistencies or errors.*

3. Other research outputs

In this section, address the management of other research outputs produced within the project, in addition to research data. These outputs can often be managed, described, and shared in line with FAIR principles, including the use of persistent identifiers. Other outputs may be digital (e.g., software, source code, workflows, protocols, models, simulations) or physical (e.g., new materials, antibodies, reagents, samples).

Why it is important

➢ Sharing other research outputs contributes to transparency, reproducibility, and the reuse of research data and results.

Guiding questions

- *In addition to data management, which **other research outputs** will be generated or reused during the project, and how will they be managed?*
- *Consider which aspects of the FAIR principles can apply to other research outputs and **how these research outputs will be shared or made available for reuse**?*

- Specify which **other research outputs** will be generated in the project and describe how they will be managed and made available for reuse, in line with the FAIR principles.

Other research outputs represent results that are neither research data nor traditional scientific publications. For example, a project may generate research software[25], source code, scripts, computational models, and simulations related to research data. These may range from a few lines of code used for data analysis to a complex software package. Other outputs may also include workflows, protocols, laboratory notebooks, etc.

Example answers:

- The project will generate research software and analysis scripts used for data processing and analysis. These outputs will be version-controlled and documented, assigned persistent identifiers (where possible), and made openly available via a trusted repository [e.g., GitHub linked with Zenodo] under an open-source licence [e.g., MIT / GPL], ensuring findability, accessibility, and reuse.

- Standard operating procedures and computational workflows developed during the project will be documented and shared in a machine-readable format [e.g., PDF/A, CWL]. Where feasible, they will be deposited in a repository [e.g., Zenodo] with descriptive metadata and a clear licence to support reuse.

[25] **Research software** means source code files, algorithms, scripts, computational workflows and executables that were created during the research process or for a research purpose, as defined by a global RDA working group FAIR for Research Software. Take into account that for code/scripts, different licences are used than for data and publications.

4. Allocation of resources

In this section, explain the resources required for FAIR data management in the project. This includes estimating the costs related to FAIRification[26] of research data, data storage and long-term preservation, as well as defining roles and responsibilities for research data management within the project, including specialised personnel (e.g., data stewards, data curators).

Why it is important

- ➢ Clearly defining costs and responsibilities related to FAIR data management is essential for effective project planning and implementation of data management activities.
- ➢ Adequate resource allocation ensures that data can be properly managed, shared, and preserved beyond the lifetime of the project.

Guiding questions

- *What **costs are expected for making FAIR data** in the project, and **how will these costs be covered**?*
- *Who will be **responsible for data management** in your project?*
- *How will **long-term data preservation be ensured**? Discuss the resources necessary to accomplish this (**costs** and potential value, **who decides** and **how**, **what data** will be preserved and **for how long**?*

Costs related to data management

- Estimate the costs associated with the FAIRification of data and long-term preservation and describe how these costs will be covered. It is recommended to provide a cost title, amount, currency, and a brief description, including information on how these costs will be covered. In the case of individuals (e.g., data stewards), you can provide only FTE.
- If no additional costs are expected, clearly state this.

Costs related to research data management may include direct and indirect costs. Where relevant, the cost can be calculated by multiplying the price (e.g., per €/TB/year) by the volume of data expected to be generated and the number of years. The estimated costs should be included in the proposal when applying for research funding.

Costs related to research data management:

- facility access, equipment, and security clearance
- hardware or software (in addition to what is usually available in the institute)
- commercial software licences (or other regular fees required to work with data or documents)
- data storage fees (in the active phase of the project)

[26] **FAIRification** means making data or other resources FAIR, aka so that they follow the FAIR principles as much as possible.

- repository charges
- data preservation and curation after the project ends (e.g., long-term archiving fees)
- staff time (e.g., data stewards)

Example answers:
- *Long-term preservation of [X] TB for 10 years at [University Name] will cost about EUR 2,000. These costs will be covered by the [project budget].*
- *Storage and backup: EUR 2,000 for secure data storage over four years, including daily incremental and weekly full backups. Costs will be covered by the project budget.*
- *Software Licences: EUR 100, fees for specialised commercial software [XY] required to process and analyse project data; covered by the project budget. Additional costs of [X] CZK will be incurred, e.g., for database administration or software maintenance.*
- *No additional costs are expected. The chosen data repositories preserve the data for free. Required hardware/software is available at the institution.*

Responsibilities for data management

- Clearly identify who is responsible for data management and data stewardship in your project (e.g., individuals, departments, or teams). It is recommended to provide the name of the responsible individual(s), their roles and contact details, where possible. Contacts and roles of responsible persons/entities can be specified in the Contributors section.
- For collaborative projects, it is recommended to explain how data management responsibilities are coordinated across partners.

Responsibilities for data management describe who is accountable for planning, implementing, and overseeing research data management throughout the entire research data lifecycle. Responsibilities may be assigned to the principal investigator, project partners, data stewards, or other designated staff. It is recommended to clearly indicate, for example, who makes decisions about data access, publication, and long-term preservation.

Example answers:
- *We have a dedicated Work Package "Data harmonisation and integration" (WP4), responsible for harmonising all data pipelines within the project.*
- *The Principal Investigator is responsible for overall data management, while each partner institution manages and documents the data they generate.*
- *Data management responsibilities are shared within the consortium, with the project coordinator ensuring alignment with the DMP and funder requirements.*
- *Data steward [Name] is responsible for overseeing data curation, metadata documentation, and compliance with FAIR principles.*

Long-term data preservation (archiving)

Specify how long-term data storage will be ensured, including:
- Which **datasets** are to be long-term preserved and for **how long**.
- **Who decides** which data to keep and **how** the decisions will be made.

Long-term preservation applies to all data identified as valuable that need to be preserved (e.g., up to 10 years after the end of the project). Preservation should be in line with legal requirements, institutional policies, or disciplinary standards.

Example answers:
- *All generated datasets will be stored for the long-term. The preservation period is 10 years for all underlying data that form the basis for a published scientific article. The decision on long-term storage is taken by the Principal Investigator in accordance with institutional policies.*
- *All relevant data for reuse and experimental replicability will be preserved for 10 years after the end of the project.*
- *Dataset [Name] will be deleted after [X] years because of the following [contractual, legal, ethical…] reasons: [explain].*
- *Only processed and anonymised datasets will be preserved for 10 years, while raw data will be retained for 5 years and then securely deleted.*
- *Final curated datasets will be stored in a trusted repository for a minimum of 10 years after project completion.*

5. Data security

In this section, describe how research data will be protected during and after the project, including secure storage and archiving solutions, access controls, data transfer methods, and backup and recovery procedures. Indicate how sensitive[27] or confidential data will be handled and whether trusted repositories/archives will be used for long-term preservation to ensure data integrity, confidentiality, and availability.

Why it is important

- ➢ Ensuring that your data is safe is crucial to any research project.
- ➢ A good storage, backup and recovery strategy will help prevent potential data loss.
- ➢ Security provisions protect research data from unauthorised access, loss, and accidental deletion, or misuse.

Guiding questions

- *What provisions are or will be in place for **data security**, including data recovery, **secure storage/archiving** and secure transfer of **sensitive data**?*
- *Will the data be **safely stored** in trusted repositories **for long-term preservation** and curation?*

Storage and backup (during the research process)

- Describe where the data will be stored and backed up during the research activities to protect data against loss, damage, or unauthorised access (e.g., institutional servers, secure cloud services).

Secure data storage refers to locations where data are safely stored and regularly backed up during the course of the project. For active ("hot") storage[28], it is preferable to use robust, managed storage solutions with automatic backup, such as systems provided by the IT support services of your institution or faculty/unit departments.

Data backup is the process of creating a duplicate copy of data in digital form and storing it on a separate device to ensure its preservation and to prevent data loss. It is recommended to store data in at least two different locations.

It is **not recommended** to store data on unsecured laptops, unencrypted external drives, or USB devices that can be easily lost or damaged. If portable devices are used, it is important to ensure that backup copies exist, e.g., on network drives and/or in dedicated backup storage.

[27] **Sensitive data** is information that should be protected against unauthorised disclosure because unauthorised access may negatively affect the privacy of an individual, trade or business secrets, or even security.
[28] **Hot storage** is for active, frequently accessed data needing fast speeds (e.g., primary storage).

Example answers:
- *All data will be stored on the institution's secure servers with regular automated backups and controlled access (institutional login, role-based permissions).*
- *Small digital files containing confidential data will be stored on a secure network drive [X], hosted and maintained by the institutional IT department. Large digital files [> XXX GB] will be stored on an external cloud platform [X], contracted by our research department. Data recovery procedures will be in place to restore data in case of loss or system failure.*
- *For secure data storage, we will use [name of the facility, e.g., CESNET storage], which includes access restrictions (e.g., passwords, encryption, and access control) as well as secure disposal of data/storage devices that are no longer needed.*

Data security/protection

- Describe how **research data security** and the protection of sensitive data will be ensured during the research. If sensitive and/or personal data (e.g., personal data, trade or business secrets) is processed, describe how you plan to protect the data from **unauthorised access**.
- Explain whether there is a **managed access procedure** in place for authorised users of protected data and who should have access to the data (if applicable).

Security measures protect research data during the project and after its completion and include, for example, secure storage, access control, secure data transfer, and data backup and recovery. Sensitive data, in particular personal data, may require anonymisation[29], pseudonymisation[30], or encryption[31], or additional protective measures. When addressing data security, it is advisable to collaborate with cybersecurity experts at your institution. If such expertise is not available, your IT department may be able to provide support.

Example answers:
- *Access to raw or sensitive data will be restricted to authorised project team members only.*
- *Personal data will be pseudonymized or anonymized as soon as possible as part of the workflow. Direct identifiers will be stored separately with restricted access.*
- *Data collected from interviews and questionnaires will be anonymised and will contain no identifiable personal information.*
- *Personal or sensitive data will be encrypted during storage and transfer.*
- *We will use dedicated secure virtual research environments where project partners have controlled access to sensitive data (e.g., SensitiveCloud environment).*
- *Transfers of sensitive data will be carried out only via secure, encrypted channels (e.g., SFTP or VPN), and no sensitive data will be shared via email or unsecured cloud*

[29] **Anonymisation** of personal data for preservation and/or sharing (truly anonymous data is no longer considered personal data).

[30] **Pseudonymisation** of personal data (the main difference with anonymisation is that pseudonymisation is reversible). Unlike anonymisation, pseudonymisation is reversible. It relies on a separately stored key or information that enables re-identification of the individual.

[31] **Encryption** is considered a special case of pseudonymisation facilitated through cryptographic transformation. The encryption key must be stored separately, e.g., by a trusted third party, to ensure that the data cannot be accessed or decrypted without authorisation.

platforms.

Attention**!** Sentences such as "*Internal data security will be managed by the participating organisations according to their internal policies/guidelines and GDPR requirements"* are too general. Try to be more specific.

Long-term data preservation (archiving)

- Provide information on where the datasets will be archived (i.e. stored for the long-term) after the end of the project (e.g., in a trusted repository, archive, or cold-storage).

Research datasets may be securely stored for **long-term preservation** (archiving) through trusted repositories or secure storage systems that ensure access control, encryption, regular automated backups, integrity checks, and data recovery procedures. These measures ensure that data remain accessible, reliable, and protected over the long term. For example, **cold storage**[32] is a method of storing data in a low-cost, long-term, offline or near-offline system. It is used for archiving data that need to be retained but are accessed only infrequently, such as older project files or regulated research data that must be preserved for **5–10 years** or longer.

Example answers:
- *All datasets will be archived for the long-term in the [ARCHIVE central repository] of [University Name], which provides managed and secure research data storage.*
- *Data backup and long-term archiving will be ensured through cold storage services provided by CESNET.*
- *Long-term archiving will be provided by storing data in trusted disciplinary repositories (as described in the Trusted repositories section).*

[32] **Cold storage** is for archival, rarely used data, prioritizing low cost (e.g., tapes) with slower retrieval.

6. Ethical and legal aspects (Ethics)

In this section, address possible ethical and legal issues related to research data management and sharing, including how these aspects will be considered, where applicable. Take into account that national legislation, institutional guidelines, and international norms may have implications for the handling of research data.

Why it is important

- ➢ Ethical and legal considerations may affect data sharing, particularly when projects handle sensitive, personal, or otherwise protected data.
- ➢ Often there is a need to balance availability and openness with confidentiality.

Guiding questions

- *Are there, or could there be, any **ethical or legal issues** that may impact data sharing? These can also be discussed in the context of the **ethics review**.*
- *Will **informed consent** for data sharing and long-term preservation be obtained in questionnaires dealing with personal data?*

Ethical aspects

- Indicate whether there are any **ethical aspects** (e.g., experiments on humans or animals, dual-use) that may impact data sharing.
- Specify if an ethics review (e.g., by an ethics committee) is required for the project.
- If any **personal data will be collected or processed**, state how compliance with applicable laws is ensured (e.g., by obtaining **informed consent** in questionnaires, or by applying **anonymisation**).
- Provide a justification if ethical issues are not applicable or were not considered in the project.

Ethical aspects are essential in all forms of research. Primary research may involve direct interaction with participants (e.g., surveys, interviews, or experiments), while secondary research involves the analysis of existing data. In both cases, researchers must adhere to ethical principles such as transparency, integrity, and the responsible use of data. It is essential to carefully assess potential risks and harm. The implications of sharing research data must be considered in line with the principle “as open as possible, as closed as necessary.”

Ethical aspects:

- Use of personal data (e.g., data identifying individuals)
- Sensitive data (health, genetic, biometric, ethnic, political, etc.)
- Data involving vulnerable groups (children, patients, marginalised communities)
- Ethical restrictions from informed consent (what participants agreed to)
- Risks of harm, misuse, or stigmatisation if data is shared openly

- Data that can affect the economic interests of communities or individuals.

Example answers:
- *No ethical or legal issues are foreseen that would prevent the open publication of the research data. The data do not contain personal or sensitive information and are not subject to intellectual property or confidentiality restrictions.*
- *Our project involves working with personal data; therefore, informed consent will be obtained from all participants. Informed consent specifies that anonymised data may be shared with third parties for research purposes.*
- *A letter explaining the purpose, methodology, and dissemination strategy of the research, including plans for data sharing, and a consent form (including consent for data sharing) will be prepared. Each respondent will also receive a clear verbal explanation.*
- *We work with data relating to the geolocation of animals that are endangered or close to extinction. For this reason, our data cannot be fully open, to avoid further endangering the species.*

Legal aspects

- Indicate whether there are any **legal aspects** that may affect data sharing. If so, explain how they will be addressed.
- If no legal restrictions apply to the data, explicitly state “no legal restrictions apply to the data”.

Legal aspects that may impact data sharing include rules and obligations that determin how research data may be collected, stored, shared, and reused. The sharing of research data must be assessed in accordance with the principle “as open as possible, as closed as necessary”.

Legal aspects that affect sharing include:
- Data protection laws (e.g., the GDPR), see above in ethical aspects
- Confidentiality and secrecy obligations
- Intellectual property rights (copyright, database rights)
- Third-party rights (data owned or co-owned by others)
- Contractual obligations (e.g., with industrial partners)
- Export control or national security regulations

Example answers:
- *Data from interviews and questionnaires will be anonymised; if anonymisation is not possible, access to the data will be restricted under clearly defined conditions (e.g., through a data use agreement). Metadata will be shared openly.*
- *Some data provided by the company [company name] are subject to non-disclosure agreements (NDAs) and therefore cannot be shared openly.*
- *Commercially sensitive data from industrial partners [partner names] may limit open sharing of raw data; only aggregated or partner-approved data may be shared.*
- *The project may use copyrighted materials from external databases [database names],*

whose licences do not allow redistribution; therefore, only derived data, metadata, or references to the original sources will be shared.

- *Data co-owned by multiple institutions [institution names] may only be shared on the basis of a joint decision by the partners, in accordance with the consortium agreement.*
- *Datasets obtained from external providers [provider name] may be subject to specific terms of use and may not be further redistributed; however, metadata may include links to the original sources.*
- *Data sharing may be delayed to protect confidential business information or potentially patentable results arising from collaboration with industrial partners [partner name].*
- *Some technical datasets related to dual-use technologies [field/area] may be subject to export control regulations and may require a compliance assessment prior to international sharing.*

7. Other

In this section, specify whether additional policies, procedures, and/or guidelines (e.g., institutional, community or national, or sector-specific) apply to your research data management practices.

Why it is important

➢ Identifying relevant policies or guidelines helps ensure compliance with, e.g., institutional or discipline requirements and provides clarity on data management procedures and rules that need to be followed.

Guiding questions

- *Will you follow any **other national, sector-specific, or departmental data management procedures**? If yes, which ones? Please list and briefly describe them.*

Policies and guidelines for research data management

- Refer to any other national, domain or sector-specific, or departmental procedures for data management that are applied in your project, if applicable. It is recommended to provide the name of the policy (e.g., institutional research data policy), a link, and a brief description of its content.

Research organisations typically have their own policies or guidelines for research data management, often as part of broader Open Science policies or frameworks. If your research unit or discipline has specific guidelines, it is advisable to follow them and refer to them in this section. This section may also include information on the tool used for preparing the Data Management Plan.

Data management policies and guidelines:

- We will adhere to the following institutional procedures at Charles University:

➢ *Research Data Policy – specifies the basic principles of research data management.*

➢ *Methodological Guidance on Data Security – describes the rules and requirements for research data security.*

Attention! Sentences such as *"The project uses the procedures and recommendations of the Czech initiative EOSC-CZ" or "The participating organisations also apply their own institutional guidelines and best practices for research data management."* are too general. Provide specific references and links where possible.

Example answer (DMP tool):

- *This Data Management Plan was created with [tool name, link to DMP], based on the [name of the DMP template used, e.g. Horizon Europe] template."*

Checklist for the Data Management Plan for P JAC beneficiaries

This checklist helps the beneficiaries to verify whether all key areas are included in the DMP. To create a DMP, it is recommended to use a suitable digital tool, such as DSW or FAIR Wizard, which allows the generation of a DMP in the required template (Horizon Europe, MEYS if available).

Data Management Plan

- DMP version
- Date of the last DMP update
- Change history table (version, date, and description of changes)

Project information

- Project title
- Project acronym (if applicable)
- Project number
- Funding information (recommended)
- Project duration (start/end date) (recommended)
- Project abstract/description (recommended)

1. Data summary

- Reuse of existing data
- Data types and formats
- Expected size of the data
- Purpose of the data collection/generation
- Data origin/provenance
- Data utility

2. FAIR data

2.1 Findability

- Persistent identifiers
- Metadata standards
- Metadata findability

2.2 Accessibility

- Trusted repositories
- Data accessibility
- Metadata availability

2.3 Interoperability

- (Meta)data standards, formats, methodologies, ontologies, or vocabularies

2.4 Reusability

- Data documentation (and metadata)
- Data licencing
- Data quality

3. Other research outputs

- Software, workflows, protocols, models, etc.

4. Allocation of resources

- Costs related to data management
- Responsibilities for data management
- Long-term data preservation (archiving)

5. Data security

- Storage and backup (during the research process)
- Data security/protection
- Long-term data preservation (archiving)

6. Ethical and legal aspects (Ethics)

- Ethical aspects
- Legal aspects

7. Other

- Policies and guidelines for research data management (name, link, a short description)

References

This Guide for the DMP preparation for P JAC beneficiaries was prepared mainly using the following resources:

- **OpenAIRE:** How to comply with Horizon Europe mandate for Research Data Management. *Used under the following terms: Unless otherwise indicated, all materials created by OpenAIRE are licensed under a CC BY 4.0 licence.*
- **Science Europe**: Practical Guide to the International Alignment of Research Data Management. *Used under the following terms: ©Copyright Science Europe 2021. This work is licensed under a CC BY 4.0 licence (except for logos and any content marked with a separate copyright notice).*
- **Elixir-Belgium RDM Guide**: Make your Data Management Plan come true. *Used under the following terms: RDM Guide by ELIXIR Belgium is licensed under a CC BY-SA 4.0, except where otherwise noted. All material under this licence can be freely used, as long as the ELIXIR Belgium RDM Guide is credited as the author.*
- **Norway (plan.research-data.no)**: DMP Support Package for Norwegian Higher Education Libraries. *Used under the following terms: except where otherwise noted, plan.research-data.no is licensed under a CC BY 4.0 licence.*
- **UK Data Service**: Research data management. *All material on this website is copyright of the UK Data Service. Duplication or sale of all or any part of it is not permitted, except that material may be duplicated for personal research use or educational purposes in electronic or print form.*
- **Swedish National Data Service**: Checklist for data management plans. *Used under the following terms: Information texts, presentations, and digital documents created by SND have a CC BY 4.0 licence.*
- **European Commission**. *Data Management Plan Template (Horizon Europe)*. The work was created using adapted parts of the Data management plan (HE): V1.1 (01.04.2022), *under the following terms: Commission Decision of 12 December 2011 on the reuse of Commission documents. Content owned by the EU on this website is licensed under the CC BY 4.0 licence.*

Document title: Guidelines for the Preparation of a Data Management Plan for P JAC Beneficiaries

Issued by: Ministry of Education, Youth and Sports of the Czech Republic (MEYS)

Editing and revision: Text prepared and edited by the Ministry of Education, Youth and Sports of the Czech Republic.

Grammar and text flow refined with assistance from ChatGPT and Copilot.

We gratefully acknowledge Eva Hnátková, Georgia Koutentaki, EOSC CZ and the National Library of Technology (NTK) for their expert contributions, helpful comments, and feedback.

Year and place of publication: Prague, 2026

Version: 1.0

DOI: https://doi.org/10.5281/zenodo.20342784

Recommended citation:
Ministry of Education, Youth and Sports of the Czech Republic (MEYS). (2026). Guidelines for DMP Checklist for P JAC Beneficiaries. Prague. Version: 1.0. DOI: https://doi.org/10.5281/zenodo.20342784.