

# OPEN SCIENCE

## RESEARCH DATA MANAGEMENT PLAN (DMP)<sup>1</sup>

<b>Project number</b>	
<b>Funder</b>	National Research, Development and Innovation Fund
<b>Project title</b>	Motivations behind housing choices
<b>Principal investigator (PI)</b>	
<b>PI ORCID identifier</b>	
<b>Beneficiary institute</b>	
<b>Project starting date</b>	
<b>Project end date</b>	
<b>Project duration</b>	
<b>Data manager/contact</b>	
<b>DMP 1<sup>st</sup> version/date</b>	
<b>DMP last update<sup>2</sup>/date</b>	

**SUMMARY** (*dataset<sup>3</sup> reference and name; origin and expected size of the data generated/collected; data types and formats*)

Provide a summary of the data addressing the following aspects:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

The project aims to better understand the motivations behind housing choices. To this end, it carries out surveys and interviews with the public and decision-makers, at local and national level.

The methodological mix applied by the project include both quantitative and qualitative methods:

- multivariate statistical analysis of publicly available data
- surveys
- interviews
- focus group research
- qualitative content analysis.

<sup>1</sup> Template for the Open Science Research Data Management Plan (DMP). The sections should describe how you plan to make the project data Findable, Accessible, Interoperable and Reusable (FAIR).

<sup>2</sup> DMP is to be regularly updated.

<sup>3</sup> Several datasets may be included into a single DMP.

Types and formats of data generated/collected:

Type	Format
databases from secondary data	.csv
databases from surveys	.sav, .dta
audio recordings of interviews	.mp3, .mp4
transcripts of interviews	.pdf/A

Expected size of the data: between 1 MB and 10 MB

It might be useful to

- Researchers
- Research communities
- Decision makers
- Education
- The general public
- Media and journalists

The project will manage data in accordance with the principles of FAIR data management — that is, Findable, Accessible, Interoperable and Reusable data. The project aims to maximise access to, and reuse of research data it has generated.

**Each of the following six issues should be addressed with a level of detail appropriate to the project. Some guiding expressions with explaining guidance help in elaboration. Please note that not all guiding thoughts are to be taken into consideration, depending on the project.**

### **1. MAKING DATA FINDABLE** (dataset description: metadata, persistent and unique identifiers e.g., DOI)

Making data findable, including provisions for metadata:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

Guidance:

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

After the research is completed, the research data will be uploaded to the ARP Data Repository:

<https://repo.researchdata.hu/>

Metadata of deposited data in the ARP Data Repository will be open under a Creative Commons licence or other licences (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles (in particular

machine-actionable) and provide information at least about the following:

- handle and – if the depositor also requests a DOI- Digital Object Identifier (DOI)
- research project/dataset (title; short abstract; keywords; data collection period/collection date; temporal coverage; date of deposit,
- author(s); ORCID(s);
- data type, data collection method (interview / focus group / fieldwork/observation / survey questionnaire / secondary analysis / text analysis / image analysis / experiment / network analysis / big data / other)
- funding (grant project name, acronym and number)
- licensing terms
- related publications

ARP has an own Schema Registry which is a Cedar-based metadata repository. In the ARP Schema Registry, users can create new metadata sets (schemas) according to their specific needs. Once they have been defined and accepted by ARP, the schemas defined by users to describe their data packages, or even the files within data packages, can be immediately used via the ARP AROMA service.

Metadata will be harvestable through the Open Archives Initiative Protocol for Metadata Harvesting system and by ARP Federated Search.

### **Naming conventions and versioning**

The best approach for project researchers is to begin implementing the provided data naming conventions and versioning policies outlined below right away, even if the data is limited to their computer or a small group of collaborators within project tasks. Having clearly structured filenames (including version numbers) facilitate sharing and archiving data throughout the research process.

The project aims to use standard naming conventions, including the following components: [task number] \_ [document type] \_ [version number] \_ [date (ddmmyy)].

Some examples are provided below:

T2.3\_nationwide\_database\_cleansed\_v2.4\_13112025

T2.3\_interview transcript\_ local government\_anonymised\_v2.1\_13112025

### **Versioning should be provided according to these conventions:**

- v1: raw data/first version of any research documentation (e.g. first version field notes; methodological approach according to a meeting memo; codebook; a raw, uncut version of an interview recording; first version of a database (uncleansed))
  - v1.x: updated version of v1 in the case of small modifications made to any of the previous versions (e.g. grammatically checked and corrected interview transcripts; database with formally checked and corrected headings, typos, fonts)
- v2: updated version of v1 in the case of more meaningful modifications made to the previous versions (e.g. cleansed database; anonymised interview transcript; field notes with comments added later)

When a final version of a file is ready (to be deposited), the components needed in the filename are the following: [task

number] \_ [document type] \_ [version number] \_ final \_ [date (ddmmyy)]. Researchers are free to add any other important metadata to the file that helps identifying it (e.g. partner institution; country identifier if relevant) in between [document type] and [version number].

Some examples are provided below:

T2.3\_nationwide\_database\_v7.2\_final\_13112026

T2.3\_interview transcript\_ local government\_v3.4\_final\_13112026

**2. MAKING DATA OPENLY ACCESSIBLE** (which data will be made openly available and if some datasets remain closed, the reasons for not giving access; where the data and associated metadata, documentation and code are deposited (repository?); how the data can be accessed (are relevant software tools/methods provided?)

Making data openly accessible:

- Specify which data will be made openly available? If some data is kept closed provide rationale for doing so
- Specify how the data will be made available
- Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
- Specify where the data and associated metadata, documentation and code are deposited
- Specify how access will be provided in case there are any restrictions

*Guidance:*

Participating in the open research data management (ORDM) does not necessarily mean opening up all your research data. Rather, the ORDM follows the principle "**as open as possible, as closed as necessary**" and focuses on encouraging sound data management as an essential part of research best practice.

The NRDIO recognises that there are good reasons to keep some or even all research data generated in a project closed. Where data need to be shared under restrictions, explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

The [Registry of Research Data Repositories](#) provides a useful listing of repositories that you can search to find a place of deposit.

To support reproducibility of research, the project will publish most of its research data openly (e.g., datasets, anonymised interview transcriptions, strings) according to FAIR principles via the ARP Data Repository (<https://repo.researchdata.hu/>). The ARP Data Repository is an enhanced version of the Dataverse data repository, which was created specifically for storing research data. The datasets will be also published in .sav and .dta format for easier access.

Interviews with dissemination level "confidential" (non-anonymous interviews) will *not* be shared due to privacy concerns. Potentially, some interviews may have their publication restricted due to commercial exploitation. These cases will be documented in the final version of the DMP.

### 3. MAKING DATA INTEROPERABLE *(which standard or field-specific data and metadata vocabularies and methods will be used)*

Making data interoperable:

- Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.
- Specify whether you will be using standard vocabulary for all data types present in your data set, to allow interdisciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

*Guidance:*

Interoperability means allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins).

The ARP Data Repository accepts only such metadata schemas that are structured according to established standards such as Dublin Core. When publishing data, we will use widely accepted file formats. The keywords will be selected from the CESSDA ELLST thesaurus.

### 4. INCREASE DATA RE-USE *(what data will remain re-usable and for how long, is embargo foreseen; how the data is licensed; data quality assurance procedures)*

Increase data re-use (through clarifying licenses):

- Specify how the data will be licenced to permit the widest reuse possible
- Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
- Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
- Describe data quality assurance processes
- Specify the length of time for which the data will remain re-usable

*Guidance:*

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

Reasons for embargoes may include time to publish or seek patents. If an embargo is sought, specify why and for how long, bearing in mind that research data should be made available as soon as possible.

In the ARP Data Repository we will use Creative Commons licences (CC), which are tools to grant copyright permissions to creative work.

Data will be published 6 months after the end of the research, without embargo.

The data package will be accompanied by an informative Readme file.

Our data quality process includes data cleansing (data scrubbing), fixing data errors, and enhancing data sets by adding missing values or providing more up-to-date information or additional records.

The published data can also remain reusable via the ARP repository for at least 10 years.

## 5. ALLOCATION OF RESOURCES and DATA SECURITY *(estimated costs for making the project data open access and potential value of long-term data preservation; procedures for data backup and recovery; transfer of sensitive data and secure storage in repositories for long term preservation and curation)*

Explain the allocation of resources, addressing the following aspects:

- Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
- Clearly identify responsibilities for data management in your project
- Describe costs and potential value of long term preservation

*Guidance:*

Note that costs related to open access to research data are eligible as part of the grant (if compliant with the Grant Agreement conditions).

Costs are eligible for reimbursement during the duration of the project under the conditions defined in the Grant Agreement.

Address data recovery as well as secure storage and transfer of sensitive data.

Also consider whether the data is safely stored in certified repositories for long term preservation and curation.

Estimated costs for making data FAIR:

- Data Collection and Storage:
  - Data storage solutions (cloud services, physical storage): HUF/year
- Data Processing and Quality Assurance:
  - Personnel costs for data entry, cleaning, and quality assurance: HUF/year
- Data sharing and archiving:
  - Depositing in ARP Data Repository and in other repositories: HUF
- Metadata Creation and Documentation:
  - Personnel costs for metadata creation and curation: HUF/year
- Total Estimated Costs: HUF

Research data is stored on password-protected computers of the researchers. Personal data can only be stored in the Centre for Social Sciences' Cloud Infrastructure (CSS Cloud), access to personal data must be kept and managed in a logbook created for this purpose.

Anonymity should be sought at the time of data collection. If personal data are included in the research, they should be anonymised as soon as possible.

**6. DATA COMPLIANCE** (*data compliance is the formal governance structure in place to ensure an organization complies with laws, regulations, and standards around its data*)

Explain legal compliance of the beneficiary institute:

- Data policy and/or strategy
- Data governance
- Describe what legal items (national and EU) and how they are followed concerning data protection

*Guidance:*

Data policy sets broad, high level principles that governs data management, data interoperability and standards, data quality, data protection and information security.

Data governance entails defining, implementing and monitoring strategies, policies and sharing the management and use of data assets.

Main legal items include, e.g., laws on the protection and management of personal data at the national level, and GDPR at the EU level

To what extent the personal data management of the project and/or institute fulfil the requirements prescribed by GDPR?

In the CSS, the RDC has been providing support in data management for many years. Policies and templates are available on RDC's website (<https://kdk.tk.hu/en>).

Research in CSS can only be launched after a data management plan has been completed and an ethics procedure has been carried out.

The processing of personal data in the research is in accordance with the provisions of the GDPR.

In all cases, participation in research is voluntary. Written consent will be requested from participants with personal/sensitive data in processing of personal/sensitive data.

**DISCLAIMER**

**It is the responsibility of the Principal Investigator to inform the NRDIO of any ethics issues/concerns regarding the collection, processing, sharing and storage of data in relation to the project.**