**OPEN SCIENCE**

**RESEARCH DATA MANAGEMENT PLAN (DMP)[[1]](#footnote-1)**

|  |  |
| --- | --- |
| **Project number** |  |
| **Funder** | National Research, Development and Innovation Fund |
| **Project title** |  |
| **Principal investigator (PI)** |  |
| **PI ORCID identifier** |  |
| **Beneficiary institute** |  |
| **Project starting date** |  |
| **Project end date** |  |
| **Project duration** |  |
| **Data manager/contact** |  |
| **DMP 1st version/date** |  |
| **DMP last update[[2]](#footnote-2)/date** |  |

|  |
| --- |
| **SUMMARY** *(dataset[[3]](#footnote-3) 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:   1. State the purpose of the data collection/generation 2. Explain the relation to the objectives of the project 3. Specify the types and formats of data generated/collected 4. Specify if existing data is being re-used (if any) 5. Specify the origin of the data 6. State the expected size of the data (if known) 7. 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.   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 for:  • Researchers  • Research communities  • Decision makers  • Educational purposes  • 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](http://rd-alliance.github.io/metadata-directory/) that can be searched for discipline-specific standards and associated tools.  After the research is completed, the - selected and prepared - research data will be uploaded to RDC repository: <https://openarchive.tk.mta.hu/>  Metadata of deposited data in RDC 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:   * Digital Object Identifier (DOI) * primary details about the research project/dataset (title; short abstract; keywords; data collection period/collection date; temporal coverage; geographic coverage; author(s); acronym; date of deposit (if available)) * creators of dataset (name(s); ORCID(s); email address(es)) * 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   **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  Metadata will be harvestable through the Open Archives Initiative Protocol for Metadata Harvesting system. The deposited dataset will be identified by DOI. Version numbers will be provided. Search keywords will be provided, all metadata (data of surveys, methodology, research tools) and the textual information will be available and searchable through an internal search engine at the RDC Repository. |

|  |
| --- |
| **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:   1. Specify which data will be made openly available? If some data is kept closed provide rationale for doing so 2. Specify how the data will be made available 3. 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)? 4. Specify where the data and associated metadata, documentation and code are deposited 5. 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](http://www.re3data.org/) 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 - selected and prepared - research data (e.g., datasets, survey results, reports, journal articles, other communication and dissemination materials) according to FAIR principles via the repository of the Research Documentation Centre (RDC): <https://openarchive.tk.mta.hu/>.  The databases are also published in .sav and .dta formats for easier access.  The metadata and many of the documents of the RDC Repository are available to all visitors. The possibility to download research raw material and publications is open to all researchers of the Centre for Social Sciences. External researchers need to ask access to restricted collections.  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. Those cases should 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:   1. Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. 2. Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary 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).  RDC Repository uses Dublin Core metadata schema in all its records.  The variables in the databases will be given easy-to-understand labels. Labelling of databases will be standard, one type of data will have the same label in each database.  Datasets will include a Readme file describing the datasets and the files they contain. |

|  |
| --- |
| **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):   1. Specify how the data will be licenced to permit the widest reuse possible 2. Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed 3. 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 4. Describe data quality assurance processes 5. Specify the length of time for which the data will remain re-usable   *Guidance*:  The [EUDAT B2SHARE](https://b2share.eudat.eu/) 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 RDC Repository we will use Creative Commons licences (CC), which are tools to grant copyright permissions to creative work.  In RDC Repository, research data goes through a quality control process and has to meet its requirements to be published.  The published dataset remains reusable via RDC Repository for at least 15 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:   1. Estimate the costs for making your data FAIR. Describe how you intend to cover these costs 2. Clearly identify responsibilities for data management in your project 3. 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 depositing for long-term preservation: * Data Processing and Quality Assurance:   + Personnel costs for data entry, cleaning, and quality assurance: HUF/year   + Personnel costs for anonimization: HUF/year * Metadata Creation and Documentation:   + Personnel costs for metadata creation and curation: HUF /year * Total Estimated Costs: HUF   During the research project research data are stored on password-protected computers of the researchers and in CSS Cloud (the institutional Cloud of CSS). Personal data will only be stored in CSS Cloud, access to personal data must be logged.  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:   1. Data policy and/or strategy 2. Data governance 3. 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?  At CSS, the RDC team has been providing support in research data management for many years. Policies and templates are available on RDC’s website (https://kdk.tk.hu/en).  Research at CSS can only be launched after the Data Management Plan has been completed and an ethic procedure has been carried out.  The processing of personal data in research projects is in accordance with the provisions of GDPR.  In all cases, participation in research projects 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.**

1. 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). [↑](#footnote-ref-1)
2. DMP is to be regularly updated. [↑](#footnote-ref-2)
3. Several datasets may be included into a single DMP. [↑](#footnote-ref-3)