**SCRIPTS: Data Management Policy. Appendix**

**Data Management Plan Templates**

*(Please choose Template A or B depending on the stage of project implementation.*

*All fields are mandatory unless otherwise is stated in questions description.*

*Rectangular forms serve for illustrative purposes only and could be modified as needed.)*

# Introduction

This template aims to assist applicants and beneficiaries of the SCRIPTS excellence cluster to describe the key aspects of data management according to open science principles and the highest ethical standards. The template is meant to be an illustrative example of a data management plan in the application phase and the project application phase and should be used alongside SCRIPTS Data Management Policy and adjusted to a specific project.

The template is based on and heavily draws on existing guidelines for data management, in particular on the Horizon Europe Programme Guide[[1]](#footnote-1) and Humboldt University’s template for Horizon-style data management plan[[2]](#footnote-2), the German Research Foundation’s Checklist for Handling of Research Data,[[3]](#footnote-3) and the Consortium of European Social Science Data Archive’s Data Management Expert Guide and the accompanying data management plan template.[[4]](#footnote-4) While preparing a data management plan please also refer to any existing discipline standards and subject-specific recommendations.

***Important:*** The data management plans shall be submitted project-based (one data management plan for the whole project).

PhD researchers affiliated with the SCRIPTS graduate program shall submit data management plans for their dissertation projects separately from the project-based data management plans.

Faculty members, postdoctoral researchers, and individual researchers affiliated with the SCRIPTS are not required to provide data management plans for any of their individual projects conducted from personal or external funds.

# TEMPLATE A.

# DATA MANAGEMENT TEMPLATE - RESEARCH DESIGN AND GRANT PROPOSALS, PROJECTS THAT COMPLETED DATA COLLECTION AND FINISHED PROJECTS

## GENERAL INFORMATION

Project name:

Principal investigator/researcher:

## DATA SUMMARY

Provide a summary of the data addressing the following issues:

Will new data be collected/generated? If yes, provide the rationale for the new data collection/generation. Will existing data be re-used? If yes, clarify the origin and any potential copyright issues. Explain how collected, generated or re-used data relates to the objectives of the project. Specify the types and formats of data that will be used (e.g. image data, text data or measurement data). Outline potential stakeholder groups that could benefit from the data.

|  |
| --- |
| The project will use both data collected via questionnaires and existing data from national registries. Collection of data was deemed necessary as no existing data is available for secondary analysis based on current information and after consultation of relevant disciplinary repositories. National registry data will be obtained from XXX following official retrieval instructions and after obtaining the necessary permissions. The collected data will be used to gain insight into XXX and the re-used data will contribute to the understanding of YYY. The dataset will be generated by the project research team with methodological approaches XYZ and will include both measurement data (e.g. endorsement of a survey item) and text data (e.g. answers to open-ended questions) information. Apart from the research team, the dataset will be useful for other research groups working on similar questions in the area of ABC. |

## OPEN SCIENCE

Provide a summary of how will the project address recommended open science practices:

Provide information on whether and how you will implement early and open sharing of project output (e.g. preregistration, preprints). What measures will the project take to increase the reproducibility of project outputs (e.g. transparent research design, the robustness of statistical analyses, addressing negative results, access to tools and instruments used)? Explain how will the project meet open access requirements (e.g. how will open access to data be ensured, are there any legitimate reasons for restricting access to data, how will the project ensure that the data are shared according to the FAIR principles, will there be embargo periods on data re-use)? How will data be preserved in the long term?

|  |
| --- |
| Research hypothesis, study design and planned analysis will be made available through preregistration in XXX platform. Scientific manuscripts will be publicly shared before peer-review and journal publication via preprint platform XXX. Preregistration will make research design and potential negative results transparent. Analytic robustness will be ensured through documenting analytical workflow and testing of alternative plausible explanatory models. The project will use open-access software XXX to analyse data and questionnaires that will be used to collect data will be shared unless restrictions apply. Open access requirements will be met by deposing data to a publicly accessible disciplinary repository ZZZ. All collected anonymised data will be made openly available. The data, metadata and documentation will comply with disciplinary standards and will use open file formats and controlled vocabularies and the standard metadata schema for easy interoperability and re-use. The data will be licensed under XXX with an embargo period of X years. All data necessary to reproduce project outputs will be archived for at least 10 years. |

## ALLOCATION OF RESOURCES

Explain the allocation of resources for data management:

What are the estimated costs for project data management and deposing the data in a repository, if applicable? Who will be responsible for data management during the project and after the project is finished. Who owns the data?

|  |
| --- |
| Estimated costs for project data management and preparing data for deposition in a repository are XXX Euro. Expenses consist of X months for one full-time equivalent for data management and documentation and the data preparation and publication costs of the repository. XY will be responsible for developing and updating the data management plan. YZ will be responsible for data archiving and publication within the repository. According to the University of XXX regulation on XXX, University is and will remain the owner of all data collected for this project. |

## DATA STORAGE AND SECURITY

Clarify how will the project ensure the storage and security of data:

How will data be saved and secured? Will data be encrypted or stored in the secure physical or password-protected digital version? How is the security of sensitive data during the project period guaranteed (access and usage management)?

|  |
| --- |
| During the course of the project, data will be automatically saved daily on an institutional server with backup on a separate offsite institutional server. The backup will be checked manually every two weeks. XY and YZ will be responsible for backup and storage. Access to all data will be password-protected and restricted to the project members. Sensitive data will be separated as early as possible to create an anonymized dataset. Only anonymized data will be shared. Access to sensitive data will only be allowed to XYZ. |

## OTHER

State other national/funder/sectorial/departmental procedures for data management that the project will adhere to, if any:

|  |
| --- |
| Research Data Management Policy of XXX University  Principles of XXX Ethics |

# TEMPLATE B.

# DATA MANAGEMENT TEMPLATE – NEW PROJECTS AT THE PRE-DATA COLLECTION STAGE OF PROJECT IMPLEMENTATION

|  |  |
| --- | --- |
| Version | Notes |
| DMP version X, YYYY-MM-DD |  |

## GENERAL INFORMATION

|  |  |
| --- | --- |
| Project name |  |
| Project description |  |
| Principal investigator/researcher |  |
| Participating researcher(s) and/or organization(s) |  |
| Project Data Contact |  |

## DATA DESCRIPTION

### Data collection/generation and re-use

Specify how and why will data be collected and re-used, if applicable:

Will new data be collected/generated? If yes, provide the rationale for the new data collection/generation. Will existing data be re-used? If yes, clarify the origin and any potential copyright issues. Explain how collected, generated or re-used data relates to the objectives of the project.

|  |
| --- |
| The project will use both data collected via questionnaires and existing data from national registries. Collection of data was deemed necessary as no existing data is available for secondary analysis based on current information and after consultation of relevant disciplinary repositories. National registry data will be obtained from XXX following official retrieval instructions and after obtaining the necessary permissions. The collected data will be used to gain insight into XXX and the re-used data will contribute to the understanding of YYY. |

### Type of data

Specify the types and formats of data that will be used (e.g. image data, text data or measurement data). What size data do you expect?

|  |
| --- |
| The dataset will include both measurement data (e.g. endorsement of a survey item) and text data (e.g. answers to open-ended questions) information. The survey will consist of XXX questions and will be sent to XXX individuals. The estimated size of the dataset is XXX MB. |

## DATA QUALITY AND DOCUMENTATION

### Data quality assurance

Describe what steps will project make to assure data quality:

How will potential flaws related to data collection and data quality be described (e.g. selection biases, skewed samples of respondents, non-responses, missing data of random and non-random nature). How reliable are the data sources? Is it possible to cross-validate (double check) the collected data? Suggested measures of how to deal with the data problems.

|  |
| --- |
| Potential flaws related to data collection and data quality will be explained in a detailed methodological report based on current best practices.[[5]](#footnote-5) The report will state for what purpose were the data collected; how were the respondents selected; how were the data collected; what information was collected; who collected the data when and where; how were the data edited, coded and weighted; and what provisions of data protection laws were respected. Data quality will be ensured by validating the sample, replication, comparison with results of similar studies and control of systemic distortion. |

### Metadata

Describe what metadata will be provided with the dataset:

|  |
| --- |
| Detailed metadata will be recorded on project- and data-level. On the project level, metadata will include the full methodological report, questionnaires and other study-related materials (unless restrictions apply) and any other used materials, such as informed consent, invitation to participate, reminders and other test materials as well as sample size and sample specifics. Questionnaires and other used materials will be provided both in machine-readable format and as screenshots to resemble the “look and feel” of the original instrument. Data-level documentation will include names, labels and descriptions of variables, codes for missing values and their descriptions, description of derived variables (if applicable), and frequencies. This metadata will be provided in a separate codebook as open source software XXX does not allow for embedding data documentation within a file. All metadata will be provided in English or in German with a corresponding English translation. |

### Terminologies and ontologies

Clarify which terminologies/ontologies will the project use, if applicable:

|  |
| --- |
| Keywords and topics will be generated by the data repository. Keywords will be based on the XXX thesaurus, and topics on XXX topical classification. |

### Reproducibility

Specify measures that will be taken to increase the reproducibility of project outputs :

How will the project ensure e.g. transparent research design, the robustness of statistical analyses, addressing negative results, access to tools and instruments used?

|  |
| --- |
| Preregistration will make research design and potential negative results transparent. Analytic robustness will be ensured through documenting analytical workflow and testing of alternative plausible explanatory models. The project will use open-access software XXX to analyse data and questionnaires that will be used to collect data will be shared unless restrictions apply. |

## STORAGE AND BACKUP OF DATA

### Folder structure, naming and versioning

Specify how conventions for folder structures, naming and versioning:

|  |
| --- |
| Folders will be organised in a hierarchical and clear structure. All material related to the data will be stored in folders, including the information on methodology and used tools. Files will be named consistently and will be uniquely identifiable and versioned. File names will include a short description of the content, version number on a one-level structure (e.g. V1, V2, V3) and date of creation in ISO 8601 standard (YYYY-MM-DD). |

### Storage and backup

Specify how data will be stored and backed up:

|  |
| --- |
| During the course of the project, data will be automatically saved daily on an institutional server with backup on a separate offsite institutional server. The backup will be checked manually every two weeks. XY and YZ will be responsible for backup and storage. |

### Data security

Specify how will project ensure the security of data:

How will data be saved and secured? Will data be encrypted or stored in the secure physical or password-protected digital version? How is the security of sensitive data during the project period guaranteed (access and usage management)?

|  |
| --- |
| Access to all data will be password-protected and restricted to the project members. Sensitive data will be separated as early as possible to create an anonymized dataset. Access to sensitive data will only be allowed to XYZ. |

## LEGAL OBLIGATIONS AND CONDITIONS

### Ethical review

State whether or not the project underwent an ethical review. If not, state why an ethical review was not necessary:

|  |
| --- |
| The research project was approved by the XXX under number XXX. |

### Informed consent

Specify how informed consent will be collected and protected:

|  |
| --- |
| Informed consent will be sent to all participants. A copy of the blank informed consent will be included in the project technical report. Informed consent will include information about the plan for data sharing. Signed informed consents will be kept separately at a lock-protected location at XXX for XXX years according to the XXX regulation. |

### Protection of the identity of participants

Clarify how will the identity of participants be protected:

|  |
| --- |
| The data will be anonymized as soon as possible. The dataset will not include any direct identifiers. XY will conduct controls for re-identification to ensure that there will not be any possibility to identify individuals in the dataset. If necessary, variables will be re-coded to broader levels before data sharing (e.g. income to income ranges or categories) to ensure that participants’ identities cannot be established. |

### Intellectual property rights/copyright

Clarify legal any legal peculiarities and intellectual property rights/copyright issues:

What are the legal peculiarities in connection with dealing with research data in your project? Are there any potential effects or limitations related to later publication or accessibility? How are usage and copyright issues as well as property issues taken into account?

|  |
| --- |
| According to the University of XXX regulation on XXX, University is and will remain the owner of all data collected for this project. Collected data will be shared under XXX licence, which allows the use of data under XXX restrictions. Data obtained from the registry XXX will be shared if relevant permissions are obtained. All survey questions and instruments developed by the project team will be shared under XXX licence. Appropriate permissions will be obtained for sharing survey questions and instruments for which University XXX does not hold copyrights. |

### Other legal obligations and conditions

State other national/funder/sectorial/departmental procedures for data management that the project will use, if any:

|  |
| --- |
| Research Data Management Policy of XXX University  Principles of XXX Ethics |

## DATA EXCHANGE AND PERMANENT ACCESS TO DATA

### Early and open sharing of data

Provide information on whether and how you will implement early and open sharing of project output (e.g. preregistration, preprints):

|  |
| --- |
| Research hypothesis, study design and planned analysis will be made available through preregistration in XXX platform. Scientific manuscripts will be publicly shared before peer-review and journal publication via preprint platform XXX. |

### Making data available

Specify how open access to data will be ensured:

Which data is particularly suitable for subsequent use in other contexts? Which data will be shared? Are there legitimate reasons for restricting access to data? Which criteria are used to select research data to make it re-usable by others? Will there be embargo periods on data reuse?

|  |
| --- |
| The entire dataset would be useful for other research groups working on similar questions in the area of ABC. Data collected by the project will be deposited at a data repository X. Data obtained by the registry XXX will be shared only upon obtaining relevant permissions. Variables XYZ will not be published publicly due to ABC legal restrictions. Dataset will be available for reuse after the embargo period of X years. Survey questions and instruments that will be used to collect data will be immediately be available for reuse, unless restrictions apply. |

### Repository

Specify the selected repository, if applicable:

Describe the characteristics of the repository (e.g. disciplinary, institutional etc.). Is the repository certified? Does it use persistent identifiers, and if so, which one?

|  |
| --- |
| Data will be deposited in repository XXX which is a disciplinary repository with XXX certification that follows the FAIR principles. The dataset will receive a Digital Object Identifier (DOI). |

### Long-term preservation

Specify how long will the data be preserved:

|  |
| --- |
| Data deposited to the repository will be preserved for as long as possible (for as long as the repository is in operation), but for a minimum of 10 years. |

## RESPONSIBILITIES AND RESOURCES

Explain the allocation of resources for data management:

Who is responsible for the handling of the research data (description of roles and responsibilities within the project)? Which resources (costs; time or other) are required to be able to deal with data adequately in the project? After the project’s expiry, who is responsible for curating the data?

|  |
| --- |
| XY will be responsible for developing and updating the data management plan. YZ will be responsible for data archiving and publication within the repository. Estimated costs for project data management and preparing data for deposition in a repository are XXX Euro. Expenses consist of X months for one full-time equivalent for data management and documentation and the data preparation and publication costs of the repository. The principal investigator XX will be responsible for curating the data. |

1. Horizon Europe (2021), “[Programme Guide](https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf)”. [↑](#footnote-ref-1)
2. Humboldt University (2016), [“Example DMP for Horizon 2020 Version 3.0”](https://www.cms.hu-berlin.de/de/dl/dataman/muster-dmp-h2020-v3). [↑](#footnote-ref-2)
3. German Research Foundation, (2021), “[Checklist Regarding the Handling of Research Data](https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_en.pdf)”. [↑](#footnote-ref-3)
4. CESSDA (n.d.), [“Data Management Expert Guide”](https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide). [↑](#footnote-ref-4)
5. Jedinger, A., Watteler, O., & Förster, A. (2018). Improving the Quality of Survey Data Documentation: A Total Survey Error Perspective. *Data, 3*(4), 45. MDPI AG. Retrieved from http://dx.doi.org/10.3390/data3040045 [↑](#footnote-ref-5)