INITIAL HORIZON 2020 DMP (v3.0)

DATA MANAGEMENT PLAN FOR PROJECT XYZ

ADMIN DETAILS

Project Name: Initial Horizon 2020 DMP (v3.0) - Data Management Plan for project XYZ
Project Identifier: Example-DMP-ID-3
Principal Investigator / Researcher: John Doe
Project Data Contact: Susan Smith, +49 (0)30 2093-70072, Susan.Smith@hu-berlin.de
Description: The example project is based on ABC and aims to investigate research questions XYZ.
Data will be collected for the purpose of quantitative analysis and evidence-based conclusions drawing.
Funder: European Commission (Horizon 2020)

1. DATA SUMMARY

Provide a summary of the data addressing the following issues:
- **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**

Based on current information and after consultation of relevant repositories, there is no existing data available for secondary analysis. The collected data will give further insight into XXX and will lead to YYY. The dataset will be generated by the project research team with methodological approaches XYZ. As no comparable data is available for secondary analysis at the moment, it is planned to make our dataset publicly available in a disciplinary research data repository. Apart from the research team, the dataset will be useful for other research groups working on similar questions in the area of ABC.

The expected data volume will be approximately 100 GB. The university storage will cover this size with no additional costs. Exchange of data with project members at different sites will be secure and redundant through use of available university cloud storage.

The data will be stored in the open format XXX. In addition, data will be saved in YYY format, which is a standard format in the discipline and widely accepted.

2. FAIR DATA

2.1 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**

We intend to share our dataset in the publicly accessible disciplinary repository ZZZ using descriptive metadata as required/provided by that repository. The repository assigns DOIs for clear identification and citability of the dataset. Additional metadata of the dataset will be offered within a separate XML file in a standardized way by using the XXX schema, which is suitable for the discipline. Folders will be organized in a hierarchical and clear structure. Files will be uniquely identifiable and versioned by using a name convention consisting of project name, dataset name, method used, ID, place and date. Keywords will be added by using the YYY thesaurus.

2.2 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**

All data will be made available. However, there will be different access levels. Anonymized data will be made openly available. Sensitive data will not be publicly available, according to data protection law. Access can be granted onsite at the repository (visiting scientist) or - with sufficient clearance - through controlled remote data processing.
The following software is needed to access the data: word/spreadsheet processing program (e.g. OpenOffice), Adobe PDF Reader, image viewing software (e.g. XnView), XML viewer
No additional documentation is needed for the software. The data, metadata and documentation will be deposited in the discipline-specific repository ZZZ as well as at the institution (for ten years).

2.3 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 inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?**

The data, metadata and documentation are compliant to disciplinary standards, open file formats and use controlled vocabularies and the standard metadata schema for easy interoperability and re-use.

2.4 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**

The data will be licensed under Creative Commons CC BY 4.0 International. All data will become available during the last three months of project end. Parts of the data can become available even before due to journal publications. There will be no embargo period.

The data can be re-used by other scientists in the field of ABC. Neighboring disciplines and interdisciplinary research groups might also be interested, because of XXX and our chosen methodological approach. Future research should also concentrate on XYZ.

The data quality is ensured by different measures. These include validation of the sample, replication, comparison with results of similar studies and control of systematic distortion.

As open formats are used for data archiving, the data will remain re-usable until the repository withdraws the data or goes out of business.

3. ALLOCATION OF RESOURCES

Explain the allocation of resources, addressing the following issues:
- **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**

Estimated costs for data preparation to be FAIR are XXX Euro. Expenses consist of additional publication and documentation costs of three months for one full time equivalent and publication costs of the repository. Associated costs for dataset preparation and data management during the project will be covered by the project itself.

Susan Smith will be responsible for data management plan updates. Jane Doe and James Smith will be responsible for backup and storage. The project leader John Doe will be responsible for data archiving and publication within the repository.

Long term preservation will result in no additional costs other than repository charges for data submission. The dataset will increase in value over the years because of its fundamental impact in the ABC field, also in the future. Long-term studies and interdisciplinary research are desirable and intended.

4. DATA SECURITY

Address data recovery as well as secure storage and transfer of sensitive data
During the project data will be automatically saved daily on an institutional server with backup on a separate offsite institutional server. Backup will be checked manually at intervals of two weeks. The two project leaders Jane Doe and John Smith will be responsible for backup and storage. They will be supported by the universities IT team. Following consultations with the institutional IT team, no additional costs are expected for storage and backup.
Sensitive data will be separated as early as possible to create an anonymized dataset. Access to sensitive data is granted only for project members with clearance through non-disclosure agreements. In addition, rightful access is ensured through secure passwords and encoding of the folders and files within cloud storage. Data transfer is secured via HTTPS protocol.
Long-term preservation will be provided by the selected disciplinary repository. Associated costs need to be covered, which will be XXX Euro. In addition, data will be stored locally for ten years at the institution with redundant, robust and secure storage. No additional costs are associated with this local storage.

5. ETHICAL ASPECTS

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
Informed consent for data sharing and long term preservation is included during data collection. Sensitive data will be separated as soon as possible and kept secure.

6. OTHER

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)
- Research Data Management Policy of Humboldt-Universität zu Berlin
- Principles of XXX Ethics