**SCRIPTS: Data Management Policy**

**Proposal Draft 21/12/2021**

**DMC**

**Background**

Open science is an approach to the scientific process that aims to disseminate knowledge – scientific publications, research data and other research outputs - as soon as it becomes available using digital and collaborative technology.[[1]](#footnote-1) Open access enhances research quality, reduces the need for duplication of research, speeds up scientific progress, helps to combat scientific fraud, and promotes economic growth and innovation.[[2]](#footnote-2)

Open science is increasingly becoming the golden standard in research. It is a policy priority and the standard method of working under the European Commission's research and innovation funding programmes such as Horizon Europe[[3]](#footnote-3) and the European Research Council.[[4]](#footnote-4) Open access to knowledge is also enshrined in the Berlin Declaration on Open Access to Knowledge in Sciences and Humanities,[[5]](#footnote-5) Open Access Strategy for Berlin[[6]](#footnote-6) and the German Research Foundation’s (DFG) Guidelines for Safeguarding Good Research Practice.[[7]](#footnote-7)

**Principles of data management in open science**

Data management includes strategies and measures for handling research data during and after a project. Research data are defined as all data that are generated or evaluated during scientific work, which includes both quantified data (includes survey data, measurement data, parsed and coded preprocessed data) as well as qualitative data and raw data (such as audiovisual information, texts, transcripts of interviews, and anthropological notes). Good data management is a key component of open science as it enables the discovery, integration and reuse of knowledge.[[8]](#footnote-8)

Data management should be based on the idea of “FAIR” (Findable, Accessible, Interoperable, Re-usable), unrestricted data access, accessibility, and replicability of research results. Hence, it puts certain conditions on project proposals, usage of grant resources, and conducting the planned research projects. The “FAIR” principles[[9]](#footnote-9) are widely accepted as the standard in data management, including by the European Commission and the DFG and are increasingly informing the data management policies of research organisations.[[10]](#footnote-10)

**Data management in the SCRIPTS Cluster of Excellence**

The present data management policy serves to specify rules and regulations that define the conditions of research implementation and data usage. We borrow the ideas of findability, accessibility, interpretability, and reusability from best data management practices to achieve as open data and products of research as possible. At the same time, we understand that certain research questions can be considered sensitive and certain data can be protected by various regulations. In these cases, we ask researchers to provide meta-data and as detailed as possible description of data specifics along with a memo stating specific conditions of why data cannot be published openly.

***Meta-data*** is descriptive data that provide information about some aspects of collected data. It summarizes the basic information about collected data that makes working with data easier. Meta-data do not reveal the collected data itself, but provides secondary information such as:

* purpose of data (collected for further qualitative or quantitative analysis);
* means of data collection (survey, deep interview, focus groups, expert interviews, archival data, ethnographical data collection, online/offline data collection, etc.;

important for ethnographic studies: researchers’ notes are not considered meta-data that is required to be provided as part of data management plan);

* for experimental studies: experimental design and the nature of randomization
* time and date of data collection;
* scope of data collection (location, number of participants, attrition rates if applicable);
* planned or established questionnaire/survey

***Applicability of the data management to individual scholars***

Data management policy applies to all the projects implemented with the direct SCRIPTS funding (under the DFG regulations). 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.

Data management policy in the SCRIPTS Cluster of Excellence is based on the following minimum requirements:[[11]](#footnote-11)

* responsible management of research data based on the FAIR principles and open access to research data under the principle “as open as possible, as closed as necessary”;
* access to information about the research outputs/tools/instruments needed to validate the conclusions of scientific publications or to validate/re-use research data (we request to provide all necessary materials to be able to replicate and validate the results of the study);
* digital or physical access to the results needed to validate the conclusions of scientific publications, unless exceptions apply;
* in cases of public emergency, if requested by the granting authority, immediate open access to all research outputs under open licenses or, if exceptions apply, access under fair and reasonable conditions to legal entities that need the research outputs to address the public emergency.

Current data management plan regulates the data management issues and concerns for each step of the project implementation.

I. Research design and grant proposals

The project description should provide information on how the project will be implemented and what design it will follow, and what the expected outcomes are, and how they can be achieved. The necessity of ethics committee approval needs to be stated at the proposal stage. Complicated research designs can be pre-registered in the preprints or pre-registration reports.

Research Data Management: to reach the possible proximity of open science and open data, we request research data management. It is mandatory for all projects that will produce or reuse data. If projects produce novel data, they are obliged to provide a 1 memo, stating the data collection plan and address all the ethical concerns related to that, and describe in what way they will be fulfilling open science principles (see the Research Proposal Stage DMP Template in the Appendix). If a project reuses the existing data, it should provide a short memo stating how reusage will be organized and managed (to avoid the potential questions of plagiarism). A full data management plan (DMP) is not required at the grant proposal stage. However, after the proposal revision, scholars have to fill in a survey form about all the concerns related to data collection, data handling and potential data storage.

The project description should outline the measures planned in the project that tend to increase reproducibility (such as transparent research design, the robustness of statistical analyses, addressing negative results, etc) or in mandatory/non-mandatory open science practices (e.g. the DMP, early sharing through preregistration and preprints, open access to software, workflows, tools, etc).

II. Project Implementation and Data Management Plan

1. Description of Data and Replicability of Results

· Data description

§ How is new data generated in the project? Will existing data be reused? Which data types, in terms of data formats (e.g. image data, text data or measurement data) is used in your project and in what way are they further processed? What size data do you expect?

· Documentation and data quality

§ Description of potential flaws related to data collection and data quality (such as 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 the collected data? Which manual and digital tools are used to collect data? Suggested measures of how to deal with the data problems.

· Storage and backup of the data

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

· Legal obligations and conditions

§ 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? Are there any important scientific codes or professional standards that should be taken into account?

· Data exchange and permanent access to data

§ Which data is particularly suitable for subsequent use in other contexts? Which criteria are used to select research data in order to make it re-usable by others? How and where can you archive your data? Are there embargo periods? When can the research data be used by third parties?

· Responsibilities and resources (including in the long-term)

§ 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 in order to be able to deal with data adequately in the project? After the project’s expiry, who is responsible for curating the data?

2. Data Privacy and Personal Data

This section concerns projects with research activities that involve processing of personal data, regardless of the method used (e.g. interviews, surveys, questionnaires, direct online retrieval etc.).

Research activities must comply with the applicable international, EU and national law (in particular, the GDPR, national data protection laws and other relevant legislation).

3. Long-Term Data Access and Storage

All the relevant data provided by the individual researchers and project teams will be archived, stored and accessible (in open or limited access) for at least 10 years.

III. Publication of the projects results at the SCRIPTS Data Portal

Open access of data necessary for replication or meta-data in cases of data restrictions is advised to be provided at the corresponding data portal. Necessary for replication procedures data do not have to include raw data (raw machine-readable or non-machine-readable files, interview recordings, text, image, or video content, individual survey responses) but unidentified pre-processed or processed, parsed data that can be used to replicate the project results. In cases of computational algorithms applications, replication codes are recommended to be provided along with the data. In cases of restricted data access (classified information, personal data protected by privacy laws, third-party data protected by the copyrights laws, etc.), scholars have to provide meta-data description, explaining the essence of the data – the scale of the data (eg., number of observations), what types of variables exist, how they were coded and why, how data was collected/purchased/retrieved and why data access is limited.

This data management plan supports the researchers towards the publication of their projects results and corresponding data in a suitable data archive or repository.

IV. Updates and Modifications

Data Management Plan needs to be updated over the course of the project whenever significant changes happen, such as (but not limited to):

* new data
* changes in consortium policies (e.g. new innovation potential, decision to file for a patent)
* changes in consortium composition and external factors (e.g. new consortium members joining or old members leaving).

1. European Commission (n.d.), [“Open science“](https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science_en), accessed 27 October 2021. [↑](#footnote-ref-1)
2. [Commission recommendation (EU) 2018/7990 of 25 April 2018 on access to and preservation of scientific information](https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32018H0790&from=EN). [↑](#footnote-ref-2)
3. 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-3)
4. European Research Council (n.d.), “[Open science](https://erc.europa.eu/managing-your-project/open-science)”, accessed 27 October 2021. [↑](#footnote-ref-4)
5. [Berlin Declaration on Open Access to Knowledge in Sciences snd Humanities](https://openaccess.mpg.de/Berlin-Declaration) (2003). [↑](#footnote-ref-5)
6. [Open Access Strategy for Berlin](http://www.open-access-berlin.de/en/strategie/index.html) (2015). [↑](#footnote-ref-6)
7. [Guidelines for Safeguarding Good Research Practice](https://zenodo.org/record/3923602" \l ".YXbGcxpByUk) (2019). [↑](#footnote-ref-7)
8. Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18 [↑](#footnote-ref-8)
9. ibid. [↑](#footnote-ref-9)
10. see for example the [Forschungsdaten-Policy der Freien Universität Berlin](https://refubium.fu-berlin.de/bitstream/handle/fub188/30821/Forschungsdaten-Policy_der_Freien_Universitaet_Berlin.pdf?sequence=2&isAllowed=y) (2021). [↑](#footnote-ref-10)
11. Modified from 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-11)