ABSTRACT
The data-driven shift in the science research leads to a wider range of research data. To manage this data in a sustainable and adequate way, data management plans (DMPs) were established as a method. However, some researchers still do not create DMPs due to lack of time, resources and understanding of the needs. Furthermore, most of the existing templates and tools are largely unknown. In this article, we investigated the benefits and challenges of DMPs in two joint research projects of several academic institutions. For this, we described the process during the DMP creation, potential challenges and benefits experienced. We showed that a DMP with completely uniform content among the partner institutions was not possible due to individual and subject differences (e.g., in storage and policies). Instead, individual texts had to be formulated in some cases to overcome the diversity. This complexity could not be handled with the existing tools. Therefore, both projects created an own adapted template with some generic contents. Existing guidelines and internal project policies helped during the generation. We experienced that fewer people work more efficiently on a DMP than many and that all researchers within the project can profit from every individual DMP. Although we were not required to produce one, we recognised the associated benefits as a guide during the research process in joint projects.
INTRODUCTION

Science depends on data. To address research questions, data need to be collected, interpreted and analysed. As their collection is time and resource consuming, funding organisations increasingly demand their sustainability and free accessibility. While scientific results are often publicly available, the underlying research data mostly remain private by the researchers or their organisation (Trippel & Zinn 2022). This prevents good scientific practices, where results can be replicated and new research can be built on existing data for similar scientific questions, contrastive studies or meta-analysis (DFG 2015) in the sense of FAIR data principles (Michener 2015). FAIR data are findable, accessible, interoperable and reusable (Wilkinson et al. 2016) and a valuable resource to support information equity, accelerate science and enhance research impact (Vision 2010). Thus, secure and efficient data sharing is essential to support and advance science; it allows researchers to save (funding) money and effort for redundant data production if comparable data already exist (Gonzales, Carson & Holmes 2022). Therefore, adequate research data management (RDM) became a prerequisite for research funding when applying for research grants (Patterson, Bothma & Van Deventer 2018). RDM is a task that includes planning, collection, storage, analysis, documentation, archiving and publication of research data (Higgins 2018, Neuroth, Putnings & Neumann 2021).

Increasingly, more funders request researchers to specify how their generated data will be managed. This task can be addressed in different ways, such as research data policies, data managements plans (DMPs) or even within a cooperation agreement (Schmiederer & Kuberek 2022). A policy contains a framework for action and orientation to create transparency and clarity in the handling of research data. Policies address ethical-legal and organisational-technical principles and framework conditions in terms of RDM (Hiemenz & Kuberek 2019). DMPs are text-based documents that orientate on policy guidelines. A DMP describes the handling of research data, how they are collected, processed, documented, analysed, stored and archived throughout as well as after the research project (Kitchener et al. 2017). The structure of a DMP thus includes project planning and data management, handling of existing and new research data, metadata to document the data and the context of its creation and their organisation, long-term archiving and access (Habohm & Müller 2011).

In the past, DMPs were not mandatory. In recent years, several funding agencies have required a detailed DMP to be submitted in grant applications to support good data practices and to promote data sharing as well as reuse (Holdren 2013). The European Framework Programme for Research and Innovation, Horizon Europe, has made their creation obligatory for funding. The National Institutes of Health has in the process of establishing a data management and sharing (DMS) policy. The specific requirements vary between funding organisations in length, detail and extent of review (Whitmire et al. 2015). This has led to an increasing need for support, guidance and appropriate tools for researchers for DMP preparation (Mannheimer 2018). Therefore, most service-providing departments in German research institutions offer various tools on how to handle research data (Dreyer, Lehmann & Odebrecht 2022).

Thus, a DMP should not be a burden but an easy-to-follow road map or guide with the opportunity for it to become an integral part of research processes and good scientific practices. This impacts and benefits everyone from researchers and publishers to funders which makes it worth the effort (Gonzales, Carson & Holmes 2022). Persistent identifiers, standardisation (metadata, vocabularies) and security (legal issues, archiving) make the research process easier and science FAIR (Blumesberger 2020). DMPs are also not fixed but evolving, living documents for all project phases, which should be started early and reviewed and revised regularly to reflect the status quo of the project and to react on needs or changes (Trippel & Zinn 2022). They likewise enable continuity in the event of staff changes, prevent double work, promote collaboration and increase the visibility and impact of research (Jones 2011).

However, there are benefits and challenges in every project, which can increase when several institutions are involved. In the following, we will describe the experiences of two projects with four to six project partners based on the process to a final DMP, with focus on the complexity, potential challenges and advantages that occurred (Table 1).

---
1 https://www.horizont-europa.de/.

Table 1
In the FDNext research project funded by the German Research Foundation (DFG), six universities from Berlin and Brandenburg are working together to evolve tools and services for a sustainable institutional RDM. In the three-year funding phase, various tools and concepts for departments, trainings for specific target groups, legal advice, policies and service management will be compiled and finally evaluated with stakeholders from the nationwide RDM community (FDNext 2020). To address those questions in a suitable manner, different methods to generate research data are used: for example, expert interviews, questionnaires, surveys and data analysis. In order to handle these data even beyond the funding phase, the FDNext project members decided to develop a project-wide DMP with a research-specific focus, although there was no formal need from the funder.

Due to the project structure, meaning different researchers from different institutions each working on small pieces of the puzzle to address the overall research questions, we decided to give everyone in the project the maximum freedom on how to handle their own research data (in the limits of FAIR and the funding directives). This means every researcher had the ability to write their own DMP. In order to still gain a project-wide narrative, a template was formulated. To meet all the requirements from the funder (DFG), we based our template on the ‘code for good scientific practice’ (DFG 2015). In addition, we oriented our template to a model plan on DMP (Heldig 2016) as well as a DMP template especially created for students of the Institute for Library and Information Science of the Humboldt-Universität zu Berlin (IBI 2022).

As a result, our template contains the main metadata regarding FDNext, such as project name, ID, short description and research focus within the project, including names and contacts of the scientists working on this task and also the main questions regarding new or reused data. Since FDNext is a very diverse joint project, every associated researcher had a slightly different vision on how to work with the (generated or existing) research data. Luckily, the questions regarding handling data could be categorised in four different sections: data strategy, data design, data transition and data storage.

Table 1 Differences between DMPs in general (left), in the FDNext project (middle) and BUA-FDM (right) with regard to seven aspects (vertical).

### PROJECT-SPECIFIC EXPERIENCE FROM FDNEXT FUNDING CODE 429828830

In the FDNext research project funded by the German Research Foundation (DFG), six universities from Berlin and Brandenburg are working together to evolve tools and services for a sustainable institutional RDM. In the three-year funding phase, various tools and concepts for departments, trainings for specific target groups, legal advice, policies and service management will be compiled and finally evaluated with stakeholders from the nationwide RDM community (FDNext 2020). To address those questions in a suitable manner, different methods to generate research data are used: for example, expert interviews, questionnaires, surveys and data analysis. In order to handle these data even beyond the funding phase, the FDNext project members decided to develop a project-wide DMP with a research-specific focus, although there was no formal need from the funder.

Due to the project structure, meaning different researchers from different institutions each working on small pieces of the puzzle to address the overall research questions, we decided to give everyone in the project the maximum freedom on how to handle their own research data (in the limits of FAIR and the funding directives). This means every researcher had the ability to write their own DMP. In order to still gain a project-wide narrative, a template was formulated. To meet all the requirements from the funder (DFG), we based our template on the ‘code for good scientific practice’ (DFG 2015). In addition, we oriented our template to a model plan on DMP (Heldig 2016) as well as a DMP template especially created for students of the Institute for Library and Information Science of the Humboldt-Universität zu Berlin (IBI 2022).

As a result, our template contains the main metadata regarding FDNext, such as project name, ID, short description and research focus within the project, including names and contacts of the scientists working on this task and also the main questions regarding new or reused data. Since FDNext is a very diverse joint project, every associated researcher had a slightly different vision on how to work with the (generated or existing) research data. Luckily, the questions regarding handling data could be categorised in four different sections: data strategy, data design, data transition and data storage.

---

The [1] data strategy on how to handle research data within FDNext is regulated in the project policy [Schmiederer et al. 2022]. If necessary, there are subject-specific concepts and measures for quality assurance, which can be described separately in the first sections of the FDNext DMP template. The [2] data design deals with the form of research data used in the project. This includes a description of the file formats and file types as well as file naming. Third-party rights can also be described in this section if the handling exceeds the provisions set out in the project policy. As long as there are no legal restrictions (e.g., third-party rights) on the [3] data transition and publication of research data, they should be published as quickly as possible. It is important that the data are made available in a form (e.g., file type) that is useful for subsequent users. If research data is released by a publisher, it must be determined how access to the data is nevertheless maintained for scientists from other fields as well as an interested public. The rules of good scientific practice regarding [4] data storage stipulate that research data must be archived for at least 10 years. This must be guaranteed in relevant, supra-regional infrastructures which will be described in the fourth and last section of the FDNext DMP template.

Once the template was reviewed, commented and revised by all project members, it was shared as a plain document in a collaborative cloud. This way every associated researcher could elaborate their own DMP regarding the special needs of their research focus within the project. Furthermore, there was a deadline for every researcher to finish their sketch of the DMP. From the day of this deadline on, we once more reviewed, commented and revised all DMPs and also seized the opportunity to gain a wider understanding of how our colleagues address our overall research questions. Due to the fact that a DMP is a living document and as thus it won't be finished before the project ends we decided to not publish our texts. In conclusion, the process is still ongoing, supporting the idea of a living document. Nevertheless, the discussion about the project-wide DMP template as well as the exchange concerning the individual DMPs helped to reach a common understanding not just of how research is to be done in FDNext but also on how we want to successfully answer our research questions. In that way, the additional work of creating, reviewing, commenting and discussing our DMPs was perfectly worth it.

PROJECT-SPECIFIC EXPERIENCE FROM BUA-FDM FUNDING CODE 501_CRDMS

The Concept Development for Collaborative Research Data Management Services (short BUA-FDM) project, funded by the Berlin University Alliance (BUA), aims to establish and strengthen sustainable RDM services and infrastructures. In order to closely align support, training, communication and services based on researchers' requirements, these were determined in the course of a survey. This enquiry also captured the researchers' needs for DMPs [Ariza de Schellenberger et al. 2022a] regarding support (e.g., in the form of tools) or reasons against their production [Jäckel, Helbig & Odebrecht 2022a; Jäckel, Helbig & Odebrecht 2022b]. Furthermore, to handle the project data, a DMP was generated, although it was not requested by the funder.

As everyone had been working on the same datasets in the project, we chose a coordinated approach for a uniform DMP. Various suitable tools were available, such as Research Data Management Organiser (RDMO), DMPTool, DMPonline or TUP-DMP, with varying advantages and disadvantages. We decided for a freely available template called RDMOkurz. RDMO is an open-source software and web application developed by a DFG project and was mentioned in our survey as a potential solution for missing technical tools regarding DMPs. It is already very well established in Germany, used or offered by various scientific institutions and within the National Research Data Infrastructure (NFDI). The RDMO template was easily implemented

---

5 https://rdmorganiser.github.io/.
6 https://dmptool.org/.
7 https://dmponline.dcc.ac.uk/.
8 https://dmp.tu-berlin.de/.
10 https://www.nfdi.de/.
into the German information RDM portal Forschungsdaten.info\textsuperscript{11} for a collaborative work. Filling out the questionnaire was intuitively feasible in a short time. However, it turned out that the questions were not suitable for us, as some only allowed yes or no answers, but the complexity of our project required a detailed description. Subsequently, the templates from Freie Universität Berlin\textsuperscript{12} and Humboldt-Universität zu Berlin\textsuperscript{13} were compared. The BUA-FDM team chose the first one and combined it with the one from RDMO. Not all questions were used, and a selection was made with regard to relevant issues, leading to an individual project template that summarised information in a continuous text and from one question group, rather than many individual answers.

Our template contained [1] administrative information on the project name and description, funding code and agency, principle investigators, participating institutions and relevant policies. In the [2] data description, we stated that we did not reuse any data but collected them ourselves through a self-evaluation with RISE-DE (Hartmann, Jacob & Weiß 2019) and the mentioned survey. We described the software as well as tools used for data collection and evaluation, the resulting datasets with their (open) formats and access rights. The [3] documentation and data quality section described the publication of the data, additional helpful information (code book, read-me file), selected metadata schema, DOI assignment and file naming. The [4] storage and technical backup during the course of the project differed depending on the institution and was presented individually. The [5] legal obligations and framework conditions included information on cross-institutional data storage and information security. [6] Data exchange and permanent accessibility described where (the open repository Zenodo\textsuperscript{14}) and how (open access) the data will be published. [7] Responsibilities and resources were divided according to the project leaders and the project staff.

For an easier collaborative work with all project members, we transferred our created template to the software Overleaf\textsuperscript{15}. Since the project had been ongoing for a while, most questions could easily be answered directly without any problems. Others (e.g., legal uncertainty) needed to be discussed. Uniform information from all institutions was combined and standardised, and differences were clearly indicated. In addition, we implemented a preliminary description with information about the institution-specific requirements (e.g., for storage or their policies). During this process, the document was kept up to date and revised as necessary. The final version was published in December 2022 (Ariza de Schellenberger et al. 2022b) and can be continuously updated as new versions in the future if required in the sense of a living DMP. Since the project members of BUA-FDM worked constantly on and with the DMP throughout the project, its preparation helped to identify and clarify open questions. The early creation of the DMP prevented us from doing redundant work and promoted cooperation; it will promote the visibility of the project results in the future.

GENERAL RECOMMENDATIONS FOR IMPROVEMENTS

The reasons against DMPs (e.g., lack of time, resources, necessity) mentioned by the researchers in the BUA-FDM survey were only partly evident in our projects. Both projects lacked suitable tools and templates and therefore created a questionnaire themselves. We understand why researchers suggested RDMO as a suitable tool for DMPs, as it is very simple, intuitive and fast to use, although it was unfortunately not sufficient enough for the BUA-FDM project. To capture the complexity of the collaboration of different institutions, more detailed DMPs are needed than the current existing templates allow. It should be clear that institutions differ in their work with (generated) research data, which means that not all contents of the DMPs can be written in a uniform way. Therefore, it was a big help in the FDNext project to categorise all questions regarding handling of research data circulating in the RDM community. In this way, we have been able to point out our research focus while still including all aspects on modern RDM.

Since the processing of the consistent answers took a lot of time in the BUA-FDM project, we

\textsuperscript{11} https://forschungsdaten.info/.
\textsuperscript{12} https://www.fu-berlin.de/sites/forschungsdatenmanagement/_dokumente/dmp-muster_v01.pdf.
\textsuperscript{13} https://www.cms.hu-berlin.de/de/di/dataman/arbeiten/dmp_erstellen/dmp-info.
\textsuperscript{14} https://zenodo.org/communities/bua-fdm/.
\textsuperscript{15} https://de.overleaf.com/.
made the whole DMP with its generic preliminary information about the respective institutions and their specifics (e.g., storage, policies) available for future projects. This can be used for subsequent DMPs, if required, to save time and resources.

In order to save personnel resources, tasks and responsibilities for the DMP should be precisely defined and delegated. Here, less is more. The great advantage of the project FDNext is and was the defined role of a coordinator. Thus, only one or two people were working on the plain template, and therefore double work could be prevented. Through the opportunity of internal reviews, everybody within the project was still able to adjust the DMP template for their needs and in the meaning of subject-specific requirements. In contrast, the BUA-FDM project experienced long processing during the development (e.g., through legal uncertainties and the long consultations with all project members). This first aspect should be better supported in the future to adequately assist researchers. For example, guidelines such as the DFG’s code for Safeguarding Good Scientific Practice about data accessibility should be considered as a help during the DMP generation. Similar, the FDNext project policy (Schmiederer et al. 2022) worked as a (also legal) framework that enabled us to freely describe our way of handling data.

DMPs have existed for years but have only recently become increasingly obligatory for research funding. Even though DMPs are not mandatory by all funding agencies, they should be prepared, as they are a road map during the research process and facilitate the work. A DMP should be generated at an early stage of a research project and be constantly updated as a living document. In addition, it should be reused as much as possible for subsequent projects. Thus, we were not able to confirm an asserted lack of relevance or benefit, as stated by several researchers from the BUA-FDM survey. Since we constantly worked on our DMPs throughout the projects, its preparation helped us to identify and clarify open questions. Thus, the elaboration of DMPs, even if not required from the funders, was a welcome support and helpful guide for our projects.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Denise Jäckel  
Humboldt-Universität zu Berlin, DE

Anna Lehmann  
Humboldt-Universität zu Berlin, DE

REFERENCES


Blumesberger, S. 2020. The data management plan—the common thread through the project. DMP consulting at the University of Vienna. In Proceedings of Science, Austrian Citizen Science Conference, Vienna, Austria, on 14–16 September 2020.


Gonzales, S, Carson, MB and Holmes, K. 2022. Ten simple rules for maximizing the recommendations of the NIH data management and sharing plan. PLOS Computational Biology, 18(8): e1010397. DOI: https://doi.org/10.1371/journal.pcbi.1010397
Hartmann, NK, Jacob, B and Weiß, N. 2019. RISE-DE—Referenzmodell für Strategieprozesse im institutionellen Forschungsdatenmanagement. DOI: https://doi.org/10.5281/zenodo.2549343


Hiemenz, B and Kuberek, M. 2019. Strategischer Leitfaden zur Etablierung einer institutionellen Forschungsdaten-Policy. DOI: https://doi.org/10.14279/depositonce-8412


Jäckel and Lehmann

Data Science Journal
DOI: 10.5334/dsj-2023-025

TO CITE THIS ARTICLE:

Submitted: 14 December 2022
Accepted: 25 May 2023
Published: 25 August 2023

COPYRIGHT:
© 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.