SERVICE, CONSULTING AND TRAINING: PORTFOLIO EXPANSION TO SUPPORT RESEARCH DATA MANAGEMENT

N. Knipprath, B. Zulauf

Heinrich-Heine-University of Düsseldorf (GERMANY)

Abstract

Making research data available is an aspect of good scientific practice. The planning, collection, processing, storage and sustainable provision of research data must meet recognised standards and high requirements in order to ensure traceability and verifiability as well as further use after publication. In this sense, research data is understood to be all data that is generated during the research process or is the result of it. The research process encompasses the entire cycle from generation (e.g. through digitisation, source research, experiments, documented observations and empirical studies), processing and analysis, to publication and archiving of the data. Research data is generated in all scientific disciplines using different methods.

The Centre for Information and Media Technology (ZIM), the University and State Library (ULB) and the Research & Transfer Department support the faculties and institutions of Heinrich-Heine-University Düsseldorf (HHU) in research data management (RDM). The ZIM operates a high-performance storage and long-term archiving infrastructure, the ULB offers researchers advice on the creation of data management plans and on questions concerning metadata and indexing options, and the Research & Transfer Department provides support in the application process.

RDM tools are an important part of our RDM portfolio. During our cooperation project with the Universities of Siegen and Wuppertal, we were able to expand and enhance our tool diversity: A new open-source software for the creation of a data management plan was set up in order to keep an optimal overview of the data management during the research project. The three universities also operate a common storage grid, which is redundantly set up at all three locations and thus offers high data security and availability. This storage grid is based on open source repository software, which is also used at all three locations. We were also one of the first universities in Germany to focus more strongly on the topic of "electronic lab notebooks" (ELN), to find a new circle of users, to offer software solutions directly and to establish a sustainable network within North Rhine-Westphalia's universities. The dialogue between the cooperation partners on the RDM solutions and service portfolios created during the collaboration supports transparency in the provision of RDM services for researchers at all three locations and has a positive effect on cooperation between the universities.

We specifically approached university researchers and invited them to weekly research data management meetings to discuss individual tasks and problems of the research project. This way, suitable solutions for the researchers could be offered and the use of the available tools could be explained directly using appropriate examples.

Our service portfolio also includes the provision of the recorded trainings. Various training formats were tried out, including short and concise "coffee lectures", training sessions at the different locations, exchange of modular topic blocks oriented on the structure of forschungsdaten.info, and all training sessions were evaluated according to specific criteria. We have based the development of our portfolio planning and the orientation of our RDM tools on the evaluation and analysis of consulting and training, and were thus able to effectively support the research projects of our university and its project partners.

Keywords: Research data management, good scientific practice, trainings, RDM tools, portfolio.
1 INTRODUCTION

Research data management (RDM) is not a new concept in the university research community. Nevertheless, its significance is becoming more and more important, its limits more and more clear and the acceptance by researchers is growing. Not only is research work improved by good research data management, but the way the researcher himself works is structurally adapted to his actual research question from the very beginning. The objectives of the researcher, his most important research results and the appreciation of his work are achieved through well thought-out, strategic planning at the beginning. "What do I want to achieve, where do I want to go?" are the central questions for entering the data management cycle (see “Fig. 1”). "Can I build on existing data, can it be re-used and where can I find it?" are the next research data management questions.

Many researchers frequently still underestimate research data management. For example, they often only think about a data management plan in the short term, which means that relevant considerations, such as those relating to the later publication of research data, are often postponed to a later date. Even if the research data should or may not be published, the ten-year period for storing such data should be considered as early as possible in the context of “good scientific practice”. If several project partners are then also to be coordinated, it becomes even more necessary to formulate all planned steps of research data management in a written data management plan (DMP). The data management plan describes for example

- which research data is generated or reused in the course of the project
- how the research data is used
- where and when they are stored and what is needed for this
- third party rights (data protection, copyrights, contractual regulations)
- responsibilities and contact persons

A good time to establish contact between researchers and the central university institutions, which will later take care of the research data, is before the project starts. And in the case of funded third-party projects, even before the application is submitted to the third-party funding body.

Figure 1. Research data lifecycle.
2 METHODOLOGY

The universities of Düsseldorf, Wuppertal and Siegen have joined forces in order to initially record the current status of their own research work at the location and, in a later step, to compare various cases and collect best practices across university boundaries. The project "FoDaKo - Research Data Management in Cooperation", funded by the Federal Ministry of Education and Research (BMBF), was intended to establish and anchor the topic of research data management within university strategies. As a result of this project, a joint final report was written in which all project goals are listed, and which can be used as a blueprint for other institutions [1]. During the course of the project, the project staff actively contacted the researchers in order to obtain the current state of knowledge about research methods, to present themselves as a central contact point for all questions concerning research data management and to arrange advisory meetings. The counselling interviews were conducted jointly by representatives of the central institutions, i.e. the Centre for Information and Media Technology (ZIM), the respective university library (UB) and the responsible research department of the administration. In this way, the researcher received information from various sources, be it assistance with applications, advice on data management plans or how, where and in what form data can be stored and what tools are available for this purpose.

The successful implementation of research data management at scientific institutions is based, on the one hand, on the provision of storage infrastructure and other technical tools for managing the data and, on the other hand, on a wide range of information and training opportunities. The online survey of the BMBF project "UNEKE - From USB stick to National Research Data Infrastructure (NFDI) - Development of a criteria-led decision model for the establishment of research data infrastructures" also showed that there was still a great need for information on the part of researchers. The results of this survey provided indications of possible topics for training courses and more targeted advice on RDM (see "Fig. 2"). This enabled the project staff to develop a concept focusing on various basic training courses. With the cooperative preparation and discussion of the materials, the know-how of the staff necessary for consulting purposes was also to be expanded at the same time.
The following topics were developed as training modules:

- Basics of Research Data Management
- Licenses for Research Data
- Requirements of Research Funding Organisations
- Data Management Plan
- DMPs with RDMO
- Persistent Identifier
- Open Science
- Repositories
- Reuse of Research Data
- Legal Aspects of Research Data Management

3 RESULTS

Since the beginning of the project, more than 50 interviews with researchers have taken place and more than 10 training sessions with 12 different modular training topics have been held. In total, more than 90 researchers from all institutions have participated in the training courses, including students, doctoral candidates, research assistants and members of the professorial staff from all faculties and many different chairs or working groups.

All content has been created by the authors themselves, or existing content has been incorporated, which has been licensed as Creative Commons Attribution 4.0 International (CC-BY 4.0) and thus made freely available. This ensures that all training materials can be published and reused by third parties. This licence also explicitly permits the editing of the materials. At the same time, the Internet presence of the central contact points of the respective institutions (for Düsseldorf, for example, [https://fdm.hhu.de](https://fdm.hhu.de)), as well as the cooperation website (to be found at [https://fodako.nrw](https://fodako.nrw)), was further improved and adapted to current enquiries, and new information and news were added.

3.1 Training evaluation

To evaluate the training courses, a short anonymous survey was developed, which had to be filled in by the course participants. The following questions were found there:

1. How do you assess the scope of the individual topics?
2. Which of the topics covered are important for your work or have you been particularly interested in?
3. What surprised you (contents, practice, discussion points, ...)
4. On which topics did you have difficulties in understanding?
5. Your further comments: ______________

When the feedback for the first question is added up, a large majority of participants (77%) consider the scope of the training topics to be appropriate. If the feedback is broken down by topic, it becomes clear that the majority (63% - 88%) of participants were also satisfied with the scope, and it is precisely here that it becomes clear that there is also a different level of knowledge or need for information for all topics because of the heterogeneous composition of the training participants already mentioned. While some would have liked to know more details, others felt that the offer was already too detailed.

Based on the qualitative feedback from the training courses and questions from interviews, FAQs were created, which could be used by all project locations for consultations and were also made available for subsequent use on the information platform forschungsdaten.info.

3.2 Result of consulting discussions

Researchers increasingly feel themselves forced to think about extensive research data management for third-party funding organizations and to formulate in a data management plan, which has often been demanded, what they intend to record and where they want to store their data. In addition to regular advice on the use of storage space for research projects, we have already been able to introduce our RDM tool portfolio to many researchers during the discussions and have found that
many opportunities were not yet known. However, we were also able to identify existing structural deficits and jointly revise and close them during the project period of the FoDaKo project.

### 3.2.1 Repository and distributed storage

For example, many researchers have been looking for a secure and reliable storage space for archiving research data (see "Fig. 3"), which we built up at all three universities simultaneously and in consultation with each other. The research data repository based on the software *DSpace* was set up and adapted as an institutional repository so that researchers who cannot use either a suitable disciplinary repository or the large multidisciplinary repository zenodo.org can store their research data in their own university and publish them directly with a Digital Object Identifier (DOI).

In addition, we have provided instances of the data management plan software solution *RDMO* (Research Data Management Organiser) for the participating institutions at the Wuppertal campus. This and the repository software solution can be used by university members via their central university identification. The respective research data is backed up locally in the repository and additionally synchronised as a copy at the other locations so that a high level of data security can be guaranteed. In this way, a number of concerns of researchers and third-party funders are addressed, such as inadequate data documentation, difficulty in finding data, data loss due to technical breakdowns and, through regular curation, the problem of obsolete, unreadable data formats.

### 3.2.2 Calendar, groupware and cloud storage

During the project duration researchers have different requirements for the storage of their work data. The size of the data ranges from a few MB of text files and CSV to mass spectrometry files that can be many GB in size. The requirements for group guidelines also differ. In Düsseldorf, for example, we set up working groups, initially in the groupware *BSCW*, later in the North Rhine-Westphalia (NRW) Cloud *Sciebo*, areas in which the researchers could only store data but not modify it. Or we discussed the opportunities offered by client-based encryption for data storage in the NRW Cloud. Researchers also asked for a group calendar that they could use within the working group, so we upgraded our wiki and collaboration software *Confluence* with a calendar plug-in.
3.2.3 Electronic laboratory books

What was completely new for many researchers, however, was the possibility of replacing their handwritten lab book with an electronic lab book (ELN). We already offer two different web applications for ELNs in Düsseldorf: The commercial variant Labfolder and the open source variant eLabFTW. Working groups that have already been using these solutions for some time are very satisfied with this central tool offering and the service is quickly getting around, especially in the medical field. Particularly here, researchers are looking for a well-organised database of their experiments, which can be searched via different tags as an improvement over a handwritten lab book. This is also reflected in an approach to establish a quality management system in preclinical research at the HHU, which we supported at the ZIM [2]. Despite the increasing use, there is often an internal threshold to actually use the electronic lab book. The motivation of the employees is the essential point to start with. For example, we at the ZIM are already in close contact with researchers in order to introduce the topic to students at an early stage and to present it attractively as part of the curriculum in order to use the positive attitude of students towards digitisation projects also for the digitisation of laboratory work [3].

4 CONCLUSIONS

The importance of research data management over the years is also reflected in a recommendation of the Council for Information Infrastructures (RfII), which states that each country is developing solutions "that are outgrowing its own national system and policy style" [4]. Therefore, international exchange to establish professional research data management is important for the communication between the applying institutions. In this way, best practices can be exchanged from which academic research can only benefit.

Interviews with researchers are seen as an important institutional aspect. In such open discussions, questions and problems can be quickly identified in order to then find concrete solutions for the research projects and to be able to provide appropriate skills for suitable tools. It is essential to achieve both a wider information on the existing tools of the RDM portfolio within the universities and to get feedback from researchers on missing tools, infrastructures or information. A good reason to improve this situation is also the Research Data Management Day planned for 2021, which is to take place simultaneously at several universities in North Rhine-Westphalia.

Obtaining feedback on the existing portfolio again and again, comparing existing solutions in a cross-university context and the continuous adaptation of the service structure are important aspects of a good RDM portfolio. We also support many different disciplines and try to keep services as generic as possible, also for efficiency reasons, and we try to generalise the improvements, if possible, to pass them on as best practices, even directly to our project partners in the FoDaKo consortium.

The objective is to bring researchers into the RDM consultation as early as possible in order to discuss a research project proposal or a larger project together. After all, funding applications are already rejected by third-party funding bodies due to a lack of research data management. Therefore, we support researchers in developing a data management plan as early as possible and advise them to do so regardless of where they are in the research data life cycle with their research project. It is important for us to explain as a narrative that only IT-technologically well thought-out third-party funding applications, for example those with so-called INF projects (in the research project integrated information management and integrated information infrastructure) have good and greater opportunities for funding.

ACKNOWLEDGEMENTS

Our thanks go to all our researchers who have come to us with their ideas and wishes for the implementation of research data management and have thus played a major role in the development of our RDM portfolio. We would also like to thank the staff of our FoDaKo cooperation, with whom we are jointly advancing the topic of RDM, especially within NRW, and the UNEKE projects for cooperating on the research data management surveys.
REFERENCES

[1] N. Knipprath, T. Rathmann, J. Stegemann, B. Zulauf, T. von Rekowski, M. Schleußinger, *Final report of the project "FoDaKo - Forschungsdatenmanagement in Kooperation"*. universi 2020.

[2] M. Hewera, A.C. Nickel, N. Knipprath et al. “An inexpensive and easy-to-implement approach to a Quality Management System for an academic research lab”, F1000Research 2020, 9:660. https://doi.org/10.12688/f1000research.24494.2

[3] B. Zulauf, N. Knipprath „ELECTRONIC LAB NOTEBOOKS - EARLY RESEARCH PRACTICE IN TEACHING“, ICERI2019 Proceedings, pp. 5651-5656, 2019.

[4] Rat für Informationsinfrastrukturen: *Entwicklung von Forschungsdateninfrastrukturen im internationalen Vergleich*. Bericht und Anregungen, Göttingen 2017, pp. 93. Accessed 24 September, 2020. Retrieved from http://www.rfii.de/?p=2346