

## Workshop Report

# Research Data Management - Current status and future challenges for German non-university research institutions

Mareike Petersen<sup>‡</sup>, Bianca Pramann<sup>§</sup>, Ralf Toepfer<sup>l</sup>, Janna Neumann<sup>¶</sup>, Harry Enke<sup>#</sup>, Jana Hoffmann<sup>‡</sup>, Reiner Mauer<sup>□</sup>

<sup>‡</sup> Museum für Naturkunde Berlin, Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany

<sup>§</sup> Georg Eckert Institute for International Textbook Research, Member of the Leibniz Association, Braunschweig, Germany

<sup>l</sup> ZBW, Leibniz Information Centre for Economics, Hamburg, Germany

<sup>¶</sup> TIB, Leibniz Information Centre for Science and Technology, University Library, Hannover, Germany

<sup>#</sup> Leibniz Institute for Astrophysics Potsdam (AIP), Potsdam, Germany

<sup>□</sup> GESIS - Leibniz Institute for the Social Sciences, Cologne, Germany

Corresponding author: Mareike Petersen ([mareike.petersen@mfn.berlin](mailto:mareike.petersen@mfn.berlin))

Reviewable v1

Received: 05 Jun 2020 | Published: 08 Jun 2020

Citation: Petersen M, Pramann B, Toepfer R, Neumann J, Enke H, Hoffmann J, Mauer R (2020) Research Data Management - Current status and future challenges for German non-university research institutions. Research Ideas and Outcomes 6: e55141. <https://doi.org/10.3897/rio.6.e55141>

## Abstract

This report describes the results of a workshop on research data management (RDM) that took place in June 2019. More than 50 experts from 46 different non-university institutes covering all Leibniz Sections participated. The aim of the workshop was the intra- and transdisciplinary exchange among RDM experts of different institutions and sections within the Leibniz Association on current questions and challenges but also on experiences and activities with respect to RDM. The event was structured in inspiring talks, a World Café to discuss ideas and solutions related to RDM and an exchange of experts following their affiliation to the different Leibniz sections. The workshop revealed that most institutions, independent of scientific fields, face similar overarching problems with respect to RDM, e.g. missing incentives and no awareness of the benefits that would arise from a proper RDM and data sharing. The event also endorsed that the Research Data Working Group of the Leibniz Association (AK Forschungsdaten) is a place for the exchange of all topics around RDM and enables discussions on how to refine RDM at all institutions and in all scientific fields.

## Keywords

infrastructure, Leibniz Association, RDM, research data, workshop, world café

## Introduction

Research data is the basis for all scientific work. The increasing digitisation of scientific processes and methods calls for new approaches in the way research data is handled. Simply publishing the conclusions resulting from an analysis of collected research data is no longer sufficient. Instead, well-structured and annotated research data is becoming an increasingly important resource for researchers. Ensuring that data is accessible and can be interpreted creates a range of diverse challenges for research funding bodies, research institutions, researchers and research support staff. It often requires a discipline-specific approach in research data management (RDM) or an adaptation of generic processes.

In 2019, the General Assembly of the Leibniz Association published Guidelines on the Handling of Research Data within the Leibniz Association (Leibniz Association 2019). Therein, the Leibniz Association emphasises the importance of responsibly and transparently handling research data within the framework of a sustainable and quality-assured research process. The association connects 95 independent German research institutions that range in focus from natural, engineering and environmental sciences to economics, spatial and social studies and the humanities. Leibniz Institutes address issues of social, economic and ecological relevance. Because of their importance for the country as a whole, the Leibniz Association institutes are funded jointly by Germany's central and regional governments. The Leibniz Institutes employ around 20,000 people, including 10,000 researchers (Leibniz Association 2018).

Against this background, the Research Data Working Group of the Leibniz Association (*AK Forschungsdaten*)<sup>\*1</sup> held an expert workshop on research data management. The one-day meeting 'Round of research data management experts' (*Expert\*innenrunde zu Forschungsdatenmanagement*) took place on June 18, 2019 at the Museum für Naturkunde Berlin, Leibniz Institute for Evolution and Biodiversity Science.

## Aims of the workshop

The aim of the workshop was the intra- and transdisciplinary exchange among RDM experts of different institutions and sections within the Leibniz Association on current questions and challenges but also experiences and activities with respect to RDM. An important objective was the strengthening of discipline specific exchange and enabling future collaboration on common questions and challenges of RDM.

## Methods

The workshop started with an introductory talk and two inspiring talks opening the exchange and discussion on research data management (RDM). The talks were followed by a World Café to collect and exchange results and opinions. Five tables with challenging and urgent topics related to RDM were arranged to facilitate intensive and direct and conversations in smaller groups. Two people of the organizing team were assigned to each table to assure a good moderation and documentation of the discussion at the same time. Prior to the concluding session, the different Leibniz Sections met individually to discuss their more domain-specific challenges and achievements. All workshop participants were personally invited. A special effort was made to ensure that all Leibniz Sections were covered and that, where possible, participants were in charge of, or had experience with RDM at their institution.

See Suppl. material 1 for a detailed programme (in German).

## Key outcomes and discussions

More than 50 experts from 46 different non-university institutes covering all Leibniz Sections participated in the workshop. After a warm welcome and an introduction to the overarching goal of the one-day workshop, Harry Enke gave a talk on research data management (RDM) in general and the planned National Research Data Infrastructure (NFDI)<sup>2</sup>. He focused on the question of whether current RDM procedures and infrastructures are ready and how they could or should contribute to NFDI (Suppl. material 2). The following talk, given by Stephanie Palek, was more practice-oriented and summarised the development of a catalogue of measures for the improvement of research data management at the Herder Institute for Historical Research on East Central Europe - Institute of the Leibniz Association (Suppl. material 3).

## World Café

In the World Café sessions, all participants were asked to stop at different stations and exchange and discuss ideas related to RDM at their institutions. The principal questions were “What is the state of affairs regarding RDM at your institution?” (Table ‘Reflection’); “Which tools do you use?”, “Do your scientists have specific technological requirements?” (Table ‘Technology’); “What are the most urgent challenges, with respect to research, awareness, implementation, and politics?” (Table ‘Society and Values’); “How are your personal experiences related to the lighting talks (improvement of institutional RDM and NFDI)?”, “For you personally, where is activity required and which decisions need to be taken?” (Table ‘Personal Objectives’); “For your scientific community, where is activity required”, “Where do you see domain specific solutions or approaches to solutions?” (Table ‘Community Perspective’). The participants were invited to participate at all five stations. So as to allow the constant and random mixing of experts present.

For a summary of the main points mentioned and discussed during the World Café and for topics of general importance mentioned at most stations please see Fig. 1.



Figure 1. doi

Main results of the World Café on experts' opinion to different aspects and challenges in research data management. Shown are the five stations ('Community Perspective', 'Reflection', 'Technology', 'Society & Values', and 'Personal Objectives') and the main points raised and discussed by the participants during the sessions.

## 'Reflection' Table

At the 'Reflection' table participants were asked about the general situation regarding RDM at their home institutions. Although for most institutions the topic of RDM is still relatively new, progress in RDM has been achieved by all represented institutions over the last few years. In many organisations, institutional research data policies have been implemented and research data management positions have been created. In some institutions researchers already receive support in managing research data, for example through the provision of training courses and help desks. However, the topic of 'Standards and Certification' reveals a more heterogeneous picture. While some institutes trust the use of internationally accepted standards according to the research practices prevailing in their discipline, other institutes need to develop their own, more specific routines. In summary, all institutes demonstrated an awareness of the necessity of professional research data management, but it still needs to be put more into practice.

## 'Technology' Table

The 'Technology' table focused on the current state of RDM at the participants' institutions, the tools and technologies being used and the technological requirements for further RDM development as perceived by the institutions' academic staff.

It became evident that in terms of developing, evaluating, and implementing institution-wide RDM policies the situation at the different institutions is currently quite heterogeneous. Most institutions are in the process of developing standards for managing scientific data, bundling it within an institutional policy and testing it in pilot studies. In a few institutions this process is quite advanced, with their policies being published and routinely applied to research projects while others have only just begun to scrutinise how the availability and interoperability of their data can be improved by enforcing RDM best practices. Notably, the development and implementation of policies is more advanced in institutions where the management takes an active interest in RDM issues and allocates resources to the development of data management and plans. The participants stressed the importance of developing policies in close collaboration with researchers to ensure their relevance and adoption.

RDM is based on a variety of technologies at the participants' institutions. While the institutional technological toolboxes have diverse components, practically all management efforts rely in one way or another on standardised metadata. This makes the development, dissemination and application of such standards essential.

Information system technologies range from relational databases, document-based data stores, such as wikis, to sharepoint services. The choice of technology is generally highly diverse and, as well as offering a choice between in-house development or customised solutions for software products, is motivated by the existing specific workflows and data flows designed by the researchers. Where experimental research is conducted, the use of electronic lab notebooks is generally perceived as desirable and a number of institutions already use them or are in the process of testing them.

The technical requirements for improving RDM practices and for fostering its adoption into the scientific workflow can be summarised in two major points:

Firstly, technology and tools are needed, which enable the integration of RDM practices into scientific workflows in a user-friendly, non-technical manner and with as little financial investment as possible. Important requirements in this respect are:

1. support for the integration of heterogeneous data from diverse and distributed sources,
2. compatibility with legacy systems,
3. automatic annotation of device-generated data,
4. support for enrichment with metadata, and
5. solutions for high-throughput and large volume data (petabytes).

Secondly, any technological solutions must be augmented by measures to increase the acceptance of RDM practices among users. This includes:

1. enabling users to make informed choices regarding usage and attribution of their data and their results,
2. addressing data privacy, licensing, and intellectual property rights concerns,

3. providing ways to measure data accuracy and trustworthiness of data sources and repositories, and
4. offering easy to understand documentation of the institution's RDM goals and practices.

### **'Society & values' Table**

The 'Society & values' table posed the question: 'In your opinion, what are the most important (social) problems and challenges that need to be addressed regarding the management of research data (research, awareness, implementation, politics)?'

The topics raised and discussed at this table can be clustered into four different areas:

One major aspect was the transparency and openness of data. The transfer, translation and dissemination of methodological knowledge was mentioned by the participants as a requirement for the 'demystification' of methods, increasing reproducibility and the avoidance of manipulated results. Although openness of science is favoured by all disciplines, patent-relevant research results are an exception. Also, this openness includes the risk of generating (incorrect) causal relationships with critical outcomes for society as a result of the partial evaluation of Big Data by third parties compare (O'Neil 2016).

In addition, it is important to protect open data from subsequent private appropriation. It may also be necessary to evaluate the risk of 'Big Data' methods producing causal relationships between formerly independent data segments.

The efficient use of data was another subject raised at this table. To reduce (personnel and financial) resources data must be available for use and reuse. This, however, requires a guaranteed level of reliability and quality of data as well as the provision of software for data evaluation and re-use. The participants stressed that excessive bureaucratisation caused by the need for performance criteria might prevent efficient data (re)use.

The participants also focused on the question of accountability. They were asked for input on how to find a balance between public funding and scientific freedom. What needs to be reported on if there is public funding and how can this be done without constraining scientific freedom? The issue of scientific reputation of published data was also a topic at this table. The acknowledgement of the value of data publication could promote its re-use, however the extension of the current rating system for scientific output might entail a yet unknown risk. The unlimited collection of digital data by various competing actors was also identified as problematic.

Finally, citizen science and the appreciation of citizen scientists was another topic at the 'Society & values' table. Participants posed the question, "To what extent can individual citizens benefit from contributing to science without being exploited and how can contributions from citizens be acknowledged?" Furthermore, there were discussions regarding whether citizen science is simply hype or whether it really takes the responsibility of society as a whole into account. Also, the data collected by citizen scientists was discussed, including its quality and (re-) usability.

## ‘Personal Objectives’ Table

At the ‘Personal Objectives’ table the experts were asked about their personal experiences, expectations and wishes related to the main topics arising from the incentive talks on RDM and NFDI.

For RDM, one thematic cluster dealt with the structure for RDM, including structured processes, structured documentation, folder and file management and consistent data formats. These are as important as the structured implementation of working groups, preferably using a top-down approach. RDM should be an integral element in the processing of projects. Most participants wished that RDM was seen by researchers as an integral part of research and not as an additional task.

RDM implementation was another major topic. In many institutions, the acceptance of RDM is higher among younger researchers, and acceptance by research group leaders and professors, for example, is urgently needed. RDM needs to be seen as an institutional task, even as a key task, within an organisation. Management must acknowledge that the implementation of RDM and RDM policies is required. As far as funding bodies are concerned, the participants would support an intense review of RDM in grant applications. All in all, the experts present stated that the status of RDM and managers should be improved.

The experts present expressed a need to motivate researchers and institutions to pay more attention to RDM by making clear:

1. that documenting the research process, as part of RDM, is an important part of good research practice and of science in general;
2. that RDM should be presented as a method for smooth work processes, as working with more structured data is much more efficient; and
3. that RDM makes their work easier in the end.

In this context, participants claimed that a cost-benefit analysis between RDM and data reuse could be helpful incentives. Researchers could be motivated to practice RDM by:

1. a reward system,
2. more pressure through (financial) incentives and
3. an improved awareness of, and reputation for RDM. All participants agreed that the publication of research data needs to be acknowledged as a valid (scientific) publication.

A general lack of willingness to share data again highlights the need for better training and awareness of the importance of RDM. As universities do not emphasize the benefits of collaboration, young researchers might need additional conversations and training workshops to adopt cooperation and share their data. Junior scientists need to be trained as RDM-promoter so that they can pass their knowledge on to other staff members.

In the participants' assessments of NFDI, collaborative working was an important point. The experts present appreciated the initiative because they hope that parallel structures will be reduced, that duplications will be avoided and that institutes can benefit from the expertise of others. Common workflows, processes and data access should be integrated into the process. The overall goal would be to achieve a community culture change and to strengthen the acceptance of RDM.

Yet, in quite a number of institutes the prevailing situation seems to be one of uncertainty and missing information. It was stated that outreach and communication as well as support and orientation are important. At the moment it seems rather unclear how to contribute best since the process seems to be difficult to understand and is seen to be too abstract and not related to daily practices. This especially applies to smaller academic communities where institutions must first address RDM before committing to NFDI.

All in all, the experts argued that NFDI should be a European initiative, and not be restricted to the national level.

### **'Community Perspective' Table**

With regard to how research data is handled, it is regularly pointed out that it is necessary to take into account discipline-specific approaches and working methods in RDM as well as the particularities resulting from the respective research data itself. The 'Community Perspective' table dealt primarily with the question which themes or fields of action are topical within the respective communities or require special attention, and whether solutions or approaches already exist in the respective fields of action. In addition, participants were asked whether there are particularities in the respective fields of work for which specific solutions are needed; or conversely, whether there are solutions that could also be relevant for other disciplines or communities.

The lack of (subject-specific and interdisciplinary) standards was regularly referred to as a central challenge. It is a deficit that affects all areas, from the collection or generation of data, its processing and analysis, through to publication and archiving. The main desires of the participants mentioned in this context were standards related to the quality of research data and standards for the description, documentation and indexing of research data (metadata, ontologies). A further important aspect in this context is that it is necessary to apply discipline-specific standards while also ensuring that they can be linked to comprehensive and global standards.

In addition to the lack of common standards there is also the issue of complexity, caused not least by the wide variety of discipline- and community-specific approaches and solutions. Participants cited the large and growing number of repositories as an example. While this fundamentally positive development leads to more research data being accessible, it also greatly reduces findability. Existing meta-searches mitigate the problem, but do not completely solve it.

Another challenge that was repeatedly mentioned by the participants is the lack of willingness on the part of researchers to share or publish their research data. However, it was not possible to determine from the discussions whether there are community-specific reasons or conditions for this. It seems that similar structural conditions and individual motives are the obstacles to data sharing in the various communities.

An additional central topic area discussed by the experts goes hand in hand with the increased expectations regarding scientific work, namely the added emphasis on open scientific practices, which produce comprehensible results. With regard to how research data is handled, all communities must be able to answer the question of how the reproducibility of research results can be ensured in concrete terms.

Finally, it is important to mention a fundamental aspect that arose during the discussion of subject-specific approaches. The experts present agreed that many areas of RDM require subject-specific approaches. At the same time, however, this poses special challenges. One is that many Leibniz Institutions comprise various communities, and sometimes also disciplines, under one roof. Due to the different subject-specific approaches used in RDM, this creates the problem of applying uniform policies and standards within the individual institutes and of developing and offering uniform solutions for dealing with data generated or processed in-house. This often creates a conflict between the need to develop a uniform institutional view of RDM and the (heterogeneous) needs of the different in-house communities.

## **Discussion by Leibniz Sections**

In the last session the experts gathered according to their institutional affiliation in the various sections of the Leibniz Association. Different aspects of RDM addressed by the previous discussions among participants with different scientific backgrounds (see above) were now re-evaluated focusing on section- or discipline-specific aspects respectively.

Experts from Section A (Humanities and Educational Research) stressed that a common research data centre for education, used by several Leibniz Institutes, would be very beneficial for the whole community. Participants of Section B (Economics, Social Sciences, Spatial Research) used the opportunity to once more discuss the benefits and challenges of NFDI for their institutions. In terms of technology, in Section C (Life Sciences) pilot projects on electronic laboratory notebooks and experiences in the development of institutional research information systems (Forschungsinformationssystem) or the use of commercial systems were important connecting factors and participants wished for closer collaboration and communication. Referring to the 'Reflection' table, experts present from Section D (Mathematics, Natural Sciences, Engineering) exchanged ideas on the current state of RDM in their institutions and asked for a better exchange of experiences (e.g. in terms of guidelines, policies, inclusion of committees). In terms of 'Personal Objectives' it was again highlighted in Section E (Environmental Sciences) that more personnel resources are needed for proper RDM services. Suppl. material 4 shows all aspects raised during the discussions in the different Leibniz Sections (in German).

The exchange among RDM experts within one section was proven to be very insightful and fruitful. Since the intra-institutional challenges are very similar, successful approaches in improving RDM in one institution can easily be transferred to other institutions as well. All sections emphasized the importance of closer collaboration and better communication in terms of RDM among institutions of the same Leibniz Section in the future.

## Conclusions & Outlook

The workshop 'Round of research data management experts' revealed that most institutions, independent of scientific fields, face similar overarching problems with respect to research data management (RDM), e.g. missing incentives and no awareness of the benefits that would arise from a proper RDM and data sharing.

Based on the exchange at the different tables and concluding discussions by Leibniz Sections, major differences in status and challenges could be either linked to the position of RDM within the institution or the stance of the institution's leadership with respect to RDM. The position within the organisation chart of staff responsible for RDM and its progress varies greatly among institutions. RDM experts are assigned to the directorate, the administration (e.g. scientific reporting), or are situated in particular scientific working groups (funded by the institution or third-party projects). RDM is most highly developed and regarded at institutions where it is the responsibility of the directorate. This is also true for institutions where the leadership itself (e.g. directors, leading scientists) appreciate the importance of good RDM. On the other hand, major differences regarding challenge sare more discipline-specific or even individual in nature. They are composed of differences in data types, formats, tools as well as research processes.

The Guidelines on the Handling of Research Data within the Leibniz Association (Leibniz Association 2019) have only recently been published. In order to assess their current influence as well as their reach and impact after one year of publication the *AK Forschungsdaten* will conduct a survey for all Leibniz Institutions in the summer of 2020. The survey will also record the general situation of research data handling at the different Leibniz Institutions. The outcomes will be summarized in the next meeting of the Working Group in 2020. This will further facilitate the exchange among different institutions within Leibniz and beyond with respect to the latest issues and challenges in RDM at non-university research institutions.

The next workshop, planned for late 2020, will address workflows for RDM and their state of implementation. In this context, both technical and organizational processes that structure and facilitate the handling of research data at research institutes will be evaluated. The aim of the workshop will be to analyse and exchange information about different standardized processes at the different Leibniz Institutes on the basis of examples and, if applicable, to develop ideas for optimising existing processes in one's own institute. This topic also addresses the need for more exchange and networking across the different institutes, which was one of the main requests formulated by the participants of the workshop 'Round of research data management experts' described herein.

The workshop in the summer of 2019 and the following conversation acknowledge the importance of the Research Data Working Group of the Leibniz Association. It is a place for the transdisciplinary exchange of all topics around RDM and enables discussions on how to improve and sustain RDM, independent of institution and scientific domain.

## Acknowledgements

We would like to thank Christian Bölling, Hans-Jürgen Schwarz and Daniel Langenhaun for moderating tables in the World Café and Christian Bölling for contributing to the text. We would also like to express our thanks to Ute Scheich and Brian Baltruschat for their assistance during the planning phase and for conducting the workshop at the Museum für Naturkunde Berlin. Anonymous reviewers significantly improved the manuscript by editing the language and proofreading. The publication of this article was funded by the Open Access Fund of the Leibniz Association.

## References

- Leibniz Association (2018) Leibniz Association. <https://www.leibniz-gemeinschaft.de/en.html>. Accessed on: 2020-5-18.
- Leibniz Association (2019) Guidelines on the Handling of Research Data within the Leibniz Association, Adopted by the General Assembly of the Leibniz Association on 29 November 2018. [https://www.leibniz-gemeinschaft.de/fileadmin/user\\_upload/Bilder\\_und\\_Downloads/Forschung/Open\\_Science/Leitlinie\\_Forschungsdaten\\_2018\\_EN.pdf](https://www.leibniz-gemeinschaft.de/fileadmin/user_upload/Bilder_und_Downloads/Forschung/Open_Science/Leitlinie_Forschungsdaten_2018_EN.pdf). Accessed on: 2020-5-18.
- O’Neil C (2016) Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown, New York.

## Supplementary materials

### Suppl. material 1: Programm FDM-Expert\*innentreffen 18.06.2019 (german) [doi](#)

**Authors:** Petersen et al.

**Data type:** Workshop Programm

**Brief description:** Programm FDM-Expert\*innentreffen 18.06.2019 (german)

[Download file](#) (71.24 kb)

### Suppl. material 2: Forschungsdatenmanagement mit NFDI: Wie sind wir vorbereitet? Wie partizipieren wir? [doi](#)

**Authors:** Harry Enke

**Data type:** presentation slides

**Brief description:** Presentation on “Expert\*innenrunde zu Forschungsdatenmanagement“, 8th workshop AK Forschungsdaten, 18. June 2019, Berlin

[Download file](#) (3.06 MB)

**Suppl. material 3: Praxisbericht: Entwicklung eines Maßnahmenkatalogs zur Verbesserung des Forschungsdatenmanagements am Herder-Institut für historische Ostmitteleuropaforschung** [doi](#)

**Authors:** Stephanie Palek

**Data type:** presentation slides

**Brief description:** Presentation on "Expert\*innenrunde zu Forschungsdatenmanagement", 8th workshop AK Forschungsdaten, 18. June 2019, Berlin

[Download file](#) (678.09 kb)

**Suppl. material 4: Photographs of notes from the discussions in the Leibniz Sessions** [doi](#)

**Authors:** Petersen et al.

**Data type:** image

**Brief description:** Shown are the aspects discussed with respect to the World Café table topics during the workshop in all five sections of the Leibniz Association (Section A - Humanities and Educational Research; Section B - Economics, Social Sciences, Spatial Research; Section C - Life Sciences; Section D - Mathematics, Natural Sciences, Engineering; Section E - Environmental Sciences).

[Download file](#) (1.06 MB)

## Endnotes

- \*1 The Research Data Working Group (*AK Forschungsdaten*) was founded in 2009. The working group actively contributes to finding solutions to processes and producing statements on different aspects of research data and its management within the Leibniz Association. Its members come from across the different Leibniz Sections (A-E, [Link](#)) and represent research institutions and research infrastructure facilities. The Working Group is led by elected spokespersons from all sections who also organise meetings and workshops. For more information please see: <https://escience.aip.de/ak-forschungsdaten>
- \*2 The aim of the national research data infrastructure (NFDI) is to enable and support a community driven framework to systematically manage scientific and research data, provide long-term data storage, backup and accessibility, and network the data both nationally and internationally. For more information please see: [https://www.dfg.de/en/research\\_funding/programmes/nfdi/index.html](https://www.dfg.de/en/research_funding/programmes/nfdi/index.html)