



How to Design a Research Data Management Platform? Technical, Organizational and Individual Perspectives and Their Relations

Lennart Hofeditz¹, Björn Ross¹ (✉) , Konstantin Wilms¹, Marius Rother¹,
Stephanie Rehwald¹ , Bela Brenger² , Ania López¹ , Raimund Vogl³ ,
and Dominik Rudolph³ 

¹ University of Duisburg-Essen, 47057 Duisburg, Germany
bjoern.ross@uni-due.de, ross.bjoern@gmail.com

² RWTH, Aachen, 52062 Aachen, Germany

³ University of Münster, 48149 Münster, Germany
d.rudolph@uni-muenster.de

Abstract. Academic research generates increasing amounts of data that needs to be shared between collaborators, made publicly accessible and/or archived for the long term, all while respecting applicable regulations on topics such as data protection. Good research data management (RDM) is challenging, and despite the growing number of available technical solutions, researchers have been reluctant to use them. We interviewed 64 academic researchers to elicit requirements and explore attitudes towards RDM. Although many funding bodies insist that each project follow published RDM guidelines, only about half the participants considered RDM relevant for their own work, and only one in three reported that they already practiced RDM or were planning to do so. The qualitative analysis of the transcripts revealed three broad categories of requirements for RDM platforms, namely technical, organizational and individual ones. We discuss how these requirements are related to, and sometimes contradict each other.

Keywords: Research data management · Open science · Qualitative research · RDM platform · Design requirements

1 Introduction

Against the background of ongoing digitization in research, an increasing amount of data is being generated by researchers at higher education institutions. As this amount of data seems to grow continuously, researchers will need to change the way they manage their data in order to keep track of it and be able to collaborate effectively. In addition, public funders of research projects such as the European Commission, the German Research Foundation (DFG), and the German Federal Ministry of Education and Research (BMBF) require researchers to maintain their data and make it publicly available. Consequently, academics of all disciplines will need to rethink their data management strategies in order to be able to manage the growing and increasingly complex

amount of data in the future. To accompany this process, the DFG has published guidelines for research data management (RDM) aimed at researchers and institutions. RDM can be defined as the organization of data, from its collection to its publication and the archiving of results [1]. RDM also includes the management of data through infrastructure, long-term storage, data security, open access, as well as communication between researchers from different disciplines [2].

Vines et al. [3] reported that the availability of data from published studies decreases with the age of publication. Specifically, the probability of data being available decreases by 17% annually after its publication [3]. Open access to research results offers the opportunity to confirm or disprove those results [4]; it consequently offers a control function that ultimately benefits quality. Furthermore, a good and transparent RDM protects against accusations of scientific misconduct [5].

RDM also provides advantages for individual researchers. Researchers who grant access to their research data are cited more often than those who do not [6]. RDM has the potential to positively support individual researchers as well as research groups [7].

Although corresponding specifications exist, there is often a lack of appropriate technical and organizational capabilities to implement a research data management infrastructure in an institution [8]. Knowledge of how to design RDM platforms and services remains very low. But how can an RDM artefact be designed so that it makes a relevant contribution to research? Are there possibly already functionalities in established systems that are used and would, therefore, be suitable to be transferred into a research data management platform? As there are no comprehensive answers to these issues, we derived the following research question:

RQ: How do RDM platforms need to be designed to support academic research?

We conducted and evaluated group interviews with researchers in workshops in order to identify requirements for a user-centered RDM platform. We conducted 16 workshops with a total of 64 participants. We evaluated the interviews with a qualitative content analysis following Mayring [9]. From the results we derived technical, organizational and individual requirements for an RDM platform and identified relationships between them.

2 Literature Review and Theoretical Background

RDM is becoming an increasingly important topic, especially in the university context. More and more institutions are beginning to develop their own technical solutions to store research data for the future [10]. One trigger for this could be the research funding of third-party funding bodies, which demand RDM accordingly [11].

Perrier et al. [12] also showed that the number of publications on RDM has increased significantly since 2010, which confirms the growing interest in the topic. Different disciplines have different requirements for research data and its management. For this reason, there is currently no uniform definition of RDM. RDM comprises the organization of data from the collection to publication and archiving of the results [1]. It also includes the management of data through an infrastructure, a long-term storage facility, data security, open access, as well as communication between researchers from different disciplines

and research fields [2]. In most cases, however, the aim is to ensure that digital research data is made available to other researchers [13]. However, RDM is not only important for sharing data, but for the entire research cycle. Starting with data collection, through analysis, to the evaluation and interpretation of the data, correct RDM can achieve significant improvements [1]. At the same time, the ongoing digitization creates ever larger amounts of data [14]. Accordingly, researchers today are confronted with an abundance of data and data types [15]. Consequently, researchers need to rethink and change their data management strategies in order to be able to manage the ever-growing and increasingly complex data volumes of the future.

Although data sharing brings many opportunities and benefits and can accelerate the research process, it is still not a common practice [16] and data is more likely to be withheld than published in scientific journals [17]. Citing other researchers is very common and is a reward for the cited researcher [18]. However, sharing data is not based on this formalism and does not offer researchers any perceptible recognition. Thus, data sharing is not yet established as a method of communication within the research community [18, 19]. Nevertheless, researchers are more inclined to share data if they expect it to be beneficial for their own careers and if the risks and effort involved appear low [20].

At the same time, researchers raise questions about privacy and the security of their own data and records [14]. There are also concerns about copyright protection in public cloud storage [21]. Researchers are skeptical as to whether collected data is not passed on to third parties under copyright protection [8].

Furthermore, they are afraid of misinterpretation of their data once they are used and interpreted in other contexts [18], and they fear losing control over their data once it is released [22]. These barriers may therefore have social rather than technical reasons [8, 23].

However, although many open access initiatives exist, there are only moderate results of applications and services [24]. In this context, there are several studies that have examined and compared existing systems and solutions [10, 12, 25, 26]. Süptitz et al. [25] showed that a general distinction can be made between functional (technical) and non-functional (framework conditions such as data security, data protection and usability) requirements [10]. Guidelines have been developed and issued at the international level on how research data should be handled [27]. In addition, researchers are required to explain at the time of application for funding how data will be handled after the conclusion of the project and what measures will be taken with regard to the sustainability of the data [11]. Wilms et al [23] compared the guidelines of ten national and international science and funding institutions. In summary, technical and non-technical factors influencing the use of RDM could be identified. According to the study, technical factors are related to the infrastructure, i.e. the platform itself, and the guidelines of the platform. Further technical factors were data security, data sharing and data maintenance. Non-technical factors were ethical factors, such as the handling of data obtained from human research, the management of, for example, false findings and factors affecting the researcher herself. The main issues here were the protection of intellectual property and incentives to use platforms for RDM.

In summary, it can be said that RDM is not only a national but also an international issue that many institutions, especially universities, are dealing with. In addition, a great deal of money is being invested by both universities and funding bodies to implement and provide technical solutions, such as virtual research environments and repositories, so that researchers can store their research data in a sustainable manner. In this context, those universities and funding institutions have developed and published guidelines for the handling of research data. There are to date no internationally accepted guidelines for handling research data. Guidelines vary by country and sometimes even within countries, which does not simplify the exchange of research data. Correspondingly, the implemented technical solutions are not or only insufficiently used, because the researchers have reservations about the new technologies and the requirements for these systems vary by discipline. Therefore, the requirements of both the disciplines and the individual researchers must be identified. Reservations and concerns must be identified and counteracted. It remains unclear how technical, social and individual requirements are interrelated, and it is essential to find out which functions researchers need in order to feel supported in their RDM.

3 Research Design

Since research in this field is still in an early stage, we chose a qualitative approach. We conducted 16 semi-structured group interviews with researchers as part of workshops on RDM. An open discussion prevented the participants from being influenced too much by given answers. A guide consisting of open-ended questions ensured that the participants' views were given the spotlight and that their statements were as unbiased as possible by the preconceptions of the workshop organizers. The workshops gave the participants a lot of freedom for discussion. The purpose of the workshops was not only to answer the research question, but also to discuss general attitudes towards RDM. The main questions in the guide resulted from the design science approach according to Hevner et al. [28] and from the current state of scientific research on functions and technical properties of research environments for RDM.

In the first part of the interviews, the interviewer gave a short presentation in which RDM was first introduced and a shared understanding of RDM was ensured with the help of a definition of the DFG. In the second part of the interviews, the aim was to find out the current status quo within the research disciplines. We asked to what extent RDM was already important in the participants' daily work, whether and to what extent they had already come into contact with RDM, and which aspects of RDM were already used within their respective disciplines.

The third part of the interview was about finding out the criteria or functionalities that a platform would have to fulfill to be attractive to the researchers and how a technical solution should be designed to support RDM in the future. If the discussion did not make progress, there were topics that were optionally addressed. Possible topics could be data security, the release of research data, long-term storage and the documentation of the research.

We conducted the group interviews in the period from 28/09/17 to 29/11/17 within the context of a project at the University of Duisburg-Essen and RWTH Aachen.

The interviews were conducted face-to-face by three different workshop organizers and the audio recorded electronically. Subsequently, all interviews were transcribed. The completely transcribed interviews were then evaluated using the deductive qualitative content analysis approach described by Mayring [29].

A total of 16 group interviews with 64 participants were conducted. The group interviews were divided into two discussion groups within the life and social sciences discipline, three group discussions within the humanities, three group discussions within the natural sciences, 5 workshops within the engineering sciences and 1 workshop included members of the Commission for Information, Communication and Media Technology at the University of Duisburg-Essen which is responsible for data protection and security. The number of participants per workshop varied between one and nine participants, and the total duration of the audio recordings varied between 12 and 73 min. The number of participants per university and discipline is shown in Table 1.

Table 1. Participants per discipline and institution

Institution	ICM	HUM	ENG	NAT	LIF	Total
RWTH Aachen	0	7	8	10	0	25
University of Duisburg-Essen	7	11	10	8	3	39
Total	7	18	18	18	3	64

ICM = Information, Communication and Media; HUM = Humanities; ENG = Engineering Sciences; NAT = Natural Sciences; LIF = Life Sciences

4 Findings

In general, we were able to determine that RDM is already relevant across disciplines and that its relevance for personal research work has increased. RDM was considered relevant for their work by about 50% of the workshop participants, while for the other half, RDM was less relevant or no precise information was given in the interviews. RDM was considered particularly relevant in engineering and life sciences and was either already applied or its introduction planned.

A majority of the researchers stated that they would store their data on local computers at their workplace or they conducted analog laboratory books to record research hypotheses and experiments. In addition, there are already common standards and set documentation in the life and social science discipline. Commercial services such as Dropbox and Google Drive with version control features were also used to store documents and their history. In the humanities, data reuse also plays an important role, since corpora (collections of texts), some of which are 40 years old, are still in use.

Basically, we could classify the requirements of the researchers into three main categories: Technical, organizational and individual requirements.

4.1 Technical Requirements

The most frequently mentioned aspects were data security and data protection. The interviewees were concerned with protection against misuse, sufficient anonymization or encryption of the data. Data should therefore be adequately protected against unauthorized access.

Another important requirement was data sharing. That means permanent access to data by others who need it in order to ensure smooth cooperation. In addition, sharing data can contribute to research transparency and this could lead to better research quality. Access and release control are also part of data sharing. It should be made clear for whom the data is released, who can access the data and in what role. Finally, technical collaboration and open access requirements were mentioned to improve the use of RDM. More detailed requirements were therefore the ability to work together with shared data, so that one can work on and with the data at the same time.

Another aspect stated by the interviewees was information retention. This includes long-term availability and the sustainability and reusability of data. Long-term availability includes the requirement for (automatic) backups and corresponding specifications from third-party funders, so that the data can still be accessed much later (for example, ten years after the project has ended). The researchers also mentioned the need for a kind of “software asset management” in order to be able to use corresponding, sometimes very special, file formats even at later times. Finally, location-dependent access to research data should ideally be possible so that the necessary data can always be accessed when required. A central point of the reusability of the data is the possibility that other researchers have access to one’s research data in order to reuse it. Researchers should be able, or even obliged, to attach the raw data to their publications. This would also have a positive effect on the reproducibility of the results, the researchers’ credibility and reputation. In this way, the data would also be tested in new contexts and could thus provide insights that were previously unthinkable.

If the data will be made openly available for other researchers, the systems that make the data available should accept many file formats or, alternatively, standardize the data formats so the researchers can handle them.

According to the interviewees, the possible application of metadata and documentation were other requirements for successfully performing RDM. Standards must be established and, at best, metadata has to be stored automatically as far as possible in order to find the data, if necessary, also with the help of “human data managers” to ensure a certain data quality.

Usability is another aspect that was often mentioned. Simple and intuitive user interfaces should be provided for a technical solution to be accepted.

Hardware and memory, a high bandwidth, faster data transfer, data security, and availability, as well as scalability, were stated as further technical requirements.

4.2 Organizational Requirements

As non-technical or organizational aspects we considered requirements that cannot be implemented through technical functionalities. For an overarching RDM, there must be

rules, policies and standards to regulate which data should be tracked and which data types are important.

The interviewees also mentioned finance, personnel structure and administration as organizational requirements. Knowledge about the solution has to be gained, appropriate marketing measures have to be carried out and clarified in order to achieve a high level of technology acceptance.

Those solutions should also be financed and organized accordingly. The system or service could therefore be offered as cheaply as possible by universities (or similar institutions). The existence of good working technical support is also important, according to our findings. This requires long-term preparation and organization by the institutions.

Legal security is also another organizational aspect, especially for data to which a large number of possible users have access. Clear responsibilities, an opportunity for legal advice and legal certainty over the entire research cycle are required.

An additional important aspect is ethics, which is fundamental for RDM since depending on the discipline, sensitive personal data about human participants is collected.

4.3 Individual Requirements

We classified all aspects that could not be classified either in the technical requirements or the organizational requirements as individual aspects. However, this does not mean that there is no overlap with the categories of technical and organizational aspects. Rather, the aim was to show and point out connections and overlaps between subject areas.

The individual aspects were mainly concerned with the individuals in the context of technology and organization. Mostly these aspects were characterized by skepticism or fear, since new developments always bring along uncertainty. We also assigned the researchers' personal attitude towards the RDM to this category.

The participants stated that, according to the specifications of the third-party funding bodies, the research data should be published as early as possible and made available to others. However, the researchers were afraid of "knowledge theft" if the data was released too early (before publication). They would like to prevent another researcher from taking credit for their research achievements.

Due to the pressures of research and publication, it is considered a clear advantage to have sole access to one's own data. This aspect influences the subsequent use of the data accordingly.

The transfer of data was another point of uncertainty mentioned in the interviews. On the one hand, researchers want to release their data according to good scientific practice. On the other hand, researchers wonder which and to what extent data may be passed on and whether a loss of control over the data would be possible. The aspects of data protection and legal certainty played a prominent role.

In the opinions of the researchers, an awareness must be developed that research is done for "eternity", i.e. for more than ten years, and that the data must be documented in such a way that they can still be used for many years to come.

However, it would also require self-discipline, as even today the researchers do not always understand the documentation of their own research from three years earlier.

In order to contribute to traceability, it is important that metadata is also used, and that detailed and comprehensible documentation will be produced.

In spite of self-discipline, the effort to use the technical solution must be kept as low as possible. The researchers described that they would probably use the solution little or not at all if the effort to store the data there or to work with them would be too high. This is mainly due to the fact that the workload of researchers is already high anyway and they want to avoid additional effort. The requirement over the entire cycle, from data sharing to reusability, from (long-term) availability to the documentation of the research, would therefore be to keep the additional effort for the individual researcher as low as possible.

4.4 Relationships of the Requirements

As already mentioned in the description of the individual aspects, there are often interrelationships within the categories. The interrelationships identified between the aspects showed that in developing an RDM platform, not only individual aspects must be considered, but the “big picture” must not be overlooked. Thus, technical aspects such as metadata and documentation, as well as sharing data in connection with individual and human aspects should be considered together. This concerns the reservations or even fears that researchers might have about certain technical functionalities. But also, the organizational aspects must not be considered separately. Especially regarding personal data, the ethical framework conditions must be discussed and taken into account. The same applies to questions of legal certainty and organization in general.

Of course, these results are only the first indications. Further research should continue to explore overarching connections between the aspects. Our results justify further research in this field to gain further insights into the perception and requirements of the researchers.

5 Discussion

With regard to the research question, it can be seen that researchers make far-reaching demands on an RDM platform. Technical functions still have to be developed to support research data management, but some of them are already available. Researchers also have detailed ideas about what organizational and individual requirements they have so that they feel supported. One important technical aspect that could be derived from the interviews was the aspect of being able to share research data online. For the researchers, this includes collaboration, access and release control and management. An important organizational aspect that we identified was the need for the exact implementation of existing rules and standards. Overarching rules provide researchers with an orientation guide on how RDM should proceed in detail and which workflows need to be run through. The rules and standards do not only refer to discipline-specific metadata, but also to the data formats and the overall workflows regarding research data. Furthermore, we identified some basic functionalities. These include functions for sharing data, for long-term availability for subsequent use and for the protection and secure use of data. In addition, metadata and file format functionalities should be integrated as these functions play a key role. The APIs of existing software were mentioned as particularly important.

Highly specialized functions from the various research disciplines could be implemented by these programming interfaces. As an example, digital laboratory books, such as those from natural sciences, highly complex calculations from the engineering sciences or text corpora considered in social science, could also be integrated. Lastly already established functions from other contexts should be adapted as far as possible.

A key contribution of this study is to show the relationships between the requirements (see Fig. 1). Not always are these easy, or even possible, to meet at the same time. For example, data sharing is an obvious requirement of a good RDM platform, yet researchers also report a need to ensure that their data is not misused. Any form of data sharing feature that allows users to download research data onto their own machine is likely to entail a small possibility that the data might fall into the wrong hands, for example through theft or careless disposal of hardware. The same holds for the requirements of archiving the data for the term and keeping costs low. A solution that meets both requirements is far from obvious: What happens to the data when the commercial cloud service goes out of business, or politicians are no longer willing to extend a publicly funded project?

For managers of research institutions, we derived organizational recommendations. In cooperation with various research disciplines, rules and standards for metadata must be established in addition to the existing guidelines for handling research data. In addition, the issue of legal certainty must be comprehensively regulated. Finally, managers must deal with the topics of human resources, financing and organization. It must be clarified to what extent new human resources must be made available, how these staff members and the provision of the technical solution can be financed and how the organization can plan for the long term, when the political context might change. In Fig. 1 we summarized all identified technical, organizational and individual requirements and their relationships.

Only about half of the participants stated that they considered RDM relevant. Only one in three reported that RDM was already practiced or that it was planned to introduce such a system. One reason for this lack of awareness could be the timeliness of the topic of RDM. It is possible that the topic has not yet “arrived” in many disciplines, as the topic has only gained momentum in recent years, which can be seen in the increasing number of publications on the topic [12]. It could also be possible that international comparisons will slowly increase attention to this topic. The technical aspects that could be derived from the interviews can already be found in similar categories in the literature that describes the requirements of the researchers [25, 30].

Especially technical functions for data sharing, collaboration or access and release control are essential for RDM to be applicable at all. It is certainly important to recognize that, especially with such centralized functions, the effort required to achieve the goal of sharing data with others must be low, since the willingness to share data decreases as the effort increases [31].

The focus on data protection and security is also currently relevant due to the EU’s General Data Protection Regulation (EU-GDPR). The principle therefore applies that the more personal the data, the more relevant data protection is. Finally, however, it must be recognized that some research data cannot be released for subsequent use because anonymization of the data would be too costly [32]. At the same time, researchers are rather skeptical about the functionalities of data provision and reusability by others, as they fear data misuse or even loss of control over their self-generated data [30].

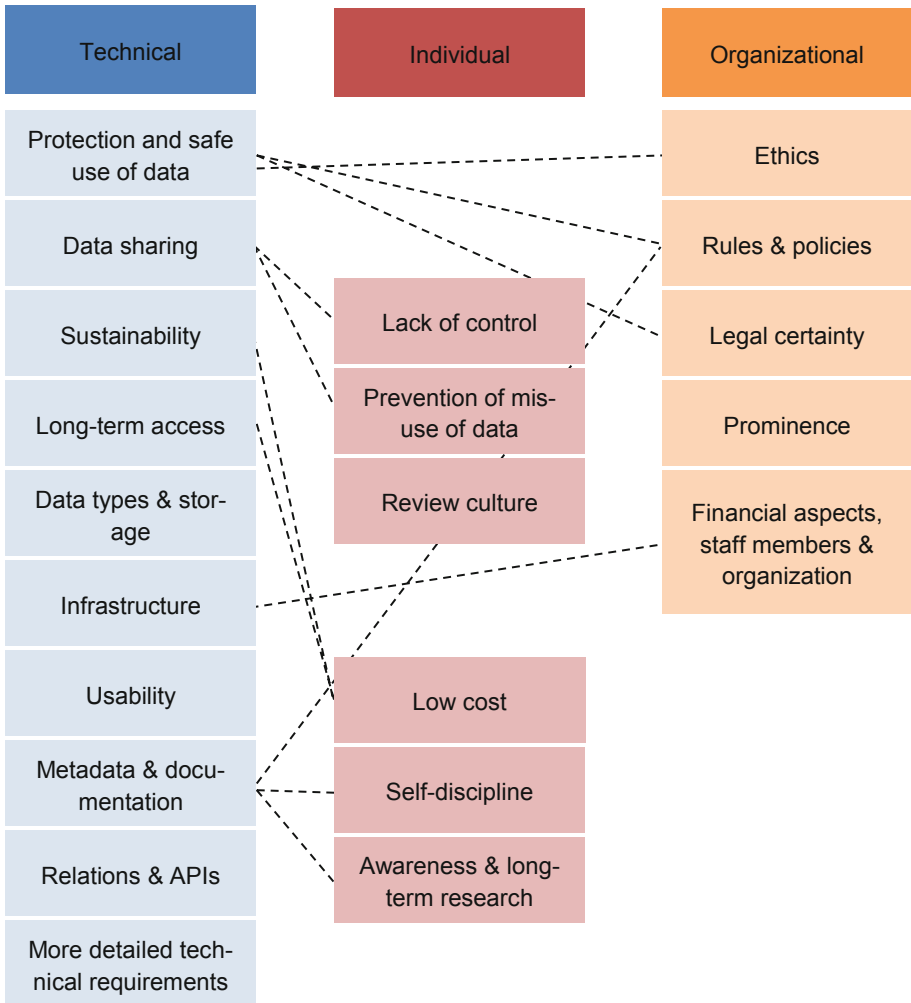


Fig. 1. Identified requirement categories and their relationships

Nevertheless, these functionalities must be in place, as they will become central to research work and are demanded by the various institutions (such as DFG or EU). However, technical functions for mere data storage will not be sufficient to make research data usable in the long term [33]. It must be clarified how the actual information content of the data can be preserved because the added value of information preservation only becomes apparent when the data can be accessed and also analyzed [34].

Research data can also benefit from metadata, as metadata can be used to contextualize and view the data [35]. Thus, the use of metadata also influences the search function and, secondarily, the effectiveness and speed of research. Metadata simplifies data sharing [36]. If there are no standards, this could prevent researchers from ultimately making their data available within the framework of open access [37]. Standards may

therefore have to be defined first. These specifications are needed simply because the unification or standardization of workflows, data formats and metadata will ensure that data can be kept clear in the long term and the associated sharing of data will be greatly simplified [36]. It is unclear who is to issue these rules. As with the aspect of metadata, the individual research communities would probably be the best option here, since they have a high level of expertise in the respective fields and thus know what is relevant, and the guidelines are then not simply prescribed by outsiders, but created involving the very communities who will later have to follow them. In addition to the actual functions, usability must also be given priority. It is therefore essential that the technical solution and all functionalities are designed intuitively so that there is as much willingness as possible among researchers to use the system. The functions and tools must therefore be usable and useful for the target group in the first instance [38].

Despite all possible obstacles, initiatives must be taken to encourage researchers to practice RDM. If one asks the researchers themselves what possible incentives might be, the answer is that one of the biggest incentives would be the increased visibility & impact of their own research [30]. For this reason, there are also three major areas that need to be stimulated in open access solutions: the publication of data, the use of published data and the value added using published data and the value added using published data [39].

In this context, individual aspects must also be considered. Aspects such as loss of control or fear of knowledge theft, as well as the perceived effort that researchers have to “take on” in order to provide research data, must be taken seriously and included if a virtual research environment is to be implemented. There is also evidence that there are individual aspects that are deeply rooted in the researcher, especially when it comes to fears of loss of control and knowledge theft. Likewise, the aspect that first, the awareness has to be developed that one’s own research will be long-term research that should also be available to others, is a point that has yet to be internalized by the researchers. Those points, in turn, seem to have an influence on technical aspects that could affect the possible acceptance and actual use of a technical solution.

The present study also has its limitations. In groups with seven participants or more, it became difficult to assign all of the statements in the interviews to the correct participants during transcription. The group discussion can also mean that the discussed topics are determined by particularly eloquent or convincing participants who strongly voice their opinion.

In summary, it can be stated that a platform for RDM should be more than just the sum of technical functions. Not only the technical requirements have to be considered, but also the organizational requirements as well as the concerns and fears of individual researchers.

6 Conclusion

The goal of this study was to evaluate how RDM platforms need to be designed to support academic researchers. Three categories of requirements were identified.

In general, basic functionalities should be given so that the system can fulfil its purpose at all. These include functions for data sharing, long-term availability for re-use, and the protection and secure use of data. In addition, functionalities for metadata

and file formats should be integrated, as these functions play a key role. In general, however, sufficient infrastructure and usability must be provided for when using the system. In addition, already established functions from other contexts should be adapted if possible, as these functions are then already field-tested and have proven themselves.

In cooperation with various research disciplines, rules and standards for metadata must be established in addition to the existing guidelines for handling research data. Furthermore, the issue of legal certainty must be comprehensively regulated.

Since this study first provides initial indications of connections between various categories and aspects, future studies should deal with the connections between technical and non-technical aspects as well as human concerns and fears in order to be able to make statements in the future as to whether and how these connections could have an influence on user behavior.

References

1. Whyte, A., Tedds, J.: Making the Case for Research Data Management. DCC Briefing Papers. Digital Curation Centre, Edinburgh (2011)
2. Schopf, J., Chaudiron, S., Jacquemin, B., Prost, H., Severo, M., Thiault, F.: Open access to research data in electronic theses and dissertations: an overview. *Libr. Hi Tech.* **32**, 612–627 (2014). <https://doi.org/10.1108/LHT-06-2014-0058>
3. Vines, T.H., et al.: The availability of research data declines rapidly with article age. *Curr. Biol.* **24**, 94–97 (2014). <https://doi.org/10.1016/j.cub.2013.11.014>
4. Agarwal, R., Dhar, V.: Big data, data science, and analytics: the opportunity and challenge for IS research. *Inf. Syst. Res.* **25**, 443–448 (2014). <https://doi.org/10.1287/isre.2014.0546>
5. Joshi, M., Krag, S.S.: Issues in data management. *Sci. Eng. Ethics* **16**, 743–748 (2010)
6. Piwowar, H.A., Day, R.S., Fridsma, D.B.: Sharing detailed research data is associated with increased citation rate. *PLoS ONE* **2** (2007). <https://doi.org/10.1371/journal.pone.0000308>
7. Link, G., Lumbard, K., Germonprez, M., Conboy, K., Feller, J.: Contemporary issues of open data in information systems research: considerations and recommendations. *Commun. Assoc. Inf. Syst.* **41**, 587–610 (2018). <https://doi.org/10.17705/1cais.04125>
8. Yamamoto, S. (ed.): HIMI 2016. LNCS, vol. 9735. Springer, Cham (2016). <https://doi.org/10.1007/978-3-319-40397-7>
9. Mayring, P.: Qualitative Content Analysis (2014). [https://doi.org/10.1016/S1479-3709\(07\)11003-7](https://doi.org/10.1016/S1479-3709(07)11003-7)
10. Amorim, R.C., Castro, J.A., Rocha da Silva, J., Ribeiro, C.: A comparison of research data management platforms: architecture, flexible metadata and interoperability. *Univers. Access Inf. Soc.* **16**, 851–862 (2017). <https://doi.org/10.1007/s10209-016-0475-y>
11. Natural Science Foundation: Data Management & Sharing FAQs (2017)
12. Perrier, L., et al.: Research data management in academic institutions: a scoping review. *PLoS ONE* **12**, 1–14 (2017). <https://doi.org/10.1371/journal.pone.0178261>
13. Reuter, C., Ludwig, T., Kotthaus, C., Kaufhold, M.-A., von Radziewski, E., Pipek, V.: Big data in a crisis? Creating social media datasets for crisis management research. *i-com* **15** (2016). <https://doi.org/10.1515/icom-2016-0036>
14. Pinfield, S., Cox, A.M., Smith, J.: Research data management and libraries: relationships, activities, drivers and influences, pp. 1–28 (2014). <https://doi.org/10.1371/journal.pone.0114734>
15. Hicks, B.: UK universities put their faith in the Google cloud
16. Piwowar, H.A., Chapman, W.W.: A review of journal policies for sharing research data (2008)

17. Campbell, E.G., Clarridge, B.R., Birenbaum, L., Hilgartner, S., Blumenthal, D.: Data withholding in academic genetics: evidence from a national survey. *J. Am. Med. Assoc.* **287**, 473–480 (2002). <https://doi.org/10.1001/jama.287.4.473>
18. Borgmann, C.L.: *Scholarship in the Digital Age*. MIT Press, Cambridge (2007)
19. Tenopir, C., et al.: Data sharing by scientists: practices and perceptions. *PLoS ONE* **6**, 1–21 (2011). <https://doi.org/10.1371/journal.pone.0021101>
20. Kim, Y., Zhang, P.: Understanding data sharing behaviors of STEM researchers: the roles of attitudes, norms, and data repositories. *Libr. Inf. Sci. Res.* **37**(3), 189–200 (2015). <https://doi.org/10.1016/j.lisr.2015.04.006>
21. Hilber, B., Reintzsch, D.: Cloud Computing und Open Source-Wie groß ist die Gefahr des Copyleft bei SaaS? In: *Computer Und Recht: Forum für die Praxis des Rechts der Datenverargei-tung, Information und Automation*. pp. 697–702. Otto Schmidt (2014)
22. Feijen, M.: What researchers want - A literature study of researchers' requirements with respect to storage and access to research data (2011)
23. Wilms, K., Brenger, B., Lopez, A., Rehwald, S.: Open data in higher education – what prevents researchers from sharing research data? In: *39th International Conference on Information System*, pp. 1–9 (2018)
24. Karkin, N., Janssen, M., Brooks, K.: Open government and data-driven policy making in the digital age. In: *Americas Conference on Information System (AMCIS)* (2018)
25. Süptitz, T., Weis, S.J., Eymann, T.: Was müssen Virtual Research Environments leisten? -Ein Literaturreview zu den funktionalen und nichtfunktionalen Anforderungen. In: *Wirtschaftsinformatik* (2013)
26. Bankier, J.G., Gleason, K.: *Institutional repository software comparison*. UNESCO Communication and Information, Paris (2014)
27. Deutsche Forschungsgemeinschaft: *Memorandum Safeguarding Good Scientific Practice*. (2013)
28. Hevner, A., Alexander, B.: Roles of digital innovation in design science research. *Bus. Inf. Syst. Eng.* **61**, 3–8 (2019). <https://doi.org/10.1007/s12599-018-0571-z>
29. Mayring, P.: Qualitative inhaltsanalyse. In: Boehm, A., Mengel, A., Muhr, T. (eds.) *Texte verstehen: Konzepte, Methoden, Werkzeuge*, pp. 159–174. UVK Univ.-Verl. Konstanz, Konstanz (1994)
30. Bauer, B., et al.: *Forschende und ihre Daten. Ergebnisse einer österreichweiten Befragung – Report 2015*. Zenodo (2015). <https://doi.org/10.5281/zenodo.31935>
31. Kim, Y., Stanton, J.M.: Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. *J. Am. Soc. Inf. Sci.* **67**(4), 776–799 (2016). <https://doi.org/10.1002/asi.23424>
32. Klemm, M., Liebold, R.: Qualitative Interviews in der Organisationsforschung. In: Liebig, S., Matiaske, W., Rosenbohm, S. (eds.) *Handbuch Empirische Organisationsforschung*, pp. 299–324. Springer, Wiesbaden (2017). https://doi.org/10.1007/978-3-658-08580-3_13-1
33. Jensen, U.: *Leitlinien zum Management von Forschungsdaten: Sozialwissenschaftliche Umfragedaten*. Cologne, Germany (2012)
34. Berendt, B., Vanschoren, J., Gao, B.: Datenanalyse und -visualisierung. In: Büttner, S., Hobohm, H., Müller, L. (eds.) *Handbuch Forschungsdatenmanagement*, pp. 139–148. Bock + Herchen, Bad Honnef (2011)
35. Link, G., et al.: Contemporary issues of open data in information systems research: considerations and recommendations. *Commun. Assoc. Inf. Syst.* **41**, 587–610 (2017). <https://doi.org/10.1016/j.anireprosci.2016.02.027>
36. Ribes, D., Polk, J.B.: Flexibility relative to what? Change to research infrastructure. *J. Assoc. Inf. Syst.* **15**, 287–305 (2014). <https://doi.org/10.17705/1jais.00360>
37. Vassilakopoulou, P., Skorve, E., Aanestad, M.: A commons perspective on genetic data governance. In: *European Conference on Information Systems (ECIS)* (2016)

38. Biljon, J. V., Pottas, A., Lehong, S., Platz, M.: Content category selection towards a maturity matrix for ICT4D knowledge sharing platforms. In: International Conference on Information Resources Management (Conf-IRM) (2016)
39. Nugroho, R.P., Zuiderwijk, A., Janssen, M., de Jong, M.: A comparison of national open data policies: lessons learned. *Transform. Gov. People, Process Policy* **9**, 286–308 (2015). <https://doi.org/10.1108/TG-03-2014-0008>