Do researchers dream of research data management?

Konstantin L. Wilms
University of Duisburg-Essen
konstantin.wilms@uni-due.de

Stefan Stieglitz
University of Duisburg-Essen
stefan.stieglitz@uni-due.de

Alina Buchholz
University of Duisburg-Essen
alina.buchholz@uni-due.de

Raimund Vogl
University of Münster
rvogl@uni-muenster.de

Dominik Rudolph
University of Münster
d.rudolph@uni-muenster.de

Abstract

The ongoing digitalization of academic work processes has led to a shift in academic work culture where researchers are supposed to take on more responsibility in term of adequate data management. Third party funding institutions as well as high class journals are increasingly asking for standardized data management processes and started to set up policies which should guide researchers to manage their data properly. In this work, we deal with the highly IS relevant topic of research data management (RDM) and provide an overview of the different existing research data management guidelines of the eight biggest governmental funded institutions and the biggest politically-independent institution. All existing guidelines of those institutions were considered in a qualitative analysis, summarized and evaluated. It has been found that non-technical requirements evolve to non-technical barriers, which institutions need to address to a greater extent within their guidelines to promote scientific research. This work shows the shift in the understanding of RDM and provides the present perspective which help researchers to better understand the ongoing trend of RDM within science.

1. Introduction

Due to the rapid digitalization of academic work processes, researcher are forced to deal with a deluge of data among their fields [1], [2]. Research data emerge over the entire research lifecycle, including the collection, analysis and evaluation of data [3]. These data are of high value, since they help to discover new insights and innovations. Researchers, educational institutions, and by this also governments are highly interested in mining available research data, since it leads to a faster progress regarding research in different fields of study [4], [5]. In addition, research data lay the foundation for any research results and constitute the basic prerequisite to prove or reject scientific hypotheses or to replicate empirical statements [6]. Hence, the management of research data is increasingly important for the utilization and preservation of such data [3], [7]. Consequently, there is an increasing demand for consistent data management policies, which can guide researchers to annotate, store, or disclose their research data properly.

However, many researcher refuse to management their research data properly. Today, there are plenty of documented cases where researchers reject to make their data available via open access, despite the fact that different tools and platforms exist, which allow them to easily share and reach research data [8], [9]. Within the work of [10], requests to authors of PLOS journals to ask for their research data are sent; but only 10% of all requests were answered. The reason to withhold research data differ: Researchers often fear security issues and a possible misappropriation of their data [6]. Additionally, ethical reasons hinder researchers to publish their data, mainly because the data was gained from human individuals, who would probably not agree with an openly publication of their personal data [11].

As a consequence, different institutions (majorly funded by governments) set up data policies and try to define norms which help researchers to conduct adequate RDM [2], [12]. Such guidelines include policies on e.g. the management of data via infrastructures, long-term storage, the disclosure of research data, preservation, ethical aspects, and communication between researchers as well as different fields of study [12], [13]. All those varying guidelines and requirements from different funding institutions made it difficult for researchers to practice proper RDM [12].

As latest studies indicate, researchers' understanding regarding RDM differs, and several researchers do not even know RDM or RDM

HICSS

guidelines in general. Moreover, it became difficult for research infrastructure developer to design suitable information systems to support researchers in this regard [9], [14]. The rapid changes in technology created a gap between researchers understanding of data management and the requirements of third-party organizations. Still, a common understanding regarding the conceptualization of RDM is missing, as well as there is a lack of scientific work to encourage researchers to discuss the shift in requirements. To support both researchers and IT developers concerning the compliance with different international RDM restrictions, this work focuses on current themes and resulting categories, such as technical and non-technical requirements. Due to that, the following research question was generated:

Which requirements are set up by (governmental) funded institutions in order to conduct research data management?

Concerning the research method, the guidelines of eight governmental funded institutions, as well as one politically independent institution are collected and compared with each other. Before, the guidelines are categorized by an inductive qualitative analysis based on the suggested process by [15]. The possible finding of new requirements are of high significance for future research, since they might be the main reason to prevent research data management on behalf of the researchers [6].

The structure of this work is presented as follows: First, the paper provides a literature review with focus on the current status quo of RDM research. In the following, the qualitative content analysis will be explained, just as the eligibility of the eight examined institutions. Next, the determined categories, as the results, will be presented. Finally, our results will be discussed and research implications will be given.

2. Literature Review

Until today, there is no common definition of the term "research data management", especially since researchers from different disciplines have varying opinions and perceptions towards RDM. Nevertheless, some definitions do exist: [3] for example define research data management as an "organization of data, from its entry to the research cycle through the dissemination and archiving of valuable results" [3](p. 1). It includes, for instance, the management of data via infrastructures, the long-term storage and security of data, and open access, but also communication between researchers and different fields of study [13]. Data constitute the basis of research, scientific progress and communication, and due to that its management is of huge importance towards the

scientific community [6]. As implied by [16] and [17], data constitute the main source of knowledge, since they are hierarchically related to both information and knowledge. Overall, RDM is said to offer different positive aspects, which mostly occur due to the increased collaboration and data sharing among researcher. As stated by [18] authors who provide access to their research data are more likely to be cited by the scientific community. According to the long tail theory, shared research data have the potential to provide increased knowledge as the data are discovered and used by new audiences [19]. Hence RDM has the potential to overcome bottleneck effects which show up if research insights are only represented within a specific community or academic field [20]. Hence, RDM is said to offer an adjusted pressure by journals, perceived benefits regarding the own career, and altruism; while on the contrary, the factor perceived effort had a negative effect on research data management [21]. Although researchers often fear that giving access to their data might lead to negative criticism about their research, studies revealed that openly shared data in fact defend against accusations by the granting of insights [22]–[24].

Mental and physical barriers that inhibit efficient RDM "are deeply rooted in the practices and culture of the research process as well as the researchers themselves" [25](p. e21101). Those barriers could either be of technical and non-technical nature [12]. Data, which remains untraceable to researchers, so called "Dark Data", might get lost over time and offers special technical and non-technical barriers regarding its management [26]. Even if technical and non-technical barriers by researchers regarding research data management are overcome, it is questionable that, for example, the supply of openly shared data due to research data management is used by other researchers [27].

Yet, it is questionable which data should be shared, since once they are loosened from the original context they might get misinterpreted or get misused due to missing standardized policies [2]. Anyhow, there are arguments for providing insights into own data and by this for using RDM stated by [2], namely the possible reproduction and verification of research, the availability of research which was funded by and its obligations towards the public, the generation of progress regarding the research with the help of other individuals, which might ask questions concerning the existing data, and finally the improvement of the current state of research in general. [28] suggested an accurate definition of sharable data "as the combined experimental data and descriptive metadata need[ed] to evaluate and/or extend the results of a study" [28] (p. 3). An analysis of researchers' experiences about

the usage of shared data collected by others reveals that standards alone do not lead to re-using data, but rather knowledge of the context, which is often left behind when it comes to research data management [29].

The importance and advantages of RDM are already widely known, and universities start building own checklists and platforms to provide among others data sharing, mainly to collect their data and to allocate todays' data for their future researchers [13], [30], [31]. It happens that existing platforms do not offer the special needs of different research institutions and fields of study, just as sufficient storage and security safeguards [9], and due to that universities, for example, develop their own platforms with individual research data management guidelines [32]. The university of Rochester (New York), as an example, spent US$200,000 in designing and implementing a digital archive to manage the data of their researchers [24]. But such investments often do not work out as they were planned, since the digital archive of the University of Rochester remained empty [24]. Researchers reported that they could not find their data, nor could offer the time to work on the management of their data and additionally have certain concerns regarding the release of their research data, like misuse and misunderstanding [24]. Likewise, libraries currently develop research data management guidelines, since they hold the view that the management of research data might become an important part in their future work [33]. Libraries' priorities make up the "provision of research data management advisory and training services" [33] (p.1). Besides, funds for open access researchers are taken into consideration by academic libraries to support data sharing, open access and researchers in general [34]. As institutional information experts, librarians should be part of data sharing efforts, since enormous administrative and operational capabilities are urgently needed [35].

Researchers often struggle with the quantity, diversity and complexity of the generated data, whether it is their own or produced one by colleagues, making future research data management inevitable [36]. [37] recommend an international framework to optimize research data management and to support the connection of the scientific community, due to an increasing amount of data in different research disciplines. It is recommended to cohere of all the professionals in research data management, such as researchers, libraries, publishers and research funders, to assure the preservation of research and data, and to secure the usability of platforms providing research data management [38]. While [39] found that most of the researchers avoid data archiving, [40] revealed that

there is an increasing awareness in regards to data storage, since the majority of researchers nowadays store their own data. Yet, the guidelines regarding research data management by nonprofit institutions influence significantly the "how, when, and whether research data are shared" [2](p. 1).

3. Method

To answer the research question asking for the requirements which are set up by (governmental) funded foundations in order to conduct research data management, a qualitative research setting was conducted. The qualitative design was chosen to gain detailed unsupported insights about the different guidelines and the specific understanding of RDM.

For the discovery of already existing international standards towards research data management guidelines and to alert on possibly missing guidelines, the biggest and most influential institutions regarding research data management within the scientific community were selected. Thus, the existing guidelines of the eight largest and widest known governmental funded institutions were collected. Additionally, Wellcome Trust, as a widely known, but politically independent institution, was consulted within the list as well. "Large" as a requirement connotes the number of the institutions' salaried employees, as well as its international influence. The considered institutions are the following:

The Australian Research Council (ARC)[41], the Bundesministerium für Bildung und Forschung (BMBF) [42], the Deutsche Forschungsgemeinschaft (DFG) [43], the European Commission H2020 Program (Horizon 2020) [44], the National Institutes of Health (NIH) [45], the National Research Council Canada (NRC) [46], the National Science Foundation (NSF) [47], the Organisation for Economic Co-operation and Development (OECD) [48] and the Wellcome Trust [49] as the only politically and financially independent institution within this list is settled in London, United Kingdom (around 2000 employees).

This work undertakes an inductive category development as research method. At first, the research question was generated to determine the definite goal of the work, as suggested by [15]. Hereafter, definitions regarding the constructed categories, just as levels of category abstractions are given. Via a step-by-step process, inductive categories are formulated by use of the constructed categories and abstractions, leading to a synopsis of categories and the creation of new ones. Following this, all categories will be revised after up to 50% of the collected material to secure a

formative check of reliability [15]. Again, all texts and guidelines are worked through, ensuring a summative check concerning the reliability [15]. The basis of the constructed categories constitutes the guidelines of the different institutions named before. Those guidelines were collected from the institutions' websites at the end of February 2017. Possible main categories were accumulated and the applicableness regarding the institutions was examined. Subsequently, sub-categories, as suggested by [15], were generated, giving a more specific representation of the institutions' concerns and guidelines regarding the management of research data.

Moreover, the established categories were divided in technical and non-technical requirements. Technical requirements involve topics like the allocation of RDM platforms, safety warranty, research data storage, memory size and suchlike. Whereas non-technical requirements represent context-oriented topics regarding the research data, such as ethical provisions, and its management. This leads to an overview of the heeded technical and non-technical requirements by the different institutions and their varieties, with special regard to the non-technical ones, since they might influence current research data management significantly [11].

4. Results

By use of the inductive category development as suggested by [15], seven main categories and fourteen sub categories were evolved. Due to the used step-by-step process, the constructed categories and their abstractions lead to the summarization of already existing categories and the development of new ones. Through the consideration of different levels of abstraction and the repeated peruse of the text, the following categorized guidelines, as a result of the qualitative analysis by [15] were formulated.

Infrastructure. Belonging to the main category "Infrastructure", the four sub categories "documentations and focus settings", "infrastructure policies", "Stabilization of RDM infrastructures", and "Tools for RDM" were determined.

The sub category "documentations, and focus settings" includes the institutions' guidelines which address the need and importance of data definitions and data documentations to simplify the reuse of data within data storage systems. As stated by the NIH: "Regardless of the mechanism used to share data, each dataset will require documentation. [..] Proper documentation is needed to ensure that others can use the dataset and to prevent misuse, misinterpretation, and confusion." [45](p. 1).

Within the sub category "infrastructure policies", all guidelines concerning the urgency of standardized policies within research data management, preferably on an international level, are collected. These standardized policies are needed to facilitate data sharing between researchers worldwide and to ease scientific communication. But solely the BMBF, NRC, and OECD provide those guidelines, while the left institutions do not mention the topic of coherent infrastructure policies.

"Stabilization of RDM infrastructures" involves guidelines that focus on research data management systems and infrastructures itself, especially on their consistency and availability. Solid infrastructures are needed to inhibit possible uncertainty of researchers and to ease the usage of RDM systems. The DFG already recognized the need for international standardized infrastructures: "[..] requirements must be defined through the cooperation of researchers and information specialists. Infrastructures are to be developed according to these requirements and, if possible, interoperably integrated in international and interdisciplinary networks from the start." [43](p.1).

The last sub category, "Tools for RDM", includes recommendations and examples of special tools for a sufficient RDM. The DFG, NIH, and Wellcome Trust offer those tools, trying to make research data management easier and more present to researchers. The Digital Curation Centre, as an example, offers different tools for research data management, such as DMPonline, which was developed by Wellcome Trust.

Security. "Security", as a main category, includes solely the sub category "Data security". Since the security of research data has been of great importance to researchers and institutions in the past, according to for example [31] and [50], and in addition is one of the mostly named advantages of research data management, it builds up its own category. "Data security" involves guidelines that face up to safety concerns and problems. Behaviors for a responsible dealing with foreign data on behalf of the researchers and the platforms developer and staff, but also the advantages of research data management systems concerning the security of data are mediated. Except the NIH and NSF, all consulted institutions provide guidelines regarding the safety of data.

Sharing. The third main category composes "Sharing" with its two sub categories "Open access of data" and "Timeliness of data sharing". Since both sub categories consider institutions requested data sharing behavior on behalf of researchers, they are collected within one main category.

"Open access of data" implies that research data is accessible for everyone, regardless if they are researchers or not. Since research is often funded by the public, it is postulated that every individual should get free access to research data [51]. Additionally, it is important for researchers to have easy access to the work and data of others, since due to that progresses in science could get achieved. All consulted institutions provide guidelines for an open access of research data, showing that standardized guidelines might already exist. The OECD states conditions which open access should fulfill: "Openness means access on equal terms for the international research community at the lowest possible cost, preferably at no more than the marginal cost of dissemination. Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based." [48](p. 15).

"Timeliness of data sharing" includes guidelines that recommend periods of time in which researchers should publish the data at the latest that relate to their already published studies. Based on the consulted institutions, two of them, namely the ARC and NIH present guidelines which suggest periods of time concerning the publishing and sharing of data. Since it is difficult to decide on a period of time to publish data because of the varying types of data, guidelines concerning the timeliness of data sharing are often unspecific, as shown as follows by the NIH: "In general, NIH considers the timely release and sharing of data to be no later than the acceptance for publication of the main findings from the final dataset. However, the actual time will be influenced by the nature of the data collected." [45](p. 1).

Storage. The main category "Storage" includes "Long-term storage of research data". Just as the security of research data, it is often named as one of the main advantages of research data management (see for example [50].

"Long-term storage of research data" means the deployment of long-term storage within research data management platforms. These long-term storage guidelines are important to legally ensure those platforms and to safely keep the data, inhibiting the loss of data. The last aspect is especially important to educational institutions, since they fear the loss of data which might be used by their future researchers according to [13]. All named institutions, except the NIH, offer guidelines for the long-term storage of research data. The OECD, as an example, does not name an exact period of time in which data should be stored. As opposed to this, the ARC names more specific time periods regarding the storage of research data from different scientific disciplines. The general recommended time period is five years from the date

of publication. For clinical trials, "retaining research data for 15 years or more may be necessary" [41](p. 2.1). The DFG specifically states that primary research data should be available for at least ten years on suitable memories [52].

Ethics. "Ethics" involves institutions guidelines that state procedures in dealing with humans and animals within science and their gained data. Since the dealing with such data causes difficulties regarding research data management [6], especially when it comes to data sharing and open access, standardized guidelines would be of great importance for researchers, but also for human participants. Within the considered institutions, nearly all of them, released guidelines that handle the dealing with such ethical issues, like for example the sharing of patients' data. By way of comparison, the statement concerning the handling of human data given by the OECD will be offered: "Privacy and confidentiality: data on human subjects and other personal data are subject to restricted access under national laws and policies to protect confidentiality and privacy. However, anonymization procedures that ensure a satisfactory level of confidentiality should be considered by custodians of such data to preserve as much data utility as possible for researchers." [48](p. 16).

Management. The category "Management" contains the two sub categories "Dealing with research misconducts" and "Education of research data management staff". Both sub categories address the management of research data management systems and platforms, which is why they are assigned to the same category.

Within "Dealing with research misconducts", guidelines concerning the dealing with misconducts made by researchers are summarized. Due to the fact that the possible finding of research misconducts inhibits the data sharing of researchers [6], it is important to counteract this problem. Solely the ARC and Wellcome Trust address this problem by offering guidelines to managers of research data management platforms.

Likewise, "Education of research data management staff" is much needed and of great importance concerning research data management. A well-educated staff is needed to foster and improve research data management systems and platforms, but also to support and help researchers who make use of research data management. Five out of nine examined institutions offer guidelines concerning this topic. The ARC, BMBF, DFG, OECD, and Wellcome Trust published guidelines concerning the education of research data management staff.

Researchers. The last main category "Researchers" consists of the three sub categories "Identification of ownership", "Inducements of recognition regarding RDM usage", and lastly "Protection of intellectual property". All these sub categories address researchers directly, which is why they were collected within one category.

"Identification of ownership" means a continuous recognition of the works' author and the data constitutor. Offered guidelines concerning that topic handle the guarantee of a stable author identification. Since researchers often fear that their work and data might get misused and that they might lose credits for their work, they will not share their data [2], institutions need to address this problem and provide assurances to enable successful research data management.

As studies already revealed, "Inducements of recognition regarding RDM usage" have a significantly positive influence on researchers' attitude towards research data management, especially on the sharing of data [53]. By means of that background knowledge, inducements of recognition because of an openly sharing of data for scientific progress must be provided within research data management. Without offering incentives to researchers for sharing their data and managing their research data, progress of research data management might be unlikely.

The last sub category "Protection of intellectual property", includes guidelines that are aimed at researchers who work for private institutions and who are, due to organizations' policies, inhibited to publish their data and research. The ARC, DFG, NRC, NSF, OECD and Wellcome Trust provide guidelines and promise support for researchers who are legally not allowed to publish and share their research data. The OECD undertakes distinctions whether the intellectual property belongs to researchers within educational institutions or companies, albeit the OECD says that a "private sector involvement in the data collection should not, in itself, be used as a reason to restrict access to the data" [48](p. 17). The OECD reveals by this implicitly, that they do not support every researcher in protecting their intellectual property.

Figure 1 presents and overview including the different main and sub-categories concerning the guidelines and whether they are offered or not by the different institutions.
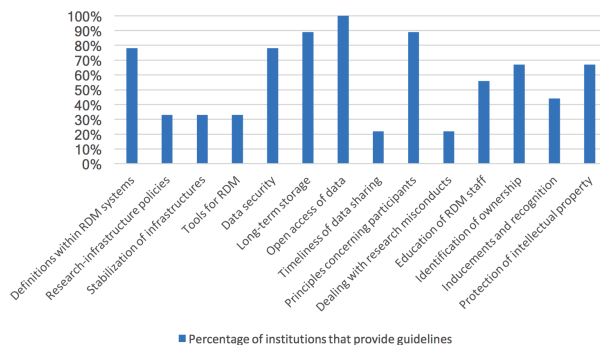

Figure 1. *Percentage of institutions that provide the categorized guidelines.*

Definitions, documentations and focus settings within research data management guidelines are provided by altogether 78% of the examined institutions. Guidelines concerning international research-infrastructure policies are provided by 33% of the institutions, while another 33% offer guidelines regarding the stabilization of infrastructures and tools for research data management. With 78%, guidelines relating to data security are released by institutions and further 89% address the long-term storage of data. Additionally, 100% of institutions provide guidelines towards an open access of research data. 22% of the institutions own guidelines for the timeliness of data sharing and another 22% own policies in regard to the dealing with research misconducts. 56% offer guidelines for education of research data management staff. 89% of the nine institutions present guidelines concerning principles towards the handling with human participants, animals, and their data. Altogether 67% address the identification of ownership as well as protection of intellectual property. Further 44% out of nine institutions offer guidelines that address inducements and recognition for researchers regarding the usage of research data management.

Based on the developed main and sub-categories, a differentiation of technical and non-technical requirements regarding the management of research data was made. The derivation of those segregated requirements is of importance, since institutions often solely address the technical requirements as well as their barriers, while various studies revealed that non-technical requirements include barriers that are equally important, since they might constitute the main reasons for researchers to not manage their research data, especially when it comes to data sharing [6], [54], [55]. Since both, technical and non-technical requirements can inhibit research data management, they are henceforth exclusively termed as technical and non-technical barriers.

The following table provides an overview of the differentiation of the main categories regarding their sub categories named before in technical and non-technical barriers:

Technical barriers	Non-technical barriers
Infrastructure	Ethics
Security	Management
Sharing	Researchers
Storage	

Table 1. *Differentiation of technical and non-technical barriers.*

5. Discussion and Conclusion

On closer consideration, it occurs that open access of data, and thus data sharing, was apprehended by all nine institutions within their guidelines, but most solely on a technical perspective, since the possibilities of data sharing for future research and the dealing with research data management infrastructures and tools are rather approached than researchers point of view and their existing fears.

Instead of giving opportunities for action regarding a secure dealing with data sharing and the removal of possible abuses, solely the importance of data sharing for a scientific progress is thematized, without any regard on researchers' perspective. This has been the case with all nine examined institutions, independently if they were funded by governments or a non-governmental institution. As already referred to in the current state of research within this work, researchers want to keep their own opportunity of sharing their data instead of giving it to research data management institutions [56].

The institutions guidelines indeed address researchers' fear of security risks on RDM platforms, but confine themselves upon technical requirements, and ignore researchers' security risks regarding the loss or possible misuse of their data. Still, it should be mentioned that respectively six institutions offer guidelines which address data security and its secure long-term storage. As a further reason why researchers avoid data sharing, a lack of career rewards was mentioned [6]. Within this work, it turned out that solely three out of nine examined institutions provide guidelines concerning the recognition and rewards to researchers for using research data management. This aspect, as a non-technical barrier, attaches no value in most of the institutions guidelines, leading to inhibition and indifference by researchers towards research data management. [54] already indicated researchers' fears concerning data sharing as non-technical barriers. Those non-technical barriers should

be negotiated in the future, since they have the same importance as the provision of technical solutions. Nevertheless, six out of nine institutions approach researchers by providing guidelines in regard to the importance of ownership and the protection of intellectual property, which has main importance to researchers [6]. However, as seen by the example of the OECD, not every researcher will be supported to protect their intellectual property. It rather depends on the type of organization they work for. But overall, those institutions miss their aim that all data will be reached via open access in the future to advance scientific progress due to the fact that they do not react to researchers' inhibitions.

In addition, by consideration of the results it gets obvious that the provision of consistent infrastructures and guidelines regarding research data management on an international level is barely detectable. Admittedly, all institutions speak out in favor of open access of research data, yet only three institutions provide guidelines and emphasize international research data management policies within infrastructures. Another three institutions argue in support of the need for RDM infrastructures stabilization within their guidelines. Those international guidelines and policies, but also stable RDM infrastructures are necessary to ease the sharing of data and its preparation on behalf of researchers. Inasmuch as institutions want to expedite science by research data management they should agree upon international RDM guidelines. It shall be said that all of the governmental funded institutions named in this work should possess consistent guidelines, since all those countries are member of the OECD, which created guidelines regarding the dealing with research data. Yet, as shown by the table within the results, the institutions seem to be discordant concerning those guidelines. For instance, some institutions dismiss guidelines, although these are provided within the OECD's principles and guidelines. The ARC, as an example, plead the OECD guidelines, while simultaneously offering own policies regarding research data management. While the OECD speaks out by way of example regarding the long-term storage of data in a very general way, and without naming specific periods of times, the ARC specializes the long-term storage of data within their own guidelines by providing principles towards periods of time for different kinds of data. Such variations mostly cease within specifications inside institutions own guidelines, but might extent sooner or later to extreme divergences. To facilitate research data management, educated staff for research data management infrastructures are particularly needed, for example to recognize and handle research misconducts. Thereby,

the quality of shared data and scientific work in general would be improved. Nevertheless, only four institutions realize the importance of educated staff for research data management. Still, more than the half of the examined institutions within this work realized the importance of data definitions, documentations and focus settings within RDM infrastructures to simplify data sharing for researchers and to increase data quality just as scientific progress.

A huge topic for researchers, as already mentioned before, are ethical concerns, for example the dealing with data which was collected from human individuals, but also the dealing with animals as participants [11]. Since ethical principles inhibit researchers to openly share their data, and since it might lead to a possible decrease of human participants in the future [57], [58], specific guidelines should be provided by institutions to decrease researchers' fears. Solely six institutions address this topic, rather generalized than providing helpful guidelines which support researchers with their dealing regarding ethical principles and possible solutions. Wellcome Trust, for example, states that "the Trust believes that the basic DNA sequence of humans and other organisms should be placed in the public domain as soon as is practical, without any fees, patents, licenses or limitations on use, [...]" [49] (p.1), but does not give attention to researchers' main ethical concerns like legitimacy. A further problem constitutes the different types of data. Medical and health related data are sorer than those gained via surveys or social media. Differences concerning the dealing with different kinds of data are solely found when it comes to data storage. The ARC specifies different periods of time for different types of data. Nevertheless, guidelines regarding the ethical aspects towards different types of data, such as the dealing with human DNA, are urgent needed and still missing. Altogether, all founded guidelines concerning non-technical barriers deal with those aspects named within the literature, but solely at a low and insufficient level.

An additional problem constitutes the general provision of research data management guidelines. Several institutions refer to a number of different institutions websites that provide research data management policies. The ARC additionally refers to the OECD, and the NRC refers in addition to the Research Data Canada institution. Instead of providing a clear depiction of research data management guidelines, as seen by the OECD, researchers must search numerous links and websites to gain guidelines of different sub topics of RDM. Most institutions are solely eager to openly share research data without answering researchers' interests, and thereby forget that the future of research data management lies within researchers' hands.

Relating to the research question, it can be said that technical requirements concerning the definition, documentation, and focus setting of data within research data management infrastructures, international research policies, stabilization of infrastructures, provision of RDM tools, open access of data and security, timeliness of sharing, and long-term storage of data are set up. The non-technical requirements constitute principles regarding the handling of human participants, animals, and their data (in this work named "Ethics"), the dealing with research misconducts and the education of research data management staff, but also the protection of intellectual property, the identification of ownership, and recognition and inducements for researchers for using research data management.

Those technical and non-technical requirements turned out to rather be barriers, since the consulted literature suggested that the as technical and non-technical discussed generalized guidelines include inhibitions and difficulties for researchers regarding research data management. Furthermore, it transpired that institutions guidelines mostly answer those difficulties rather general, without any regard to researchers concerns and opinions. Admittedly, the examined institutions argue that research data management is inalienable and much needed for scientific progress, yet themselves inhibit the usage of research data management by researchers due to a lack of support and comprehension. More specific guidelines are needed to still researchers' fears and to give them an understanding of the advantages of research data management. In terms of technical requirements, we recommend future IS investigations to focus on motivational IS designs (e.g. [59], [60]) to increase researchers' acceptance towards RDM.

It should be mentioned that the examination within this work had limitations. The biggest governmental funded institutions and the biggest independent institution concerning the dealing with research data management were consulted indeed, yet those are overall solely nine out of thousands existing. Beyond, some worked through guidelines included overlaps, since institutions referred to guidelines from other institutions (for example the ARC invokes on the OECD), but simultaneously offer own guidelines which dissent from those invoked ones. It is necessary to consider that the classification and collection of guidelines could be more specific or more general, depending on the used process of qualitative analyses.

6. References

[1] G. Bell, T. Hey, and A. Szalay, "COMPUTER SCIENCE: Beyond the Data Deluge," *Science (80-.).*, vol. 323, no. 5919, pp. 1297–1298, Mar. 2009.

[2] C. L. Borgman, "The Conundrum of Sharing Research Data," *J. Am. Soc. Inf. Sci. Technol.*, vol. 63, no. 6, pp. 1059–1078, 2012.

[3] A. Whyte and J. Tedds, "Making the case for research data management," *Digital Curation Centre*, 2011.

[4] R. Devarakonda, G. Palanisamy, J. M. Green, and B. E. Wilson, "Data sharing and retrieval using OAI-PMH," *Earth Sci. Informatics*, vol. 4, no. 1, pp. 1–5, 2011.

[5] R. Higman and S. Pinfield, "Research data management and openness: The role of data sharing in developing institutional policies and practices," *Program*, vol. 49, no. 4, pp. 364–381, Sep. 2015.

[6] L. Corti, V. Van den Eynden, L. Bishop, and M. Woollard, "Managing and Sharing Research Data," *Sage Publ. Ltd*, no. November, 2014.

[7] G. Bell, "Foreword," in *The Fourth Paradigm*, T. Hey, S. Tansley, K. Tolle, and H. T. Tolle, Eds. Washington: Microsoft Research, 2009, pp. xi–xv.

[8] J. Becker, T. Heide, R. Knackstedt, and M. Steinhorst, "Supporting Knowledge Management and Collaboration in Research Communities Using Automatically Created Research Portals," *Int. J. Web Portals*, vol. 5, no. 2, pp. 1–16, 2013.

[9] R. C. Amorim, J. A. Castro, J. Rocha da Silva, and C. Ribeiro, "A comparison of research data management platforms: architecture, flexible metadata and interoperability," *Univers. Access Inf. Soc.*, pp. 1–12, Jun. 2016.

[10] C. J. Savage and A. J. Vickers, "Empirical Study of Data Sharing by Authors Publishing in PLoS Journals," *PLoS One*, vol. 4, no. 9, pp. 70–78, 2009.

[11] S. Childs, J. McLeod, E. Lomas, and G. Cook, "Opening research data: issues and opportunities," *Rec. Manag. J.*, vol. 24, no. 2, pp. 142–162, 2014.

[12] K. Wilms, C. Meske, S. Stieglitz, D. Rudolph, and R. Vogl, "How to Improve Research Data Management," in *Human Interface and the Management of Information: Applications and Services. HIMI 2016. Lecture Notes in Computer Science*, 9735th ed., Y. S., Ed. Springer, Cham, 2016, pp. 434–442.

[13] J. Schopfel, S. Chaudiron, B. Jacquemin, H. Prost, M. Severo, and F. Thiault, "Open access to research data in electronic theses and dissertations: an overview," *Libr. Hi Tech*, vol. 32, no. 4, pp. 612–627, Nov. 2014.

[14] K. Wilms *et al.*, "Digital Transformation in Higher Education – New Cohorts, New Requirements?," *AMCIS 2017 Proc.*, Aug. 2017.

[15] P. Mayring and T. Fenzl, "Qualitative Inhaltsanalyse," in *Handbuch Methoden der empirischen Sozialforschung*, N. Baur and J. Blasius, Eds. Wiesbaden, 2014, pp. 543–556.

[16] R. L. Ackoff, "From data to wisdom," *J. Appl. Syst. Anal.*, vol. 16, no. 1, pp. 3–9, 1989.

[17] M. E. Jennex and S. E. Bartczak, "A revised knowledge pyramid," *Int. J. Knowl. Manag.*, vol. 9, no. September, pp. 19–30, 2013.

[18] H. A. Piwowar, R. S. Day, and D. B. Fridsma, "Sharing detailed research data is associated with increased citation rate," *PLoS One*, vol. 2, no. 3, p. e308, Mar. 2007.

[19] C. L. Palmer *et al.*, "Data curation for the long tail of science : The case of environmental sciences," in *Third International Digital Curation Conference*, 2007, pp. 1–6.

[20] M. Witt, "Institutional Repositories and Research Data Curation in a Distributed Environment," *Libr. Trends*, vol. 57, no. 2, pp. 191–201, 2008.

[21] Y. Kim and J. M. Stanton, "Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis," *J. Assoc. Inf. Sci. Technol- ogy*, vol. 67, no. 4, pp. 776–799, 2012.

[22] M. Joshi and S. S. Krag, "Issues in Data Management," *Sci. Eng. Ethics*, vol. 16, no. 4, pp. 743–748, 2010.

[23] S. H. Koslow, "Should the neuroscience community make a paradigm shift to sharing primary data?," *Nat. Neurosci.*, vol. 3, no. 9, pp. 863–866, 2000.

[24] B. Nelson, "Data Sharing: Empty archives," *Nature*, vol. 461, no. 7261, pp. 160–163, 2009.

[25] C. Tenopir *et al.*, "Data Sharing by Scientists: Practices and Perceptions," *PLoS One*, vol. 6, no. 6, 2011.

[26] P. Bryan Heidorn, "Shedding Light on the Dark Data in the Long Tail of Science," *Libr. Trends*, vol. 57, no. 2, pp. 280–299, 2008.

[27] I. M. Faniel and T. E. Jacobsen, "Reusing Scientific Data: How Earthquake Engineering Re-searchers Assess the Reusability of Colleagues' Data," *Comput. Support. Coop. Work*, vol. 19, no. 3, pp. 355–375, 2010.

[28] D. Gardner *et al.*, "Towards Effective and Rewarding Data Sharing," *Neuroinformatics*, vol. 1, no. 3, pp. 289–295, 2003.

[29] A. S. Zimmermann, "New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data," *Sci. Technol. Hum. Values*, vol. 33, no. 5, pp. 631–652, 2008.

[30] G. Pryor, S. Jones, and A. Whyte, "Delivering Research Data Management Services: Fundamentals of Good Practice.," in *Delivering Research Data Management Services: Fundamentals of Good Practice.*, London: Facet Publishing, 2013.

[31] G. Stöckle, "A Checklist for Planning Research Data Management," *Astrostatistics Data Min.*, pp. 247–251, 2012.

[32] S. Liang, V. Holmes, G. Antoniou, and J. Higgins, "iCurate: A Research Data Management System.

Multi-disciplinary Trends in Artificial Intelligence," *Lect. Notes Comput. Sci.*, pp. 39–47, 2015.

[33] A. M. Cox and S. Pinfield, "Research data management and libraries: Current activities and future priorities," *J. Librariansh. Inf. Sci.*, vol. 46, no. 4, pp. 299–316, 2014.

[34] R. Nariani and L. Fernandez, "Open Access Publishing: What Authors Want.," *Coll. Res. Libr.*, vol. 73, no. 2, pp. 182–195, 2012.

[35] J. Giffels, "Sharing Data is a Shared Responsibility," *Sci. Eng. Ethics*, vol. 16, no. 4, pp. 801–803, 2010.

[36] B. Hanson, A. Sugden, and B. Alberts, "Making Data Maximally Available," *Science (80-.).*, vol. 331, no. 6018, p. 649, 2011.

[37] Arzberger, P. *et al.*, "Promoting Access to Public Research Data for Scientific, Economic, and Social Development," *Data Sci. J.*, vol. 3, pp. 135–152, 2004.

[38] E. Smit, J. Van Der Hoeven, and D. Giaretta, "Avoiding a Digital Dark Age for data: why publishers should care about digital preservation," *Learn. Publ.*, vol. 24, no. 1, pp. 35–49, 2011.

[39] P. A. Soranno, K. S. Cheruvelil, K. C. Elliott, and G. M. Montgomery, "It's Good to Share: Why Environmental Scientists' Ethics Are Out of Date," *Bioscience*, vol. 65, no. 1, pp. 69–73, 2015.

[40] D. Rudolph, A. Thoring, and R. Vogl, "Research Data Management: Wishful Thinking or Reality?," *PIK - Prax. der Informationsverarbeitung und Kommun.*, vol. 38, no. 3–4, pp. 113–120, Jan. 2015.

[41] Australian Research Council (ARC), "Research Data Management – Australian Code for the Responsible Conduct of Research," 2017. .

[42] Bundesministerium für Bildung und Forschung (BMBF), "RFII Empfehlungen 2016: Leistung aus Vielfalt," 2017. [Online]. Available: http://www.rfii.de/?wpdmdl=1998.

[43] DFG Deutsche Forschungsgemeinschaft, "Memorandum Safeguarding Good Scientific Practice," 2013.

[44] European Commission, "H2020 Programme - Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020," 2017.

[45] National Institutes of Health (NIH), "NIH Data Sharing Policy and Implementation Guidance," 2017. [Online]. Available: https://grants.nih.gov/grants/policy/data_sharing/data_shar- ing_guidance.htm.

[46] National Research Council Canada (NRC), "Data Analytics Centre," 2017.

[47] NSF National Science Foundation, "Data Management & Sharing FAQs | NSF - National Science Foundation." [Online]. Available: https://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp. [Accessed: 15-Jun-2017].

[48] Organisation for Economic Co-operation and Development (OECD), "Data Management Procedures," 2016. [Online]. Available: https://www.oecd.org/pisa/pisaprod- ucts/PISA 2012 Technical Report_Chapter 10.pdf.

[49] Wellcome Trust, "Policy on data management and sharing." [Online]. Available: https://wellcome.ac.uk/funding/managing-grant/policy-data-management-and-sharing. [Accessed: 15-Jun-2017].

[50] T. Viangteeravat, V. R. Nagisetty, M. N. Anyanwu, E. Kuscu, I. M. Brooks, and C. S. McDonald, "Protected Research Information Management Environment (PRIME) provides a secure open source data management option for clinical and scientific research," *BMC Bioinformatics*, vol. 12, no. 7, 2011.

[51] F. Berman and V. Cerf, "Who Will Pay for Public Access to Research Data?," *Science (80-.).*, vol. 341, no. 6146, pp. 616–617, 2013.

[52] Deutsche Forschungsgemeinschaft, "Deutsche Forschergemeinschaft," 2017. [Online]. Available: www.dfg.de.

[53] D. S. Sayogo and T. A. Pardo, "Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data," *Gov. Inf. Q.*, vol. 30, pp. S19–S31, 2013.

[54] M. Feijen, W. Horstmann, P. Manghi, M. Robinson, and R. Russell, "DRIVER: Building the Network for Accessing Digital Repositories across Europe," *Ariadne*, no. 53, p. np ST-DRIVER: building the network for accessin, 2007.

[55] N. Pearce and A. H. Smith, "Data sharing: not as simple as it seems," *Environ. Heal. 1*, vol. 10, no. 10, pp. 1–7, 2011.

[56] E. G. Campbell, B. R. Clarridge, L. Birenbaum, S. Hilgartner, and D. Blumenthal, "Evidence From a National Survey," *J. Am. Med. Assoc.*, vol. 287, no. 4, pp. 473–480, 2002.

[57] A. Carusi and M. Jirotka, "From data archive to ethical labyrinth," *Qual. Res.*, vol. 9, no. 4, pp. 285–298, 2009.

[58] H. A. Piwowar, "Who shares? Who doesn't? Factors associated with openly archiving raw research data," *PLoS One*, vol. 6, no. 7, 2011.

[59] C. Meske, T. Brockmann, and K. Wilms, "Social Collaboration and Gamification," in *Gamification, Progress in IS*, Springer, Cham, 2017, pp. 93–109.

[60] C. Meske, T. Brockmann, K. Wilms, and S. Stieglitz, "Gamify Employee Collaboration - A Critical Review of Gamification Elements in Social Software," *Australas. Conf. Inf. Syst.*, pp. 1–15, Jun. 2015.