Armel Lefebvre*, Baharak Bakhtiari, and Marco Spruit*

Exploring research data management planning challenges in practice

Abstract: Research data management planning (RDMP) is the process through which researchers first get acquainted with research data management (RDM) matters. In recent years, public funding agencies have implemented governmental policies for removing barriers to access to scientific information. Researchers applying for funding at public funding agencies need to define a strategy for guaranteeing that the acquired funds also yield high-quality and reusable research data. To achieve that, funding bodies ask researchers to elaborate on data management needs in documents called data management plans (DMP). In this study, we explore several organizational and technological challenges occurring during the planning phase of research data management, more precisely during the grant submission process. By doing so, we deepen our understanding of a crucial process within research data management and broaden our understanding of the current stakeholders, practices, and challenges in RDMP.

Keywords: Research data management, Data management plan, Open science, Data reuse, FAIR principles

ACM CCS: Social and professional topics → Computing/technology policy → Government technology policy → Governmental regulations

1 Introduction

Public funding agencies and research institutions are facing novel challenges related to the management of research outputs produced in Academia. In recent years, governing bodies across the world have started to promote new open science policies and practices for managing scientific information. Unlike open access policies

*Corresponding authors: Armel Lefebvre, Marco Spruit, Utrecht University, Department of Information and Computing Sciences, 3584 CC Utrecht, Netherlands, e-mails: a.e.j.lefebvre@uu.nl, m.r.spruit@uu.nl, ORCID: https://orcid.org/0000-0002-7428-1728 (A. Lefebvre)
Baharak Bakhtiari, Utrecht University, Department of Information and Computing Sciences, 3584 CC Utrecht, Netherlands, e-mail: baharak.bakhtiari@gmail.com

that exclusively promote the public availability of scientific articles [1], open science (OS) policies expand open access to a more extensive set of research outputs into consideration. Accordingly, OS includes scientific publications and their corresponding artifacts, such as software, data, and sample material, into the scope of scientific information [2, 3]. As a result, research data management (RDM) is introduced by funding agencies as a critical capability of research institutions that benefit from training programs, dedicated IT services, and new roles in research organizations [4]. Nevertheless, previous research on research data management planning (RDMP) practices has reported challenges related to a lack of knowledge about the usefulness and best practices of RDMP. So far, other studies have shown that (1) funder policies for RDMP are quite general [5], (2) researchers are often reluctant to disseminate curated data [6], and (3) there is a lack of knowledge and detailed guidelines to support RDMP in research institutions [4]. Moreover, in the United States (US), [7] showed that requirements from US funding agencies are inconsistent and emphasize post-publication data management rather than foster an up-stream data strategy, which could guarantee more robust data management from the start of a research project. Likewise, Science Europe, a European association of national public funding agencies, acknowledged the complexity of current policies and recently proposed standardized RDMP guidelines [8]. In this work, we investigate research data management planning (RDMP) as a function of research data management (RDM). RDMP is an essential part of the grant application process. In Europe, for instance, the European Commission has (partially) incorporated data management and planning in its grant application procedures for the Horizon 2020 (H2020) funding program [3]. In Europe, for instance, the European Commission has (partially) incorporated data management and planning in its grant application procedures for the Horizon 2020 (H2020) funding program [3]. As part of the application procedure of H2020, applicants submit data management plans (DMP). A DMP is an additional document in which grant applicants outline how data is acquired for their research project, which technology and standards they intend to use (e. g., storage, back-up, software), how data will be preserved and, possibly, shared and, the costs induced by additional resources and services needed to manage data [9].

Therefore, we seek to investigate current RDMP practices in academia from two perspectives: a funder perspective and a research data service perspective. By doing so, we aim at shedding light upon existing practices and challenges in RDMP. Besides, we suggest potential solutions where information technology can play a crucial role in improving the review of RDMP deliverables. The following research question drives this study: **What are the current challenges and practices in research data management planning?** To answer this question, we follow a case study approach and collect experiences from representatives of public funding agencies, grant reviewers, and data stewards. Second, as part of the case study, we analyzed 98 data management sections in (draft) research proposals of projects submitted to NWO, the Dutch national science foundation. The goal of the analysis was to investigate whether current data management paragraphs reflect the ambition of producing reusable research data.

2 Background and related work

From the start of the digital revolution, data management has played a crucial role in organizations. At that time, managers became aware of the potential business value of data stored in companies. However, it appeared that data was also not fully integrated across information systems (IS) [10]. Strategic data planning (SDP) was established as a response to the absence of integration of information systems in firms, which is one of the earlier attempts to plan information architectures integrating data sources in organizations [10, 11].

The ambition to align data systems with corporate needs is still actively pursued in new data management practices. Accordingly, business and industry environments view data planning as a strategic process. Thus, the business value of data depends upon the capacity to integrate corporate information systems. Research data management planning (RDMP) share similar ambitions with data planning as RDMP aligns the production and use of research data with an open science strategy. However, data management in academia has not been as thoroughly investigated as data management in businesses [4, 12].

On the one hand, research data management (RDM) refers to the management of research data during the lifecycle of a research project, from data creation to dissemination. Funders and research institutions' policies state that RDM is the responsibility of researchers. Researchers are supported by research data management services (RDS). RDS consists of library services specializing in data management matters and metadata curation, IT services for hardware and software and, data stewards as an emerging role in academia, which assist researchers and research groups in managing data.

On the other hand, research data management planning (RDMP) is defined here as the process of planning costs, (storage) technology, formats, documentation, legal matters and openness of data to effectively manage research data during and after (publicly funded) research projects [3, 5]. Often, the main deliverable of research data planning is a data management plan (DMP) [9]. Additionally, data stewards are supporting scientists in establishing a data management plan and curate data [13]. The reason why data stewards hold this supporting role is that data management planning is a new practice for scientists who might ignore fundamental data management concepts [4]. For instance, funding agencies in the Netherlands rely upon a set of guidelines focused on extending the life cycle of research data and producing reusable data. These guidelines are based on FAIR principles, which are a set of guidelines used to create Findable, Accessible, Interoperable, and Reusable research data [14].

DMP requirements may vary for each (public) funding agency, as shown in Table 1. For instance, depending on the funder, grant applicants need to fill in a data management section in the proposal, and submit an additional DMP before the grant is disbursed (e. g., NWO) or after (e. g., H2020). For the Dutch Research Council (NWO), the first RDMP deliverable is a section called data management, which includes many questions about reusable data. The analysis of data management sections we present in Section 4 uses those data management sections as input. Then, applicants submit a full DMP. A DMP answers a more comprehensive set of questions about RDM matters for the funded project.

For the European Commission (H2020), the paragraph is labeled "as a data approach." The data approach statement is a sub-section of the grant proposal. Additionally, both H2020 and NWO make use of DMPs for improving data management.

2.1 Evaluation of research data planning

Earlier, we touched upon strategic data planning (SDP) to show how business and industry have sought to increase the quality of (strategic) data and extract value from their data by integrating information systems. We showed that a similar planning approach is recently used in academia to produce reusable scientific data. A significant difference between strategic planning (SDP) and RDMP, though, is

Table 1: Comparison of RDM processes of national funding agencies.

Agency	ST	DMP	Process	Deliverable(s)
NWO	NL	Yes	Data management section in the application then Data management plan four months after the project is granted	Data management section, and the structured data management plan
ZONMW	NL	Yes	DMP after the project has been granted	The structured data management plan
ESRC	UK	Yes	DMP as part of the initial application	The structured data management plan
AHRC	UK	Yes	DMP as part of the initial application	The structured data management plan
NSF	US	Yes	DMP as part of the initial application	The data management plan of a maximum of two pages, directorate-specific

that SDP programs are deployed inside the boundaries of firms, while RDM is often an inter-organizational effort between funders and research institutions. Hence, it is quite complex to define generic criteria and rules for evaluating the quality of research data management due to the variety of data produced in science.

In RDM, quality criteria are mainly relying on FAIR principles. Recently, Science Europe standardized the areas of data management that data management plans should address. According to Science Europe, DMPs ideally cover the following aspects of data management: Data description and collection or re-use of existing data, documentation and data quality, storage and backup during the research process, legal and ethical requirements, codes of conduct, data sharing and long-term preservation and, data management responsibilities and resources [8].

3 Method

We conducted an exploratory case study in the Netherlands. We collected data using ten semi-structured interviews with representatives of funding agencies and research data management services in the Netherlands (national level) and one interview with the European Commission, which funds research at the European level through ERC and H2020 funding programs. The list of interviewees is shown in Table 2. Interviews lasted approximately 45 minutes, and were either conducted face to face or remotely on Skype. All interviews were recorded after obtaining the approval from the interviewees, who signed an informed consent form.

The interview protocol contains seven items. The first two items focus on the grant application process and aim at comparing different review processes per agency. The next three items question the criteria for judging the quality of data management planning. These questions help to determine what criteria are used for evaluating data management plans. Finally, the last three items falling under

Table 2: Organization, experience, and role of the ten interviewees. Identifiers (ID) are used in the results sections to refer to the interviewees.

ID	Organization	Experience	Role
F1	Government Agency	4 years	Grant support
F2	Dutch Funder	7–8 years	Grant evaluation
F3	Dutch Funder	8 months	Policymaking
F4	Government Agency	7–8 years	Policymaking
F5	European Funder	4 years	Grant evaluation
R1	University/Medical Centre	2 years	Research support
R2	University/Medical Centre	3 months	Research support
R3	University/Medical Centre	10 years	Research support
R4	University/Medical Centre	5 years	Research support
R5	University/Medical Centre	4 years	Research support

the challenges of data management planning ask the interview participants to describe technical and organizational challenges that occur in practice. The interview protocol was structured, as shown in Table 3.

The ten interviews were recorded, transcribed, and anonymized. The transcribed text version was subsequently analyzed with support from NVivo software (version 12.5). Interview transcripts were classified by type of organization. There are two organizations in the sample: funding agencies (F code) and research data management services (R code). In total, we coded the transcripts with 299 nodes in NVivo to structure the interview data from the (English) transcripts. Examples of such nodes are tools (such as DMPOnline), feedback on DMP, legal issues, and metadata. Next, we categorized those nodes into overarching groups of concepts, namely: checklist/completeness, openness, FAIR items, data archiving requirements, metadata, domain subjectivity, learning process and, institutional support.

Based on the interview results, which are shown in Section 4, we manually screened 98 proposals where a data management paragraph section is to be filled by grant applicants. There, we have seeking to classify proposals according to the reusability criteria communicated by the

Table 3: RDMP interview protocol.

Grant application process
1. Would you please describe the grant application procedure within your agency and how you proceed with the applications?
2. Why is it important to have a DMP in the proposal phase?

Quality of data management plans
1. Do you believe that grant applicants pay enough attention to the DMP?
2. What feedback do you provide to researchers about DMPs?
3. Would you please explain in detail how DMPs are reviewed before submission (or evaluated after submission)?

Challenges of data management planning in practice
1. What are the challenges in reviewing the DMP?
2. What could help you more to overcome these challenges?
3. Are you aware of any suitable software for RDMP?

interviewees. The results of the grant proposal analysis are found in Section 4.2.

4 Results

In this section, we present the results of the interviews (see Section 4.1) and the analysis of the data management statements in 98 grant proposals (see Section 4.2).

4.1 Interviews

In this section, we present the results of the interviews. This section is divided according to the main categories of the interview protocol, which are: grant application process (Section 4.1.1), quality of data management plans (DMPs), in Section 4.1.2, and challenges of research data management planning (RDMP) (Section 4.1.3). We noted that there is a variety of grant application processes. In our study, two types of processes were identified. The first process is when a data section is submitted with the proposal, and more complete DMPs are only required after the project has been granted. The second process only requires a data management plan.

4.1.1 Grant application process

As the interviewees explain, the goal of these two deliverables is different. The data management paragraph is used to create awareness, and the data management plan is an elaborated document where successful applicants give more detail about data management matters. "so those are very general questions, and NWO also accepts general answers to those questions they are not going into too much detail because that is something that will

come along when the grant is awarded. [...] if the grant is awarded, the researcher has to submit the data management plan within four months, and he or she cannot start the project until the DMP has been approved, so then things will be in more detail in DMP" [R4].

An interviewee F3 employed by a funding agency confirms that data management plans are more elaborated than paragraphs and have to be submitted later on: "But the main aim is to create awareness of the importance of good research data management. After the project is awarded, only the project leaders are requested to write a DMP according to the template the funding agency has, and, in that plan, they have to elaborate on the four questions they have answered in the main application form".

When a paragraph is not needed, funding agencies still expect researchers to be aware of data management matters though this requirement is not formalized in a specific section in the proposal. As explained by [F2], the paragraph has been removed from the proposal as "we did it before, but we did not do anything with it, and then we said if we do not do anything with it then why we do it, so we put it at the responsibility of researcher to take care of this part. I must admit that the researchers are very busy. They will not do that, but we strongly believe, and we strongly encourage that this is the own responsibility of researchers."

4.1.2 Quality of data management plans

Next, we seek to extract a set of quality criteria for research data management plans, the main deliverable of RDMP during the submission of grants. We grouped the quality criteria in three categories: completeness of the document, openness, and showing that a data management plan is a living document, which means that researchers announce that they will keep the DMP up to date.

The completeness of DMPs is a recurrent criterion appearing in the interviews. For funders and data services, a complete DMP shows that researchers thought seriously about data management matters. As explained by the interviewees: "Our main concern is completeness and whether the impression that the researchers have really thought about their data" [F3]. "if this type of basic things are missing or are explained in a very general way and not detailed, so this kind of check is more like a completeness check" [F5], "When it comes to DM plan I try to check whether they answered the questions and whether their answer is complete because if the questions have several parts, they [= researchers] tend to answer the first part and they tend to forget about the second part" [R4].

Regarding openness and exchange of research data, grant applicants might state in a DMP that their data will be exchanged between partners in a research consortium but not be necessarily made available outside the consortium, as explained by [F1], "there is still the option of opting out from openness of data, and a DMP is also focused on the exchange with consortium partners." The absence of justification on why data is not made openly available might lead to the rejection of the data management plan, as it happened to [R2]: "researchers say they will not open their data and they do not say why. So, the agency will get back and say, please explain why your data is not made openly available". When grant applicants mention closed data without justification, it happens that funders reject the DMP as they believe that researchers should at least share a part of their data "we always advocate this idea of maybe you cannot share all of your data, but certainly, there is some data that you can share. Then you choose a mixed access regime: part of it is open access, part of it restricted access. But it always needs to be findable, and it is always good description and license or how to access it and, who can access it and what is possible to do with it, who to contact if you cannot get direct access and if they like to use it in follow-up research and so on... So, research funders could be more demanding than they are at this moment it may be a matter of time". Nevertheless, according to [R5], competition for grants play a role against openness: "in my opinion, there is a controversy between open science and the competition within the disciplines as well because in one hand you are competing with others to get the grant then you are asked to put all your data open."

Submitting a data management plan is not the end; funders expect to receive future updates during the project. As explained by [R4], "you can simply start with the project, and within the six months, you were supposed to deliver your first version of DMP where the two standards agree on that the DMP is a living document. You

have strict deadlines for delivering the first version, but you should also deliver an updated version when there is a major change that is relevant, for instance where you planned to collect certain data, which is no longer possible or when for instance a partner that has strong impact on the data quite the consortium and stuff like that". In practice, we did not collect evidence that researchers indeed do DMP updates once the grant is received, on the contrary. Interviewee [R2] states that "No, we usually lose track of the researchers after they submit their DMP. Sometimes I run into them and ask them, and they keep me up to date. But we do not actively ask them about the DMPs after they had come to us for help".

4.1.3 Challenges of data management planning in practice

As mentioned earlier, the review of DMPs is still in an early phase, and that involves a learning process between funders, data stewards, scientists. More specific challenges identified in the interviews are discussed hereafter.

First, reviewers have no specific guidelines, especially for helping researchers with FAIR principles. "So, these principles need to be operationalized, maybe more concrete, but at the moment they do not. Often, it is not clear what it is, what do we need with findable, accessible, interoperable, reusable. We have some ideas, but we still need some guidelines, you talked about guidelines, but with this, we still need a lot more guidance" [R3].

Also, the absence of specific guidelines affect funding agencies as funders cannot provide feedback, "I think in the procedure we have right now it does not happen that often that researchers receive detailed feedback on their DMP as I explained earlier, we have more check on the completeness and adequateness of the answers but the [organization employing F3] has not this capacity as the average employee is not a data manager". Some funders promised feedback but also encounter issues with the required expertise: "Another problem with encounter is that we promised to give feedback on a DMP so that means that my colleagues have to know something about DM because they have to give feedback on DMPs" [F2].

Besides, there are only limited efforts put in planning data due to the low probability of obtaining a research grant. [R1] often sees researchers dealing with data management as a last-minute effort: "my experience at data section in the university was more or less the very last part of the proposal that was written, it was mainly a copypaste effort because we provided some template so there is not many individuals or thoughtful effort there because

that is when the DMP actually needs to be written. So, for the data section, it is more or less a copy-paste effort that the RDM team or other support team provided". [F1] concurs as this phenomenon is quite apparent in projects depending on the organization of consortia: "And then it is really difficult to also include their opinion in the data and data management plan in time for submission. You see the same thing actually happening also with the consortium agreements which are most of the time also signed after the proposal is granted and it is kind of interesting that also the commission looks for the additional effort before submitting a proposal without even knowing whether a proposal will get granted".

Another point of concern that funding agencies are aware of is that researchers lack the budget for RDM to make data available for the long term (after the project). At the same time, funding agencies expect the funding of long-term preservation to be a contribution from research institutions, as explained by F3: "a researcher can budget the management of research data within the project but of course not after the project has ended... Then it comes really to the budget of research institutions themselves, and I do not yet have a clear view on the availability of funds on an institutional level yet. I really can imagine that for some research, this is a problem, but also, we think that the university should invest in providing funds to store data for 10 years" [F3].

Lastly, there are academic disciplines where researchers lack knowledge about research data management solutions. Moreover, not only researchers lack insights into RDM matters. Also, when legal issues such as privacy are involved, it appears to be challenging for employees of research data services to provide the right support to researchers: "sometimes researchers are not aware of the technical possibilities of storing data on the network. When they do field research, they store data on the laptop, for instance, with all the possible dangers. Second, they are not aware of the technical possibilities when it comes to storing large data sets. They are very much surprised if they hear they can put up to terabytes without any costs on the network disks of our institution [...] sometimes we notice that they buy their own external hard disk because they just have the assumption that they do not have enough disk space.". In addition to proper infrastructure, legal knowledge is also not trivial to find at the side of research data services, as [R2] states "and personal data for the most parts, it is not my strong point I mean I am really trying to dive in to it. But that is the most delicate one because there is an actual legal requirement to protect data and these things and not only, I might not be knowledgeable about it. I mean... I know quite a lot about it, but I am
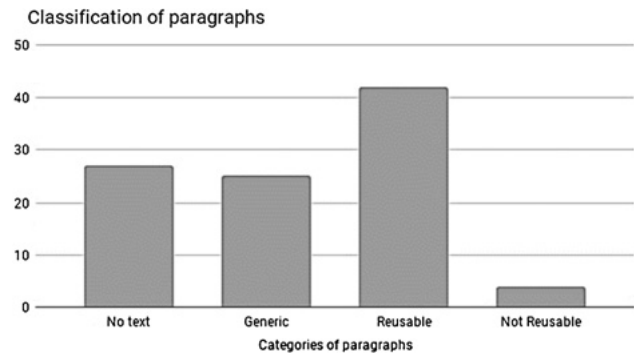

Figure 1: The results of the manual inspection of 98 data management paragraphs.

not a lawyer. But also, researchers often come to you with I do not have personal data. And you know that clearly, they do, they do interviews, so then you are telling them hey you have personal data and sometimes trying to explain that there are some issues. There are quite a lot of things that you change because you have personal data. These can possibly be the problem".

## 4.2 Grant application analysis

The study of the data management section in grant proposals aimed at exploring how grant applicants intend to plan research data management, and more specifically, how they intend to satisfy the reusability criterion set by funders. What can be seen from Figure 1 is that 42 data management sections in grant proposals met the requirements of reusability (as shown in Table 4). Nevertheless, 52 grant proposals (out of 98) did not contain an answer or a very generic answer to the proposal section dedicated to RDM.

5 Reflection and suggestions

We conceptualized the context of research data management planning in Figure 2, where we see that data management plans play a central role in research data management planning (RDMP). The main criteria presented by the interviewees are that DMPs must be complete, describe relevant metadata standards, and address the (absence of) openness of the research data produced during the funded project. Further, the ideal situation is that data management planning remains synchronized with the project, so any updates must be communicated to the funder. Research data services are depicted as a "boundary-spanning role" [15], where the objective of data manage-

Table 4: Criteria for the grant application analysis.

Criteria	Description
No text	The grant applicants have not answered the data management paragraph
Generic	The modalities of data reusability are not explained clearly, or the grant applicants indicate that the data is partially not reusable
Reusable	Grant applicants explain how their data will be reusable
Not reusable	Grant applicants explicitly state the data will not be reusable and open
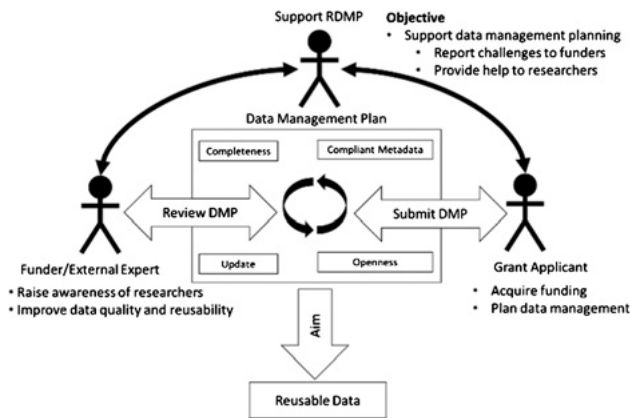
Figure 2: RDMP context where funders and grant applicants interact with research data management planning (RDMP) supporters at universities.

ment services is to translate the funder's requirements into acceptable DMPs and, subsequently, proper RDM in practice. Unfortunately, a key lesson from the interviews is that this "boundary-spanning role" is currently limited by the unavailability of appropriate guidelines and the capacity to provide feedback.

Furthermore, we provide an overview of the criteria which lead to satisfactory data management sections in grant proposals, i. e., the items that should be addressed by grant applicants. A list of items corresponding to the reusability criteria is shown in Table 5. The suggestion is to use the corresponding feedback to grant applicants before the proposal is submitted, with automated feedback technology using natural language processing (NLP) techniques. With automated feedback, the paragraphs in a grant proposal are divided based on a template of the proposal containing the questions to be answered by applicants. Then, sentences are selected, and queries are run on each sentence related to the RDM section. Examples of queries written to extract data from full-text paragraphs are: ["LOWER": "Publicly", "LOWER": "available", "LOWER": "archive", "POS": "NOUN"], ["LOWER": "partly", "DEP": "advmod", "LOWER": "available"]. In case an item, as shown in Table 5, is missing, feedback

corresponding to the category found in the document can be shown to grant applicants. In the end, the user can see the comments on the DM section as well as the rest of the proposal on his/her screen. So far, this approach has been preliminarily evaluated with "Impacter," an automated feedback tool on grant proposals (https://impacter.eu/). Hence, software solutions such as Impacter could further implement the items and corresponding feedback to also address research data management planning in grant proposals.

From the interviews, we observed that research data management planning is still at an early stage, where the goal of reusable scientific data is not yet fully supported by efficient planning processes and clear quality criteria. A lot of expectations are put on the shoulders of grant applicants, i. e., the researchers, but, at the same time, the technological infrastructure for RDMP is not yet fully functional. The fact that RDM and RDMP are complex for researchers has been investigated by a number of previous studies, for instance, [16, 17]. Our study confirms the view of RDM as a network of stakeholders, institutions and individual grant applicants with competing interests, sometimes leads to the production of non-reusable data.

Moreover, we covered a perspective of RDM where little previous research exists, as most studies focus on funders, researchers or, data services independently [18]. As such, we depicted that RDMP is an ongoing effort where data services also possess a mediating role between researchers and funding agencies, keeping funders informed of concrete issues occurring in practice, which then lead to the revision of funding policies.

As a suggestion of future work, one might investigate how technological solutions like automated feedback generation with natural language processing might circumvent some of the knowledge issues experienced by funders, data services and, researchers. Data reusability can be operationalized using the quality criteria of funders, as shown in Table 5. For instance, funders might expect that trusted digital repositories support compliant metadata standards, which then leads to an increasing the findability of the data. Such technology could implement en-

Table 5: Summary of items expected for reusable data by funders and data management services.

Item	Corresponding feedback
Data format	What documentation do you need to make your data format more authentic and transparent [R1]? If your data do not have a common form, please provide a relevant convert tool [F1]
Persistent Identifier	The DOI code and catalog and repository and data format are the critical requirements of Data Management for your project [F2]. "Thinking of a good license is indeed not really a challenge but you need to do it!" [F4]
Metadata	Metadata is important for the findability of data [R4]; to produce reusable data you need to be more specific about metadata. You can follow Dublin Core or Data Cite standards [R4, R2, F3]
Repository standard	"Most repositories they do confirm at least Dublin core standards" [R4]. You can find a list of trusted repositories on the funder's or university's website [R4, F3, R3]
Domain-specific Repository	Data management in different research fields varies a lot and domain experts will be reviewing your data management plan later [F5]; therefore, you have to comply with the demand of your faculty [R5] and get domain-specific guidance [F4]
Certified repository	Open access repositories are able to keep your data even after the project! It is better to calculate the cost of your data maintenance during and after the project [F3, F5, R1, R3]
Licenses	Dataset needs to be findable and have a good description and license or how to access it and who to access it and what is possible to do with it who to contact if you can't get direct access [F4]
Empty Paragraph	At this point in time, you need to talk about the basics of your data management [F1]. The DOI code and catalog and repository and data format are the critical requirements of Data Management for your project [F2]

tity recognition or more straightforward dictionary-based approaches to detect the presence or absence of these elements in data paragraphs and data management plans. The underlying reason is that we are far from being able to evaluate the effectiveness of RDMP without having insights into the outcome, i. e., the reusability of the data. Therefore, future studies should seek to systematically study data management plans to obtain a better view of the extent to which DMPs comply with the criteria for reusable data.

Additionally, we covered a perspective of RDM where little previous research exists, as most studies focus on funders, researchers or, data services independently [12, 19]. In Figure 1, we summarize the findings of the case study. As such, we depicted how RDMP is an ongoing effort where data services also possess a mediating role between researchers and funding agencies, keeping funders informed of concrete issues occurring in practice.

However, a limitation of our exploratory study is the limited number of funding agencies and representatives present in our sample. Still, we aimed at offering insights from outside of the Netherlands by including the perspective of the European Commission by interviewing a representative of the H2020 funding program. As we provided the questionnaire, we hope it will foster follow-up studies to conduct additional interviews in other EU countries and elsewhere, to provide a deeper understanding of the variety of grand application processes, and RDMP more specifically. Next, we saw that our findings corroborate studies covering other states than the Netherlands in terms of challenges found in Williams [7]. Williams et al. found a

similar trend indicating that funders and services focus on reusability and sharing but less on particular aspects during the project. That being said, none of the previous work attempted to cover the perspectives of funding agencies and data services, highlighting the reciprocal learning process in which both funders and data services co-evolve.

Another limitation is that we focused on the sharing aspect of research data, as it is the aspect that is emphasized in grant proposals. As such, the knowledge gained from the interviews is limited to data sharing, but other elements of RDM could be considered as well. Corti et al. [12, p. 27] described additional aspects than data sharing, which are essential to address during the planning phase: responsibilities, formatting, storing, ethics, copyrighting and, sharing. These aspects could be investigated more in-depth by analyzing data management plans instead of grant proposals, which were limited to data sharing and reusability in our case.

6 Conclusion

To conclude, research data management planning (RDMP) has many ongoing challenges, which makes the evaluation of its soundness and effectiveness to generate reusable data a complex task. Furthermore, while there is an agreement of our interviews to expect research data to be reusable, the practicalities and criteria might differ. So, we have identified differing planning processes, nonstandard quality criteria, and a series of complex challenges occurring during the planning phase. At the same

time, we have integrated recurring points of improvements from our respondents into actionable criteria to ensure that RDMP is addressing data reusability properly. By doing so, we have strived to contribute to a better understanding of RDMP as a crucial process within research data management.

References

1. L. Chan *et al.*, "Budapest Open Access Initiative | Budapest Open Access Initiative," 2002. [Online]. Available: http://www.citeulike.org/group/15400/article/11855295. [Accessed: 18-Apr-2016].
2. S. B. Ardestani et al., "B2SHARE: An Open eScience Data Sharing Platform," in *2015 IEEE 11th Int. Conf. e-Science*, pp. 448–453, 2015.
3. European Commission, *EU Open Innovation, Open Science, Open to the World*, 2016.
4. A. Lefebvre, E. Schermerhorn, and M. Spruit, "How Research Data Management Can Contribute to Efficient and Reliable Science," in *The 25th European Conference of Information Systems*, 2018.
5. D. Dietrich, T. Adamus, A. Miner, and G. Steinhart, "De-mystifying the data management requirements of research funders," *Issues Sci. Technol. Librariansh.*, vol. 70, 2012.
6. K. Wilms, S. Stieglitz, A. Buchholz, R. Vogl, and D. Rudolph, "Do Researchers Dream of Research Data Management?," in *Proc. 51st Hawaii Int. Conf. Syst. Sci.*, pp. 4411–4420, 2018.
7. M. Williams, J. Bagwell, and M. Nahm Zozus, "Data management plans, the missing perspective," *Journal of Biomedical Informatics*, vol. 71, pp. 130–142, 01-Jul-2017.
8. Science Europe, "Practical Guide to the International Alignment of Research Data Management," 2018.
9. W. K. Michener, "Ten simple rules for creating a good data management plan," *PLOS Comput. Biol.*, 2015.
10. D. L. Goodhue, J. A. Quillard, and J. F. Rockart, "Managing the data resource: a contingency perspective," *MIS Q.*, vol. 12, no. 3, pp. 373–392, 1988.
11. G. Shanks, "The challenges of strategic data planning in practice: An interpretive case study," *J. Strateg. Inf. Syst.*, 1997.
12. L. Corti, V. Van den Eynden, L. Bishop, and M. Woollard, *Managing and sharing research data: A guide to good practice*, 2014.
13. J. Hartter, S. J. Ryan, C. A. MacKenzie, J. N. Parker, and C. A. Strasser, "Spatially Explicit Data: Stewardship and Ethical Challenges in Science," *PLoS Biol.*, 2013.
14. M. D. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, 160018, Mar. 2016.
15. D. Hislop, R. Bosua, and R. Helms, *Knowledge Management in Organizations: A Critical Introduction*, 4th ed. Oxford, UK: Oxford University Press, 2018.
16. C. L. Borgman, *Big Data, Little Data, No Data: Scholarship in the Networked World*. The MIT Press, 2015.
17. K. G. Akers, "Going beyond data management planning: Comprehensive research data services," *Coll. Res. Libr. News*, 2017.
18. C. L. Borgman, "The conundrum of sharing research data," *Journal of the American Society for Information Science and Technology*, vol. 63, no. 6, pp. 1059–1078, 2012.
19. K. G. Akers and J. Doty, "Disciplinary differences in faculty research data management practices and perspectives," *Int. J. Digit. Curation*, 2013.

Bionotes

Armel Lefebvre
Utrecht University, Department of Information and Computing Sciences, 3584 CC Utrecht, Netherlands
a.e.j.lefebvre@uu.nl

Armel Lefebvre is a Ph.D. candidate in the organization and information group at Utrecht University. His research focuses on the socio-technical aspects of reproducible and open science.

Baharak Bakhtiari
Utrecht University, Department of Information and Computing Sciences, 3584 CC Utrecht, Netherlands
baharak.bakhtiari@gmail.com

Baharak Bakhtiari graduated from Utrecht University with an M.Sc. Business Informatics. She currently works as a Business Analyst in the banking sector in the Netherlands. Her graduation project focused on applying Natural Language Processing techniques to the Research Data Management framework. She works with data management systems and workflow management applications in her current job.

Marco Spruit
Utrecht University, Department of Information and Computing Sciences, 3584 CC Utrecht, Netherlands
m.r.spruit@uu.nl

Marco Spruit is an associate professor in the Natural Language Processing group within the Information and Computing Sciences department of the Faculty of Science at Utrecht University in the Netherlands. As principle investigator in the department's Applied Data Science Lab, his research primarily focuses on Self-Service Data Science.