



Delft University of Technology

On a quest for cultural change

Surveying research data management practices at Delft University of Technology

Mancilla, Heather Andrews; Teperek, Marta; Dijck, Jasper van; Heijer, Kees Den; Eggermont, Robbert; Plomp, Esther; Turkyilmaz - van der Velden, Yasemin ; Kurapati, Shalini

DOI

[10.18352/lq.10287](https://doi.org/10.18352/lq.10287)

Publication date

2019

Document Version

Final published version

Published in

LIBER Quarterly

Citation (APA)

Mancilla, H. A., Teperek, M., Dijck, J. V., Heijer, K. D., Eggermont, R., Plomp, E., Turkyilmaz - van der Velden, Y., & Kurapati, S. (2019). On a quest for cultural change: Surveying research data management practices at Delft University of Technology. *LIBER Quarterly*, 29(1). <https://doi.org/10.18352/lq.10287>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.



On a Quest for Cultural Change — Surveying Research Data Management Practices at Delft University of Technology

Heather Andrews Mancilla

Delft University of Technology, Delft, The Netherlands

H.E.AndrewsMancilla@tudelft.nl, orcid.org/0000-0002-6637-2830

Marta Teperek

Delft University of Technology, Delft, The Netherlands

M.Teperek@tudelft.nl, orcid.org/0000-0001-8520-5598

Jasper van Dijck

Delft University of Technology, Delft, The Netherlands

J.H.A.vanDijck@tudelft.nl

Kees den Heijer

Delft University of Technology, Delft, The Netherlands

C.denHeijer@tudelft.nl, orcid.org/0000-0003-0314-2779

Robbert Eggermont

Delft University of Technology, Delft, The Netherlands

R.Eggermont@tudelft.nl

Esther Plomp

Delft University of Technology, Delft, The Netherlands

E.Plomp@tudelft.nl, orcid.org/0000-0003-3625-1357

Yasemin Turkyilmaz-van der Velden

Delft University of Technology

Y.Turkyilmaz-vanderVelden@tudelft.nl, orcid.org/0000-0003-2562-0452

Shalini Kurapati

Delft University of Technology

S.Kurapati@tudelft.nl, orcid.org/0000-0003-1492-4100

Abstract

The Data Stewardship project is a new initiative from the Delft University of Technology (TU Delft) in the Netherlands. Its aim is to create mature working practices and policies regarding research data management across all TU Delft faculties. The novelty of this project relies on having a dedicated person, the so-called 'Data Steward,' embedded in each faculty to approach research data management from a more discipline-specific perspective. It is within this framework that a research data management survey was carried out at the faculties that had a Data Steward in place by July 2018. The goal was to get an overview of the general data management practices, and use its results as a benchmark for the project. The total response rate was 11 to 37% depending on the faculty. Overall, the results show similar trends in all faculties, and indicate lack of awareness regarding different data management topics such as automatic data backups, data ownership, relevance of data management plans, awareness of FAIR data principles and usage of research data repositories. The results also show great interest towards data management, as more than ~80% of the respondents in each faculty claimed to be interested in data management training and wished to see the summary of survey results. Thus, the survey helped identified the topics the Data Stewardship project is currently focusing on, by carrying out awareness campaigns and providing training at both university and faculty levels.

Keywords: Research Support; Data Management; FAIR principles; Data Management Plans; Library; University

1. Introduction

The importance of effective research data management (RDM) and sharing practices in research is nowadays highly recognised by funding bodies, governments, publishers and research institutions. The commitment to the Findable, Accessible, Interoperable and Re-usable (FAIR) principles (Wilkinson et al., 2016) is not only a requirement for all projects funded by the European Commission's Horizon 2020 funding scheme (European Commission, 2017), but they are also one of the fundamental principles of the European Open Science Cloud (European Commission, 2018). In addition to that, in the Netherlands, the Dutch government declared that Open Science and Open Access should be the norm (Regeerakkoord, 2017–2021). The two major national funding bodies, the Dutch Research Council (NWO) and the Netherlands Organisation for Health Research and Development (ZonMW), have detailed requirements for data management and data sharing as part of their research grant conditions (NWO, 2016; ZonMW, 2018). In parallel, more and more journals and publishers require that research data supporting research articles are made available (e.g., Nature research, 2016; PLOS, 2014). Last but not least, research institutions have also recognised the importance and necessity of good data management and transparency in research. In the Netherlands, this has been reflected in the National Plan Open Science¹ (NPOS), signed in 2017 by the Association of Universities in the Netherlands (VSNU), and in the Netherlands Code of Conduct for Research Integrity published in October 2018.²

Consequently, in order to ensure that high-level policies are reflected in day-to-day research practices, research institutions have started offering additional support services for RDM. At TU Delft, central library support for RDM and data sharing has been in place already for several years. Furthermore, TU Delft is part of the 4TU consortium of technical universities in the Netherlands and it is home to the 4TU.Centre for Research Data archive³ (4TU.ResearchData), which functions as a certified, trusted repository (Data Seal of Approval⁴) for long-term preservation and sharing of research data. Both, the TU Delft Research Data Services and 4TU.ResearchData Services have been evaluated using the Research Infrastructure Self-Evaluation Framework (RISE) (Rans & Whyte, 2017). This framework helped assessing the maturity levels of the provided services regarding research data management. From this, it was clear that more effort had to be injected into policy

development and training.⁵ In line with the fact that bottom-up community-driven approaches are favoured at TU Delft,⁶ we believe that data management support needs to be discipline-specific in order to be truly relevant to our research communities.

Heading in such direction, TU Delft's executive board provided funding for three years (2018–2020) to initiate the Data Stewardship project at TU Delft. A dedicated Data Steward with a subject specific background (a PhD or an equivalent experience in the faculty-related research area) was hired at every TU Delft faculty. All Data Stewards are coordinated by the Library at TU Delft, and constantly interact with other support staff in order to develop mature working practices for RDM across the campus.

How can we approach such a task? We reasoned that we first need to understand what the current practices are, and based on that, develop a system which allows us to improve such practices and regularly assess their progress. Hence, two main strategies were adopted: 1) conducting qualitative, semi-structured interviews with researchers across the faculties;⁷ and 2) run quantitative surveys about data management practices at TU Delft in a periodic fashion. The semi-structured interviews provide an important in-depth insight into researchers' needs and are necessary for building trust and connections with the research community. Additionally, having a broader quantitative overview of RDM practices is necessary to provide robust benchmarking of the project.

This paper presents the results of the first RDM quantitative survey carried out at TU Delft. The survey is partly based on the Data Asset Framework (DAF) (Johnson, Parsons, & Chiarelli, 2016). The DAF survey is a comprehensive tool that allows institutions to assess researchers' data management practices and identify gaps in service provisions. However, since the DAF survey is rather lengthy (consisting of over 60 questions), it was decided that the general principle of the DAF framework would be followed, but that the questionnaire itself would be substantially simplified into a survey containing a total list of 22 questions.

2. Method

The survey was developed as a web-based questionnaire and it was distributed via email to all staff members of 6 out of the 8 faculties of the TU Delft.

The 2 remaining faculties did not have a Data Steward before July 2018 (Data Stewards were incorporated at different times, and the survey was carried out only at the Faculties that had a Data Steward in place).

The survey was sent in two runs. The first run was carried out in November 2017 at the Faculties of Aerospace Engineering (AE), Civil Engineering and Geosciences (CEG), and Electrical Engineering, Mathematics and Computer Science (EEMCS). The second run was carried out in the months of May-June 2018 at the Faculty of Technology, Policy and Management (TPM), and in the months of May-July 2018 at the Faculties of Mechanical, Maritime and Materials Engineering (3mE) and Applied Sciences (AS).

The survey consisted of 22 questions about RDM, aside questions asking for demographic information (e.g., position, institution, faculty, department, among others). The topics included automatic data backups, time frame and frequency of data loss, use of dedicated tools for RDM, data ownership, data stewardship, data management plans (DMPs), awareness of FAIR data principles, and use of research data repositories. The response scheme was mostly multiple choice with categorical answers (e.g., 'Yes,' 'No' and 'Not sure' options). The analysis shown in this article was carried out using the software Tableau Reader v2018.2.

In order to encourage responses, the respondents were given the possibility to be included in a draw for vouchers of a known commercial house in the Netherlands. Those who wanted to participate in the draw and/or wanted to receive information about the results were asked to provide their email addresses. The results of the draw carried out at each Faculty were disseminated accordingly by each Data Steward. Data was anonymized by removing identifiable features, and the raw files were destroyed.

2.1. Data Availability

A description of the survey and the questions are publicly available in Open Science Framework under the name 'Quantitative assessment of research data management practice' (Teperek et al., 2019). The anonymized data is publicly available in Zenodo under the title 'Quantitative assessment of research data management practice' (Krause et al., 2018), and a visualization of the survey is available at Tableau Public under the name 'TU Delft Quantitative Assessment of Research Data Management Practice 2017–2018.'⁸

The survey was also carried out at the Ecole Polytechnique Federale de Lausanne (EPFL) at the end of 2017. The report of those results can be found in the website of the EPFL Library.⁹ The results given in this work correspond to those for TU Delft only.

2.2. Response Rates

The survey was sent to all staff members per faculty. The total number of respondents was 680. Among these, 628 respondents correspond to ‘Full Professors,’ ‘Associate Professors,’ ‘Assistant Professors,’ ‘Postdocs/ Researchers’ and ‘PhD candidates.’

Table 1 lists the response rates per academic position per faculty. Considering Full Professors, Associate Professors, Assistant Professors, Postdocs/ Researchers and PhDs candidates, the total response rates per faculty varied from 8% at EEMCS to 37% at AE. The majority of the respondents were PhD candidates, representing 52% of the responses (see Figure 1). The response rate from Full Professors on the other hand was of 5% (varying from no responses at CEG to 48% at AE).

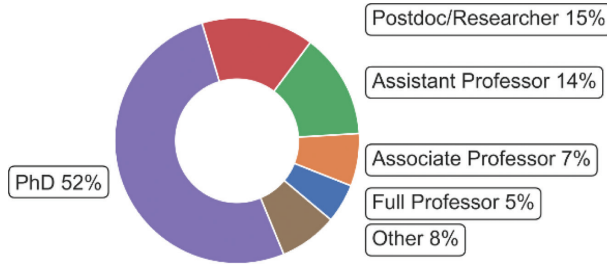
In the following section, the results will be presented considering the responses from Full/Associate/Assistant Professors, Postdocs/Researchers and PhD candidates in order to restrict the answers to data associated with research.

Table 1: Response rates per academic position at each faculty (%).

Position	Response Rates					
	AE	AS	CEG	EEMCS	TPM	3mE
Full Professor	48	7	–	9	17	6
Associate Professor	79	24	-	9	25	10
Assistant Professor	47	16	13	30	33	14
Postdoc/Researcher	36	4	8	2	16	8
PhD candidate	30	10	21	10	17	12
Total	37	9	13	8	20	10

For the total response rates we have considered Full Professors, Associate Professors, Assistant Professors, Postdocs/Researchers and PhD candidates. No Full Professors nor Associate Professors from CEG replied to the survey.

Fig. 1: Total number of responses per academic position (%). The 'Other' category includes MSc students, Guest Researchers, Lecturers, among others.



3. Results

3.1. Data Backup & Data Loss

Figure 2 presents the responses regarding automatic backups of research data. About 43% of the respondents do not have the data automatically backed up, while the percentage of people answering 'Yes' to the question 'Is your research data automatically backed up?' is 42% on average, ranging from 39 to 47% across faculties (see Table 2).

Fig. 2: Responses regarding automatic backups of research data. On average, 42% of the respondents have the research data automatically backed up, against a 43% of respondents that claim not to have the data automatically backed up. See also Table 2.

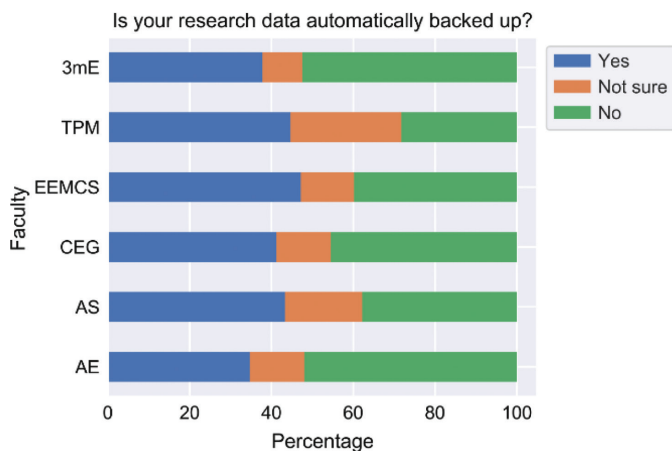


Table 2: Results to the question ‘Is your research data automatically backed up?’

Is your Research Data Automatically Backed up?						
Answer	AE	AS	CEG	EEMCS	TPM	3mE
Yes	36	42	41	47	45	39
Not sure	14	19	13	13	27	10
No	51	39	46	41	28	52

Results are given in percentages relative to the total number of respondents from each faculty. The percentages have been rounded to the nearest integer.

Responses from different faculties appear to be similar, with the exception of the responses from TPM faculty, where the percentage of respondents not doing automatic backups is the lowest across all faculties (28% compared to 39 to 52% for the other faculties). However, the overall share of those who do not know if the data is backed-up at TPM is the highest.

Focusing on the answers per position, the percentage of respondents in higher positions of the academic career (i.e. Full/Associate/Assistant Professors) that do automatic backups is greater than that of the PhD candidates that replied to the survey (see Table 3).

Regarding data loss, Figure 3 shows the responses per faculty to the question ‘Did you lose any research data in the past year?’ Table 4 lists the responses per academic position. According to Figure 3, answers across all faculties appear to behave similarly. On average, about 13% of the respondents in each faculty claim to have lost data in the past year. Also percentages of data loss

Table 3: Percentage of respondents that do automatic backups per position.

Position	Respondents that do automatic backups (%)
Full Professor	43
Associate Professor	49
Assistant Professor	56
Postdoc/Researcher	41
PhD candidate	36
Other	34

The numbers are given relative to the total number of respondents in each academic position considering all faculties. All percentages have been rounded to the nearest integer.

Fig. 3: Responses regarding research data loss in the past year. On average, 13% of the respondents claim to have lost research data in the past year.

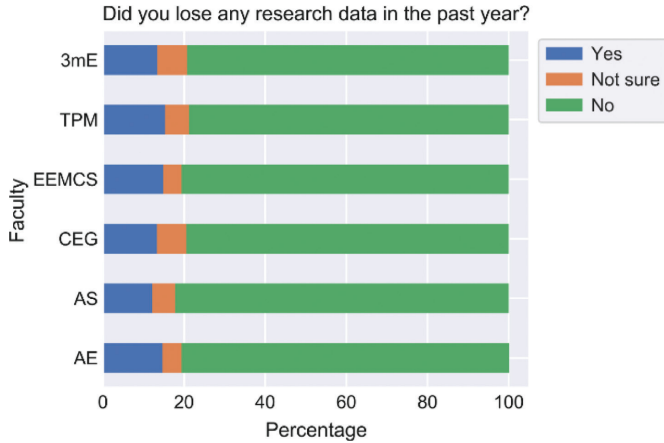


Table 4: Percentage of respondents who have lost data in the past year.

Position	Respondents that do automatic backups (%)
Full Professor	11
Associate Professor	11
Assistant Professor	15
Postdoc/Researcher	9
PhD candidate	14
Other	13

The percentages are given per academic position considering all faculties. Percentages have been rounded to the nearest integer.

are at a similar level considering the responses per academic position (see Table 4). Interestingly, PhD candidates and Assistant Professors show the largest percentages of data loss (14 and 15% respectively).

Cross-correlating the responses between doing automatic backups and losing data, it is interesting to see that in almost all faculties, the percentage of data loss (in the past year) indicated by respondents that do automatic backups is lower than the percentage of data loss indicated by respondents that do not do automatic backups (see Table 5). Only for the TPM faculty it turned out to

Table 5: Comparison of data loss percentages between respondents that do automatic backups, and those who claim not to have their research data automatically backed up.

Faculty	Do automatic backups and have lost data in the past year (%)	Do not do automatic backups and have lost data in the past year (%)
AE	2	23
AS	8	15
CEG	10	16
EEMCS	10	21
TPM	13	8
3mE	6	16

All percentages have been rounded to the nearest integer. The average data loss percentages of respondents that do automatic backups is of 8%, while the average data loss percentage of respondents that do not automatically backup the data is of 17%.

be the other way around. As listed in Table 5, data loss percentages of respondents that do automatic backups is of 8% on average, while that indicated by respondents that do not do automatic backups is of 17%.

3.2. Research Data Repositories

When queried about being aware of research data repositories, respondents could choose one of the following answers: ‘Yes, I am already using them to find existing datasets or to share my own data;’ ‘Yes, I am aware of research data repositories, but I have not used them;’ ‘Not sure;’ ‘No, I have no idea what these are.’ Results show respondents appear to be aware of research data repositories, but are not necessarily using them (see Figure 4 for responses per position, and Table 6 for responses per faculty). The most common answer in all faculties was ‘Aware but not using,’ ranging from 42% of the replies at AS faculty to 61% at TPM. Only about 16% of all respondents per faculty claim to be using research data repositories to find existing datasets or to share data.

Participants were also asked whether they had heard about the 4TU. ResearchData, for which respondents could reply ‘Yes,’ ‘No,’ or ‘Not sure.’ Inspection of those results shows that between 4 (AS) to 31% (TPM) of the respondents who replied ‘Not sure’ to being aware of research data repositories, claim to have heard about the 4TU. ResearchData repository

Fig. 4: Responses regarding awareness of research data repositories. The answers respondents could choose from have been shortened to ‘Using’ (option ‘Yes, I am already using them to find existing datasets or to share my own data’); ‘Aware but not using’ (‘Yes, I am aware of research data repositories, but I have not used them’); and ‘Not aware’ (‘No, I have no idea what these are’). The results are given in percentages considering all faculties. In general, respondents tend to be aware of research data repositories, but claim not to be using them.

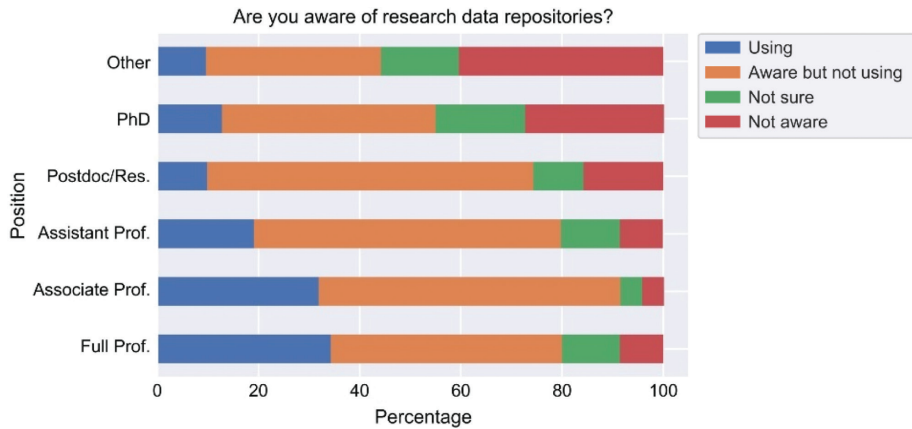


Table 6: Results to the question ‘Are you aware of research data repositories?.’

Are you aware of research data repositories?						
Answer	AE	AS	CEG	EEMCS	TPM	3mE
Using	17	15	14	24	11	17
Aware but not using	45	42	52	46	61	45
Not sure	18	15	13	11	13	18
Not aware	19	28	21	20	15	19

Answers have been shortened as defined for Figure 4. Results are given in percentages relative to the total number of respondents from each faculty. All percentages have been rounded to the nearest integer.

(Table 7). Moreover, among the respondents who have heard about the 4TU. ResearchData, an average of 8% replied ‘Not aware’ (i.e., chose the option ‘No, I have no idea what these are’) when asked about research data repositories (Table 7). These contradictions suggest respondents either do not know what repositories are, or do not know very well what the 4TU. ResearchData is (see more in Discussion).

Table 7: Comparison of answers from survey respondents regarding awareness of research data repositories, and awareness of the 4TU.ResearchData.

Faculty	Respondents who use repositories and have heard about the 4TU.ResearchData	Respondents who are 'not sure' of being aware of repositories, and have heard about the 4TU.ResearchData	Respondents who have heard about the 4TU.ResearchData and claim not to know what repositories are
AE	64	25	11
AS	77	4	4
CEG	43	9	6
EEMCS	58	16	10
TPM	67	31	10
3mE	55	10	7

The numbers correspond to percentages per faculty. All percentages have been rounded to the nearest integer.

Fig. 5: Responses to the question 'Does your project have a data management plan?.' Responses are given as percentages with respect to the total number of respondents per faculty.



3.3. Data Management Plans & FAIR Data

Figure 5 shows most respondents stated they were not working on a project with a DMP by the time they replied to the survey. Only ~19% of the respondents claim to be working in projects with a DMP, and a similar percentage is not sure whether the project they are working on has a DMP or not.

Table 8: Comparison of responses between researchers who work on projects with a DMP, and those who do not.

Respondents	Use repositories	Aware or use repositories	Have lost data in the past year	Do automatic backups	Aware of FAIR data
Work on projects with a DMP	27±11	81±11	10±6	48±12	51±15
Do not work on projects with a DMP	15±4	62±7	12±3	44±6	24±6

The numbers represent the average and the standard deviation calculated from considering responses per faculty. All percentages have been rounded to the nearest integer.

Table 9: Results regarding awareness of FAIR data.

Are you aware of funders' expectations for Findable, Accessible, Interoperable and Reusable (FAIR) data?						
Answer	AE	AS	CEG	EEMCS	TPM	3mE
Yes	23	24	24	30	40	30
Not sure	21	32	27	19	18	20
No	56	44	48	51	42	49

Results are given as percentages relative to the total number of respondents from each faculty. All percentages have been rounded to the nearest integer.

Interestingly, among the respondents who are either aware or using research repositories (see Table 6), we find that the percentage of respondents working on projects with DMPs is greater than the percentage of respondents who do not work with DMPs (see Table 8). This also holds among the respondents who are aware of FAIR data (see Table 8).

Concerning FAIR data awareness alone, more than 50% of the respondents at each Faculty are not 'aware' or are 'not sure' of funders expectations for FAIR data (see Table 9). In general, the percentage of respondents who answered to be aware of FAIR, is at the 20–30% level across faculties (except at TPM faculty; see Table 9). Most of these answers are from staff members in higher positions of the academic ladder (see Figure 6).

Results also show that respondents who are aware of FAIR data tend to also be 'aware of or using' research data repositories, as opposed to the respondents who are not aware of what FAIR data is. However no significant

Fig. 6: Awareness of FAIR data. The percentages are given with respect to the total number of respondents per academic position (across all faculties).

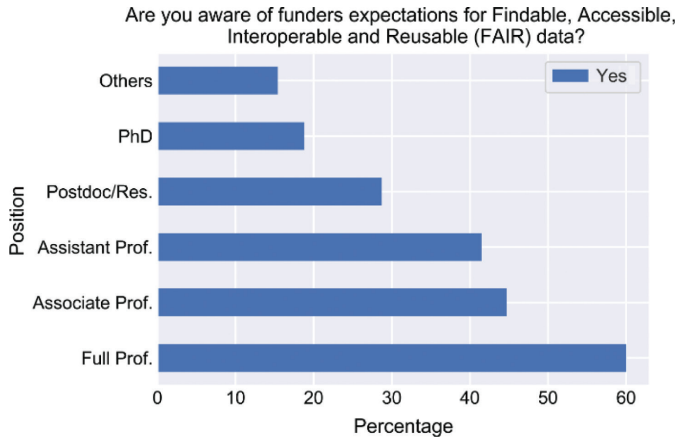


Table 10: Comparison between respondents who are aware of FAIR data and those who are not.

Respondents	Aware or use repositories	Use repositories	Aware of 4TU.ResearchData
Aware of FAIR	87±7	27±14	60±11
Not aware of FAIR	53±6	9±5	28±9

The numbers represent the average and the standard deviation calculated from considering responses per faculty. All percentages have been rounded to the nearest integer.

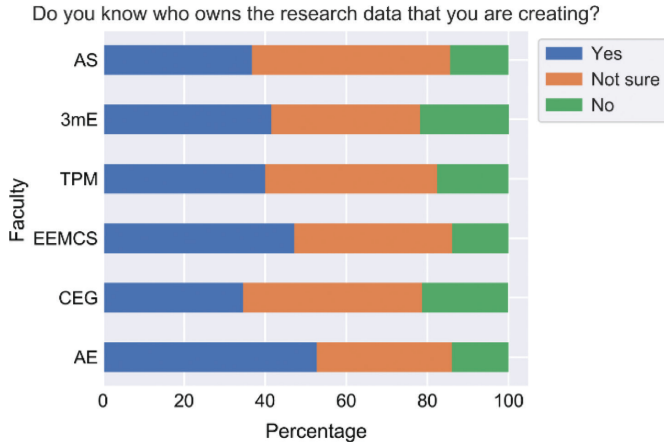
difference is detected when comparing directly with usage of research data repositories alone (see Table 10).

This positive trend of FAIR data and research data repositories awareness is also seen when comparing the answers to the question about having heard of the 4TU.ResearchData archive (see Table 10).

3.4. Data Ownership

Overall researchers -particularly PhD candidates- show little awareness about who owns the data. Participants were specifically asked 'Do you

Fig. 7: Results regarding data ownership awareness. Responses are given as percentages considering the total number of responses per faculty.



know who owns the data you are creating?.' Only those who responded 'yes' to that question were asked to specify who the owner(s) of the data was(were). The results show that at least ~50% of all the respondents of each faculty do not know or are not sure of who the owner(s) of the data is(are) (see Figure 7).

Researchers in higher academic positions appear to be more aware of data ownership, particularly Full Professors and Associate Professors (>60%; see Table 11). Less than 50% of the Postdocs claim to know who owns the data. PhD candidates on the other hand, appear to be the least aware of data

Table 11: Responses regarding data ownership.

Do you know who owns the research data that you are creating?			
Position	Yes	Not sure	No
Full Professor	66	31	3
Associate Professor	64	28	8
Assistant Professor	57	35	8
Postdoc/Researcher	45	41	14
PhD candidate	33	44	23

Results are given as percentages relative to the total number of answers per academic position (considering all faculties). All percentages have been rounded to the nearest integer.

Table 12: Data ownership responses among PhD candidates.

Faculty	Claim to know who owns the data	Claim full or partial ownership of the data	Claim full or partial ownership with respect to total PhD responses
AE	44	17	8
AS	25	22	6
CEG	35	28	10
EEMCS	40	24	10
TPM	24	67	16
3mE	22	27	6

Only respondents who answered ‘yes’ to the question ‘do you know who owns the research data that you are creating?’ were asked to specify who the owner(s) of the data was(were). The last column on the right lists the percentage of PhD respondents who claimed full or partial ownership, considering the total number of PhD responses per faculty. All percentages have been rounded to the nearest integer.

ownership, with a ‘Yes’ percentage of 33% considering the responses from all faculties (Table 11). Furthermore, between 17 (AE) and 67% (TPM) of the respondent PhD candidates who affirm knowing who the owner(s) is(are), claim some degree of ownership on the data they manage (see Table 12). This translates to an average of ~9% of all respondent PhD candidates claiming to have either full or partial ownership of the data (right column of Table 12); where partial ownership appears to be shared with many different stakeholders (e.g., TU Delft, supervisor, research group, company, public, funder, etc.) and combinations thereof.

The unawareness regarding this topic is also apparent from the written comments added to the answer of ‘You said you know who owns the research data that you are creating. Who is it?’. Examples of such comments are: ‘Me! Well the university I guess’ (PhD candidate), ‘Department and supervisors’ (PhD candidate), and ‘The regulations are not completely clear on this, but as far as I remember it’s the authors’ (answer from Associate Professor).

3.5. Stewardship of Research Data

Respondents were also asked ‘Who do you think is responsible for the stewardship of research data resulting from your project?’. However confusion about the term ‘stewardship’ was apparent from the answers, suggesting not everyone is familiar with this term in the first place. This was clear from

the first run of the survey at AE, CEG and EEMCS faculties. Thus, it was decided that the question would be modified to ‘Who do you think is responsible for the management of the research data resulting from your project?’ for the surveys carried out later at the faculties of 3mE and AS. Interestingly, such change in formulation of the question had no significant impact on the results: the term ‘management’ was found to be similarly confusing to the term ‘stewardship.’

Considering the above, most staff members (84% at AE; 94% at AS; 87% at CEG; 77% at EEMCS; 91% at TPM; and 92% at 3mE) acknowledge their role in being responsible of taking care of the data in the projects they are involved in. However, this responsibility is also said to be shared with other university stakeholders. In this regard, PhD candidates indicated their supervisor is either full or partially responsible for the data stewardship throughout the research projects (e.g., 37% at TPM, 50% at CEG, 40% at EEMCS and 37% at AE).

Participants were also asked whether they had heard about the Data Stewardship project and data management support at their faculties. Among the answers, respondents from TPM appear to be more familiar with the Data Stewardship project and dedicated support (45%; see Figure 8), while such answer in the other faculties varied from 15 to 27% (Figure 8).

Fig. 8: Responses regarding Data Stewardship project and dedicated support on RDM at the faculties. The results are given as percentages relative to the total number of responses per faculty.

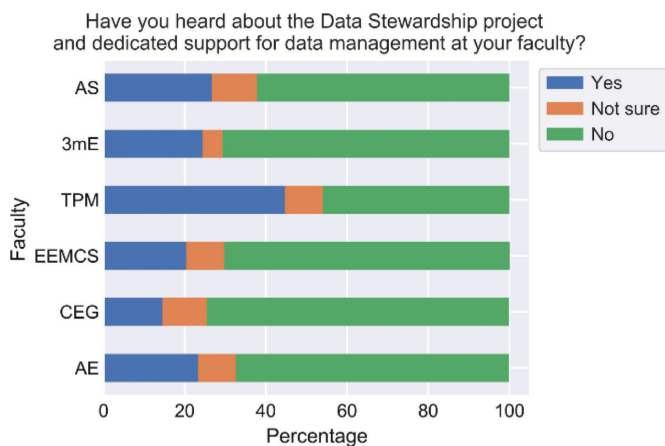
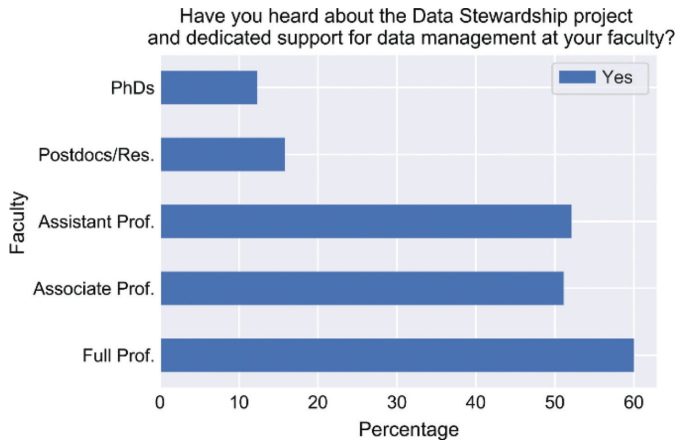


Fig. 9: Responses regarding Data Stewardship project and dedicated support on RDM at the faculties. The results are given as percentages considering the total number of respondents per academic position (from all faculties).



Breaking down the answers by academic position, we find that in general (Full/Associate/Assistant) Professors are more aware of the Data Stewardship Project and dedicated support for RDM than the other staff members (see Figure 9). On the other hand, <20% of the total number of Postdocs/Researchers and PhD candidates respectively, claim to be aware of the Data Stewardship Project and dedicated support.

3.6. Interest in Training

Regarding training in RDM topics, researchers were asked 'Please indicate if you (or related staff/students) would be interested in any potential training on research data management.' Figure 10 shows the results considering the total number of answers per academic position. Among the offered training topics were: 'General introduction to research data management;' 'Data management plan preparation;' 'Data backup and storage solutions;' 'How to use repositories for data sharing and searching for existing datasets;' 'Data ownership and licensing;' 'Using version control software;' 'Funders' requirements for data management and sharing;' 'Working with confidential

data (personally identifiable, commercially sensitive etc.); ‘Data carpentry’;¹⁰ ‘Software Carpentry’;¹¹ among others. The names of such trainings have been shortened in Figure 10 for the sake of better visualization. Respondents were allowed to choose multiple topics if desired.

Fig. 10: Interest in RDM training. Surveyed participants were presented different training options, and were asked to choose the ones that would be of interest to them or related staff. Different panels show the preferred options of the respondents per academic position. From first to last panel, answers are shown for: Full Professors, Associate Professors, Assistant Professors, Postdocs/Researchers and PhD candidates. The names of the offered options have been shortened for better visualization of the results (e.g., Funders’ requirements refers to funders’ requirements for data management and sharing).

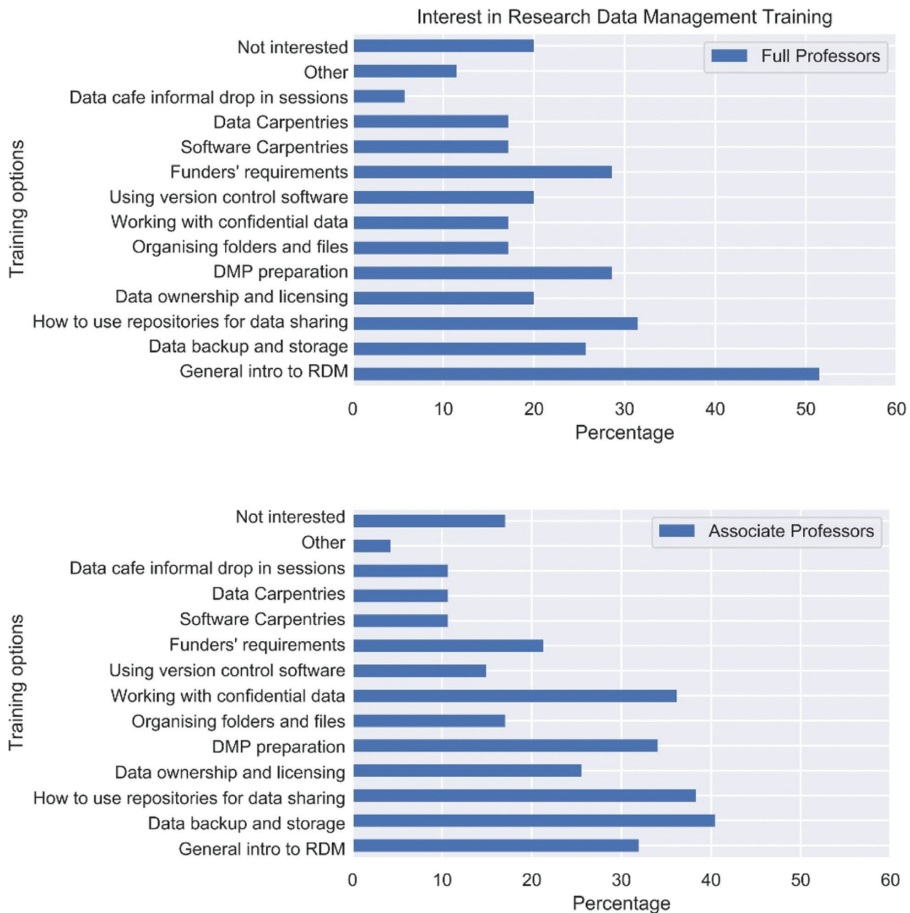
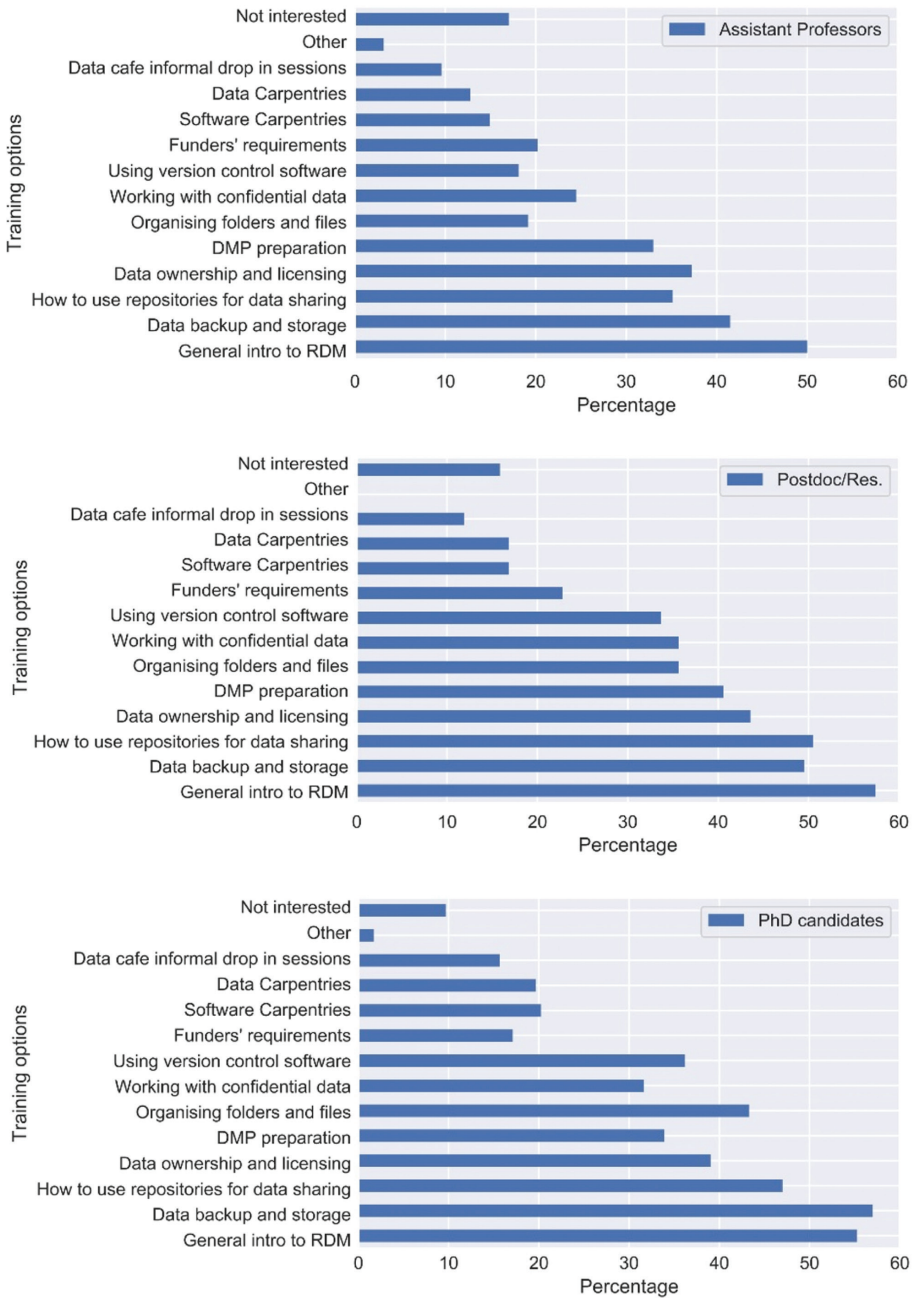


Fig. 10: (Continued)



According to the results, there appears to be great interest among the surveyed researchers: more than 80% of the respondents are interested in RDM training. Interestingly, researchers in different academic positions expressed interest in different topics: Full Professors are mostly interested in a 'General Introduction to Research Data Management.' Associate and Assistant Professors expressed more interest in 'Working with confidential data,' and 'Data Ownership.' While Postdocs/Researchers and PhD candidates appear to be mostly interested in a 'General Introduction to Research Data Management,' but also in 'Data Backup and Storage.' These results appear to be consistent with what each academic position faces at work on a daily basis in terms of RDM.

4. Discussion

The questions in this survey aimed to target general RDM practices, and not necessarily faculty-specific ones. Hence, it is not surprising the results of this survey showed similar trends across the different faculties of the university.

In general, we find some concerning practices that might suggest researchers are not familiar with what the university has to offer regarding RDM; and/or there is little education about what data management is, and how research can benefit from it.

The fact most respondents do not have the data automatically backed up or do not know if the data is automatically backed up, indicates a great fraction of the respondents might be performing manual backups, and/or do not know very well what TU Delft ICT solutions are regarding (at least) data backups (e.g., poor use of the TU Delft network drives).

The possibility of manual backups being a common practice among researchers (especially PhD candidates) is of great concern, since such practice leads to a substantial higher risk of data loss, than when relying on automatic backups. Percentages of data loss registered in the last year are at the 10% level, however such data loss occurrences have caused delays of up to 6 months of work. In addition to this, the percentage of data loss indicated by respondents that do automatic backups is lower than that indicated by respondents that do not do automatic backups. Hence, the Data Stewardship project has the mission to encourage researchers not to rely only (even less mainly) in

manual backups. Along with that, researchers should be encouraged to make use of TU Delft ICT resources and RDM services.

The lack of use of the TU Delft network drives and/or the little understanding of these solutions is quite apparent from the text comments written by participants who 'claim to do automatic backups.' When asked how those automatic backups are done, examples of typical answers are: '*Managed by the ICT department at our faculty. The frequency I don't know. I put the data on the project drive (U);*' '*Once a day, usually backed up in a harddisk or a usb disk, myself manages the backup;*' '*Twice a week, my data is backed up in my mobile hard disk;*' '*On USB hard drives separate from the systems I work on, or remotely.*' Moreover, only 34% of the respondents doing automatic backups, mention the university network drives (most of the times using them together with other backup solutions). About 28% of the respondents doing automatic backups mention Surfdrive¹² (most of them mentioning Surfdrive alone); 16% mention Dropbox (either alone or together with other platforms); and 7% mention Google Drive (either alone or together with other platforms).

On a more concerning note, the free-text comments about how automatic backups are done show that some respondents who 'have' the research data automatically backed up, are doing it by themselves. Hence it is not clear what definition of an 'automatic backup' the respondents considered when answering this question (only respondents who claimed to do automatic backups, were asked how the backups are done). It is the aim of the Data Stewards then, to increase awareness regarding the sensitivity and security of data, and which data storage, backup and processing solutions are the most suitable ones for each data type.

Another example of the lack of awareness about TU Delft RDM services, comes from the responses about the 4TU.ResearchData. Even though TU Delft researchers claim to have heard about 4TU.ResearchData, the survey results suggest respondents might not necessarily know what the 4TU.ResearchData archive is. The contradictions mentioned in section 3.2 indicate some respondents might not fully understand what a research repository is, and/or what the 4TU.ResearchData is for.

At last, this lack of knowledge about TU Delft RDM services is also apparent when asked about Data Stewardship project awareness, and knowledge of ICT support for RDM (Figures 8 and 9). Only 15 to 27% of the respondents

claimed to have heard about them (Figure 8). On one hand, such unfamiliarity with the Data Stewardship project is not surprising, since the Data Stewards had recently been introduced at their respective faculties when the survey was sent out. On the other hand, the question also mentioned the university ICT support, and the replies from specially early career researchers were still rather poor. This reveals another challenge for Data Stewards which is: bringing RDM to the day to day practices of (specially) early career researchers.

The issue mentioned above also brings up the lack of education regarding RDM. This is also clear from: the confusion about the terms 'stewardship' and 'management;' the contradictions on research data repositories; and the comments on how automatic backups are done. In addition to that, when asked about what 'data management tools' respondents use, some of the tools that were mentioned (as free text responses) included 'Mendeley,' 'hard-drives,' 'Google files,' 'Google drive,' 'MyBrain,' 'Dropbox,' 'OneDrive,' 'Onenote,' among others aside 'Git,' 'Github,' 'Gitlab,' 'Subversion,' 'Bitbucket' and 'Mercurial.' Interestingly 'papers,' 'Digital computer,' 'slack,' and 'plain simple ASCII text files,' were also mentioned as 'data management tools.'

From the results of this survey, we see the need for further awareness raising and education with respect to RDM topics. This should be addressed at both an early career stage (e.g., PhD candidates) and among established researchers (i.e., Professors). Senior researchers are clearly more familiar with policies and regulations, however they are not necessarily aware of the daily RDM practices these policies imply.

In addition to that, the survey results pose a new question for us: do researchers value proper RDM practices? Or are these only seen as new funder/institutional mandates? This question is guided by the relation found between the responses about 'FAIR data awareness' and 'awareness or use of research data repositories;' while no relation with solely 'use of research data repositories' was observed (Table 10). In addition to that, only 19% of the respondents claimed to be working on a project with a DMP, and a similar percentage is observed for respondents 'not being sure of' whether they are working on a project with a DMP or not (Figure 5). Hence, it is not clear whether researchers see the benefits of following FAIR principles and DMPs, or if these are only viewed as regulatory requirements from (mainly public) funders. Regardless of that, the results show that DMPs are indeed great tools to increase awareness about adequate RDM practices. Based on this, the Data Stewardship

project is currently focusing on bringing awareness into actual practice: encouraging researchers to recognize tools such as DMPs not only as funder deliverables, but also as useful instruments to take good care of the data.

A relevant aspect of data management that also raises concerns is data ownership. As seen in section 3.4, over 50% of the respondents 'do not know' or 'are not sure of' who the owner of the data is. Researchers in higher academic positions appear to be more aware of data ownership than early career researchers. This might be related to the fact that established researchers are the ones directly involved in the contractual phase of research projects. From the survey results, it is not clear if such information is accordingly disseminated to the early career researchers, who manage relevant research data on a daily basis. This we find a relevant subject, since once data ownership is clearly established, and well communicated to all team members from the beginning of a project, it makes things clearer when deciding on how the data should be managed throughout the project and the restrictions thereof (e.g., data encryption, data sharing, protected storage).

Clarifying responsibilities regarding data is also relevant. In this aspect, most staff members do recognize their role in being (either fully or partially) responsible for the data in the projects they are involved in (section 3.5). Among PhD candidates, between 37 and 50% claim their supervisor is either fully or partially responsible for data management. Respondents who claim they have either full or partial ownership on the data tend to also recognize responsibility on the data. Such responsibility is assumed either alone or shared with other university stakeholders (mostly supervisor and ICT manager). However, this also holds for the respondents who 'do not know' or 'are not sure of' who the owner of the data is. In other words, respondents acknowledge responsibility regardless of ownership. This in addition to the great interest respondents show about RDM training (section 3.6), definitely help setting up the proper environment for the Data Stewardship project to work on improving the RDM at the different faculties of the TU Delft.

5. Conclusions

In a machine-readable data driven era, RDM is becoming an increasingly important topic for researchers. Proper data management practices are not only beneficial for research, as it facilitates research and promotes verifiability

and transparency in the field. But it is also useful for researchers themselves, as it promotes effective research throughout their careers, and it makes it far easier for them to share data with others. In that sense, proper data management practices facilitates the path for Open Science and responsibly data sharing.

All such benefits are becoming quite clear to the community, to the point that researchers and research institutions/universities are becoming more aware about the need for further RDM support, in terms of both infrastructure and guidance.

The survey results presented in this work have shown two main things: 1) lack of awareness (and quite likely, understanding) about some RDM topics, such as data ownership and what 'FAIR data' implies; and 2) researchers show great interest about RDM. More experienced researchers appear to be more aware about funders' requirements such as DMPs and FAIR data principles, than the early career researchers. This can be explained by the fact senior researchers are the ones dealing with policies, regulations and mandates. However it is not clear whether 'awareness' in this case, directly implies 'understanding' or furthermore, actual adoption of such practices. The results also suggest that such high-level topics are not be necessarily communicated/disseminated to the research groups (more specifically, to the early career researchers).

Based on the findings of this survey, the Data Stewardship project at TU Delft has focused on understanding researchers' needs concerning data management, and spreading awareness about adequate RDM practices, and RDM services available for TU Delft researchers. We expect to carry out the survey at a periodic basis in order to also benchmark the evolution of the Data Stewardship project at a university level; and we encourage other institutions to reuse this survey and/or build upon it, to help evaluate RDM awareness at their own institutions/universities.

Acknowledgement

Special thanks to Munire van der Kruijk and all support staff at the TU Delft faculties and Library, who provided advice and help disseminating the survey at the faculties.

We thank the Editorial Board of LIBER Quarterly for their critical remarks about this template.

We also acknowledge the continuing support from the LIBER Executive Board, whose financial contribution allows us to publish this journal in Open Access.

References

- European Commission. (2017). *H2020 programme. Guidelines to the rules on open access to scientific publications and open access to research data in Horizon 2020*. Version 3.2. Retrieved March 21, 2017, from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.
- European Commission. (2018). *Implementation roadmap for the European open science cloud*. Retrieved March 14, 2018, from https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf#view=fit&pagemode=none.
- Johnson, R., Parsons, T., & Chiarelli, A. (2016). *Jisc Data Asset Framework Toolkit 2016*. Zenodo. <http://doi.org/10.5281/zenodo.177876>.
- Krause, J., Lambeng, N., Andrews, H., Boehmer, J., Cruz, M., van Dijck, J., Teperek, M. (2018). *Quantitative assessment of research data management practice (Version 3) [Data set]*. Zenodo <http://doi.org/10.5281/zenodo.1164397>.
- Nature research. (2016). *Data availability statements and data citations policy: guidance for authors*. Retrieved September 1, 2016, from <https://www.nature.com/documents/nr-data-availability-statements-data-citations.pdf>.
- NWO. (2016). *Open (FAIR) data*. Retrieved July 22, 2019, from <https://www.nwo.nl/en/policies/open+science/data+management>.
- PLOS. (2014). *Data availability*. Retrieved March 15, 2014, from <https://journals.plos.org/plosone/s/data-availability>.
- Rans, J., & Whyte, A. (2017). *Using RISE, the Research Infrastructure Self-Evaluation Framework*, v.1.1. Edinburgh: Digital Curation Centre. Retrieved July 22, 2019, from <http://www.dcc.ac.uk/resources/how-guides/RISE>.
- Regeerakkoord 2017 – 2021. (2017). *Vertrouwen in de toekomst*. VVS, CDA, D66 en ChristenUnie. Retrieved July 22, 2019, from <https://www.rijksoverheid.nl/regering/documenten/publicaties/2017/10/10/regeerakkoord-2017-vertrouwen-in-de-toekomst>.
- Teperek, M., Krause, J., Lambeng, N., Blumer, E., van Dijck, J., Eggermont, R., ... der Velden, Y.T. (2019). *Quantitative assessment of research data management practice*. Retrieved from <https://osf.io/mz3fx>.
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., ..., Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3, 160018 (n.p.). <https://doi.org/10.1038/sdata.2016.18>.

ZonMw. (2018). *ZonMw is changing to a new approach for data management*. Retrieved December 20, 2018, from https://www.zonmw.nl/fileadmin/zonmw/documenten/Toegang_tot_data/ZonMw_is_moving_to_a_new_DMP_policy_2019.pdf.

Notes

¹ Available at https://www.openscience.nl/files/openscience/2019-02/nationalplanopenscience_en.pdf.

² Netherlands Code of Conduct for Research Integrity, available at <https://www.vsnul.nl/files/documents/Netherlands%20Code%20of%20Conduct%20for%20Research%20Integrity%202018.pdf>.

³ <https://researchdata.4tu.nl/en/>.

⁴ Currently, the Data Seal of Approval (DSA) certification and the World Data System of the International Science Council (WDS) Regular Members certification have been replaced by the CoreTrustSeal Data Repository certification. See more information in <https://www.coretrustseal.org/about/>.

⁵ See <https://openworking.wordpress.com/2017/06/28/2017-self-assessment-of-research-data-services-and-4tu-centre-for-research-data-services-with-rise/>.

⁶ TU Delft Strategic Framework 2018-2024: what does it mean for Open Science? See <https://openworking.tudl.tudelft.nl/2018/02/04/tu-delft-strategic-framework-2018-2024-what-does-it-mean-for-open-science/>.

⁷ See <https://openworking.wordpress.com/2019/01/25/data-stewardship-at-tu-delft-2018-report/>.

⁸ See <https://public.tableau.com/profile/jasper.van.dijck#!/vizhome/20180809TUDelftResearchDataManagementSurvey2017-2018/TUDelftRDMsurvey2017-2018>.

⁹ See EPFL Research Data Management Survey Results in https://researchdata.epfl.ch/wp-content/uploads/2018/09/AbbreviatedReportRDSurvey_ENG_20180409_Final.pdf.

¹⁰ <https://datacarpentry.org/>.

¹¹ <https://software-carpentry.org/>.

¹² Surfdrive is a personal cloud-service available for researchers associated to Dutch universities. It offers staff and researchers 250 GB of storage space, with backup up synchronisation possibilities. See <https://www.surf.nl/en/store-and-share-your-files-securely-in-the-cloud-with-surfdrive>.