RESEARCH PAPER

Research Data Management for Master's Students: From Awareness to Action

Daen Adriaan Ben Smits and Marta Teperek
TU Delft, NL
Corresponding author: Daen Adriaan Ben Smits (daen.smits@outlook.com)

This article provides an analysis of how sixteen recently graduated master's students from the Netherlands perceive research data management. It is important to study the master's students' attitudes towards this, as students in this phase prepare themselves for their career. Some of them might become future academics or policymakers, thus, potentially, the future advocates of good data management and reproducible science.

In general, students were rather unsure what 'data management' meant and would often confuse it with data analysis, study design or methodology, or ethics and privacy. When students defined the concept, they focussed on privacy aspects. Concepts such as open data and the 'FAIR' principles were rarely mentioned, even though these are the cornerstones of contemporary data management efforts. In practice, the students managed their own data in an ad hoc way, and only a few of them worked with a clear data management plan. Illustrative of this is that half of the interviewees did not know where to find their data anymore. Furthermore, their study programmes had diverse approaches to data management education. Most of the classes offered were limited in scope. Nevertheless, the students seemed to be aware of the importance of data management and were willing to learn more about good data management practices.

This report helps to catch an important first glimpse of how master's students (from different scientific backgrounds) think about research data management. Only by knowing this, accurate measures can be taken to improve data management awareness and skills. The article also provides some useful recommendations on what such measures might be, and introduces some of the steps already taken by the Delft University of Technology (TU Delft).

I. Introduction

The adequate management of research data has always been an indispensable element of trustworthy scientific research, but the interest for research data management practices, skills and experiences has flourished in the last decade. This increased recognition for data management and the sense of urgency attached to it (Feijen 2011), is partly fuelled by questions about research reproducibility and the perceived existence of a reproducibility crisis in research. Monya Baker's survey of 1,500 researchers in Nature (2016), revealed that 90% of researchers felt that there was a reproducibility crisis (Baker 2016). It was found that, in some disciplines, over 80% of respondents experienced problems when reproducing other people's results. The survey also investigated the reasons behind such irreproducible research. In addition to a toxic publication culture ('selective reporting', 'pressure to publish' etc.), respondents also referred to (a lack of) data availability, indicating that 'methods and code were unavailable' and that 'the raw data from the original lab was not available'. The fact that poor data management practices contribute to irreproducible research, was re-emphasised in a recent study in Molecular Brain (Miyakawa 2020). The author noted that 'more than 97% of the 41 manuscripts did not present the raw data supporting their results, when requested by an editor'. To prevent such flaws and to increase the reproducibility of scientific studies, reliable data management throughout the entire research cycle is essential.

Not surprisingly, multiple stakeholders intervened with policies to improve research data management practices. National (e.g. the Dutch NWO (n.d.) and ZonMw (n.d.)) and international funders, for example, the European Commission (n.d.), require their grantees to submit a data management plan at the beginning of the research project. Moreover, grantees need to share research data, generated during the project, as openly as possible. In line with these measures, publishers also reacted with updated policies. *Science* was called to strengthen its data management policies for data reporting and publication (Science, n.d.) after a high profile paper retraction in May 2017 (Berg 2017). In a commentary on this matter, Roche (2017) wrote: 'The only computer containing the study's raw data was allegedly stolen and no backups existed on another machine or an online repository. Many are left wondering how this could happen in an era of cloud computing and open data'. Many other publishers, for example PLOS (Bloom, Ganley & Winker 2014) and Springer Nature (n.d.), are similarly committed to better data reporting. They increasingly require that the research data, which supports findings described in an academic article, is made available.

Governments also took the initiative to improve research data management practices at national levels. In the Netherlands, a revised Netherlands Code of Conduct for Research Integrity has been published in 2018. That research data and its management play an important role in the revised document, is illustrated by the fact that the code mentions the word 'data' 49 times. The code obliges researchers to 'manage the collected data carefully' and to 'contribute, where appropriate, towards making data findable, accessible, interoperable and reusable' (FAIR, KNAW et al. 2018: 17). Here, the code of conduct refers to the FAIR principles (Wilkinson et al. 2016), a concept that has been widely adopted by various stakeholders who are interested in good data management across Europe (European Commission 2018; Wittenburg et al. 2019), but also on other continents (Van Reisen et al. 2020; Sales et al. 2020). Hence, the new standard for researchers, according to the code of conduct, is to 'make research findings and research data public subsequent to the completion of the research' (KNAW et al. 2018: 18). However, institutions also have their 'duties of care' towards data management. It is made explicit that the institutions have to 'provide a research infrastructure', 'ensure that all data, software codes and research materials, published or unpublished, are managed and securely stored' and 'that, in accordance with the FAIR principles, data is open and accessible to the extent possible and remains confidential to the extent necessary' (KNAW et al. 2018: 20–21).

In parallel, research institutions have also been developing their own policies aimed at improving research data practices. For example, the Delft University of Technology (TU Delft) has a policy framework for research data management (Dunning, 2018). This document determines the roles and responsibilities of various stakeholders within the university. It addresses both central university support units (such as IT, legal, library) and introduces discipline-specific faculty-policies (e.g. TU Delft faculty of Mechanical, Maritime, and Materials Engineering 2019). Some smaller departments within TU Delft, also defined their own, even more specific policies on how to manage their research data assets (Akhmerov & Steele 2019). To further stimulate such initiatives on institutional level, the European Commission funded a project called the Leaders Activating Research Networks (2017), which defines a framework for a research data management policy for research performing institutions (LEARN, 2017).

The paragraphs above underline that research data management has attracted the attention of policymakers. However, these policies are only effective when they stay close to the daily research practice (Cruz et al. 2019) and adhere to the researcher's perceptions of good research data management. When the policies are not truly embedded and do not appeal to the intrinsic motivation of researchers, behavioural change at a university level might be difficult to achieve (Carlson & Stowell-Bracke 2013). Maria Cruz (2019) highlighted that this gap between policy and practice needs to be addressed, before good research data management can be properly implemented across various research domains. Feijen (2011: 28) agrees: 'There is a growing awareness that there should be more focus [on research data management], but it seems that this awareness is stronger in circles *outside* the research groups'.

Illustrative in this regard is that researchers are, in contrast to policymakers in this field, not very aware of the FAIR data principles. This is visible on smaller and larger scales. A survey from TU Delft, conducted in 2017 and 2018 (almost 700 respondents), revealed that researchers were largely unaware of these FAIR principles. Moreover, more than 50% of the respondents at each faculty were not 'aware' or were 'not sure' of the funders expectations related to FAIR data (Mancilla et al. 2019). In the State of Open Data Report, Fane et al. (2019) found something similar, as they reported that more than 50% of the 8,500 researchers that responded to their questionnaire never heard of the FAIR principles.

Interestingly, most research on research data management practices tends to focus on practices of research staff (including PhD candidates). The experiences of master's students are often overlooked. The lack of particular attention for master's students is at least partially explained by the fact that, most of the time, master's students are not the leading authors on research papers and grant proposals. In addition, the way

intellectual property rights apply to master's students is complex. For example, master's students in the Netherlands retain ownership of the research data and any intellectual property they create. Therefore, it is unclear to what extent the research data management policies of universities, publishers or funders actually apply to them. Despite all this, data literacy is very important for this group of students. As Carlson et al. (2011: 629) noted: 'Researchers increasingly need to integrate the disposition, management and curation of their data into their current workflows, but it is not yet clear to what extent faculty and students are sufficiently prepared to take on these responsibilities'. This causes an interesting friction, because it was learned that most data management-related curricula are not openly accessible and are not targeted on students outside of information science programs (Piorun et al. 2012: 47).

In this light, it is interesting to gain insight into how university master's students perceive research data management. They form an important group to study, as they conduct their first major research project but do not receive the same amount of research data training as PhD students often do. Moreover, the master's students might be the future academics or policymakers and, thus, potentially, the future advocates of good research data management and reproducible science. Carlson & Stowell-Bracke (2013: 5), who interviewed master's students in a water field station, argue that: 'A significant gap in efforts to understand the practices of researchers through case studies, surveys or other means of investigation, is the overall lack of attention given to the role of graduate students and their work in generating, processing, analysing and managing data'.

This exploratory study addresses the gap in the understanding of data management perceptions and practices by master's students. It presents the qualitative outcome of sixteen interviews with master student's in the Netherlands. In the next section, the methodology of the study is explained. The findings will be discussed in section III. In the discussion (section IV), suggestions for follow up actions are included, together with some concrete steps already undertaken by TU Delft.

II. Methodology

To learn more about the data management attitudes and perceptions of master's students, sixteen semi-structured interviews were conducted in September and October 2019. All interviewees have obtained a master's diploma (all after 2015) from a Dutch university. In the Netherlands, master's programmes have a minimum length of one year. For admission to a master's programme, a bachelor's degree (or a recognised equivalent) is mandatory (NUFFIC, 2018). Most disciplines offer one-year master programmes, but two-year and three-year programmes are common in applied disciplines, such as physics, applied sciences or medicine. Master's programmes offer the students further academic specialisation. In a master's curriculum, a lot of emphasis is put on the education and the acquiring of academic research skills. In the vast majority of programmes, the completion of a master thesis research is obligatory (NUFFIC, 2018), which generally comprises at least 25% of the length of the programme. All interviewees included in this study have performed such a master's thesis research.

To invite the interviewees, the purposive sampling method was used, meaning that the researcher used own judgments to select the participants. According to Tongco (2007: 147), 'the purposive sampling technique, also called judgment sampling, is the deliberate choice of an informant due to the qualities the informant possesses'. In this case, the informants are all persons with a master's diploma and they are member of a strategic alliance of three universities in the Netherlands. The interviewees had diverse study backgrounds, although most of them studied social sciences. Interviewing students from different disciplines and different universities allows for diverse views and experiences to be included. The results of these interviews help to catch an important first glimpse of how diversely educated master's students think about research data management. The participant's university backgrounds (master's studies only) and their fields of study are included in a table in the supplementary material of this article.

To select the interviewees, twenty ex-students were emailed in advance (in early September 2019), and were asked if they wanted to participate. Four participants decided not to do so, or did not reply to the email. All others consented via email to participate in the study. Before starting the interviews, the ex-students were informed about the procedure and they were orally asked for their consent again. They were told that the interviews were not recorded, but that detailed notes were taken. Recording was considered but was eventually decided against. It was believed that this approach would lead to better exploratory information on the students' real attitude towards data management and that this would ensure that the interviewees felt more comfortable and at ease when answering personal questions about their own data management. All participants gave the researcher permission to use direct quotes in the publication, given that all interview notes would be anonymised. The interviews were conducted in Dutch, so the notes were also taken in Dutch. The open and axial codes (method will be explained in the next paragraph) in the data file and the

direct quotes used in the report were translated into English. At TU Delft, where this study was conducted, MSc procedures regarding ethics are overseen by the faculties/units where the research is conducted. Only if these units deem it appropriate, applications are made to the central TU Delft Human Research Ethics Committee. In this case, following a discussion with the traineeship supervisor, the supervisor's decision was that the ethical risks to the study participants are negligible, therefore a formal application to the TU Delft Human Research Ethics Committee was deemed not essential.

The participants all answered the same eleven questions (see **Table 1** below) about research data management, in an interview that took between thirty and forty-five minutes. The interview questions were divided in three categories: a) the participant's experiences with data management, b) attention for data management during the study curriculum and c) data in today's world.

To analyse the interviews, open and axial coding methods were used. Coding helps to distinguish patterns, repeated concepts and categories in interview data (Given 2008). Open coding is the act of closely interpreting interview data, in order to summarize the main idea of the text and to be able to make a first selection of concepts (Given 2008). In the axial coding phase, the data is structured in such a way that it becomes possible to make theoretical connections between categories, questions, answers and concepts (Kolb 2012). Due to the semi-structured interviews, the axial coding is very valuable to distinguish patterns and concepts across interview questions and participants. The full interview notes, the open coding (with English translations) and the axial coding tags are openly accessible and downloadable from the 4TU.ResearchData repository (Smits, 2020). The full supplementary material includes tables with the participants' study backgrounds, the (anonymised) distribution of universities, the translations of the full quotes used in this article and an extensive table that shows how many times an axial code was applied by the researcher. For privacy reasons, the question about the interviewees' studies and thesis topics was left out.

Limitations of the study

This study was carried out to gain a better understanding of how graduated master's students in the Netherlands perceive research data management and how they put data management into practice during their master's studies. The interviewees studied at different universities and within different fields. Due to the exploratory nature of the study, the semi-structured interview approach and the fact that the students have shared their own opinions, the results should be interpreted with this specific context in mind and care should be taken with any potential generalisations. Nevertheless, the results outline how diversely educated master's students think about research data management. Only by knowing (and further exploring) this, accurate measures can be taken to improve data management awareness and skills. For a more university-specific view, it is important to carry out this study within a more focused group.

Table 1: Interview questions asked to participants.

A)	Experiences with data management
1	Can you tell me more about your master thesis?
2	What do you think 'data management' means (in the context of scientific research)?
3	What kind of data did you collect?
4	How did you manage your data during your master thesis?
5	Why do you think data management is/is not important?
B)	**Attention for data management during the study curriculum**
6	What would be/are the biggest challenges to improve your data management skills (think about GDPR, digital literacy, coding skills, computer skills, ethics, lack of technical skills, human errors)?
7	Can you tell me more about how you learnt about data management? What did you do well/not well?
8	What tools would you need to become (even) better in data management? Was there enough attention for data management during your studies?
9	What would you do differently if you had the chance to do your research again?
C)	**Data in today's world**
10	What do you think the role of data management is in our current society?
11	How do you see the role of data in society?

III. The interview findings

In this section, the findings of the study are presented. In their interviews, the students defined data management, they talked about their experiences, their attitudes and how they learnt (or not) about data management during their curricula. The participants also indicated what they needed to become better in data management and how they thought about 'data' in the current day and age.

All the interviewed students have dealt with data during their research, whether quantitative or qualitative in nature. Five of them indicated to have used only qualitative methods, nine of them solely utilised quantitative methods. Two of the participants mixed these approaches. In their research, the students processed (pre-existing) datasets in SPSS,[1] collected longitudinal patient data, held interviews, analysed documents or conducted surveys. The findings, and their possible relation to the data management literature identified in the introduction, have been grouped under the five main subheadings to come.

Confusion about the definition of 'research data management'

Before diving into more specific questions about research data management, the students were asked to define the concept. The Technical University Eindhoven (TU/e, n.d.), uses the following definition:

> 'Research data management (RDM) is the careful handling and organization of research data during the entire research cycle, with the aim of making the research process as efficient as possible and to facilitate cooperation with others. More specifically, RDM helps to protect data, it facilitates in sharing the data with others and it ensures that research data is findable, accessible and (re)usable'.

Even though the participants, right at the start of the interviews, did not give such an extensive and comprehensive answer, some of the aforementioned factors surfaced in their definitions too. One participant, who studied languages, answered: 'I believe that data is just a modern word for information. Everything around you can be data. So data management is the processing of all the information that you find relevant for a certain purpose'. All the interviewees actually believed that data management describes how researchers handle and safely store their data, during but also after the research. The 'FAIR' principles, which some members of the data management community see as the cornerstone of data management, were only mentioned by one interviewee. This is in line with the earlier reported findings by the State of Open Data Report (Fane et al. 2019) and Mancilla et al. (2019). Only one interviewee, who studied educational sciences, said that data sharing was an inherent part of the definition of good data management: 'It would be very valuable if data is stored in a standardised manner, so that it is very easy to combine and compare different datasets. Also, it is important to make the data visible to the outside world'.

That said, many interviewees also confused data management with other aspects of good research practice, such as methodology or study design. For example, a psychology student commented:

> 'My data and the management of it was a mess. The survey questions did not correlate well with one another, and as too many participants did not answer to a question, or filled in 'not applicable', I got into trouble with my sample. My sample was too small so it could not lead to a significant test. I should have asked better questions, on which the answer 'not applicable' was not a possibility. After all these problems, I asked for external help and I wrote a longer discussion because I had so little data'.

Other students, when asked about their data management experiences, often spoke about being unsure how to code data, what kind of statistical test to run, how to hold scientific interviews or when to use consent forms.

Students are aware of privacy issues associated with data processing

When the students recalled their data management experiences, they often referred to privacy issues. In fact, twelve students predominantly focussed on the privacy concerns that came along with their data collection. Privacy-related issues have been addressed fifty-eight times throughout all these interviews about data management, by fifteen different interviewees. Only one interviewee did not explicitly mention privacy

---

[1] SPSS (Statistical Package for the Social Sciences) is a computer programme for statistical analyses, particularly used in social sciences.

aspects. In general, students asked for consent, they anonymised their data and destroyed it when the sensitive information was no longer essential. Thus, the students who worked with personal data were very conscious about its sensitivity and the importance of anonymising what they found. The overwhelming attention for privacy is not surprising, due to the relatively recent introduction of the GDPR[2] and the attention this has generated.

The students also came back to these privacy concerns when answering questions about how they felt about the role of data in today's world. Eleven students were concerned about the data gathering practices of both technology companies and the government. They were unified in the opinion that big data and privacy play an important role in our society. Generally, however, the students' opinions on data collection seemed to depend on instinctive feelings, rather than on facts. One interviewee used the analogy that your personal data belongs to you as much as your hair or DNA, so you should have the right to sell it yourself if you wish. Five students realised that big data collection also offers chances to solve societal problems and improve services and technologies. Moreover, four students believed that collecting (personal) data for the sake of collecting data had to be minimised. This is where the link to their own research data came in, because two students also collected more data than they actually needed. While doing surveys, they collected all the data that could potentially be of relevance to their research questions, without knowing whether they would actually need these additional data.

An ad hoc approach to data management

When facing problems or challenges with their data, the students dealt with them intuitively, whether or not with the help of their thesis supervisor. Most of the students (13/16) could not rely on a data management plan, as only three interviewees within the sample mentioned that they developed such a plan. In fact, nine diversely educated interviewees were explicit that they managed their data in an ad hoc way. The students also did not always foresee the implications of the data choices they made in the beginning of their project.

One psychology student reflected on the lack of proper documentation during the data processing: 'The data that we stored in the digital environment did not show up as we wanted in SPSS. We used certain formulas which changed something in the data, but we could not find the error. We knew it was not something substantial to our dataset, so we continued with the mistake still in there'. A sociology student also mentioned that she never knew how reliable the data she used actually were: 'One researcher gave me the results of a quantitative study related to my topic. I used these results, which were already processed and analysed in SPSS, without questioning anything. I have never seen the raw dataset. Who knows these data were actually reliable?'

Illustrative of this ad-hoc approach to data management is the fact that half (8/16) of the 16 interviewees did not know how to access their data anymore. This lack of raw data availability makes their studies irreproducible, a flaw that Miyakawa (2020) found among researchers too. When thinking about their own research, the sixteen participants did not see their data as a research output on its own, or a valuable set of information that could help science forward. The students did not directly see the point of publishing their own data to provide evidence for the work they had done. Carlson & Stowell-Bracke (2013: 19) also underlined this, as in their sample 'none of the students had really given the long term maintenance of their digital datasets much thought or taken action to ensure long term access to their data'. The interviewed students were only concerned about finishing their thesis and the dataset was merely an instrument to achieve this.

One of the students stated: 'I did not have a plan what to do with my data. Of course I anonymised my interviews, but I have no clue what I did with the transcripts. I also don't know where the data are anymore. You use the data for your analysis, but that's it. Then it does not feel like your own problem anymore. Nobody is interested in the process after publishing the report'. In addition, a pedagogy student reflected: 'It was clear that fast graduation was a big concern for our study. My master thesis felt as an obligatory goal to reach.'

Feijen (2011: 27) found a similar pattern. He wrote: 'In most cases, data from the previous project will stay where it was at the end of the research project – where, more often than not, the storage situation is unreliable and the data is likely to deteriorate over time. Although all who were involved know that data will probably be lost forever, there is no time to take protective measures. Researchers have the feeling that they are not in a position to solve this problem, nor do they tend to accept responsibility for it. Their willingness to take responsibility is not highly developed'.

---

[2] The General Data Protection Regulation came into effect as of May 2018. For more information please visit the website of the European Commission: https://ec.europa.eu.

Students are aware of the importance of data sharing for research reproducibility, but they do not publish their data

There was unanimity among the students that good data management during research is important for research reproducibility. Eleven participants stressed that good data management and sharing is essential to reproduce studies. However, they did not always translate these overall principles into practical actions. A languages student said: 'I was quite aware of the importance of reproducibility. However, during the coding of my data, I realised that nobody would actually be interested in my data ever again. As a result, I did not manage my data with utmost care, which could have harmed the credibility of my research'.

None of the interviewed students published their data. Two interviewees suggested the creation of a data archive for master's students, to facilitate data sharing and knowledge flow.[3] One of the international relations students stated:

> 'My data was not published anywhere. For an outsider, it is impossible to retrieve my data. I have the feeling that many other students did research on similar issues, so many fished in the same pool of information. But none of this data was findable for me to work with. That is a missed chance. It would be good to have a data archive in which master's students can store their data.'

When discussing data sharing, similarly to the study conducted by Carlson & Stowell-Bracke (2013), some interviewees underlined the importance of openly accessible research data, while others pointed out the potential risks. They thought that open data can foster new research, help increase transparency and improve the reliability of scientific claims. One student stated, for instance, that when certain issues get sensationalised in the media, the underlying dataset helps to retrieve the exact context in which things happened, so that a more balanced picture could be put forward.

A sociology student also mentioned the positive role open data could play in the competitive research climate: 'It is good to be more transparent about research data. Research is also a lot about raising funds and getting grants. That is why I think data checks are so important. I think the market has a bad influence on research'. The moral duty to openly publish data also plays a role. An art history student reflected: 'why does society need to pay twice, by subsidising the research first and then paying again for access to the results?'

Students also discussed the risks associated with data sharing. Some were concerned about the abuse of sensitive data, or that others might not be skilled enough to interpret the data in the way it was intended, leading to unreliable new results. There was a feeling that caution was needed when interpreting data, because it could be hard to reproduce the exact same circumstances as when the data were gathered, even when rich metadata is available. Some also stressed that the data re-users could have a different social background than the creator of the content, leading to other assumptions or beliefs. One student, who studied educational science, reflected on her own data: 'it is crucial to know what your data really means. After some time, I realised that my quantitative data had to be interpreted in a different way than I thought. If I wouldn't have found this, I would have come to the wrong conclusions'.

Gaps in data management training for Master's students

The students also answered questions on whether there was enough attention for data management during their curriculum, what their biggest challenges were and what they needed to become better at data management. In essence, students did not receive dedicated training on data management. Various studies had elements of data management incorporated into the curricula, but the focus was typically on other topics, such as ethics, statistical analysis or research methods. Sometimes, data documentation or safe data handling was discussed in thesis groups, during seminars or in specific classes.

A comment from a pedagogy student nicely illustrates the issue that data management was confused with other topics, as she mentioned that 'data management education was all about analysing in SPSS.' During the interviews, thirteen out of the sixteen interviewees declared that their study did not have enough attention for data management in general, and that they wanted to learn more about the subject. A similar observation was made in the study by Piorun et al. (2012), in which their students also confirmed their demand for data management education. One student, who studied an international relations specialisation, said: 'I wish there was more attention for data management. It was all quite unclear to us and that caused a lot of

---

[3] The 4TU Centre for Research Data actually hosts a section in which also master's students can submit their data. For more information, please visit: https://data.4tu.nl/repository/collection:masterthesis_data.

unrest and confusion. It would have been good if the instructors would have shown us why and how to do thorough data management. Now I just kept doing my own thing'.

Remarkably, there was a media studies student who questioned the increased attention for data management in science and education: 'Data management is a trend. If it was so important, why wasn't it such a hot topic earlier? It seems that universities now put so much effort into it, only to ensure that they are not responsible anymore when something goes wrong. Then the university is innocent because they have taught us about the risks. I don't think that data were not well managed before the attention increased, it was just more inherent to the research and we trusted more on common sense'.

IV. Discussion and final remarks

This preliminary study captured the attitudes of sixteen master's students towards data management. Overall, the results suggest that, with the exception of awareness of data privacy issues and GDPR, the students had a rather fragmented knowledge of data management. Interestingly, many of them were confused about what data management meant, and seemed to associate data management issues with other research topics, such as data analysis, methodology and study design. Consequently, most of the students managed their data in an ad hoc fashion, without any dedicated planning upfront. This intuitive approach also surfaced in their attitudes towards data sharing. While students were aware that data availability is essential for reproducibility, none of them shared their own research data, knew how to do it, or felt this was important for their study. Given the relatively low awareness about data management among the students, it was not surprising to note that none of them has received comprehensive training on this matter. Data management education, if any, seemed to be added to existing courses and study discussions, but without a coherent approach. Nevertheless, almost all students understood the importance of data management and wished they had received better data management training.

The results presented in this exploratory study suggest that academic institutions could invest more resources into the data management education of master's students. While approaches to data management can differ in different research fields, a comprehensive overview of research data management is needed for master's students, regardless of scientific discipline. This is particularly important given that, as indicated in this study, students tend to process sensitive data that needs to be managed responsibly. Good data management and data sharing could also enhance the visibility of the students' work, improve the rigour of the research and improve the overall transparency.

Even though these findings are exploratory in nature, they already prompted TU Delft to embark on several initiatives aimed at improving data management awareness among master's students. At the Faculty of Architecture and the Built Environment, the faculty data steward is now running a pilot to provide data management education in one of the master's courses. The data steward teaches the students about data management, just before the students start their thesis projects. The advantage of partnering with an experienced faculty data steward guarantees that all key aspects of data management education are addressed in a coherent fashion. The results of this pilot are preliminary – the data steward only joined one course so far – but the feedback received from the students and the course coordinator was positive. The data steward was asked to regularly present to the students participating in this course. This faculty data steward is now also initiating discussions with the coordinators of other MSc courses. Pending the outcomes of this pilot at the Faculty of Architecture and the Built Environment, similar approaches might be adapted at other faculties at TU Delft.

The DelftOpenHardware community[4] is also a promising development in this field. This community is a bottom-up, community driven initiative by TU Delft researchers, to encourage collaboration on Open Hardware projects. Hardware and design projects are quite common at TU Delft (e.g. the designs of new machines, equipment, tools etc.), especially among master's students who conduct such efforts for their thesis projects. One of the core missions of the DelftOpenHardware community is to teach good documentation and to promote the sharing of data in hardware designs. The majority of its members are master's students, who value the informal support they receive through the DelftOpenHardware community. The community is meeting every week and students regularly join the drop-in sessions, to receive support on data management and documentation. Data management help is offered directly by experts in the field, while administrative support (such as venue booking) is offered by library staff. That way, students receive very practical, hands-on data management support in their projects.

---

[4] For more information, please visit: https://delftopenhardware.nl/

Finally, this study also touched upon the importance of strict privacy measures when processing sensitive and personal research data. At TU Delft, the compliance of research data with GDPR is achieved by following a dedicated workflow, which starts with a data management plan (DMP). Every project that processes personal information needs to have a DMP (TU Delft, n.d.). These DMPs are created in a dedicated tool, called DMPonline.[5] Whenever a new DMP is created, the faculty data steward is notified about it and then reviews the plan. Following the review, the data steward advises the researcher on appropriate data management steps, such as an ethical review or a data protection impact assessment. So far, however, the focus of this service was on researchers and PhD candidates. The outcomes of this study highlight that the data processing of master's students also needs to be addressed, in particular because most master's students have been following very ad hoc data management procedures. However, given the sheer number of master's students, it is impossible to ask all master's students to follow the aforementioned workflow. That is why the 'GDPR Research Data Working Group' at TU Delft is currently conducting community consultations to explore the best possible solution for master's students.

Overall, the findings of this exploratory study provided important insights on data management practices among master's students. It highlighted that data management awareness among master's students is rather low. Therefore, research institutions need to invest in more thorough education on this matter. TU Delft has already undertaken some preliminary steps in the right direction, but more work needs to be done, including extended research on this topic.

Supplementary material
All research material is openly available and accessible for any interested reader. The supplementary material includes tables with the participants' study backgrounds, the involved universities, the translations of the full quotes used in this article and an extensive table that indicates how many times a certain code, to label the interview answers, was applied by the researchers. An explanation on how to interpret the coding is also included in the supplementary material. Please click this link to access the full dataset: http://doi.org/10.4121/uuid:ee978f4b-4b2a-4fb1-aeed-829f773eb316.

Acknowledgements

Competing Interests
The authors have no competing interests to declare.

References
Akhmerov, A and **Steele, G.** 2019. TU Delft Open Data Policy of the Quantum Nanoscience Department. *TU Delft*. DOI: https://doi.org/10.5281/zenodo.2556949

Baker, M. 2016. 1,500 Scientists Lift the Lid on Reproducibility, 25 May 2016. *Nature News*. [Last accessed 30 April 2020]. DOI: https://doi.org/10.1038/533452a

Berg, J. 2017. Editorial Retraction. *Science*, 356(6340): 812. DOI: https://doi.org/10.1126/science.aan5763

Bloom, T, Ganley, E and **Winker, M.** 2014. Data Access for the Open Access Literature: PLOS's Data Policy. *PLOS Biology*, 12(2): 1–3. DOI: https://doi.org/10.1371/journal.pbio.1001797

Carlson, J, Fosmire, M, Miller, CC and **Nelson, MS.** 2011. Determining Data Information Literacy Needs: A Study of Students and Research Faculty. *Portal: Libraries and the Academy*, 11(2): 629–657. DOI: https://doi.org/10.1353/pla.2011.0022

Carlson, J and **Stowell-Bracke, M.** 2013. Data Management and Sharing from the Perspective of Graduate Students: An Examination of the Culture and Practice at the Water Quality Field Station. *Portal: Libraries and the Academy*, 13(4): 343–361. DOI: https://doi.org/10.1353/pla.2013.0034

Cruz, M. 2019. Bringing Researchers Along on the Road to FAIR data. *Zenodo*. DOI: https://doi.org/10.5281/zenodo.3249802

Cruz, M, Dintzner, N, Dunning, A, Kuil, A. van der, Plomp, E, Teperek, M, der Velden, YT and **Versteeg, A.** 2019. Policy Needs to Go Hand in Hand with Practice: The Learning and Listening Approach to Data Management. *Data Science Journal*, 18(1): 45. DOI: https://doi.org/10.5334/dsj-2019-045

---

[5] The DMP online portal is accessible through https://dmponline.tudelft.nl/?perform_check=false.

Dunning, A. 2018. TU Delft Research Data Framework Policy. *TU Delft*. Available at: https://www.tudelft.nl/en/library/current-topics/research-data-management/r/policies/tu-delft-faculty-policies/. [Last accessed 1 May 2020].

European Commission. 2018. Turning FAIR into Reality. Brussels, Belgium: European Commission. Available at: http://op.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF [Last accessed 30 April 2020].

European Commission. n.d. Data management – H2020 Online Manual. Belgium: European Commission. Available at: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm [Last accessed 30 April 2020].

Fane, B, Ayris, P, Hahnel, M, Hrynaszkiewicz, I, Baynes, G and **Farrell, E.** 2019. The State of Open Data Report 2019. *Digital Science*. DOI: https://doi.org/10.6084/m9.figshare.9980783.v2

Feijen, M. 2011. What Researchers Want. *SURF Foundation*. Available at: https://www.bvekennis.nl/wp-content/uploads/documents/11-0198-What-researchers-want.pdf [Last accessed 30 April 2020].

Given, LM. 2008. The SAGE Encyclopedia of Qualitative Research Methods. Thousand Oaks: SAGE Publications. DOI: https://doi.org/10.4135/9781412963909

KNAW; NFU; NWO; TO2-federatie; Vereniging Hogescholen; VSNU. 2018. Wetenschappelijke Gedragscode Integriteit [Netherlands Code of Conduct for Research Integrity]. The Hague, Netherlands. Available at: https://www.vsnu.nl/files/documents/Netherlands%20Code%20of%20Conduct%20for%20Research%20Integrity%202018.pdf [Last accessed 1 May 2020].

Kolb, SM. 2012. Grounded Theory and the Constant Comparative Method: Valid Research Strategies for Educators. *Journal of Emerging Trends in Educational Research and Policy Studies*, 3(1): 83–86. Available at: http://jeteraps.scholarlinkresearch.com/articles/Grounded%20Theory%20and%20the%20Constant%20Comparative%20Method.pdf [Last accessed 1 May 2020].

Leaders Activating Research Networks. 2017. Model Policy for Research Data Management (RDM) at Research Institutions/Institutes: 133–136. Available at: https://discovery.ucl.ac.uk/id/eprint/1546606/1/25_Learn_Model%20Policy_133-136.pdf [Last accessed 1 May 2020].

Mancilla, HA, Teperek, M, van Dijck, J, den Heijer, K, Eggermont, R, Plomp, E, der Velden, YT and **Kurapati, S.** 2019. On a Quest for Cultural Change – Surveying Research Data Management Practices at Delft University of Technology. *LIBER Quarterly*, 29(1): 1–27. DOI: https://doi.org/10.18352/lq.10287

Miyakawa, T. 2020. No Raw Data, No Science: Another Possible Source of the Reproducibility Crisis. *Molecular Brain*, 13(24). DOI: https://doi.org/10.1186/s13041-020-0552-2

NUFFIC. 2018. Onderwijssysteem Nederland [Education system in The Netherlands]. The Hague, Netherlands: Nuffic. Available at: https://www.nuffic.nl/publicaties/onderwijssysteem-nederland/ [Last accessed 1 May 2020].

NWO. n.d. Open (FAIR) data. The Hague, The Netherlands: NWO. Available at: https://www.nwo.nl/en/policies/open+science/data+management [Last accessed 1 May 2020].

Piorun, M, Kafel, D, Leger-Hornby, T, Najafi, S, Martin, E, Colombo, P and **LaPelle, N.** 2012. Teaching Research Data Management: An Undergraduate/Graduate Curriculum. *JESLIB*, 1(1): 46–50. DOI: https://doi.org/10.7191/jeslib.2012.1003

Roche, DG. 2017. Evaluating Science's Open-data Policy. *Science*, 357(6352): 654. DOI: https://doi.org/10.1126/science.aan8158

Sales, L, Henning, P, Veiga, V, Costa, MM, Sayão, LF, da Silva Santos, LOB and **Pires, LF.** 2020. GO FAIR Brazil: A Challenge for Brazilian Data Science. *Data Intelligence*, 2(1): 238–245. DOI: https://doi.org/10.1162/dint_a_00046

Science. n.d. Science Journals: editorial policies. Washington, DC, United States of America: Science. Available at: https://www.sciencemag.org/authors/science-journals-editorial-policies [Last accessed 1 May 2020].

Smits, DAB. 2020. Research Data Attitudes of Recently Graduated Master Students (interview notes and code) [Dataset]. DOI: http://doi.org/10.4121/uuid:ee978f4b-4b2a-4fb1-aeed-829f773eb316

Springer Nature. n.d. Research Data Policies. Berlin, Germany: Springer Nature. Available at: https://www.springernature.com/gp/authors/research-data-policy [Last accessed 1 May 2020].

Technische Universiteit Eindhoven. n.d. Wat is Research Data Management? Eindhoven, The Netherlands: TUE. Available at: https://www.tue.nl/universiteit/bibliotheek/ondersteuning-onderwijs-onderzoek/wetenschappelijk-publiceren/data-coach/begrippen-en-achtergrond/wat-is-research-data-management/ [Last accessed 1 May 2020].

Tongco, M. 2007. Purposive Sampling as a Tool for Informant Selection. *Ethnobotany Research and Applications*, 5(1): 147–158. Available at: Ethnobotany Research Journal [Last accessed 1 May 2020]. DOI: https://doi.org/10.17348/era.5.0.147-158

TU Delft. n.d. Personal Data: Personal Research Data Workflow. Delft, The Netherlands: TUD. Available at: https://www.tudelft.nl/en/library/current-topics/research-data-management/r/manage/confidential-data/personal-data/ [Last accessed 1 May 2020].

TU Delft Faculty of Mechanical Maritime and Materials Engineering. 2019. The Faculty of Mechanical, Maritime and Materials Engineering Research Data Management Policy. Delft, The Netherlands: TU Delft. DOI: https://doi.org/10.5281/zenodo.3524106

Van Reisen, M, Stokmans, M, Mawere, M, Basajja, M, Ong'ayo, AO, Nakazibwe, P, Kirkpatrick, C and **Chindoza, K.** 2020. FAIR Practices in Africa. *Data Intelligence*, 2(1): 246–256. DOI: https://doi.org/10.1162/dint_a_00047

Wilkinson, MD, Dumontier, M, Aalbersberg, IJ, Appleton, G, Axton, M, Baak, A, Bouwman, J, et al. 2016. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data*, 3(160018). DOI: https://doi.org/10.1038/sdata.2016.18

Wittenburg, P, Lautenschlager, M, Thiemann, H, Baldauf, C and **Trilsbeek, P.** 2019. FAIR Practices in Europe. *Data Intelligence*, 2(1): 257–263. DOI: https://doi.org/10.1162/dint_a_00048

ZonMw. n.d. ZonMw-procedure Datamanagement. The Hague, The Netherlands: ZonMW. Available at: https://www.zonmw.nl/nl/over-zonmw/toegang-tot-data/zonmw-procedure-datamanagement/ [Last accessed 1 May 2020].