Cultural obstacles to research data management and sharing at TU Delft

Plomp, Esther; Dintzner, Nicolas; Teperek, Marta; Dunning, Alastair

DOI
10.1629/uksg.484

Publication date
2019

Document Version
Final published version

Published in
Insights: the UKSG Journal

Citation (APA)
https://doi.org/10.1629/uksg.484

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.
Cultural obstacles to research data management and sharing at TU Delft

Research data management (RDM) is increasingly important in scholarship. Many researchers are, however, unaware of the benefits of good RDM and unsure about the practical steps they can take to improve their RDM practices. Delft University of Technology (TU Delft) addresses this cultural barrier by appointing Data Stewards at every faculty. By providing expert advice and increasing awareness, the Data Stewardship project focuses on incremental improvements in current data and software management and sharing practices. This cultural change is accelerated by the Data Champions who share best practices in data management with their peers. The Data Stewards and Data Champions build a community that allows a discipline-specific approach to RDM. Nevertheless, cultural change also requires appropriate rewards and incentives. While local initiatives are important, and we discuss several examples in this paper, systemic changes to the academic rewards system are needed. This will require collaborative efforts of a broad coalition of stakeholders and we will mention several such initiatives. This article demonstrates that community building is essential in changing the code and data management culture at TU Delft.

Keywords
Data; software; support; RDM; data stewardship; TU Delft

Introduction
Recent reports of a reproducibility crisis in science led to increased demands for transparency in research practices and open data. Before data sets and related code can be opened, they will have to be appropriately managed and stored to be reusable for other researchers. This process involves research data management (RDM): implementing standard practices for accurate data collection and processing, documentation and analysis. RDM practices improve the reusability of data sets, as well as increase the efficiency, transparency and reproducibility of research. While RDM is beneficial to the scientific process as well as to individual researchers, it can be difficult for researchers to know how to improve their data and code management. Several surveys asking researchers about...
barriers to data management and sharing indicated that the main obstacles are cultural, not technological. Obstacles include limited encouragement of data management and sharing within a research field, a preference of researchers to share data upon request, the perception of data management and sharing as time consuming, and the lack of training available to improve these practices. These cultural barriers are not easy to breach as cultural change is a slow process, particularly in academia.

A key part of Delft University of Technology’s (TU Delft) approach to this cultural change is its Data Stewardship project. The Data Stewardship project focuses on incremental improvements in researchers’ data management and sharing practices, by increasing awareness and providing support to researchers.

RDM: the Case of TU Delft

TU Delft is the largest technical university in the Netherlands, with ~5000 employees (including PhD students), ~23,500 students and eight separate faculties. Quite like at other universities, researchers struggle to improve their data management and sharing practices due to a lack of resources and time, expertise and incentives. However, there are also specific characteristics to TU Delft that influence the research data ecosystem. The focus on technical subjects means large quantities of numerical data are gathered from both physical experiments and computer-based simulations, and the development of dedicated software tools to process these data is very common. In addition, researchers are often engaged with industry collaborations and partnerships with governmental institutions. Confidential data from such projects, either commercially sensitive data or personally identifiable information, presents additional challenges to data management and sharing.

Perhaps because of the technical nature of the university, a comprehensive technical infrastructure is in place to help with RDM practices. A variety of secure storage solutions for managing data during the project are offered. TU Delft also utilizes the data storage and sharing services of SURF, the collaborative organization for Information and Communication Technology (ICT) in Dutch education and research. Researchers are furthermore supported by TU Delft Library, that hosts DMPonline, an online platform to create data management plans (DMPs) and provides templates for DMPs. TU Delft also hosts 4TU.Centre for Research Data (or 4TU.ResearchData), a certified (Data Seal of Approval) archive for long-term preservation and sharing of research data. Finally, researchers can make use of dedicated funds from the Library to prepare their data for deposit at 4TU.ResearchData.

Despite the availability of these services, a survey conducted at six out of eight faculties of TU Delft in 2017–2018 (the remaining two Faculties did not have a Data Steward at the time that the survey was conducted) indicated that data management practices could still benefit from improvements. For example, only around 40% of the 628 respondents backed up their data automatically (Figure 1). This was striking, given that all data storage solutions offered by TU Delft ICT and SURF come with automated back-up. The majority of the researchers (between 42–61% across the six faculties) were aware of data repositories but indicated that they did not use them. Similarly, responses to open questions indicated a lack of awareness of the facilities in place, for example:

‘People don’t tell us anything, we don’t know the options, we just do it ourselves.’

‘I think data management support, if it exists, is not well known among the researchers.’

‘reports of a reproducibility crisis in science led to increased demands for transparency in research practices and open data’

‘only around 40% of the 628 respondents backed up their data automatically’

‘The majority of the researchers ... were aware of data repositories but ... did not use them’
'I think I miss out on a lot of possibilities within the university that I have not heard of. There is too much sparsely distributed information available and one needs to search for highly specific terminology to find manuals.'

Figure 1. Responses regarding automatic back-ups of research data on the data management survey in 2017/2018 (with response rates varying from 8% for Electrical Engineering, Mathematics and Computer Science to 37% for Aerospace Engineering). On average, 42% of the 628 respondents indicated they have their research data automatically backed up, compared with 43% of respondents that did not.

Furthermore, researchers are not aware of the terms that they need to find the information they require and are thus unable to find the right place to ask their questions. For example, only 20–30% of the TU Delft researchers indicated they were aware of the FAIR principles (Findable, Accessible, Interoperable, Reusable) or data ownership. If researchers are unaware of principles and regulations, they will not be able to adhere to them. This lack of awareness should not be confused with a lack of interest in the topic. The majority of respondents (between 78–94% across the six faculties) to the survey indicated that they considered themselves as responsible for the stewardship of their research data and 80% of the respondents were interested in data management training. We therefore reasoned it was essential to better connect researchers with the research data management and sharing solutions they sought.

In parallel to the survey, we also conducted qualitative, informal interviews with researchers. These prompted the realization that despite TU Delft's overarching focus on research in various domains of engineering and technical sciences, there are significant differences between faculties in the type of methodology applied and the types of data generated. For example, at the Faculty of Electrical Engineering, Mathematics and Computer Science, almost every project has a strong computational component, often relying on big data processing. Researchers from the Faculty of Technology, Policy and Management gather a lot of personal data coming from quantitative and qualitative surveys. Moreover, two faculties (Industrial Design and Architecture and the Built Environment) have a specific focus on design processes, and the role and even definition of data in design is subject to much discussion.

What we learnt from both approaches was that:

• researchers are unaware about TU Delft data management facilities and RDM terminology

• different faculties approach RDM differently, have diverse needs and require dedicated support.
Data Stewards: generalists with research background and excellent communication skills

In response to the issues identified above, the Data Stewardship project at TU Delft was initiated in 2017.\textsuperscript{26} The Data Stewardship project focuses on incremental improvements in current data management and sharing practices, by implementing relevant changes within the faculty and providing support for researchers. Each of the eight faculties has had a full-time Data Steward since the end of 2018. The Data Stewardship project was initially centrally supported via strategic funding from the University’s Executive Board and co-ordinated by the Data Steward Co-ordinator working from TU Delft Library. At the time of writing, financial responsibility for the Data Stewards is being adjusted, and each individual faculty will be financially responsible for its own Data Steward. Data Stewards are increasingly hired in the Netherlands,\textsuperscript{27} but TU Delft is one of the first universities in the world to provide such support with this capacity at the faculty level.

The decision to embed the Data Stewards at the faculty level was a conscious way of addressing the communication issues mentioned above. Rather than being based centrally (e.g. at ICT or the Library), positioning at the faculty level enables a close connection to researchers: a local, dedicated, easily findable point of contact for any questions they may have regarding data management. These questions focus on storage solutions, data management tools, data sharing, DMPs, and budgeting for data management. Data Stewards are able to answer most questions related to these topics, and, where necessary, they connect researchers with other subject experts (e.g. on the General Data Protection Regulation [GDPR], ICT and legal teams). The majority of support offered is through personal consultations with researchers at the time when the researcher requires this support. Researchers either request help themselves or are offered support by the Data Steward (e.g. after grants are awarded that require DMPs). By focusing on providing expert advice and guidance and increasing awareness, instead of chastising researchers for failing to meet requirements, Data Stewards aim to build the trust of the research community.

To successfully engage with researchers and drive improvements in RDM practices, the Data Stewards must have a very specific skill set. At TU Delft, they all have a PhD degree (or equivalent) in a subject area that is relevant to the faculty. This background in research allows the Data Stewards to communicate more efficiently with researchers, as they are familiar with the research practices, struggles and requirements. Next to an understanding of the requirements and tools that researchers need, it is essential that Data Stewards have excellent communication skills and understand the views of different stakeholders within the university (Figure 2). They function as a connection point between their faculty and the broader University. Therefore, strong interpersonal skills are crucial to effectively translate and understand different policies and requirements from the faculty point of view, while looking for opportunities for cross-University collaborations and synergies. Good communication within the Data Stewards team is also essential. Without regular contact the chances of each Data Steward giving conflicting advice, missing the opportunities for synergy or being pushed in a specific direction by their own faculty is much higher. A Data Stewardship Coordinator oversees the Data Stewards team to facilitate effective co-operation between the team members. For example, there are weekly meetings with the full team, a dedicated Slack channel, an online communication platform, for short communication and one-to-one meetings between Data Stewards and the Data Stewardship Coordinator.
Figure 2. Various stakeholders that the Data Stewards interact with at TU Delft

It is also worth emphasizing what Data Stewards are not intended to do. They do not act as compliance police. This is re-emphasized by the University’s Research Data Framework Policy (mentioned below): final responsibility for how research data is collected, analysed and shared should be with the researcher and not the supporting staff. Equally, Data Stewards cannot dedicate themselves to helping specific research groups and projects at length as they work across a faculty. Not only are there too many researchers and too many varied requests, but, more importantly, Data Stewards must have a holistic overview of the faculty data management needs in order to advise on the most effective ways to address them. So, Data Stewards are not technical experts that can dive in and manage a project’s data or code. Rather, they are skilled ‘generalists’ with a research background. They provide broad advice – or point to other experts – that then allows researchers to make more fine-grained decisions about how they manage their data.

Data Champions are leading the way

As one Data Steward cannot be familiar with all the discipline-specific practices within their faculty, and peer-to-peer learning is more effective,28 the Data Champion initiative was started in 2018,29 inspired by a similar endeavour at the University of Cambridge.30 TU Delft Data Champions are leaders in the research community that practise and advocate good RDM. They are willing to share their experiences, tools and tips with their peers and can provide the discipline-specific support that the Data Stewards cannot. In return, the Data Champion initiative offers (international) networking and funding opportunities for training and workshops, increased visibility of researchers and recognition for their work in code and data management.

The growing community consists of over 45 Data Champions at the time of writing, with representatives of all the faculties and almost all departments. The Data Champions are interested in a broad range of topics, and involved in initiatives such as improving research reproducibility in geosciences,31 software reproducibility32 and data sharing.33 In 2019, ten interviews with Data Champions were conducted to highlight their work, which were then published as blog posts.34 The interviews offered the Data Champions an opportunity to
talk about various aspects of their work, such as promoting open hardware, using Electronic Lab Notebooks, providing training for other researchers, leading citizen science projects, overcoming challenges with data sharing, and many others.

Data Champions help accelerate the improvement of RDM practices by contributing to a shared vision and highlighting the need for change. At the same time, they demonstrate how to implement these changes and form a community that establishes best practices. Having Data Champions teaming up with the Data Stewards facilitates peer-to-peer learning strategies and the creation of tailored data management workflows, specific to individual research groups. Examples of these collaborations are the development of discipline-specific data management policies and Data Champions and Data Stewards working together to teach researchers programming skills, as outlined in more detail in the next sections.

**A shared vision: policy development**

Allied to the appointment of the Data Stewards, the TU Delft Research Data Framework Policy was published in 2018. The framework policy outlines the roles of the Library, ICT Department, University Services, the Graduate School and the Executive Board at TU Delft. To ensure it respects different research practices in different disciplines, it asks the faculties to create their own research data policies. The faculty policies will specify the responsibilities of faculty-level stakeholders: deans, heads of departments, researchers and PhD students. The Data Stewards are leading the development of the Faculty Research Data Policies and are tasked with ensuring cross-campus coherence in the faculty-specific policies.

The development of the faculty policies is achieved through discussions in meetings between the Data Stewards, management support staff, the dean, heads of departments and researchers. Regular consultations with researchers during the policy development period also presented an opportunity for raising awareness about data management. The Data Champions are also actively involved in the development of the policy. For example, Data Champions from the Faculty of Applied Sciences led the development of the data management policy for their department, Quantum Nanosciences. Their policy then inspired the development of the faculty policies for Applied Sciences and Mechanical, Maritime and Materials Engineering.

The bottom-up approach in which feedback was gathered was greatly appreciated by researchers. The direct involvement and investment of researchers’ time in improving the RDM guidelines and requirements may increase their commitment to the success of changing the practices, increase awareness of the benefits that come with the change, and it also creates a sense of ownership of the policy within the faculty. A change in practices is easier when the community agrees on why these changes are important.

**Need for agility: moving from data to code**

The Data Stewards were initially asked to provide support for data, but through interactions with researchers it became increasingly apparent that software support was just as important. At a university of technology such as TU Delft, a large percentage of the researchers are dependent on in-house software tools for their research, but they do not necessarily have the software development background required to update or maintain them, and can therefore experience various difficulties. With improved software skills, researchers can manage and share their data more easily. As a result, their research overall becomes more reproducible. Moreover, there are similar barriers for the uptake of both coding and data management practices, as both outputs are currently undervalued.

Providing related support for software and data therefore became part of the Data Stewards’ core work. The Data Stewards responded to this demand in various ways, for
example they are learning software support skills themselves to transfer them to researchers through Software Carpentry and Data Carpentry training. The Carpentries being a non-profit project that promotes reproducible computational research through teaching basic computing skills in an inclusive environment to researchers worldwide. The Data Stewards also organize Coding Lunch and Data Crunch walk-in sessions and encourage ongoing initiatives that provide more in-depth software support such as a Coding Assistant, a dedicated person addressing specific coding questions. In these initiatives, both the Data Stewards and the Data Champions play a crucial role in the organization and communication of the events, as they can reach out to researchers about these activities and encourage them to participate, act as trainers, and address specific questions researchers may have on data and code management.

The challenge of rewards and incentives

Beyond TU Delft, there are many issues that affect the practice of RDM. Crucially, the academic reward system needs to change, and to change at a global level. Researchers value data management practices but will only give these practices a higher priority and change their current norms when these activities become key in hiring criteria, performance evaluations and funding rewards. At TU Delft, the Data Champion initiative acknowledges researchers with good RDM practices and boosts appreciation of these skills in their annual reviews. The Data Stewards promote the work of new roles such as data managers and software engineers and encourage their hiring. While the TU Delft Data Stewardship project is only the work of one institution, it can help with actions that set an example for broader change. Indeed, in 2018 alone the team have attended 46 national and international conferences and meetings, including 33 occasions when team members were asked to talk about their work as invited speakers or keynote speakers.

Furthermore, the Data Stewards promote the importance of software and data management and teamwork by actively engaging with NWO, the Netherlands Organization for Scientific Research with funding instruments, on these topics. In 2018 the Data Stewards organized a workshop on data management and open science skills that researchers need to have at different stages in their careers. The results of this workshop were incorporated in the EOSCpilot (European Open Science Cloud, a cloud service offering a catalogue of resources and services for open science). The Data Stewards’ work is part of a larger set of initiatives being taken by TU Delft (in its forthcoming Open Science programme for 2020–2024) and the Netherlands as a whole, via the National Platform for Open Science (a collaboration of Dutch organizations on realizing open science). For systematic change in rewards and incentives, the Data Stewards work closely with TU Delft Library to engage a broad coalition of stakeholders at a national and international level. While changing academic rewards and incentives is not part of the official job description of the Data Stewards, they facilitate these conversations at various national and international fora and help to increase the recognition of data management activities.

Conclusion

TU Delft is privileged to already have an appropriate technical infrastructure in place, enabling the Data Stewardship project to drive the cultural change required to RDM practices. Without the right people that understand the needs and requirements of researchers regarding their data and code management practices, these practices will not improve and available tools will remain underutilized. The Data Stewardship project reaffirms existing values of the research community and allows researchers to commit to the changing norms of TU Delft, their faculties, the funders and the broader research community. By working together, the Data Stewards and Data Champions are building a
community that paves the way for cultural change in research data management and sharing practices at TU Delft. Even when resources are limited, it is possible to build a community of individuals, such as the Data Champions, that are engaged with data management practices and to facilitate and support them to enable cultural change. At TU Delft this cultural change will still take time, even with the Data Stewards and Data Champions in place. While the road to cultural change in improving research data management practices is long, TU Delft has covered a considerable distance since the introduction of the Data Stewardship project.

Acknowledgements
The authors are thankful for the feedback received on this case study by an anonymous reviewer and for the editorial guidance of the publications associate, Ally Souster.

Abbreviations and Acronyms
A list of the abbreviations and acronyms used in this and other Insights articles can be accessed here – click on the URL below and then select the ‘full list of industry A&As’ link: http://www.uksg.org/publications#aa

Competing interests
The authors declare that they have no competing interests.

References
5. Julia S. Stewart Lowndes et al., “Our path to better science in less time using open data science tools,” Nature Ecology & Evolution 1, no. 6 (June 2017), DOI: https://doi.org/10.1038/s41559-017-0160 (accessed 26 September 2019).


24. Marta Teperek, Views on Data Stewardship – report of preliminary findings at the Faculty of Technology, Policy and Management (TPM) at TU Delft (2018), DOI: https://doi.org/10.31219/osf.io/8ce5v (accessed 16 September 2019).


39. Teperek, Views on Data Stewardship; Cruz et al., “Policy needs to”; Gilley, Gilley, and McMillan, “Organizational Change.”


