IJDC | *General Article*

Data Stewardship: Addressing Disciplinary Data Management Needs

Marta Teperek
Research Data Services
TU Delft Library

Maria J. Cruz
4TU.Centre for Research Data
TU Delft Library

Ellen Verbakel
Research Data Services
TU Delft Library

Jasmin Böhmer
4TU.Centre for Research Data
TU Delft Library

Alastair Dunning
Research Data Services and 4TU.Centre for Research Data
TU Delft Library

Abstract

One of the biggest challenges for multidisciplinary research institutions which provide data management support to researchers is addressing disciplinary differences (Akers and Doty, 2013). Centralised services need to be general enough to cater for all the different flavours of research conducted in an institution. At the same time, focusing on the common denominator means that subject-specific differences and needs may not be effectively addressed. In 2017, Delft University of Technology (TU Delft) embarked on an ambitious Data Stewardship project, aiming to comprehensively address data management needs across a multi-disciplinary campus.

In this article we describe the principles behind the Data Stewardship project at TU Delft, the progress so far, identify the key challenges and explain our plans for the future.

Received 22 January 2018 ~ *Accepted* 20 February 2018

Correspondence should be addressed to Marta Teperek, Research Data Services, TU Delft Library, Prometheusplein 1, 2628 ZC Delft, Netherlands. Email: m.teperek@tudelft.nl

An earlier version of this paper was presented at the 13th International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

 doi:10.2218/ijdc.v13i1.604

Introduction

The high potential of research data and of data-driven approaches led to data stewardship being recognised internationally as a key foundation of future science. Carlos Moedas from the European Commission said that Open Science "is a move towards better science, to get more value out of our investment in science and to make research more reproducible and transparent. (...) Recent advances (...) required thousands of scientists to collaborate on (...) on data. And that implies that research data are findable and accessible and that they are interoperable and reusable" (Ayris et al., 2016). In support of this, the European Commission estimated that about 5% of research expenditure should be spent on properly managing and stewarding data (Ayris et al., 2016). Barend Mons, chair of European Commission's high level expert group on the European Open Science Cloud, estimated that 500,000 data stewards will be needed in Europe to ensure effective research data management (Versweyveld, 2016). In addition, the European Open Science Cloud promises new tools, and related EU strategy papers suggest new rewards and grant funding schemes (such as FP9) to benefit those practising open science (O'Carroll et al., 2017).

TU Delft's College van Bestuur (CvB – TU Delft Executive Board) made a strategic decision to be a frontrunner of this global move and a dedicated Data Stewardship programme was initiated (Teperek, 2017a). The long-term goal of this programme is to comprehensively address research data management needs across the whole campus in a disciplinary manner. To achieve this, subject-specific data stewards are to be appointed at every TU Delft faculty. Strategic funding from the CvB was allocated to support 0.5 FTE of a data steward per faculty until December 2018, and 1.0 FTE of a data steward per faculty from January 2019 to December 2020. Subsequently, faculties are to themselves decide how to best address their researcher data management needs (van Wezenbeek, 2018).

Principles Behind the Data Stewardship Project

TU Delft has eight faculties, each conducting different types of research. Therefore, to truly understand and address data management needs at faculty level, we believed there should be a dedicated, embedded person – a data steward – employed to look after research data at each of the faculties. In addition, to ensure consistent and aligned messages and to allow coherent service development, there had to be a mechanism to foster the creation of active links between faculty data stewards, the Library-based central Research Data Support team and other institutional service providers. Therefore, the role of Data Stewardship Coordinator was created, based centrally at the Library.

Data Stewards: Disciplinary Experts Who Look After Research Data

Data stewards are disciplinary experts with knowledge of data management who are employed at faculties in order to advise researchers and faculty members on the various aspects of research data management. Specifically, the data stewards are tasked with the following:

- **Analyse data management needs** – through undertaking a mixture of semi-structured qualitative interviews and quantitative surveys;
- **Provide advice and consultancy** – meet with researchers, discuss their data management practices, make suggestions for possible improvements and become the trusted person for any questions about data management;
- **Liaise with key faculty stakeholders** – ensure that the various faculty service providers (such as contracts managers or faculty information coordinators) are aware of good data stewardship and that requirements of good data stewardship are aligned with their workflows (for example, budgeting for data management in grant applications);
- **Train and inspire** – advocate for good data management, deliver information sessions, analyse training needs, develop and deliver workshops to ensure that researchers have the skills necessary for responsible data stewardship;
- **Help comply with funders' and journals' policies** – assist researchers with drafting their data management plans, preparing their research data for deposit and advise them on changes to data policies;
- **Develop faculty research data policies** – organise and facilitate policy consultations across the faculty, help faculty define roles and responsibilities of the different faculty-level stakeholders, and drive policy implementation, evaluation and revision;
- **Prepare the faculty for the future** – keep the faculty up to date with new developments and policy changes related to data stewardship, and keep abreast of new developments in the faculty's research area to ensure that researchers have the right skills to manage their data, despite of evolving research methodologies;
- **Liaise with the Data Stewardship Coordinator and other stewards** – liaise with other members of the Data Stewardship programme to exchange practice and to discuss relevant issues;
- **Deliver regular reports** – regularly evaluate, monitor and report on data management practices within the faculty.

In addition, we believed that disciplinary expertise, reflected in a PhD degree (or equivalent experience) in the area of faculty's research, was necessary for the stewards to provide relevant and tailored advice to their communities.

Central Coordination

Furthermore, it was important to ensure that data stewards, while embedded within faculties, worked together as a team and shared the same goals. Team working also implied establishing mechanisms for effective practice exchange, creating a peer support network and being able to utilise complementary skills of the different team members.

In addition, data stewards were primarily subject specialists and needed training and support in data management aspects of their role. Finally, given that TU Delft Library already had a central Research Data Support team, it was important to ensure effective coordination of work between the data stewards and the central team. For this reason, it

 doi:10.2218/ijdc.v13i1.604

was crucial to have the position of the Data Stewardship Coordinator, appointed at the Library. The responsibilities of the Data Stewardship Coordinator are:

- **Building and sustaining a team of data stewards** – create and facilitate an ongoing forum for all data stewards to meet, support each other and exchange practice;
- **Motivate and support data stewards** – encourage and support the data stewards and identify opportunities for their professional development;
- **Ensure appropriate training for data stewards** – assess and monitor the needs and possible skill gaps of data stewards and ensure that appropriate training is provided;
- **Facilitate interactions between the data stewards and the central Research Data Services team** – ensure that work provided by both teams is complementary and that there is communication flow between team members;
- **Identify synergies across the university** – consolidate the faculty-specific information obtained by data stewards in order to identify and leverage synergies within the university;
- **Liaise with other university services** – liaise with the various university-level service providers in relation to research data management and ensure coordination of efforts and coherence of workflows and policies, and ensure that other service providers share their expertise, as needed, with the data stewards;
- **Influence and coordinate implementation of a research data policy framework** – contribute to the development of the research data policy framework at TU Delft and coordinate its implementation across the faculties;
- **Monitor and inform policy development** – monitor changes in recommended data management practices on international level, inform policy development and ensure that data stewards are aware of the most important developments and advise their faculties accordingly;
- **Establish and maintain effective national and international networks** – collaborate with colleagues nationally and internationally to share and inform best practice in data stewardship.

Research Data Policy Framework

We also believed that if the Data Stewardship project was to become truly embedded across the campus, advocacy, training and other work delivered by data stewards needed to be accompanied by policy changes. Many institutions, in particular in the UK, have central policies on research data management (University of Vienna, 2016). However, at TU Delft it was decided that each faculty would develop its own policy on research data management, based on a common template. Our rationale for deciding on a different approach was the belief that one-size-fits-all solutions for research data management were difficult to implement in practice and came with a risk of becoming too aspirational and detached from the day-to-day practice. Ensuring faculty's leadership in policy development would also likely result in greater ownership and engagement with research data management within the faculties.

Coherence between the different faculty policies would be ensured by the use of a common policy template. The template would determine which policy aspects cannot be changed (for example, policy for working with personal sensitive data), and which aspects can be adjusted, depending on the faculty's preferences (for example, the requirement for every project to have a data management plan). In addition, an overarching TU Delft data management policy framework would be the central reference point for all faculty policies and it would also determine the responsibilities of the central TU Delft support services regarding research data.

Progress So Far

Between August and October 2017, the Data Stewardship Coordinator and the first data stewards were appointed at three TU Delft faculties: the Faculty of Electrical Engineering, Mathematics and Computer Science, the Faculty of Civil Engineering and Geosciences and the Faculty of Aerospace Engineering (Dunning and Teperek, 2017). At the beginning of 2018, data stewards are to be appointed at the five remaining TU Delft faculties: the Faculty of Architecture and the Built Environment, the Faculty of Industrial Design Engineering, the Faculty of Technology, Policy and Management, the Faculty of Applied Sciences and the Faculty of Mechanical, Maritime and Materials Engineering.

Following the initial appointments, ways of working, shared project spaces and communication tools were established. In addition, in order to ensure information flow and team building, regular weekly meetings were set up between all the data stewards, as well as meetings three times per year for faculty secretaries to discuss data stewardship.

The data stewards have also started investigating data management needs and organising advocacy events at their faculties. As of 10 January 2018, a total of 34 semi-structured qualitative interviews were conducted. This was accompanied by a quantitative survey sent out to researchers at all three faculties (see below). Ten separate information sessions were delivered at faculties and departments. Several types of advertisement materials were also developed. In addition, five blog posts and newsletters and four progress reports were published on the Open Working blog[1]. Data stewards also engaged with colleagues nationally and internationally, including five presentations given at national and international conferences.

Training Needs

Data stewards were recruited primarily as disciplinary experts, who, although all interested in data management, had various degree of awareness of the different aspects of good data management practice, such as funders' requirements, FAIR principles (Wilkinson et al., 2016) etc. Therefore, an intense training programme was designed. In addition to completing the Essentials 4 Data Support[2] course provided externally (Grootveld and Verbakel, 2015), the data stewards attended multiple sessions delivered by internal and external experts (local ICT provisions, national ICT provisions, data management planning, use of repositories, selfish benefits of data management, workflows for working with big data, the principles of open science and others)

1 Data Stewardship – Open Working: https://openworking.wordpress.com/category/data-stewardship
2 Essentials 4 Data Support: http://datasupport.researchdata.nl/en

 doi:10.2218/ijdc.v13i1.604

(Teperek, 2017b). Completion of the training programme helped to ensure that all of the stewards are equipped with the necessary knowledge and skills to advise their faculty researchers on data management.

Challenges

Despite the fact that the project was only launched in 2017, several challenges have been already identified. The top three are:

1. Effective communication and collaboration between data stewards and the central Research Data Support team;
2. Establishing measures for project evaluation;
3. Addressing the need for more granular disciplinary experts.

All three challenges are briefly discussed below.

Collaboration Between the Disciplinary Data Stewards and the Central Research Data Support

One of the biggest challenges for the project was to establish a framework for effective communication and collaboration between the data stewards and the central Research Data Support team. As mentioned before, the central Research Data Support team at TU Delft Library had been already providing services to the research community, such as advice on data management planning, advice on data archiving and others. With the appointment of data stewards, it became crucial that research support tasks are effectively allocated between the two teams and also that the two teams communicate with each other.

First, to split the tasks effectively, it was decided that data stewards would be the first contact point for their researchers due to their awareness of subject-specific practices. The central team would be the source of expertise in the more general matters, such as data management plan support, or the use of the central data archive '4TU.Centre for Research Data' for data publication. Second, an in order to ensure effective flow of information, two members of the central Research Data Support team who are responsible for liaison with researchers attend the weekly data stewards meeting. In addition, all members of the Research Data Support team are invited to progress update meetings by the data stewards.

Handling data management plans is one example of how the collaborative approach was implemented in practice. First, members of the Research Data Support team delivered training to data stewards on supporting researchers with writing data sections in grant proposals, as well as preparing data management plans. Subsequently, data stewards attended meetings between the Research Data Support team and researchers to learn from the experts on how to best deliver the advice on data management plans. Once the data stewards got confident enough about their expertise on data management planning, they became the first contact points for researchers at their faculties who needed assistance with data management plans. The data stewards now explain to researchers the principles behind data management plans, discuss with them their data management strategy and provide suggestions which are the most relevant to the type of

research. After receiving the thorough advice from a data steward, researchers draft their data management plan, which is then sent to the central Research Data Support team for any additional feedback. In that way tailored support for data management plans can be provided to a larger number of researchers and the central Research Data Support team undertakes quality assurance and offers additional advice as needed.

Metrics for Project Evaluation

In order to judge the progress of the project, it was necessary to develop and agree on an effective set of metrics, which would allow us to decide whether the project is moving towards its goal of improving good research data management practice. In addition to qualitative, semi-structured interviews, we also created a quantitative survey in collaboration with EPFL in Switzerland and the University of Cambridge in the UK. The survey consisted of ten primary questions about data management practice, which were selected based on existing data management assessment tools (Johnson, Parsons and Chiarelli, 2016) and which were complemented with some institution-specific questions. We decided on a minimal number of agreed questions for two reasons. First, we would like to re-run the survey periodically to assess the progress of the project and we thought that keeping the number of questions to the minimum would more likely help us ensuring better response rates over the years. Second, using a set of agreed questions by several different institutions meant that cross-institutional comparisons and benchmarking would be possible.

As of January 2018, the first three TU Delft faculties and EPFL, had completed their first survey runs. The University of Cambridge was in preparation for the survey dissemination. As soon as the initial datasets from the three institutions are available, they would be made openly available, together with the survey itself.

Is One Data Steward Per Faculty Enough?

Finally, in order to be truly discipline-specific, one data steward per faculty might not be enough. There is substantial diversity in the research topics and disciplines within the faculties themselves. For example, research groups at the Faculty of Applied Sciences work on topics ranging from Imaging Science, through Biotechnology, Chemical Engineering all the way through to Quantum Nanoscience. All these topics have different methodologies and work with diverse types of research data. Data stewards should thus be supported by departmental data champions (Higman, Teperek and Kingsley, 2017) who could also act as local community advocates for data management. We are planning to initiate a similar Data Champions programme at TU Delft in 2018.

Future Outlook

In 2018 the data stewards will primarily focus on awareness raising and will start developing training for research communities. They will also initiate policy development work and, importantly, launch the Data Champions programme. Data stewards will also analyse and publish the initial quantitative survey results and they will summarise findings from the qualitative, semi-structured interviews. Both will help faculties decide on top priorities for addressing their data management needs. Finally,

2018 will see data stewards at the remaining five faculties joining the team and will thus enable the project to start moving at its full speed.

The final goal of the data stewardship project is to ensure that researchers at TU Delft adhere to good data management practice on a day-to-day basis. However, a realistic short-term goal is to assess whether the proposed solution of having dedicated data stewards network leads to improvements in data management practice. Initial metrics for this should become available towards the end of 2018, when the three faculties re-run their surveys.

The ultimate evidence of the success of the project would be the judgement of the research community itself. Will researchers perceive data stewards as the trusted sources of data management expertise? If so, one would expect to see data stewards becoming permanent, key faculty staff members. And if subject-specific expertise is the solution for good data management practice, perhaps one day the presence of a dedicated data steward in every research group will become the norm, similar to the presence of lab managers or project managers nowadays.

References

Akers, K.G. & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation 8*(2), 5-26. doi:10.2218/ijdc.v8i2.263

Ayris, P., Berthou, J-V., Bruce, R., Lindstaedt, S., Monreale, A., Mons, B., Murayama, Y., Södergård, C., Tochtermann, K., Wilkinson, R. (2016). Realising the European Open Science Cloud. European Commission. Retrieved from https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf

Dunning, A., & Teperek, M. (2017). Update on the Data Stewardship project at TU Delft: 3 months' progress report. Open Working Blog. Retrieved from https://openworking.wordpress.com/2017/11/28/update-on-the-data-stewardship-project-at-tu-delft-3-months-progress-report/

Grootveld, M., & Verbakel, E. (2015). Essentials for data support: Training the front office. *International Journal of Digital Curation 10*(1). doi:10.2218/ijdc.v10i1.364

Higman, R., Teperek, M., & Kingsley, D. (2017). Creating a community of data champions. *BioRxiv*, 104661. doi:10.1101/104661

Johnson, R., Parsons, T., & Chiarelli, A. (2016). Jisc data asset framework toolkit. doi:10.5281/zenodo.177876

O'Carroll, C., Rentier, B., Cabello Valdes, C., Esposito, F., Kaunismaa, E., Maas, K., et al. (2017). Evaluation of research careers fully acknowledging open science practices: Rewards, incentives and/or recognition for researchers practicing open science. European Commission Working Group on Rewards under Open Science. Retrieved from https://ec.europa.eu/research/openscience/pdf/os_rewards_wgreport_final.pdf

Teperek, M., (2017a). *Data stewardship – addressing disciplinary data management needs.* Open Working Blog. Retrieved from https://openworking.wordpress.com/2017/08/29/data-stewardship-addressing-disciplinary-data-management-needs/

Teperek, M., (2017b). Training for data stewards. Open Working Blog. Retrieved from https://openworking.wordpress.com/2017/09/18/training-for-data-stewards/

University of Vienna. (2016). Evaluation grid for RDM policies in Europe. Retrieved from http://learn-rdm.eu/wp-content/uploads/LEARN_Evaluation-Grid-for-RDM-Policies-in-Europe_Version-1.0-August-2016.pdf

van Wezenbeek, W. (2018). *Open science – our way forward.* TU: Librarian Blog. Retrieved from https://tulibrarian.weblog.tudelft.nl/2018/01/11/open-science-our-way-forward/

Versweyveld, L. (2016). We need 500.000 respected data stewards to operate the European Open Science Cloud. e-Infrastructures Reflection Group. Retrieved from http://e-irg.eu/news-blog/-/blogs/we-need-500-000-respected-data-stewards-to-operate-the-european-open-science-cloud

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data 3*. doi:10.1038/sdata.2016.18