Vision for Research Data & Software management training at TU Delft

Date: 23-10-2019
Authors: Paula Martinez Lavanchy, Alastair Dunning, Marta Teperek, Nicole Will and Ellen Verbakel

Introduction

Research has become a highly data-intensive activity. New technologies and instruments allow researchers to collect and produce data at exceptionally fast rates, with the consequence that new analysis techniques and methodologies are emerging. Relatedly, the education programmes at universities have evolved and have started to include data analysis and programming skills within their curricula. However, there is less focus on the good management, documentation and preservation of those data when teaching those skills.

Funding agencies, scientific journals and universities have recognised the value of making high-quality research data available, whenever possible, for reuse. Researchers also recognise the value of the data they produce and publish. However, often best practice for RDM is not part of their research workflows. Therefore, TU Delft Library believes that an overarching vision for training can help embed Research Data and Software Management as part of the education and skills development of students and researchers.

This vision and its implementation should help to provide a tailored programme of skills and knowledge development that serves different disciplinary needs regarding RDM across the institution. An effective training strategy and implementation will help to increase the quality of the data collected and produced at TU Delft. It will help TU Delft researchers comply with funder requirements, do research in line with the institutional Research Data Framework Policy, do research more efficiently and to build more trusted and transparent alliances/collaborations.

Scope

This vision focuses on the skills needed by PhD students and early career researchers (from now on referred to as researchers) because they are the ones facing increasing requirements from funders, journals and institutional policies regarding research data and software management.

Although the vision does not directly target senior researchers, it is still relevant to consider a way of making available to them the content of the courses here described. Senior researchers should be able to consult the material whenever they are confronted with a

question or if they want to broaden their knowledge on the subject of research data and software management.

Masters students are not considered in this strategy for the time being. After a successful implementation of this training strategy, it should be evaluated how to extend the training to Master students.

Pre - requisites for a successful implementation of this vision:

1. **Building on existing courses.** Data and software management training has to be built upon existing generic and faculty-specific courses[1] in order to ensure complementarity with existing skills learned by the PhD students and to avoid reinventing existing courses .

2. **New courses and materials.** In order to build and implement this vision, it will be necessary to create new courses, new teaching material. Depending on the type of course, it will need to be prepared/provided by the library and/or the faculties (i.e. graduate schools, education teams, researchers, etc.).

3. **Resources and sustainability.** A collaborative effort between the library and the faculties and other services (including external/international collaborators) is needed to provide the training and to make it sustainable. TU Delft Library should ensure the resources for assuming a continuous coordinating role of the training, allocate resources for the generation of relevant material and provide its employees with the relevant skills needed for providing training. Faculties should recognize the value of their data stewards and researchers as skilled trainers to deliver the courses described in this vision and to ensure that the proper incentives (see next point) are in place for them to get involved in the training activities.

4. **Incentives**. Researchers have to receive the proper incentives to take part in the training. In the case of PhD students, they should receive credits recognized by their respective graduate schools according to the 'Doctoral Education competencies model' designed by TU Delft Graduate School. For early stage and senior researchers that decide to join the training, the new skills acquired should be recognised as part of their curriculum and competence evaluations. For those data stewards and researchers that provide training, this activity should be recognized as part of their teaching duties and be considered in the 'Result & Development Cycle'.

5. **Revision and iterative improvement.** The library and graduate schools have to continuously engage in consultation processes with PhD students and researchers (including PhD supervisors), and use the feedback to iteratively improve and update the content and the way the courses are delivered. For example, it should be evaluated if those courses initially provided in a face-to-face format can be

[1] Digital Skills: Training on basic programming skills using Python for Bsc programs. Starts in September 2019 - approx. 1 EC.

transferred to a blended or online format. The feedback should be collected at the different courses, but also from other events related to the topic of Research Data and Software management.

6. **Licensing and Copyright.** All material produced for the courses described in this vision will be published and made available under a Creative Commons - Attribution License (CC-BY 4.0). This license will allow for the material to be publicly available in order to engage, for example, senior researchers that are not formally taking part in the training program. It will also allow the re-use of the material by all researchers who would like to include different aspects of RDM within their courses. Finally, it will also allow the re-use of the material by RDM support services or researchers outside TU Delft.

7. **Lesson Plans.** All newly developed courses will be accompanied by clear learning objectives, a lesson plan and a description of the methods selected for the training.

Schematic representation of the training vision.

This vision consists of a modular approach that covers the different levels of knowledge needed in the different stages of a PhD project and takes account of the different types of data and disciplines found at TU Delft.

TU Delft library will work on developing guidance and tools to help PhD students and researchers to choose the right courses according to their background knowledge.

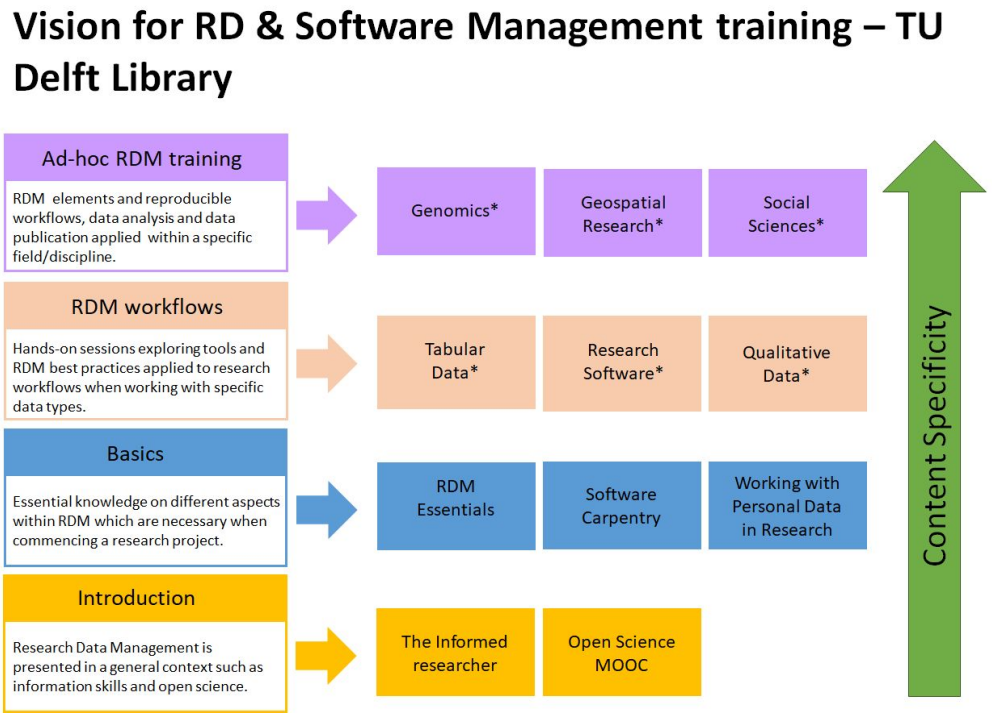

Figure 1. Schematic representation of the Vision for Research Data & Software management training at TU Delft. *Examples of envisioned courses.

The courses (some already available and some in preparation) are organised in four different modules, which build upon each other (Fig. 1). The different levels (from bottom to top) increases the specificity of the content from considering data into a general context (e.g. open science) to skills that apply to a specific data type or a research discipline.

From bottom to top:

Introduction: Research Data Management is presented in a general context such as information skills and open science.

Basics: Essential knowledge on different aspects within RDM which are necessary when commencing a research project.

RDM workflows: Hands-on sessions exploring tools and RDM best practices applied to research workflows when working with specific data types.

Ad-hoc RDM training: RDM elements and reproducible workflows, data analysis and data publication applied within a specific field/discipline.

Description of Existing Courses

The Informed Researcher

This course is intended for PhD students in their first year. Learners will get the skills to locate, evaluate, organize and use the right information at the right time during their research. These skills will allow them to save time and effort by learning on how to organize their work processes efficiently from the beginning.
Link:
https://www.tudelft.nl/onderwijs/opleidingen/phd/doctoral-education-programme-de/training-programme/r1a2-the-informed-researcher-information-and-data-skills/

MOOC: sharing your research with the world (2019)

This course helps researchers to grasp the key principles of Open Science. They will learn the objectives, main concepts, and benefits of Open Science principles along with practices for open data management and open data sharing.
Link: https://www.edx.org/course/open-science-sharing-your-research-with-the-world-2

Software carpentry

The Software Carpentry workshops aims to help researchers get their work done in less time and with less pain by teaching them basic research computing skills. This is hands-on two day workshop where the students learn to: automate tasks with the Unix shell, plot and Programme in Python, use version control with Git. All the courses websites are collected here: https://github.com/4TUResearchData-Carpentries

Ad-hoc RDM training

TU Delft library in collaboration with researchers actively advocating for best practices on RDM are collaborating to fulfill the needs of deeper knowledge at certain disciplines regarding reproducible workflows, data analysis and documentation, and whenever relevant regarding data publication.

An example of this is the recent Genomic Data Carpentries workshop (June 2019). It was organised by TU Delft library, data champions from the Faculty of Applied Sciences and the data stewards from the Faculty of Electrical Engineering, Mathematics and Computer Science and the Faculty of Applied Sciences. This is the link to the workshop website: https://github.com/4TUResearchData-Carpentries/2019-06-04-delft