Proceedings of the IATUL Conferences   2019 IATUL Proceedings

Closing the Skills Gap - the Basics of the Research Data Management (BRDM) Course: Case University of Turku

Jukka Rantasaari
Turku University Library Finland, jukka.rantasaari@utu.fi

Heli Kokkinen
Turku University Library Finland, heli.kokkinen@utu.fi

CLOSING THE SKILLS GAP - THE BASICS OF THE RESEARCH DATA MANAGEMENT (BRDM) COURSE: CASE UNIVERSITY OF TURKU

Jukka Rantasaari
Turku University Library
Finland
jukka.rantasaari@utu.fi

Heli Kokkinen
Turku University Library
Finland
heli.kokkinen@utu.fi

Abstract

The current challenge for researchers at the University of Turku is that there is a substantial gap between the level of targeted and present research data management (RDM) skills. In order to better understand this challenge and to develop a training course in RDM, we examined the importance of RDM competencies vs. perceived competencies of doctoral students in different stages of research data life cycle. The RDM importance and competencies were examined by interviewing doctoral students, supervisors and biostatisticians. So far, 34 interview sessions on RDM skills have been conducted, covering research data life cycle topics such as collection, organization, documentation, processing and sharing the data. The interviewees' average estimate of the importance of RDM skills in different stages of research data life cycle was 4.1 (very important) on Likert scale 1 to 5. An average estimate of the competencies of doctoral students was 2.6 (have somewhat skills). Targets for competencies have been set – besides by the interviewees themselves – by the Data Policy of the University of Turku, Finnish and EU level Open Science principles and research literature.

Based on the results we developed a three-credit RDM course for doctoral students and post-doctoral researchers. The course was developed by a working group consisting of university teacher-researchers, lawyers, library's open science specialists, data protection officer, IT Services, and biostatisticians.

Three different study programmes of the BRDM are initiated: Health Sciences programme, Natural Sciences programme and Survey and Interview Studies programme. Each study programme has 7 modules, of which 3 are mutual for all the study programmes. During the course, students complete a study plan and build a data management plan for a research project and learn e.g. to take care of data privacy and to collect, store, protect, process, document and share data.

In this preliminary paper we will discuss the conducted interviews and their key results, the RDM course planning and implementation, the student feedback and the lessons we have learned so far.

Keywords: Research Data Management, RDM, Doctoral Students, Graduate Students, Post Doc Researchers, Training, University of Turku

This document is a preliminary, abridged version of the full article, which will be published later this year.

1. Introduction

1.1 Research Data Management (RDM)

There is a substantial gap between targeted and present research data management (RDM) skills. This gap have been stated in international policy papers and reports (e.g. European commission 2017) and in research literature (e.g. Carlson & Stowell Bracke 2015; Van Tuyl and Whitmire 2016 as cited in Sapp Nelson 2017; Jahnke & Asher 2012).

RDM is systematic handling of the information generated or re-used in an academic research project (The University of Edinburgh. (2019, May 10). Why research data management? Information Services, Research data service. Retrieved from: https://www.ed.ac.uk/information-services/research-support/research-data-service/research-data-management.

1.2 Library's role

Research data can be seen as part of information ecosystem and thus as one of the information sources. It is important that library takes a central role in planning RDM, curation, sharing, re-use and preservation of research data. Library can encourage and guide researchers to use good RDM practices, teach how to use external repositories, how to cite and make data sets citable and thus get credit of sharing their data. (Carlson, Fosmire, Miller & Nelson 2011; Calzada Prado & Ángel Marzal 2013; Mannheimer, Sterman & Borda 2016)

We have promised in the Data Policy of University of Turku (https://www.utu.fi/en/research/open-science/research-data-and-data-policy) that researchers and students will be provided with training and support in creating data management plans and in data management throughout the research life cycle. Turku University Library has main responsibility in coordinating data management trainings.

1.3 Aim of the Study

The aim of the whole research project is to find out doctoral students' RDM principles and practices as well as present and needed competencies at the University of Turku. Besides of that, our aim is also to plan, implement and evaluate RDM training, taking full advantage of the findings of the interviews and surveys.

In this first part of the study we will focus on the RDM importance and competencies measured with five level Likert scale based on the rankings of doctoral students, supervisors and biostatisticians at the University of Turku. Moreover, we will use the analysis of the answers to Likert scale questions to plan, implement and evaluate RDM training for doctoral students.

2. Purpose of This Study and Used Research Methods

2.1. Purpose of This Study

The purpose of this study is

- To find out the RDM importance and doctoral students' competencies as ranked by doctoral students and faculty members
- to plan and implement RDM training for doctoral students to meet the competency needs
- To evaluate the implementation of the first RDM training

In this preliminary draft of our study we will concentrate on the first, quantitative results and their analysis of the interviews. In the full paper, which will be published later this year, we will analyze quantitatively both the interviews and the pre- and post-survey filled by the participants

of the BRDM course. We will tell more about pre- and post-survey of the BRDM course in the chapter 4.21.

2.2.   Research Methods

2.2.1.   Interviewees and interviews

Our target group is doctoral students, who, as fresh researchers, are in the process of learning research work principles and practices. Besides of them, to get faculty members views, we have also interviewed their supervisors and biostatisticians who are processing research data together with doctoral students. So far we have interviewed fifteen doctoral students and nineteen faculty members from six faculties. In the interviews we have used a combination of a structured and a semi-structured interview method. A structured interview is a one in which the interview questions together with relevant response categories are pre-specified on a questionnaire. The semi-structured interview is a qualitative data collection strategy in which the researcher asks informants a series of predetermined but open-ended questions. (Ayres, 2012). In this paper we will focus on the structured part (Likert scale questions) of the interviews.

We adapted the interview form from the Data Information Literacy Toolkit by Jake Carlson, Megan Sapp Nelson, Marianne Stowell Bracke and Sarah Wright (2015). It includes both an interview form and instructions for interviewer.

2.2.2.   Interview Forms

The original data information literacy interview worksheets includes 9 modules for faculty and 10 modules for graduate students (http://dx.doi.org/10.5703/1288284315510). Modules mainly have open ended questions and assessments on importance of certain skills related to data information literacy.

We made some changes to the interview form:

- We added the RDM competence dimension besides of the questions measuring RDM importance.
- We added informant consent by which we tell how gathered information will be used and added information of data privacy. Also confidentiality of the study is explained.
- We made the interview form electronic so all answers are saved in Webropol database (https://new.webropolsurveys.com/).

2.2.3 Quantitative analysis

In this study we will focus on the interviewees' rankings of the importance of RDM skills and doctoral students' RDM competencies in the different phases of the research data life cycle measured with five level Likert Scale. Interviewees were asked to rank the importance between 1 to 5 where 1 is "not important" and 5 is essential. Likewise, they were asked to rank the competence between 1 to 5 where 1 is "don't have skills" and 5 is "ultimate skills".

So far we have compiled the mean values of the answers to the Likert questions with all respondents' answers together and doctoral students' and faculty members' answers separately.

3.   Preliminary quantitative findings of the interviews

3.1.   Importance of RDM

All interviewees' average views on the importance of RDM competencies were above very important (4,1) on Likert scale. Among the three most important stages there were ethics and attribution (4,5), data processing and analysis (4,5) and data quality and documentation (4,4).

The views on the importance of RDM competencies between faculty members and doctoral students differed somewhat. The doctoral students' average estimate of importance were above very important (4,2), whereas faculty members' average estimate were very important (4) (graph 1).
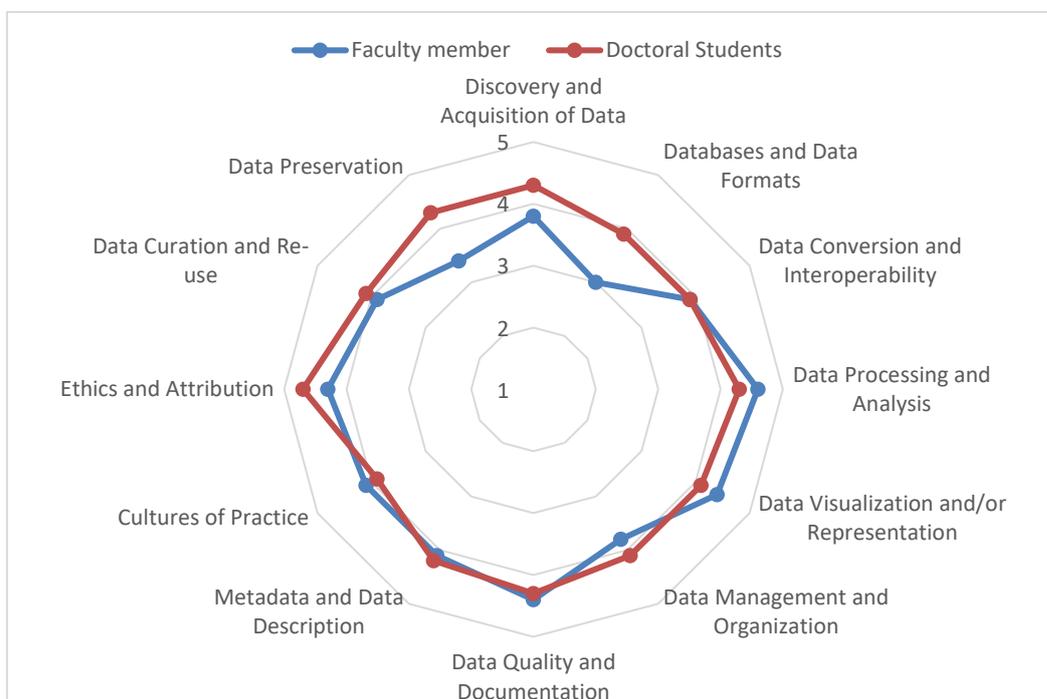
Graph 1: Importance of RDM competencies in different stages of research data life cycle as evaluated separately by faculty members and doctoral students.

3.2. RDM skills

All interviewees' average views on the doctoral students' RDM skills in different stages of research data life cycle (graph 1) were above somewhat (2,5) on Likert scale. The three stages, where skills were seen at their best, were data processing and analysis, when ranking was a little above good (3,1), data visualization and representation, when ranking was a little below good (2,9) and ethics and attribution in which case ranking was somewhat below good (2,8).

The views between doctoral students and faculty members differed notably on doctoral students' level of skills. Doctoral students' own ranking of their RDM skills was almost good (2,8) on average, whereas faculty members' ranking was a little above "have somewhat skills" (2,2) on average.

The three stages with most notable differences, were data preservation, ethics and attribution and databases and data formats. In these stages doctoral students' own evaluation of their skills was almost one whole digit higher in Likert scale than faculty members' evaluation.

4. The Basics of Research Data Management Course

4.1. Planning and structure

When interviewing the leader of the biostatistician team, came out that she had had a RedCap – database training lessons for about a year. We decided to begin planning a whole RDM basics training on the basis of her database training lessons.

We set up the Basics of Research Data Management (BRDM) working group and invited some teacher-researchers, Head of University Legal Services, Head of Research of University

Hospital Clinical Research Center, Data Protection Officer and Development Specialist of Research Development Services to the group.

On the ground of the results of the quantitative analysis of the Likert scale questions (chapter 3), we have so far found that there have been paid only little attention to RDM in faculties. There are of course differences between disciplines, but it has come clear, that there is definitely a need for very basic RDM training regardless of discipline. That's why we decided to build a common RDM basics course in the University of Turku Graduate School (UTUGS).

The Basics of Research Data Management (BRDM) Course (https://opas.peppi.utu.fi/en/course/UGSL0001/13417) includes preliminary assignments and contact sessions. The Course has three study programmes: Health Sciences, Natural Sciences and Survey and Interview Study programme. Every study programme has seven modules, three of which are common.

The BRDM working group ended up with three study programme course structure, because solutions concerning research contracts, permits and licences as well as processing, storing and long term preserving are different depending on the nature of data. Considering that the course will be common RDM basics training for all disciplines, the working group came to conclusion that we can best match these needs with this kind of course structure. E.g. in Health Sciences data usually is identifiable, confidential and highly sensitive; in Nature Sciences data for the most part is not confidential and sensitive, but there can be other very rigorous demands for data handling and storing, in Surveys and Interviews the nature of data can vary a lot depending on each respondent's and interviewee's answers.

In the heart of the BRDM course there is a study plan of a made-up research project, which teacher-researchers have created before the first lesson. The role of the made-up study plan is to help demonstrate how to apply RDM principles in practice through one representative research project. On the ground of the study plan teacher-researcher and his/her students will write a description of the data that has to be gathered and processed in the project.

After the first module the project's data will be examined from different relevant aspects like

- data management planning of the whole project,
- legal aspects like IPR issues, contracts, permits and licences,
- is there need for privacy notice and risk analysis,
- how to build a database for gathering and organizing data,
- what kind of storage, protect, process and describing are needed for the data and finally,
- how to apply FAIR principles and different long term preservation platforms for project data.

4.2.   Survey and feedback

4.21 Survey

We asked students to fill pre-survey before the course and will ask them to fill post-survey after the course. In survey form, students rank the importance of RDM competencies in different stages of research data life cycle and also rank their own skills in these stages. With the exception of questions concerning data analyzing and visualizing that were left out, questions are identical with the Likert scale importance and competence questions used in interviews (chapter 3).

Because the BRDM Course is still ongoing when writing this article, we have not yet thoroughly analyzed results of the pre-survey, let alone the post-survey, which results we don't yet have.

Preliminary quantitative analysis of the pre-survey in which we have 45 answers, hints that students of the course have ranked both the importance of RDM competencies and their present skills pretty much on the same level that were doctoral students' and faculty members' average rankings in interviews (see chapter 3).

4.2.2 Module based feedback
Besides pre- and post-survey, students have also been asked to give formal feedback after each module.

So far students have given over 130 formal module based feedbacks. Most of them are very positive with some suggestions how to make the module even better. Most common development suggestion is that it should be possible for students to write their own study plan and DMP instead of the made-up study plan and DMP. Teachers have been praised for good expertise and interactive style in most of the feedbacks.

5. Conclusions

Our aim was to find out the importance of RDM skills and present RDM competencies of doctoral students ranked by doctoral students themselves and by supervisors and biostatisticians. Based on the quantitative results of the interviews and of course on the know-how of the BRDM working group our second aim was to plan and implement RDM training for doctoral students.

The doctoral students and faculty members we interviewed ranked the importance of RDM competencies as "very important". At the same time they saw that doctoral students had only somewhat competencies. Shortage of skills were found all along the line of research data life cycle.

The Basics of research data management (BRDM) training is the first formal RDM skills training programme at the University of Turku. It contains many central issues from most common research data life cycle phases, with the exception of analysis and visualization, which were excluded. The BRDM is a common RDM basics study module for graduate students and post doc researchers in all faculties. This is the first step on the long way to data fluent research community in our university.

After we have finalized the first implementation of the BRDM Course in June 2019, we will continue analyzing the quantitative and qualitative results of the interviews, we will also thoroughly analyze the results of the pre- and post-surveys and the formal and informal feedback of the students as well as the experiences of the teachers. The BRDM working group will continue planning the second implementation, which will take place in Spring 2020. Planning and realization will be made in co-operation with experts of Åbo Akademi, the Swedish speaking university in Turku.

Our long term strategic objective is to build a RDM training path to support good quality data management in undergraduate, graduate and post doctorate phases in the University of Turku.

References

Ayres, Lioness (2012). Semi-Structured Interview In: The SAGE Encyclopedia of Qualitative Research Methods. https://dx.doi.org/10.4135/9781412963909

Calzada Prado, J. & Ángel Marzal M. (2013). Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. Libri 63(2): 123-134. http://dx.doi.org/10.1515/libri-2013-0010

Carlson, J., Fosmire M., Miller C. & Nelson M. S. (2011). Determining Data Information Literacy Needs: A Study of Students and Research Faculty. Portal: Libraries & the Academy 11(2): 629–657. http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1031&context=lib_fsdocs

Carlson, J., Sapp Nelson, M. R., Bracke, M. S., Wright, S. J. 8 (2012, last update April 6 2015), v 1.0, Data Information Literacy Toolkit. http://dx.doi.org/10.5703/1288284315510

Carlson, J., Stowell Bracke, M. (2015). Planting the seeds for data literacy: lessons learned from a student-centered education program. International Journal of Digital Curation 10 (1): 95-110. http://dx.doi.org/10.2218/ijdc.v10i1.348

The Data Policy of the University of Turku, 2016, https://www.utu.fi/en/research/open-science/research-data-and-data-policy

European commission (2017). Providing researchers with the skills and competencies they need to practice Open Science: Open Science Skills Working Group Report. https://ec.europa.eu/research/openscience/pdf/os_skills_wgreport_final.pdf

Jahnke, L.M. & Asher, A. (2012). The problem of data. CLIR Publication No. 154. Washington D.C.: Council on Library and Information Resources (CLIR). https://www.clir.org/wp-content/uploads/sites/6/pub154.pdf

Mannheimer, S., Sterman, L.B. & Borda S. (2016). Discovery and Reuse of Open Datasets: An Exploratory Study. Journal of eScience Librarianship 5 (1): e1091. http://dx.doi.org/10.7191/jeslib.2016.1091

Sapp Nelson, M. (2015). Data Management Outreach to Junior Faculty Members: A Case Study. Journal of eScience Librarianship 4 (1): e1076. http://dx.doi.org/10.7191/jeslib.2015.1076

Sapp Nelson, M.R. (2017). A Pilot Competency Matrix for Data Management Skills: A Step toward the Development of Systematic Data Information Literacy Programs. Journal of eScience Librarianship 6 (1): e1096. https://doi.org/10.7191/jeslib.2017.1096