

Doctoral Students' Educational Needs in Research Data Management: Perceived Importance and Current Competencies

Jukka Rantasaari
University of Turku
Åbo Akademi University

Abstract

Sound research data management (RDM) competencies are elementary tools used by researchers to ensure integrated, reliable, and re-usable data, and to produce high quality research results. In this study, 35 doctoral students and faculty members were asked to self-rate or rate doctoral students' current RDM competencies and rate the importance of these competencies. Structured interviews were conducted, using close-ended and open-ended questions, covering research data lifecycle phases such as collection, storing, organization, documentation, processing, analysis, preservation, and data sharing. The quantitative analysis of the respondents' answers indicated a wide gap between doctoral students' rated/self-rated current competencies and the rated importance of these competencies. In conclusion, two major educational needs were identified in the qualitative analysis of the interviews: to improve and standardize data management planning, including awareness of the intellectual property and agreements issues affecting data processing and sharing; and to improve and standardize data documenting and describing, not only for the researcher themselves but especially for data preservation, sharing, and re-using. Hence the study informs the development of RDM education for doctoral students.

Received 24 February 2020 ~ Revision received 31 January 2021 ~ Accepted 1 June 2021

Correspondence should be addressed to Jukka Rantasaari, Lotilanranta 8 as. 2, 37630 Valkeakoski, Finland. Email: jukka.rantasaari@utu.fi

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/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution Licence, version 4.0. For details please see <https://creativecommons.org/licenses/by/4.0/>



Introduction

Although the amount of data, data platforms, and cloud services have increased manifold, researchers' data management practices have not changed at the same pace. According to many studies, a major reason for this disparity is missing or insufficient education on research data management (RDM) for researchers, resulting highly varied skills in different phases of the data lifecycle, such as collecting, storing, documenting, organizing, preserving, and sharing. This is threatening the development of eResearch (Carlson, Fosmire, Miller, and Sapp Nelson, 2011; Jahnke, Asher, and Keralis, 2012; Tenopir, Birch, and Allard, 2012): in the digitizing, networked research environment, it is necessary that researchers have new kinds of technological skills as well as skills to manage growing, diverse, and collaboratively produced data (Qin and D'Ignazio, 2010).

Definitions of the Key Concepts

Applying the widely accepted Key Competences for Lifelong Learning (European Commission, 2019) we define competence as a combination of knowledge, skills, and attitudes:

- **Knowledge** is composed of the concepts, facts and figures, ideas and theories which are already established, and support the understanding of a certain area or subject.
- **Skills** are defined as the ability to carry out processes and use the existing knowledge to achieve results.
- **Attitudes** describe the disposition and mindset to act or react to ideas, persons or situations.

On a generic level, it is possible to identify four stages in the lifecycle of research data in a research project: raw data, processed data, analyzed data, and published data (Witt, Carlson, and Brandt, 2009). Though there are many different definitions of research data, Pryor (2012) has captured the essence of it by stating that data is all the information systematically acquired and processed into new knowledge in academic research. Another concise definition is by Briney (2015), according to whom data is anything that you “perform analysis upon”. Furthermore, according to the definition of the Research Data Network (RIN), data is emphasized as a means to validate research results “as a product of research and an essential part of the evidence necessary to evaluate research results, and to reconstruct the events and processes leading to those results” (Research Information Network, 2008). RDM is the systematic treatment of data, involving operations to make data easier to find, understand, and use in present and future projects (Briney, 2015).

In the research literature, data management competencies and skills have been examined from different angles, such as from their significance on furthering active citizenship and democracy (Atenas, Havemann, and Timmermann, 2020; Warschauer, 2011), from their importance as central working skills in datafied and digitized society (Ridsdale et al., 2015; Schäfer and van Es, 2017), and from securing high quality

research by taking care of the integrity and FAIRness¹ of the data in the academic research process. In this study we focus on the latter viewpoint: the importance of data management skills and competencies in fostering the integrity and high quality of research.

Intensive use of data necessitates people to have competencies related to the creation, management, and preservation of data (Semeler, Pinto, and Rozados, 2019). RDM competencies have often been referred to in terms of literacy skills. So far, there is no agreement on a unified terminology pertaining to the concept (Borgman, 2016; Bowler et al., 2017; Klenke et al., 2020). Related terms and concepts include “data literacy” (Prado and Marzal, 2013; Stephenson and Caravello, 2007), “research data literacy” (Schneider, 2013), “statistical literacy” (Schield, 2004; Twidale, Blake, and Gant, 2013), “science (data) literacy” (Koltay, 2014; Qin and D’Ignazio, 2010; Shorish, 2015), and “visual literacy” (Hattwig, Bussert, Medaille, and Burgess, 2013; Womack, 2014). Carlson et al. (2011) have introduced the concept of “data information literacy” (DIL), which, according to authors, builds upon and reintegrates several literacies, including science data literacy and data literacy into an emerging skill set. “Data information literacy, then, merges the concepts of researcher-as-producer and researcher-as-consumer of data products”.

Concerning the elements of data (information) literacy, although there is no definitive list of the competencies involved (Koltay, 2015; Wanner, 2015), there is strong consensus in the research literature of the main topics that should be included in the data management education (Sapp Nelson, 2017). Carlson et al. (2011) studied the data management needs of the research faculty and students of the Earth and Atmospheric Sciences, filtered the information through the perspective of the ACRL’s information literacy competency standards (Association of College and Research Libraries, 2000), and produced 12 core data management competencies for data information literacy (DIL): databases and data formats; discovery and acquisition; management and organization; conversion and interoperability; quality assurance; metadata; curation and re-use; cultures of practice; preservation; analysis; visualization, and ethics.

Although there are various terms, definitions, and conceptualizations of data management skills and competencies, researchers have to take a stand on generic data management operations, such as collecting, quality assurance, storing, documenting, organizing, processing, analyzing, preserving, and sharing data when planning their research projects (Strasser, Cook, Michener, and Budden, 2012). The exact way these RDM operations and practices are carried out varies between disciplines because of different research processes and methods (Coates, 2014; Molloy and Snow, 2012; Weller and Monroe-Gulick, 2014). Moreover, the execution of RDM operations depends also on the types of research and data (Lefebvre, Schermerhorn, and Spruit, 2018; Scholtens et al., 2019).

High quality data management practices demand data management education (Koltay, 2019). Data (information) literacy training is seen as a way to educate data fluent researchers and students (Sapp Nelson, 2017; Schneider, 2013). In the crosswalk exercise through five different competency frameworks, Sapp Nelson (2017) indicated Carlson et al.’s (2011) competency framework with 12 DIL competencies as the most comprehensive so far². Although the DIL framework is founded on information collected from the research faculty and students of the Earth and Atmospheric Sciences,

¹ FAIR Principles: <https://www.go-fair.org/fair-principles/>

² Besides RDM competency frameworks aimed at researchers and students there are extensive domain-agnostic frameworks like **EOSC**, **DAMA**, and **EDISON** focusing especially on the data stewardship level RDM and data science competencies and capabilities.

it incorporates the central, generic phases of the data lifecycle with discipline-agnostic descriptions of the skills and competencies needed in those phases. Thus, according to Carlson et al. (2011), DIL competencies can be adapted and modified to other disciplines as well. As a starting point, we base our study on these 12 DIL competencies. Besides covering most of the data lifecycle phases on a generic level, we found the framework to be a sound basis for developing a modular data management course structure in collaboration with many stakeholders of RDM.

However, to emphasize the skills and competencies in the focus of the study versus an inquiry into the broader issue of data (information) literacy, we prefer to use the terms “RDM skills” and “RDM competencies” instead of “DIL literacy” in this article. Furthermore, when we have conducted interviews and surveys, and have planned and developed data management services and training together with different RDM stakeholder groups like faculty researchers and research support specialists, we noticed that they understand, accept, and adopt the terms “research data management” (RDM) and “RDM competencies” better than “data information literacy” or “DIL competencies”.

The Aim of the Study

According to research literature, there seems to be inconsistency between the high importance of RDM often perceived by researchers and graduate students and the actual highly varied RDM practices and skills in everyday research work (Carlson, Johnston, and Westra, 2015; Coates, 2014; Gabridge, 2009; Jahnke, Asher, and Keralis, 2012; Tenopir, Birch, and Allard, 2012; Thielen, Samuel, Carlson, and Moldwin, 2017). The aim of this study is to investigate doctoral students’ current research data management competencies, as self-rated by doctoral students (DSs) and as rated by faculty members (FMs), as well as the importance of these competencies as rated by DSs and FMs on the five-point Likert-like scale. The quantitative analysis of the ratings is complemented by a qualitative content analysis of open-ended answers and additional information obtained from answers to close-ended questions provided by the study participants. Moreover, based on the potential gap between the average level of rated importance of the competencies and the average level of rated or self-rated DSs’ current competencies, the aim is to find the educational RDM needs of DSs. This study will help answer the following RQs:

- [RQ-1] How important are RDM competencies as rated by doctoral students (DSs) and faculty members (FMs)?
- [RQ-2] How did FMs rate and DSs self-rate DSs’ current competencies?
- [RQ-3] What kinds of educational RDM needs do doctoral students have?

The analysis is founded on a structured interview study at the University of Turku (UTU).

Literature Review

Importance of RDM Competencies

Earlier studies note that researchers and students have often considered RDM skills and competencies highly important (Carlson, Jeffryes, Johnston, Nichols, Westra, and Wright, 2015; Parsons, Grimshaw, and Williamson, 2013; Pouchard and Bracke, 2016; Qin and D'Ignazio, 2010; Thielen et al., 2017; Wilson, Jeffreys, Patrick, Rumsey, and Jefferies, 2015), although there also have been found some differences concerning the perceived importance of competencies according to disciplines and preferred methodologies (Akers and Doty, 2013; Joo and Peters, 2019; Weller and Monroe-Gulick, 2014). Most of the competencies are discipline-agnostic (Frugoli, Etgen, and Kuhar, 2010; Molloy and Snow, 2012; Specht et al., 2015). Competencies that are closely related to producing research results, such as data processing, analyzing, and visualizing, are considered the most important (Joo and Peters, 2019; Parsons, Grimshaw, and Williamson, 2013). Additionally, organizing, documenting, and describing of data are perceived as very important (Knight, 2013; Qin and D'Ignazio, 2010). Likewise, legal, ethical, and data security competencies are viewed as very important or important (Carlson, Jeffryes et al., 2015; Knight, 2013). In the research literature, respondents and interviewees are typically asked to rate the importance of RDM competencies, not the current RDM competencies respondents or interviewees believe they or their students have. However, in a recent survey by Pasek and Mayer (2019), graduate students and faculty members were asked to rate the knowledge and skill levels for the 12 DIL competencies of graduate students. On a five-point scale, respondents rated graduate students' "ethics and attribution" competencies as highest, between three and four, and most of the other competencies like "cultures of practice"; "data visualization"; "quality assurance"; and "data processing and analysis" between two and three. Although graduate students' self-ratings were higher than faculty members' ratings, the authors found the difference statistically insignificant.

It has been well documented in the research literature that graduate students usually have little or no formal education about RDM (Adamick, Reznik-Zellen, and Sheridan, 2013; Alexogiannopoulos, McKenney, and Pickton, 2010; Carlson, Jeffryes et al., 2015; Goben and Griffin, 2019; Griffin, 2020; Johnston and Jeffryes, 2014; Krahe, Toohey, Wolski, Scuffham, and Reilly, 2020; Maienschein, Parker, Laubichler, and Hackett, 2019; Molloy and Snow, 2012; Parsons et al., 2013; Peters and Vaughn, 2014; Shorish, 2015; Wiley and Kerby, 2018). Consequently, graduate students' RDM competencies and skills vary greatly depending on their earlier education and experience. A claim presented by some supervisors is that RDM competencies of graduate students are somewhat superficial, and that they have a limited picture of the research data lifecycle and the importance of data as part of research process (Carlson, Jeffryes, et al., 2015). Because of the missing RDM education, students develop their RDM practices ad hoc, through trial and error (Thielen and Hess, 2017; Wright and Andrews, 2015). In the research community, data has not been considered as valuable as articles are from the researcher's career perspective. This has resulted in omission of long-term RDM planning, and in non-standardized practices: data processing is seldom properly documented and described; storage platforms are heterogeneous; institutionally recommended platforms may be unknown to researchers, or researchers and students distrust them; data formats and types may be outdated or non-standard; version

management and file naming practices are miscellaneous; and data ownership and funders' mandates are unclear (Andrews Mancilla et al., 2019; Humanities Advanced Technology and Information Institute (HATII), 2009; Knight, 2013; Krahe et al., 2020; Parsons et al., 2013; Piorun et al., 2012).

In the UK, RDM practices in higher education institutions have been evaluated by several Data Asset Framework (DAF) surveys (Humanities Advanced Technology and Information Institute, 2009). When asked about educational needs of the researchers and students, respondents mentioned needs for training and support for making DMPs; where and how to store, preserve and share data; the creation of metadata and documentation; to become familiar with funders' mandates; managing sensitive data; and intellectual property rights (IPR) issues and citing data (Knight, 2013; Parsons et al., 2013). Moreover, outside the list of the DAF survey form, students and researchers requested support and training for more technical oriented RDM needs like backup of large data sets; storing platforms of sensitive data; shared servers for inter-organizational collaborative research projects; building, collecting and organizing data in relational databases; and analyzing data (Knight, 2013; Parsons et al., 2013).

The Role of Library and Other Research Support Units

Research data can be seen as part of an information ecosystem and thus as an information source (Carlson, 2015). Because the primeval goal of libraries has been to connect information sources and persons in need of information, it can be seen as justified that library takes a central role in supporting good RDM practices, encouraging, guiding and helping researchers in discovering, planning, curation, sharing, re-use, and preservation of research data. Library staff can teach how to find and use external repositories, how to cite and make data sets citable and show researchers how to get credit for sharing their data (Mannheimer, Sterman, and Borda, 2016; Prado and Marzal, 2013).

According to several surveys on research data management services (RDSs), academic libraries have concentrated mainly on consultative, informational RDSs, e.g. supporting writing of data management plans, training researchers, and making RDM instructions, whereas fewer libraries are serving, or planning to serve technical, "hands-on" RDM services like maintaining repositories and preparing data for deposit into it, creating metadata, storing, and preserving data sets (Cox et al., 2017; Federer, 2017; Joo and Peters, 2019; Tenopir et al., 2017). In some universities, the library already helps researchers with "hands-on" RDM services, e.g. to negotiate to access closed data sets, licensing data sets, fixing data sets, visualizing data, etc. (Federer, 2016).

Though most often RDSs are led by the library or research office, or these two units together, producing RDSs for the whole research data lifecycle is a multi-unit task which requires diversified knowledge (Cox et al., 2017). If the library, research office, research IT, legal services, and academic contributors can co-operate, plan and produce RDSs together for researchers and students, there is no need for library to carry the lion's share of supporting RDM. Research support networks should combine efforts in producing comprehensive RDSs (Castle, 2019; Revez, 2018; Verbaan and Cox, 2014; Yu, 2017).

Research Methods

Structured Interviews

The data collection method was a quantitative structured interview or survey interview (Groves et al., 2009; Gubrium and Holstein, 2011; Leinonen, Otonkorpi-Lehtoranta, and Heiskanen, 2017; Mittenfelner and Ravitch, 2018) in which all the respondents responded to close-ended and open-ended predetermined questions in the same order by completing an online questionnaire form together with the interviewer during the interview.

A survey interview method was chosen instead of conventional survey study because we knew from previous research (e.g. Carlson, Johnston, Westra, and Nichols, 2013) that RDM and relative concepts are possibly unfamiliar to many researchers and that it can be important to discuss the meaning of different terms to get valid answers to our questions.

In addition to Likert-like and other close-ended questions, respondents answered predetermined open-ended questions, further defining or enlarging their answers to fixed-choice questions and, depending on their answers, to predetermined follow-up questions in which they were asked to justify, specify, or enlarge their answers. A primarily quantitative data collection method was selected because this study will focus on the Likert-like ratings of the interviews. Additionally, we used the answers to other close-ended and open-ended questions as sources to provide complementary insights and illustrations for quantitative ratings (Bryman, 2006; Greene, Caracelli, and Graham, 1989). The data collection and its preliminary analysis have been presented before by Rantasaari and Kokkinen (2019a, 2019b).

The interview questionnaire form³ was adapted and modified from the DIL Interview Toolkit⁴ by Carlson et al. (2013), which is based on the structure of the Data Curation Profiles Toolkit⁵.

The author of this study (the head librarian for research services) acted as principal interviewer. Besides a respondent and the principal interviewer the following were involved in interviews when possible: The head librarian for learning services, the head of research IT, and a data librarian. The notes were taken by the two head librarians separately. After each interview the principal interviewer proofread and combined the notes with the interview questionnaire form completed by a respondent in the interview session. In five interview sessions there was only one interviewer (the author) taking the notes. After these five interviews the interviewer sent the notes to the respondent for approval. Attendance from both the library and the research IT in planning and implementing the interviews was fruitful firstly by bringing wider data management expertise to the interviews and secondly because of the responsibility of these two units to develop and maintain RDSs and data infrastructure at UTU. The average length of each interview session was two hours. In total, we carried out 35 interviews.

The competence areas covered in the structured interviews were:

- Discovery and acquisition of data
- Databases and data formats

³ See: <https://doi.org/10.5281/zenodo.4429480>

⁴ DIL Interview Toolkit: <https://docs.lib.purdue.edu/dilsymposium/2013/interviewinstruments/1/>

⁵ Data Curation Profiles Toolkit: <https://docs.lib.purdue.edu/dcptoolkit/>

- Data conversion and interoperability
- Data quality and documentation
- Data management and organization
- Data processing and analyzing
- Data visualization and presentation
- Metadata and data description
- Cultures of practice
- Data curation and re-use
- Ethics and attribution
- Data preservation

Interview questionnaire forms

The original Interview Worksheet by Carlson et al.⁶ includes nine question categories for faculty and ten categories for graduate students. Some of the questions are Likert-like and other fixed-choice questions, and some are open-ended. The Toolkit also includes instructions on how to conduct the interview and additional questions to be asked.

In this study, some changes were made to the interview forms (Rantasaari, 2020):

- “Research data management” (RDM) competence was used instead of “data information literacy” (DIL).
- Besides Likert-like ratings of the importance of each competence in the original form, we added Likert-like ratings of the DSs’ current competencies as rated by FMs and self-rated by DSs.
- An informed consent and explanation about the use of gathered information was added. We also added information about data privacy. Confidentiality of the study was explained.
- The question concerning agreements, permits and licenses was added in the interview form for FMs, because we wanted to know what kind of agreements, permits, and licenses have been made in the projects, whether they allow long-term preservation and sharing of data, and the roles of faculty vs. doctoral students in preparing them.
- Instead of using a print form, the questionnaire was implemented as an online form developed in the Webropol⁷ -software.

⁶ DIL Toolkit, Interview Worksheet: <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1058&context=dilsymposium>

⁷ Webropol survey and reporting tool: <https://webropol.com>

Respondents

To gain answers to our research questions we conducted structured interviews with 15 DSs and 20 FMs in 34 interview sessions during the spring and fall of 2018 and the spring of 2019 (Table 1 in Appendix A). FMs were mainly doctoral supervisors, but we also interviewed four biostatisticians of a medical faculty who process and analyze research data together with doctoral students and, thus, have a good vantage point of doctoral students' data management practices.

As is indicated (Table 1 in Appendix A), the number of respondents from different disciplines varies. In the faculties of Medicine, Social Sciences, and Education, the share of respondents is about equal to the share of these faculties' doctoral students in relation to all doctoral students at the UTU; in Science and Engineering, and in the Turku School of Economics, the share is bigger; in Humanities it is smaller; and there are no respondents from the Faculty of Law. We chose the respondents first for the expected data intensity in their discipline and, second, to obtain information about the management with different types of data and data sources. According to the Community Capability Model Framework (CCMF), the necessary characteristics for data-intensive research are intense computational analysis of data, analysis of large quantities of data using specific software, and combining data from different sources for re-use (Lyon, Ball, Duke, and Day, 2012). Still, to contrast different research contexts, the sample includes some researchers from non-data-intensive disciplines – for example, theoretical physics.

Analysis

The importance of the competencies and DSs' current management of the 12 RDM competencies developed by Carlson et al. (2011) were rated by FMs and DSs on a Likert-like scale (Table 1).

We analyzed the Likert-like scale ratings of the respondents and possible differences of the ratings and self-ratings between FMs and DSs using JMP Pro 15 to produce descriptive statistics with a two sample t-test (assuming equal variances). The paired t-test or Wilcoxon signed rank test (depending of the current distribution) was used to compare the ratings of the competencies' importance vs. the ratings or self-ratings of DSs' current competencies. A significance level of 0.05 (two-tailed) was used. Additionally, we categorized and coded the open-ended answers in NVivo 12, and we conducted qualitative content analysis of the data of these answers to help us understand and interpret the Likert-like ratings or self-ratings. Moreover, numerical data from fixed-choice answers were used to inform the ratings and open-ended answers.

Table 1. The 12 RDM competencies and the five-point Likert-like scale for importance of RDM and current RDM competencies of DSs

12 RDM (DIL) competencies	Five-point Likert-like scale
discovery and acquisition of data	1 = “not important”
	1 = “do not have competence”

12 RDM (DIL) competencies	Five-point Likert-like scale
databases and formats	2 = "somewhat important"
	2 = "have some competence"
data conversion and interoperability	3 = "important"
	3 = "have good competence"
data processing and analysis	4 = "very important"
	4 = "have very good competence"
data visualization and representation	5 = "essential"
	5 = "have ultimate competence"
data management and organization	
data quality and documentation	
metadata and data description	
cultures of practice	
ethics and attribution	
data curation and re-use	
data preservation	

Results

Quantitative Results

The mean of the perceived importance of RDM competencies was 4.07 in the five-point Likert-like scale as rated by FMs and DSs. The mean of DSs' perceived current RDM competencies as rated by FMs and as self-rated by DSs was 2.47. The difference between perceived importance and DSs' current competencies is statistically highly significant ($p < 0.0001$) with all of the twelve competencies (Figure 1; Table 1 in Appendix B).

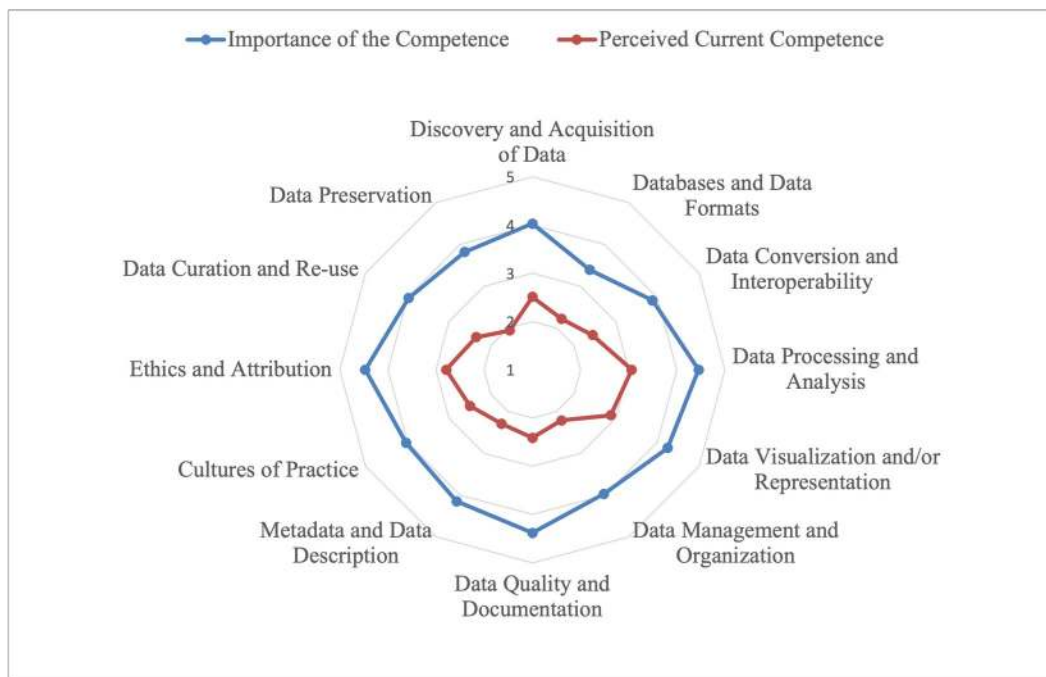


Figure 1. The importance of RDM competencies vs. doctoral students’ current RDM competencies as rated/self-rated by all respondents.

The difference between FMs’ and DSs’ ratings on the importance of RDM was statistically insignificant ($p=0.33$). The mean rating of FMs was 3.98, whereas the mean rating of DSs was 4.17 (Figure 2; Table 2 in Appendix B). The only statistically significant difference concerned “data preservation” ($p=0.02$), in which FMs rated the importance as 3.42 (mean), and DSs rated the importance as 4.33 (mean).

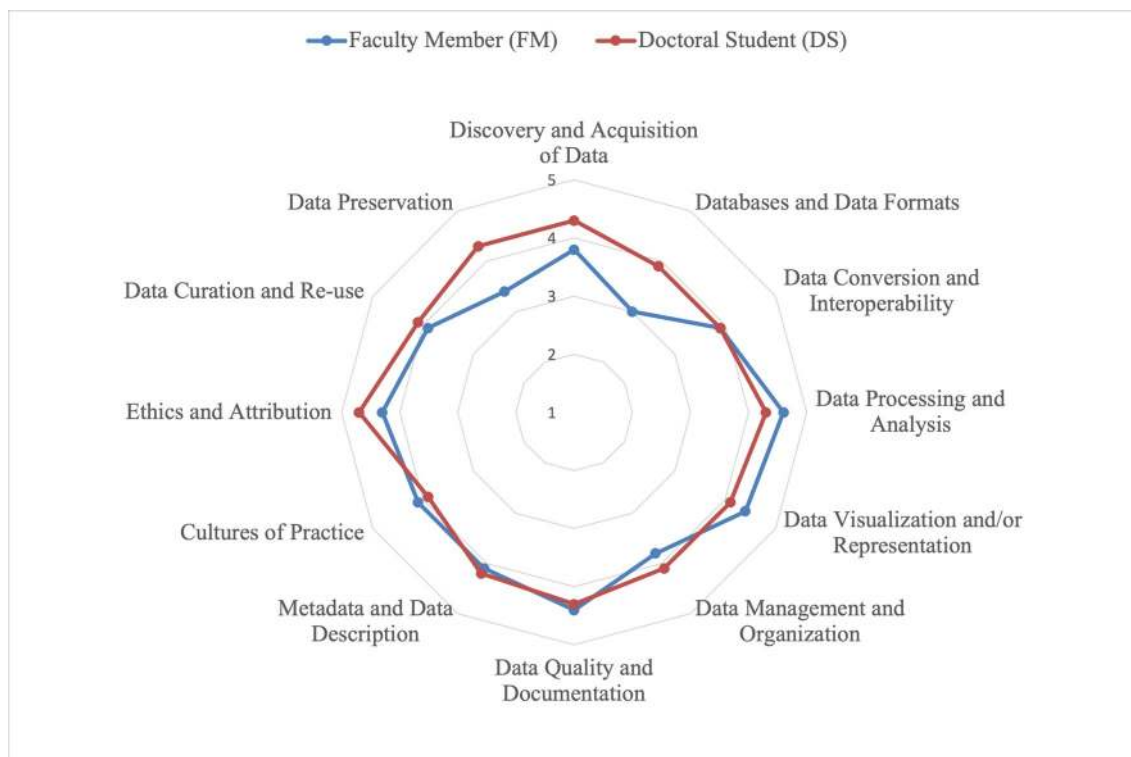


Figure 2. The importance of RDM competencies as rated separately by faculty members (FMs) and doctoral students (DSs).

The difference between FMs' ratings and DSs' self-ratings of DSs' current RDM competencies was statistically significant ($p=0.0018$). DSs' average self-rating was 2.82, whereas FMs' average rating was 2.19 (Figure 3; Table 3 in Appendix B). Both groups rated or self-rated DSs' competencies as best with “data processing and analysis”, 2.94 (FMs) and 3.20 (DSs), “data visualization and presentation”, 2.74 (FMs) and 3.07 (DSs), and “ethics and attribution”, 2.37 (FMs) and 3.33 (DSs). Statistically, the most significant differences were in “data preservation” ($p<0.0001$), “data curation and re-use” ($p<0.0001$), and “ethics and attribution” ($p<0.0001$). In these three competencies the difference between the ratings of FMs and the self-ratings of DSs was approximately one point in the five-point Likert-like scale: FMs rated DSs current competencies from 1.42 to 2.37, and DSs self-rated their competencies from 2.6 to 3.33.

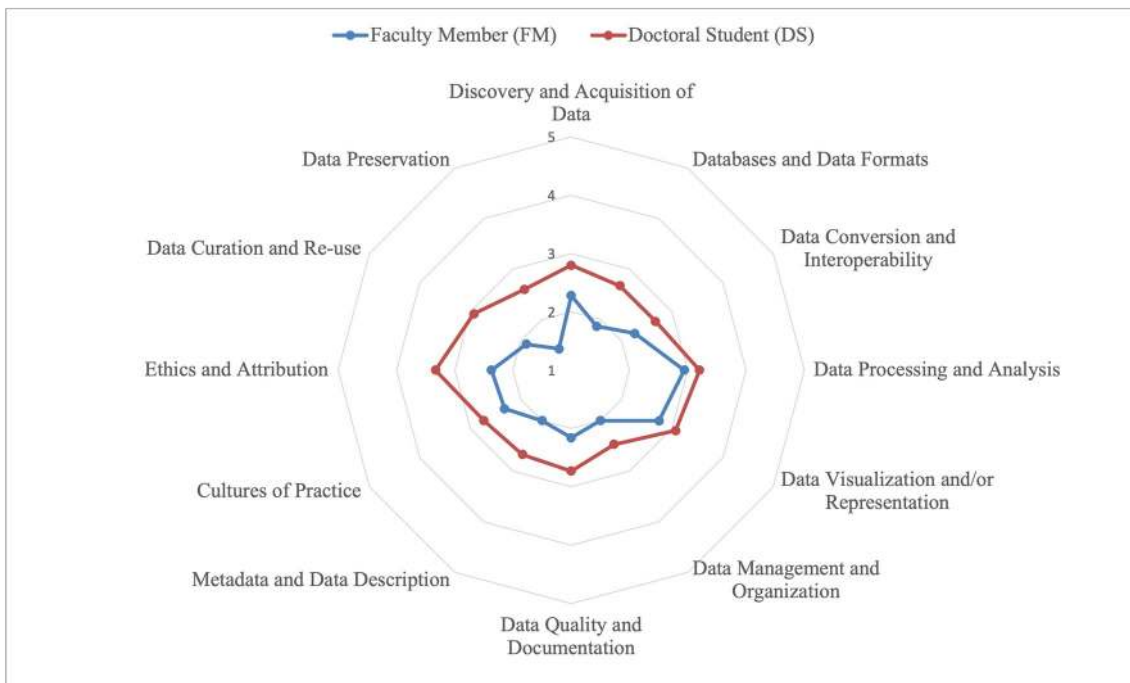


Figure 3. DSs' current RDM competencies as rated by FMs and self-rated by DSs.

Detailed Analysis of Selected Competencies

Next, we will scrutinize the five RDM competencies selected based on the biggest difference between the rated importance and DSs' current rated/self-rated competence. Besides using the quantitative analysis of the Likert-like scale ratings and other fixed-choice answers, we have used the qualitative content analysis of the data gathered from the study participants' open-ended answers. The wider the gap between the perceived importance and current competence, the stronger is the evidence of the educational need.

Data quality and documentation

The gap between the rated importance (4.38) and DSs' rated or self-rated current competence (2.41) was 1.97 points. When FMs rated DSs' competence as “have some” (2.16), DSs self-rated their competence close to “have good competence” (2.73). However, judging by DSs' comments and answers to open-ended questions it seems that

documentation is mainly aimed at the DSs themselves. “Documentation is made for myself to follow my data during the project.” (Doctoral student, Turku School of Economics). Still, 80 percent (12) of responding DSs felt that their documentation was good enough for someone outside the project to understand and use their data. By contrast, only 15 percent (3) of FMs agreed that DSs’ documentation enable outsiders to understand and use the data. “Most DSs don’t document because they don’t think anyone else would use their data after the current project” (Biostatistician, Faculty of Medicine). FMs saw skills gaps, especially in DSs’ RDM planning, data structure, documentation of revisions, coding of variables and labels, and cross-checking of data. A faculty member from the Faculty of Medicine stated that because the need for sound documentation and data management has not been found to be intrinsic but comes instead from policymakers, statisticians, and data analysts, the level of documentation has remained weak.

Data preservation

The gap between the rated importance (3.82) and the DSs’ rated or self-rated current competence (1.94) was 1.88 points. On average, DSs saw data preservation as significantly more important (4.33) than FMs (3.42). However, FMs from HSS disciplines rated the competence as “very important” (4.14), whereas FMs from STEM disciplines rated it as “important” (3.00). In HSS disciplines, especially in history, culture, and arts, and in some social sciences disciplines such as in economic sociology, there is an established long-term preservation culture. “The value of the collected and produced materials not only for the researcher her/himself but also for others were understood when the cultural studies’ practices were formerly created. Data were seen as important in building national identity” (Supervisor, Faculty of Humanities).

DSs’ self-rating (2.6) was significantly higher than FMs’ rating (1.42) of DSs’ average current competencies. Many doctoral students seemed to have a positive attitude toward long-term preservation of data: for the question, “How long would your data set be useful or have value for you or others if it were to be preserved,” 73 percent of doctoral students (11) answered “indefinitely,” or “from 50 to 100 years”. Still, they typically had not taken long-term preservation into account (or did not know if it had been taken into account) in agreements concerning their project data. Additionally, they often had not documented and described data looking ahead at possible future users. “What is needed is to have straightforward but not compulsory instruction in how to manage qualitative data during and after the research project not only from a technical but also from an ethical viewpoint” (Doctoral student, Turku School of Economics). Many faculty members admitted that there is no formal education or instructions on long-term data preservation, and it has not been considered in the contracts. “Researchers’ focus is here and now, and they don’t pay so much attention to re-use and long-term preservation issues” (Supervisor, Faculty of Science and Engineering).

Metadata and data description

The gap between the rated importance (4.15) and DSs’ rated or self-rated current competence (2.29) was 1.86 points. DSs self-rated their competence closer to “good” (2.67) whereas supervisors rated it as “have some competence” (2.00). There was also a difference in the ratings of DSs’ current competence between STEM and HSS disciplines’ FMs: STEM FMs (12) rated DSs’ current competencies as “have some competence” (2.25), whereas HSS FMs (7) rated them between “do not have” and “have some competence” (1.57). Many FMs from different disciplines emphasized the importance of describing and documenting data, but especially FMs from HSS

disciplines commented that there are no formal trainings and standards for description and documentation. “In principle there has not been a thought that anyone other than researchers themselves would use their focus group interview data. As far as qualitative research data are concerned, there is not that kind of culture [of preserving data] as there is of preserving quantitative research data.” (Supervisor, Faculty of Social Sciences). Some FMs were also skeptical about DSs’ understanding of the importance of administrative or descriptive metadata that are external to informational content of data and that help to use the data; they also believed that DSs are more competent in structural or technical metadata (how the data components are organized and named). Some supervisors supposed that DSs think metadata are more important to others than to themselves, and therefore, DSs do not spend time describing data.

Data management and organization

The gap between the rated importance (3.97) and DSs’ rated or self-rated current competence (2.21) was 1.76 points. Some FMs commented that many DSs are missing an overall perspective of data and its lifecycle in a research project. “It would be important to have a big picture of the data and its relevance to understand the importance of preservation and re-use” (Supervisor, Faculty of Humanities). However, in many projects, DSs obtain ready-made data collected by others, and they do not necessarily have to plan data management and organization as a unity. Typically, there are no standardized RDM procedures. Practical decisions on data management and organization are often the DSs’ own responsibility. “It would have been a huge benefit if there had been some training on data management” (Doctoral student, Faculty of Social Sciences). Despite the highly rated importance, research project directors also have only recently begun to pay attention to the need for standardized data management and organization practices, mainly because of the increased amount of digital data and data platforms and collaborative research projects.

Ethics and attribution

The gap between the rated importance (4.47) and DSs’ rated or self-rated current competence (2.79) was 1.68 points. There was also a highly significant difference between DSs’ average self-rating of 3.33 (“good competence”) and FMs’ average rating of 2.37 (“some competence”). According to some FMs from the Turku School of Economics, DSs have learned to handle sensitive data fairly well. On the other hand, one FM from the Faculty of Social Sciences commented that DSs have problems in coordinating data openness and privacy. Moreover, based on the DSs’ answers to open-ended questions concerning data ownership and intellectual property rights of the research projects’ data, these issues seemed to be unclear to almost all DSs. “Everything that has something to do with the letter of law is unclear and scary” (Supervisor, Faculty of Social Sciences).

Overview of the Results

This study focused on the following research questions: How important are RDM competencies as rated by DSs and FMs? How did FMs rate and DSs self-rate DSs’ current competencies? What kinds of educational RDM needs do DSs have?

How important are RDM competencies as rated by DSs and FMs?

On average, all the competencies were rated as “very important” or close to it (3.82–4.47) by all respondents, with the exception of “databases and data formats,” which was

rated closer to “important” (3.39). DSs rated all the competencies as “very important” or close to it (3.87–4.33). With the exception of “data preservation” (3.42) and “databases and data formats” (3.00), FMs rated all the competencies as “very important” or close to it (3.83–4.61) as well.

How did FMs rate and DSs self-rate DSs’ current competencies?

Regarding the current competencies of DSs, only “data processing and analysis” were rated or self-rated on average as “have good competence” (3.06), while all the other competencies were rated between “have some competence” and “have good competence” (1.94–2.88) by all respondents. DSs’ self-ratings of their current competencies were higher than FMs’ ratings of DSs’ current competencies all along the line, and the difference was statistically significant. On average, FMs did not rate any of the DSs’ current competencies as “good,” although “data processing and analysis” (2.94) were close to it. FMs rated DSs’ current competencies as “have some competence” (2.00–2.94), except for “databases and data formats,” “curation and re-use,” and “data preservation,” which they rated between “do not have competence” and “have some competence” (1.42–1.89). DSs self-rated their current “data processing and analysis,” “visualization and presentation,” and “ethics and attribution” competencies as “have good competence” (3.07–3.33); they rated all other current competencies between “have some competence” and “have good competence” (2.60–2.93). Some differences in the ratings or self-ratings between respondents of STEM and HSS disciplines were also found. However, due to the small number of respondents in this study, these differences have no statistical relevance.

What kinds of educational RDM needs do DSs have?

We found a highly statistically significant difference between DSs’ current rated/self-rated RDM competencies and the rated importance of these competencies (Table 1, Appendix B). Though the difference is between one and two Likert-like scale points with all the competencies, signifying that DSs have educational needs in all RDM competencies, the gap is widest in “data quality and documentation”, “data preservation”, “metadata and data description”, “data management and organization”, and “ethics and attribution”.

Discussion

The High Importance of RDM Competencies

The average rating of RDM competencies as “very important” (4.06 in the five-point scale from one to five) by all respondents is in line with the results of several previous studies in which the importance of RDM competencies were reported as important or very important (Carlson et al., 2013; Pasek and Mayer, 2019; Pouchard and Bracke, 2016; Qin and D’Ignazio, 2010; Thielen et al., 2017; Wilson et al., 2015). In the study by Carlson, Jeffryes, et al. (2015), with the same 12 competencies, the importance of all competencies was rated between 3.8 and 4.44 in the five-point scale.

The General Level Difference Between Perceived Importance and DSs' Current Competencies

Regarding the whole research data lifecycle, we found that the difference between perceptions of DSs' current RDM competencies as “have some competence” (2.47 in the scale from 1 to 5) and the perceived importance of these competencies as “very important” (4.06) had high statistical significance ($p < 0.0001$). A probable reason for the gap is missing or minor training of skills. A substantial amount of literature exists on graduate students' little or missing formal education of RDM, entailing high variation in RDM competencies when students learn the skills ad hoc by trial and error (Adamick et al., 2013; Alexogiannopoulos, McKenney, and Pickton, 2010; Carlson, Jeffryes, et al., 2015; Johnston and Jeffryes, 2014; Krahe et al., 2014; Maienschein et al., 2019; Molloy and Snow, 2012; Parsons et al., 2013; Peters and Vaughn, 2014; Shorish, 2015; Wiley and Kerby, 2018). Standard deviations of the self-ratings of DSs' current RDM competencies were larger (0.64–1.23) than FMs' ratings of DSs' RDM competencies (0.6–0.89), which may stem from a high variation of skills among DSs (Table 3, Appendix B). It is noteworthy that the difference between FMs' and DSs' ratings of the importance of RDM competencies was insignificant ($p = 0.33$), but the difference between FMs' ratings and DSs' self-ratings of DSs' current competencies was significant ($p = 0.0018$). On average, DSs self-rated their competencies close to “have good competencies” (2.82) when FMs rated DSs' competencies slightly higher than “have some competencies” (2.19). Though there are very few studies comparing faculty members' and graduate students' views of graduate students' RDM competencies, the finding is in line with the survey of Pasek and Mayer (2019), who found that in the science-based disciplines, graduate students' self-assessments were higher than faculty ratings. However, Pasek and Mayer found that the difference was statistically insignificant. Possible reasons for the differences in statistical significance between the study of Pasek and Mayer and this study can stem from different research subjects (graduate students vs. doctoral students), methods (survey vs. structured interview), number of respondents (210 vs. 34), and selected disciplines (science-based disciplines vs. STEM and HSS disciplines). In this study, it is possible that DSs overrated their own competencies. The possible overrating could be due to the lack of knowledge of RDM competencies, though we tried to clarify RDM and each competency before respondents answered the questions by filling in the questionnaire during the interviews. Another reason for over-estimating competencies could be that if one does not know what they do not know, there is a great tendency that they overestimate their abilities (Kruger and Dunning, 1999). DSs also could have given disproportionately high ratings because they were uncertain about the skills they thought the interviewers expected them to have. This “interviewer bias” is a widely discussed factor in the research literature (Groves et al., 2009; Waterfield, 2018).

Disparities in Perceptions about Specific Competencies

On average, FMs rated DSs' current competencies as “have some competence” or below. The competencies they rated between “no competence” and “have some competence” are “data curation and re-use” (1.89) and “data preservation” (1.42). DSs self-rated their current competencies on average between “have some competence” and “have good competence” or higher. The competencies they self-rated as “good” or higher were “ethics and attribution” (3.33), “data processing and analysis” (3.20), and “data visualization and presentation” (3.07). The biggest differences between FMs'

ratings and DSs' self-ratings of DSs' current competencies were in “data preservation” (1.18 points), “data curation and re-use” (1.04), and “ethics and attribution” (0.96) (Table 3, Appendix B).

The gap between the perceptions of DSs' current RDM competencies and the perceived importance of these competencies is widest, at nearly two points in the five-point Likert-like scale, in “data quality and documentation” (1.97 points), “data preservation” (1.88), and “metadata and data description” (1.86). Especially FMs from HSS disciplines rated DSs' “metadata and data description” competencies low (1.57), between “no competence” and “some competence.” A possible reason for the low rating could be that in HSS disciplines in which qualitative methods are more prevalent than in STEM disciplines, the software and infrastructure used for analysis often do not automatically produce metadata and descriptive information. Secondly, according to some FMs from the Faculty of Education and the Turku School of Economics, there is no prevailing culture for standardized documentation, description, and long-term preservation of qualitative interview data in their disciplines. This finding is in line with Tenopir et al. (2015), who found that “those who work with human subject data were more likely to use no metadata to describe their datasets,” and researchers working especially with medicine/health science, business, social sciences, and psychology did not perceive access to others' data as essential for science.

The perceived high importance of documentation, description, and preservation of data by study participants does not necessarily seem to be conveyed to a practical level. Although especially FMs stressed the importance of data and systematic data management as part of conducting high-quality research, respondents noted being busy with more immediate research practices like collecting, processing, and analyzing data and publishing research results as reasons why only little attention was paid to other RDM practices such as documentation, description, and preservation of data. The finding is supported by several studies in which researchers have expressed a lack of time and money to improve the documentation of their data for long-term preservation, sharing, and re-use (Read, Larson, Gillespie, Oh, and Surkis, 2019; Tenopir et al., 2015; Yu, Deuble, and Morgan, 2017). The documentation is typically made for the researcher themselves to get the project through. Up to 73 percent (11) of DSs did not know whether there are standards for organizing, documenting, and describing data in their disciplines. This finding parallels Schumacher and VandeCreek's (2015) study, in which faculty members were largely unaware of basic data management principles.

Competencies that are most closely related or elementary to producing research results such as data processing, analyzing, and visualizing, have usually been considered most important (Carlson, Jeffryes et al., 2015; Joo and Peters, 2019; Parsons et al., 2013). Additionally, legal and ethical considerations have been rated as very important (Johnston and Jeffryes, 2014; Pasek and Mayer, 2019; Thielen et al., 2017). In this study, the importance of these competencies was rated between 4.2 and 4.5 in the five-point Likert-like scale. When we asked the respondents to rate or self-rate the level of DSs' current competencies, study participants generally rated DSs' skills highest in “data processing and analysis” (3.06), “data visualization and representation” (2.88), and “ethics and attribution” (2.79). Differing from the ratings or self-ratings of most other RDM competencies by all respondents, STEM disciplines' DSs and FMs rated DSs' processing, analyzing, and visualizing competencies as “good” or close to good (2.67–3.08). Interestingly, DSs of HSS disciplines (6) self-rated their “processing and analyzing” as “very good” (4.0) and “visualization and presentation” between “good” and “very good” (3.5), while FMs of HSS disciplines (7) rated DSs “visualization and presentation” competencies only as “have some” (2.14) and “data processing and

analysis” competencies as nearly “good” (2.94). Judging by the comments of some HSS disciplines’ DSs, who said that data visualization in their discipline is not so important in qualitative research and is given little attention, it is possible that HSS disciplines’ DSs overestimated their competencies. Self-rating the competencies of “data processing and analysis” as “very good” by HSS DSs could possibly stem from students’ thinking that qualitative methods, such as content analysis, are somehow more straightforward to learn and possess than the quantitative methods used more in STEM disciplines. This could point in the same direction as the surveys of Weller and Monroe-Gulick (2014) and Joo and Peters (2019) – who found that the need for assistance with data analysis was expressed by researchers who conducted quantitative, qualitative statistical, and experimental research – as well as health scientists and social scientists, but much less by humanities researchers for any kind of data analysis.

Concerning “ethics and attribution,” only 40 percent of DSs said they had participated in an ethics basics course or data privacy lecture, although ethics is a mandatory subject at the moment at UTU. Besides, most DSs were unaware of the agreements and of the owner of the data produced and used in the projects in which they were involved. The finding is in line with the results of the survey of Andrews Mancilla et al. (2019) who found that among the academic researchers PhD candidates appeared to be the least aware of data ownership. In our study DSs were usually not responsible for managing legal and ethical issues like agreements, permits, and privacy notices in projects in which they were not the principal investigator, which was typical especially in STEM disciplines. Finally, judging from the widely used but not safe and secure data storage platforms such as laptops, Dropbox, and external hard drives, data privacy practices are not necessarily as good as the DSs self-rating would suggest. Hence, FMs’ rating of DSs’ current competence as “have some competence” (2.37) in “ethics and attribution” seems more realistic than DSs’ average self-rating of having “good competence” (3.33) (Table 3, Appendix B).

The competencies that have been rated as slightly less important, between “important” and “very important” – like “databases and data formats” (3.39), “data preservation” (3.82), and “data conversion and interoperability” (3.89) – are also close to the ratings of the studies of Carlson, Jeffryes, et al. (2015) and Pasek and Mayer (2019). Yet, rating the competencies as “important” does not seem to reflect the DSs’ current data practices, considering their minimum experience using databases as data organizing tools and their minor experience depositing and discovering data from repositories. A possible reason for rating the data preservation as somewhat lower in importance than most other competencies by FMs (3.42) could be that long-term preservation and sharing of data is still not a prevailing practice in many disciplines. However, unlike STEM disciplines’ FMs (12), who rated the importance of long-term preservation as “important” (3.0), FMs of HSS disciplines (6) rated it as “very important” (4.14). This may be because there are an established data preservation culture, practices, and repositories, especially in humanities, where the Archives of History, Culture, and Arts Studies of UTU⁸ are much used, as well as in some social science disciplines such as economic sociology, which uses the Finnish Social Science Data Archive⁹ to deposit and discover data. According to many FMs, the researchers’ primary interest is typically to get their current project through and to obtain results from the data, not necessarily long-term preservation and possible re-use of data in future projects (Kowalczyk, 2018). Concerning the significantly different rating of the

⁸ Archives of History, Culture, and Arts Studies of UTU: <https://www.utu.fi/en/university/faculty-of-humanities/school-of-history-culture-and-arts-studies/services>

⁹ Finnish Social Science Data Archive: <https://www.fsd.tuni.fi/en/>

importance of “data preservation” ($p=0.02$) between FMs and DSs, some possible reasons for the higher appreciation of data preservation by DSs (4.33) compared to FMs (3.42) could be that, for many DSs, this is probably their first research project, whereas for FMs, this is one project among many. Secondly, as novice researchers without comprehensive data sets of their own, DSs could benefit more than FMs from qualified data sets collected, curated, and preserved by other researchers in repositories. Based on DSs’ answers to some close-ended questions concerning their knowledge and practices, most DSs seemed to be unaware or to have limited knowledge of organizational principles and disciplinary cultures of data practices, including preservation policies in UTU and in their disciplines, and they may rate the importance of the competence highly just in case: 67 percent of DSs (10) were ignorant of the data policy of UTU; 73 percent (11) were not aware of organizing, documenting, and description standards in their disciplines; and 67 percent (10) had not deposited data to data repositories. Those who had used repositories had deposited their data most often to the GenBank,¹⁰ the Finnish Social Science Data Archive, or to the Archives of History, Culture, and Arts Studies of UTU. FMs with long-term experience of the cultures of practices in their disciplines are probably also better informed about whether there are principles and practices for preservation of data sets in their disciplines. However, we must account for the high standard deviation (1.22) of the rating of the importance of “data preservation” by FMs (Table 2, Appendix B), stemming mostly from different ratings of the importance of the competence between HSS and STEM disciplines’ FMs.

Limitations

The analysis and conclusions of the results of this study have at least two restrictions. Because of the small number of participants in the study and the use of convenience sampling, the results cannot be generalized outside of this group. However, we believe that the quantitative analysis of the Likert-like ratings, together with additional numerical data from answers to fixed-choice questions and the qualitative content analysis of the answers to open-ended questions, will give valuable information about the perceptions of DSs’ current RDM competencies and the perceived importance of these competencies. The study will be a good starting point to further studies concerning DSs’ current competencies and competence needs. Besides, we have found the DIL interview toolkit¹¹ as a useful tool for the library to increase faculty members’ knowledge about RDM and to establish contacts and collaboration with them to build RDM education to doctoral students.

Conclusions and Implications

The aim of this study was to investigate FMs’ ratings and DSs’ self-ratings of DSs’ current RDM competencies, as well as their ratings of the importance of these competencies on a five-point Likert-like scale. Moreover, based on the potential gap between the average level of the competencies’ rated importance and the average level of the DSs’ rated or self-rated current competencies, the aim was to discover the

¹⁰ GenBank: <https://www.ncbi.nlm.nih.gov/genbank/>

¹¹ DIL Interview Toolkit: <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1058&context=dilsymposium>

educational RDM needs of DSs. The quantitative analysis of Likert-like scale ratings was complemented by insights from open-ended comments and other fixed-choice answers of the study participants. On average, the perceived importance was rated as “very important” (4.07 in the five-point Likert-like scale), while DSs’ current competencies were rated or self-rated between “have some competence” and “have good competence” (2.47). The difference between DSs’ current RDM competencies and the perceived importance of these competencies was statistically highly significant ($p < 0.0001$) in all the 12 competencies, signifying that DSs have educational needs in all of these competencies.

When planning the education based on the results of this study, it is important to note that, with the exception of “data preservation”, there were no statistically significant differences in the ratings of the importance of RDM competencies between FMs and DSs (Table 2, Appendix B). However, when it comes to DSs’ current competencies, the difference between FMs’ ratings of DSs’ average current competencies and DSs’ self-ratings was statistically significant. This is not surprising per se, as self-ratings and non-self-report measures of competence are not identical constructs. As a rule DSs self-rated their current competencies higher than FMs rated DSs’ average competencies (Table 3, Appendix B). Reasons for the DSs’ higher self-ratings may stem from the lack of knowledge of the meaning of RDM competencies in practice and/or from the over-confidence of their competencies (Kruger and Dunning, 1999). The “interview bias” (Groves et al., 2009) may also be a possible reason for disproportionately high self-ratings of DSs because they were uncertain about the skills they thought the interviewers expected them to have. These possible sources of errors are important to recognize when planning the education. However, the gap between the rated/self-rated DSs’ current competencies and the rated importance of competencies is statistically significant with both groups (Tables 4-5, Appendix B). This means that both FMs and DSs recognized the need for education concerning all the 12 competencies. Six competencies having the largest gap were almost the same in both groups: data quality and documentation (FMs, DSs), metadata and data description (FMs, DSs), data preservation (FMs, DSs), data curation and re-use (FMs), ethics and attribution (FMs, DSs), data management and organization (FMs, DSs), and discovery and acquisition of data (DSs). Thus it is possible to utilize the results of the analysis when planning RDM education: major needs emerged to improve and standardize data management planning, and data documenting and describing, not only for the ongoing research project but especially for data preservation, sharing, and reusing. Though in principle 87 percent of DSs had a positive attitude to data sharing, they typically made the (unstandardized) documentation only for themselves, not for other persons. Most of the DSs were unaware of the intellectual property and contracts issues affecting the possibility to share the data. Neither did 60 percent of DSs know any discipline-specific or general open repositories to find other researchers’ datasets and to deposit their own data.

However, practical application of the generic principles varies depending on discipline; type and format of data, including legal, ethical, storage, preservation, and sharing considerations of data; and other circumstances like policies and mandates of publishers and funders. There is also need for campus-wide collaboration in planning and implementing of the curriculum, because learning and applying RDM competencies in different research settings requires multi-professional expertise by many specialists like researchers, teachers, lawyers, data librarians, research IT professionals, biostatisticians, and repository specialists. In a separate article under way we will tell about development, implementation, and assessment of RDM education that we have

built in collaboration with multi-professional working group, taking advantage of this analysis (see also Rantasaari and Kokkinen, 2019a, 2019b).

Sound data management skills and practices make it possible to produce, maintain, preserve, and share high-quality, coherent research data. On the one hand, this may require more time and effort from researchers in data processing, documenting, and curating to produce well-organized, reliable, reusable, and FAIR data because in many disciplines researchers have not taken account of the long-term preservation, sharing, and re-use of data (Boté and Termens, 2019; Krahe et al., 2020; Read et al., 2015; Tenopir, 2011; Tenopir et al., 2015; Yu, Deuble, and Morgan, 2017). On the other hand, sound RDM practices will secure the high quality of data, which enables its sharing and re-use, and thus, will have more impact.

The lack of graduate students' RDM skills has been found in previous research literature (e.g. Goben and Griffin, 2019; Wiley and Kerby, 2018). The investigation of faculty members' and graduate students' perceptions of graduate students' current RDM competencies and the perceived importance of these competencies is under-examined; this would help emphasize the need for RDM education. This study will help fill this gap. However, due to a small number of the respondents in this study and the scarcity of previous research, there is a need for further research on how the perceived importance of RDM competencies and the perceived current competencies of DSs varies between disciplines, methods, and data types. Furthermore, when applying the findings of this research to the development of RDM curricula, the educational impact on researchers' RDM practices should be studied further based not only on the self-assessments of the participants but also on the observations of their data management practices.

Acknowledgements

I would like to thank my supervisors, Professor Gunilla Widén at the Åbo Akademi University and Professor Isto Huvila at the Uppsala University, who read and commented on several drafts of this study; Biostatistician Eliisa Löyttyniemi at the University of Turku for discussions on statistical analysis; the Head of Library Services Heli Kokkinen at the Turku University Library who helped with conducting interviews by taking notes and reading and commenting the text; and the Head of Research IT Meri-Tuulia Kaarakainen at the University of Turku for commenting the interview forms and participating in interviews.

Data

The quantitative data underlying this study can be accessed through Zenodo (Rantasaari, J. (2020)).

References

Adamick, J., Reznik-Zellen, R., & Sheridan, M. (2013). Data management training for graduate students at a large research university. *Journal of EScience Librarianship*, 1(3), 180–188. doi:10.7191/jeslib.2012.1022

- 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
- Alexogiannopoulos, E., McKenney, S., & Pickton, M. (2010). *Research data management project: A DAF investigation of research data management practices at The University of Northampton*. Retrieved from The University of Northampton website: <http://nectar.northampton.ac.uk/2736/1/Alexogiannopoulos20102736.pdf>
- Andrews Mancilla, H., Teperek, M., Van Dijck, J., Den Heijer, K., Eggermont, R., Plomp, E., Turkyilmaz-van der Velden, Y., & 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:10.18352/lq.10287
- Association of College and Research Libraries. (ACRL). (2000). *Information literacy competency standards for higher education*. Retrieved from <http://hdl.handle.net/11213/7668>
- Atenas, J., Havemann, L., & Timmermann, C. (2020). Critical literacies for a datafied society: Academic development and curriculum design in higher education. *Research in Learning Technology*, 28, 1–14. doi:10.25304/rlt.v28.2468
- Borgman, C. L. (2016). *Big data, little data, no data: Scholarship in the networked world* (First MIT). London: The MIT Press.
- Boté, J.-J., & Termens, M. (2019). Reusing data technical and ethical challenges. *DESIDOC Journal of Library & Information Technology*, 39(06), 329–337. doi:10.14429/djlit.39.06.14807
- Bowler, L., Acker, A., Jeng, W., & Chi, Y. (2017). “It lives all around us”: Aspects of data literacy in teen’s lives. *Proceedings of the Association for Information Science and Technology*, 54(1), 27–35. doi:10.1002/pra2.2017.14505401004
- Bresnahan, M., & Johnson, A. (2014). *Building a unified data and information literacy program: A collaborative approach to instruction* [PowerPoint slides]. Colorado University Libraries. Retrieved from <https://scholar.colorado.edu/concern/presentations/5d86p0976>
- Briney, K. (2015). *Data management for researchers : Organize, maintain and share your data for research success*. Exeter, UK: Pelagic Publishing.
- Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, 6(1), 97–113. doi:10.1177/1468794106058877
- Carlson, J. (2015). Future directions for data information literacy: Growing programs and communities of practice. In J. Carlson & L.R. Johnston (Eds.), *Data information literacy: Librarians, data, and the education of a new generation of researchers* (pp. 247-260). West Lafayette, Indiana: Purdue University Press. Retrieved from www.jstor.org/stable/j.ctt6wq2vh.16

- Carlson, J., Fosmire, M., Miller, C. C., & Sapp Nelson, M. (2011). Determining data information literacy needs: A study of students and research faculty. *Portal: Libraries and the Academy*, 11(2), 629–657. doi:10.1353/pla.2011.0022
- Carlson, J., Jeffryes, J., Johnston, L. R., Nichols, M., Westra, B., & Wright, S. J. (2015). An exploration of the data information literacy competencies: Findings from the project interviews. In J. Carlson & L.R. Johnston (Eds.), *Data information literacy: Librarians, data, and the education of a new generation of researchers* (pp. 51–70). West Lafayette, Indiana: Purdue University Press. Retrieved from <https://www.jstor.org/stable/j.ctt6wq2vh.8>
- Carlson, J., Johnston, L. R., & Westra, B. (2015). Developing the data information literacy project: Approach and methodology. In J. Carlson & L.R. Johnston (Eds.), *Data information literacy: Librarians, data, and the education of a new generation of researchers* (pp. 35–50). West Lafayette, Indiana: Purdue University Press. Retrieved from <https://www.jstor.org/stable/j.ctt6wq2vh.7>
- Carlson, J., Johnston, L., Westra, B., & Nichols, M. (2013). Developing an approach for data management education: A report from the data information literacy project. *International Journal of Digital Curation*, 8(1), 204–217. doi:10.2218/ijdc.v8i1.254
- Castle, C. (2019). Getting the central RDM message across: A case study of central versus discipline-specific research data services (RDS) at the University of Cambridge. *Libri*, 69(2), 105–116. doi:10.1515/libri-2018-0064
- Coates, H. L. (2014). Ensuring research integrity: The role of data management in current crises. *College and Research Libraries News*, 75(11), 598–601. doi:10.5860/crln.75.11.9224
- Cox, A. M., Kennan, M. A., Lyon, L., & Pinfield, S. (2017). Developments in research data management in academic libraries: Towards an understanding of research data service maturity. *Journal of the Association for Information Science & Technology*, 68(9), 2182–2200. doi:10.1002/asi.23781
- Eisenberg, M. B. (2008). Information literacy: Essential skills for the information age. *DESIDOC Journal of Library & Information Technology*, 28(2), 39–47. doi:10.14429/djlit.28.2.166
- European Commission. (2019). *Key competences for lifelong learning*. Luxembourg. <https://doi.org/10.2766/569540>
- Federer, L. (2016). Research data management in the age of big data: Roles and opportunities for librarians. *Information Services & Use*, 36, 35–43. doi:10.3233/ISU-160797
- Federer, L. (2018). Defining data librarianship: a survey of competencies, skills, and training. *Journal of the Medical Library Association*, 106(3), 294–303. doi:10.5195/jmla.2018.306

- Frugoli, J., Etgen, A. M., & Kuhar, M. (2010). Developing and communicating responsible data management policies to trainees and colleagues. *Science and Engineering Ethics*, 16(4), 753–762. doi:10.1007/s11948-010-9219-1
- Gabridge, T. (2009). The last mile: Liaison roles in curating science and engineering research data. *Research Library Issues*, (265), 15–21. doi:10.29242/rli.265.4
- Goben, A., & Griffin, T. (2019). In aggregate: Trends, needs, and opportunities from research data management surveys. *College and Research Libraries*, 80(7), 903–924. doi:10.5860/crl.80.7.903
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255–274. doi:10.3102/01623737011003255
- Griffin, T. M. (2020). Centering graduate students' research projects in data management education: A pilot program. *Journal of Librarianship and Scholarly Communication*, 8(1), 1–21. doi:10.7710/2162-3309.2365
- Groves, R. M., Fowler, J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2nd Edition). Hoboken: John Wiley and Sons Inc.
- Gubrium, J., & Holstein, J. (2011). Survey interviewing. In R. A. Singleton & B. C. Straits (Eds.), *Handbook of Interview Research* (online, pp. 59–82). SAGE Publications, Inc. doi:10.4135/9781412973588
- Hattwig, D., Bussert, K., Medaille, A., & Burgess, J. (2013). Visual literacy standards in higher education: New opportunities for libraries and student learning. *Portal*, 13(1), 61–89. doi:10.1353/pla.2013.0008
- Humanities Advanced Technology & Information Institute. (2009). *Data asset framework implementation guide*. Retrieved from http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf
- Hunt, K. (2005). The challenges of integrating data literacy into the curriculum in an undergraduate institution. *IASSIST Quarterly*, 28(2), 12–16. doi:10.29173/iq791
- Jahnke, L., Asher, A., & Keralis, S. (2012). *The problem of data*. Retrieved from https://digitalcommons.bucknell.edu/fac_pubs/52
- Johnston, L., & Jeffryes, J. (2014). Data management skills needed by structural engineering students: Case study at the University of Minnesota. *Journal of Professional Issues in Engineering Education and Practice*, 140(2). doi:10.1061/(ASCE)EI.1943-5541.0000154
- Joo, S., & Peters, C. (2019). User needs assessment for research data services in a research university. *Journal of Librarianship and Information Science*, 1–14. doi:10.1177/0961000619856073

- Klenke, C. M., Auch Schultz, T., Tokarz, R. E., & Azadbakht, E. S. (2020). Curriculum data deep dive: Identifying data literacies in the disciplines. *Journal of ESscience Librarianship*, 9(1), 1–16. doi:10.7191/jeslib.2020.1169
- Knight, G. (2013). Research data management at London School of Hygiene and Tropical Medicine: Web survey report. Retrieved from <http://blogs.lshtm.ac.uk/rdmss/files/2013/04/LSHTM-RDM-Web-Survey-Report.pdf>
- Koltay, T. (2014). Big data, big literacies? *Tema*, 24, 3–8. Retrieved from https://www.researchgate.net/publication/265334966_Big_data_big_literacies
- Koltay, T. (2015). Data literacy: In search of a name and identity. *Journal of Documentation*, 71(2), 401–415. doi:10.1108/JD-02-2014-0026
- Koltay, T. (2017). Research 2.0 and research data services in academic and research libraries: priority issues. *Library Management*, 38(6/7), 345–353. doi:10.1108/LM-11-2016-0082
- Koltay, T. (2019). Data literacy in academia: Basics and pedagogical views. *Opus et Educatio*, 6(4). doi:10.3311/ope.343
- Kowalczyk, S. T. (2018). Modelling the research data lifecycle. *International Journal of Digital Curation*, 12(2), 331–361. doi:10.2218/ijdc.v12i2.429
- Krahe, M. A., Toohey, J., Wolski, M., Scuffham, P. A., & Reilly, S. (2020). Research data management in practice: Results from a cross-sectional survey of health and medical researchers from an academic institution in Australia. *Health Information Management Journal*, 49(2–3), 108–116. doi:10.1177/1833358319831318
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. doi:10.1037/0022-3514.77.6.1121
- Lefebvre, A., Schermerhorn, E., & Spruit, M. (2018). How research data management can contribute to efficient and reliable science. In *26th European Conference on Information Systems: Beyond Digitization – Facets of Socio-Technical Change, ECIS 2018* (pp. 1–15). Association for Information Systems. Retrieved from https://aisel.aisnet.org/ecis2018_rp/35
- Leinonen, M., Otonkorpi-Lehtoranta, K., & Heiskanen, T. (2017). Kyselyhaastattelu. In M. Hyvärinen, P. Nikander, J. Ruusuvaori, & A.-L. Aho (Eds.), *Tutkimushaastattelun käsikirja* (online, p. 394). Tampere: Vastapaino. Retrieved from <https://www.ellibslibrary.com/cepos/9789517686112>

- Lyon, L., Ball, A., Duke, M., & Day, M. (2012). Developing a community capability model framework for data-intensive research. In R. Moore, K. Ashley, & S. Ross (Eds.), *Proceedings of the 9th International Conference on Preservation of Digital Objects* (pp. 9–16). Toronto: University of Toronto, Faculty of Information. Retrieved from <https://researchportal.bath.ac.uk/en/publications/developing-a-community-capability-model-framework-for-data-intens>
- MacMillan, D. (2010). Sequencing genetics information: Integrating data into information literacy for undergraduate biology students. *Issues in Science and Technology Librarianship*, 61(Spring). doi:10.5062/F44F1NNK
- Maienschein, J., Parker, J. N., Laubichler, M., & Hackett, E. J. (2019). Data management and data sharing in science and technology studies. *Science Technology and Human Values*, 44(1), 143–160. doi:10.1177/0162243918798906
- Mannheimer, S., Sterman, L., & Borda, S. (2016). Discovery and reuse of open datasets: An exploratory study. *Journal of EScience Librarianship*, 5(1), 1–14. doi:10.7191/jeslib.2016.1091
- Mittenfelner, C. & Ravitch, S. M. (2018). Interviews. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (p. 3). doi:10.4135/9781506326139
- Molloy, L., & Snow, K. (2012). The data management skills support initiative: Synthesising postgraduate training in research data management. *International Journal of Digital Curation*, 7(2), 101–109. doi:10.2218/ijdc.v7i2.233
- Parsons, T., Grimshaw, S., & Williamson, L. (2013). *University of Nottingham research data management survey*. Nottingham. Retrieved from <http://eprints.nottingham.ac.uk/1893/>
- Pasek, J. E. E., & Mayer, J. (2019). Education needs in research data management for science-based disciplines: Self-assessment surveys of graduate students and faculty at two public universities. *Issues in Science and Technology Librarianship*, (92). doi:10.29173/istl12
- Peters, C., & Vaughn, P. (2014). Initiating data management instruction to graduate students at the University of Houston using the New England Collaborative Data Management Curriculum. *Journal of EScience Librarianship*, 3(1), 86–99. doi:10.7191/jeslib.2014.1064
- Piorun, M., Kafel, D., Leger-Hornby, T., Najafi, S., Martin, E., Colombo, P., & LaPelle, N. (2012). Teaching research data management: An undergraduate/graduate curriculum. *Journal of EScience Librarianship*, 1(1), 46–50. doi:10.7191/jeslib.2012.1003
- Pouchard, L., & Bracke, M. S. (2016). An analysis of selected data practices: A case study of the purdue college of agriculture. *Issues in Science and Technology Librarianship*, (85). doi:10.5062/F4057CX4

- Prado, J. C., & Marzal, M. Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri*, 63(2), 123–134. doi:10.1515/libri-2013-0010
- Pryor, G. (2012). Why manage research data? In G. Pryor (Ed.), *Managing research data* (p. 224). Cambridge: Facet Publishing. doi:10.29085/9781856048910
- Qin, J., & D'Ignazio, J. (2010). The central role of metadata in a science data literacy course. *Journal of Library Metadata*, 10(2–3), 188–204. doi:10.1080/19386389.2010.506379
- Rantasaari, J., & Kokkinen, H. (2019a). Closing the skills gap: The basics of the research data management (BRDM) course: Case University of Turku. In *Proceedings of the IATUL Conferences* (pp. 1–8). Perth: 2019 IATUL Proceedings. Retrieved from <https://docs.lib.purdue.edu/iatul/2019/fair/5/>
- Rantasaari, J., & Kokkinen, H. (2019b). *Closing the skills gap: The basics of the research data management (BRDM) course: Case University of Turku*. [Powerpoint slides]. LIBER 2019. Retrieved from <https://zenodo.org/record/3259596#.YLhUSpMzaWg>
- Rantasaari, J. (2020). Doctoral students' educational needs in research data management: Quantitative data of perceived importance and current competencies. [Data set]. Meyrin, Switzerland: Zenodo. doi:10.5281/zenodo.3668647
- Read, K. B., Larson, C., Gillespie, C., Oh, S. Y., & Surkis, A. (2019). A two-tiered curriculum to improve data management practices for researchers. *PLoS ONE*, 14(5), 1–14. doi:10.1371/journal.pone.0215509
- Read, K. B., Surkis, A., Larson, C., McCrillis, A., Graff, A., Nicholson, J., & Xu, J. (2015). Starting the data conversation: Informing data services at an academic health sciences library. *Journal of the Medical Library Association*, 103(3), 131–135. doi:10.3163/1536-5050.103.3.005
- Research Information Network. (2008). *Stewardship of digital research data: A framework of principles and guidelines*. Retrieved from <http://shapingthefuture.pbworks.com/f/Stewardship+of+digital+research+data+by+RIN.pdf>
- Revez, J. (2018). Opening the heart of science: A review of the changing roles of research libraries. *Publications*, 6(11), 1–13. doi:10.3390/publications6010009
- Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., & Wuetherick, B. (2015). *Strategies and best practices for data literacy education: Knowledge synthesis*. Halifax. doi:10.13140/RG.2.1.1922.5044
- Sapp Nelson, M. (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), 1–11. doi:10.7191/jeslib.2017.1096

- Schild, M. (2004). Information literacy, statistical literacy and data literacy. *IASSIST Quarterly*, 28(2–3), 6–11. doi:10.29173/iq790
- Schneider, R. (2013). Research data literacy. In S. Kurbanoglu, E. Grassian, D. Mizrahi, R. Catts, & S. Spiranec (Eds.), *Worldwide commonalities and challenges in information literacy research and practice* (pp. 134–140). Istanbul: Springer International Publishing. doi:10.1007/978-3-319-03919-0_16
- Scholten, S., Anbeek, P., Böhmer, J., Brullemans-Spansier, M., Geest, M. van der, Jetten, M., Staiger, C., Slouwerho, I., & Gelder, C. W. G. van. (2019). *Life sciences data steward function matrix, version 1.1*. doi:10.5281/ZENODO.2561723
- Schumacher, J., & VandeCreek, D. (2015). Intellectual capital at risk: Data management practices and data loss by faculty members at five American universities. *International Journal of Digital Curation*, 10(2), 96–109. doi:10.2218/ijdc.v10i2.321
- Schäfer, M. T., & van Es, K. (Eds.). (2017). *The datafied Society: Studying culture through data*. (online). Amsterdam: Amsterdam University Press. Retrieved from <https://library.oapen.org/handle/20.500.12657/31843>
- Semeler, A. R., Pinto, A. L., & Rozados, H. B. F. (2019). Data science in data librarianship: Core competencies of a data librarian. *Journal of Librarianship and Information Science*, 51(3), 771–780. doi:10.1177/0961000617742465
- Shorish, Y. (2015). Data information literacy and undergraduates: A critical competency. *College and Undergraduate Libraries*, 22(1), 97–106. doi:10.1080/10691316.2015.1001246
- Society of College, National and University Libraries (SCONUL). (2011). *The SCONUL seven pillars of information literacy: Core model for higher education*. Retrieved from <https://www.sconul.ac.uk/sites/default/files/documents/coremodel.pdf>
- Specht, A., Guru, S., Houghton, L., Keniger, L., Driver, P., Ritchie, E. G., Lai, K., & Treloar, A. (2015). Data management challenges in analysis and synthesis in the ecosystem sciences. *Science of the Total Environment*, 534, 144–158. doi:10.1016/j.scitotenv.2015.03.092
- Stephenson, E., & Caravello, P. S. (2007). Incorporating data literacy into undergraduate information literacy programs in the social sciences: A pilot project. *Reference Services Review*, 35(4), 525–540. doi:10.1108/00907320710838354
- Strasser, C., Cook, R., Michener, W., & Budden, A. (2012). *Primer on data management*. Retrieved from http://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE*, 6(6), 1–21. doi:10.1371/journal.pone.0021101

- Tenopir, C., Birch, B., & Allard, S. (2012). *Academic libraries and research data services: Current practices and plans for the future* (Vol. 46). Retrieved from http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLOS ONE*, *10*(8), 1–24. doi:10.1371/journal.pone.0134826
- Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., Schmidt, B., Baird, L., Sandusky, R. J., & Allard, S. (2017). Research data services in European academic research libraries. *LIBER Quarterly*, *27*(1), 23–44. doi:10.18352/lq.10180
- Thielen, J., Samuel, S. M., Carlson, J., & Moldwin, M. (2017). Developing and teaching a two-credit data management course for graduate students in climate and space sciences. *Issues in Science and Technology Librarianship*, (86). doi:10.5062/F42Z13HQ
- Thielen, J., & Hess, A. N. (2017). Advancing research data management in the social sciences: Implementing instruction for education graduate students into a doctoral curriculum. *Behavioral and Social Sciences Librarian*, *36*(1), 16–30. doi:10.1080/01639269.2017.1387739
- Twidale, M. B., Blake, C., & Gant, J. (2013). Towards a data literate citizenry. In *iConference Proceedings* (pp. 247–257). Fort Worth: University of Illinois. Retrieved from <http://hdl.handle.net/2142/38385>
- Verbaan, E., & Cox, A. M. (2014). Occupational sub-cultures, jurisdictional struggle and third space: Theorising professional service responses to research data management. *Journal of Academic Librarianship*, *40*(3–4), 211–219. doi:10.1016/j.acalib.2014.02.008
- Wanner, A. (2015). Data literacy instruction in academic libraries: Best practices for librarians. See Also: *The UBC School of Library, Archival and Information Studies Student Journal*, (1), 1–17. doi:10.14288/sa.v1i1.186335
- Warschauer, M. (2011). A Literacy approach to the digital divide. In M. A. Pereyra (Ed.), *Las multialfabetizaciones en el espacio digital*. Malaga: Ediciones Aljibe. Retrieved from <http://education.uci.edu/uploads/7/2/7/6/72769947/lit-approach.pdf>
- Waterfield, J. (2018). Interviewer bias. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (p. 3). doi:10.4135/9781506326139
- Weller, T., & Monroe-Gulick, A. (2014). Understanding methodological and disciplinary differences in the data practices of academic researchers. *Library Hi Tech*, *32*(3), 467–482. doi:10.1108/LHT-02-2014-0021
- Wiley, C. A., & Kerby, E. E. (2018). Managing research data: Graduate student and postdoctoral researcher perspectives. *Issues in Science and Technology Librarianship*, (89). doi:10.5062/F4FN14FJ

- Wilson, J., Jeffreys, P., Patrick, M., Rumsey, S., & Jefferies, N. (2015). *Results of the 2012 University of Oxford research data management survey*. Oxford: Mendeley. doi:10.17632/7C9CP6KKMJ.1
- Witt, M., Carlson, J., & Brandt, D. S. (2009). Constructing data curation profiles. *The International Journal of Digital Curation*, 4(3), 93–103. doi:10.2218/ijdc.v4i3.117
- Womack, R. (2015). Data visualization and information literacy. *IASSIST Quarterly*, 38(1), 12–17. doi:10.29173/iq619
- Wright, S. J., & Andrews, C. (2015). Developing a for-credit course to teach data information literacy skills: A case study in natural resources. In J. Carlson & L. R. Johnston (Eds.), *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers* (pp. 73–99). Retrieved from <https://www.jstor.org/stable/j.ctt6wq2vh.9>
- Yu, F., Deuble, R., & Morgan, H. (2017). Designing research data management services based on the research lifecycle – a consultative leadership approach. *Journal of the Australian Library and Information Association*, 66(3), 287–298. Retrieved from <https://tandfonline.com/doi/full/10.1080/24750158.2017.1364835>
- Yu, H. H. (2017). The role of academic libraries in research data service (RDS) provision. *The Electronic Library*, 35(4), 783–797. doi:10.1108/EL-10-2016-0233

Appendix A: Respondents

Table 2. Respondents by their roles and faculties in relation to number and percentage of all doctoral students.

Faculty	DS	SV	BS	Total	IV %	FDS	PhD %
FM	5	3	4	12	34	637	33
FSE	4	6		10	28	363	19
TSE	3	2		5	14	160	8
FH	1	2		3	9	360	19
FSS	1	2		3	9	182	10
FE	1	1		2	6	126	7
FL				0	0	74	4
Total				35	100	1902	100

FM = Faculty of Medicine + Turku PET Centre

FSE = Faculty of Science and Engineering + Finnish Centre for Astronomy with ESO

TSE = Turku School of Economics

FH = Faculty of Humanities

FSS = Faculty of Social Sciences

FE = Faculty of Education

FL = Faculty of Law

DS = Doctoral Student

SV = Supervisor

BS = Biostatistician

IV % = Percentage of all respondents

FDS = Amount of doctoral students in faculty

PhD % = Faculty's doctoral students' percentage of all doctoral students at UTU

Appendix B: Tables of Results

Table 1. The importance of RDM competencies vs. doctoral students' current RDM competencies as rated/self-rated by all respondents (mean, standard deviation (SD), and p-value).

Competence	Mean of the perceived importance of competence	SD	Mean of the perceived current competence	SD	p-value
Ethics and attribution	4.47	0.75	2.79	0.93	<0.0001
Data processing and analysis	4.45	0.67	3.06	0.92	<0.0001
Data quality and documentation	4.38	0.74	2.41	0.73	<0.0001
Data visualization and presentation	4.24	0.78	2.88	0.76	<0.0001
Metadata and data description	4.15	0.8	2.29	0.82	<0.0001
Cultures of practice	4.03	0.87	2.50	0.98	<0.0001
Discovery and acquisition of data	4.03	1.1	2.52	0.96	<0.0001
Data management and organization	3.97	0.97	2.21	0.73	<0.0001
Data curation and re-use	3.97	1	2.35	0.97	<0.0001
Data conversion and interoperability	3.89	1.04	2.44	0.85	<0.0001
Data preservation	3.82	1.14	1.94	0.91	<0.0001
Databases and data formats	3.39	1.46	2.22	1.07	<0.0001
MEAN	4.07	0.56	2.47	0.62	<0.0001

Rated importance of competence: 1=not important; 2=somewhat important; 3=important; 4=very important; 5=essential.

Rated/self-rated current competence: 1=do not have; 2=some; 3=good; 4=very good; 5=ultimate.

Table 2. The importance of RDM competencies as rated separately by faculty members (FMs) and doctoral students (DSs)

Competence	Mean of the perceived importance of competence (FMs)	SD	Mean of the perceived importance of competence (DSs)	SD	p-value
Ethics and attribution	4.32	0.82	4.67	0.62	0.18
Data processing and analysis	4.61	0.61	4.27	0.7	0.14
Data quality and documentation	4.42	0.61	4.33	0.9	0.74
Data visualization and presentation	4.37	0.76	4.07	0.8	0.27
Metadata and data description	4.11	0.76	4.20	0.86	0.75
Cultures of practice	4.11	0.88	3.93	0.88	0.58
Discovery and acquisition of data	3.83	1.15	4.27	1.03	0.27
Data management and organization	3.84	1.07	4.13	0.83	0.39
Data curation and re-use	3.89	1.1	4.07	0.88	0.63
Data conversion and interoperability	3.89	1.1	3.87	0.99	0.94
Data preservation	3.42	1.22	4.33	0.82	0.02
Databases and data formats	3.00	1.49	3.93	1.27	0.07
MEAN	3.98	0.54	4.17	0.58	0.33

Perceived importance of competence: 1=not important; 2=somewhat important; 3=important; 4=very important; 5=essential.

Table 3. DSs' current RDM competencies as rated by FMs and self-rated by Dss.

Perceived competence	Mean (FMs)	SD	Mean (DSs)	SD	p-value
Data processing and analysis	2.94	0.8	3.20	1.08	0.44
Data visualization and presentation	2.74	0.81	3.07	0.7	0.22
Ethics and attribution	2.37	0.83	3.33	0.82	<0.0001
Cultures of practice	2.32	0.89	2.73	1.1	0.23
Discovery and acquisition of data	2.28	0.75	2.80	1.15	0.13
Data conversion and interoperability	2.26	0.56	2.67	1.11	0.22
Data quality and documentation	2.16	0.6	2.73	0.8	0.02
Metadata and data description	2.00	0.67	2.67	0.9	0.02
Data management and organization	2.00	0.77	2.47	0.64	0.07
Data curation and re-use	1.89	0.81	2.93	0.88	<0.0001
Databases and data formats	1.87	0.81	2.67	1.23	0.03
Data preservation	1.42	0.69	2.60	0.74	<0.0001
MEAN	2.19	0.43	2.82	0.65	0.0018

Perceived current competence: 1=do not have; 2=some; 3=good; 4=very good; 5=ultimate

Table 4. FMs' ratings of the importance of RDM competencies vs. their ratings of DSs' average current RDM competencies.

Competence	Mean of the perceived importance of competence	SD	Mean of the perceived current competence	SD	p-value
Data processing and analysis	4.61	0.61	2.94	0.80	<0.0001
Data visualization and presentation	4.37	0.76	2.74	0.81	<0.0001
Ethics and attribution	4.32	0.82	2.37	0.83	<0.0001
Cultures of practice	4.11	0.87	2.32	0.89	<0.0001
Discovery and acquisition of data	3.83	1.15	2.28	0.75	<0.0001
Data conversion and interoperability	3.89	1.10	2.26	0.56	<0.0001
Data quality and documentation	4.42	0.61	2.16	0.60	<0.0001
Metadata and data description	4.11	0.76	2	0.67	<0.0001
Data management and organization	3.84	1.07	2	0.77	<0.0001
Data curation and re-use	3.89	1.10	1.89	0.81	<0.0001
Databases and data formats	3	1.49	1.87	0.81	0.001
Data preservation	3.42	1.22	1.42	0.69	<0.0001
MEAN	3.98	0.54	2.19	0.43	<0.0001

Table 5. DSS' ratings of the importance of RDM competencies vs. their self-rated current RDM competencies.

Competence	Mean of the perceived importance of competence	SD	Mean of the perceived current competence	SD	p-value
Data processing and analysis	4.27	0.70	3.2	1.08	0.002
Data visualization and presentation	4.07	0.80	3.07	0.70	0.003
Ethics and attribution	4.67	0.62	3.33	0.82	<0.0001
Cultures of practice	3.93	0.88	2.73	1.1	0.001
Discovery and acquisition of data	4.27	1.03	2.8	1.15	<0.0001
Data conversion and interoperability	3.87	0.99	2.67	1.11	0.02
Data quality and documentation	4.33	0.90	2.73	0.80	<0.0001
Metadata and data description	4.2	0.86	2.67	0.90	0.0001
Data management and organization	4.13	0.83	2.47	0.64	0.0002
Data curation and re-use	4.07	0.88	2.93	0.88	0.002
Databases and data formats	3.93	1.27	2.67	1.23	0.0095
Data preservation	4.33	0.82	2.6	0.74	<0.0001
MEAN	4.17	0.58	2.82	0.65	<0.0001