# Multi-Stakeholder Research Data Management Training as a Tool to Improve the Quality, Integrity, Reliability and Reproducibility of Research

Rantasaari, Jukka

*Published in:*
LIBER Quarterly: The Journal of the Association of European Research Libraries

*DOI:*
10.53377/lq.11726

Published: 12/07/2022

*Document Version*
Final published version

*Document License*
CC BY

LIBER QUARTERLY
The Journal of the Association of European Research Libraries
Vol. 32, (2022) 1–54 | e-ISSN: 2213-056X

# Multi-Stakeholder Research Data Management Training as a Tool to Improve the Quality, Integrity, Reliability and Reproducibility of Research

**Jukka Rantasaari**

University of Turku, Åbo Akademi University
jukka.rantasaari@utu.fi, orcid.org/0000-0001-5927-3781

## Abstract

To ensure the quality and integrity of data and the reliability of research, data must be well documented, organised, and described. This calls for research data management (RDM) education for researchers. In light of 3 ECTS Basics of Research Data Management (BRDM) courses held between 2019 and 2021, we aim to find how a generic level multi-stakeholder training can improve STEM and HSS disciplines' doctoral students' and postdoc researchers' competencies in RDM. The study uses quantitative, descriptive and inferential statistics to analyse respondents' self-ratings of their competencies, and a qualitative grounded theory-inspired approach to code and analyse course participants' feedback. **Results**: On average, based on the post-course surveys, respondents' (n = 123) competencies improved one point on a four-level scale, from "little competence" (2) to "somewhat competent" (3). Participants also reported that the training would change their current practices in planning research projects, data management and documentation, acknowledging legal and data privacy viewpoints, and data collecting and organising. Participants indicated that it would be helpful to see legal and data privacy principles and regulations presented as concrete instructions, cases, and examples. The most requested continuing education topics were metadata and description, discipline specific cultures, and backup, version management, and storage. **Conclusions**: Regarding to the widely used criteria for successful training containing 1) active participation during training; 2) demand for RDM training; 3) increased participants'


Uopen Journals | http://liberquarterly.eu/ | DOI: 10.53377/lq.11726

knowledge and understanding of RDM and confidence in enacting RDM practices; and 4) positive post-training feedback, BRDM meets the criteria. This study shows that although reaching excellent competence in a RDM basics training is improbable, participants become aware of RDM and its contents and gain the elementary tools and basic skills to begin applying sound RDM practices in their research. Furthermore, participants are introduced to the academic and research support professionals and vice versa: Stakeholders will get to know the challenges that young researchers and research students encounter when applying RDM. The study reveals valuable information on doctoral students' and postdoc researchers' competencies, the impact of education on competencies, and further learning needs in RDM.

**Keywords**: Research data management; Training; Competencies; Early career researchers; PhD students; Doctoral students; Postdoc researchers

## 1. Introduction

During the second decade of the 2000s, many international, national, and institutional principles and policies and an increasing number of funders and publishers started recommending or mandating researchers to write data management plans (DMPs) and share data (e.g., Academy of Finland, 2019; "Amsterdam call for action on open science", 2016; European Commission, 2018a,b; European University Association, 2017; National Science Foundation, 2011; UNIFI, 2016; Wellcome, 2017). Researchers need education, guidance, and support in research data management (RDM) to help fulfil this task. At the core of these principles, policies and demands is to obtain data of publicly funded research openly accessible and reusable, when possible. In principle, research data, or at least metadata, should be Findable, Accessible, Interoperable and Reusable (FAIR[1]). Sound RDM practices advance the integrity of data, reliability of research results, transparency of the research process, and reproducibility of research (e.g., Chiarelli et al., 2021). However, research transparency and data reuse may only be fully realised if data is opened and shared (Borghi et al., 2018). Shared research data also avoids the gathering of duplicate data and enables combined efforts to find solutions to complicated interdisciplinary research issues like climate change and pandemics (Doucette & Fyfe, 2013; Shearer, 2009). Moreover, sharing research data can significantly shorten the time it takes to move from an initial scientific

discovery to practical applications (Federer, 2016). Nevertheless, it is only useful to share well-documented, described, and organised data (Borghi et al., 2018; Rieser, 2018) that provides clear data sharing parameters, including intellectual property rights (IPR) and agreements. (Rantasaari, 2021).

Though RDM[2] is perceived as important or very important by researchers and graduate students (Pasek & Mayer, 2019; Thielen et al., 2017), many researchers are not managing their data according to recommended RDM guidelines. For example, graduate students are often given substantial data management responsibilities in research projects though they usually have received little or no education in RDM (Goben & Griffin, 2019; Krahe et al., 2020; Maienschein et al., 2019; Wiley & Kerby, 2018). Thus, they tend to develop ad hoc solutions with the trial-and-error method (Thielen & Hess, 2017; Wright & Andrews, 2015). Therefore, RDM practices are often unstandardised, and IPR and contract issues may be unfamiliar. Also, documentation made to carry out the ongoing research that does not consider other uses and users does not enable data sharing and reusing, undermining research reproducibility (Rantasaari, 2021).

In this article, our goal is to find how generic, multi-stakeholder training can improve participants' competencies and further comprehension of the relevance of sound research data management practices to the quality and integrity of data and reliability of the research.

In practice, we will report the outcomes of the Basics of Research Data Management (BRDM) course over three years (2019–2021), held at two Finnish universities. The learning objectives and contents of BRDM were developed based on an interview study on doctoral students' RDM competencies and learning needs (Rantasaari, 2021; Rantasaari & Kokkinen, 2019), discussions with the leader of the biostatistician team of the University of Turku (UTU), and research literature and lessons learned from previous RDM trainings (e.g., Piorun et al., 2012; Qin & D'ignazio, 2010; Thielen et al., 2017; Whitmire, 2015; Wright & Andrews, 2015).

We aim to answer the following questions:

- RQ1: How did course participants self-rate their RDM competencies before and after BRDM course?

- RQ2: What kind of educational impact did the course have on participants' RDM competencies (knowledge, skills, and abilities) based on participants' self-ratings and collected and categorised feedback?
- RQ3: What kind of further learning needs did the respondents express after the course?

After the introduction, we will discuss specific contents and lessons learned in previous RDM basic trainings directed specifically for graduate students or researchers. The methods section will describe BRDM's objectives, structure, and learning methods. Research methods used to answer research questions RQ 1 to 3 will be described. Section four contains the results of the study, and section five the discussion and conclusions.

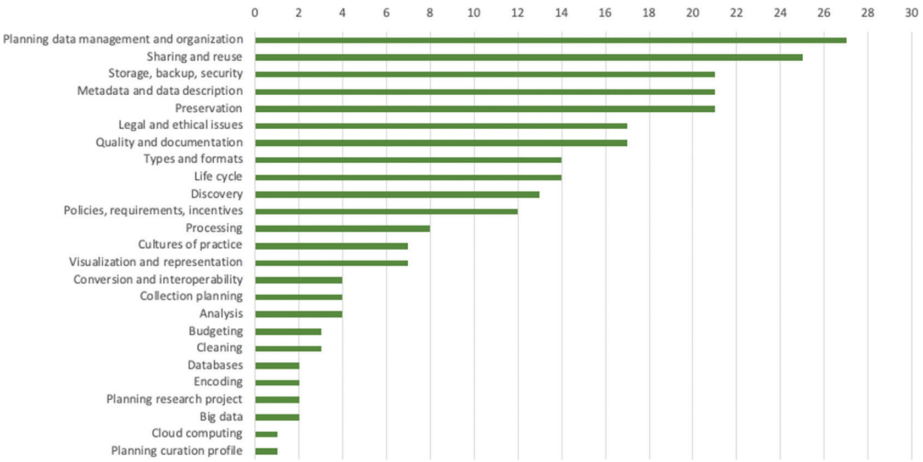
## 2. Literature Review

### 2.1. Common Contents in RDM Education

We collected the information of 30 RDM trainings from research articles and conference proceedings directed at graduate students or researchers between 2010 and 2021. Using the trainings' descriptions, the author categorised their contents as RDM topics and listed the topics handled in each training (Table 1 in Appendix A). The number of trainings addressing each topic is charted below (Figure 1). The most common topics covered in over 50% of the trainings were "Planning data management and organisation" (27); "Sharing and reuse" (25); "Storage, backup, and security" (21); "Metadata and data description" (21); "Preservation" (21); "Legal and ethical issues" (17); and "Quality and documentation" (17).

Overwhelmingly, the most common topics were "Planning data management and organisation" and "Sharing and reuse" which is understandable as the need for RDM became widespread after big funders like Wellcome (2017) and National Science Foundation (2011) began mandating DMPs and recommending data sharing in funding applications. Finnish major research funder Academy of Finland has required DMPs and data sharing in principle since 2015 (Academy of Finland, 2019). In RDM educational programs, data have been noted as a validator of research. Also, the reuse of data, as well as the policies, permits, and licenses demanding the data sharing, and the importance of becoming familiar with data sharing culture and infrastructure, have

*Fig. 1: RDM topics handled in 30 trainings held between 2010 and 2021.*




been discussed (Piorun et al., 2012; Read et al., 2019; Research data service, n.d.; Wright & Andrews, 2015).

Besides the informational type of contents, some courses and workshops include more technically oriented RDM topics such as data analysing and visualising, wrangling, merging, cleaning, and publishing data sets, as well as building and using relational databases for data gathering, organising, and querying (Carpentries, n.d.; Pascuzzi & Sapp Nelson, 2018; Qin & D'ignazio, 2010; Read et al., 2019; Research data service, n.d.; Wright & Andrews, 2015).

### 2.2. Lessons Learned in RDM Education

Though a lack of comprehensive and specific reporting of the results of educational RDM efforts exists (Goben & Griffin, 2019; Perrier et al., 2017), the feedback and results appear to be satisfactory or good. Typically, the attendants have been reported to have given good feedback (e.g., Chew et al., 2021; Whitmire, 2015), with their satisfaction varying from medium to high (Muilenburg et al., 2014).

As a result of training, competencies usually improve by one step, typically from "no competency" to "little competency" or from "some competency"

to "good competency" (Qin & D'ignazio, 2010; Schmidt & Holles, 2018; Wright & Andrews, 2015). According to Peters and Vaughn (2014), based on the participants' self-assessment (n=65) after the NECDMC workshop, the competencies were mostly good. In a survey after five RDM courses held in 2013–2017, 77% of the respondents (n=31) considered the course useful, and 58% said they were interested in advanced education when available (Wiljes & Cimiano, 2019). In the feedback of the four clinical RDM workshops, respondents (n=113), who were mainly project coordinators, faculty members, and managers, expressed a need to learn RDM from many viewpoints and aspects like IPR, data security and privacy, and data curation (Read, 2019).

Participants have typically requested more practical exercises, discipline-specific cases, hands-on learning, and interactivity to concretise generic RDM principles to develop the training and deepen their competencies (Adamick et al., 2013; Byatt et al., 2013; Chew et al., 2021; Pascuzzi & Sapp Nelson, 2018; Wiljes & Cimiano, 2019). Nevertheless, fictitious cases not closely connected to participants' own research have been stated as uninteresting in feedback (Peters & Vaughn, 2014). Participants were most interested in learning more about data types and formats, archiving and long-term preservation, and metadata in the post-NECDMC workshop survey by Peters and Vaughn (2014), as well as data sharing, IPR, and legal issues. Participants were also interested in gaining more information on metadata and data security issues in the post-course survey of the NECDMC application by Muilenburg et al. (2014). In general, interactivity, discussion, peer supporting, and letting students apply generic principles in their own data are ways of concretising RDM (Peters & Vaughn, 2014; Read et al., 2019; Wright & Andrews, 2015).

Educational interventions in RDM are usually coordinated and led by libraries in academic institutions. Ideally, they begin with contextualising the education and determining the researchers' practices and needs via interviews, surveys, work shadowing, or focus groups (Kafel et al., 2014; Oliver, 2017; Qin & D'ignazio, 2010). In some cases, there has been a multi-professional steering group or committee, under which library is leading, and usually also carrying out the implementation (Kafel et al., 2014; Piorun et al., 2012). The library has been the main, and many times, the only actor arranging and implementing education on RDM. However, in interviews and surveys with students and researchers, the fact that data management needs are unrestricted to informational and consulting services typically delivered by the

library has become evident (Joo & Peters, 2020; Oliver, 2017). Examples are creating RDM guidance, helping with data management plans, and planning and implementing education. Librarians may lack expertise in technical RDM assistance or using data science tools for data analysing, visualising, coding, cleaning, and database building (Cerny, 2021; Read, 2019). Librarians are not necessarily the best advisors on ethical and legal issues or safe and secure storage, either (Cerny, 2021; Peters & Vaughn, 2014). Thus, many educators are planning an increased collaboration in data management training and support with researchers, libraries, research IT, legal services, research funding, and research offices (Castle, 2019; Cox & Pinfield, 2014; Joo & Peters, 2020; Latham, 2017; Oliver, 2017; Peters & Vaughn, 2014; Read, 2019; Revez, 2018; Verbaan & Cox, 2014; Wittenberg & Elings, 2017; Yu, 2017).

## 3. Methods

### 3.1. Course Backgrounds

Our research goal is to find how generic, multi-stakeholder training can improve participants' competencies and further comprehension of the relevance of sound research data management practices regarding the quality and integrity of data and reliability of the research. The methods section will describe how we aimed at these goals with the versatile expertise of the course designers and teachers, as well as the learning objectives, course structure, and contents. We will also describe how we analysed the results of the training.

The analysed BRDM course was developed and implemented at the University of Turku (UTU), the third-largest research-intensive university in southwestern Finland with eight faculties, five independent units, and 21,000 students including 2,000 doctoral students and 3,300 staff members. The data policy of UTU (2016) motivated the planning of the studied course, according to which researchers would be offered training and support for writing DMPs and data managing during the project's lifecycle. The OpenUTU project group, containing members from the research office, library, research computing services, legal affairs, and communications unit of UTU, created the data policy. The library was responsible for creating and coordinating trainings and support in RDM for researchers. Because developing education for all researchers

was impossible, the head of library services (the author) suggested starting with doctoral students (DSs) and postdoc researchers (PdRs) in a prime position to learn sound RDM practices from the beginning of their career. The author interviewed 35 doctoral students, supervisors, and biostatisticians in UTU to learn the perceived importance of RDM competencies and doctoral students' current competencies (Rantasaari, 2021; Rantasaari & Kokkinen, 2019). Data management planning, documentation of data processing, and managing IPR and contract issues contained the most profound skills gaps. However, participants also lacked knowledge of different issues throughout the data lifecycle. Therefore, the author, with the leader of UTU's biostatistician team, set up a working group and invited researcher-teachers from different faculties, a grant writer, data librarians, lawyers, a data security officer, and an IT computing specialist to plan and teach a course on RDM for DSs and PdRs. In 2020 we extended the course to Turku's other university – Åbo Akademi University (ÅAU) – the only Swedish language multi-faculty university with 5,500 students and 1,100 staff members in Finland and with whom UTU has a long tradition of joint projects.

### 3.2. Learning Objective, Course Structure and Data Management Plans in BRDM

A participant's learning objective was to familiarise themselves with RDM's central concepts and develop a high-class research plan and data management plan (DMP). After completing the course, a participant comprehends the significance of well-documented FAIR data for the ongoing study and other potential use and users, applying safe and secure practices in collecting, producing, handling, storing, sharing, and preserving the data, and acknowledging IPR, privacy, and sensitivity considerations when needed.[3]

Though BRDM is a generic and introductory course, we separated the course for different tracks. The preliminary idea behind the track-based division was that the data management actions needed and applied depend partly on the type of the data, partly on research methods, and partly on discipline (Aker & Doty, 2013; Joo & Peters, 2020; Lefebvre et al., 2018; Scholtens et al., 2019; Weller & Monroe-Gulick, 2014). These underlying factors delineate what kind of contracts, usage rights, storing solutions, processing, reuse, and preserving is needed or possible. For example, the methods in the clinical

health sciences are usually experimental or observational[4]; data are often identifiable, confidential, and highly sensitive. In the natural sciences, methods are typically experimental, observational, or simulation-based[5]; data is largely not confidential and sensitive. However, there can be other rigorous demands for handling, storing, and preserving large data sets. In survey and qualitative research, the data and its needed and possible actions can vary greatly, depending partly on discipline and partly on each respondent's or interviewee's answers, the study subject's activity, and so forth.

In 2019, the first year, the BRDM course consisted of three tracks (Clinical Health Sciences, Survey Research, and Natural Sciences), with seven face-to-face modules in Finnish for DSs and PdRs at UTU. In each track, participants were to prepare a shared DMP together during the course. A DMP was based on a fictitious research plan delivered by the faculty teacher-researchers in Module One. The participants learned by familiarising themselves with pre-class materials and preparing assignments on Moodle, after which they attended a lecture on the module.

In 2020, the course began with a joint introductory lecture with all the four tracks (Clinical Health Sciences, Survey Research, Interview Research, and Natural Sciences). The course was developed for DSs and PdRs of UTU and ÅAU. Clinical Health Sciences and Survey Research tracks were held in Finnish, whereas Interview Research and Natural Sciences were held in English. The course was turned fully online via Moodle after the three first modules because of the COVID-19. Instead of preparing a fictitious research plan and DMP, everyone created their own research plan and a DMP. Course modules were linked by mapping each module with the sub-section(s) of the General Finnish DMP template[6] and adding an assignment to prepare and update a relevant section of the DMP before and after each module's workshop session. The last assignment was to return the DMP and give an anonymous peer review of another participant's DMP. Finally, the author of this article assessed and rated each DMP and gave a general level feedback of all the DMPs using Finnish DMP Evaluation Guidance (FDEG) (Aalto et al. 2021). Otherwise, the learning followed the 2019 pattern, consisting of pre-class activities followed by a lecture on Zoom.

In 2021, the course was online from beginning to end and adapted a flipped classroom method for teaching. The course continued with the same four

track structure used in 2020 except the Interview Research track was turned to the Qualitative Research track. In each module, participants introduced themselves with the modules' pre-class materials in Moodle and drafted a relevant section of their DMP for themselves. The participants also added questions to the discussion forum based on the pre-class materials and their own data. After pre-class activities, the module's Zoom workshop session was reserved for discussion based on the questions that participants had written beforehand or asked during the workshop. As in 2020, the modules' post-class assignment was to update a DMP's relevant section, informed by the discussion in the modules' workshop. Each participant returned their DMP and peer-reviewed another participant's DMP as a final assignment for the course, after which the author assessed and scored the DMPs (Table 1).

### 3.3. Formative Assessment: Feedback

Following each module, participants were asked to give formal feedback through an online form (Appendix B). Module-based feedback was used as a formative assessment to control the participants' learning, receive information on experienced challenges, and collect proposals for improving the course. Hence, feedback produced ongoing information for the teachers to

*Table 1: The structure, contents, and responsible teachers of the four-track BRDM course.*

| CLINICAL HEALTH SCIENCES | SURVEY RESEARCH | QUALITATIVE RESEARCH | NATURAL SCIENCES | TEACHERS |
|---|---|---|---|---|
| Introductory Lecture:<br>Background and concepts; Characteristics of a high quality research plan; Course practicalities | | | | Head of library services;<br>Data librarian; Grant writer |
| Research plan:<br>Objective; design;<br>implementation; expected results | Research plan:<br>Objective; design;<br>implementation; expected results | Research plan:<br>Objective; design;<br>implementation; expected results | Research plan:<br>Objective; design;<br>implementation; expected results | Researchers;<br>Lectors;<br>University teachers |
| Data management plan (DMP):<br>data life cycle; DMP-tool; roles; resources; metadata;documentation | | | | Self-study module;<br>Data librarian |
| IPR, agreements and licenses<br>(in Finnish) | | IPR, agreements and licenses<br>(in English) | | Head of legal affairs unit;<br>Assistant legal counsel;<br>Data librarian |
| Data privacy:<br>privacy notice; risk analysis;<br>(in Finnish) | | Data privacy:<br>privacy notice; risk analysis; anonymization<br>(in English) | | Data protection officer;<br>Data archive specialist |
| RedCap:<br>building a form based database | RedCap:<br>building a survey form | NVivo:<br>organizing and coding data | RedCap:<br>building a form based database | Head of biostatistician team; Lector |
| Data storage, protection, processing, describing and IT Service solutions | | | | IT system architect |
| Data preservation, sharing and citing; Open data repositories | | | | Data librarian |
| A voluntary Q&A Session | | | | Module teachers |
| DMP:<br>returning and peer-reviewing | DMP:<br>returning and peer-reviewing | DMP:<br>returning and peer-reviewing | DMP:<br>returning and peer-reviewing | Head of library services;<br>Data librarian |
| a general level feedback on DMPs | | | | Head of library services |

edit and enhance the modules and their contents. Moreover, halfway through the course, the author compiled the feedback with answers and information about remedies that were made or would be made.

The author used the grounded theory-inspired approach to code and analyse the feedback for this study (Bryant & Charmaz, 2016; Cassidy, 2012; Timonen et al., 2018). Sub-categories were created based on the topics that emerged from the coded comments. Grounded theory as an analysing approach is well-suited for processing and analysing the feedback as the aim was to let the feedback data speak for itself and not use an existing theoretical framework and categories formulated according to the framework.

### 3.4. Summative Assessment: Survey

The participants were asked to participate in a survey to self-rate their competencies before and after the course. In 2019, the survey was carried out twice – before and after the course – on a scale from 1 to 5: 1=no competence, 2=some competence, 3=good competence, 4=very good competence, 5=top competence (Appendix C). In 2020–2021, we performed a post-course survey in which participants were asked to self-rate their competencies before and after the course on a scale from 1 to 4: 1=no competence, 2=little competence, 3=somewhat competent, 4=very competent (Appendix D). Participants were also asked to give a course rating from 1–100, if they would recommend the course to other DSs and PdRs, and choose the topics about which they would like to have more education.

The survey served as a summative assessment for collecting participants' perceptions of their learning, the quality of the course, and further education needs.

The respondents' self-ratings of their competencies were analysed using JMP Pro 16 to produce descriptive and inferential statistics with medians, custom quantiles, Wilcoxon signed-rank test (one group), Wilcoxon rank-sum test (two independent groups), and Steel-Dwass test (multiple comparisons). Frequencies and Chi-square test were used for announcing further learning needs. A significance level of 0.05 (two-tailed) was used. Also, module- and course-based feedback comments were coded and categorised in NVivo 12.

### 3.5. Summative Assessment: Data Management Plans

In Module One, each participant created their own research plan. During the course and based on their research plan, they wrote a DMP using a Finnish General DMP template and guidance. The course participants' DMPs will be analysed in a later study.

## 4. Results

### 4.1. Participants

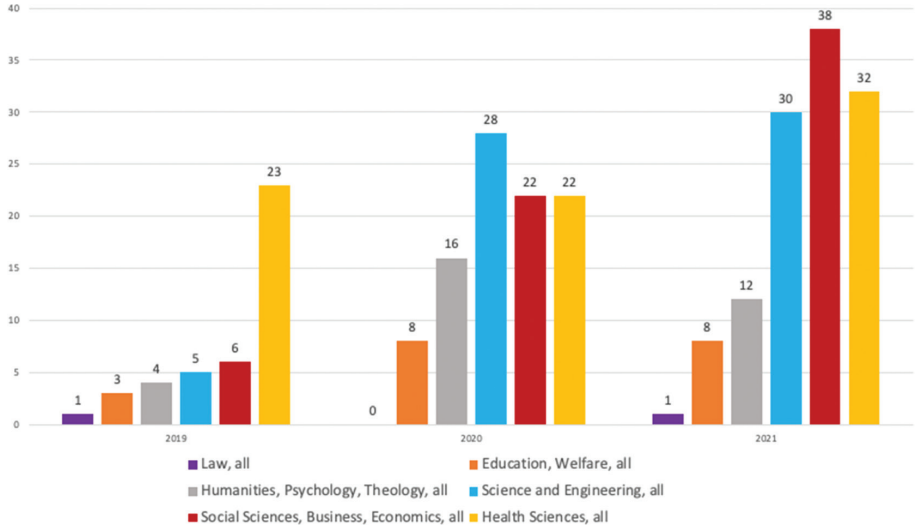
Of the 386 enrolled participants in 2019–21, 346 (90%) were DSs, 37 (10%) were PdRs, and 3 (1%) were university employees. Of those who completed the full course with 3 ECTS credits, 154 (91%) were DSs, 14 (8%) were PdRs, and 1 (1%) was a university employee. Of the participants who did not complete the full course but (on average) half the modules, 72 (80%) were DSs, 17 (19%) were PdRs, and 1 (1%) was a university employee. In 2019, participants who did not complete the full course performed (on average) 3 out of 7 modules; in 2020, 3 out of 8; and in 2021, 4 out of 8. Performing only part of the modules does not mean that participants interrupted the course but that the modules were performed evenly between modules 0 (introduction) and 8 (final assignment). PdRs, in particular, picked modules according to their interests, without needing to earn the 3 ECTS credits (Table 2).

The largest disciplines represented regarding the number of participants in 2019–2021 (n=259) were Health Sciences with 77 (30%) participants; Social, Business and Economics with 66 (26%) participants; and Science and

*Table 2: Enrolled, fully completed, and (on average) approximately half the modules performed in 2019–2021 courses.*

| Status | Enrolled | Fully completed | Performed on average half of the modules |
|---|---|---|---|
| Doctoral Student | 346 | 154 | 72 |
| Postdoc Researcher | 37 | 14 | 17 |
| University Employee | 3 | 1 | 1 |
| | 386 | 169 | 90 |

*Fig. 2: The number of all participants by discipline in 2019–2021 courses (n=259).*




Engineering with 63 (24%) participants. Markedly fewer participants came from Humanities with 32 (12%) participants; Education with 7 (7%) participants; and Law with only 2 (1%) participants. (Figure 2; Table 2 in Appendix A).

### 4.2. Feedback

We asked participants to fill in a feedback form after each module on the Moodle course platform. In 2019, the module-based feedback was a mandatory course assignment. This task was voluntary in 2020 and 2021, echoed in the number of feedback forms we received: 133 forms in 2019, 114 in 2020, and 69 in 2021. In 2019 and 2020, participants were given a time slot at the end of the classes or workshops to provide feedback; in 2021, we simply reminded participants to give feedback after the live Zoom workshop sessions.

The feedback form contained three main categories:

1. What are the three things you have learned?
2. How will the things you have learned change your practices?
3. How would you suggest the module be developed?

Under these main categories, the author created sub-categories and sorted the comments using a grounded theory-inspired approach. The five biggest sub-categories stand for 90 to 100% of all comments in the main categories (Figures 3 to 5; Tables 3 to 5 in Appendix A). Because a respondent's comment in a feedback form could include several aspects, it could be placed accordingly in two or more sub-categories. For example, the comment "I am now more aware of IPR issues and GDPR, which enables me to plan my next research in more detail" has been placed in the sub-categories "I will pay notice to IPR, agreements and licenses", and "I will pay more notice to data privacy and data security". Hence, the total number of comments in different sub-categories is bigger than the number in the original four main categories.

#### *4.2.1. What are the Three Things you have Learned?*

Based on the feedback given in this main category, we identified 252 separate comments in 2019, 194 in 2020, and 119 in 2021. The sub-category "What, why

*Fig. 3: Five top sub-categories based on the feedback given in the main category "What are the three things you have learned" in 2019–2021?".*

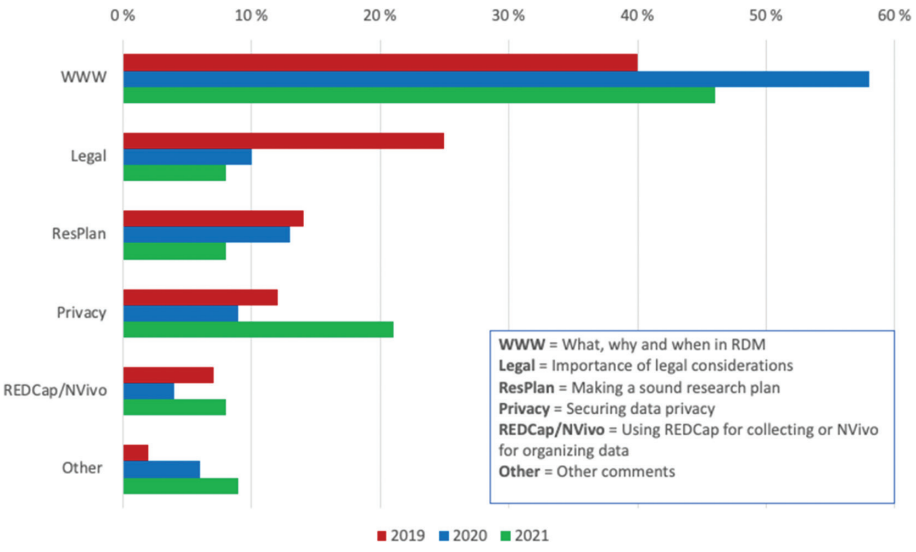


*Fig. 4: Five top sub-categories based on the feedback comments given in the main category "How will the things you have learned change your practices?".*

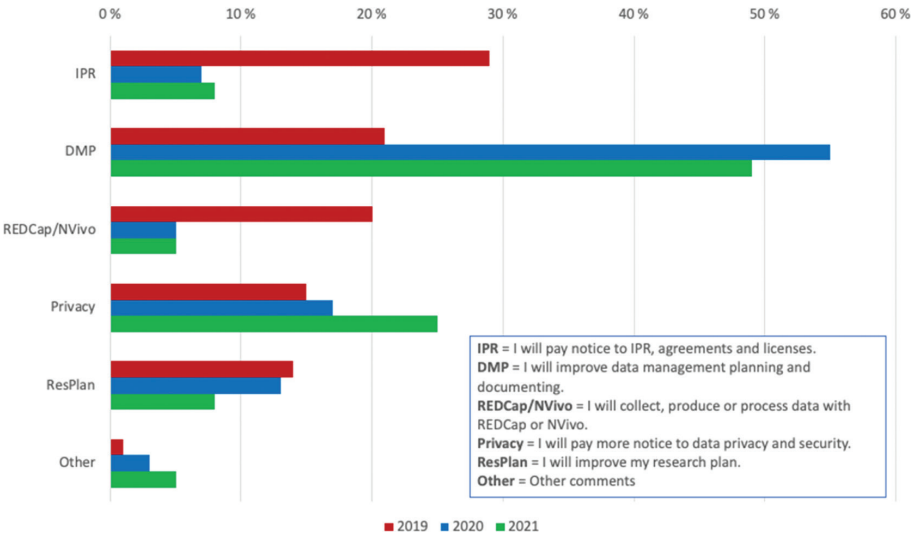
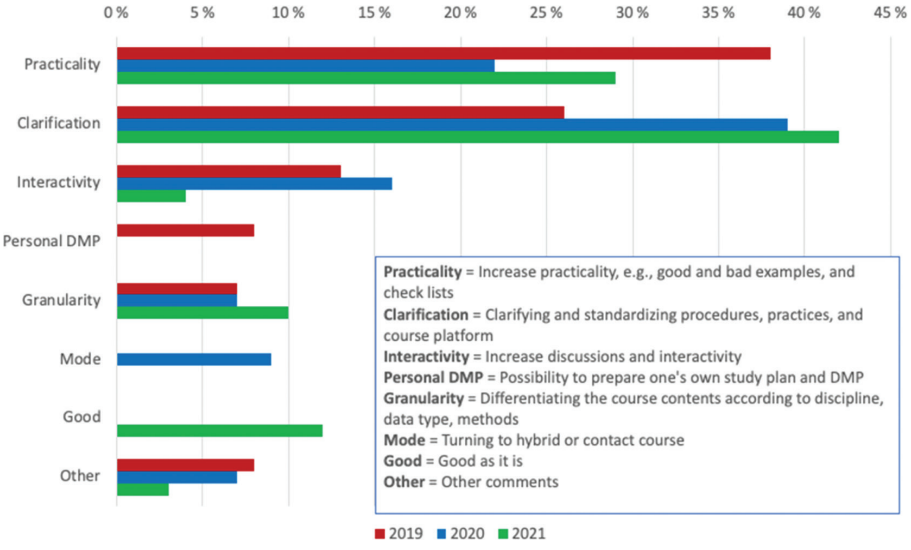



*Fig. 5: Seven top sub-categories based on the feedback comments given in the main category "How would you suggest the module be developed?".*



and when in RDM" contains comments of RDM essentials such as the learning of the rationale and tools to plan data management, the central concepts of RDM, and the different phases such as storing, documenting, preserving, and sharing of data:

> *"(I learned) ways to conceptually approach Research Data Management and the practices and perspectives related to it."* (Module 2, Qualitative Research, ID 16/2021).

The number of the comments concerning data management planning and documentation grew in 2020 and 2021:

> *"Documentation of data in a clear and readable form is a crucial step in the data management and processing."* (Module 5, Natural Sciences, ID 91/2020).

At the same time, the percentage of the comments belonging to the sub-category "The importance of legal considerations" dropped from 25% in 2019 to 8% in 2021. (Figure 3; Table 3 in Appendix A).

#### *4.2.2. How will the Things you have Learned Change your Practices?*

Based on the feedback given in this main category, we identified 87 comments in 2019, 132 in 2020, and 63 in 2021. The comments concerning improving data management planning and documenting practices increased from 21% to 49%, whereas the comments about the intention to focus more on IPR, agreements, and license issues decreased from 29% to 8%. The following quotation illustrates the increased number of comments concerning documentation's importance:

> *"I learned a lot about the importance of documentation and metadata as well as publishing datasets. I will apply the FAIR principles when my research work needs to be checked and will review the data management all the time." (Module 7, Qualitative Research, ID 68/2021).*

At the same time, the percentage of data privacy comments (e.g., data privacy notice, informed consent, and GDPR), and data security comments (safe and secure data storing platforms), increased from 15% to 25%. (Figure 4; Table 4 in Appendix A).

#### *4.2.3. How would you Suggest the Module be Developed?*

We identified 90 (2019), 136 (2020), and 101 (2021) proposals to develop the modules. Most of the respondents wished for practicality such as more discipline-specific instruction, checklists, and cases, along with the clarification and standardisation of course practicalities, schedules, and course platforms, and how to balance the workload between different modules. More practicality and concreteness were desired, especially in law-related modules three and four:

> *"All the law-related sections could explain things in less of a law-speech manner as law speech is generally really vague and does not provide any practical knowledge. In general, the wideness of topics was really good." (Post-Course Survey, Natural Sciences, ID 53/2021)*

In 2019 and 2020 (but not in 2021), many comments expressed a desire for more interactivity and discussions. Unlike in 2020–2021 courses, participants preparing their own research plan and DMP in 2019 – visible in the 7 (8%) comments – was impossible. For the first time in 2021, we received 12 (12%) answers that the module is good as is. (Figure 5; Table 5 in Appendix A).

#### *4.2.4. The Overall Score*

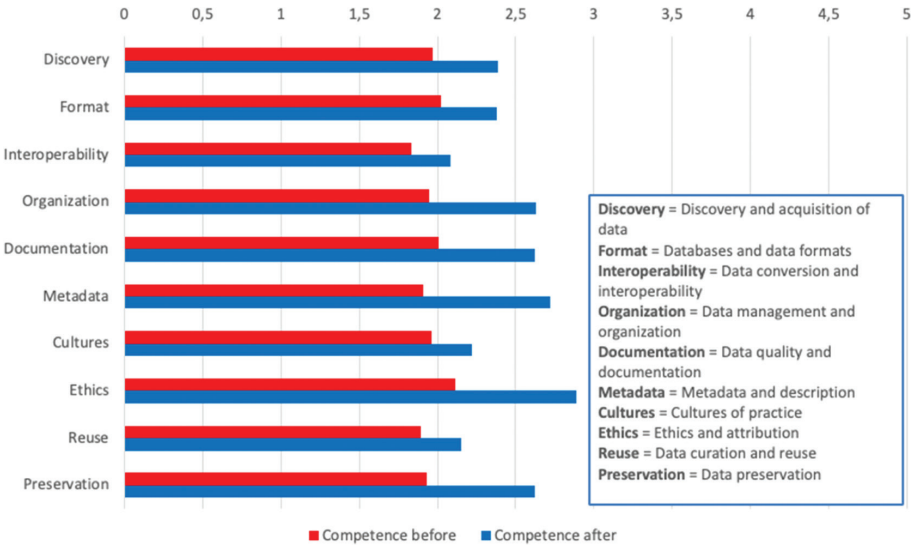
In the post-course surveys after the 2020–2021 courses, participants were asked to score the course between 0 and 100. After the 2021 course, participants were also asked if they would recommend the training to other DSs or PdRs. Based on the survey respondents' general score of 68 out of 100 in 2020 (n=53) and 74 out of 100 in 2021 (n=64), the course lived up to the reasonable expectations of a general level introductory education. Equally, 92% of the post-course survey respondents in 2021 expressed they would recommend the course to other DSs and PdRs.

### 4.3. Surveys 2019–21: Competencies in RDM before and after the Course

#### *4.3.1. BRDM 2019*

Participants were asked to rate their current RDM competencies on a five-point scale from 1 to 5 before and after the 2019 course (Appendix C). Hence,

*Fig. 6: Respondents' median self-ratings of their competencies before and after the BRDM 2019 course.*




45 (82%) enrolees answered the pre-course survey, and 17 (41%) of those who completed at least part of the modules answered the post-course survey. Before the course, participants' median self-rating of their RDM competence was 1.96 (Q1:1.82, Q3:2.09). After the course, participants' median self-rating of their competence was 2.32 (Q1:2.12, Q3:2.84). The improvement was statistically significant, $p=0.003$ (Wilcoxon rank-sum test), or 0.36 points. (Figure 6; Table 6 in Appendix A).

#### *4.3.2. BRDM 2020–2021*

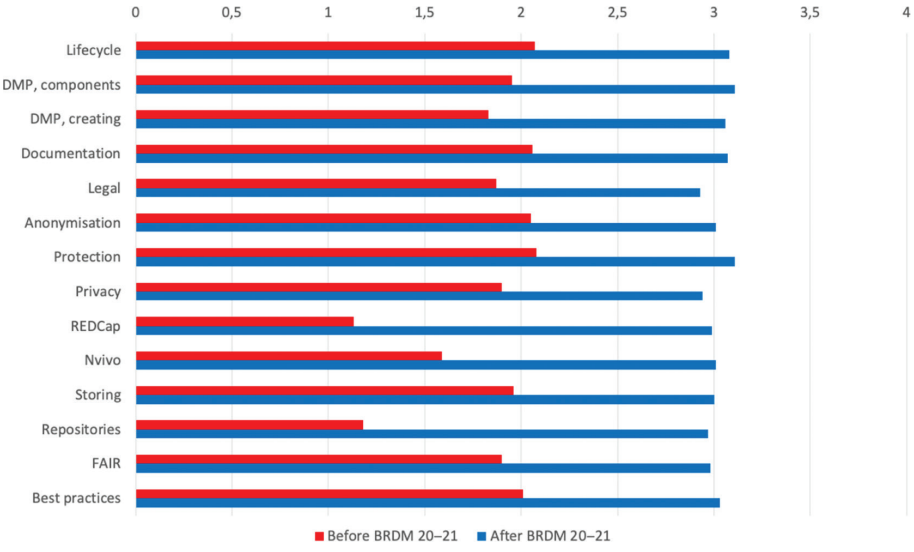
The surveys in 2020 and 2021 (Appendix D) differed from the 2019 survey related to the contents and execution:

- Instead of pre- and post-course surveys, we only had a post-course survey.
- The competencies were specified to respond more closely to the learning objectives of the modules (Rantasaari et al. 2021).
- The scale was 1 to 4 instead of 1 to 5.

The combined response rate to the surveys was 49% (106 respondents out of 217 participants) after the 2020–2021 courses. On the 1 to 4 scale, the median self-rated competence before and after the courses was 1.97 and 3.03, respectively. Thus, the median self-rated competencies improved statistically highly significantly, $p<0.0001$ (Wilcoxon signed-rank test), or 1.06 points (Figure 7; Table 7 in Appendix A).

Regarding the variance in the results between disciplines and course tracks, differences were statistically insignificant at the level of total medians, although some were found concerning a few specific competencies before the course. First, respondents in the "Qualitative Research" track and the "Humanities, psychology, and theology" discipline self-rated their competence higher than those in the "Clinical Health Sciences" track in identifying the data life cycle and recognising a DMP's components ($p=0.02$, Steel-Dwass). Second, respondents in the "Qualitative Research" track and the "Humanities, Psychology, and Theology" and the "Social Sciences, Business, and Economics" disciplines self-rated their competence in applying anonymisation higher than those in the "Natural Sciences" track ($p=0.01$,

*Fig. 7: Based on median self-ratings, respondents' competencies related to the specified learning objectives before and after the BRDM 2020–2021 courses. Light blue bars represent the competencies before the courses and dark blue bars after the courses. Full descriptions of the learning objectives can be found in the survey form (Appendix D).*

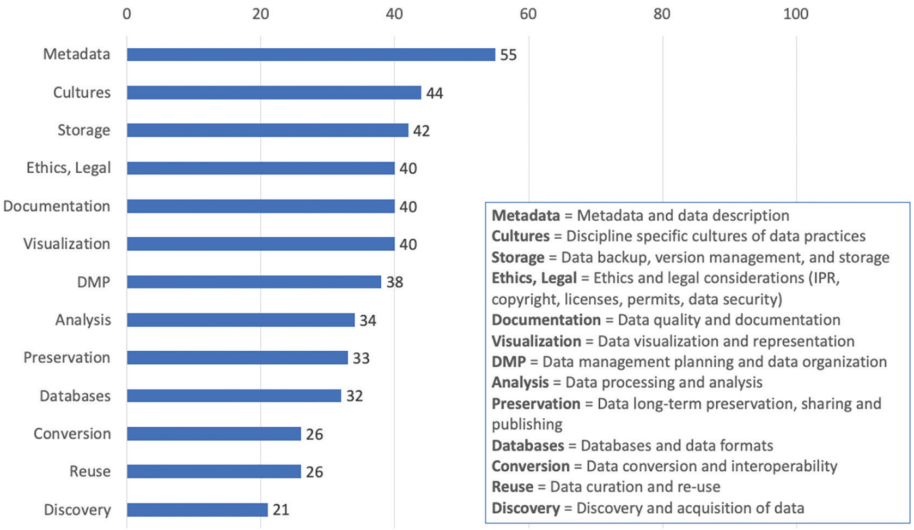


Steel-Dwass) and the "Science and Engineering" discipline (p=0.02, Steel-Dwass). Third, respondents in the "Social Sciences, Business, and Economics" discipline and the "Qualitative Research" track self-rated their competence in applying data privacy higher than those in the "Science and Engineering" discipline and the "Natural Sciences" track (p=0.02, Steel-Dwass). All differences after the course were insignificant.

### 4.4. Subjective Educational Needs in 2020–2021: What would you like to Learn more about?

Participants were asked to choose the topics they wanted to learn more about in the post-course surveys. As much as 102 respondents (96%) expressed interest in advanced training. Six topics receiving over half (261) of all mentions (471) were "Metadata and description" (55), "Discipline-specific cultures" (44), "Backup, version management, storage" (42), "Ethics and legal considerations" (40), "Quality and documentation" (40), and "Visualisation and representation" (40). However, interest for advanced training in "Discovery and acquisition" (21) and "Data curation and reuse" (26) were the lowest. Differences were statistically insignificant related to respondents'

*Fig. 8: The topics respondents would like to learn more about.*



discipline or course track. The frequencies of mentions for further learning needs are illustrated in Figure 8.

## 5. Discussion and Conclusions

### 5.1. How did the Course Succeed?

In this article, our goal was to find how generic, multi-stakeholder training could improve participants' competencies and further comprehension of the relevance of sound research data management practices to the quality and integrity of data and reliability of the research. Furthermore, the questions we aimed to answer were as follows: RQ1) How did the course participants self-rate their RDM competencies before and after the course? RQ2) What kind of educational impact did the course have on participants' RDM competencies (knowledge, skills, and abilities) based on participants' self-ratings and the collected and categorised feedback? RQ3) What further learning needs did the respondents express after the course? These questions will be discussed with the help of the criteria for successful training as created by Oo et al. (2021).

Based on the systematic review of 28 RDM trainings between 2012 and 2019, Oo et al. (2021) introduced a four-part criterion for successful training consisting of 1) active participation during training; 2) demand for RDM training; 3) increased participants' knowledge and understanding of RDM and confidence in enacting RDM practices; and 4) positive post-training feedback. How BRDM matched these criteria will be discussed below.

Concerning the participation during training, BRDM was based on active learning: Participants read and listened to course materials, completed assignments, developed their own research plan and a DMP, peer-reviewed each other's DMP, drafted questions based on course materials and their own data management issues, and participated in the workshop discussions. The activities sought to help participants link the principles and other theoretical contents to their research practices (see also Thielen et al., 2017; Whitmire, 2015; Wiljes & Cimiano, 2019; Wittenberg & Elings, 2017). Judging by the feedback during and after the 2021 course, we succeeded in bringing interactivity and discussion to modules. However, there was still a demand for turning, especially legal and data privacy principles and regulations, into concrete instructions, cases, and examples when possible.

After completing the course, almost all respondents expressed interest in further education for RDM training. The most frequently mentioned topics for further learning were "Metadata and description", "Discipline-specific cultures", "Backup, version management, storage", "Ethics and legal considerations", "Quality and documentation", and "Visualisation and representation". Metadata, ethics, and legal issues were also the most wanted topics for continued learning in the courses of Muilenburg et al. (2014) and Peters and Vaughn (2014). Conversely, despite emphasising FAIR principles, as well as data sharing and reuse throughout BRDM, advanced training in "Discovery and acquisition" and "Data curation and reuse" were the least preferred topics. Minor interest in these might be comprehensible concerning cultures of practices in many disciplines where researchers' primary interest is getting their current project through and obtaining results from the data rather than long-term preservation and the possible data reuse in future projects (Kowalczyk, 2017; Rantasaari, 2021).

Concerning increased knowledge, understanding, and confidence in enacting RDM practices, participants highlighted that they had learned RDM essentials such as understanding the rationale and learning the tools to plan data management, RDM's central concepts, and storing, documenting, preserving, and sharing data. Moreover, participants learned legal and data privacy issues and how to use REDCap and NVivo in data collecting and organising. Correspondingly, they reported that the training would change their current practices in planning research projects, managing and documenting data, acknowledging legal and data privacy viewpoints, and using REDCap and NVivo in data collecting and organising. The median self-rated improvement in RDM competencies was 0.36 points in 2019 and 1.06 in 2020–2021 – one level up from "little competence" to "somewhat competent". One-step improvement during a generic RDM course is a typical result that has been documented in several post-course surveys (e.g., Qin & D'ignazio, 2010; Wright & Andrews, 2015).

As far as respondents' disciplines or course tracks are concerned, the differences were statistically insignificant at the level of total medians. However, some significant differences were found concerning a few specific competencies before the course. Respondents in the "Qualitative Research" track and the "Humanities, Psychology, and Theology" and "Social Sciences, Business, and Economics" disciplines self-rated their anonymisation competencies before the course as higher than those in the "Natural Sciences" track and the

"Science and Engineering" discipline. Likewise, respondents in the "Social Sciences, Business, and Economics" discipline and the "Qualitative Research" track self-rated their data privacy management competencies before the course as higher than those in the "Natural Sciences" track and the "Science and Engineering" discipline. These differences are comprehensible because data in qualitative research and social sciences, more often than in the natural sciences and engineering, contain personal or even sensitive contents. However, the differences had disappeared after the course. This indicates that the course had bridged the gaps in respondents' competencies regarding their disciplines and course tracks. On applying the RDM principles in participants' own data management planning, the results of the assessment and rating of the returned DMPs in the BRDM 2020–2022 courses will be reported in another upcoming article.

Pertaining to feedback, the course was perceived as a solid and important introduction to RDM's different aspects. Teachers – including a grant writer, researchers, data librarians, lawyers, a data privacy officer, a data archive specialist, a biostatistician, and an IT professional – were appreciated as real domain experts. However, regarding propositions for course development, respondents asked for a more down-to-earth approach, concretising, and examples, especially in legal and data privacy issues. Clarification of the course platform and course practicalities were also requested.

BRDM can be determined as one of the few trainings (so far) that meet all parts of the four-part criteria for successful training as defined by Oo et al. (2021). However, because of BRDM's limited number of participants, we cannot generalise our study's results and the factors affecting them outside the studied group. Furthermore, we cannot know the long-term impact of the participants' self-rated competencies on their RDM activities without follow-up. Still, 319 returned module-based feedback forms, and 168 survey responses revealed valuable, indicative information of doctoral students' and postdoc researchers' competencies, the impact of the education on competencies, and further learning needs in RDM.

### 5.2. The Value of BRDM and Lessons Learned

BRDM is an educational effort bringing value to RDM training. So far, academic libraries have been the main, and many times the only, actor arranging

and implementing education on RDM in research-intensive universities. As a further development need, educators have often mentioned a need for collaboration with multiple stakeholders (Castle, 2019; Cox & Pinfield, 2014; Joo & Peters, 2020; Latham, 2017; Oliver, 2017; Peters & Vaughn, 2014; Read, 2019; Revez, 2018; Verbaan & Cox, 2014; Wittenberg & Elings, 2017; Yu, 2017). In BRDM, using versatile expertise in planning and teaching has been embedded from the beginning: Academic and research support experts planned and taught the course. Second, the contents of BRDM were wide-ranging containing most of the phases of data life cycle, beginning from the writing of a high-class research plan – which makes this course unique – to the sharing and long-term preservation of the data. However, limited resources excluded more technical data science contents such as analysing, visualising, cleaning, merging, and programming data. Third, participants applied sound RDM principles in their data management by writing a DMP during the course. Hence, assessing BRDM's results is based not only on the feedback and self-rating of the participants' competencies with further learning needs (typical measures of success in many previous trainings) but the returned DMPs. Fourth, a flipped classroom approach that is rarely used as a teaching method in previous RDM training (Griffin, 2020; Johnston & Jeffryes, 2015; Mithun & Luo, 2020), was adapted in the BRDM 2021 course. Fifth, many previous RDM trainings have been criticised for inadequate reporting (Goben & Griffin, 2019; Perrier et al., 2017). In this study, we aimed for extensive and precise reporting.

Next, we will present some concrete lessons that we have learned during planning, implementing, and analysing the results of BRDM in 2019–2021.

Planning and implementing training with multiple RDM stakeholders enable acknowledging all relevant aspects of the data life cycle. Participants will get an overall view of the numerous factors affecting RDM, while stakeholders' overall understanding of the RDM and the challenges doctoral students and postdoc researchers confront increase. The downside of a large working group and many teachers is the administrative burden in coordinating the training. Moreover, the pedagogical skills of multi-professional specialists can be diverse. Therefore, keeping the training coherent by reaching a consensus on the learning objectives, teaching methods and contents, course practicalities, and deadlines with all the teachers and working groups is paramount.

RDM is an organic part of a research project – from planning the goal and research questions and proceeding to the methods of collecting, producing,

processing, storing, sharing, and preserving the data. Thus, recalling and updating or, preferably, rewriting a research plan is important when developing a DMP. Otherwise, research and data management plans can be asynchronous for example regarding data types to be collected, produced, and reused in a project.

Collecting feedback throughout training serves as a formative assessment to control the participants' learning, receive information from experienced challenges, and gather proposals to quickly improve the training.

Though planning and implementing the flipped classroom approach takes a lot of work from teachers, it pays back by increasing flexibility and helping activate participants. Still, quizzes or follow-up tasks are essential to show that participants learned the pre-class materials, as Mithun and Luo (2020) have pointed out.

Measuring the learning results should not be based solely on the participants' self-assessment or feedback, but on the assessment of assignments such as DMPs developed during training. Moreover, a follow-up intervention would be needed to collect empirical evidence on how the planned actions in DMPs have been applied in research practice (see also Perrier et al., 2017).

A modular structure enables cherry-picking the training, reducing the drop-out rate. For example, PdRs do not necessarily need credits or certification by completing a training but want to bridge their knowledge gaps by choosing the modules that interest them.

Finally, according to this study and many others (e.g., Chew et al., 2021; Pascuzzi & Nelson, 2018; Wiljes & Cimiano, 2019), there is never too much IPA (Interaction, Practice, and Application) in training. Still, participants achieving excellent competence in a basic training are improbable. Instead, discipline, data type or research method specific workshops with fewer participants will help deepen the elementary skills (e.g., Petters et al., 2019; Read, 2019; Thielen & Hess, 2017). However, as highlighted in research literature, training without synchronised incentives, policies, processes, and infrastructure is insufficient to bring about behavioural change (Chew et al., 2021; Perrier et al., 2020). A realistic target for a generic training could be that participants become aware of RDM and its contents and gain the elementary tools and basic skills to begin applying sound RDM practices in their research processes. Moreover,

introducing participants to support services of multiple RDM stakeholders is important. That stakeholders learn what kind of challenges researchers and research students encounter when applying RDM is equally important.

## 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 analyses; and Information Specialist Päivi Kanerva for acting as a responsible teacher of the BRDM course.

## Availability of Data

The quantitative data underlying this study can be accessed through Zenodo: https://doi.org/10.5281/zenodo.6526121.

## References

Aalto, S., Ahokas, M., Friman, J., Fuchs, S., Korhonen, T., Kuusniemi, M. E., Laakso, K., Lennes, M., Manninen, S., Ojanen, M., Rantasaari, J., Virtanen, M. E., & Xu, Q. (2021). *Finnish DMP evaluation guidance*. [Working paper]. Zenodo. https://doi.org/10.5281/zenodo.4729831

Academy of Finland. (2019). *The Academy of Finland's funding terms and conditions 2019–2020* (Vol. 1, Issue September 2019). https://www.aka.fi/globalassets/10rahoitus/liiteet/rahoitusehdot_2019-2020_en.pdf

Adamick, J., Reznik-Zellen, R. C., & Sheridan, M. (2013). Data management training for graduate students at a large research university. *Journal of eScience Librarianship*, *1*(3), 180–188. https://doi.org/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. https://doi.org/10.2218/ijdc.v8i2.263

Amsterdam call for action on open science. (2016, April 5). In *Wikipedia*. https://en.wikipedia.org/wiki/Amsterdam_Call_for_Action_on_Open_Science

Borghi, J., Abrams, S., Lowenberg, D., Simms, S., & Chodacki, J. (2018). Support your data: A research data management guide for researchers. *Research Ideas and Outcomes 4*, Article e26439. https://doi.org/10.3897/rio.4.e26439

Borycz, J., & Carroll, A. J. (2021). COVID-19 as an opportunity to expand the instructional portfolio of STEM librarians. *Issues in Science and Technology Librarianship*, *98*. https://doi.org/10.29173/ISTL2609

Briney, K. (2015). *Data management for researchers: Organize, maintain and share your data for research success* (1st ed.). Pelagic Publishing.

Bryant, A., & Charmaz, K. (2016). *The SAGE Handbook of grounded theory*. SAGE Publications Ltd. https://doi.org/10.4135/9781848607941

Byatt, D., Scott, M., Beale, G., Cox, S. J., & White, W. (2013). *Developing researcher skills in research data management: Training for the future - a DataPool project report*. University of Southampton. https://eprints.soton.ac.uk/351026/1/REPORT-Supporting_Researchers-RDM-Training-Final.docx

Carpentries. (n.d.). *Data carpentry: Building communities teaching universal data literacy*. https://datacarpentry.org/

Cassidy, A. (2012). Vermin, victims and disease: UK framings of badgers in and beyond the bovine TB controversy. *Journal of the European Society for Rural Sociology*, *52*(2), 192–214. https://doi.org/10.1111/J.1467-9523.2012.00562.X

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. https://doi.org/10.1515/libri-2018-0064

Cerny, M. (2021). Digital Competences of students of library studies: Comparison of research results for 2018-2020. *Education Sciences*, *11*(11), 1–13. https://doi.org/10.3390/educsci11110729

Chew, A. W., Oo, C. Z., Wong, A. L., & Gladding, J. (2021). An initial evaluation of research data management online training at the University of New South Wales. *IFLA Journal*, *OnlineFirst*. https://doi.org/10.1177/03400352211054120

Chiarelli, A., Loffreda, L., & Johnson, R. (2021). *The art of publishing reproducible research outputs: Supporting emerging practices through cultural and technological innovation*. [Report]. Zenodo. https://doi.org/10.5281/ZENODO.5521077

Clement, R., Blau, A., Abbaspour, P., & Gandour-Rood, E. (2017). Team-based data management instruction at small liberal arts colleges. *IFLA Journal, 43*(1), 105–118. https://doi.org/10.1177/0340035216678239

Cole, G., & Evans, J. (2014). University of Exeter research data management and open access training for staff. *ALISS Quarterly, 10*(1), 22–25. https://alissnet.files.wordpress.com/2012/11/vol-10-no-1-oct-2014.pdf

Consortium of European Social Science Data Archives. (2017). *CESSDA training: Data management expert guide*. https://www.cessda.eu/DMGuide

Corti, L., & Van den Eynden, V. (2015). Learning to manage and share data: Jump-starting the research methods curriculum. *International Journal of Social Research Methodology*, *18*(5), 545–559. https://doi.org/10.1080/13645579.2015.1062627

Cox, A. M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, *46*(4), 299–316. https://doi.org/10.1177/0961000613492542

Doucette, L., & Fyfe, B. (2013). Drowning in research data: Addressing data management literacy of graduate students. *Imagine, Innovate, Inspire: The Proceedings of the ACRL 2013 Conference*, 165–171. https://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2013/papers/DoucetteFyfe_Drowning.pdf

European Commission. (2018a). *OSPP-REC: Open science policy platform recommendations*. OSPP-REC. https://doi.org/10.2777/958647

European Commission. (2018b). Commission recommendation (EU) 2018/790 of 25 April 2018 on access to and preservation of scientific information. *Official journal of the European Union L*, *134*, 12–18. http://data.europa.eu/eli/reco/2018/790/oj

European University Association. (2017). *Towards full open access in 2020: Aims and recommendations for university leaders and national rectors' conferences*. http://www.eua.be/Libraries/publications-homepage-list/towards-full-open-access-in-2020-aims-and-recommendations-for-university-leaders-and-national-rectors-conferences?platform=hootsuite

Federer, L. (2016). Research data management in the age of big data: Roles and opportunities for librarians. *Information Services & Use*, *36*(1-2), 35–43. https://doi.org/10.3233/ISU-160797

Goben, A., & Griffin, T. (2019). In aggregate: Trends, needs, and opportunities from research data management surveys. *College and Research Libraries*, *80*(7), 903–924. https://doi.org/10.5860/crl.80.7.903

Griffin, T. M. (2020). Centering graduate students' research projects in data management education: A pilot program. *Journal of Librarianship and Scholarly Communication 8*(1), Article eP2365. https://doi.org/10.7710/2162-3309.2365

Johnston, L., & Jeffryes, J. (2015). Civil engineering/ graduate students. *Data Information Literacy Case Study Directory 3*(1), Article 1. https://doi.org/10.5703/1288284315479

Joo, S., & Peters, C. (2020). User needs assessment for research data services in a research university. *Journal of Librarianship and Information Science*, *52*(3), 633–646. https://doi.org/10.1177/0961000619856073

Kafel, D., Creamer, A.T, & Martin, E.R. (2014). Building the New England Collaborative Data Management Curriculum. *Journal of EScience Librarianship*, *3*(1), 60–66. https://doi.org/10.7191/jeslib.2014.1066

Kowalczyk, S. T. (2017). Modelling the research data lifecycle. *International Journal of Digital Curation*, *12*(2), 331–361. https://doi.org/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. https://doi.org/10.1177/1833358319831318

Latham, B. (2017). Research data management: Defining roles, prioritizing services, and enumerating challenges. *The Journal of Academic Librarianship*, *43*(3), 263–265. https://doi.org/10.1016/j.acalib.2017.04.004

Lefebvre, A., Schermerhorn, E., & Spruit, M. (2018). How research data management can contribute to efficient and reliable science. *ECIS 2018 Proceedings*, 1–15. https://aisel.aisnet.org/ecis2018_rp/35

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. https://doi.org/10.1177/0162243918798906

Mithun, S., & Luo, X. (2020). Design and evaluate the factors for flipped classrooms for data management courses. *2020 IEEE Frontiers in Education Conference (FIE) 2020*, 1–8. https://doi.org/10.1109/FIE44824.2020.9274201

Muilenburg, J., Lebow, M., & Rich, J. (2014). Lessons learned from a research data management pilot course at an academic library. *Journal of eScience Librarianship*, *3*(1), 67–73. https://doi.org/10.7191/jeslib.2014.1058

Mustajoki, H. (n.d.). *Finnish open science online resource*. Finnish Open Science Online Resource. Ministry of Education and Culture, Finland. https://findocnet.fi/course/view.php?id=136

National Science Foundation. (2011). *Grant proposal guide* (NSF 11-1). https://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp

Oliver, J. C. (2017). Bioinformatic training needs at a health sciences campus. *PLoS One*, *12*(6), Article e0179581. https://doi.org/10.1371/journal.pone.0179581

Oo, C. Z., Chew, A. W., Wong, A. L. H., Gladding, J., & Stenstrom, C. (2021). Delineating the successful features of research data management training: A systematic review. *International Journal for Academic Development*. https://doi.org/10.1080/1360144X.2021.1898399

Pascuzzi, P. E., & Sapp Nelson, M. R. (2018). Integrating data science tools into a graduate level data management course. *Journal of eScience Librarianship*, *7*(3), Article e1152. https://doi.org/10.7191/jeslib.2018.1152

Pasek, J. 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*. https://doi.org/10.29173/istl12

Perrier, L., Blondal, E., Ayala, A. P., Dearborn, D., Kenny, T., Lightfoot, D., Reka, R., Thuna, M., Trimble, L., & MacDonald, H. (2017). Research data management in academic institutions: A scoping review. *PLoS One, 12*(5), Article e0178261. https://doi.org/10.1371/journal.pone.0178261

Perrier, L., Blondal, E., & MacDonald, H. (2020). The views, perspectives, and experiences of academic researchers with data sharing and reuse: A meta-synthesis. *PLoS One*, *15*(2), Article e0229182. https://doi.org/10.1371/journal.pone.0229182

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), 86–89. https://doi.org/10.7191/jeslib.2014.1064

Petters, J. L., Brooks, G. C., Smith, J. A., & Haas, C. A. (2019). The impact of targeted data management training for field research projects - A case study. *Data Science Journal, 18*(1), Article 43. https://doi.org/10.5334/dsj-2019-043

Piorun, M. E., Kafel, D., Leger-Hornby, T., Najafi, S., Martin, E.R., Colombo, P., & LaPelle, N. R. (2012). Teaching research data management: an undergraduate/graduate curriculum. *Journal of eScience Librarianship*, *1*(1), 46–50. https://doi.org/10.7191/jeslib.2012.1003

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. https://doi.org/10.1080/19386389.2010.506379

Rantasaari, J. (2021). Doctoral students' educational needs in research data management: Perceived importance and current competencies. *International Journal of Digital Curation*, *16*(1), 1–36. https://doi.org/10.2218/IJDC.V16I1.684

Rantasaari, J., & Kokkinen, H. (2019, June 23-27). *Closing the skills gap: The basics of the research data management (BRDM) course: Case University of Turku*. [Conference presentation]. The fortieth IATUL conference, Perth, Australia. https://docs.lib.purdue.edu/iatul/2019/fair/5/

Rantasaari, J., Löyttyniemi, E., Cooper, H., Fredriksson, M., Henriksson, B., Huttunen, S., … Wilson, L. (2021). *Basics of the Research Data Management (BRDM) Course: Course Structure and Learning Objectives 2019-22* [Presentation]. Zenodo. https://doi.org/10.5281/zenodo.3692224

Read, K. B. (2019). Adapting data management education to support clinical research projects in an academic medical center. *Journal of the Medical Library Association*, *107*(1), 89–97. https://doi.org/10.5195/jmla.2019.580

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), Article e0215509. https://doi.org/10.1371/journal.pone.0215509

Research data service. (n.d.). *MANTRA: Research data management training*. University of Edinburgh. https://mantra.ed.ac.uk/

Revez, J. (2018). Opening the heart of science: A review of the changing roles of research libraries. *Publications*, *6*(1), 1–13. https://doi.org/10.3390/publications6010009

Rieser, A. (2018, August 21). How to handle big data, before it handles you. *Industry Week*. https://www.industryweek.com/technology-and-iiot/how-handle-big-data-it-handles-you

Schmidt, L., & Holles, J. (2018). A graduate class in research data management. *Chemical Engineering Education*, *52*(1), 52–59. https://journals.flvc.org/cee/article/view/105451

Scholtens, S., Anbeek, P., Böhmer, J., Brullemans-Spansier, M., van der Geest, M., Jetten, M., Staiger, C., Slouwerhof, I., & van Gelder, C. W. G. van. (2019). *Life sciences data steward function matrix*. [Project deliverable]. Zenodo. https://doi.org/10.5281/ZENODO.2561723

Schöpfel, J., & Prost, H. (2016). Research data management in social sciences and humanities: A survey at the University of Lille (France) - LIBREAS. Library Ideas. *Libreas: Library Ideas*, *29*, 98–112. https://libreas.eu/ausgabe29/09schoepfel/

Shadbolt, A., Konstantelos, L., Lyon, L., & Guy, M. (2014). Delivering innovative RDM training: The immersive informatics pilot programme. *International Journal of Digital Curation*, *9*(1), 313–323. https://doi.org/10.2218/ijdc.v9i1.318

Shearer, K. (2009). *Research data: Unseen opportunities*. CARL ABRC. https://portagenetwork.ca/wp-content/uploads/2016/04/Data_mgt_toolkit_EN.pdf

Southall, J., & Scutt, C. (2017). Training for research data management at the Bodleian Libraries: National contexts and local implementation for researchers and librarians. *New Review of Academic Librarianship*, *23*(2–3), 303–322. https://doi.org/10.1080/13614533.2017.1318766

Surkis, A., LaPolla, F. W. Z., Contaxis, N., & Read, K. B. (2017). Data day to day: Building a community of expertise to address data skills gaps in an academic medical center. *Journal of the Medical Library Association*, *105*(2), 185–191. https://doi.org/10.5195/jmla.2017.35

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*. https://doi.org/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), 1–15. https://doi.org/10.1080/01639269.2017.1387739

Timonen, V., Foley, G., & Conlon, C. (2018). Challenges when using grounded theory: A pragmatic introduction to doing GT research. *International Journal of Qualitative, 17*(1), 1–10. https://doi.org/10.1177/1609406918758086

UNIFI. (2016). *Open science and data: Action programme for the Finnish scholarly community*. https://www.unifi.fi/wp-content/uploads/2019/04/UNIFI_Open_Science_and_Data_Action_Programme.pdf

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. https://doi.org/10.1016/j.acalib.2014.02.008

Wang, M., & Fong, B. L. (2015). Embedded data librarianship: A case study of providing data management support for a science department. *Science & Technology Libraries*, *34*(3), 228–240. https://doi.org/10.1080/0194262X.2015.1085348

Wellcome. (2017, July 10). *Data, software and materials management and sharing policy*. https://wellcome.ac.uk/funding/guidance/data-software-materials-management-and-sharing-policy

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. https://doi.org/10.1108/LHT-02-2014-0021

Whitmire, A. L. (2015). Implementing a graduate-level research data management course: Approach, outcomes, and lessons learned. *Journal of Librarianship and Scholarly Communication*, *3*(2), 1–22. https://doi.org/10.7710/2162-3309.1246

Wiley, C. A., & Kerby, E. E. (2018). Managing research data: Graduate student and postdoctoral researcher perspectives. *Issues in Science and Technology Librarianship*, *89*. https://doi.org/10.5062/F4FN14FJ

Wiley, C. A., Mischo, W. H., Schlembach, M. C., & Imker, H. J. (2017, June 24–28). An integrated data management instructional program. ASEE Annual Conference and Exposition, Columbus, USA https://doi.org/10.18260/1-2--27572

Wiljes, C., & Cimiano, P. (2019). Teaching research data management for students. *Data Science Journal*, *18*(1), 1–9. https://doi.org/10.5334/dsj-2019-038

Wittenberg, J., & Elings, M. (2017). Building a research data management service at the university of California, Berkeley: A tale of collaboration. *IFLA Journal*, *43*(1), 89–97. https://doi.org/10.1177/0340035216686982

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). Purdue University Press. https://www.jstor.org/stable/j.ctt6wq2vh.9

Yu, H. H. (2017). The role of academic libraries in research data service (RDS) provision: Opportunities and challenges. *The Electronic Library*, *35*(4), 783–797. https://doi.org/10.1108/EL-10-2016-0233

## Notes

[1] https://www.go-fair.org/fair-principles/.

[2] Systematic treatment of data, involving operations to get easier to find, understand and use data in the present and future projects (Briney, 2015, p. 7).

[3] Specified learning objectives of BRDM and its' modules (Rantasaari et al., 2021).

[4] https://www.nia.nih.gov/health/what-are-clinical-trials-and-studies.

[5] https://www.amnh.org/explore/videos/the-scientific-process.

[6] https://www.dmptuuli.fi/template_export/476471047.pdf.

# Appendix A: Tables of Results

*Table 1: Topics of the trainings in 2010–2021.*

| Source[a] | Planning and organisation | Sharing and reuse | Storage, backup, security | Metadata and data description | Preservation | Legal and ethical issues | Quality and documentation | Types and Formats | Life cycle | Discovery | Policies, requirements, incentives | Processing | Cultures of Practice |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Adamick et al., 2013 | | 1 | 1 | 1 | 1 | 1 | | | | | | | |
| ARDS, 2018[b] | 1 | 1 | | | 1 | 1 | | | 1 | 1 | 1 | | |
| Borycz & Carroll, 2021 | 1 | 1 | | 1 | 1 | | 1 | | 1 | | | | |
| Carpentries, n.d. | 1 | | | | | | | | | | | 1 | |
| Castle, 2019 | 1 | 1 | 1 | | | | | | | | | | |
| CESSDA, 2017[c] | 1 | 1 | 1 | | 1 | 1 | 1 | | | 1 | | 1 | |
| Clement et al., 2017 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | | | 1 |
| Cole & Evans, 2014 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | |
| Corti & Van den Eynden, 2015 | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | | | | | |
| EDINA[d] | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | 1 | 1 | |
| Kafel et al., 2014 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | | | 1 |
| Mustajoki, n.d. | 1 | 1 | 1 | 1 | 1 | 1 | | | | 1 | 1 | | |
| Pascuzzi & Nelson, 2018 | 1 | 1 | 1 | | | 1 | 1 | | 1 | 1 | | | |
| Petters et al., 2019 | 1 | | | 1 | | | 1 | | 1 | | | 1 | |
| Piorun et al., 2012 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | | | 1 |
| Qin & D'ignazio, 2010 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Rantasaari & Kokkinen, 2019 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | |
| Read, 2019 | | 1 | 1 | 1 | | | 1 | 1 | 1 | | 1 | 1 | 1 |
| Read et al., 2019 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | | 1 | | 1 |
| Schmidt & Holles, 2018 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | | |
| Schöpfel & Prost, 2016 | 1 | | 1 | 1 | 1 | | | 1 | 1 | | | | |
| Shadbolt et al., 2014 | | 1 | 1 | | 1 | 1 | | | 1 | | 1 | | |
| Southall & Scutt, 2017 | 1 | | | | 1 | | | | | 1 | 1 | | |
| Surkis et al., 2017 | 1 | | | | | | | | | | | 1 | |
| Thielen et al., 2017 | 1 | 1 | | 1 | 1 | 1 | 1 | | | 1 | | | 1 |
| Wang & Fong, 2015 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | | | 1 | | |
| Whitmire, 2015 | 1 | 1 | | 1 | | | | 1 | | 1 | | | |
| Wiley et al., 2017 | 1 | 1 | 1 | 1 | 1 | | 1 | | | | 1 | | |
| Wiljes & Cimiano, 2019 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | 1 | | | |
| Wright & Andrews, 2015 | 1 | 1 | | | | | 1 | 1 | | | 1 | 1 | |
| SUM | 27 | 25 | 21 | 21 | 21 | 17 | 17 | 14 | 14 | 13 | 12 | 8 | 7 |

*Table 1: Continued*

| Source | Visualisation and representation | Conversion and interoperability | Collection planning | Analysis | Budgeting | Cleaning | Databases | Encoding | Planning research project | Big data | Cloud computing | Planning curation profile |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Adamick et al., 2013 | | | | | | | | | | | | |
| ARDS, 2018 | | | | | | | | | | | | |
| Borycz & Carroll, 2021 | | | | | | | | | | | | |
| Carpentries, n.d. | 1 | | | 1 | | 1 | | 1 | | | 1 | |
| Castle, 2019 | | | | | | | | | | | | |
| CESSDA, 2017 | | | | | | | | | | | | |
| Clement et al., 2017 | | | | | 1 | | | | | | | |
| Cole & Evans, 2014 | | | | | | | | | | | | |
| Corti & Van den Eynden, 2015 | | 1 | | | | | | | | | | |
| EDINA | 1 | 1 | | 1 | | | | | | | | |
| Kafel et al., 2014 | | | | | 1 | | | | | | | |
| Mustajoki, n.d. | | | | | | | | | | | | |
| Pascuzzi & Nelson, 2018 | 1 | | 1 | | | | | | | 1 | | |
| Petters et al., 2019 | | 1 | | | | 1 | | | | | | |
| Piorun et al., 2012 | | | | | 1 | | | | | | | |
| Qin & D'ignazio, 2010 | 1 | 1 | 1 | | | 1 | 1 | 1 | | | | |
| Rantasaari & Kokkinen, 2019 | | | 1 | | | | | | 1 | | | |
| Read, 2019 | | | | | | | | | | | | |
| Read et al., 2019 | | | 1 | | | | | | | | | |
| Schmidt & Holles, 2018 | | | | | | | | | 1 | | | 1 |
| Schöpfel & Prost, 2016 | | | | | | | | | | | | |
| Shadbolt et al., 2014 | | | | | | | | | | | | |
| Southall & Scutt, 2017 | | | | | | | | | | | | |
| Surkis et al., 2017 | 1 | | | 1 | | | | | | 1 | | |
| Thielen et al., 2017 | 1 | | | | | | | | | | | |
| Wang & Fong, 2015 | | | | | | | | | | | | |
| Whitmire, 2015 | | | | | | | | | | | | |
| Wiley et al., 2017 | | | | | | | | | | | | |
| Wiljes & Cimiano, 2019 | | | | | | | | | | | | |
| Wright & Andrews, 2015 | 1 | | | 1 | | | 1 | | | | | |
| SUM | 7 | 4 | 4 | 4 | 3 | 3 | 2 | 2 | 2 | 2 | 1 | 1 |

[a]For details of source, see references of main text.
[b]Australian Research Data Commons: https://ardc.edu.au/.
[c]https://www.cessda.eu/DMGuide.
[d]https://mantra.ed.ac.uk/.

*Table 2: Participants by their disciplines 2019–2021.*

| Year | 2019 | 2020 | 2021 | Sum |
|---|---|---|---|---|
| Law | 1 | 0 | 1 | 2 |
| Education, Welfare | 3 | 8 | 8 | 19 |
| Humanities, Psychology, Theology | 4 | 16 | 12 | 32 |
| Social Sciences, Business, Economics | 6 | 22 | 38 | 66 |
| Science and Engineering | 5 | 28 | 30 | 63 |
| Health Sciences | 23 | 22 | 32 | 77 |
| Sum | 42 | 96 | 121 | 259 |

*Table 3: What are the three things you have learned?*

| Category | 2019 | 2020 | 2021 |
|---|---|---|---|
| What, why and when in RDM | 100 | 112 | 55 |
| Importance of legal considerations | 64 | 19 | 9 |
| Making a sound research plan | 38 | 26 | 10 |
| Securing data privacy | 29 | 17 | 24 |
| Using data collecting or organizing software | 17 | 7 | 10 |
| Other comments | 4 | 13 | 11 |
| Sum | 252 | 194 | 119 |

*Table 4: How will the things you have learned change your practices?*

| Category | 2019 | 2020 | 2021 |
|---|---|---|---|
| I will pay notice to IPR, agreements and licenses | 25 | 9 | 5 |
| I will improve data management planning and documenting | 18 | 73 | 31 |
| I will collect, produce or process data with REDCap or Nvivo | 17 | 6 | 3 |
| I will pay more notice to data privacy and security | 13 | 22 | 16 |
| I will improve my research plan | 13 | 17 | 5 |
| Other comments | 1 | 5 | 3 |
| Sum | 87 | 132 | 63 |

*Table 5: How would you suggest the module be developed?*

| Category | 2019 | 2020 | 2021 |
|---|---|---|---|
| Increase practicality, e.g., good and bad examples and check lists | 34 | 30 | 29 |
| Clarifying and standardizing procedures, practices, and course platform | 23 | 53 | 42 |
| Increase discussions and interactivity | 12 | 22 | 4 |
| Possibility to prepare one's own study plan and DMP | 7 | 0 | 0 |
| Differentiating the course contents according to discipline, data type, methods | 6 | 9 | 10 |
| Turning to hybrid or contact course | 0 | 12 | 0 |
| Good as it is | 0 | 0 | 12 |
| Other comments | 8 | 10 | 4 |
| Sum | 90 | 136 | 101 |

*Table 6: Competencies before and after BRDM 2019 (medians, custom quantiles, and p-values).*

| Competence | Median, before | Q1; Q3 | Median, after | Q1; Q3 | p-value (Fit Y by X; Wilcoxon rank-sum test) |
|---|---|---|---|---|---|
| Discovery and acquisition of data | 1.97 | 1.78; 2.14 | 2.39 | 2.12; 2.84 | 0.02 |
| Databases and data formats | 2.02 | 1.82; 2.22 | 2.38 | 2.07; 2.89 | 0.04 |
| Data conversion and interoperability | 1.83 | 1.17; 1.98 | 2.08 | 1.81; 2.65 | 0.07 |
| Data management and organization | 1.95 | 1.76; 2.14 | 2.63 | 2.16; 3 | 0.001 |
| Data quality and documentation | 2.01 | 1.98; 2.06 | 2.62 | 2.11; 2.94 | 0.02 |
| Metadata and data description | 1.91 | 1.76; 2 | 2.72 | 2.21; 2.91 | <0.001 |
| Cultures of practice | 1.96 | 1.81; 2.08 | 2.22 | 1.86; 3 | 0.07 |
| Ethics and attribution | 2.11 | 2.03; 2.69 | 2.89 | 2.37; 3.10 | 0.01 |
| Data curation and reuse | 1.89 | 1.31; 1.97 | 2.15 | 2.04; 2.68 | 0.001 |
| Data preservation | 1.93 | 1.80; 2 | 2.62 | 2.11; 2.94 | 0.001 |
| Median, custom quantiles, p-value | 1.96 | 1.82; 2.09 | 2.32 | 2.12; 2.84 | 0.003 |

*Table 7: Competencies before and after BRDM 2020–2021 (medians, custom quantiles, and p-values).*

| Knowledge, skill, or ability | Median, before | Q1; Q3 | Median, after | Q1; Q3 | p-value (Distributions; Wilcoxon signed-rank test) |
|---|---|---|---|---|---|
| Describing your research and data collection process to identify your data lifecycle | 2.07 | 2.03; 2.14 | 3.08 | 3.05; 3.13 | <0.0001 |
| Recognizing the necessary components of a data management plan | 1.95 | 1.89; 1.99 | 3.11 | 3.07; 3.82 | <0.0001 |
| Creating a data management plan to manage and curate your own data | 1.83 | 1.12; 1.93 | 3.06 | 3.03; 3.10 | <0.0001 |
| Documenting your data for yourself and others | 2.06 | 2.02; 2.13 | 3.07 | 3.04; 3.12 | <0.0001 |
| Applying the relevant laws, agreements, permits, and licenses to your data | 1.87 | 1.14; 1.93 | 2.93 | 2.88; 2.97 | <0.0001 |
| Applying the basic anonymization methods for qualitative and quantitative research when needed | 2.05 | 2; 2.13 | 3.01 | 2.97; 3.06 | <0.0001 |
| Recognizing the importance of data protection for collecting, processing, storage and sharing of data | 2.08 | 2.03; 2.20 | 3.11 | 3.06; 3.83 | <0.0001 |
| Creating a data privacy statement and a risk analysis when needed | 1.9 | 1.16; 1.97 | 2.94 | 2.89; 2.98 | <0.0001 |
| Creating a database or a survey for capturing and maintaining your data using REDCap software | 1.13 | 1.06; 1.23 | 2.99 | 2.83; 3.14 | <0.0001 |
| Organizing and coding your qualitative data for analyzing using NVivo software | 1.59 | 1.29; 2.23 | 3.01 | 2.88; 3.16 | <0.0001 |
| Creating a storage and backup plan, and applying it to your data using the services of your organization, or the services of The IT Center for Science (CSC) | 1.96 | 1.82; 2.09 | 3 | 2.88; 3.12 | <0.0001 |
| Evaluating data repositories for depositing and publishing your data and discovering other researchers' data for re-use | 1.18 | 1.1; 1.92 | 2.97 | 2.93; 3.01 | <0.0001 |
| Applying FAIR principles to your data when possible | 1.9 | 1.15; 1.97 | 2.98 | 2.86; 3.08 | <0.0001 |
| Applying data management best practices concerning collecting, organizing, documenting, storing, long-term preserving, and sharing (when possible) to your own data | 2.01 | 1.91; 2.12 | 3.03 | 2.98; 3.09 | <0.0001 |
| Median, custom quantiles, p-value | 1.97 | 1.93; 2.01 | 3.03 | 2.98; 3.08 | <0.0001 |

# Appendix B

## Give a Session Specific Feedback for the Course

We kindly ask you to give a brief feedback of a module you have completed.

1. Study Programme

◯ Health Sciences

◯ Survey Research

◯ Natural Sciences

◯ Qualitative Research

2. Choose the module for what you are giving feedback:

◯ Introductory lecture

◯ Module 1: Research plan

◯ Module 2: Data management plan

◯ Module 3. IPR issues, permits and licences

◯ Module 4: Privacy notice and risk analysis

◯ Module 5: Building database with REDCap

◯ Module 5: (in qualitative research): Learn the basics of NVivo software

◯ Module 6: Data storage, protection, processing, describing, and IT service solutions

◯ Module 7: Data preservation, sharing and citing. General and discipline specific open data repositories

3. What are the three things you have learned?

4. How will the things you have learned change your practices?

5. How would you suggest the module be developed?

Send

# Appendix C

## Research Data Management - BRDM 2019 Course

Dear Participant,

To obtain information of participants' perceptions of the importance of research data management in the differenct phases of data lifecycle, and to obtain information of their current skills and competencies, we kindly ask you to answer the questions below.

We will use the participants' anonyme answers in teaching and developing the course, and as a part of research project on RDM. The anonyme data will be preserved in Finnish Social Science Data Archive for research, teaching and learning purposes.

Thank you for you participation in the survey!

**1. Faculty**

- ◯ Humanities
- ◯ Education
- ◯ Science and Engineering
- ◯ Medicine
- ◯ Law
- ◯ School of Economics
- ◯ Social Sciences

**2. Study Programme**

- ◯ Health Sciences
- ◯ Survey
- ◯ Natural Sciences

Please indicate how important you believe it is for you to be knowledgeable in each of the competencies listed below by the time you graduate. Please also tell how well do you think you´ll manage the competence now.

**Discovery and Acquisition of Data**

Skills may include:
Locates and utilizes disciplinary data repositories or other external data sources. Evaluates the quality of the data available from external sources. Not only identifies appropriate external data sources, but also imports data and converts it when necessary, so it can be used locally.

**3. Discovery and Acquisition of Data: Importance**

- ◯ Not important
- ◯ Somewhat Important
- ◯ Important
- ◯ Very Important
- ◯ Essential
- ◯ I don't know or NA

**4. Discovery and Acquisition of Data: Competence**

- ◯ Don't have competence
- ◯ Somewhat competence
- ◯ Good competence
- ◯ Very good competence
- ◯ Ultimate competence

**Databases and Data Formats**

Skills may include:
Understands the concept of relational databases, how to query those databases, and becomes familiar

with standard data formats and types for their discipline. Understands which formats and data types are appropriate for different research questions.

**5. Databases and Data Formats: Importance**

- ◯ Not important
- ◯ Somewhat Important
- ◯ Important
- ◯ Very Important
- ◯ Essential
- ◯ I don¨t know or NA

**6. Databases and Data Formats: Competence**

- ◯ Don't have
- ◯ Somewhat
- ◯ Good
- ◯ Very good
- ◯ Ultimate

**Data Conversion and Interoperability**

Skills may include:
Is proficient in migrating data from one format to another. Understands the risks and potential loss or corruption of information caused by changing data formats. Understands the benefits of making data available in standard formats to facilitate downstream use.

**7. Data Conversion and Interoperability: Importance**

- ◯ Not important
- ◯ Somewhat Important
- ◯ Important

- ◯ Very Important
- ◯ Essential
- ◯ I don’t know or NA

**8. Data Conversion and Interoperability: Competence**

- ◯ Don't have
- ◯ Somewhat
- ◯ Good
- ◯ Very good
- ◯ Ultimate

**Data Management and Organization**

Skills may include:
Understands the lifecycle of data, develops data management plans, and keeps track of the relation of subsets or processed data to the original data sets. Creates standard operating procedures for data management and documentation.

**9. Data Management and Organization: Importance**

- ◯ Not Important
- ◯ Somewhat Important
- ◯ Important
- ◯ Very Important
- ◯ Essential
- ◯ I don’t know or NA

**10. Data Management and Organization: Competence**

- ◯ Don't have

- ◯ Somewhat
- ◯ Good
- ◯ Very good
- ◯ Ultimate

**Data Quality and Documentation**

Skills may include:
Recognizes, documents, and resolves any apparent artifacts, incompletion, or corruption of data sets. Utilizes metadata to facilitate an understanding of potential problems with data sets. Documents data sufficiently enough to enable the reproduction of the research results and the data by others. Tracks data provenance and clearly delineates and denotes versions of a data set.

**11. Data Quality and Documentation: Importance**

- ◯ Not Important
- ◯ Somewhat Important
- ◯ Important
- ◯ Very Important
- ◯ Essential
- ◯ I don’t know or NA

**12. Data Quality and Documentation: Competence**

- ◯ Don't have
- ◯ Somewhat
- ◯ Good
- ◯ Very good
- ◯ Ultimate

**Metadata and Data Description**

Skills may include:
Understands the rationale for metadata and proficiently annotates and describes data so it can be understood and used by self and others. Develops the ability to read and interpret metadata from external disciplinary sources. Understands the structure and purpose of ontologies in facilitating better sharing of data.

**13. Metadata and Data Description: Importance**

- ◯ Not Important
- ◯ Somewhat Important
- ◯ Important
- ◯ Very Important
- ◯ Essential
- ◯ I don't know or NA

**14. Metadata and Data Description: Competence**

- ◯ Don't have
- ◯ Somewhat
- ◯ Good
- ◯ Very good
- ◯ Ultimate

**Cultures of Practice**

Skills may include:
Recognizes the practices, values, and norms of his/her chosen field, discipline, or subdiscipline as they relate to managing, sharing, curating, and preserving data. Recognizes relevant data standards of his/her field (metadata, quality, formatting, etc.) and understands how these standards are applied.

**15. Cultures of Practice: Importance**

- ◯ Not Important

- ◯ Somewhat Important
- ◯ Important
- ◯ Very Important
- ◯ Essential
- ◯ I don't know or NA

**16. Cultures of Practice: Competence**

- ◯ Don't have
- ◯ Somewhat
- ◯ Good
- ◯ Very good
- ◯ Ultimate

**Ethics and Attribution**

Skills may include:
Develops an understanding of intellectual property, privacy and confidentiality issues, and the ethos of the discipline when it comes to sharing and administering data. Acknowledges data from external sources appropriately. Avoids misleading or ambiguous representations when presenting data.

**17. Ethics and Attribution: Importance**

- ◯ Not Important
- ◯ Somewhat Important
- ◯ Important
- ◯ Very Important
- ◯ Essential
- ◯ I don't know or NA

**18. Ethics and Attribution: Competence**

- ◯ Don't have
- ◯ Somewhat
- ◯ Good
- ◯ Very good
- ◯ Ultimate

**Data Curation and Re-use**

Skills may include:
Recognizes that data may have value beyond the original purpose, to validate research, or for use by others. Is able to distinguish which elements of a data set are likely to have future value for self and for others. Understands that curating data is a complex, often costly endeavor that is nonetheless vital to community-driven e-research. Recognizes that data must be prepared for its eventual curation at its creation and throughout its lifecycle. Articulates the planning and activities needed to enable data curation, both generally and within his/her local practice. Understands how to cite data as well as how to make his/her data citable.

**19. Data Curation and Re-use: Importance**

- ◯ Not Important
- ◯ Somewhat Important
- ◯ Important
- ◯ Very Important
- ◯ Essential
- ◯ I don't know or NA

**20. Data Curation and Re-use: Competence**

- ◯ Don't have
- ◯ Somewhat
- ◯ Good

- ◯ Very good
- ◯ Ultimate

**Data Preservation**

Skills may include:
Recognizes the benefits and costs of data preservation. Understands the technology, resources, and organizational components of preserving data. Utilizes best practices in preparing data for its eventual preservation during its active lifecycle. Articulates the potential long term value of his/her data for him/herself or others and is able to determine an appropriate preservation timeframe. Understands the need to develop preservation policies and is able to identify the core elements of such policies.

**21. Data Preservation: Importance**

- ◯ Not Important
- ◯ Somewhat Important
- ◯ Important
- ◯ Very Important
- ◯ Essential
- ◯ I don’t know or NA

**22. Data Preservation: Competence**

- ◯ Don't have
- ◯ Somewhat
- ◯ Good
- ◯ Very good
- ◯ Ultimate

---

# Appendix D

*Confidential*



## BRDM Participant Survey 2021

Dear BRDM 2021 Course participant,

Please give an anonymous rating of your research data management (RDM) knowledge, skills or abilities a) before taking the BRDM Course, and b) after taking the BRDM Course.

Please answer, even if you enrolled but did not start the course. In that case you are only asked to answer two questions.

We will use the results of this survey to develop the course and its contents. The survey is also part of my dissertation project at Åbo Akademi University.

Anonymous data of the survey will be preserved openly accessible in the Finnish Social Science Data Archive (https://www.fsd.tuni.fi/en/) and in the Zenodo Archive by CERN (https://zenodo.org/).

At the end of the survey you can also give feedback for the Course.

Answering takes about 10 to 15 minutes.

Best regards,

Jukka Rantasaari

Choose your role

◯ PhD student
◯ Postdoc researcher
◯ Other

**Choose your organization**

University

◯ University of Turku (UTU)
◯ Åbo Akademi University (ÅAU)

**Choose your faculty**

Faculty

◯ Faculty of Education / UTU
◯ Faculty of Humanities / UTU
◯ Faculty of Law / UTU
◯ Faculty of Medicine / UTU
◯ Faculty of Science and Engineering / UTU
◯ Faculty of Social Sciences / UTU
◯ Turku School of Economics / UTU
◯ Other unit

Faculty

◯ Faculty of Arts, Psychology and Theology / ÅAU
◯ Faculty of Education and Welfare Studies / ÅAU
◯ Faculty of Science and Engineering / ÅAU
◯ Faculty of Social Sciences, Business and Economics / ÅAU
◯ Other unit

*Confidential*



**What is your study programme in the Basics of Research Data Management (BRDM) Course?**

Study programme

◯ Health Sciences
◯ Qualitative Research
◯ Natural Sciences
◯ Survey Research

Modules you participated in? Check all the modules you participated. If you enrolled, but did not participate in any modules, choose "No participation"

☐ Module 0: Introductory lecture
☐ Module 1: Research plan
☐ Module 2: Data management plan
☐ Module 3: IPR, permits and licences
☐ Module 4: Privacy notice and risk analysis
☐ Module 5: REDCap OR NVivo
☐ Module 6: Data storage, protection, processing, describing and IT service solutions
☐ Module 7: Data preservation, sharing and citing
☐ No participation

If you started the course, but then interrupted, would you please tell us the reasons:

◯ The BRDM course did not meet my expectations
◯ I was busy elsewhere
◯ Other reasons

Would you like to specify the reasons for interruption? What would have make you to continue?

**How would you rate your RDM knowledge, skills or ability before and after taking the BRDM Course?**

**1. Describing your research and data collection process in order to identify your data lifecycle.**

| | No competence | Little competence | Somewhat competent | Very competent | N/A (not applicable) |
|---|---|---|---|---|---|
| Knowledge, skills or ability, before | ◯ | ◯ | ◯ | ◯ | ◯ |
| Knowledge, skills or ability, after | ◯ | ◯ | ◯ | ◯ | ◯ |

**2. Recognizing the necessary components of a data management plan.**

| | No competence | Little competence | Somewhat competent | Very competent | N/A (not applicable) |
|---|---|---|---|---|---|
| Knowledge, skills or ability before | ◯ | ◯ | ◯ | ◯ | ◯ |
| Knowledge, skills or ability, after | ◯ | ◯ | ◯ | ◯ | ◯ |

*Confidential*



| **3. Creating a data management plan in order to manage and curate your own data.** | | | | | |
|---|---|---|---|---|---|
| | No competency | Little competency | Somewhat competent | Very competent | N/A (not applicable) |
| Knowledge, skills or ability, before | ○ | ○ | ○ | ○ | ○ |
| Knowledge, skills or ability, after | ○ | ○ | ○ | ○ | ○ |

| **4. Documenting your data for yourself and others.** | | | | | |
|---|---|---|---|---|---|
| | No competence | Little competence | Somewhat competent | Very competent | N/A (not applicable) |
| Knowledge, skills or ability, before | ○ | ○ | ○ | ○ | ○ |
| Knowledge, skills or ability, after | ○ | ○ | ○ | ○ | ○ |

| **5. Applying the relevant laws, agreements, permits, and licenses to your data.** | | | | | |
|---|---|---|---|---|---|
| | No competence | Little competence | Somewhat competent | Very competent | N/A (not applicable) |
| Knowledge, skills or ability, before | ○ | ○ | ○ | ○ | ○ |
| Knowledge, skills or ability, after | ○ | ○ | ○ | ○ | ○ |

| **6. Applying the basic anonymization methods for qualitative and quantitative research when needed.** | | | | | |
|---|---|---|---|---|---|
| | No competence | Little competence | Somewhat competent | Very competent | N/A (not applicable) |
| Knowledge, skills or ability, before | ○ | ○ | ○ | ○ | ○ |
| Knowledge, skills or ability, after | ○ | ○ | ○ | ○ | ○ |

| **7. Recognizing the importance of data protection for collecting, processing, storage and sharing of data.** | | | | | |
|---|---|---|---|---|---|
| | No competence | Little competence | Somewhat competent | Very competent | N/A (not applicable) |
| Knowledge, skills or ability, before | ○ | ○ | ○ | ○ | ○ |
| Knowledge, skills or ability, after | ○ | ○ | ○ | ○ | ○ |

| **8. Creating a data privacy statement and a risk analysis when needed.** | | | | | |
|---|---|---|---|---|---|
| | No competence | Little competence | Somewhat competent | Very competent | N/A (not applicable) |
| Knowledge, skills or ability, before | ○ | ○ | ○ | ○ | ○ |
| Knowledge, skills or ability, after | ○ | ○ | ○ | ○ | ○ |

| **9a. Creating a database or a survey for capturing and maintaining your data using REDCap software.** | | | | | |
|---|---|---|---|---|---|
| | No competence | Little competence | Somewhat competent | Very competent | N/A (not applicable) |
| Knowledge, skills or ability, before | ○ | ○ | ○ | ○ | ○ |
| Knowledge, skills or ability, after | ○ | ○ | ○ | ○ | ○ |

*Confidential*



| **9b. Organizing and coding your qualitative data for analyzing using NVivo software.** | | | | | |
|---|---|---|---|---|---|
| | No competence | Little competence | Somewhat competent | Very competent | N/A (not applicable) |
| Knowledge, skills or ability, before | ○ | ○ | ○ | ○ | ○ |
| Knowledge, skills or ability, after | ○ | ○ | ○ | ○ | ○ |

| **10. Creating a storage and backup plan, and applying it to your data using the services of your organization, or the services of The IT Center for Science (CSC).** | | | | | |
|---|---|---|---|---|---|
| | No competence | Little competence | Somewhat competent | Very competent | N/A (not applicable) |
| Knowledge, skills or ability, before | ○ | ○ | ○ | ○ | ○ |
| Knowledge, skills or ability, after | ○ | ○ | ○ | ○ | ○ |

| **11. Evaluating data repositories for depositing and publishing your data, and discovering other researchers' data for re-use.** | | | | | |
|---|---|---|---|---|---|
| | No competence | Little competence | Somewhat competent | Very competent | N/A (not applicable) |
| Knowledge, skills or ability, before | ○ | ○ | ○ | ○ | ○ |
| Knowledge, skills or ability, after | ○ | ○ | ○ | ○ | ○ |

| **12. Applying FAIR principles to your data when possible.** | | | | | |
|---|---|---|---|---|---|
| | No competence | Little competence | Somewhat competent | Very competent | N/A (not applicable) |
| Knowledge, skills or ability, before | ○ | ○ | ○ | ○ | ○ |
| Knowledge, skills or ability, after | ○ | ○ | ○ | ○ | ○ |

| **13. Applying data management best practices concerning collecting, organizing, documenting, storing, long-term preserving and sharing (when possible) to your own data.** | | | | | |
|---|---|---|---|---|---|
| | No competence | Little competence | Somewhat competent | Very competent | N/A (not applicable) |
| Knowledge, skills or abilities, before | ○ | ○ | ○ | ○ | ○ |
| Knowledge, skills or abilities, before | ○ | ○ | ○ | ○ | ○ |

**Learning needs**

*Confidential*



Which of these topics would you like to learn more about?

- ☐ Discovery and acquisition of data
- ☐ Databases and data formats
- ☐ Data conversion and interoperability
- ☐ Data backup, version management, and storage
- ☐ Data management planning and data organization
- ☐ Data quality and documentation
- ☐ Data processing and analysis (not covered on BRDM 2020 Course)
- ☐ Data visualization and representation (not covered on BRDM 2020 Course)
- ☐ Metadata and data description
- ☐ Discipline specific cultures of data practices
- ☐ Ethics and legal considerations (IPR, copyright, licenses, permits, data security)
- ☐ Data curation and re-use
- ☐ Data long-term preservation, sharing and publishing

Something else you would like to learn more about?

Course rating

We are grateful that you chose this course and have stayed with us until this final module.

Please give us your rating. Poor — Excellent

*(Place a mark on the scale above)*

Would you recommend this course to other PhD Students or PostDoc Researchers?

- ◯ Yes
- ◯ No

You can also give us final verbal feedback for the whole course.

What went well? How could we improve?

**Did you enroll the course, but did not participate?**

If you enrolled, but did not started the course, would you please tell us the reasons:

- ◯ I was busy elsewhere
- ◯ Other reasons (you can specify below)

If you did not start the course, would you like to specify the reasons? What would have make you to start?