



RDM 101 Impact Evaluation

**A Study on Training Effectiveness
and Research Practice Changes**

Authors: Narmin Rzayeva, Francesca Morselli,
Carla Strubbia, Nikki Grens,
Gargi Kulkarni, Paula Martinez Lavanchy

Date: 01.12.25

RDM 101 Impact Evaluation

**A Study on Training Effectiveness
and Research Practice Changes**

Contents

1	Executive Summary	6
1.1	Overview	6
1.2	Objectives	6
1.3	Methodology	6
1.4	Key Findings	6
1.4.1	Short-Term Impact	6
1.4.2	Long-Term Impact	7
1.5	Areas for Improvement	7
1.6	Recommendations	7
2	Introduction	8
2.1	About RDM 101	8
2.2	Study Objectives	8
3	Data Source & Methodology	9
4	Short-term impact of RDM 101	10
4.1	Graduate School Feedback Forms	10
4.1.1	Objectives of the Analysis	10
4.1.2	Graduate school Feedback Forms Overview & Structure	10
4.1.3	Methodology & Software	11
4.1.4	Results	12
4.2	Pre- and Post-training Survey	23
4.2.1	Objectives of the Analysis	23
4.2.2	Survey Overview & Structure	23
4.2.3	Methodology & Software	24
4.2.4	Results	25
5	Long-term impact of RDM 101	38
5.1	Objectives of the Analysis	38
5.2	Interview Overview & Structure	38
5.3	Methodology & Software	39
5.4	Results	40
6	Main Highlights of the Study	45
6.1	Short-Term Impact Summary	45
6.1.1	Course Perception and Engagement	45
6.1.2	Feedback-Driven Improvements	45
6.1.3	Training Effectiveness and Shifts in RDM Practices	45
6.1.4	Online vs In-Person Delivery	46
6.2	Long-Term Impact Summary	47
6.2.1	Motivations and Expectations	47
6.2.2	Knowledge Acquisition and Impact on Practice	47
6.2.3	Course Design and Delivery	47

6.2.4 Emotional and Social Dimensions	47
7 Conclusion	48
Data and Code Availability	49
References	50
Acknowledgments	51
Author Contributions	52

1 Executive Summary

1.1 Overview

The RDM 101 course (Martinez-Lavanchy et al., 2024) at TU Delft is a foundational course designed to equip PhD candidates with essential research data management (RDM) skills. This impact study evaluates the course's effectiveness in achieving its learning objectives and influencing research practices, both in the short and long-term.

1.2 Objectives

The study aimed to:

1. Assess short-term knowledge and skill acquisition.
2. Evaluate long-term changes in research practices.
3. Inform improvements to course design and delivery.

1.3 Methodology

A mixed-methods approach was used, including:

- Analysis of Graduate School feedback forms (2020–2025).
- Pre- and post-training surveys (2024/2025).
- Semi-structured interviews with past participants and data stewards.

1.4 Key Findings

1.4.1 Short-Term Impact

- **High Satisfaction:** Participants consistently rated the course highly for content, trainer support, and relevance.
- **Increased Awareness:** Awareness of institutional RDM support rose from 45% to 98% post-training.
- **Improved Practices after the course:**
 - 90% changed their data storage strategies.
 - 81% implemented new backup strategies.
 - Use of secure institutional storage (e.g., Project Data U: drive) increased significantly.
- **Enhanced Understanding:**
 - Confidence in RDM skills and familiarity with FAIR principles improved markedly.
 - 98% felt better equipped to develop or improve their Data Management Plans (DMPs), with the Data Flow Map (DFM) assignment cited as a particularly valuable component.

1.4.2 Long-Term Impact

- **Sustained Behavioural Change:** Participants reported continued use of RDM practices up to 18 months post-training.
- **Positive Shift in Attitudes:** Participants moved from viewing RDM as a compliance task to recognizing its value in research quality and efficiency.
- **Mindset Shift:** Many described a transformation in how they approach data management, emphasizing organization, documentation, and reproducibility.
- **Emotional and Social Benefits:** The course reduced anxiety around RDM and fostered a sense of community and peer learning.
- **Trainer Influence:** Trainers were praised for their approachability and support, contributing significantly to the learning experience.

1.5 Areas for Improvement

- **Workload vs. Credits:** Participants felt the workload was high relative to the credits awarded.
- **Discipline-Specific Content:** There is demand for more tailored examples and case studies.
- **Delivery Format:** While in-person class sessions were preferred for engagement, online and hybrid formats improved accessibility.

1.6 Recommendations

- Introduce modular, discipline-specific learning paths.
- Rebalance workload and credit allocation.
- Keep the blended format of the course, and work on enhancing online delivery with interactive tools and structured sessions.
- Investing in expert-led training is key to embedding RDM practices in research at TU Delft.
- Continue investing resources on course evaluation to keep it relevant for PhD candidates.

2 Introduction

2.1 About RDM 101

Research Data Management 101 (RDM 101) (Martinez-Lavanchy et al., 2024) is a basic course offered to PhD candidates as part of the Doctoral Education programme at TU Delft. The course aims to equip researchers with essential knowledge and practical skills for managing research data and other relevant research objects in accordance with best practices. Below are the main learning objectives of the course:

- *realise the important role that good data management plays in research*
- *identify different types of research data and recognize the regulations, policies and/or legal requirements associated with them*
- *list the main components of the FAIR data principles and connect them to your own research workflows*
- *employ the acquired knowledge to design an efficient research data management strategy for your projects according to best practices.*

The course is delivered in a blended format, combining self-paced online modules in Brightspace LMS with three in-person or online (via Zoom) class sessions. It is taught by four RDM experts from the TU Delft Library's Research Data Services (RDS) training team. The online content includes videos, interactive images, quizzes, and opportunities for peer and trainer interaction through discussion forums. Over the three-week duration of the course, participants use their own research data as case studies to gradually develop a Data Flow Map (DFM) for their projects. This map can later serve as the basis for their project's Data Management Plan (DMP).

2.2 Study Objectives

The training team places high value on providing extensive feedback and delivering face-to-face, interactive sessions, as this approach is believed to support the long-term development of data management skills. However, this format is time- and resource-intensive, particularly due to the continuous feedback and detailed review of practical assignments it requires. This often raises the question: does the current course design and time investment of trainers effectively achieve the intended outcomes and impact?

To address this, the training team conducted an evaluation of the course, focusing on both its short-term impact, meaning what knowledge and skills participants gain immediately after the training, and its long-term effects - what changes in RDM practices the training helps researchers to bring about.

To achieve this broader aim, three main goals were identified: .

Goal 1 (short-term impact). To evaluate the extent to which participants have **gained knowledge and developed skills related to RDM best practices**.

Goal 2 (long-term impact). To explore how participants **apply RDM principles to their daily research tasks**, such as data collection, analysis, and interpretation. This could involve assessing changes in **data organization methods, metadata completeness, documentation quality, version control adherence, data sharing behaviours, etc.**

Goal 3. To inform improvements to the course **design, delivery format, and content**, by identifying areas where the course could better **align with participants' needs and expectations**, and ensuring its **continued relevance and effectiveness**.

3 Data Source & Methodology

The short-term impact of the RDM 101 course was evaluated using two data sources. First, data collected through *Graduate School feedback forms* from 2020 onwards was analysed. These forms provided **insights into participants' immediate reactions to the course content and delivery**.

Second, for this study a series of *pre- and post-training surveys* were conducted across four training sessions during the 2024/2025 academic year. These surveys aimed **to measure changes in participants' knowledge, attitudes, and confidence related to research data management as a result of the training**.

To assess the long-term impact of the RDM 101 course, *interviews were conducted with 11 participants* who had completed the course 12 to 18 months earlier. These interviews explored **how the training influenced their research practices over time**. In addition, two data stewards were interviewed to gather **their perspectives on any observable changes in the course participants' approach to research data management**.

As seen, we employed a mixed-methods approach, combining both quantitative and qualitative data sources. The figure below summarizes the main data sources used in this study, including feedback forms, surveys, and interviews with PhD candidates and data stewards.

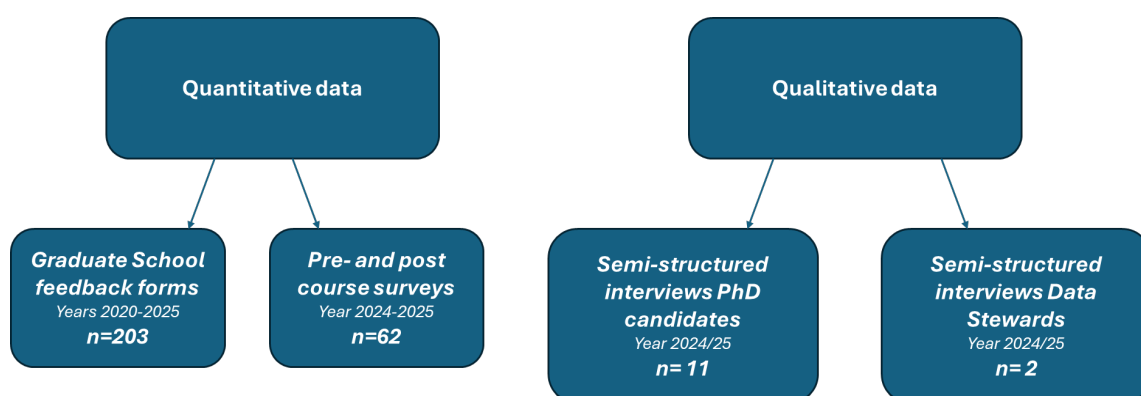


Figure 1. Overview of qualitative and quantitative data sources used in the study.

To perform this study, we submitted a Human Research Ethics Committee (HREC) application, which included the Data Management Plan (DMP) of the project. The preparation of documentation and the HREC application process was guided by the Data Steward at the Faculty of Technology, Policy and Management.

Due to ethical and privacy considerations, the raw research data containing potentially confidential and identifiable personal information are not publicly available and are securely stored on a university network drive. To support transparency and reproducibility of the study, the anonymised quantitative datasets and the high-level qualitative categories, together with their conceptual descriptions and containing no personally identifiable information about participants, are published in the 4TU.ResearchData repository (Rzayeva et al., 2025). The scripts used for the analysis are openly available in the 4TU.ResearchData repository (Grens, 2025) and are linked to a public GitHub repository to ensure the reproducibility of the analysis. For more information, please see the Data and Code Availability section of the report.

4 Short-term impact of RDM 101

4.1 Graduate School Feedback Forms

4.1.1 Objectives of the Analysis

To evaluate the effectiveness and reception of the courses, TU Delft Central Graduate School employed a standardized feedback form. This form is part of the Doctoral Education programme's quality assurance process. As RDM 101 is part of this programme, this feedback form has been distributed to RDM 101 participants since the course's first edition in 2020. Completion of the form is voluntary, and participants do not receive any incentives for providing feedback. The aim of the form is to gather participants' input on various aspects of the course, including its structure, content, and delivery by trainers. This study incorporates an analysis of the data collected from the feedback forms **to highlight the evolving perceptions and opinions of participants, providing insights into how the course has been received and developed over time.**

4.1.2 Graduate school Feedback Forms Overview & Structure

Over the years, the structure of the feedback forms has been modified, with changes in section names and specific questions. Despite these changes, the core objective has remained consistent: **to collect structured feedback from participants to inform continuous improvement.**

In the most recent edition, the feedback form is organized into eight key sections:

1. **Participant Starting Date of PhD.** Indication of the participant's current year in their PhD programme, providing context for interpreting their feedback.
2. **Overall Grade.** Participants assign an overall numerical grade to the course, offering a high-level indicator of satisfaction.
3. **Course Content.** This section measures the perceived quality of the course content, its alignment with participants' prior knowledge, and whether it met their expectations.
4. **Course Setup.** Participants assess how engaging the course was, how well the assignments aligned with the course content, and how the workload compared to the GS credits awarded.
5. **Trainers.** This section evaluates the trainers' effectiveness in presenting material clearly, fostering a safe and open environment, offering valuable feedback, and being receptive to suggestions.
6. **Course Information.** Participants rate the clarity of the information they received before the course started.
7. **Course Relevance.** Participants provide feedback on how relevant the course content is to the participant's current PhD research and future career.
8. **Recommendation & Clarification.** Finally, participants indicate how likely they are to recommend the course to fellow PhD candidates, offering insight into the perceived value of the course overall.

4.1.3 Methodology & Software

All of the above-mentioned sections, except two, included questions that were quantitative in nature, such as participant grading, voting, or rating. Sections 3 and 8 included open-text responses.

The feedback forms were distributed to participants by the Graduate School via EvaSysResearch application. For this study, the Graduate School exported the data in an excel format and share it with us. From these spreadsheets we selected questions that were relevant for the goal of our study and then analysed them using R programming language. The scripts used for the analysis are openly available on 4TUResearch Data repository (Grens, 2020) integrated with public GitHub repository to ensure reproducibility of analysis

For open-text questions, an initial round of manual coding of a sample of open-text responses was conducted by one principal investigator (NR). Subsequently, the coding scheme was refined, consolidated, and validated in collaboration with a second investigator (PML). Categories were developed separately for the pre-training and post-training sample datasets, which resulted in distinct category sets where this was methodologically appropriate. Once categories were defined the remaining survey datasets were cleaned and coded following the same scheme (as described in the accompanying codebooks). Categories were then assigned either through direct command-line input in RStudio or via pre-generated Excel files that allowed iterative revision. This workflow ensured consistent, transparent categorisation across datasets, and the final categorised responses were subsequently summarised and visualised to support interpretation.

To enhance transparency and reproducibility, the resulting codebooks, containing the final category structures and their conceptual descriptions, have been published in the 4TU.ResearchData repository (Rzayeva et al., 2025).

For certain course runs, the feedback forms results were not generated by the software, due to a low number of responses (less than two), which could compromise participant anonymity. At the time of writing this report, data from only one run of the 2025 course was available and has been included in the analysis. Based on the above mentioned factors, the number of feedback forms analysed for the report is as follows: 2020 year – 1 run, 2021 year – 5 runs, 2022 year – 2 runs, 2023 year– 3 runs, 2024 year – 6 runs, and 2025 year – 1 run.

When interpreting the results presented below in the next section, several contextual factors should be taken into account:

1. RDM 101 was delivered by two trainers from 2020 till the first half of 2023. Starting from the second half of 2023, the training team was expanded up to four trainers. Therefore, the number of course runs and the coverage of PhD candidates changed over the years, with a total of 190 PhD candidates (210 spots were offered) completing the course in 2024 - the year when all new trainers were involved in the training delivery.
2. Following the expansion of the training team, the course content, in-class exercises, and tools used in the training were revised, resulting in both minor and major updates to the course materials and delivery approach, which can influence differences in the perception of the course by participants.
3. Due to the COVID-19 pandemic, the course was delivered entirely online from 2020 to 2022. In the first semester of 2023 through the first semester of the 2024 academic year, the training shifted to an exclusively in-person format. Starting from the second semester of 2024, RDM 101 has been offered in a dual modality, with both online and in-person class sessions available.
4. As mentioned earlier, the structure of the feedback forms has evolved over the years. Some questions were removed, and new ones were added. For this reason, results of the analysis of certain sets of questions are available for the entire period of 2020-2025, while for others

results are only available starting from 2023, when the most recent revision of the feedback form was implemented.

4.1.4 Results

Subsections below represent the results of the analysis of feedback forms distributed to participants of the RDM 101 course in 2020-2025.

Participant Starting Date

The opening question of the feedback form asked participants to indicate which stage of their PhD trajectory they were in. Among those who completed the feedback form:

- In **2020** (n = 5), responses were spread across all years of the PhD trajectory: 2 were in their 1st year, and the others in later years (2nd to 5th).
- In **2021** (n = 52), around 70% (approx. 36) were in their 1st year, with a smaller number in the 2nd year and very few in later years.
- In **2022** (n = 11), about 60% (approx. 7) were 1st-year PhDs, and the remaining were mostly in their 2nd year.
- In **2023** (n = 33), about two-thirds (approx. 22) of the participants were in their 1st year, followed by some in the 2nd year and one in the 3rd year.
- In **2024** (n = 88), the majority—nearly 90% (approx. 79)—were in the 1st year, and a few in the 2nd and 3rd years.
- In **2025** (n = 14), all participants reported being in their 1st year of the PhD.

As observed, the vast majority of respondents were in the early stages of their PhD trajectory. This aligns well with the recommendation for researchers to take the RDM 101 course early in their research journey. Early participation ensures that they learn best practices in research data management before collecting actual data. Many of these practices relate to important considerations that need to be addressed prior to data collection or analysis—such as ethical approvals, legal requirements, and estimating data size to choose appropriate storage solutions.

Overall Course Grade

As a starting point for the course assessment, participants were asked to provide an overall grade for the course. The bar plots in Figure 2 illustrate the distribution of these grades across the analysed years. Throughout this report, “n” refers to the number of respondents who answered a particular question, “av.” indicates the average value, and “dev.” stand for the standard deviation.

Give an overall grade for this course (on a scale from 1 - 10):

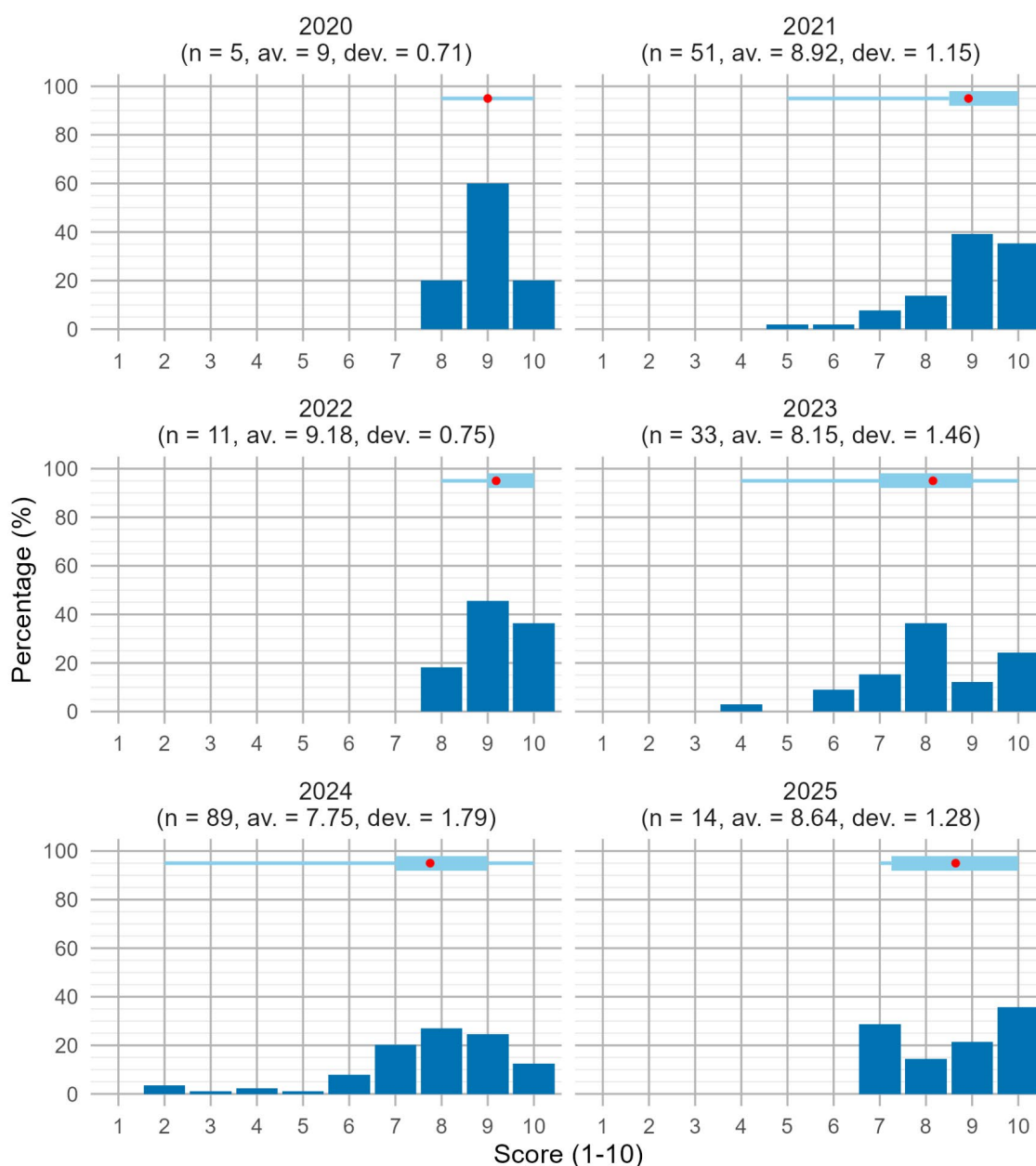


Figure 2. Overall course rating by participants.

As previously noted, the training team was expanded in the second half of 2023, and in 2024, the course underwent significant updates in both content and delivery format. This includes the introduction of an online modality alongside the existing in-person class sessions. These modifications enabled the highest coverage of participants in 2024, with 89 respondents completing the feedback forms. The expansion of the trainer team, substantial course revisions, and increased participant numbers may contribute to the relatively lower average (compared to the earlier years) and high variation (standard deviation) observed in the course grades for 2024.

As mentioned, for 2025, feedback data is available from only one course run. This limits the ability to draw robust conclusions about the longer-term impact of the updates. 2025 year results are of high interest since this year marks the first full year in which all course changes have been fully

implemented and stabilized. As seen from the figure the initial results from 2025 are promising, showing a high average grade from the participants.

Course Content

The overall content of the course was another subject for participants' grading. Figure 3 illustrates the results of grading of the course content from "very poor" to "excellent" by participants.

In my opinion the content of this course is:

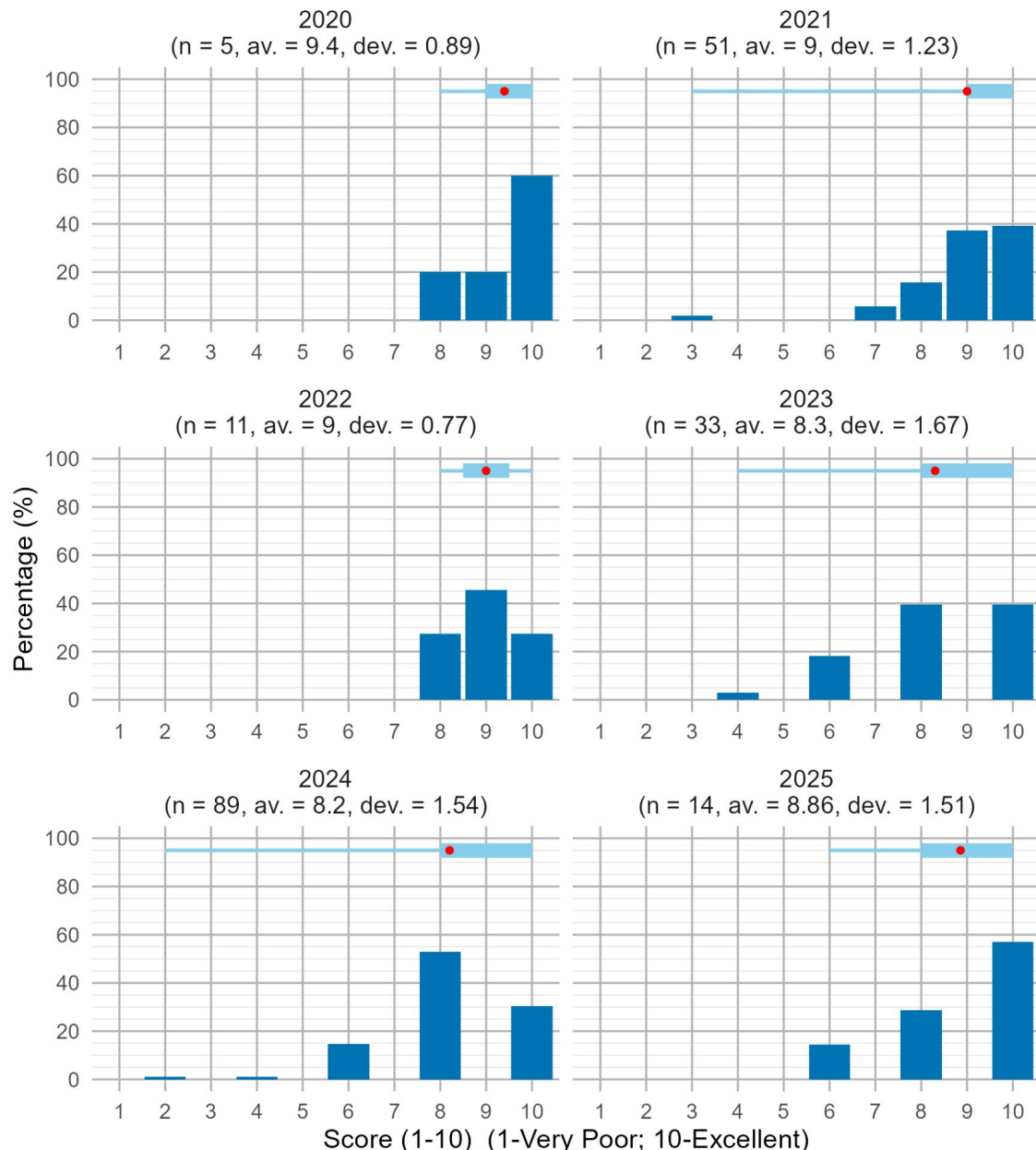


Figure 3. Course content rating by participants.

As seen, the content of the course received high grades from the very first year. Relatively lower average and higher deviation in more recent years can be explained by the reasons provided for the previous question. In our opinion, this trend also demonstrates a shift in the participants' requirements

regarding the course. Figure 4 shows whether the course content met participants' expectations. This data is available only from 2023 onwards, as this question wasn't included in the feedback forms in earlier years.

The content of this course matched my expectations.

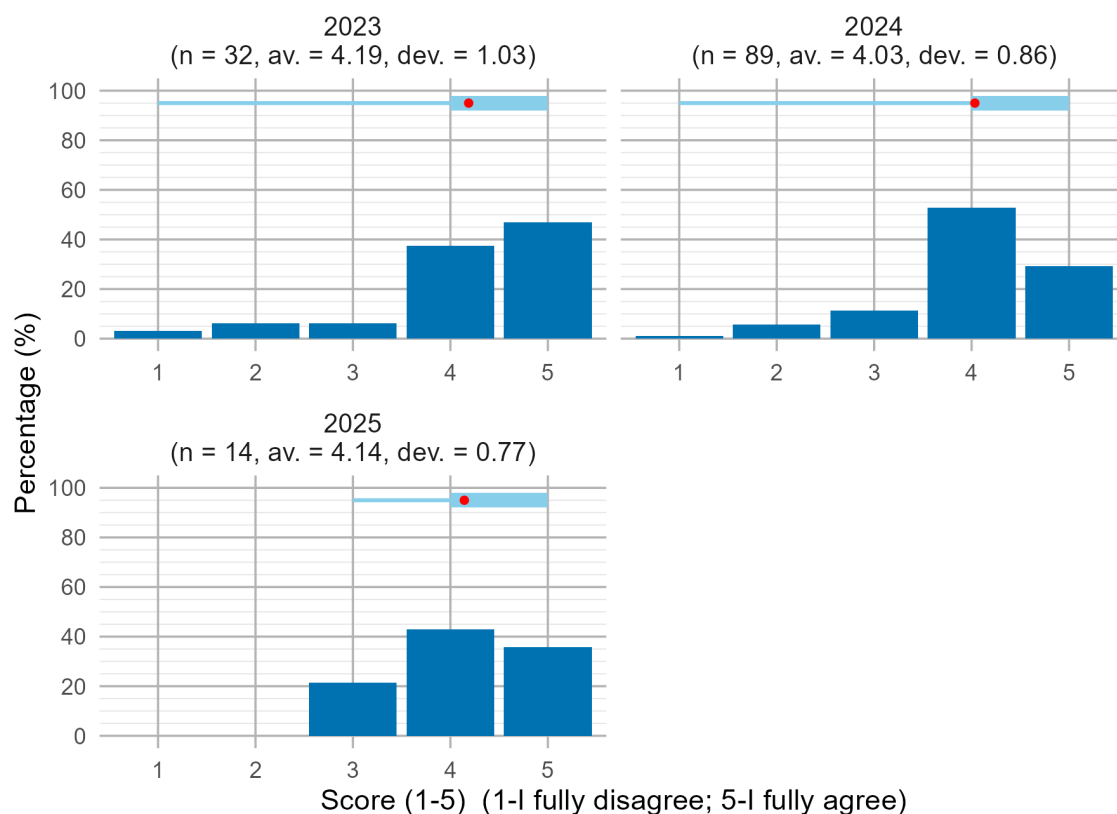


Figure 4. The agreement level with the course content matching participants' expectations.

Although the overall majority of students (82%) agreed or strongly agreed that the course met their expectations, we believe that, when considered together with responses to the previous question, the results indicate that there is still room for improvement in the course content to better align with participants' needs and expectations. Some insights into potential areas for improvement can be found in the suggestions provided by participants, which are presented later in the report in Table 1.

Course Setup

For us as the trainers team, it was also interesting to look more closely at the question where participants were asked about their engagement experience. This question is especially relevant to assess whether the updates to the training, such as changes to the in-class exercises and delivering sessions in two different modes (online and in-person), had any impact on participant engagement. As mentioned earlier, these changes were gradually implemented from the beginning of 2024.

How engaging did you experience this course to be?

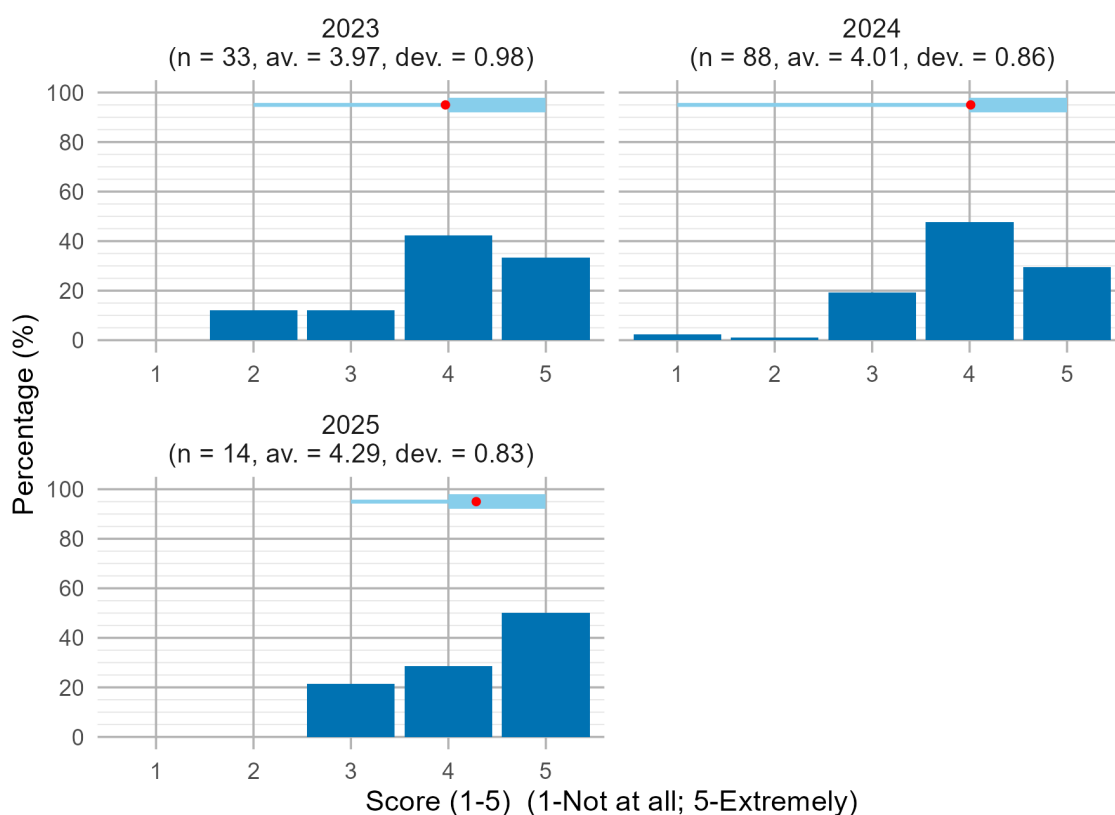


Figure 5. Course engagement level rating by participants.

When comparing 2024 with 2023 data, we observe that the average engagement level increased for a much larger number of participants, and the deviation decreased. The data from one run in 2025 can also serve as an initial supporting indicator of this trend. We believe this suggests that the recent modifications to the training delivery design and format may have had a positive effect on participant engagement.

Feedback on Trainers

Participants also evaluated the trainers from various perspectives. For instance, they were asked to rate their level of agreement with the statement “I felt safe to share personal experiences/opinions” on a scale from 1 (fully disagree) to 5 (fully agree). Across all years, responses were highly positive, with the vast majority selecting 5. In 2020, the average score was 4.8; in 2021, it rose slightly to 4.87. Similarly high levels of agreement continued in subsequent years: 4.91 in 2022, 4.58 in 2023, 4.57 in 2024, and again 4.57 in 2025. This is a particularly relevant indicator for us, as one of our main goals is to foster the exchange of knowledge and best practices between peers and trainers.

We analysed their responses from 2020 regarding the trainers’ ability to present course materials in a clear and structured way (Figure 6), as well as the clarity and value of feedback provided by trainers (Figure 7). This data was available only for the recent years.

The trainer(s) presented the material in a clear and structured way.

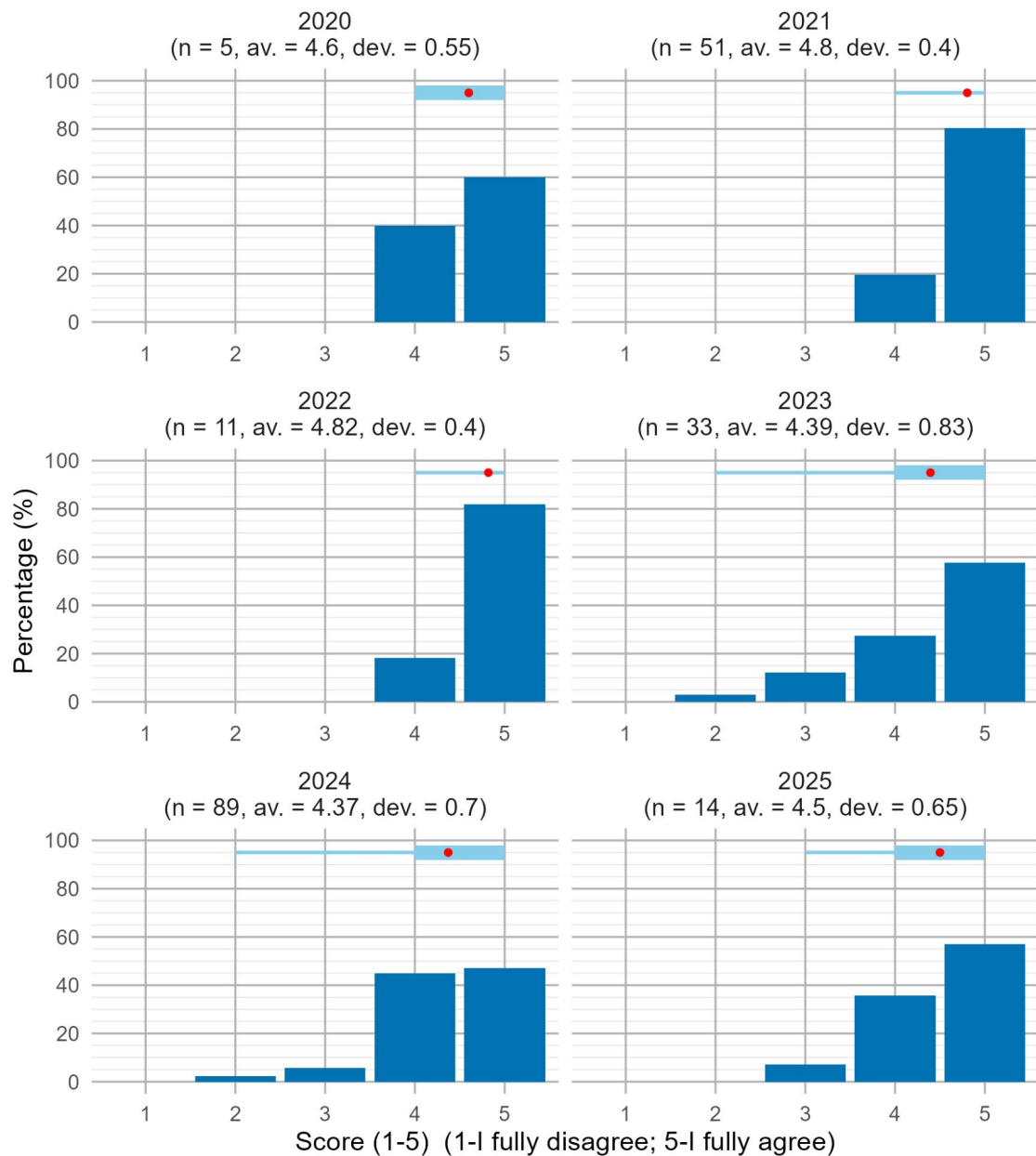


Figure 6. Participants' agreement level with the clarity and structure of the trainers' material presentation.

The trainer(s) provided clear and valuable feedback.

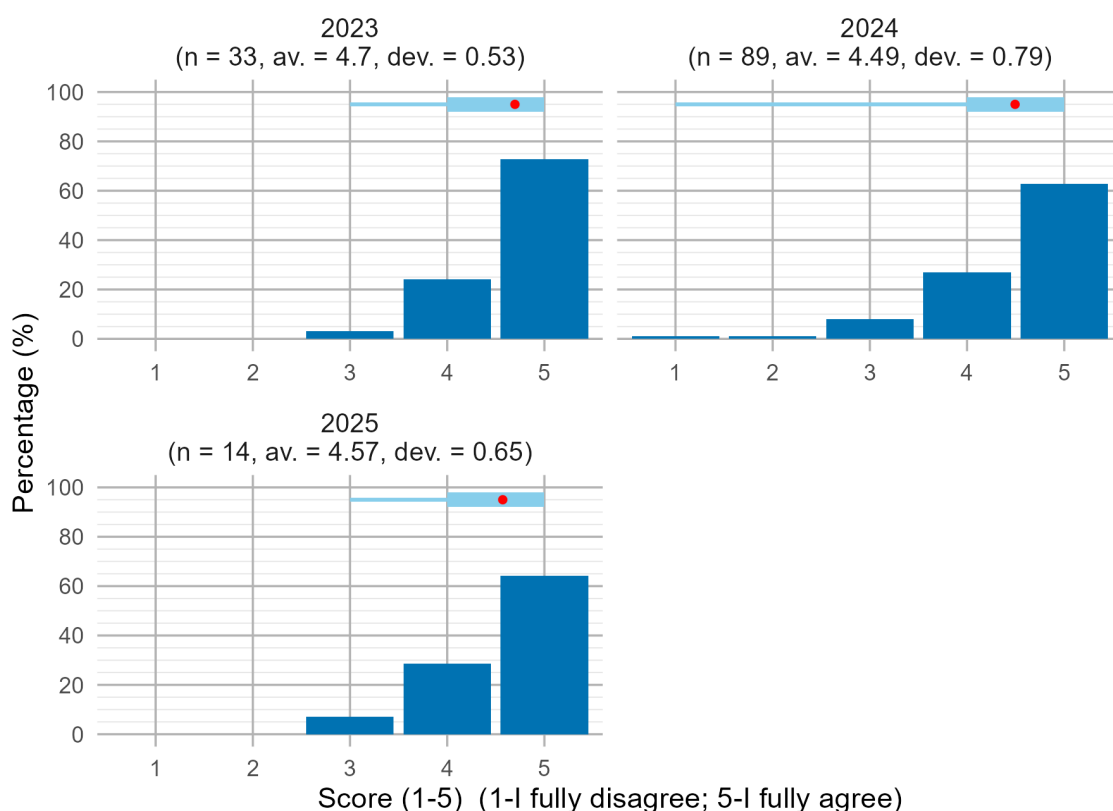


Figure 7. Participants' agreement level with the clear and valuable feedback provided by trainers.

As observed in the figures above, the vast majority of participants agreed or fully agreed that the trainers performed well across the given aspects in all observed years. This suggests that the course setup—including class sessions with trainer's explanations and interactions, feedback provided during sessions and on assignments, as well as other interactions with trainers via Brightspace is of high value to participants.

However, the slight decrease in average scores observed in 2023 and 2024 may indicate the effect of the training team still "landing and gaining experience". It also suggests that there is room for improvement for the training team, possibly through exploring new delivery methods and pedagogical approaches, feedback practices, as well as testing new session dynamics to further enhance the participant experience.

Course relevance

93% of participants agreed or strongly agreed that the skills provided in the course are relevant to their PhD project and/or trajectory, and 87% reported that for their future career. The data for these questions is available only for recent years.

This course provided me with skills that are relevant to my PhD project and/or to my PhD trajectory.

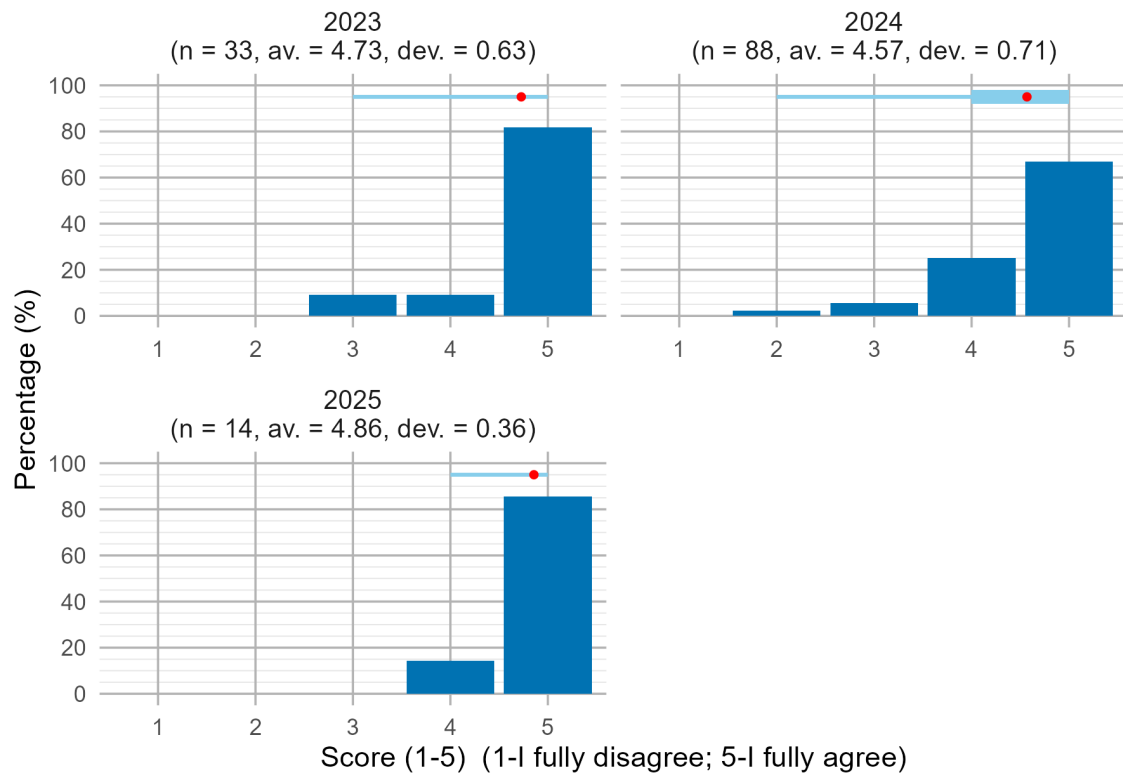


Figure 8. Participants' agreement level with the relevance of the skills provided by the course to PhD projects and trajectories.

This course provided me with skills that I expect to be relevant for my future career.

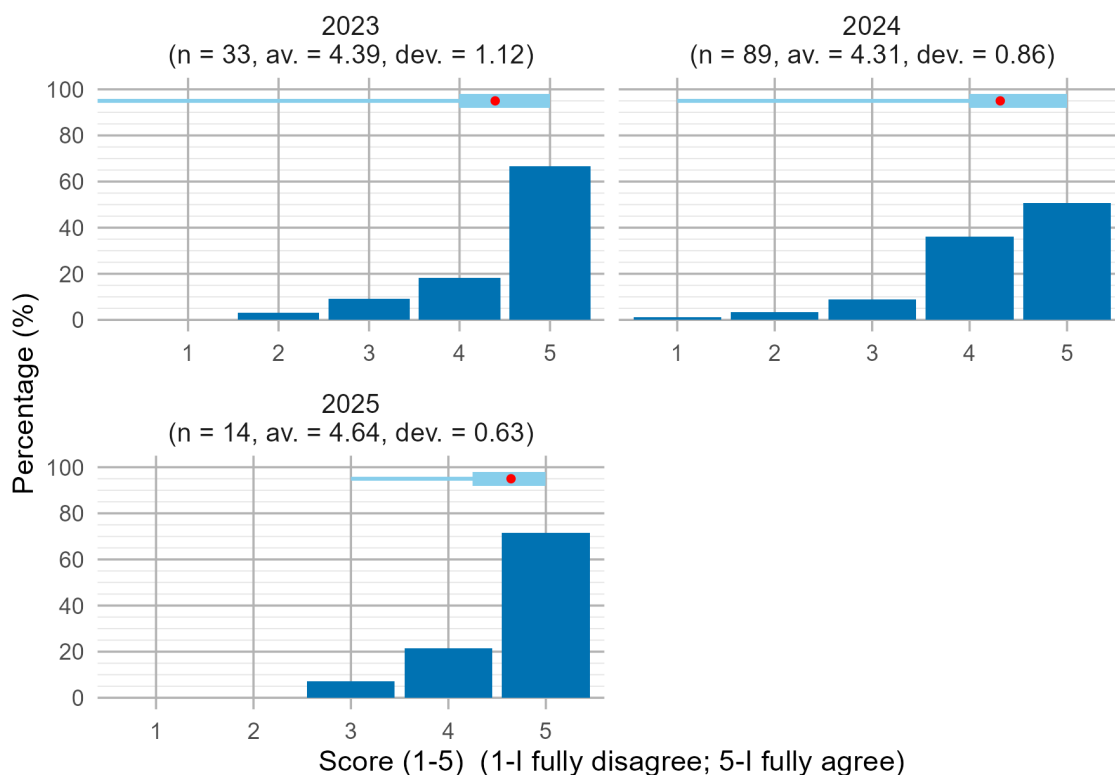


Figure 9. Participants' agreement level with the relevance of the skills provided by the course to their future career.

These results suggest that, upon completing the course, PhD candidates feel equipped with the necessary skills for effective research data management during their PhD project. The results presented in Figure 9 are the first indication that participating in the course is considered to provide skills that can be useful in the long-term.

We consider these two indicators as strong evidence supporting the importance of continuing the course, while also recognizing the need for further improvements to ensure the course remains relevant, with specific areas for enhancement to be identified through this study.

Participants Highlights and Suggestions

As previously mentioned, Section 8 of the feedback form included an open-text response designed to gather participants' insights on the most valuable aspects of the training and suggestions for improvement. Both these questions were optional. The responses on the question about the most valuable parts of the course were available from the first edition of the training in 2020 and are illustrated in Figure 10.

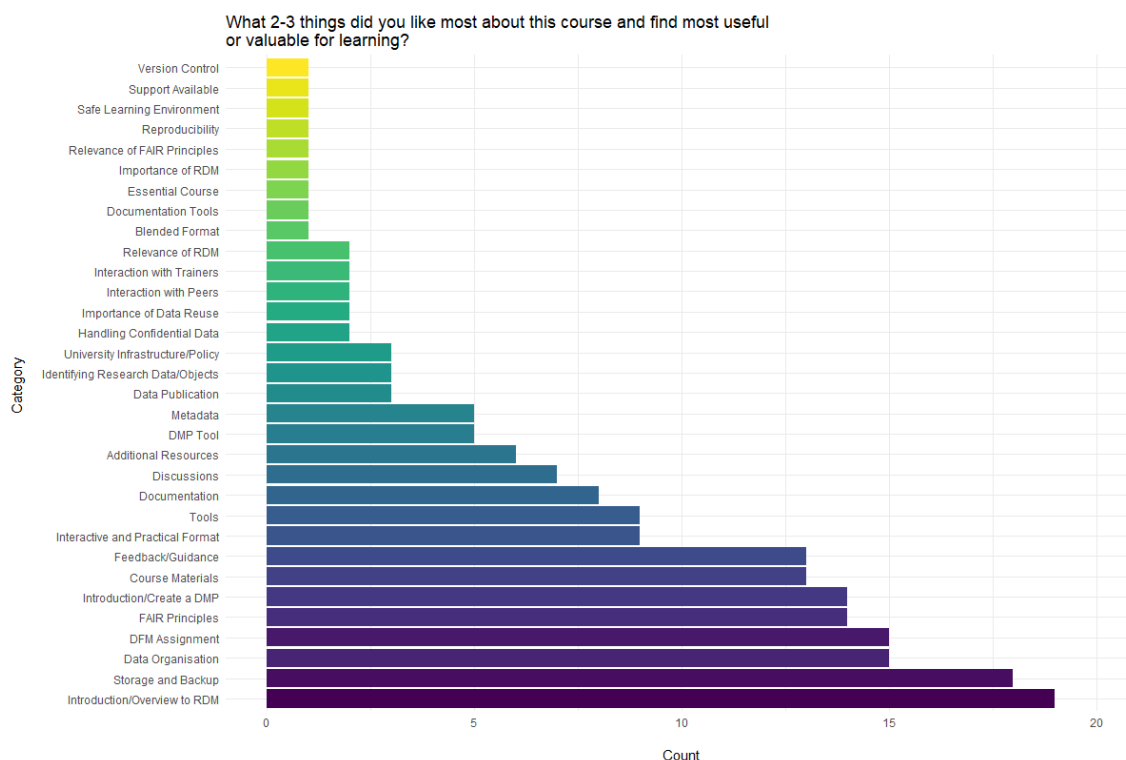


Figure 10. Aspects of the course participants found most useful and valuable for learning (n = 85). This plot was generated based on an open-ended question indicated in the title. The responses of participants were grouped and categorised manually.

Participants were also asked to indicate suggestions that would make the course a better learning experience. We grouped the main responses of the participants into five broad categories: “Workload”, “Content and Relevance”, “Class design”, “Interactivity and Format”; “General positives.” The table below demonstrates the summary of the responses as well as main highlights extracted from them.

Table 1. Participant suggestions to improve the course (n=71)

Category name	Summary and Highlights
Workload	<p>Participants highlighted that the workload is too high relative to the course credits. Many reported that the amount of self-study, homework, and preparation time exceeds their expectations for a PhD-level course. Specific suggestions include:</p> <ol style="list-style-type: none"> <i>1. Reducing the volume of assignments.</i> <i>2. Extending deadlines or allowing more preparation time.</i> <i>3. Assigning more credits to reflect the actual time commitment.</i>

Content and Relevance	<p>Some participants felt that the course content lacked discipline-specific case studies and practicality in terms of examples and templates:</p> <ol style="list-style-type: none"> <i>1. Include more discipline-specific case studies, tools and templates.</i> <i>2. Provide examples of real data management plans and data preservation.</i>
Class design	<p>Participants had mixed opinions about the number and format of class sessions. While some advocated for more meetings to allow deeper discussion, others preferred fewer sessions or even a fully self-paced format. Few participants also suggested improving the structure of the classes, particularly by reducing the emphasis on group discussions in favour of more direct instruction (lectures) and practical exercises:</p> <p><i>Reduce discussion time in favour of more focused lectures or hands-on exercises.</i></p>
Interactivity and Format	<p>Several participants found class sessions crowded and preferred in-person sessions over online classes:</p> <ol style="list-style-type: none"> <i>1. Suggest smaller class and more structured breakout sessions for better discussion.</i> <i>2. Prefer to have an option of offline (in-person) classes.</i>
General positives	<p>Substantial number of respondents found the course valuable and essential with strong recommendation on:</p> <ol style="list-style-type: none"> <i>1. Making the course mandatory for all the PhD candidates.</i> <i>2. Emphasizing its role in the early stages of PhD study.</i>

Participant comments related to “Content and Relevance” provide additional context for the results presented in Figure 4. Respondents expressed a need for content, use cases, feedback, and examples that are more closely aligned with their specific research topics. While the course already offers personalized feedback through discussion forums and extensive feedback on their assignments by the trainers and peers, recent cohorts, who often enter the course more aware about RDM and its importance, appear to seek more in-depth knowledge that supports direct application to their own research projects.

Therefore, as a training team, we would like to explore the possibility of introducing tailored learning paths within RDM 101. That seems to be the most suitable approach, given the existing resource limitations and concerns about the time required from trainers to provide intensive feedback. It may also help balance participants’ growing expectations for advanced content with the practical limitations of the current course structure.

4.2 Pre- and Post-training Survey

4.2.1 Objectives of the Analysis

Pre- and post-training surveys were designed as part of this research, to collect data aligned with the study's objectives. The primary objective of pre- and post-training surveys was to assess the immediate knowledge, skills, and changes in perspective that participants gained through the RDM 101 course. To guide this evaluation, the following research questions were formulated:

- To what extent does students' understanding of research data management change after taking the course?
- What are the expectations of the participants before the course? And are they met during the training?
- Which parts of the training students find more effective for gaining knowledge and developing skills?

4.2.2 Survey Overview & Structure

The survey was designed around four main sections:

1. General information, including information about participants (e.g., PhD year, academic discipline), as well as questions assessing participants' confidence and familiarity with RDM
2. Data practices - aimed at evaluating participants' knowledge, practices, and satisfaction in the following areas:
 - a. Data Storage and Backup
 - b. FAIR Data
 - c. Data Management Plan
 - d. Data Publication
3. Learning expectations - questions assessing participants' expectations prior to the course and perceived learning gains and training effectiveness after the course

The surveys were distributed to participants across four runs of the RDM 101 course: three delivered in the first semester and one in the second semester of the academic year 2024/2025. They were conducted at two key time points: the pre-training survey was conducted four days prior to the start of the course, and the post-training survey was shared at the end of the final class session, and a reminder was sent the next day. No incentives were offered to participants, and participation was entirely voluntary, with no impact on course credits. No personal data of participants was collected. The privacy statement and consent text added at the beginning of the survey was drafted with the support of the data steward from the Faculty Technology, Policy and Management to ensure alignment with GDPR requirements.

For convenience, from here onward, RDM 101 runs that were delivered with class sessions in online format will be referred to as "online runs," and those with in-person class sessions will be referred to as "in-person runs". For certain questions where it is relevant, we distinguish results of participants from online and in-person runs.

In total 115 participants attended the course across the four considered runs, with 58 participants joining the online runs and 57 participants in the in-person runs. Of these, 95 completed the pre-training survey and 101 completed the post-training survey, resulting in **response rates of 83% and 88%, respectively**. To ensure the robustness of the results, we included in the analysis only responses of participants who completed both the pre- and post-training surveys (here and after

“matched responses”). The table below shows the distribution of participants who took the pre- and post-training the surveys.

Table 2. Number of RDM 101 participants who completed pre- and post-training surveys.

Mode	Pre-training Survey	Post-training Survey	Matched Responses
Online runs	49	49	33
In-person runs	46	52	32

Three participants (two from online runs and one from in-person runs) did not complete the post-training survey, and their responses were therefore excluded from the analysis. As a result, the total number of matched participants included in the analysis is 62. Therefore, we consider that the effective response rate is 54%.

4.2.3 Methodology & Software

Both, pre- and post- training surveys include quantitative and qualitative mixture of questions. To match pre- and post-training survey responses from the same participant while maintaining anonymity, we included a question allowing respondents to generate a unique code. This code was based on information known only to them—that is, the first letter of the city where they were born, the number of siblings they have, and the first two letters of their current street. This three-elements ID code showed to be the most effective to avoid confusion between the participants.

All surveys were designed and distributed using the TU Delft-licensed Qualtrics survey platform. The analysis of the results was conducted using the R programming language. The scripts used for the analysis are openly available on 4TUResearch Data repository (Grens, 2025) integrated with public GitHub repository to ensure the reproducibility of analysis.

For open-text questions, an initial round of manual coding of a sample of open-text responses was conducted by one principal investigator (NR). Subsequently, the coding scheme was refined, consolidated, and validated in collaboration with a second investigator (PML). Categories were developed separately for the pre-training and post-training sample datasets, which resulted in distinct category sets where this was methodologically appropriate. Once categories were defined the remaining survey datasets were cleaned and coded following the same scheme (as described in the accompanying codebooks). Categories were then assigned either through direct command-line input in RStudio or via pre-generated Excel files that allowed iterative revision. This workflow ensured consistent, transparent categorisation across datasets, and the final categorised responses were subsequently summarised and visualised to support interpretation.

To enhance transparency and reproducibility, the resulting codebooks, containing the final category structures and their conceptual descriptions, have been published in the 4TU.ResearchData repository (Rzayeva et al., 2025).

4.2.4 Results

General Information

Figure 11a and Figure 11b provide an overview of the survey participants. Figure 11a shows their faculty distribution, while Figure 11b represents their distribution by year in the PhD trajectory.

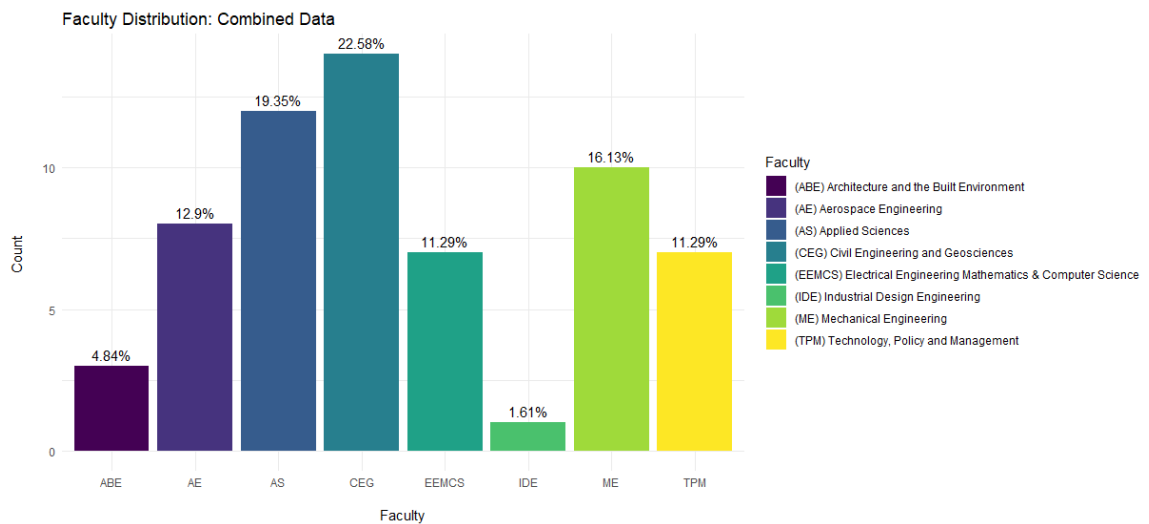


Figure 11a. Faculty distribution of survey participants.

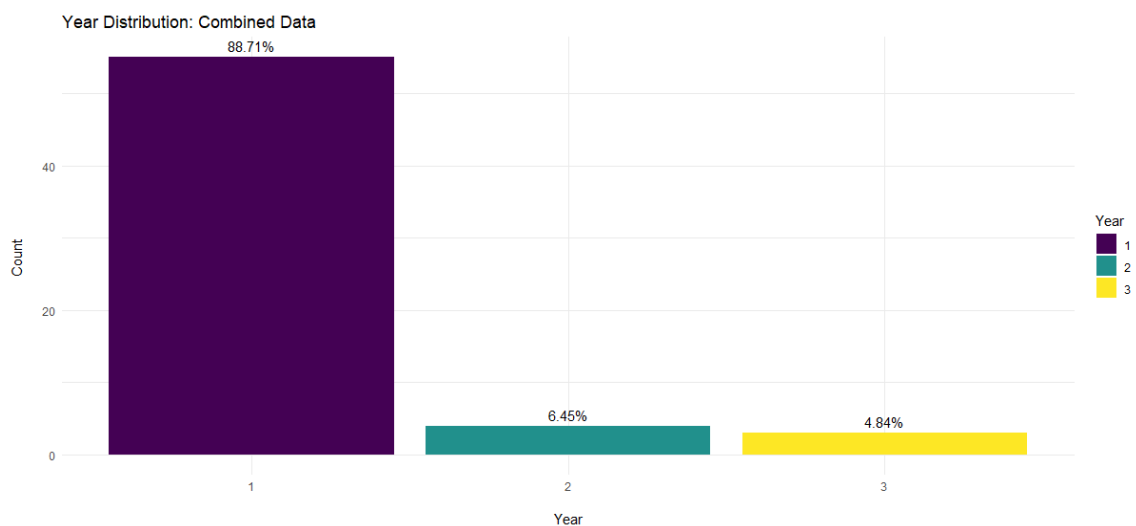


Figure 11b. PhD trajectory year distribution of survey participants.

To start with, we asked participants if they were aware about dedicated support available at the university/faculty for research data/software-related questions. Before the training 55% (n=34) of respondents reported they were not aware about any support services/resources available at TU Delft, and this percentage dropped to 2% (n=1) after taking the course. This was an open-ended question, allowing participants to indicate multiple options. Figure 12a and Figure 12b show the distribution of institutional support reported by participants before and after taking the course, respectively.

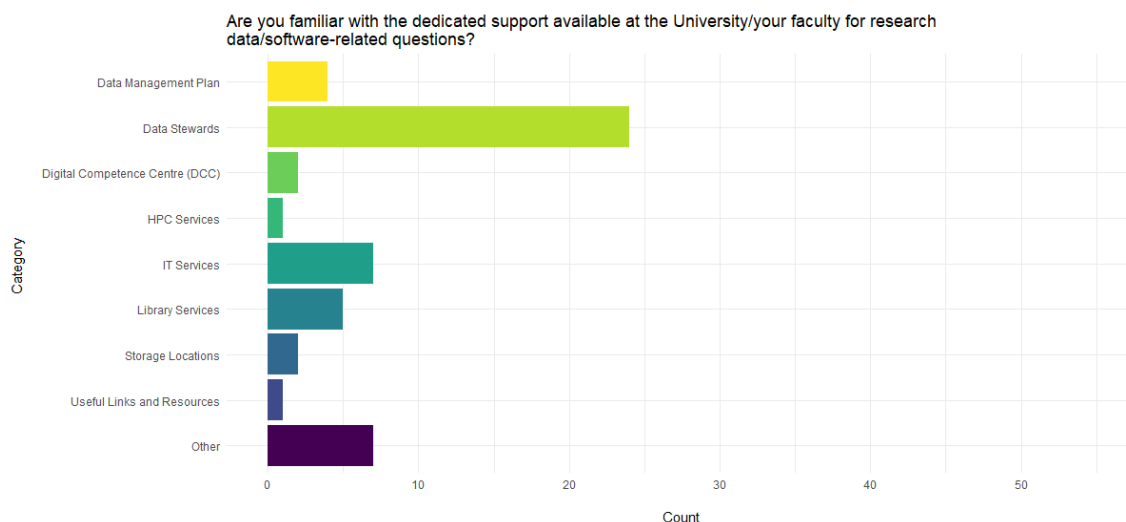


Figure 12a. Data/software-related support mentioned by participants in the pre-training survey. This plot was generated based on an open-ended question indicated in the title. The responses of participants were grouped and categorised manually.

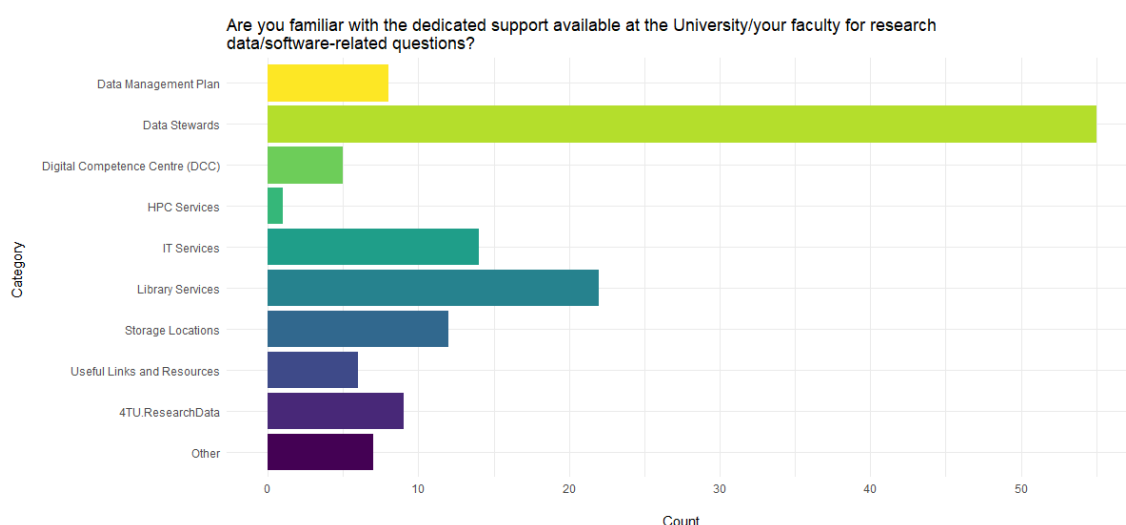


Figure 12b. Data/software-related support mentioned by participants in the post-training survey. This plot was generated based on an open-ended question indicated in the title. The responses of participants were grouped and categorised manually.

As seen from the figures the number of participants familiar with the support of data stewards almost tripled after the training, rising from approximately 24 to nearly 60 responses. Recognition of storage locations also substantially rose along with awareness of library services and IT services which roughly doubled. 4TU.ResearchData was only mentioned in the post-training responses, indicating that the training introduced this resource to participants who were previously unaware of it. All these shifts demonstrate that the training impacted in improving participants' knowledge and awareness of available institutional support regarding research data and software management.

Data practices

Data Storage and Backup

One of the topics and assignments in the RDM 101 course is devoted to storage and backup practices for research data. The key message is that when choosing a storage location, researchers must ensure that the data of their project is stored in a secure and sustainable manner. Special emphasis is placed on the importance of securely storing sensitive, confidential, or personal data. Researchers also need to consider data longevity and proper backup strategies. Furthermore, after the project ends, PhD candidates must ensure that the data remains accessible to TU Delft employees. In contrast to laptops, external hard drives, and commercial cloud services which do not meet above mentioned requirements, the course recommends using the Project Data (U:) drive as the primary storage location for research data. The Project Data (U:) drive is a network drive backed up by TU Delft ICT. The course also introduces other institutionally supported storage, backup, and data-sharing options, such as the Project Data (M:) drive and SURFdrive while emphasizing their limitations. For research software/scripts/codes the course materials and trainer's recommendations suggest using GitHub and GitLab.

90% of respondents on the post-training survey reported they changed their strategy for research data/code (scripts) storage during or after taking after the course. The figures below show what were the main preferred data/code storage locations reported by participants before and after taking the course.



Figure 13a. Storage locations for research data and code indicated by participants in the pre-training survey.

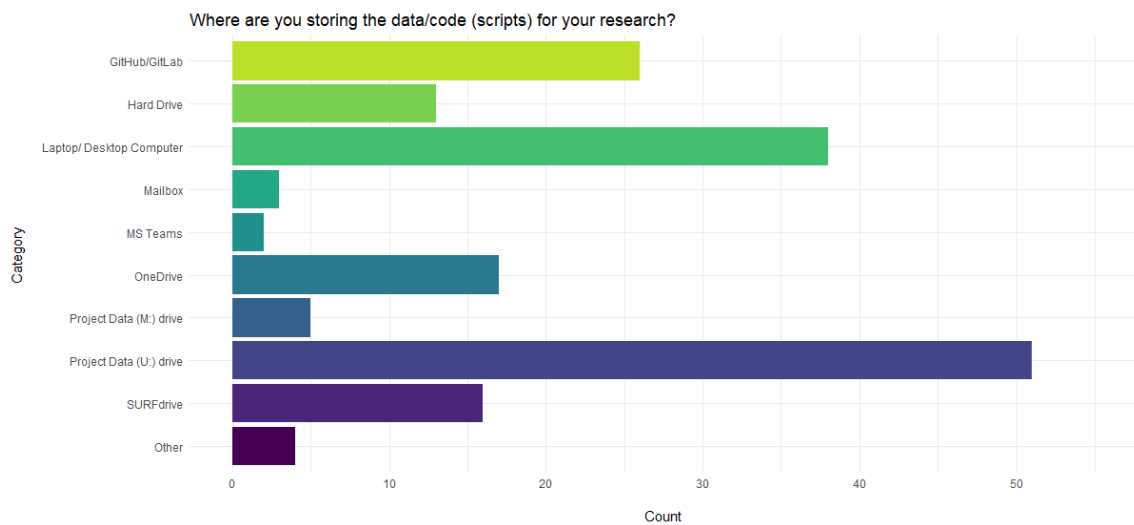


Figure 13b. Storage locations for research data and code indicated by participants in the post-training survey.

81% of participants reported that they had established a backup strategy during or after completing the course. The figure below illustrates the primary data/code backup locations reported by participants before and after the training.

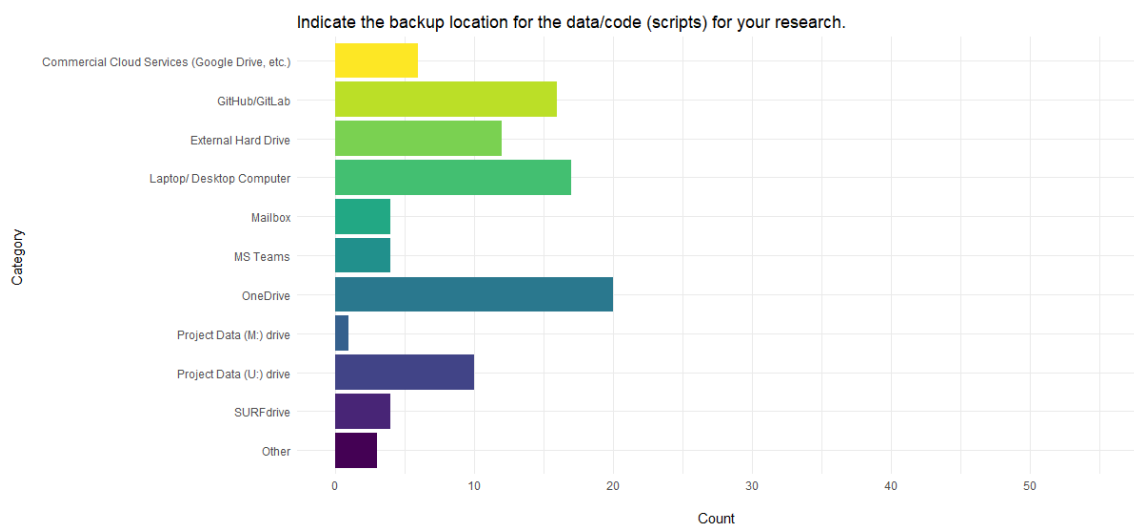


Figure 14a. Backup locations for research data and code indicated by participants in the pre-training survey.

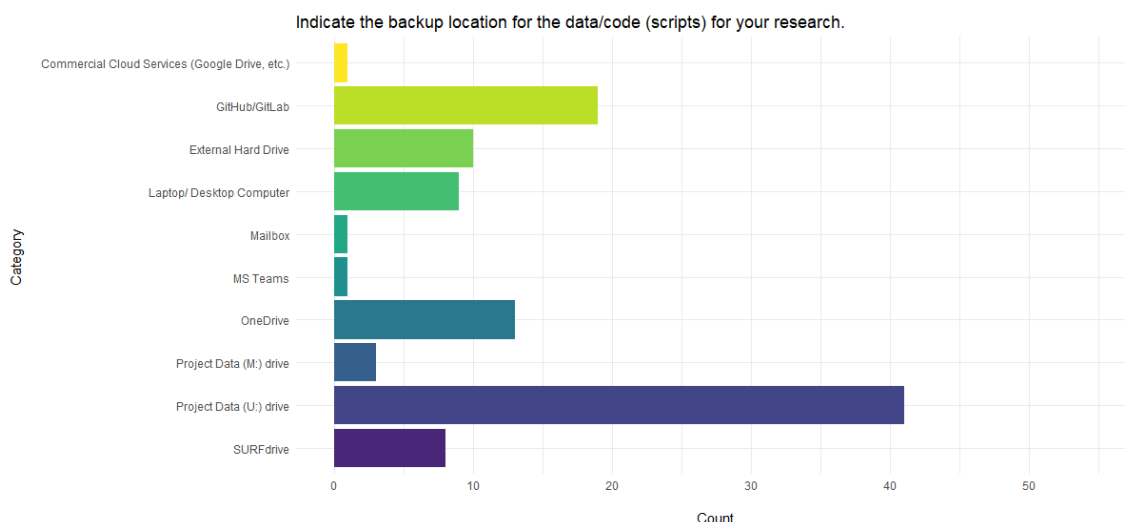


Figure 14b. Backup locations for research data and code indicated by participants in the post-training survey.

As seen from the figures, recommended institutional solutions such as the Project Data (U:) drive became significantly more relevant for both storage and backup among participants. Its usage increased from around 20 to 50 responses for storage, and from 10 to nearly 40 responses for backup. There was also a modest increase in the use of SURFdrive, indicating growing awareness of this institutional storage promoted during the training. OneDrive and Laptop/Desktop computer, which had been the most preferred backup options before the course, were overtaken in the post-training survey by Project Data (U:) drive.

Use of Git platforms (e.g., GitHub, GitLab) did not increase substantially, suggesting that participants were already familiar with version control systems before the training. Their consistent presence and slight increase in responses suggests stable interest in version control systems among participants. In contrast, the use of commercial cloud services (e.g., Google Drive) for primary storage and backup almost disappeared after the training, suggesting a shift away from less secure platforms.

Familiarity, Importance and Confidence in Key RDM Concepts

The surveys did not include a dedicated section focused on key RDM concepts in a combined way. Instead, the questions were organized thematically, covering topics such as FAIR data, Data Management Plans, and others. However, these sections included “mirroring questions”, identically worded questions in both the pre- and post-training surveys designed to capture changes in participants’ views, awareness, and confidence regarding the key concepts addressed in the course.

Figure 15 presents a combined analysis of the responses to these questions, illustrating shifts in participants’ perspectives before and after the training. Participants were asked to rate each of the following aspects on a scale from 1 to 5, both before and after taking the course:

- Confidence in their Research Data Management skills
 - 1 – Very low confidence
 - 2 – Low confidence
 - 3 – Neutral
 - 4 – High confidence
 - 5 – Very high confidence

- Familiarity with the responsibilities of PhD candidates regarding Research Data Management according to the TU Delft Research Data Framework Policy
 - 1 – Not at all familiar
 - 2 – Slightly familiar
 - 3 – Moderately familiar
 - 4 – Very familiar
 - 5 – Extremely familiar
- Familiarity with the FAIR principles for data and/or software
 - 1 – Not at all familiar
 - 2 – Slightly familiar
 - 3 – Moderately familiar
 - 4 – Very familiar
 - 5 – Extremely familiar
- The importance of creating a Data Management Plan
 - 1 – Not at all important
 - 2 – Slightly important
 - 3 – Moderately important
 - 4 – Very important
 - 5 – Extremely important

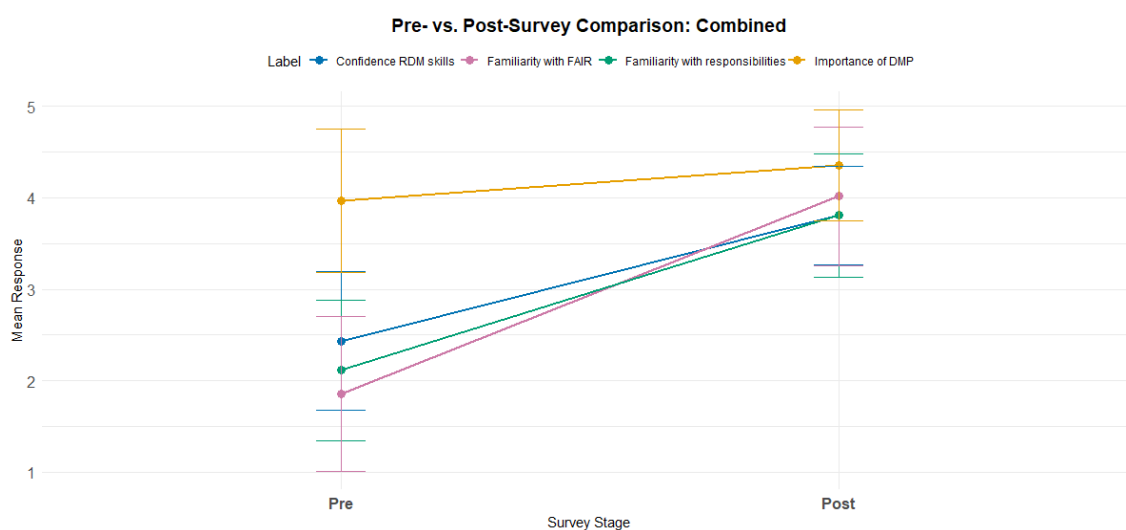


Figure 15. Pre- and post-training ratings of participants' confidence and familiarity with key RDM topics, including RDM skills, the FAIR principles, PhD responsibilities, and the importance of DMP.

We observe positive trends of increasing awareness and confidence within the participants about the RDM concepts mentioned above. The most significant change is seen in participants' familiarity with the FAIR principles, where the average rate increased from less than 1.9 to 4.0. We think this result is of a big value, as FAIR principles are there to enhance transparency and reusability of research outputs supporting the goals of Open Science. Average rate for confidence in RDM skills and familiarity with PhD candidates' responsibilities increased from around 2.4 and 2.1, respectively, to 3.8. A less pronounced change was observed in participants' ratings of the importance of DMPs. However, as seen from Figure 15, DMPs were already highly valued before the training, with an average rating of almost 4.0. After the training, this number increased to around 4.4. Additional

information to explain this result is provided by the subsequent question, which explored participants' views on training impact to their DMPs in more detail.

Data Management Plan

In the pre-training survey 19% of participants reported that their projects already had a DMP, and 31% reported that the work on a DMP had started, but had not been completed yet. We have to remember that according to the Research Data Framework policy of TU Delft (Ahlers et al., 2020), all PhD candidates have to submit a DMP as part of the documentation of the Go/no-Go step of the PhD Development Cycle (Delft University of Technology, n.d.) established by the Central Graduate School. Combination of this result with the high level of perceived importance of DMPs before the course reported in the previous section suggests that participants became aware of the role and importance of a DMP once it was introduced to them as part of the RDM requirements set by the policy and Graduate School. However, after the training, 98% of participants also reported that the course provided helpful knowledge and skills to start developing/improving their DMP. The figure below illustrates main thematic categories related to this question.

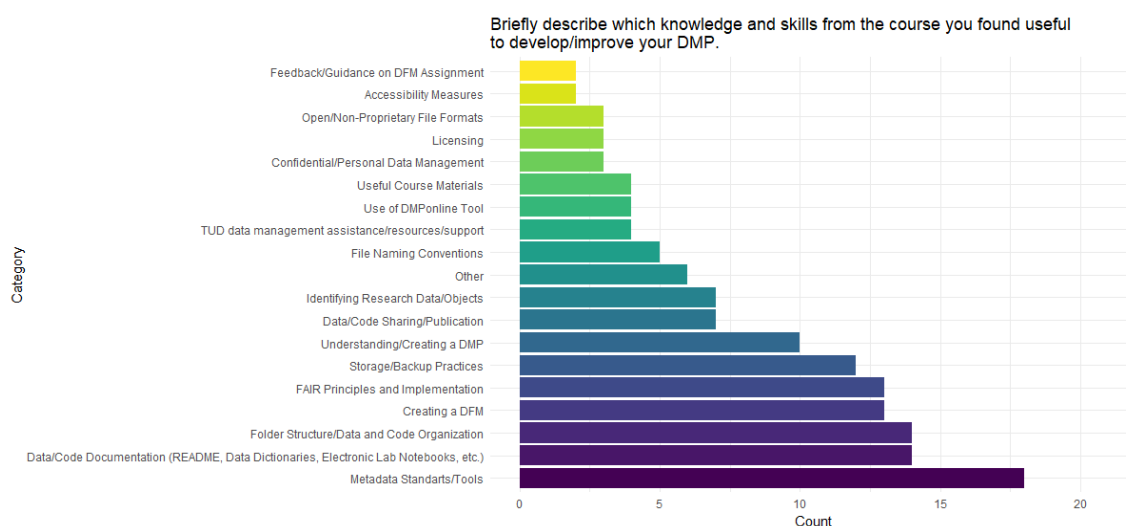


Figure 16. Knowledge and skills that helped participants improve or develop their DMP. This plot was generated based on an open-ended question indicated in the title. The responses of participants were grouped and categorised manually.

Based on the figure, in our opinion, RDM 101 appears to have played a significant role in equipping participants with practical and applicable competencies. The most frequently mentioned areas relate to concrete tools, templates, standards, practices and implementations. This suggests that participants particularly valued hands-on, practical aspects of the course that directly support the creation and enhancement of DMPs. The prominence of the DFM exercise among the most useful elements for participants is particularly valuable for us as trainers. DFM is an assignment that requires participants to apply gained knowledge in a structured way by mapping the data flows for their own research projects. The DFM themes were designed in alignment with the sections of a DMP, helping participants bridge their theoretical knowledge with real-world application. All DFM assignments are reviewed by the trainers, who provide feedback to support participants in refining their data management practices.

that while participants seemed to be already familiar with the general definitions of FAIR, the course helped them gain more practical knowledge and skills for implementing these principles.

After completing the course, all participants, except one, reported that they had planned or implemented specific measures/actions to make their research data and/or code (scripts) more FAIR, based on the knowledge and skills gained during the training. Participants' responses were manually grouped into thematic categories, which are illustrated in the figure below.

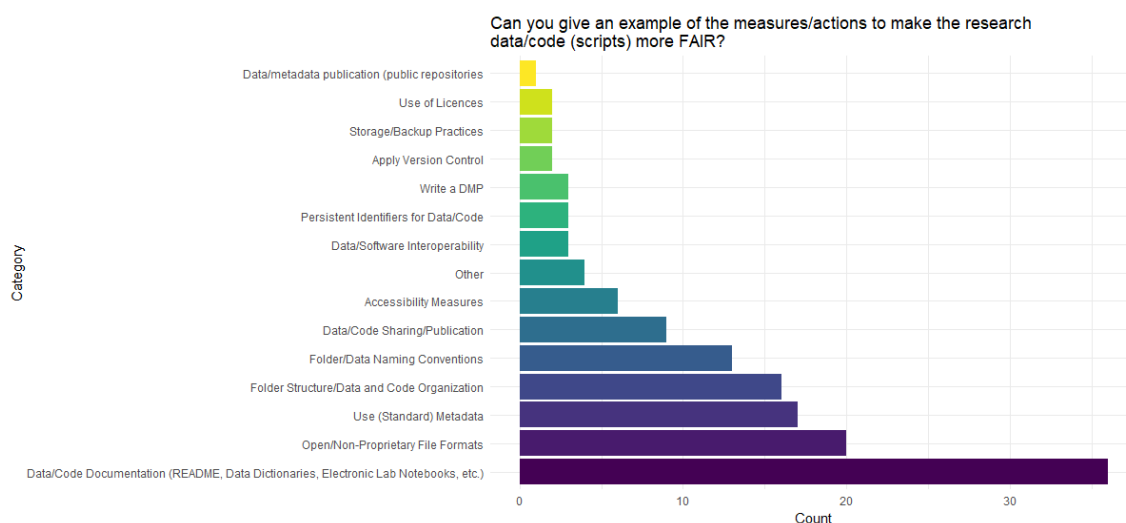


Figure 18. Examples of measures/actions to make research data/code more FAIR, based on knowledge and skills gained during the training. This plot was generated based on an open-ended question indicated in its heading. The responses of participants were grouped and categorised by the training team members.

The only participant who reported not having planned or implemented any specific measures explained that their software already adhered to the FAIR principles, thanks to the prior knowledge before the training.

Data publication

Before the training 76% of participants planned to publish the data/code (scripts) of their research projects. This number grew to 95% after completing the course. The figure below illustrates which repositories participants indicated to publish their data/code (scripts) in pre- and post- training surveys. Participants were allowed to indicate multiple repositories.

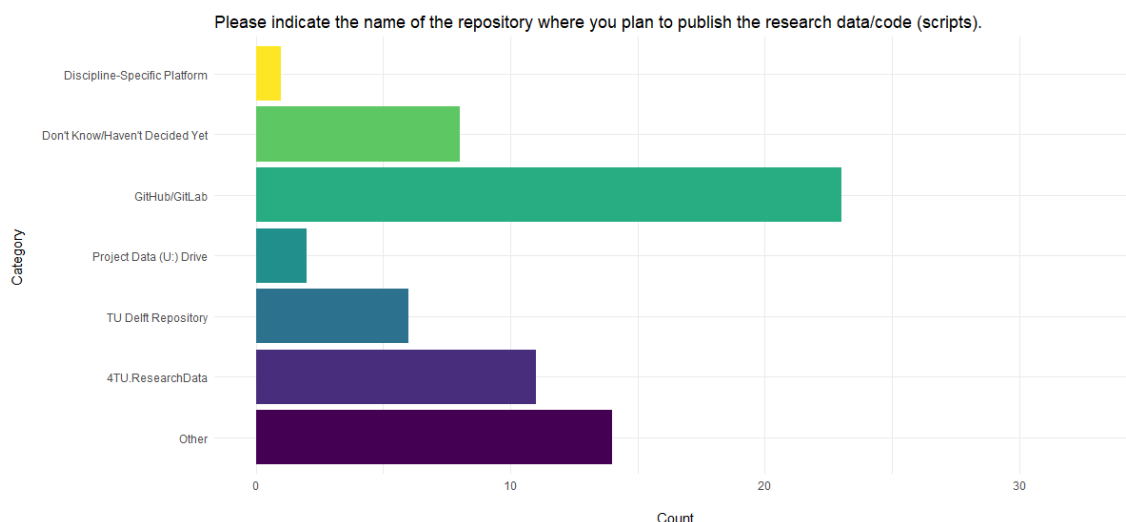


Figure 19a. Planned repositories for publishing research data/code (scripts) indicated by participants in the pre-training survey. This plot was generated based on an open-ended question indicated the title. The responses of participants were grouped and categorised manually.

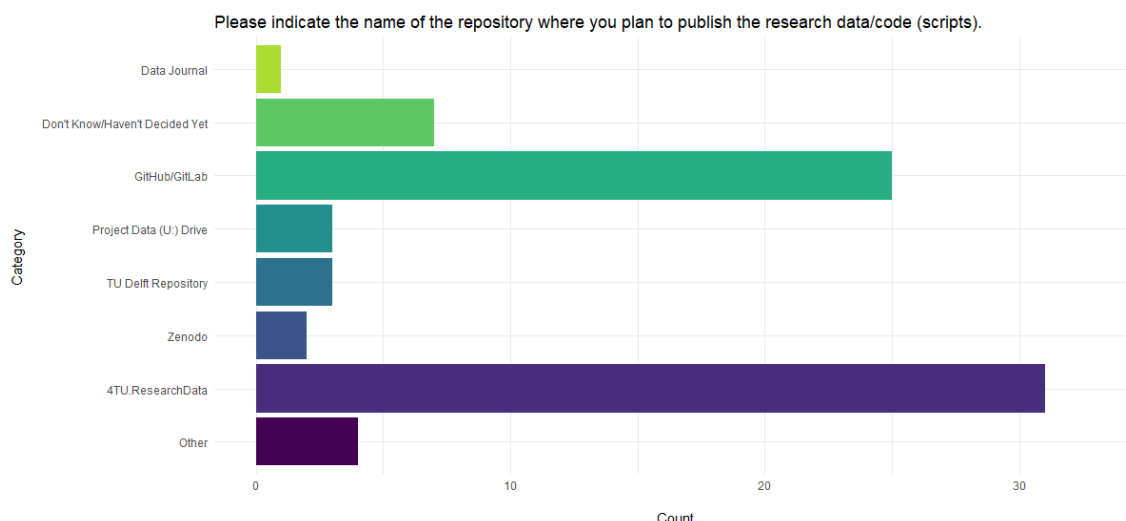


Figure 19b. Planned repositories for publishing research data/code (scripts) indicated by participants in the post-training survey. This plot was generated based on an open-ended question indicated the title. The responses of participants were grouped and categorised manually.

Although the vast majority of participants already intended to publish research data or code (scripts) of their projects before the training, Figure 19a and Figure 19b show a shift in their choice of repositories. For instance, we see a twice doubled increase in preference for the 4TU.ResearchData repository after completing the course. This repository is specifically recommended during the training as a reliable option that meets key requirements related to institutional governance, data management best practices, and technical infrastructure, while also ensuring long-term data preservation. Zenodo, another recommended repository highlighted in the course materials, appeared in the post-training responses only, though at a low rate.

As also discussed in the section on data storage and backup, the relevance of platforms like GitHub and GitLab for participants remained relatively stable before and after the training. This suggests that most participants were already familiar with these platforms prior to RDM 101.

We also see some confusion in the survey responses regarding the use of the Project Data (U:) drive and the TU Delft Repository for data publication, both of which are not intended for this purpose. Although this misunderstanding decreased after the course as indicated in the post-training survey, it still persisted among some participants.

Before the training 20 participants reported that they do not plan to publish the data/code (scripts) of their research project. When asking participants about their reasons for not planning to publish their research data/code (scripts), the most common themes included concerns around privacy and confidentiality, especially when dealing with confidential data from external partners or data-type specific databases, which restricts public sharing due to contractual or legal obligations. Another frequent reason was uncertainty or indecision when participants indicated that they had not yet discussed data publishing plans with their supervisors or advisors. Some respondents explained that the data/code (scripts), in their view, would not be broadly beneficial to the scientific community. Several respondents mentioned working with personal data.

After the training, the number of participants who reported that they do not plan to publish any data or code (scripts) from their research project decreased to 3. The reasons behind these responses matched the same as before the training: confidentiality concerns and uncertainty or indecision.

Learning Expectations

As a team of trainers, we were interested in understanding how different training components performed across various delivery formats — specifically during online and in-person runs. The figure below presents participants' ratings of the helpfulness of specific training components in both formats.



Figure 20. Participants' ratings of how helpful each training component was in facilitating effective learning, presented separately for in-person and online formats. Results are shown as percentages on a 1–5 scale, where 1 = Not helpful at all and 5 = Extremely helpful. Participants could rate each component independently.

Most course components are considered effective for facilitating learning, with the majority of ratings falling between 3 and 5. Looking separately at the delivery formats, we see that the effectiveness of assignments and the feedback provided, as well as the materials and activities provided through the

LMS appeared to be equally effective across both formats, with the balanced distribution of ratings between “4” and “5” points.

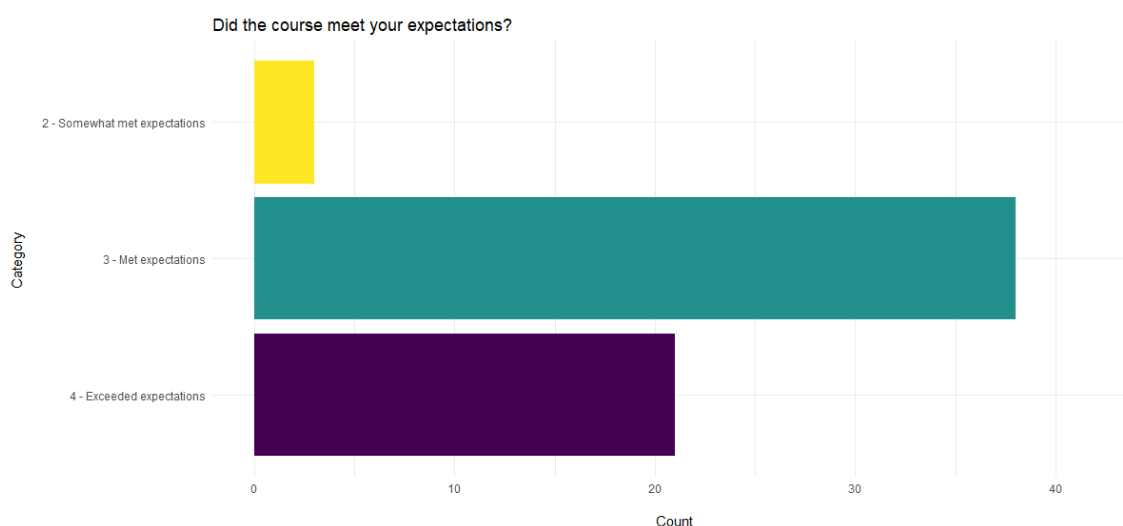
Notably, in-class activities and interaction with peers and trainers received higher ratings from almost twice as many participants in in-person sessions compared to online sessions. These results provide us with valuable insights on which components we should prioritise and invest resources in when updating the course.

Before the training we asked participants if they can provide an example of how they think this course could help them with their research. This question was optional, the number of participants replied was 33. It was an open-ended question and the responses of participants were grouped and categorised by the training team members. The top five categories we could identify among responses include efficient data/code organization/structure/management, guidance in writing/finalising a DMP, data sharing/publication, data storage/backup, and accessibility measures.

After the course, 74% of participants reported that the training helped them address specific challenges or problems related to data and/or code (scripts) management in their research projects. This question was open-ended and the responses of participants were grouped and categorised by the training team members. The top five categories that emerged were: folder structure/data and code organization, use (standard) metadata, data/code documentation (readme, data dictionaries, etc.), file naming conventions, and research data/software management practices.

Before the training, participants were also asked what they expected to learn most from the course. The question was open ended and participants’ responses were again manually categorised. The aspects that participants expected to learn more about from the course include efficient data/code organisation/structure/management, storage/backup practices, DMP writing and guidance, TU Delft RDM guidelines/regulations, and accessibility measures.

To finalise the post-training survey, and get a global impression about satisfaction of the participants with the course we asked participants to rate how the course met their expectations. The results are illustrated in Figure 21 combined overall for all participants (top part), and separately for the online and in-person runs (bottom part).



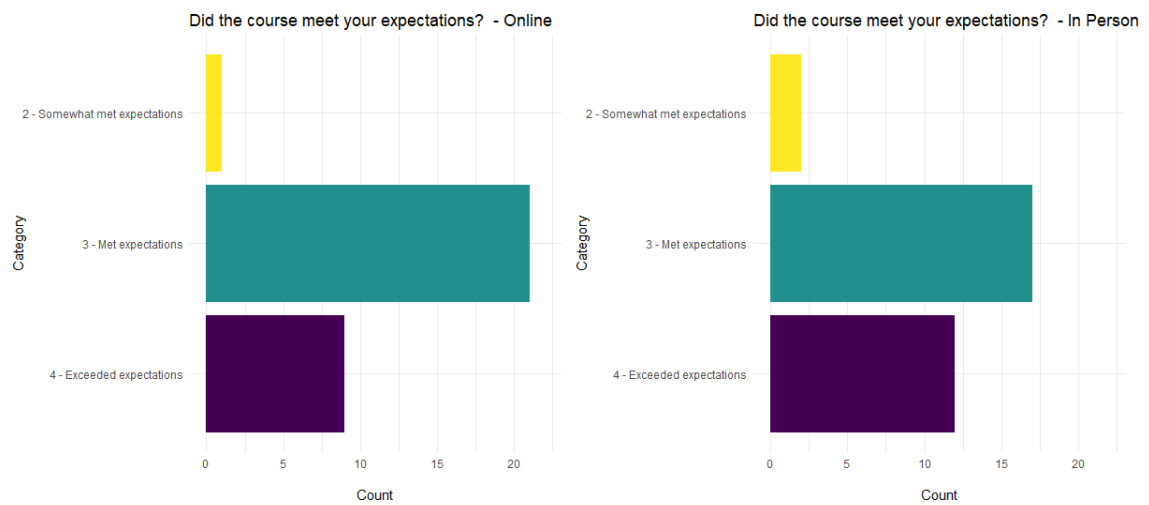


Figure 21. Extent to which the course met participants' expectations.

Despite the differences in the evaluation of certain course components between the in-person and online formats observed in Figure 20, the results in Figure 21 show that the course meets or exceeds learners' expectations, regardless of the delivery mode.

5 Long-term impact of RDM 101

5.1 Objectives of the Analysis

To assess the long-term impact of the RDM 101 course on participants' knowledge, attitudes, skills, and behaviours related to research data management, we conducted semi-structured interviews with PhD candidates who had completed the course RDM 101 between 1 and 1.5 years prior to the interview. Additionally, two data stewards were interviewed to provide external perspectives on any observable changes in participants' approaches to RDM.

5.2 Interview Overview & Structure

Each interview lasted up to one hour and followed a semi-structured format. We first asked participants general questions about when they attended the RDM 101 course and whether they took it in person or online. The conversation then explored their motivations and expectations for joining the course, as well as their prior knowledge and experience with research data management practices.

We then invited participants to reflect on what they learned during the course, how they interacted with trainers, data stewards, and peers, and how they experienced the training overall. They were also asked to describe any changes in their RDM practices after completing the course, how they applied the knowledge gained, and whether they continued to use the skills they had learned.

The interviews concluded with a request for suggestions and recommendations on how the course could be improved.

Below is a copy of the interview guide we used as a prompt for the discussions. While the guide provided a structure, additional ad-hoc follow-up questions were asked to the interviewed participants as the conversation developed. These spontaneous prompts were tailored to each interview and are not included in the guide.

PART A - INTRODUCTION

- Can you introduce yourself?
- Which year of the PhD are you?
- Which faculty?

PART B - Experience before or during the course

- When did you attend the RDM 101 course?
- Did you attend in person or the online course?
- Why did you choose to attend the RDM 101 course?
- What was the level of the RDM knowledge before you took the course?
- How and how much do you think the course helped you to organize the data management of your PhD study? Can you please recall an event/ episode when you thought that what you were learning was crucially important? Wakeup call

PART C - Experience after the course

- Can you describe what you have learned from this course? What would be different in your research if you didn't take the course?
- Were you in touch with the course educators or the data steward at your faculty after the RDM 101 course? Do you remember for which reason?
- [if they took the course online]: how would you describe the online course experience?

- [if they took the course in person]: how would you describe the in-person course experience
- Do you think you will use the knowledge acquired during the course in the future? If so, how?
- Do you have any recommendations/ advice, or feedback for trainers of the course? On the design of the course?

5.3 Methodology & Software

We used a qualitative descriptive study design, informed by Braun and Clarke (2006). For the thematic analysis framework and reporting the Consolidated Criteria for Reporting Qualitative Research (COREQ) (Tong, Sainsbury, & Craig, 2007) was used.

Interviews were conducted with 11 PhD candidates who had completed the course 12 to 18 months earlier and two data stewards from different faculties at TU Delft. Participants represented diverse disciplinary backgrounds and worked with various data types, including personal and sensitive data, source code, large-scale simulations, and large language models (LLMs). This variation reflects the interdisciplinary reach of the RDM 101 course and its relevance across different research contexts.

Participants were selected through purposeful sampling, a technique used to identify and recruit individuals who can provide rich, relevant, and diverse information. This approach ensured variation across faculties, types of research projects, and data types (e.g., personal data, code, large language models). Importantly, selection was not based on participants' performance in the RDM 101 course. All participants were recruited voluntarily and provided informed consent. Participants were not remunerated for their time. We pseudo-anonymised data collected from participants to protect their identities and ensure confidentiality, particularly because direct quotes are used throughout the report. Throughout the report, we removed or generalised any identifiable information to ensure that participants could not be recognised based on their words, faculty, or project context.

We conducted the data analysis using TU Delft–licensed Atlas.ti (version 23), a commercial qualitative data analysis software used to facilitate coding and organise qualitative data.

Data was analysed thematically using an inductive approach. Coding followed a constant comparative method, a process of continuously comparing new data with existing codes to refine patterns and categories. Participants' own words were used to guide code development, strengthening the credibility of the analysis.

Two principal investigators (CS and GK) independently reviewed and manually coded the transcripts to enhance familiarity with the data and to strengthen the trustworthiness of the findings. A third researcher (NR) joined later to review sections and contribute to refining the themes. Discrepancies in coding were resolved through discussion among the team until consensus was reached. Initially, we analysed the responses from PhD candidates and data stewards separately; however, as the analysis progressed, both commonalities and differences across the two groups were examined. We identified six main themes that correspond closely to the research goals and objectives and the participants' experiences.

Reflexivity was applied throughout the process to remain aware of researcher influence and to prioritise participant voices. Trustworthiness of the findings was supported through triangulation (multiple coders), rich description of context (the impact of RDM 101 over time), and a transparent audit trail from protocol to reporting.

Due to ethical and privacy considerations, we do not share verbatim transcripts or detailed coding linked to individual participants, in order to maintain confidentiality. The interview protocol, a blank version of the participant information sheet and consent form, and documentation explaining the methodology are openly available via the 4TU.ResearchData repository (Rzayeva et al., 2025).

5.4 Results

In the analysis we followed an inductive approach. This allowed themes to emerge directly from the participants' narratives. This method ensured that the findings remained grounded in their lived experiences and perspectives. The findings are organized into four key themes:

- **Motivations and Expectations** – exploring why participants enrolled in the course and how well it aligned with their needs.
- **Knowledge Acquisition and Impact** – examining what participants learned and how it influenced their data management behaviours.
- **Course Experience and Feedback** – capturing participants' impressions of the course structure, delivery, and learning tools.
- **Emotional and Social Dimensions** – highlighting the role of emotions and peer interactions in shaping the learning experience.

We believe that, taken together, these themes offer a deeper understanding of the course's effectiveness from the participants' perspectives and highlight areas for potential improvement.

Theme 1: Meeting Motivations and Expectations and Navigating Challenges

Participants enrolled in RDM 101 with a wide range of motivations, shaped by both institutional requirements and personal interests. These motivations generally fell into three categories:

- **External Motivators:** Many participants attended the course to fulfill Graduate School requirements, such as earning credits or completing a Data Management Plan. For some, this was perceived as a bureaucratic necessity. As participant C noted, *“At first, it felt like a box-ticking exercise; I needed to complete my Data Management Plan.”*
- **Internal Motivators:** Others were driven by a genuine desire and aspiration to improve their research practices. Participant A shared, *“I was keen to learn how to make my data FAIR and reusable in the future.”* Similarly, participant M explained, *“It's something that I wanted to learn early on in the research process because I know that data management is important at every stage, so I enrolled.”*
- **Practical Needs:** Several participants were already working with large or complex datasets and needed immediate, applicable strategies. Participant B, for example, reflected: *“As soon as I started, I had to do a lot of simulation and there was a lot of data. I was already missing some simulations, some codes, so I said, ‘OK, it's better to follow a course and see what they suggest and how to organize the data.’”* Participant E added *“One important aspect is how to work with companies. If you focus only on data management that can be done at TU Delft, but don't talk about security protocols or how to ensure data is secure; or working with commercially constrained data, that's a gap.”*

This diversity of motivations shaped how PhD candidates engaged with the RDM 101 course. Some approached it with minimal expectations, while others hoped for a transformative learning experience. Data stewards also reflected on this range of engagement among PhD candidates, noting, *“We see the course as both a necessity and an opportunity”* [Participant E].

Overall, participants appreciated the course's broad and flexible design, which offered both foundational knowledge and practical tools. However, a notable difference in expectations emerged between those seeking valuing general foundational knowledge and those looking for discipline-specific data management practices and guidance. Particularly participants from the social sciences and humanities expressed a desire for more tailored content. One participant, who worked with sensitive survey and interview data, shared: *“As a PhD student, you have no colleagues to rely on. I*

thought this course would teach me everything I need to know to be compliant [with ethical and privacy requirements]" [Participant N].

Despite receiving individual feedback on assignments, some participants found it unclear whether the course addressed social science data needs or focused primarily on technical software domains. And while trainers were described as helpful and motivated some social science-specific questions were often redirected to faculty data stewards. As one participant noted, *"Some questions they could answer and some they could not. I guess they were also more from a technical background"* [Participant C].

This feedback highlights a key tension: while RDM 101 is designed as a general introduction to research data management, it cannot fully meet the specific needs of every discipline within its limited scope. Nevertheless, several participants found value in the interdisciplinary nature of the course, which encouraged reflection on their own practices through exposure to diverse research contexts. One participant remarked, *"Even though it wasn't about my specific field, hearing about other disciplines' data challenges made me rethink my own practices"* [Participant H].

Importantly, this broader perspective seemed to encourage participants to continue engaging with research data management even after the course ended. Many reached out to their faculty data stewards and supervisors with follow-up questions, suggesting a lasting shift in awareness and a more proactive attitude toward managing their data. At the same time, the variation in participants' expectations highlights an ongoing challenge: balancing a general, foundational approach with the specific needs of different disciplines. This is especially relevant given that RDM 101 is designed to provide broad principles rather than specialised, discipline-specific instruction.

Theme 2: Knowledge Acquisition, Impact on Practice and Cognitive Shifts.

The RDM 101 course appeared to enhance participants' theoretical understanding and practical application of research data management principles, as reflected in their descriptions of changes in attitude and behaviour. Many reported gaining new competencies in areas such as version control (e.g., Git), data storage strategies, file naming conventions, and documentation standards like READMEs. For instance, the in-person session on data documentation tools helped participants better record experiments and their versions and ensure compliance with repository requirements like 4TU.ResearchData. Participants noted how these tools streamlined their workflow and improved data traceability. Beyond acquiring technical skills, the course also encouraged a deeper mindset shift, participants began to approach data management more thoughtfully and deliberately in their day-to-day research. One participant shared: *"Before the course, I never really thought about how I stored my data or named files. Now [after one year] I'm much more deliberate and consistent."* [Participant M]

Others noted how the course helped initiate an iterative process of revisiting data and refining methods. Participant C explained *"That kind of feedback between thinking about your data and then your methods, and then back to your data, really helped me. It was a process that was started by this course..."*

One of the most appreciated parts of the course was the introduction to the FAIR data principles (Findable, Accessible, Interoperable, and Reusable). Learning about these concepts encouraged participants to reflect critically on their own data practices. Participant A admitted, *"I definitely think it's useful because I wasn't fully aware of data management before. Even if I was struggling with poor data management from someone who worked on my part of the project before me."* Several mentioned that the frameworks and tools introduced during the course helped them save time and work more efficiently, especially when preparing to publish their research. *"My second project [I] already had a ready a folder called data management. So every time I collect more data I immediately put it the right way in there, so I know that as soon as it's finished and I want to publish, I can just basically do copy paste and put it in the for TU directory, which is really helpful because it saves you a lot of time if you have to do that last minute - when you find out that that is required."* [Participant N]

Some even described a broader shift in how they saw themselves as researchers. One participant said: *“I think it opened my eyes to the things you should be aware of when you're trying to involve yourself with these kinds of technologies and trying to become the best sort of data scientist that is inside of you.”* [Participant H]

Data stewards noticed this shift too. They reported that participants who attended the RDM 101 course often returned with questions months after the course, showing they were actively applying what they had learned, especially when working on their Data Management Plans. As participant E noted, *“They remember the course content and often come back with questions”*.

The timing of the course within the PhD journey came up repeatedly. Many participants felt that taking RDM 101 earlier would have helped them avoid mistakes and save time. As participant L. put it, *“I wish I had this course earlier; I made so many mistakes.”* Participant O echoed this: *“If this was before data collection started, it would have saved time.”* One participant even suggested making the course mandatory: *“You will have to do this. So this is the fastest way to do it right. And the decision to take the course should not be left solely to the student.”* [Participant N].

This early exposure to RDM not only benefits students but also eases the workload for data stewards, who otherwise need to spend more time reviewing and correcting DMPs. As participant E observed, *“The main point is the awareness of preparing the data management plan. I received questions from people who did the RDM 101 course and then, sometime after a question arises about data management and they contact me.”*

In the end, participants saw RDM 101 as more than just a skills course. It helped them build better habits, think more critically about their research, and feel more prepared for the data challenges of their PhD journey.

Theme 3: Feedback and General Impressions of the RDM 101 Course

Participants emphasized that the course effectively helped them initiate data management practices by providing a comprehensive overview of key skills through practical learning, group assignments, peer reviews, personalised feedback, and a structured methodological approach. Many valued the in-person format for enabling immediate feedback and stimulating peer discussions, which enriched the learning experience.

One participant reflected: *“I really liked that there was a balance between a part of the class that was a bit more theoretical from what I remember a bit more. About sharing insights and explaining what the activities were going to be about. And then a really big part of the class was this type of group work or even independent work on, yeah, filling out our own plan [data flow map] and thinking about the type of information that we'd be needing.”* [Participant M]

The delivery of the course was widely praised, with participants highlighting the practical activities and interactive elements as key factors in maintaining engagement and aiding knowledge retention. The LEGO game, in particular, was frequently mentioned as an effective tool for raising awareness around data documentation practices: *“The LEGO game made me realise how unclear my documentation was. It was a simple but powerful exercise.”* [Participant N]

Another crucial learning tool noted by several participants was the data flow map. This tool, developed, customised and used by the training team, helped them visualize their data's journey through the research process, including decisions about storage, backup, metadata, and data documentation. Participants also appreciated the thoughtfully designed assignments and the flexibility of using various templates Excel, Miro, and PowerPoint, which allowed them to choose formats that suited their working styles. While some initially questioned its relevance, especially those with simpler data structures, participants mentioned that later recognised its value in clarifying workflows and supporting data management plan preparation. For example, one participant reflected: *“It was quite elaborate. I thought especially during some exercises, you were like “for my type of data, do I have to fill out this form?” Because it seems a bit silly and it is only a few CSV files.”* [Participant D].

Eventually, the same participant later acknowledged its usefulness stating that *“...But now, just the fact, knowing that I did the data documentation at some point just helps when I publish stuff. Maybe on GitHub to just quickly go back without having to think about, “where should I publish it? and what should I think about?” It’s a nice thing to fall back on because I thought about it back then, and it’s been nicely written down”* [Participant D]

Peer learning and dialogue among them were consistently highlighted as valuable components of the course. The opportunity to discuss data challenges with peers from diverse disciplines was seen as enriching and illuminating, fostering a deeper understanding of data management complexities.

While all interviewed participants attended the in-person version of RDM 101, several reflected on the potential benefits and limitations of different delivery formats. In-person sessions were generally preferred for their structure, immediacy, and opportunities for deeper engagement. As participant F shared, *“I like being able to ask follow-up questions and go a bit more in-depth, maybe also questions about the ethics, like the philosophy behind it. That’s why I like in-person more.”* Others noted that being physically present helped them stay focused: *“In-person you are kind of obliged to follow and listen. But online, I usually put the meeting on one screen and do other things on another.”* [Participant C]

At the same time, participants acknowledged that online or hybrid formats could increase accessibility, especially for external PhD candidates or those seeking a basic introduction. Offering flexible options was seen as a way to better accommodate different learning needs and schedules.

In addition to the course content and learning activities, the role of the trainers emerged as a key factor in shaping participants’ overall experience. Participants, in fact, consistently described the trainers as approachable and supportive. The presence of multiple trainers during in-person sessions was seen as a strength of the in-person class as this enabled more personalised feedback and facilitated effective group work. The approach of the trainers and teaching assistant was seen as friendly and responsive, which contributed to enhanced learner support and motivation. One participant noted: *“The trainers and assistants were very friendly and helpful. They were open to answer all the questions and follow-up.”* [Participant A]

Another shared: *“I really liked having that balance and also during the independent work, it was really nice to know that the educators were around to ask them questions for whatever help we needed”* [Participant F]. Finally, *“I had a good feeling about the course... I don’t use it daily, but it’s a good thing to fall back on. And I think when the need is there, you have the course instructors, but also the faculty data stewards that are quite easily accessible.”* [Participant G]

Theme 4: Emotional and Social Dimensions of RDM Learning

An unexpected but important theme that came up in the interviews was how participants felt during the course and how their social experiences affected their learning. While the course mainly focused on teaching practical skills, many participants shared that their emotions and interactions with others played a big role in how they learned and stayed engaged.

Some participants said they felt nervous or overwhelmed at the beginning of the course, especially when they thought about how hard it might be to manage large or sensitive datasets. For them, research data management had seemed confusing or like just another rule they had to follow. *“Data is quite complicated...we have to deal with two institutional and legislative contexts. I had to put together a data management plan before attending the course, which was very challenging for me.”* [Participant D]

The course helped clear up these worries by giving them clear steps and guidance, which made them feel more confident and less stressed. *“It [RDM 101] gave me fundamental knowledge about managing data during my PhD, writing a data management plan, using different backup tools, and uploading datasets to repositories.”* [Participant B]

As their confidence grew, many participants said they started to feel more in control of their data. Instead of seeing data management as something they had to do just to meet rules from funders or institutions, they began to see it as a useful and important part of their research. One participant said: *“The course made me realise that it [RDM] isn’t just ticking boxes for compliance. It’s about owning my research data and making sure it’s organised and accessible not just for others, but for myself too.”* [Participant A]

This feeling of ownership was often linked to hands-on tools they learned in the course, such as the data flow map. These tools helped them feel more organised and in charge of their work. *“The most useful thing I would say was really the data flow map and starting with the implementation of GIT for my coding. I am glad because it helped me.”* [Participant F]

Another big part of the course experience was the sense of community. Many participants said they enjoyed learning with others, especially during in-person sessions. Talking with trainers, teaching assistants, and other peers PhD candidates helped them feel less alone in their work. They liked being able to share problems, hear how others handled data, and get new ideas. Participants also expressed how group experiences made learning more enjoyable and helped build a support network that many found very helpful. As one participant said: *“That was interesting for me to learn about other people’s work... it’s a nice moment to get out of your bubble.”* [Participant F]

These responses indicate that emotional reactions and social interactions were commonly reported positive aspects of participants’ experiences with the RDM 101 course.

6 Main Highlights of the Study

This study aimed to assess both the short-term and long-term impacts of the RDM 101 course on participants' knowledge, confidence, and research data management practices. Findings from Graduate School feedback forms, pre- and post-training surveys, and interviews **consistently indicate that the course has had a positive influence, while also revealing areas for further refinement.**

6.1 Short-Term Impact Summary

6.1.1 Course Perception and Engagement

Feedback from the Graduate School evaluations showed generally high satisfaction with the course content, trainers, and delivery. However, **variations in participant responses, particularly in 2023 and 2024, coincide with significant course changes**, such as expansion of the training team, revised materials, and dual delivery formats (online and in-person). These changes may explain the slightly lower average scores and higher standard deviation in ratings. At the same time, data suggests **a positive trend in participant engagement, particularly in the most recent sessions**, indicating that the new design and content updates may be yielding benefits.

Participants also rated the trainers highly across all years, particularly **in creating a supportive learning environment and delivering structured content**. Slight **declines in some ratings during the transition period (2023–2024)** suggest the need for ongoing **pedagogical development and feedback process enhancement**.

The analysis results strongly indicate that participants perceive the skills gained in the course as relevant and applicable. Specifically, **93% of participants agreed or strongly agreed that the skills provided were relevant to their PhD project or trajectory, and 87% found them valuable for their future careers**. We find this finding very valuable, as it indicates that participants who took the course recognize the course's practical importance not only in the immediate context of doctoral research but also in supporting long-term professional development.

6.1.2 Feedback-Driven Improvements

Participants welcomed the course's structure, especially its practical components and trainer feedback/guidance. However, some recurring concerns were:

- **Workload:** Many participants felt the course workload was too high relative to the number of credits. Adjusting credit allocation or redistributing assignments could help.
Discipline-Specific Content: Requests for more tailored examples and tools indicate a need for modular or domain-specific case studies.
Delivery Format Preferences: Preferences varied, but smaller class sizes and more structured discussions were common suggestions for improvement.

6.1.3 Training Effectiveness and Shifts in RDM Practices

One of the most significant outcomes of the survey analysis can be observed in **shifts in RDM awareness and practices** among participants, clearly demonstrating the impact of the course.

- **Awareness of Institutional Support:** Prior to training, **only 45% of participants were aware of TU Delft's RDM support structures**. After the course, **awareness rose dramatically to 98%**, including substantial increases in recognition of dedicated roles and tools, including data stewards, institutional storage locations, library and IT services.
- **Changes in Storage and Backup Practices:** **90% of participants changed the storage strategy for research data, and 81% established a backup strategy during or after the training**. The responses evidence that after the course participants changed insecure storage options to recommended institutional solutions. **Use of the Project Data (U:) drive for research data storage rose from ~20 to ~50 responses; for backup, from ~10 to ~40**. Use of Git platforms remained stable, indicating prior awareness, while **commercial cloud usage nearly disappeared**.
- **Confidence and Understanding of RDM concepts and practices:** Survey responses showed substantial increases in participants' self-reported **confidence in RDM skills (from ~2.4 to 3.8)** and **familiarity with their responsibilities as PhD candidates (~2.1 to 3.8)**. We value these shifts as an indicator that participants leave the training not only more informed, but also **more capable and motivated to take ownership of RDM-related tasks**. The most dramatic improvement was in **familiarity with the FAIR principles (~1.9 to 4.0)**. Additionally, post-training word clouds about FAIR keywords shifted from general terms to more practical, specific concepts like "metadata", "documentation" and "README", suggesting **a move from conceptual to applied understanding**. Also, nearly all participants reported **implementing or planning specific FAIR-related actions** in their research projects after the training. The thematic distribution of responses on FAIR measures to be taken in relation to their own research further illustrates how learners are capable to **translate the course knowledge into actionable changes in their own projects**.
- **Impact on Data Management Planning:** After the course, **98% of participants agreed that the training helped them start or improve their DMPs**. The Data Flow Map creation assignment, which aligns with DMP sections and is supported by continuous trainer feedback, was especially valued. Participants appreciated practical tools, templates, and step-by-step guidance in DMP development/enhancement, reinforcing the course's hands-on design.
- **Data Publication and Repository Choice:** The proportion of participants intending to **publish data/code increased from 76% to 95% post-training**. There was a marked shift toward institutional and trusted repositories like 4TU.ResearchData and Zenodo. Confusion about inappropriate repositories (e.g., TU Delft Repository or Project U: drive for publication) declined but still appeared in some responses, highlighting an area for clarification.
- **Addressing Specific Challenges:** **74% of participants reported that the training helped them resolve real research problems related to data/code management**. Many pointed to better organization, documentation and application of RDM best practices, indicating translation of course content into applied research support.

6.1.4 Online vs In-Person Delivery

While both delivery modalities proved effective, participants in the in-person runs reported that **in-class activities and interactions with peers and trainers were particularly beneficial to their learning experience**. Preference of in-person classes over online was also reflected in the open responses to Graduate School feedback forms. However, **no significant differences were**

observed between the two formats in terms of participants' engagement levels or acquisition of knowledge and skills.

6.2 Long-Term Impact Summary

6.2.1 Motivations and Expectations

Participants joined RDM 101 for a variety of reasons, including **fulfilling Graduate School requirements, improving their research practices, or addressing immediate data challenges**. These motivations influenced how they engaged with the course. While some viewed it as a formal requirement, others hoped for a more **transformative learning experience**. The course's **flexible design** allowed it to meet many of these needs, though participants from social sciences expressed a desire for **more discipline-specific content and support**.

6.2.2 Knowledge Acquisition and Impact on Practice

The course effectively **strengthened both theoretical understanding and practical skills** in research data management in the long-term. Participants who took the course up to 18 months ago reported **improvements in RDM areas such as version control, documentation, and data organization**. The introduction to FAIR principles was particularly impactful, prompting **critical reflection and long-term changes in data practices**. Many participants described a **shift in mindset from seeing RDM as a compliance task to recognizing it as an integral part of their research process**.

6.2.3 Course Design and Delivery

Participants praised the course's structure, especially its **balance of theory and hands-on activities**. Tools like the **LEGO game and data flow map** were frequently cited as helpful for understanding **documentation and workflow planning**. **Group work, peer and trainers feedback, and the presence of multiple trainers contributed to a supportive and engaging learning environment**. While all interviewed participants attended the in-person format, several expressed interest in a **hybrid or fully online version to increase accessibility**.

6.2.4 Emotional and Social Dimensions

An important and unexpected finding was **the emotional and social impact of the course**. Some participants initially felt overwhelmed by RDM concepts, especially when dealing with complex or sensitive data. However, the course helped **reduce anxiety by providing clear guidance and practical tools**. **Social interactions, both with peers and trainers**, played a key role in building long-term confidence and motivation. Participants appreciated **the sense of community and the opportunity to learn from others**, which helped them feel **less isolated in their research journey**.

7 Conclusion

Overall, the results of this study demonstrate that RDM 101 is achieving its core objectives - **improving participants' knowledge, confidence with, and application of best practices in research data management**. The study evidences that the course **not only introduces foundational theoretical concepts but also equips participants with practical knowledge and skills that can be directly applied to their research workflows**. We also observed a lasting shift in mindset, with many participants coming to see RDM as an integral part of the research process.

The report also provides concrete evidence of the importance of **investing resources and having an expert training team** at the library to design and deliver the course (and other courses) in **an interactive and practical way** to make a **long-term impact on the integration of RDM practices in research at TU Delft**. Because of the abstract nature of concepts like RDM, FAIR data, and FAIR software, there is a need **to provide opportunities for participants to learn from each other, receive guided feedback, and actively share challenges and solutions**. This approach helps **translate abstract concepts into practical relevance** for participants' research, motivating them to **apply what they have learned**.

Additionally, having sufficient resources within the training team allows us to dedicate time **for evaluating our courses in detail and updating them in a timely manner**, which is essential for fast-evolving topics such as RDM, FAIR data and software, and **to meet the needs of an evolving community of learners**.

At the same time, there is room for further refinement, especially in **managing the course workload and credit balance, implementing a more modular, and discipline-tailored learning paths within the course, and adopting a customizable format**.

The report also provides input for thinking about **the format of course delivery**. The feedback received through the Graduate School feedback forms showed that **in-person delivery was generally preferred by participants**, while during the interviews, participants acknowledged that formats such as **online classes and a completely self-paced format could increase accessibility**. The results of this report confirm that **a blended format of the course is the best way to maintain the short-term and long-term impact** of the course. However, as a team, we acknowledge that further exploration is needed to improve the blended format with online classes through **improved session design, delivery strategies, and the use of interactive tools**.

The findings of this study provide **strong support for the current design of RDM 101 while highlighting promising directions for future development of the course**. We think that the results of this study, along with the ongoing analysis of new feedback data and continued collaboration with stakeholders, such as researchers, data stewards, and faculties, will provide **valuable insights for future updates to the course, ensuring it remains relevant and effective**. With this, we believe that **newly structured and well-supported RDM 101 training will enhance its role as means of fostering Open Science and responsible research practices at TU Delft**.

Data and Code Availability

Due to ethical and privacy considerations, the raw research data containing potentially confidential and identifiable personal information is securely stored on the university-curated network storage, the Project Data (U:) drive, and are available upon request. Researchers can request access to the raw data by contacting the TU Delft Library Research Data Services Training Team at rdmtraining-lib@tudelft.nl.

To support transparency and reproducibility of the study, the anonymized quantitative datasets and the high-level qualitative categories, together with their conceptual descriptions and containing no personally identifiable information about participants, are published in the 4TU.ResearchData repository (Rzayeva et al., 2025). These materials include:




- Pre- and post-training survey questionnaires and opening statement
- Cleaned pre- and post-training qualitative survey data
- Codebooks for qualitative survey data
- Cleaned GS feedback form data
- Codebooks for qualitative GS feedback data
- Interview protocol form
- Interview methodology
- The participant information sheet and consent form
- Code group data

The scripts used for the analysis are openly available on 4TUResearch Data repository (Grens, 2025) integrated with public GitHub repository to ensure reproducibility of analysis.




References

1. Ahlers, I., Andrews, H., Cruz, M., van Dijck, J., Dintzner, N., Dunning, A., Eggermont, R., den Heijer, K., Ilamparuthi, S., van der Kruyk, M., van der Kuil, A., Kurapati, S., Love, J., Martinez Lavanchy, P., Plomp-Petersen, E., de Smaele, M., Teperek, M., Turkyilmaz-van der Velden, Y., Versteeg, A., & Wang, Y. (2020). *TU Delft Research Data Framework Policy* (V2). Zenodo. <https://doi.org/10.5281/zenodo.4088123>
2. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
3. Delft University of Technology. (n.d.). *The PhD development cycle*. <https://www.tudelft.nl/en/ae/research/ae-graduate-school/the-phd-development-cycle>
4. Grens, N.(2025). *Analysis Code Supporting "RDM 101 Impact Evaluation: A Study on Training Effectiveness and Research Practice Changes"*. 4TU.ResearchData. <https://doi.org/10.4121/f7e9b088-cf91-464d-a6a9-772fde5d8bff>
5. Martinez-Lavanchy, P. M., Rzaeva, N., Dauwels, S., Barrera-Coronel, M., Utrilla Guerrero, C., Dace, H., Strubbia, C., van Schöll, P., & Zormpa, E. (2024). *TU Delft Research Data Management 101 course* (Version v2) [Lesson]. Zenodo. <https://doi.org/10.5281/zenodo.10732095>
6. Rzaeva, N., Morselli, F., Strubbia, C., Grens, N., Kulkarni, G., & Martinez-Lavanchy, P. M. (2025). *Data supporting "RDM 101 impact evaluation: A study on training effectiveness and research practice changes"* [Data set]. 4TU.ResearchData. <https://doi.org/10.4121/90a1d562-8496-44f3-b8a5-5610ddc871d5>
7. Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, 19(6), 349–357. <https://doi.org/10.1093/intqhc/mzm042>

Acknowledgments

 	<p>We would like to express our gratitude to Selin Kubilay and Elviss Dvinskis (Data Managers at the Digital Competence Center - DCC), for their great support, expertise and enthusiasm. Their guidance in ensuring that the code for analysing the data were developed according to best practices and properly documented was essential to ensure reusability and reproducibility of our analysis.</p>
	<p>Our other big gratitude is to Nicolas Dintzner (TPM Data steward) for his invaluable support throughout the entire project. His expertise was crucial in developing the Data Management Plan for this project, preparing the opening statement for surveys and informed consent for interviews, and navigating the HREC application process. We are very thankful for his flexibility and guidance in addressing several RDM tasks during the study.</p>
	<p>A big thank you goes to Judith Linnemann-Karel (Central Graduate School) who took the time to provide us with the feedback forms data in the format we needed for this study.</p>

Author Contributions

	<p>Narmin Rzayeva (NR) (RDS trainer) was the project lead of the RDM 101 Impact study. She oversaw the design of the project plan, and also kept the coordination of all the activities of the different contributors. She contributed to the design of the methodology with a particular focus on quantitative aspects of the research, designing and analysing pre- and post-training surveys, as well as supporting the analysis of the Graduate School feedback forms.</p>
	<p>Francesca Morselli (FM) (Postdoc VU Amsterdam) contributed to this investigation with a particular focus on the qualitative research component, supporting the analysis of the theoretical framework, the design of the methodology, as well as the development of interview protocols, conducting interviews and designing of pre- and post-training surveys.</p>
	<p>Carla Strubbia (CS) (RDS trainer) contributed to the project as a research data management training expert. She was responsible for guiding the analysis of qualitative data and shaping the overall evaluation approach. Her experience in research training, qualitative methodologies and commitment to improving data practices ensured that the study was both rigorous and impactful.</p>
	<p>Nikki Grens (NG) (RDS student assistant) contributed to the project as a student assistant from its start. She was responsible for preprocessing the data, developing the codes in R, generating plots, and documenting the analysis. Her prior knowledge, combined with her enthusiasm and commitment to learning, enabled the implementation of best practices acquired during the project, ensuring a smooth and efficient analysis process.</p>



Gargi Kulkarni (GK) (RDS student assistant) contributed to the project as a student assistant to support us with the analysis of the interviews. She was part of the team working on the qualitative part of the study. Her curiosity and commitment to learning supported the implementation of thoughtful and rigorous analysis throughout the research project.



Paula Martinez Lavanchy (PML) (RDS training coordinator) is the initiator and facilitator of this research study. She also contributed to the development of the methodology for quantitative data analysis and provided subject-matter expertise for validating the qualitative results.

