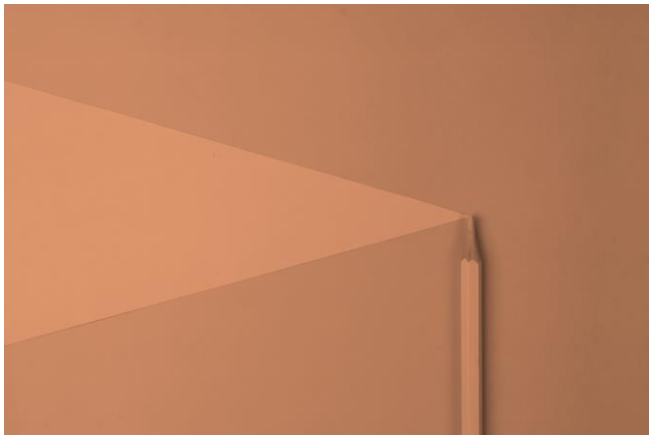

Survey of the Finnish research data management training 2020 and 2021

Päivi Kanerva (University of Turku), Hannele Keckman-Koivuniemi (Finnish Social Science Data Archive), Tanja Lindholm (University of Helsinki), Laura Mure (Aalto University), Maria Söderholm (Finnish Environment Institute).

Contents

The aim of the study 2
Introduction 2
RDM practises in Finland 3
Methodology 4
Survey 4
Categorisation of the training activities 4
Format of the training 5
Results 6
Cross-cutting perspective on data 6
Organizations and types 6
Overview of training activities 7
Starting year and end year 12
Detailed analysis of the training offerings 13
General 13
Plan 14
Document 15
Store 17
Protect 17
Preserve 19
Open and re-use 19
Discussion 21
Limitations 22
Conclusions 23
Abbreviations 25
References 26

The aim of the study

Our aim in this study was to gain a deeper understanding of what kind of research data management (RDM) training activities are provided by different Finnish organizations, in particular, we tried to get the answers to the following questions:

- What type of research data management training organizations provide?
- Are training offerings formal courses which are integrated into study programmes or more informal in nature?
- Are there any credits awarded for completion?
- Which RDM topics are covered and are there any gaps in the training offerings?
- Are training offerings offered regularly?
- Are training offerings provided for different levels of learners, e.g., beginner vs. professional or undergraduate vs. researcher level?
- Which organizational units and members of the staff are involved at the institutional level?

This report is intended to give an insight of the current situation of RDM training offerings for the organizations that participated in the survey. The report helps organizations to plan their future training and helps to develop their teaching and training initiatives in RDM as well as provide comprehensive base for those yet in a planning phase.

Introduction

The importance of research data management has grown substantially over the last decades. The reasons are obvious. Due to the digital revolution, the amount of data has risen rapidly (Helbing, 2015), it is easier to share, store and re-use the data. At the same time the RDM responsibilities and duties have risen as well. Data needs to be well organized, maintained, curated, stored, and documented (just to mention a few) so that it can be accessed and used by other researchers or the general public.

One big game changer in demand for responsible RDM is funder demands. Nowadays most funders, (e.g. European Commission https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/agr-contr/general-mga_horizon-euratom_en.pdf p. 111) require data management plans (DMPs), which are formal documents that specify how research data is handled during and after a research project. Funders see the research data as one of the most important and valuable outputs of publicly funded research (https://www.aka.fi/en/research-funding/apply-for-funding/how-to-apply-for-funding/az-index-of-application-guidelines2/data-management-plan/data-management-plan/).

The demand for responsible RDM is not just something that funders want. It is, or at least it should be, a key learning objective on the road to become a researcher. Research data management might sound like a new term, but it is nothing new, it is something that researchers around the world and throughout the time have done. Nowadays it has just become a necessity because of the amount of data growing rapidly. Not only is the total amount of data created, captured, copied, and consumed globally increasing, but also data created by a single researcher is increasing and it makes it important to tame that flood.

In the past five years' time the so-called FAIR principles (Wilkinson et al., 2016) has come along to guide the process of doing research and producing data. The aim is to "provide guidelines to improve the Findability, Accessibility, Interoperability, and Reusability of digital assets". FAIR principles guide for a better re-use of the data in the long term (Wilkinson et al., 2016).

When we look closely at the FAIR principles, it is easy to see how profound they are for responsible RDM per se. If we think for instance reproducibility in research, which has shown in many surveys to be in crises (Baker, 2016), it is obvious what kind of efforts should be concentrated on data documentation, one of the key elements of FAIR principles and responsible data management.

Detailed documentation enables research data to be understandable for the researchers themselves during and after the research project but also for others: fellow researchers, collaborators, funders, evaluators etc. The data will be in better overall shape and the risk of false interpretation of the data will decrease. (Fuchs et al, 2018, p. 3)

Knowing all the advantages of implementing responsible data management, we should not waste data any longer. Nevertheless, the question is how and who can respond to the new demands. Funder requirements and policies for DMPs and open science demand reactivity among research organizations and services built around RDM practices.

Luckily the importance of data management skills and competencies in fostering high quality research is acknowledged widely in research organizations, and they are developing different kinds of training and guidance to train their staff and researchers to fulfil the requirements of responsible science.

RDM practises in Finland

Finland has been one of the pioneers in RDM. For example, the Finnish Social Science Data Archive (FSD) started providing presentations, lectures and learning materials on research data re-use, data management, data documentation, digital long-term preservation and other FSD and Consortium of European Social Science Data Archives (CESSDA) basic services already at the beginning of 2000.

In the autumn of 2008, the Academy of Finland renewed its funding application guidelines by adding a new requirement concerning research data: "The Academy requires that applicants give an account of how the project's research material will be obtained, how it will be used and stored and how its later use will be made possible." (Kuula et al., 2008, p. 9). Since the September 2016 Academy of Finland call, it was mandatory to include a separate DMP in addition to the application. It was recommended to draft the DMP by using data management planning tool DMPTuuli (Academy of Finland, 2016, p. 51-53).

DMPTuuli was provided by the Tuuli project of the Finnish Ministry of Education, Science and Culture. The project was used to create a national network of research organizations for RDM in Finland, whose activities are a continuum.

With both DMP requirements of the Academy of Finland and the DMPTuuli project, Finnish organizations quickly began to develop their RDM services. Staff training was of great importance in order to form a comprehensive RDM service network, e.g. The University of Helsinki first started with the training of library staff and finally formed a comprehensive co-operation network between different service units, which are responsible for the university's RDM training (Koskinen, 2019).

Methodology

Survey

The survey was designed by an open data training working group of the expert panel in Open data (https://avointiede.fi/en/open-science-expert-panels/open-data) comprising of Päivi Kanerva (University of Turku), Hannele Keckman-Koivuniemi (Finnish Social Science Data Archive), Tanja Lindholm (University of Helsinki), Laura Mure (Aalto University), Taina Saarenpää (University of Lapland) and Maria Söderholm (Finnish Environment Institute).

This is mainly descriptive study due to the small sample size of the survey (n = 106). However, the aim is to provide a comprehensive picture of current RDM training opportunities in the Finnish organizations.

The survey was designed by using Google Forms. The Research Data Management Training Survey -questionnaire was sent to 74 Finnish research organizations (see Appendix 1)

The questionnaire (see Appendix 2) covered three parts and there were 17 questions in total (some of them were mandatory). The questionnaire, in particular the section on training descriptions, provided relevant information and guidance on how to answer the questions aiming to facilitate answering and produce consistent data. (e.g., Figure 1 and 2).

The questionnaire was sent via email to those persons in the organizations, who we assumed to have the best information and general view of research data management training. The web survey was open from November 26th until 12th of December 2021.

We were also encouraging organizations, which did not offer any training on RDM at the time of the survey or were just in a planning phase, to answer the questionnaire. This was important information for later comparisons and monitoring of the training offerings for coming years.

Categorisation of the training activities

Training activities were categorized based on the research data management life cycle described in Figure 1. The life cycle model is widely used and hence provided understandable classification.

The concept of training had a wide meaning in the questionnaire: it covered everything from contact teaching to digital materials, basically all the practices the organization develops and uses to guide and train their researchers and students in research data management planning, storing, documenting, sharing and opening research data.

General training covered all aspects from planning to opening the data. If the training was more specific focusing on a certain content or theme, the respondents were able to choose the given category (Figure 1). There was also a possibility to choose multiple choices. It was important to allow multiple choice option, as it is common for RDM training to cover multiple topics.

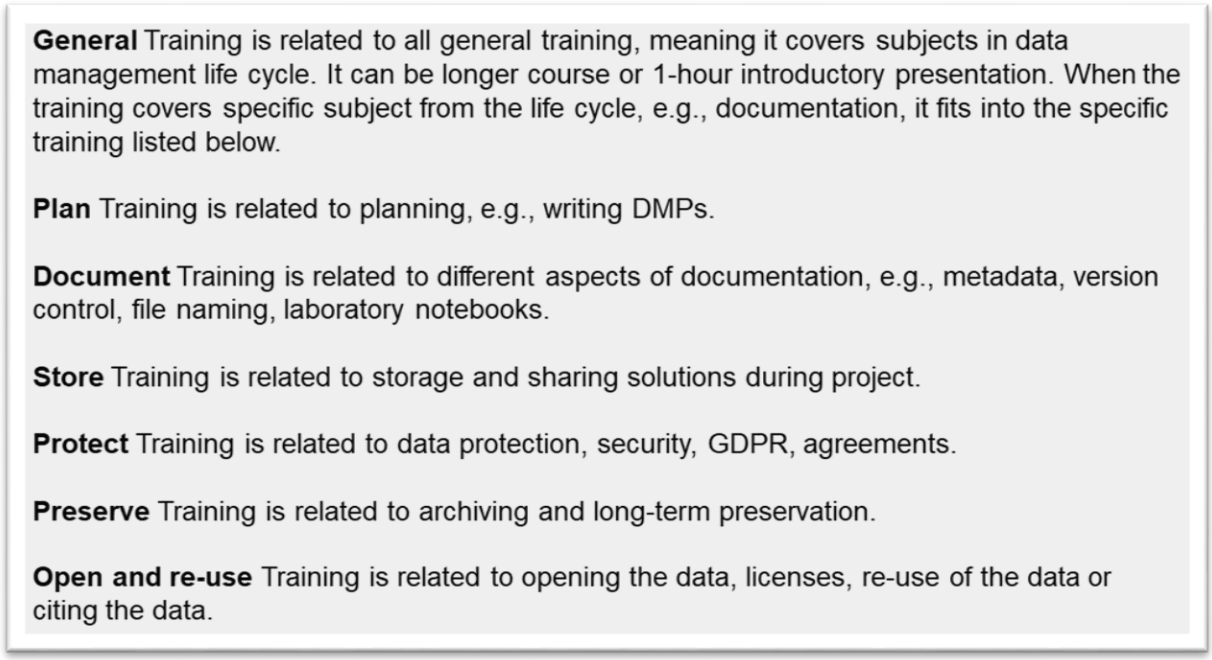

Figure 1 Categorisations of the training offerings.

Format of the training

We looked at different training classifications and based on them we modified our own classification model. The aim was to cover as much as possible the training provided by the organizations and to provide an easy-to-use tool for reporting on training provision.

The training offerings were distinguished into six different formats depending on whether credits or feedback were given, whether the training was arranged online or whether the material could be utilized in other training formats (Figure 2).

First, we wanted to make a clear distinction between **contact teaching** and **lecture format**. In contact teaching the main difference from the lecture type of training was that students would get credits and possible feedback from the training.

We also wanted to see how many organizations offered **digital online training**. Criteria for this format was that the training was organized only online, and feedback and credits were received for the performance.

In **courses, workshops** and **lectures**, the idea was that the same materials could be utilized for both remote and contact teaching, while in **webinars**, materials were specifically designed for the session in question.

We could also identify differences in modes of delivery between **webinars** or **seminars** versus **lectures** or **courses**. Webinars and seminars usually happen only once or irregularly and quite often have several speakers, whereas the independent lectures and courses run on a regular basis and the material does not change significantly.

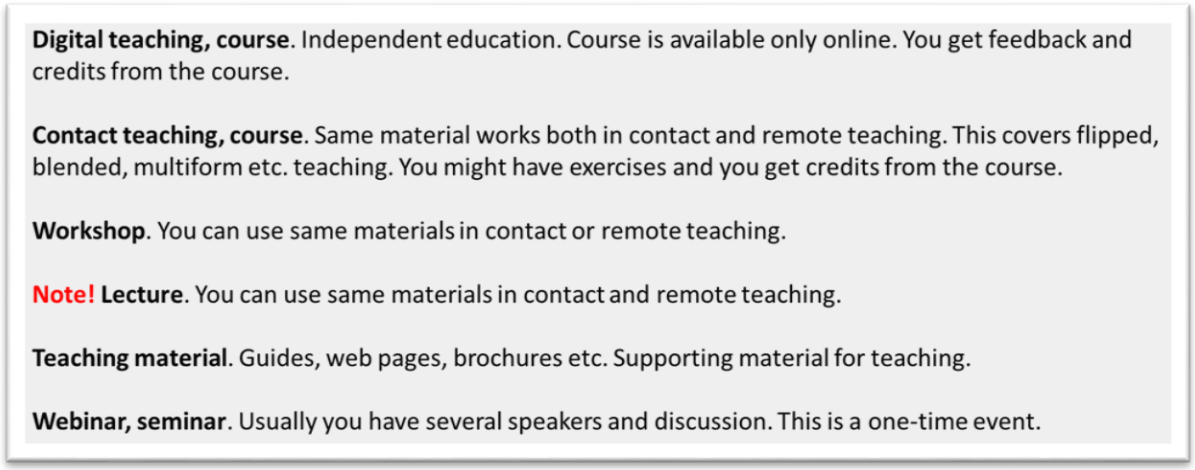

Figure 2 Training formats.

Results

In this section, we have first evaluated the results at a general level and then divided different training offerings by the topics and analysed them in more detail. The classifications of the training offerings have been made according to the respondents' specifications and, if necessary, the name or the description of the training.

There is some overlapping in the reported training sessions. Respondents categorized many of the training sessions into multiple categories indicating that the topic of the training handled one or more different aspects of research data management (e.g., plan and document).

The overlaps have been handled as follows:

- If the training was reported both in **general** and in most of the **topic-specific** categories, it was considered as **general** training according to the definition above.
- Where the **general** part of the training was just an addition to more detailed training, these training sessions are dealt in the **topic-specific** training categories.
- If the respondent chose multiple **topic-specific** training categories to the training, it has been handled in all sections reported (e.g., both in **document** and **plan**).

A broader analysis was conducted of the **general** topic category, as it was the most common and had the most comprehensive selection of training types.

Cross-cutting perspective on data

Organizations and types

Total 36 organizations responded to our data management training survey. The organizations reported 96 training offerings (see Appendix 2).

The responses were received from 5 different organization types (Table 1). The universities (n=14) and the universities of applied sciences (n=13) formed the largest group of respondents (75%). As 38 organizations did not respond to the survey, the response rate remained low.

Table 1 Organization type of the respondent organizations.

Organization type	n	%
University	14	39
University of applied sciences	13	36
Research institute	6	17
Funder	2	6
CSC	1	3
Total	**36**	**100**

Vast majority of organizations did provide training or at least training materials (91%), whereas only 5 % did not provide any training and 4 % were only just in a planning phase.

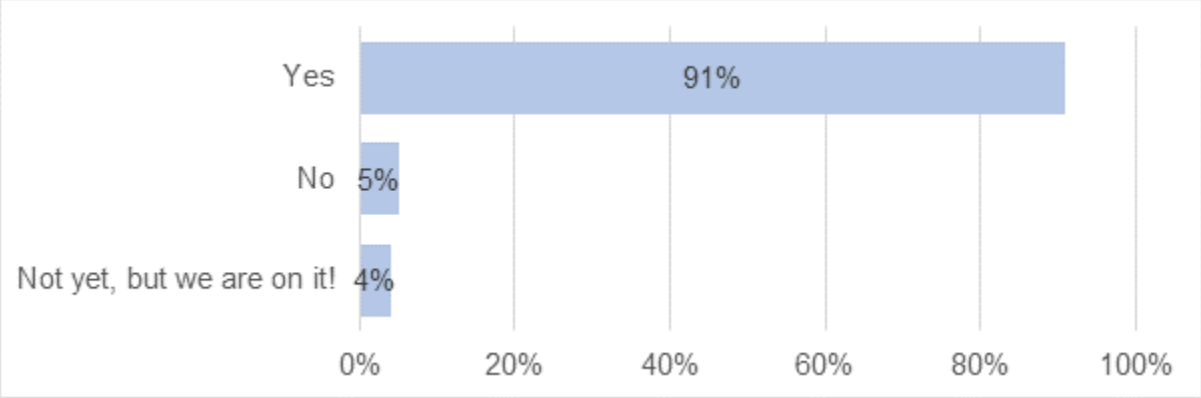

Figure 3 Does your organization provide training or materials? (%)

Universities of Helsinki (UH) and Tampere (TAU) mentioned 14 different training offerings, University of Turku (UTU), and Aalto University (AU) 13, Åbo Akademi University (ÅA) 7, and University of Eastern Finland (UEF) 5 offerings and University of Lapland (ULA) 4 offerings. Finnish Social Science Data Archive (FSD) offered 4 training sessions, CSC – IT Center for Science (CSC) 3 and Language Bank of Finland (FIN-CLARIN) 2. Other organizations mentioned one training.

Overview of training activities

Half of the training sessions focused on **general** level training that covered all aspects from planning to opening the data (n=59) (Table 2). **Topic-specific** training was provided to a similar extent (**plan** and **protect** n=14, **document** and **open** n=11) except **preserve** (n=5) and **store** (n=7).

The majority of training sessions were organized by universities (n=103). Moreover, almost all **topic-specific** training sessions were offered by universities (n=47).

Table 2 Training types provided by respondent organizations.

Topic	n	Universities	Other	Funder	Research institutes	Uni of Applied sciences
General	59	46	1		4	7
Document	11	10	2			
Open	11	10	2			
Plan	14	13		1		
Preserve	5	4	1			
Protect	14	14				
Store	7	6	1			
Total	121	103	7	1	4	7

Primary service providers were libraries (61%) and research services (26%). Another significant service providers mentioned: IT services (11%), data support units (14%), legal (7%) and data protection (8%) units (Figure 5).

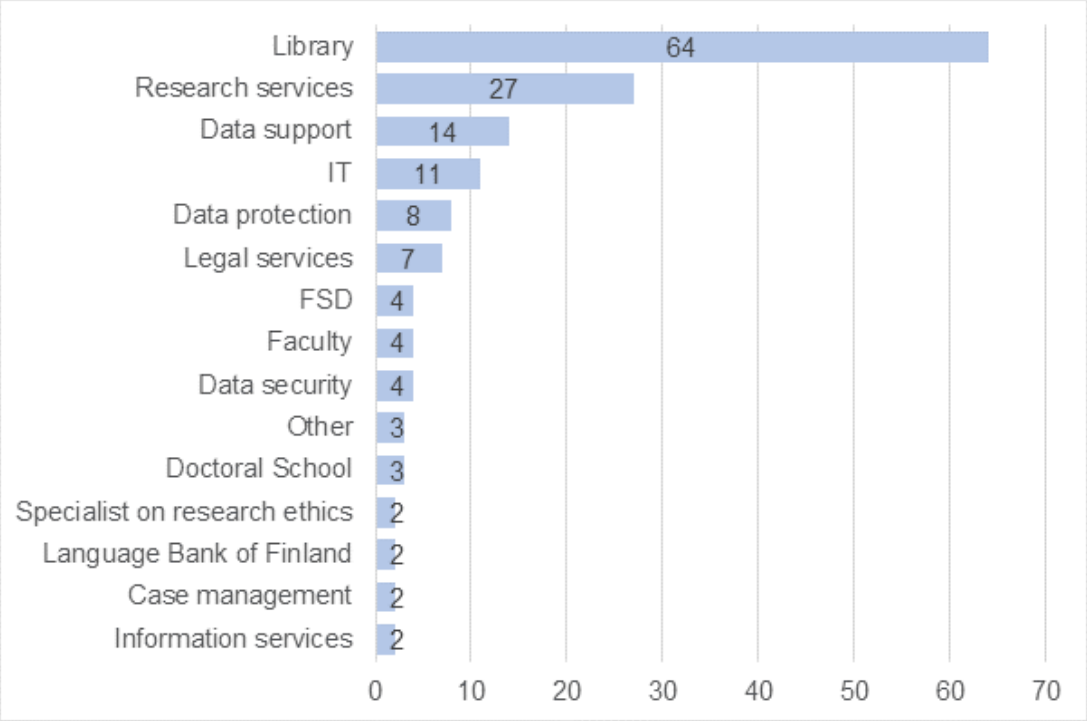

Figure 4 Primary service providers organizing the training offerings (n=96). Note: On average 1,5 *service providers at the same training.*

Most of the training offerings were aimed for researchers (71) and doctoral students or equivalent level (55), where only low number of training offerings were offered for other personnel (26), master's (or equivalent) (20), RDM professionals (12) and bachelor's (or equivalent) (12) (Figure 6).

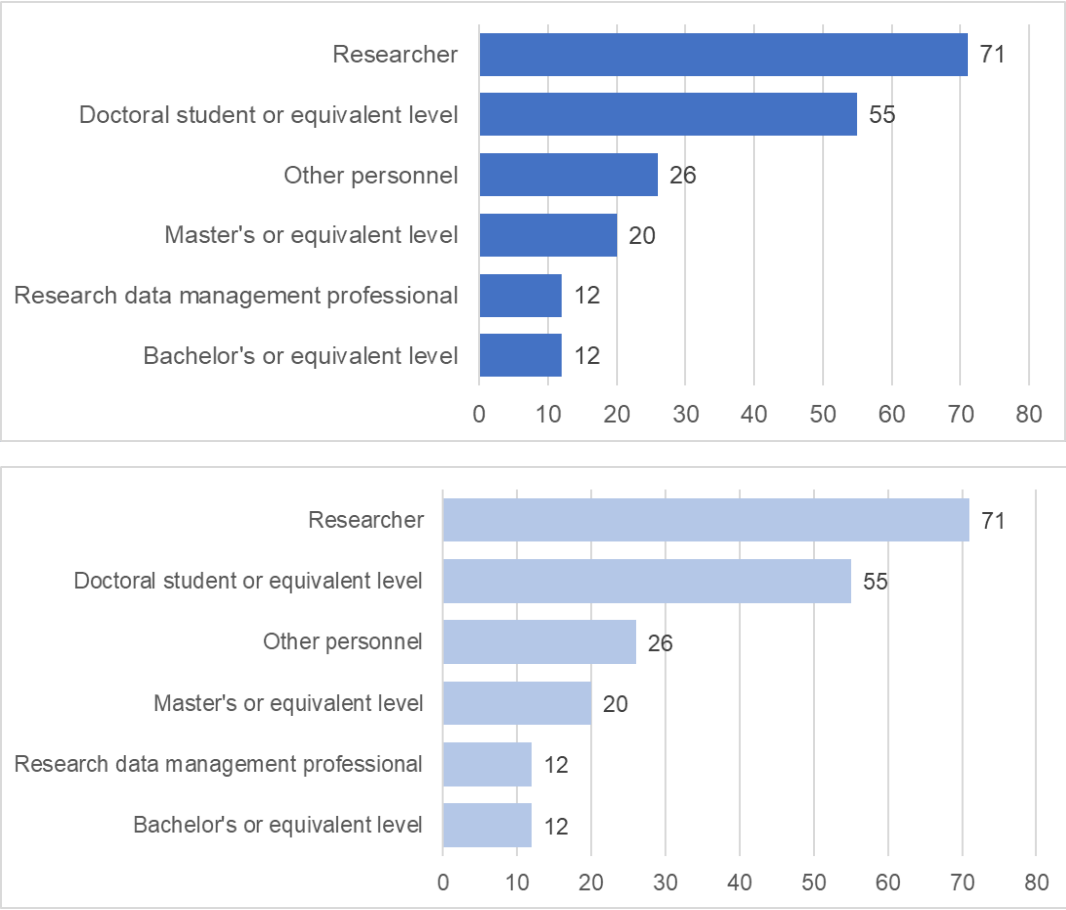

Figure 5 Educational level of the primary target group.

Most of the training offerings were targeted to either beginner (n=61) or advanced (n=47) level, whereas only few sessions were considered as professional level (n=8).

Table 3 Education level, learning level of the target group (n=96).

Education level	**n**
Beginner	61
Advanced	47
Professional	8
Total	**116**

Majority of the training listed were targeted for all disciplines (over 60%). Social sciences and humanities and arts provided most of the discipline level training.

Some large universities that have campuses for different disciplines provided also discipline level training, i.e., the University of Helsinki offered campus level training targeted for specific discipline, for example medicine and health sciences.

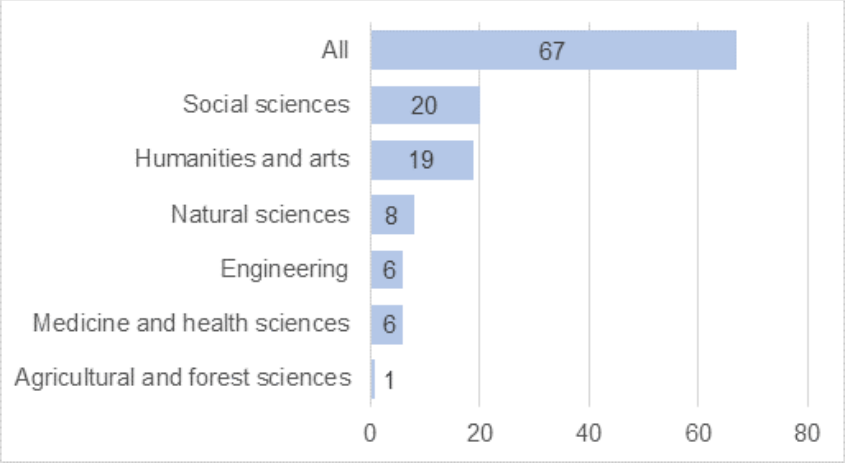

Figure 6 Discipline (n). Note: On average same training fits for an average of 1.2 different disciplines.

Different formats of the training sessions were quite equally represented, although lectures, workshops and teaching materials formed 70 % of the training formats.

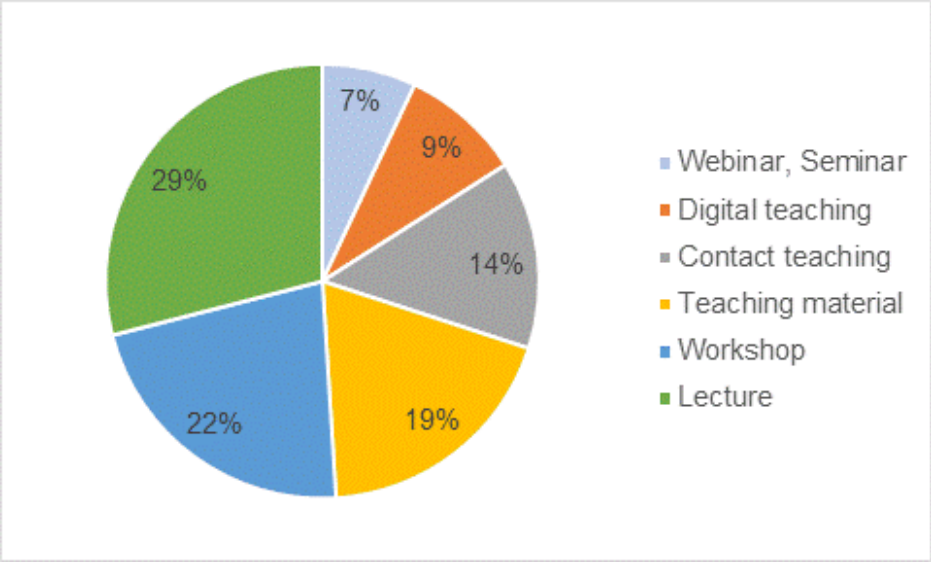

Figure 7 Training format (%).

The most common language for the training was English. The Academy of Finland and the University of Vaasa provided their training sessions in three languages: Finnish, Swedish and English. Åbo Akademi University provided their training sessions in Swedish and English. No other languages were mentioned.

Table 4 Training language. Note: On average the same training is offered in 1,4 different languages (n=96).

Training language	n
Swedish	5
Finnish	59
English	80
Total	**144**

Typically, the duration of the training varied from one to two hours (70%). Over 4 hours were 15% of the sessions and less than hour 7% (Table 5).

Table 4 Length of training in hours.

Length	**n**	**%**
Less than one hour	5	7
One hour	23	30
1,5-2 hours	30	40
2,5-3 hours	7	9
4 hours -	11	15
Total	**76**	**100**

Most training did not give any credit points for completions (70%). Typically, if the credit points were given, the maximum of 1 point was credited (20%) and in only 10% of the sessions were possible to get 2 or more credit points. (Table 6).

Table 5 How many credits in numbers.

Credits	**n**	**%**
0	52	70
Max 1	15	20
3	3	4
2	2	3
5	2	3
Total	**74**	**100**

Typically, the training offerings were offered termly or yearly (77%). Only 23% reported organizing sessions irregularly (Figure 9).

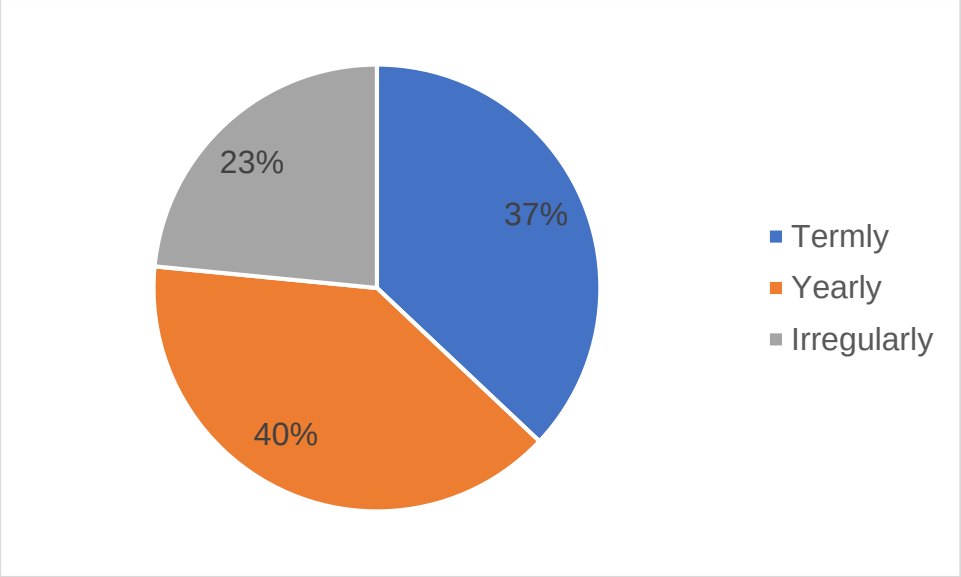

Figure 8 How often training is provided (%).

Starting year and end year

The starting year of the training was marked in 76 individual training sessions. The starting year varied between 2000 and 2021. The end year was mostly left blank indicating training was ongoing. The rest of the given answers stated that the training sessions were ongoing (93%) (Table 7).

FSD and FIN-CLARIN reported longest running ongoing courses, FSD from 2000 and FIN-CLARIN from 2004.

Table 6 Starting year. The first-time training was provided.

Starting year	**n**	**%**
2000	2	3
2004	1	1
2008	1	1
2009	1	1
2012	2	3
2016	4	5
2017	4	5
2018	6	8
2019	15	20
2020	16	21
2021	23	30
2022	1	1
Total	**76**	**100**

Detailed analysis of the training offerings

General

All respondents provided general level training or training materials. There were altogether 54 training offerings or training materials listed. In general, target groups were either doctoral students or researchers whereas the contents were mainly aimed for beginner level. In most cases training materials covered all target groups. Yet, the training provided varied from beginner level one hour lecture about data management plans provided by the Academy of Finland to 5 credit online course provided by FIN-CLARIN. Some training providers listed several training offerings at the same time.

Format of the training offerings

Courses (digital and contact) were the most common format for the training offerings, where the amount of credits varied from half a credit to three credits. In 13 of the training sessions RDM was part of a larger course, often related to information management and open science.

There were five organizations that offered courses that solemnly focused on research data management covering all aspects from planning phase to achieving the data.

Training providing credit points

University of Turku and Åbo Akademi University jointly arranged **Basics of Research Data Management** -course for doctoral students and post-doc researchers covering all aspects of RDM. This 3 credits course has been running since 2019. The digital course is targeted for all disciplines represented at both Universities.

University of Eastern Finland offered one credit **Basics of research data management**-digital course for doctoral students for all disciplines and the course has been running since 2021.

The Language Bank of Finland (organizationally part of University of Helsinki) arranged 5 credits **Data Clinic**- online course. The focus on this course was on handling, processing, and analysing language data and hence it was targeted for humanities and social sciences where the target groups are master's and doctoral students.

Tampere University offered one credit **Research data management** – course for doctoral students which focused on survey and interview data. The course is suitable for anyone working with survey and interview data, but it is traditionally focused on humanities and social sciences. The course has been running since 2020.

University of Lapland will arrange a new course starting 2022. The course has wide target audience from master students to all personnel at the university. The course is targeted for humanities, arts and social sciences.

Other training formats

In the general training, lectures were the most common training format of those not providing credits (n=13). They varied from 45 minutes lectures into 6,5-hour lecture series **Basics of Research Data Management** for researchers provided by University of Eastern Finland.

Webinars, seminars, and workshops (n=6) formed a minority of the training offerings and were mainly subject-specific training sessions or clinics, such as the National Library of Finland’s

data clinic where you get to know the National Library of Finland's digitised entities, data catalogue and open data services and have exercises related to the possibilities offered by digi.kansalliskirjasto.fi for digital humanities and the data generated by digitization.

Teaching materials

Only 9 organizations listed different kinds of training material. They were generally targeted for all levels, all disciplines and mainly for beginner level. Materials varied from data management guides to training videos. All materials were openly available for anyone, e.g. Finnish Social Science Data Archive's data management guidelines have been openly available since 2009. See the Appendix 3 for more details on all openly available training offerings.

General, several training offerings

Five of the organizations listed several training offerings where some of the sessions were in general level and some more institution-specifically focused.

JAMK University of Applied Sciences provided different kinds of webinars and seminars related to RDM for example research data management in research, development, and innovation projects.

Both University of the Arts Helsinki and LAB University of Applied Sciences provided tailored workshops on-demand-basis related to different aspects of RDM.

Finnish Meteorological Institute provided training materials on open science, open data, open access to publications, data management plans and licensing as well as their data repository, such as https://opensciencemooc.eu/ and https://mantra.ed.ac.uk/.

Geological Survey of Finland provided short workshop in natural sciences in advanced and professional level.

Plan

Respondents reported 14 ongoing or past specific data management planning related training courses. Three of these has also been reported in other topic specific RDM training offerings. The workshops are said to be hands-on training for writing a DMP in which the plans made by participants are reviewed and they are advised on how to make a plan. The respondents reported four lecture sessions aiming to support in writing the plan. One of them presents the structure of the DMP and gives tips for writing a plan, focusing on making the plan based on the Academy of Finland template. The second lecture training is part of a larger 16-hour course provided by the University of Turku, where the making of a plan is one part. Otherwise, the training is described to cover a very wide range of elements related to data management planning. Finnish Social Science Data Archive (FSD) provides several lectures dealing with planning and presenting their services. The fourth reported lecture aim to help thesis supervisors in guiding the writing of the plan and evaluate the plans. In addition, three of the training courses concern materials published online: Academy of Finland's guide and template for writing a data management plan for a funding application, the DMP template for undergraduate students published by the University of Turku and a comprehensive Data Management Guidelines provided by FSD.

The target audience:

- Researchers (10)
- Doctoral students (2)
- Other personnel (2)
- RDM professionals (1)
- Master's or equivalent level (1)
- Bachelor's or equivalent level (1)

Table 7 Overview of the reported plan related training sessions.

Name	Org.	Org. type	Topic	Education level	Discipline	Format	Duration	Regularity
Webpage information on open data and data management plans	AoF	Funder	Plan	Advanced	All	Teaching material		
RDM Advanced: Resources and responsibilities	UH	University	Plan	Advanced	All	Lecture	2	Yearly
Data Management Plans: a How-to	AU	University	Plan	Advanced	All	Workshop	1	Termly
How do I write a data management plan	TAU	University	Plan	Beginner	All	Lecture	1,5	Yearly
How to evaluate the DMP	TAU	University	Plan	Advanced	All	Lecture	1	Yearly
DMP Workshop for Researchers [DMP työpaja tutkijoille]	UEF	University	Plan	Advanced	All	Workshop	2	Yearly
Data Management Plan (DMP) Workshop	UH	University	Plan	Advanced	All	Workshop	2	Yearly
Research Data Management [Tutkimusaineistojen hallinta]	JYU	University	Plan	Beginner	All	Contact teaching	3	Termly
Basics of research data management	UTU	University	Plan	Advanced	All	Contact teaching	16	Yearly
DMP-template for undergraduate students	UTU	University	Plan	Beginner	All	Teaching material		
Academy of Finland/DMP-workshops	UTU	University	Plan	Beginner	All	Workshop	1	Yearly
Ethical pre-evaluation in the human - sciences ethics training at CEAS	UTU	University	Plan; Protect	Advanced	Humanities, arts;Social sciences	Workshop	3	
Data Management Guidelines	FSD	Other	Plan;Document;Protect; Preserve;Open & reuse	Beginner; Advanced; Professional	Humanities, arts;Social sciences	Teaching material		
Presentations on research data reuse, data management, data documentation, digital long-term preservation and other FSD basic services.	FSD	Other	Plan;Document;Protect; Preserve;Open & reuse	Beginner; Professional	Humanities, arts;Social sciences	Lecture	1	Termly

Document

Eleven organizations reported to provide training in documentation. The contents and topics of training courses varied to a great extent and six of these has also been reported in other topic specific RDM training offerings. Seven training courses are described as dealing with a specific tool, method, or data. In this group University of Helsinki provides three courses: two Redcap lectures, one on the basic functions of the software and the other on its use in the

collection of clinical research data. The third course handles Speech Analysis and is provided by the Language Bank of Finland. Aalto University offers two special training courses. They are workshops, one is an introduction to GitHub and the other a course on reproducibility of code and research data. In addition, the Finnish Social Science Data Archive (FSD) provides two training offerings especially helpful for research utilising personal data. Other documentation training offerings are more general in nature.

The target audience:

- Researchers (10)
- Doctoral students (6)
- Other personnel (4)
- RDM professionals (4)
- Master's or equivalent level (1)
- Bachelor's or equivalent level (1)

Table 8 Overview of the reported data documentation related training sessions.

Name	Org.	Org.type	Topic	Education level	Discipline	Format	Duration	Regularity
Introduction to GitHub	AU	University	Document	Advanced	Natural sciences; Engineering	Workshop	1	Irregularly
How to Make Your Research/Code Reproducible and Reusable	AU	University	Document	Advanced	All	Workshop	1.5	Termly
Documentation and metadata-learn how to make your research data understandable and re-usable	ÅA	University	Document	Beginner	All	Lecture	1	Irregularly
Metadata and Data Description [Metadata ja aineiston kuvailu]	TAU	University	Document	Beginner	All	Lecture	0,5	Irregularly
RDM Advanced: Documentation	UH	University	Document	Advanced	All	Lecture	2	Yearly
Introduction to Speech Analysis	UH	University	Document; Open & reuse	Beginner; Advanced	Humanities, arts;Social sciences	Digital teaching		Yearly
Training about Fairdata services	CSC	Other	Document; Store;Open & reuse	Beginner; Advanced	All	Workshop	1	Termly
REDCap training – overview and basic functions	UH	University	Document; Store;Protect	Advanced	All	Lecture	2	Termly
RedCap Basics - Clinical Research	UH	University	Document; Store;Protect	Beginner; Advanced	Medicine and health sciences	Lecture	2	Termly
Presentations on research data reuse, data management, data documentation, digital long-term preservation and other FSD basic services.	FSD	Other	Plan;Document;Protect;Preserve;Open & reuse	Beginner; Professional	Humanities, arts;Social sciences	Lecture	1	Termly
Data Management Guidelines	FSD	Other	Plan;Document;Protect;Preserve Open & reuse	Beginner; Advanced Professional	Humanities, arts;Social sciences	Teaching material		

Store

Four respondents answered that they offer data storage training, and altogether they were organizing seven training sessions. All the organizers were universities apart from one training provided by CSC (Table 8).

Training contents were focusing on data storing solutions on a practical level, introducing organization-specific solutions or data management tools. The training by CSC consisted of introducing their Fairdata services for organizations.

The target audience:

- Researchers (7)
- Doctoral students (3)
- Other personnel (3)
- RDM professionals (3)

Table 9 Overview of the reported data storage related training sessions

Name	Org.	Org. type	Topic	Education level	Discipline	Format	Duration	Regularity
Training about Fairdata services	CSC	Other	Document;Store; Open & reuse	Beginner; Advanced	All	Workshop	1	Termly
REDCap training – overview and basic functions	UH	University	Document;Store; Protect	Advanced	All	Lecture	2	Termly
RedCap Basics - Clinical Research	UH	University	Document;Store; Protect	Beginner; Advanced	Medicine and health sciences	Lecture	2	Termly
How to Store Research Data	AU	University	Store	Beginner	All	Workshop	1	Termly
Where should I store my research data and other files? Seminar on storage solutions and collaboration tools for ÅAU researchers	ÅA	University	Store	Beginner	All	Webinar, Seminar	2,5	Irregularly
RDM Advanced: Storing solutions	UH	University	Store	Advanced	All	Lecture	2	Yearly
Working with Restricted Datasets	AU	University	Store;Protect	Advanced	All	Workshop	1,5	Irregularly

Protect

Fourteen training sessions could be identified handling the topic of data protection as part of the training, and six of these sessions were focusing solely on data protection issues.

Most of the training sessions handled data protection topics from a legal and ethical point of view (5 sessions). The approach of the other training sessions was very practical, as the topics handled data anonymization (3 sessions), the use of the specific RDM tool (2 sessions) or working with restricted data (1 session). The beginner-level training organized by the University of Tampere and two training offerings from FSD were assessing the topic at a general level.

The target audience:

- Researchers (12)
- Doctoral students (5)
- Other personnel (3)
- RDM professionals (3)
- Master's or equivalent level (1)
- Bachelor's or equivalent level (1)

Table 10 Overview of protect related training events.

Name	Org.	Org. type	Topic	Education level	Discipline	Format	Duration	Regularity
REDCap training – overview and basic functions	UH	University	Document; Store;Protect	Advanced	All	Lecture	2	Termly
RedCap Basics - Clinical Research	UH	University	Document; Store;Protect	Beginner; Advanced	Medicine and health sciences	Lecture	2	Termly
Research data from a legal perspective	ÅA	University	Protect	Beginner	All	Lecture	1	Irregularly
Basics of anonymization	FSD	University	Protect; Preserve	Advanced; Professional	Humanities, arts; Social sciences	Workshop	2	Yearly
Ethical pre-evaluation in the human - sciences ethics training at CEAS	UTU	University	Plan;Protect	Advanced	Humanities, arts; Social sciences	Workshop	3	
Handling of Personal Data	AU	University	Protect	Advanced	All	Workshop	1,5	Termly
Basics of Data Anonymisation	AU	University	Protect	Advanced	All	Workshop	1	Termly
Legal Aspects of Research Data	AU	University	Protect	Advanced	All	Workshop	1	Termly
Data management and data protection	TAU	University	Protect	Beginner	All	Lecture	1,5	Termly
RDM Advanced: Law & Ethics	UH	University	Protect	Advanced	All	Lecture	2	Yearly
Open Science and licenses	UTU	University	Protect	Beginner	All	Webinar, Seminar	0,5	Yearly
Working with Restricted Datasets	AU	University	Store;Protect	Advanced	All	Workshop	1,5	Irregularly
Data Management Guidelines	FSD	Other	Plan;Document;Protect; Preserve;Open & reuse	Beginner; Advanced; Professional	Humanities, arts;Social sciences	Teaching material		
Presentations on research data reuse, data management, data documentation, digital long-term preservation and other FSD basic services.	FSD	Other	Plan;Document;Protect;Preserve;Open & reuse	Beginner; Professional	Humanities arts;Social sciences	Lecture	1	Termly

Preserve

Three organizations offered training on preserving issues: CSC, FSD and University of Helsinki. The topics focused on digital preservation and choosing data for preservation, data anonymization and archiving of research data.

The target audience:

- Researchers (4)
- Doctoral students (3)
- Other personnel (3)
- RDM professionals (3)
- Master's or equivalent level (2)
- Bachelor's or equivalent level (2)

Table 11 Overview of preserve related training events.

Name	Org.	Org. type	Topic	Education level	Discipline	Format	Duration	Regularity
Basics of anonymization	FSD	University	Protect; Preserve	Advanced; Professional	Humanities ,arts; Social sciences	Workshop	2	Yearly
Manage well and get preserved - videos	CSC	Other	Preserve	Advanced	All	Teaching material	1	Irregularly
RDM Advanced: Publishing and archiving data	UH	University	Preserve; Open & reuse	Advanced	All	Lecture	2	Yearly
Presentations on research data reuse, data management, data documentation, digital long-term preservation and other FSD basic services.	FSD	Other	Plan;Document; Protect;Preserve ;Open & reuse	Beginner; Professional	Humanities arts;Social sciences	Lecture	1	Termly
Data Management Guidelines	FSD	Other	Plan;Document; Protect; Preserve;Open & reuse	Beginner; Advanced; Professional	Humanities , arts;Social sciences	Teaching material		

Open and re-use

Eleven training sessions handled the topic of opening and re-using data. Three training sessions focused on both open access publishing and open research data, two sessions handled just open research data and three sessions opening data from the FAIRdata perspective. Numerous topics of data management were covered in the two training sessions. One training session focused on specific speech tools for speech analysis (training organised by Language Bank of Finland).

The target audience:

- Researchers (6)
- Doctoral students (7)
- Other personnel (1)
- RDM professionals (1)
- Master's or equivalent level (1)
- Bachelor's or equivalent level (1)

Table 12 Overview of the reported opening and re-using data related training sessions.

Name	Org.	Org. type	Topic	Education level	Discipline	Format	Duration	Regularity
Introduction to Speech Analysis	UH	University	Document; Open & reuse	Beginner; Advanced	Humanities, arts;Social sciences	Digital teaching		Yearly
Training about Fairdata services	CSC	Other	Document; Store;Open & reuse	Beginner; Advanced	All	Workshop	1	Termly
How to Share Research Data Through a Data Repository: A Zenodo Example	AU	University	Open & reuse	Advanced	All	Workshop	1,5	Termly
Current Trends in Academic Publishing: Plan S and Overlay Journals	AU	University	Open & reuse	Advanced	All	Workshop	1,5	Irregularly
FAIR and/or open data - how to get started (there are some variations of the course name, but content quite similar)	ÅA	University	Open & reuse	Beginner	All	Lecture	1,5	Yearly
How to make research data FAIR	TAU	University	Open & reuse	Advanced	All	Lecture	1	Irregularly
Basics of Open and Responsible Science: Open Access Publishing and Research Data	UEF	University	Open & reuse	Beginner	All	Digital teaching	15	Yearly
Open access for research outputs in University of Turku	UTU	University	Open & reuse	Beginner	All	Contact teaching	1	Yearly
RDM Advanced: Publishing and archiving data	UH	University	Preserve; Open & reuse	Advanced	All	Lecture	2	Yearly
Presentations on research data reuse, data management, data documentation, digital long-term preservation and other FSD basic services.	FSD	Other	Plan;Document;Protect;Preserve;Open & reuse	Beginner; Professional	Humanities, arts;Social sciences	Lecture	1	Termly
Data Management Guidelines	FSD	Other	Plan;Document;Protect;Preserve;Open & reuse	Beginner; Advanced; Professional	Humanities, arts;Social sciences	Teaching material		

Discussion

The aim of the study was to obtain a comprehensive understanding about the RDM training offerings of various Finnish research and higher education organizations, and which organizational units are involved in providing training. It was also important to find out the key focus areas of the training offerings: whether the training provides general understanding about RDM or whether the training focuses on specific topics such as documentation, publishing, or storage of data. In addition, we wanted to see if training offerings had become more formal in terms of academic performance by awarding credits and whether the training offerings were available to learners of various levels.

We sent the survey into 74 organizations, of which 36 responded so the response rate was low. Furthermore, most of the respondents were from universities (14) and universities of applied sciences (13). Only 6 respondents were from research institutes. Thus, we can say this survey gives a comprehensive information mostly about the higher education organization (HEO) data management training providers and the training they provide but not about the training provision of other organizations.

General RDM training is well available in all organizations. It is also noteworthy that **topic-specific** training sessions such as **storage** (13 sessions) and **documentation** (11 sessions) were offered to a great extent. However, there were only a few advanced-level training offerings that concerned with, for example, a particular method, data type, or data management from the perspective of a particular discipline. Therefore, we can clearly identify a training gap and the need for more advanced and targeted training. The offering of those training sessions was also provided by very few organizations (4). For example, both Language Bank of Finland and Finnish Social Science Data Archive (FSD) provide detailed discipline level training offerings. FSD has several training offerings that support research utilizing personal data. The training deal with issues such as anonymisation, storage and publishing as well as the documentation of personal data. The Language Bank of Finland, on the other hand, offers training that supports the handling and analysing of language data. Also, other organizations could provide discipline-specific training for their focus areas, for example in natural sciences.

Based on the survey, we can say that the provision of training in organizations has become a permanent and systematic part of the provision of services and support. Three-quarters reported that training is provided on a regular basis (77%). It is also gratifying that more formal RDM training seems to be quite common in respondent organizations. The results showed that 22 of the training sessions were granting study points when concluding the training. While credits were given, it was mostly **general** type of training (18 training sessions).

RDM training was provided by people working in libraries and information services, as well as workers with similar experience and education from research services. Thus, we can say that staff with a library background play a key role in producing RDM training in Finland.

The question of which other units are involved in organizing the training is also interesting, especially since it is recognized that understanding of several topics and practical implementation often requires expertise in conducting research. An assessment of the extent of the co-operation and the contribution of the different units showed that libraries and

research services co-organized training offerings, mainly with IT services (10 mentions) and legal services (7 mentions).

Most of the training offerings were provided for researchers and doctoral students or equivalent level. This is understandable where the DMP's for example are mandatory for those groups and need guidance and training. What got less attention in training offerings where bachelor's and master's or equivalent level and research data management professionals. The lack of training for RDM professionals was reflected also in educational level where only 8 training sessions were offered.

Limitations

The representativeness of the data is affected by the facts that response rate was low and furthermore the most active respondents were higher education organizations (HEO). We may have missed out on relevant information, or the results may be biased. First, only part of the training offerings may have been reported. We cannot be sure about the prevalence of under-reporting, but in terms of training materials, it is quite likely. Only 13 organizations reported providing some kind of **training material** (20% of the training offerings). The amount is unexpectedly low considering that e.g., establishing websites is one of the most common ways to provide RDM support within organization (Garbuglia et al., 2022, p. 16). Second, we possibly did not reach all the units providing RDM training because we submitted a request to participate mainly to well-known training providers. As a result, we may be missing information on training arranged by, for example, teachers, researchers, and faculties. The activity of HEOs makes sense as they have a large number of academic staff and thus the demand for RDM training in these organizations is most likely high. They also have more resources available, such as library/information service/research services and legal and IT staff, to organize systematic training activities compared to smaller research institutes.

In cross-checking the responses, we also found that some questions on the form might have been interpreted differently than intended and thus we should have paid more attention to guidance. This concerned questions on Educational level (the target group of a training) and Education level. (the competence level of a participant). For example, training for researchers may have been interpreted automatically as training at the professional level. However, the level of training should have been chosen according to whether participation in training required a little (beginner) or a lot of knowledge (advanced or professional levels) on the RDM topic in question.

The possibility of selecting multiple choices for the training categories may have led to different interpretations in the classification among the respondents. It was quite common that if the respondent chose category general for the training, they also chose the rest of the categories as well instead of just all applicable. Also quite often, the other way around, when choosing a particular category, such as plan, the category of general was also chosen. We recategorized the respondents' choices by comparing them with names of the training offerings and course descriptions as well as by checking the website if possible.

Conclusions

In the future, RDM will be part of basic education and working life competence same as IT skills are. Everyone needs to know where the data can be stored, where to find it, who has the right to use it or how and why the data should be destroyed.

The results of the survey gives an overview of the present situation of the RDM training opportunities in Finland, but does not cover all RDM related training organized. Apart from that, the results presented in this survey report indicate that Finnish organizations are aware of the needs in training provision and are providing various types of RDM training offerings to different target audiences and educational levels.

The survey showed that the main training providers were libraries and equivalent service units. This is understandable when we think about the fundamental goal which libraries historically have and hence have taking a central role in supporting good RDM practices. In organizations where a one desk service was established for research data management services, most of the training offerings were provided together with IT and legal services. Even some organizations have created a solid base for RDM practises and training offerings, they cannot answer the growing demand for different discipline level training offerings for various levels. The growing need should be supported at all levels of the organizations and more RDM experts need to be involved. Organizing training offerings and developing materials would also increase collaboration within the organization. This would also be a chance for national and international co-operation.

We propose more efforts to be invested in these areas:

- RDM training need to be implemented into the existing study programmes and training must be provided on all degree levels. This can be partly done by integrating the RDM practises into suitable existing courses, such as information seeking courses.
- Especially the training offerings intended to students, should be awarded with credits to increase the motivation to participate. That would eventually enable RDM practises becoming daily practises. One possibility would be to develop the current national training opportunities into this direction, e.g., MOOC.
- More training offerings are needed in documentation, opening and re-using the data on discipline level. This could be done as national and international co-operation.
- The requirements for the content of training are likely to increase, and it is therefore necessary to ensure that competent trainers are available and that trainers come with variety of backgrounds. Professional level training as well as career opportunities should be provided by the organizations.
- Organizations need to provide enough resources to co-ordinate organized training offerings and more co-operation is needed between units within organizations.

This survey provides valuable information for those that work with RDM related training. In addition, the results could be used on making a common roadmap for future RDM education in Finland. In this report we provide a comprehensive list of open-source materials provided by the participants of this survey and examples of how they have organized their training offerings. This would especially benefit those that are planning their training. With this report

we also wish to increase the collaboration between different organizations and within organizations.

We wish to thank all participants and those that were in the initial face of this work: Siiri Fuchs (CSC), Liisa Hallikainen (ULA), Harri Ollikainen (University of the Arts Helsinki) and the open data expert group.

Abbreviations

AoF	Academy of Finland
AU	Aalto University
CSC	CSC – IT Center for Science
DMP	Data management plan
FSD	Finnish Social Science Data Archive
FIN-CLARIN	Language Bank of Finland
HEO	Higher education organization
JYU	University of Jyväskylä
RDM	Research data management
TAU	Tampere University
UEF	University of Eastern Finland
UH	University of Helsinki
ULA	University of Lapland
UTU	University of Turku
ÅA	Åbo Akademi University

References

Academy of Finland. *September Call. 1-28 September 2016*, p. 51-53. https://www.aka.fi/globalassets/10rahoitus/hakuilmoitukset/hakuilmoitus_syyskuu_2016_en.pdf

Avoin tiede (julkaisuaika tuntematon). *Open Science and Research Initiative*. Haettu 5.5.2022 osoitteesta https://avointiede.fi/en/policies/policies-open-science-and-research-finland/open-science-and-research-initiative

Baker, M. 1,500 scientists lift the lid on reproducibility. (2016). *Nature* **533,** 452–454. https://doi.org/10.1038/533452a

Fuchs, S. & Kuusniemi, M. (2018). *Making a research project understandable - Guide for data documentation* (1.2). p. 3. Zenodo. https://doi.org/10.5281/zenodo.1914401

Garbuglia, F., Morais, R., Berghmans, S., Gaillard, V. (2022). *A closer look at research data practices in European universities Follow-up to the 2020-21 - EUA Open Science survey*. p. 16. https://eua.eu/downloads/publications/research%20data%20follow-up%20report.pdf

Helbing, D. (2015). What the Digital Revolution Means for Us. *In: Thinking Ahead - Essays on Big Data, Digital Revolution, and Participatory Market Society*. Springer, Cham. https://doi.org/10.1007/978-3-319-15078-9_12

Koskinen, K. (13.6.2019). Institutional RDM Support at the University of Helsinki. *OpenAire blog.* https://www.openaire.eu/blogs/institutional-rdm-support-at-the-university-of-helsinki

Kuula, Arja & Borg, Sami (2008). Open Access to and Reuse of Research Data. The State of Art in Finland. University of Tampere. Finnish Social Science Data Archive; 7. p. 9. https://www.fsd.tuni.fi/fi/tietoarkisto/julkaisut/julkaisusarja/fsdjs07_oecd_en.pdf

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3,** 160018. https://doi.org/10.1038/sdata.2016.18