NFDI4Earth

Deliverable D1.3.1

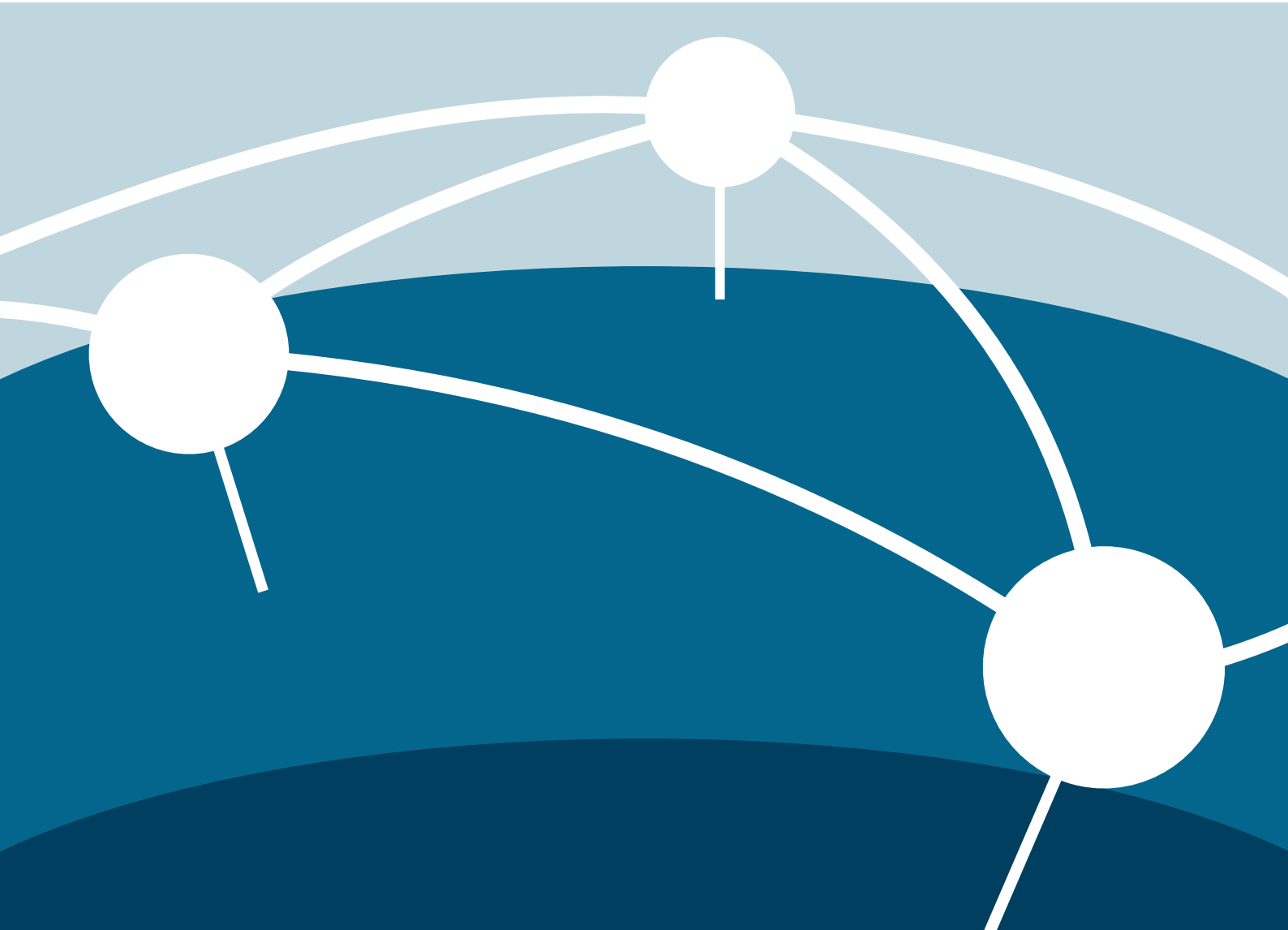
Mapping of existing educational resources and initial education and training needs within the Earth system science community

Farzaneh Sadeghi (farzaneh.sadeghi@hs-bochum.de), Carsten Keßler, Yomna Eid, Edzer Pebesma, Balthasar Teuscher, Martin Werner, Christopher Purr, Remon Sadikni

2022-12

DOI: 10.5281/zenodo.7940195

nfdi4earth.de



Citation

Farzaneh Sadeghi, Carsten Keßler, Yomna Eid, Edzer Pebesma, Balthasar Teuscher, Martin Werner, Christopher Purr, Remon Sadikni. 2023. *Mapping of existing educational resources and initial education and training needs within the Earth system science community (NFDI4Earth Deliverable D1.3.1)*. Zenodo. https://doi.org/10.5281/zenodo.7940195

Acknowledgement

This work has been funded by the German Research Foundation (DFG) through the project NFDI4Earth (DFG project no. 460036893, https://www.nfdi4earth.de/) within the German National Research Data Infrastructure (NFDI, https://www.nfdi.de/).

Executive summary

This report summarizes the need assessment undertaken to clarify the research data management-related training needs of the Earth system science (ESS) researchers and maps out existing educational resources on research data management (RDM) and research data analysis tailored for ESS researchers.

To ensure that not only the training topics needed by the Earth system science researchers but also how to make maximum use of the existing and future educational materials are taken care of by M1.3 activities, the needs were gathered and analyzed in two steps. First NFDI4Earth members were surveyed and asked to share their needs and then a focus group discussion with experts in research data management was conducted.

Results of the needs assessment indicated that the ESS researchers need practical research data management training. Open and reusable training materials are needed to facilitate educators and accelerate the RDM training within the community. In addition, existing and coming educational materials should become more findable and be associated with rich metadata. Offering personalized and adjustable learning methods and materials is needed, to encourage researchers to practice RDM in their own scientific work. Developed materials and curricula should be flexible to keep up with fast changes in digital training needs, to enable a long-term service. Currently, a collaboration between similar projects is needed to boost the impact of our work as well as exchange experiences.

In addition to training need assessment, the existing educational resources that can benefit the users of our educational services have been collected. The collected resources are mapped out into three main categories. Firstly, we map our resources to the main four components of the earth. Secondly, we follow the classification of the German Research Foundation (DFG) and map our resources to various subject areas within the ESS. Lastly, they correspond with each stage of the research lifecycle. By implementing a comprehensive mapping strategy, we gain a better understanding of the areas where new material is necessary and when we can modify existing material. This approach enables us to optimize our resources and improve our overall productivity.

The list of collected educational materials covering ESS-related research data management and data analysis is attached to this report. The list also includes detailed information on each resource. This information will be used later to create the NFDI4Earth educational catalog. The outcomes of this report will be used to form and prioritize the NFDI4Earth educational and training activities.

Glossary and terminology

CEN - Center for Earth System Research and Sustainability, Universität Hamburg

D1.3.1 - Deliverable. 1.3.1: Mapping of existing educational resources and initial education and training needs within the Earth system science community

DFG - German Research Foundation

DMP - Data Management Plan

EduHubs - An open network of education and training hubs

EduPilots - Educational Pilots

EduTrain - Education and Training Materials and Services, see https://www.nfdi4earth.de/2participate/education-training

EOSC - European Open Science Cloud, see https://www.eosc-portal.eu

EPOS - European Plate Observing System, see https://www.epos-ip.org

ES - Earth System

ESS - Earth System Sciences

FAIR - Findability, Accessibility, Interoperability, and Reusability

GFZ - Helmholtz Centre Potsdam – GFZ German Research Centre for Geosciences

HSBO - Bochum University of Applied Sciences

ICAML - Interdisciplinary Center for Applied Machine Learning

ICDC - Integrated Climate Data Center

M1.3 - Measure1.3: Education and Training Materials and Services

NFDI - National Research Data Infrastructure (Nationale Forschungsdaten Infrastruktur), see http://www.nfdi.de

NFDI4Earth - NFDI Consortium Earth System Sciences / NFDI Konsortium Erdsystemforschung, see https://www.nfdi4earth.de

OER - Open Educational Resources

RDI - Research Data Infrastructure

RDM - Research Data Management

RI - Research Infrastructure

SICSS - School of Integrated Climate System Sciences

TA - Task Area

TUM - Technische Universität München

UNIHH - Universität Hamburg

UNIMS - Universität Münster

Contents

1. Introduction **1**
1.1. Measure 1.3: education and training materials and services 1
1.2. M1.3 participants 1
1.3. Deliverable 1.3.1 2

2. Mapping of education and training needs within the ESS community **3**
2.1. Review of previous work 3
2.2. User stories 6
2.3. Focus group interview 7
2.4. Challenges 8

3. Mapping of existing educational resources **10**
3.1. Identifying relevant training topics 10
3.2. Categorization 12

4. Future work **15**
4.1. Synergy building 15
4.2. Metadata objects for educational content 16
4.3. Input for the NFDI4Earth training e-catalog 17
4.4. Impact on the measure 17

5. Conclusions **18**

References **19**

Acknowledgements **19**

Appendix **20**

A. User stories **20**

B. Existing educational resources **21**

1. Introduction

1.1. Measure 1.3: education and training materials and services

The goal of measure 1.3 (M1.3) is to enable widespread uptake of NFDI4Earth services for conducting open and sustainable research. The users of our services are predominantly researchers in the Earth System Sciences, ranging from early career to experienced senior scientists and professors, ESS educators and training professionals, and ESS university students.

The primary objective of our materials and services is to encourage and assist researchers and educators to practice and promote sustainable, FAIR, and open research from the early stages.

Additionally, the NFDI4Earth training platform, jointly with the NFDI section training and education, and our under-development network of educational institutes (EduHubs), will continue to identify emerging gaps in training provision across Germany and ensure that appropriate training solutions are developed and delivered. These gaps will be addressed either by the NFDI4Earth team, the EduHubs, or the educational pilots (EduPilots) funded through this measure. The approach of the M1.3 in this regard is to:

1. Discover the training needs and future trends around research data management and analysis in research data infrastructures for the ESS community
2. Map existing, yet largely scattered relevant open educational resources tailored for ESS
3. Provide rich metadata for the collected material and add them to the NFDI4Earth knowledge hub
4. Develop new materials to meet the training needs of the ESS community
5. Develop target group-specific curricula
6. Revise the developed contents and curricula regularly based on user feedback

1.2. M1.3 participants

Measure 1.3 is coordinated by the team at Bochum University of Applied Sciences, and implemented in collaboration with the following participants:

- **Technische Universität München, Big Geospatial Data Management**: provides education and training activities and materials based on course material gathered from their lectures “Big Geospatial Data”, “Principles of Programming”, “Principles of Spatial Data Mining and Machine Learning” and the Interdisciplinary Center for Applied Machine Learning (ICAML). Beyond this project, they are contributing to a larger activity to build a curriculum to educate students for mobile AI chip design covering for the first time the span from computer science and electronic circuits (e.g., chip

design, operating systems, FPGAs) all the way to geospatial applications (autonomous drones, deadwood reckognition in large scale, autonomous geospatial observation). Furthermore, foundational computational course elements can be contributed from the two-semester course “Computational Foundations” held for the bachelor aerospace.

- **Universität Hamburg, Center for Earth System Research and Sustainability (CEN)**: provides expertise based on lectures dealing with data analysis in ESS held in different master programs of the UHH. CEN fosters an interdisciplinary exchange, e.g., the Integrated Climate Data Center (ICDC) is involved in the master program of the SICSS (School of Integrated Climate System Sciences) showing how to use EOS data with the available tools.
- **University of Münster, Institute for Geoinformatics**: offers training and education activities in the context of their regular teaching programs (BSc Geoinformatik, MSc Geoinformatics and Spatial Data Science, MSc Geospatial Technologies (Erasmus Mundus joint degree), and graduate school in geoinformatics), and provides open teaching materials in the OpenGeoHub summer schools (https://opengeohub.org/summer-school/) as well as on-line teaching materials such as the open course book ”Spatial Data Science, with applications in R” (https://r-spatial.org/book).

1.3. Deliverable 1.3.1

The purpose of this first deliverable (D1.3.1) from M1.3, entitled ”Mapping of existing educational resources and initial education and training needs within the Earth system science community”, is to collect existing educational resources tailored for supporting RDM in ESS. Moreover, the report summarizes the identified training needs of the ESS community for enabling FAIR research data management. The remainder of this report is structured as follows: Section 2 describes the mapping of education and training needs within the ESS community, including a literature review based on the results from previous projects, the development of user stories, and the outcome of a focus group interview. Section 3 describes how the collection of relevant existing training and education materials has been performed and how the corresponding catalog has been categorized. Section 4 then summarizes the main outcomes of our analysis, followed by brief conclusions in Section 5.

2. Mapping of education and training needs within the ESS community

A needs assessment has been performed to sharpen the future objectives of M1.3 and to enable efficient planning to meet the training needs of the users. In addition, the needs assessment intends to enable the education and training team to provide their services relying on existing developments, but at the same time expanding and upgrading them according to future trends. Therefore, besides the topics requested to be covered by ESS researchers, the needs assessment has also focused on the educational content presentation methods and their shortcomings, as well as insights from similar projects and why the educational content that is currently available is not used optimally.

Three actions were taken to identify the digital training needs of the ESS community. First, similar projects and initiatives were detected, and their publications were reviewed to have an insight into what the outcomes of their training need assessment have been. Second, to gain information about what future NFDI4Earth users will require to efficiently use the infrastructure, the project consortium team has been surveyed and asked about their needs and requirements. Since the NFDI4Earth community often acts in the dual role as users and providers of Earth System (ES) data [1], involving them from the very beginning was a high priority. This was done jointly with the user expert group. Lastly, a focus group discussion was conducted with representatives from other RDM projects and national as well as international research infrastructures. The goal of this discussion was to collect the current training needs of the community beyond NFDI4Earth and to exchange ideas on innovative training approaches and environments to address these needs.

2.1. Review of previous work

In the context of Big Earth data and the uptake of FAIR data principles, a preliminary step to tackling the necessary needs of the Earth System Sciences community is to conduct a needs assessment and gap analysis. Before conducting our needs assessment, we surveyed the most relevant current literature and extracted the following key information.

A report produced by the highly-relevant ELIXIR project [2] identified the need for establishing a domain-agnostic job description for the data steward role. With an increase in the data volumes generated by and produced for ESS, the community needs a clear definition of the skills framework for data stewardship. Three different roles were characterized for data stewards depending on their three partially overlapping stakeholder fields. These are policy (i.e., defining policies at institutions for specific data/use cases, FAIR data, etc.), research (i.e., adopting the appropriate data workflows, tools, standards, and infrastructure), and finally infrastructure (i.e. facilitating software, hardware, services, and technical infrastructure for users). All three

data steward roles are often experienced as in-between different disciplines and professionals, aiding in the translation between the different stakeholders and professionals. The eight competence areas were defined as policy and strategy, compliance, alignment with FAIR data principles, services, infrastructure, knowledge management, network, and data archiving.

According to the study, "the extensive training materials created by groups largely working in isolation are highly distributed and are themselves not adherent to FAIR principles" [2]. The key challenge identified was the presence of a lot of spread expertise that was hard to find, as well as the development of a shared understanding of data stewardship requirements between the wide variety of different initiatives.

Thus, an agreed terminology for the necessary skills to make and keep the data FAIR is missing and is highly needed to assist in the construction of training frameworks and curricula for FAIR data stewards that will enable the more precise and fine-grained annotation and discovery (findability) of the training materials. Additionally, the design of a computer-readable semantic model that will be used for aligning existing resources and improving their findability is needed.

Spread across projects from 2016-2019, the identified needs in summary were:

- A high need for training, as most training is currently done 'on the job'.
- Difficulty in navigating and assessing the quality of information present – thus a need for a condensed, structured way to offer the available information and courses would be very helpful to data stewards.
- Providing courses with quality indication including certification for data stewards, to show that they have the competencies and for employers to be able to hire qualified persons.
- Much more tailored guidance and discipline-specific examples to help them apply the Data Management Plan (DMP) questions to their context, i.e., provide domain-level guidance.
- Demand for reward structures that ensure data stewardship is evaluated on suitable criteria, beyond those used to evaluate research itself (e.g., publication of peer-reviewed papers), as well as providing incentives for training, professional development, and good practice from research funders, learned societies, professional bodies, and publishers.

Another study [3] regarding the motivation for migration to cloud-based services was run as a web-based survey among Big Earth data users to better understand the challenges and opportunities that might arise concluded that "an insufficient literacy in cloud systems and a lack of trust due to security concerns and opacity of emerging costs" exists. The results of the survey highlight the following needs:

- Tailor services to intermediate users: Cloud services that are targeted at intermediate users instead of policy- and decision-makers; and the development of open rather than over-engineered systems with a high level of abstraction, which should be avoided.

- Invest in capacity-building: More substantial capacity-building efforts are required to decrease the existing gap in cloud skills and uptake. Following the principles of reproducibility, collaboration, and open science, training activities with the help of interactive tools (e.g. computation notebooks) in addition to modular self-paced and online learning activities are likely suggestions.
- Develop cloud certification standards: A cloud certification mechanism that could help in building up overall trust in cloud-based services.

Lastly, another notable study [4] was conducted by the German Research Centre for Geosciences (GFZ) within its institution, to identify the gaps and needs in their data management process. With 411 respondents, the proportion of the different profiles was section head/group leader (36%), Post-doc/Senior scientist (28.9%), Ph.D. student (24.4%), and other (13.9%). The survey concluded that there was a pressing need for:

- Provision of adequate storage space (finding an adequate repository) for long-term archiving and data preservation (~50% of study survey respondents), advice and guidance on technical issues (~40%) such as metadata standards, improvement, and structuring of research data handling, and advice and guidance on general research data management issues (~35%), backup schedules, and familiarization with good scientific practice.

The findings of this study likely reflect community-wide issues, and the most important gaps identified are grouped into six categories discussed in the following paragraphs.

In terms of **data safety** practices, 55% of the respondents regard backup as their personal responsibility, whereas 41% rely on dedicated staff. While the common backup frequency was reported as daily and/or weekly, one-fifth had no schedule. The identified locations of back-ups were external hard drives, the institution/GFZ backup service, or group and section servers. Particularly during active project phases, backups were (76%) on local data storage, (60%) on the central group server, and (42%) on external hard drives.

The survey investigated **data acquisition**, identifying a large variety of data types (from the sensor to modeling data) and formats that often require proprietary or own software – which was a key issue in terms of interoperability.

Data documentation was also surveyed with approximately a quarter of respondents not recording any metadata. Thus, the respondents said they would benefit from data documentation tools, clearly defined workflows, community-agreed standards, and project-related data management resources. A variety of training needs in the data management process were recognized as well, including archiving, metadata standards, general research data management, backup, and licensing.

The **data preservation and storage** procedures were explored as well, with respondents

reporting more than 100GB of generated data in the past 5 years. Additionally, only 40% were aware that research data underlying a scientific publication must be archived at the institution of origin for at least 10 years, with 50% of Ph.D. students unaware of this requirement. Moreover, only 32% of respondents indicated data from completed projects were archived in a research data repository.

Data sharing habits during active project phases reported that data are often shared within project and research groups only. Data sharing requirements by journals were largely unknown by 48% of the respondents as well. However, most respondents preferred to publish data as an article supplement, with already 42% following the best practice of data sharing via a domain-specific research data repository. While 65% of the respondents state that they would share their data publicly if they were given credit (e.g. via citation), around (62%) would require dedicated embargo periods or personal permission as a condition for data sharing. Reportedly, 67% of the respondents rely on recommendations from colleagues or use the institute's infrastructure, with 20% of respondents unaware of the possibility to obtain data via research data repositories.

Finally, querying the general awareness of **data management** aspects including workflows, practices, and policies revealed that a large percentage of respondents were unaware that existing data policies or guidelines apply to them. Around 67% of all respondents were not familiar with the German Research Foundation 2006 Guidelines for Safeguarding Good Scientific Practice. However, it is worth noting that the currently running version was last updated in April 2022, but no statistics were presently found regarding the new version.

2.2. User stories

During May and June 2022, with the cooperation of the user expert group, the needs of the community were collected in the form of user stories. The result of this collection was 183 user stories written by partners of NFDI4Earth. For this survey, the participants were asked to think of different characters who may be users of our services and write a request for each one and the reason for that request. Out of these 183 stories, 15 directly expressed the user expectations from the education and training team (Appendix B).

The majority of stories expressing the training needs were concerning research data management training and the FAIR principle. The educational material format, the language of the materials, as well as materials fit for learners and educators were among the requests of the participants from the NFDI4Earth education and training services; one request was on curricula for inspiration with no further details (Appendix B). Overall, the user stories did not reveal any new content needs, innovative formats, or specific training services that were not already under consideration by the M1.3 team.

2.3. Focus group interview

The educational needs detected in this survey have turned out to be quite generic and mostly focus on the current situation, therefore lacking a perspective of future training and education needs. To fill this gap, a focus group discussion including experts with research data management experience and awareness of future trends and upcoming developments has been conducted.

To complement the needs mapping, collect input on future training and education needs, and generate innovative ideas for the development of the corresponding materials and services, we invited scientists and research data managers from a diverse set of Earth science fields to participate in this discussion. On 18th October 2022, email invitations were sent to candidates for the discussion. The candidates were professionals from similar projects all over Europe, as well as NFDI4Earth, working in academic institutions and research organizations. We contacted individuals who were either educators or worked together with researchers on RDM improvement. Out of 12 invited researchers, eight were able to attend the discussion on the 1st of December 2022 (Table 1). The discussion was held in English and lasted two hours as per schedule.

Table 1: professional background of the focused group discussion participants

ID	Institute	Project	Background	Location
1	CSC (IT center for science)	EOSC	Sector engagement, Data infrastructure	Finland
2	Deutsches Klimarechenzentrum	NFDI4Earth	Data management, Climate science	Germany
3	Aalborg University	Skills4EOSC	Intersection study of IT and people, Change management and digital innovation	Denmark
4	University of Turku	Finnish national research infrastructure for geoinformatics	Capacity-building ideas and the community engagement, Professor in Geospatial research	Finland
5	GARR	Skills4EOSC ICDI	Library sciences, Digital cultural heritage	Italy
6	GARR	Skills4EOSC ICDI competence center	Researcher engagement, Research infrastructure	Italy
7	Deutsches Klimarechenzentrum	EOSC Future, NFDI, NFDI4Earth	Data management, Metadata repository and discovery service	Germany

ID	Institute	Project	Background	Location
8	National Institute of Geophysics and Volcanology	ICDI, Skills4EOSC, EPOS	Physics, Geoscience, Earthquake mechanics	Italy

Discussions among participants with a diverse set of roles and responsibilities helped us map out what is coming in terms of functionality in Research Data Infrastructures (RDIs) in the following years. This was to ensure a complete picture of what skills and needs must be covered by our materials and activities, and also what functionalities of the technical platforms should be considered.

First, participants were given an introduction to the NFDI and NFDI4Earth projects, with a focus on education and training activities. The participants were then asked to introduce themselves and talk about their background and current projects. In the invitation email, the topic of the meeting was explained, so after the introduction, the participants started the conversation according to their related knowledge and experiences.

This part of the discussion was fully recorded and transcribed with the permission of the participants. This report summarizes the main results of the discussion.

2.4. Challenges

One of the issues that were brought up several times in the discussion was **metadata and the findability** of educational materials. Participants collectively agreed that there is plenty of educational content, but it is neither findable nor visible. The absence of a fitting standard metadata scheme for educational objects can be one of the reasons for this issue. Efforts have been made by similar projects to develop metadata schemas for educational material, yet an accepted standard schema, that provides a balanced number of fields big enough to make the material findable, yet small enough to encourage creators to use it, is lacking. Educational content, like any other type of data, gets copied and republished or sometimes creators prefer to set up their own portals to publish content instead of using platforms and services that already exist. This is the nature of the Web and this phenomenon is hence inevitable; therefore, it is important to make the material available in a way that is easy to reach the public, independent of how they look for such content (e.g., through a simple Web search, specialized Open Educational Resources (OER) catalogs, or an institutional learning management system). Findability becomes more challenging when dealing with non-text-based content, which is the case for plenty of educational resources.

The matter that was mentioned the most during the discussion was the **need for collaboration with similar projects**. Exchanging experiences and providing mutual feedback is the key to

establishing a sustainable service. Although there is a large amount of educational content on open science and research data management, many projects are creating their own replica, which indicates a lack of synergy between similar initiatives. This project-driven approach is evidently not sustainable, as education and training platforms generally lack maintenance and are eventually shut off some time after the end of the funding period. The importance of not reinventing the wheel, instead leveraging existing materials and services from other countries or organizations, should not be overlooked. Besides establishing a network of similar projects, it is essential to establish an effective relationship with the research community as well as with data providers and policymakers. Training younger researchers is different from decision-makers; therefore, we must pay attention to terminology and to semantics used to train different target groups from early on.

The **reusability of the training materials** was another discussion topic. There are tons of training materials in the form of PowerPoint slides whereas training encompasses more than that – exercises and the corresponding datasets, readings, and tools for evaluation (both self-evaluation and for use by a trainer). However, these training materials can be of high quality and still won't be reused if information on *how* to use the material is not provided. To enable the reuse of such material, they must be presented as packages including additional information and resources including individualized instructions, target groups, guidelines for slides, exercises, and practice data – plus the already mentioned metadata for findability. Providing training content as a reusable package is a superb way to boost the number of quality courses among platforms. Throughout developing training packages, their application should be taken into consideration. Training materials could be used in two different ways. Either by other educators to adopt them and use them as part of their classes, or as self-training materials. To improve the adaptation of the developed materials, training packages should be created in a way that is useful for both educators and self-learners without resulting in double work. It is worth noting that training is only one of the tools needed to raise competencies. Training, combined with involving researchers in the process of data generation and data curation, can provide the highest efficiency; consequently, educational services can be most effective when researchers are encouraged to participate in these processes as their routine work, and this can only be realized when, in addition to the outputs of research, data generation and curation are acknowledged and rewarded as a scientific work. Only then can we train people who are capable of data management, data generation, and data value production, in the research context within the infrastructures that we are offering them.

Personalized learning is currently one of the major and most talked about subjects when it comes to education and training, which was also addressed during the meeting. The following challenges and methods were discussed to address the needs of Research Infrastructure (RI) users in terms of customized learning. Although educational systems are loaded with many traces of failed tries by learners, and far more sophisticated training can be offered based

on the evidence of these failed efforts, the traces learners leave behind often get neglected instead of following up and analyzing them. Solutions like applying micro-learning strategies are helpful to provide learners with the right training at the right time. Facilitating informal learning is also an important target for RIs, as it is essential to effectively support their users, who are mostly already in the professional world, and are possibly more interested in practical information. One of the reasons that personalized learning has not been widely enabled is that most of the infrastructure efforts are top-down efforts, and quite often there is an obvious gap between top-down and bottom-up approaches. That is, the creators of RIs are not in touch with the users, so they are not aware of their (personalized) needs. Currently, there is an evident mismatch between RI providers and the actual needs of graduate-level operators like students or scientists.

One major challenge brought up in the discussion was the **ever-changing nature of the digital world**, which also applies to research data management. Therefore, by the time competence objectives are crystallized, they might be outdated. As a solution, the community needs flexible, modularized curricula developed in a way that allows maneuvering and gives a frame without a rigid structure, so the content can always be redesigned ensuring a pathway of growth.

Finally, two other issues were raised briefly because all the participants unanimously agreed upon them. One is the **missing middle-level open educational resources** between very generic OERs and the highly domain-specific OERs, which is equivalent to information gained at the end of a bachelor's studies or the beginning of a master's studies. The other raised issue was the **acronym fatigue**. The current overuse of acronyms by RI providers particularly affects educational content, leading to missing context, and making training material hard to follow.

3. Mapping of existing educational resources

The training materials collected and developed by M1.3 should provide elaborate information about the management of research data, particularly the application of FAIR principles. These materials should explicitly help ESS researchers manage and process their data efficiently and in a FAIR environment. This section describes how relevant existing materials have been retrieved and categorized. The corresponding catalog of existing materials is attached to this report in Excel format so that it can easily be searched and filtered (Appendi C).

3.1. Identifying relevant training topics

The collection of existing educational materials dealing with research data management for Earth system scientists was done using Google and Bing search engines. For this purpose, the following keywords, as well as their combination, were used:

- Class
- Competencies in Earth System Sciences
- Course
- Data science online course “spatio temporal”
- Earth and environmental science
- Earth Science Data Management
- Earth system data science
- Earth system data science
- Earth system data science curriculum
- Earth system science
- Earth system science
- Earth system science training
- Education
- Exercise
- FAIR
- Free
- Learning
- Lecture
- Online
- Online course
- Online learning
- Open course
- Research data management
- Research data management curriculum
- Spatio-temporal
- Spatio-temporal analytics
- Spatio-temporal data literacy
- Spatio-temporal data science
- Spatio-temporal data science
- Spatiotemporal
- Training
- Tutorial

Considering that the Google search engine often shows search results regionally, after the initial search, the advanced search option was used to repeat the search for the exact keywords in English-speaking countries such as India and the United States.

The mapping of existing educational content was conducted between June and August 2022. The search results are 79 educational objects, ranging from single webpages and YouTube videos to a whole course or curriculum, which can be seen in Appendix C. The topics of these

materials include the basic concepts of research data management up to complex analysis and research data management specific to ESS researchers. 64 of the 79 collected resources cover RDM tailored for one or more subsystems of the Earth system and the rest general RDM training. Our effort was to include a limited number of the most complete RDM training packages in the list. 46 of the 64 ESS-relevant are suitable for all the ES subject areas.

This list is by no means exhaustive but provides a snapshot of the available open and/or free educational materials and their features regarding research data management for ESS researchers.

3.2. Categorization

The collected educational materials were categorized based on different criteria. These materials are categorized according to subject areas of ESS, stages of research data management, difficulty, type, and quality of presentation. In addition to this, more information such as the license, and contact information of the content creator was collected so that the collected content can be evaluated best and prepared for the NFDI4Earth educational catalog.

The earth system is composed of four subsystems that are the geosphere, the hydrosphere, the atmosphere, and the biosphere. Figure 1 shows the share of educational materials collected for each of these subsystems.

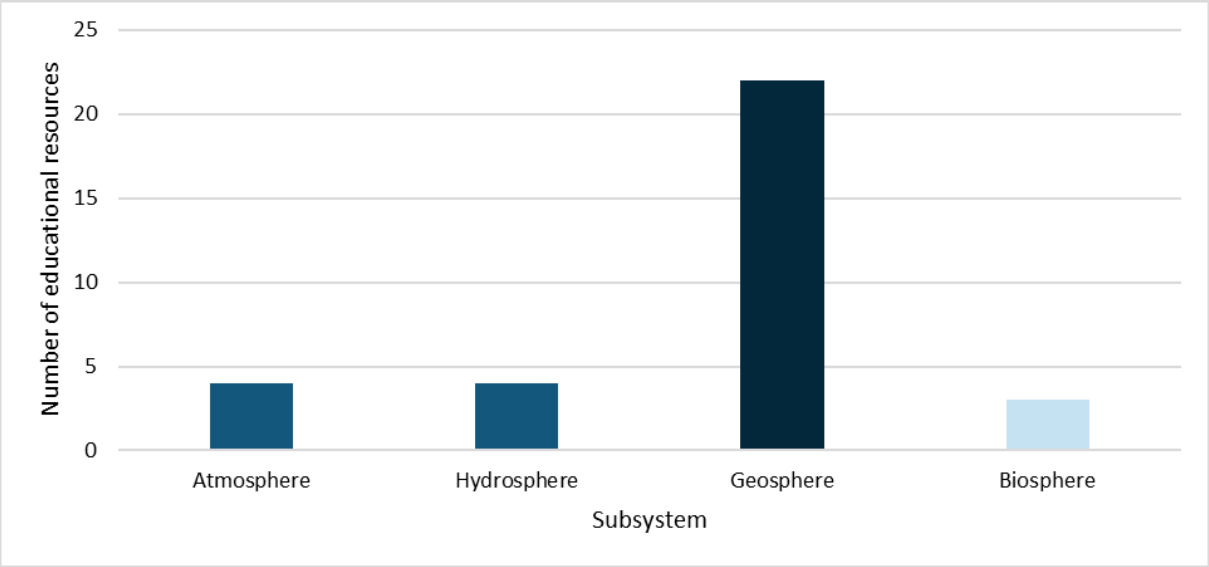


Figure 1: Number of educational resources identified for each of the earth subsystems

Each of the four subsystems was divided into the following subject areas proposed by the DFG to describe the materials more precisely [5]. The collected materials will be labeled based on one or more of these categories. To ease the evaluation process, codes were defined to tag

each subject area. The codes are based on the earth subdomain each of the subject areas refers to. For subject areas not present in the mapped material, the codes are based on their earth subdomain; given that some resources meet the needs of the whole ESS community and some only one of the subdomains of the earth. In the following, you can see the list of subject areas, as well as the codes:

AT.1 climate research
AT.A atmospheric science
BI.A paleontology
GE.1 geology
GE.2 geochemistry
GE.3 geophysics
GE.4 geoinformatics
GE.4 remote sensing
GE.A cartography
GE.A geodesy
GE.A geography
GE.A human geography
GE.A hydrogeology
GE.A mineralogy
GE.A petrology
GE.A photogrammetry
GE.A physical geography
HY.1 hydrology
HY.2 oceanography
HY.A integrated water resources management
HY.A limnology
HY.A urban water management
HY.A water chemistry
HY.A water research

Figure 2 shows the distribution of these educational materials according to the different subject areas of earth sciences. Basic training for managing ES data, usable for all the subject areas has the greatest number of courses, and more complex training tailored for one or a few subject areas is mostly available for geoinformatics and remote sensing.

Each of the training materials was tagged according to the stage of research data management that they cover. Using different resources, the following steps were identified for tagging the educational materials [6, 7]:

RD.1 Planning data management, including DMP, DMP tools: type, form, file formats, standards,

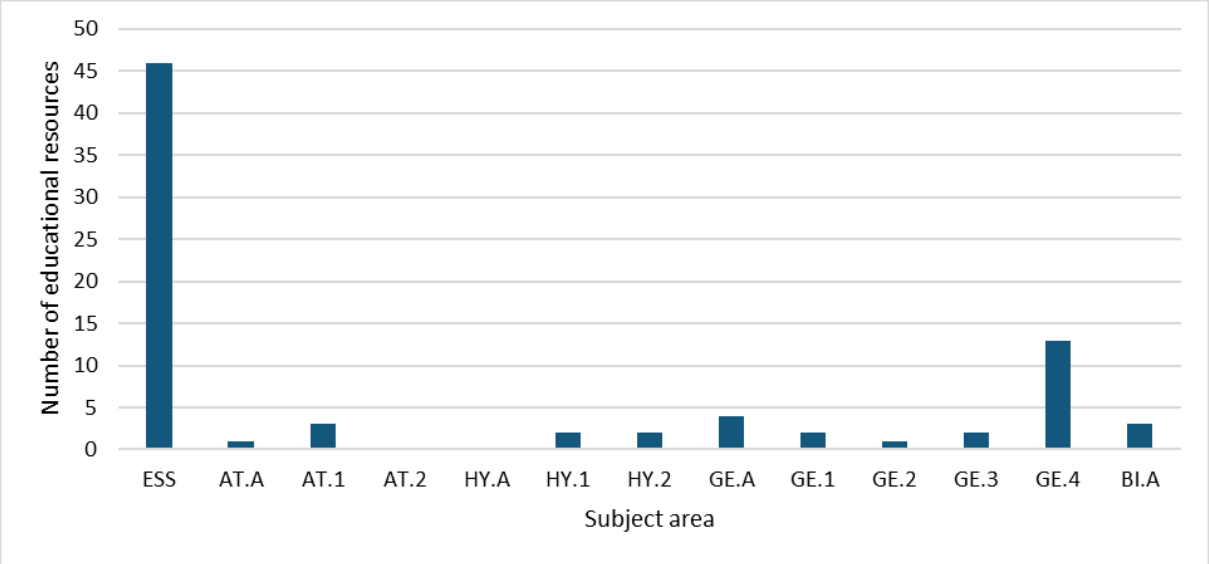


Figure 2: Number of educational resources identified for different subject areas of earth sciences and educational resources beneficial to all ESS researchers

preservation format, size, sensitivity, metadata, research objects to be tracked and documented

RD.2 Data acquisition and collection, including locating existing data, data collection and management, capturing and creating metadata, clean up

RD.3 Data processing, including entering, transcribing, checking, validating, and cleaning data, anonymizing data, describing data, managing and storing data

RD.4 Data analysis, including data security, interpreting, & deriving data, producing outputs, authoring publications, preparing for sharing, data analysis, statistical methods, data visualization & exploration options, and research computing

RD.5 Data preservation, including data repositories, data storage, backup & archiving, migrating to the best format & medium, creating metadata and documentation

RD.6 Giving access to data, distributing data, sharing data, controlling access, establishing copyright, promoting data

RD.7 Reusing data, including follow-up research, new research, undertaking research reviews, scrutinizing findings, teaching & learning, FAIR data

RD.8 Data disposal

ESS and FAIR were also added as two stand-alone codes to identify and evaluate materials that are suited for the whole community and the materials that focus on FAIR principles, respectively.

Resources on research data management for ESS researchers are typically generic and at an introductory level. Figure 3 displays an overview of educational materials categorized based on

different stages of research data management.

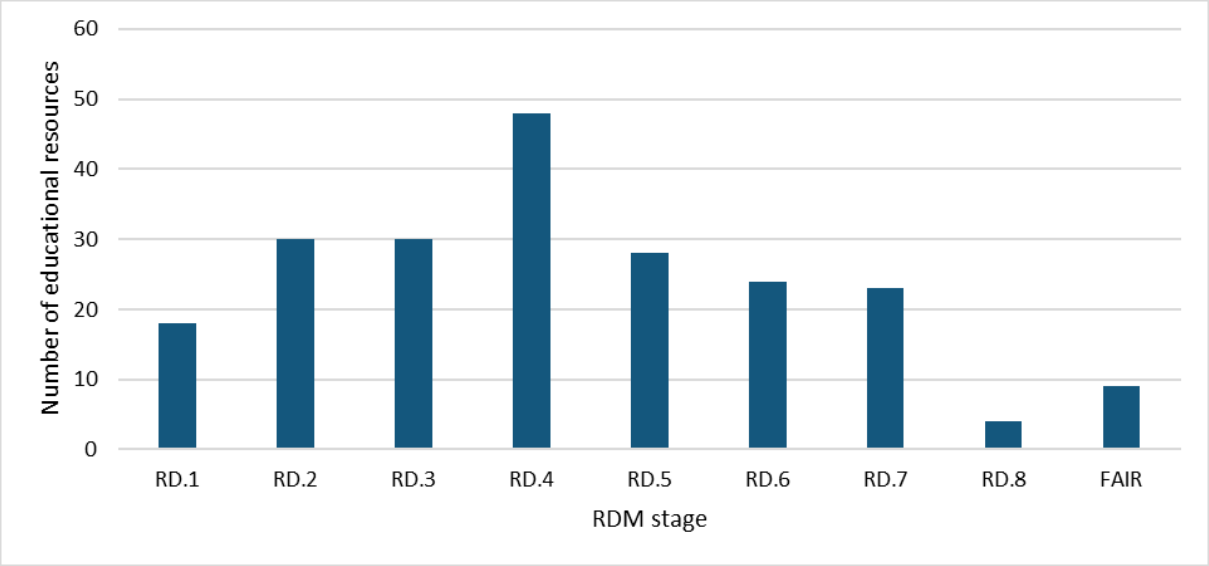


Figure 3: Number of educational resources identified for each stage of the research data life cycle and educational resources with FAIR-related topics

This diagram indicates that the largest amount of educational content is focused on data analysis for ES data, although this does not mean that this amount of educational content is sufficient or detailed enough because data analysis is much broader than other steps of research data management, and also data analysis methods for each of the ESS subject areas can be very different from other subject areas.

4. Future work

4.1. Synergy building

To avoid duplication of effort, M1.3 should build on experiences from comparable projects, such as Skills4EOSC. We will continue planning experience exchanges and gap analysis discussions with experts from similar projects. The expert panel who contributed to the discussion on the 1st of December will be the first members of this synergy group. The expert panel will continue to elaborate on the needs briefly discussed in the first gathering. The suggested topics for the next discussion are:

1. How to improve the findability of educational material and development of a single metadata schema for training content, acceptable for all subject areas. Currently, many acceptable suggestions are available from different projects, yet there is no agreed-upon

and frequently used metadata schema, and there are several competing open formats to enable integration of the actual content into different learning management systems.

2. Methods and approaches for activation of personalized learning, both in online and presence training.
3. Designing flexible and adjustable curricula to cope with rapidly evolving topics.

4.2. Metadata objects for educational content

For the NFDI4Earth training catalog, a profile with 26 metadata elements is planned. The elements of the metadata for educational material are chosen based on categories repeated in numerous stories (Section 2.2), metadata provided by owners of the mapped existing educational resources, the IEEE Standard for Learning Object Metadata [8], and Metadata Schema for Research Data Management Training Materials [9]. Due to these existing metadata schemata and vocabularies, we foresee the re-use of the corresponding terms without the need to reinvent the wheel and create an NFDI4Earth-specific schema. The following metadata elements for describing the educational material in our e-catalog will be used, more details on these elements will be provided in deliverable 1.3.12.

- Title
- Subject
- Description
- Publisher
- Date
- Learning resource type
- Identifier
- Source
- Language
- Rights
- Workload
- Requirement/prerequisite
- Difficulty
- Keyword
- Login required
- Code
- Interactivity
- Subtitle
- Version
- Authors
- Last update
- Contact
- Logo
- Dependencies/different modules

4.3. Input for the NFDI4Earth training e-catalog

The mapped educational material and additional information about them (Section 4.2) will be used to improve their visibility through the NFDI4Earth training e-catalog. All resources collected for the e-catalog will be assigned with their determined subject area(s) and their role(s) in the research data life cycle, which boosts the findability and reuse of the material.

4.4. Impact on the measure

This deliverable is intended as a foundation for structuring our content and services. The next step will be to evaluate, harmonize and strengthen the existing training materials. Further, evaluated materials will be published as an e-catalog, and some will be adapted to the NFDI4Earth curriculum. The NFDI4Earth curriculum will be a tailor-made training plan based on the outcomes of this work. The needs will be adjusted for the different ESS subject areas. Training needs similar for different subsystems will be addressed first. Educational projects will be funded through annual calls to fill the identified training and technical gaps.

5. Conclusions

The user stories, the survey within NFDI4Earth, the review, and the focus group discussion have all underlined the need for research data management training in the ESS community. A general insight from the preparation of this first deliverable is also that a better understanding of the application of data management in Earth sciences is required to develop training that addresses the tasks that our community performs during the research process. Therefore, the priority of M1.3 will be on preparing and conducting Earth data management training. These pieces of training should challenge the users and encourage them towards doing practical work and implementing FAIR data management methods in their research. Arguably, more accurate information can be obtained only by involving learners in the research data management process (Section 2.2, Appendix B). Only after empowering the users with tailored RDM skills and encouraging them to apply those skills in managing their research data, and then redoing the need assessment, we can move from general answers to more detailed topics.

The need for educational materials that can be used by trainers was also visible in the collected user stories and mentioned in the group discussion. The issue can be solved by improving the findability of existing material and providing training as packages (Section 2.4). Our effort is to improve the findability of collected OERs by the NFDI4Earth e-catalog, which provides rather comprehensive and detailed metadata for the collected material. In addition, the materials will be evaluated and quality controlled. The NFDI4Earth curriculum aims to help improve the reusability of the data by clustering OERs from different sources, and modular presenting them instead of standalone pages. Furthermore, the curricula metadata provides learners and teachers with additional information, such as learning objectives and course prerequisites.

Creating new educational content by M1.3 and the educational pilots will be concerned with the missing middle level between generic RDM tasks and highly specialized tasks that are performed only by a small number of experts. (Section 2.4).

The need for collaboration among similar projects and addressing the digital training needs by an international community effort is urgent. Our goal is to make the focus discussion a regular information exchange, our panel of experts was eager to continue the collaboration with the M1.3 team and very well aware of this need.

References

[1] L. Bernard et al., "NFDI Consortium Earth System Sciences - Proposal 2020 revised," 2021, doi: 10.5281/ZENODO.5718944.

[2] Salome Scholtens et al., "Final report: Towards FAIR data steward as profession for the lifesciences. Report of a ZonMw funded collaborative approach built on existing expertise," 2022, doi: 10.5281/ZENODO.3471707.

[3] J. Wagemann, S. Siemen, B. Seeger, and J. Bendix, "A user perspective on future cloud-based services for Big Earth data," International Journal of Digital Earth, vol. 14, no. 12, pp. 1758–1774, 2021, doi: 10.1080/17538947.2021.1982031.

[4] B. Radosavljevic et al., "Report on the Survey of Digital Data Management Practices at the GFZ German Research Centre for Geosciences," 2019, doi: 10.2312/GFZ.B103-19029.

[5] <www.dfg.de>, DFG, German Research Foundation. [Online]. Available: https://www.dfg.de/en/dfg_profile/statutory_bodies/review_boards/subject_areas/index.jsp (accessed: Dec. 22 2022).

[6] Research Data Management [@Harvard]. [Online]. Available: https://researchdatamanagement.harvard.edu/ (accessed: Dec. 22 2022).

[7] What is research data management? | uOttawa Library. [Online]. Available: https://www.uottawa.ca/library/research-data-management/what-research-data-management (accessed: Dec. 22 2022).

[8] 1484.12.1-2020 - IEEE Standard for Learning Object Metadata. [S.l.]: IEEE, 2020.

[9] K. Biernacka et al., "Metadatenschema für Schulungsmaterialien zum Thema Forschungsdatenmanagement", 2020, doi: 10.5281/ZENODO.3760398.

Acknowledgements

We want to thank all user expert group members: Sibylle Hassler, Ivonne Anders, Jonas Kuppler, Veronika Grupp, Thomas Rose, Ralf Klammer, Auriol Degbelo, and Christin Henzen, for collecting and working on user stories.

Appendix

A. User stories

Table: User stories expressing the training needs of the ESS researchers and educators; persona requirements will then translate to certain functionalities and features

Persona					**Need**	**Reason/goal**
Name (**Gender**, **Age**)	**Occupa-tion**	**Discipline**	**Institution**	**Education level**	**has a certain wish/need/requirement towards the N4E**	
Marry (female, 24)	Master Student	Geo-informatics	Research institution	Bachelor degree	wants to learn about uncertainty handling.	
					is looking for methods to include **data management** into Master/PhD courses.	
					wants to have Outreach Materials on **NFDI4Earth**	
					needs tools to combine teaching-**materials** to e.g. a presentation	
Hannes (male, 40)	Lecturer, Senior Scientist	Paleontol-ogy	University	PhD	looks for **RDM** training **materials**	because he has to setup the course in his institute
Gregor (male, 39)	Researcher,Committee member	Mineral-ogy	University	PhD	looks for inspiration for **module descriptions**	because he has to re-design the Soft Skill module to include RDM
Martin (male, 52)	Employee		Forschungs-daten-kontaktstelle Hamburg	Higher Education Qualification/A-level	wants information on **RDM** basics in german **language**.	as teaching material
Martin (male, 52)	Employee		Forschungs-daten-kontaktstelle Hamburg	Higher Education Qualification/A-level	wants information on **NFDI4Earth** and **RDM** basics in english **language**.	as teaching material
Berta (female, -)	Employee, Research Support		Forschungs-daten-kontaktstelle Hamburg		needs reusable educational **resources** that can be adopted to different target groups and users (handbooks, slides, e-learning)	for education of students and technical staff
					is looking for ready-to-use **training data** sets	for teaching RDM and/or Data Science at university
Birte (female, 45)	Researcher,Professor	Geodesy	University	PhD	is looking for self-teaching courses on **FAIR** ESS **RDM**	for their PhD students
					is looking for teaching **materials** in video **formate**	because is visual learner
					is looking for discipline specific **data management** teaching **materials**.	
					is looking for **FAIR** teaching **materials**, which can be easily re-used/ be combined	
					is looking for a trainer in **RDM**	to give a workshop at their institution

B. Existing educational resources

For the detailed catalog of relevant existing educational resources, see the attached Spreadsheet `NFDI4Earth_D1.3.1.appendix_existing-educational-resources.xlsx`.

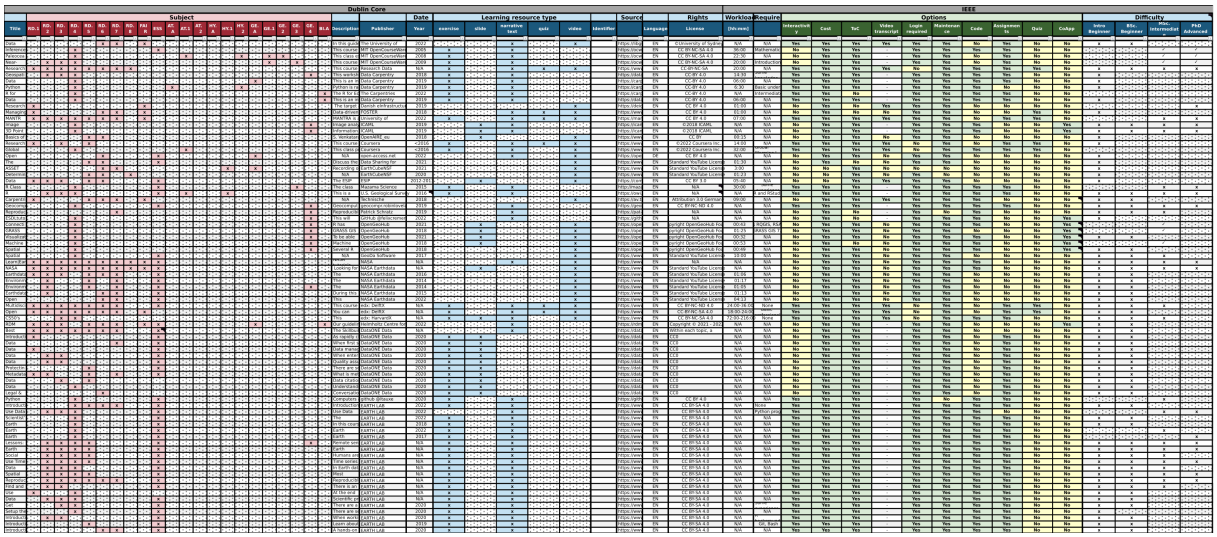


Figure 4: Preview of the collected educational materials and the specifications associated with them