ing.grid preprints

RESEARCH ARTICLE

How to teach good research data management to next generation researchers?

Syed Ashfaq Hussain Shah [1], Frank Petzold [1]

1. Chair of Architectural Informatics - TUM School of Engineering and Design, Technical University of Munich, Munich.

Date Submitted:
2024-03-18

Keywords:
Research data management, Information infrastructure, Federated research infrastructure, Collaborative research, Teaching, Training, Methods, Curriculum development, Active learning

Data availability:
Data can be found here: mediaTUM

Software availability:

Corresponding Author:
Syed Ashfaq Hussain Shah
syed.Hussain@tum.de

Abstract. These days research work is subject to comply with FAIR principles. Additionally, it is subject to the practices of Open Science. Different stakeholders e.g. DFG are setting the goals of reproducible research work. This not only requires adequate handling of data but also the record of related information and practices during the research work. In this way, different tools and workflows are being developed and suggested to achieve the goals of good research and its data management. Those tools and workflows facilitate researchers and ease the research management tasks e.g. by means of standardisation, automation of processes and record of corresponding information. The researches of now a days are interdisciplinary and work collaboratively where participants are located at distinct locations, belong to different domains and have different levels of competencies. In such cases, provision of tools and specification of workflows is not enough. Just like other management, good research data management is a skill that need to be taught to the researchers systematically with details. So that they could make right decisions where and when needed. As a result, the contents and the materials for the education of good research data management become important.

This paper presents contents and materials, approaches and skills which address the challenges of teaching and guiding good research data management in, in person, digital and hybrid environments. These were prepared for and imparted to the participants of collaborative research centre during the four years' period. The objectives of the presented case of teaching and guidance of research data management have been applied mostly than classic theory learning.

1 Introduction

In contemporary debate of knowledge organisation data is considered as foundation of knowledge hierarchy. In this hierarchy it is processed to form information and then knowledge. It is a bidirectional relation where knowledge also influences in determining about data [1]. Thus, its management is not about trying to capture all knowledge or wisdom instead only the considered one. The research, teaching and practice of knowledge organisation is mostly considered affiliated with Library and information science (LIS). There it is about describing, representing, filing and organising documents and document representations as well as subjects and concepts

[2]. The term knowledge management is defined differently in different fields. However, it is broadly considered relating to the use, create, share and manage e.g. data and information [3] In the field of research these activities are referred to as research data management (RDM) where these activities are about storage, organisation, documentation, and dissemination of research data. These are continuous activities which are needed to be performed during the period of research project [4]. These are important part of modern research activities regardless of the method e.g. quantitative, qualitative used for the research.

The research works have been influenced by the adoption and the development of technologies. This situation also lead to the changes and modernisation of teaching and guiding methods for research practices. In this case its data management. Those changes and state of the art are being discussed one by one in the remaining part of this section.

1.1 Research landscape

Modern researches are not merely desk, lab or field based work carried out by one, few or group of likeminded investigators. Instead, these have become more interdisciplinary and cooperative work of investigators from diverse fields and organisations. These are also not merely paper and materials etc. based work. Due to the technological development and its adoption these have become more than before data-driven works too. Apart from that researches have now been subject to as open businesses, reproducible and more useful to the practitioners [5], [6]. This situation also laid the foundation for the debate about novel research practices and agendas e.g. e-science, e-research [7].

In the case of such developments and progresses new genre of research organisations are being formed by the state, continental or other funding sources. Collaborative research centres (CRC) formed by the funding of German research council (DFG) are examples of such recent formations. In these centres data need to be FAIR (Findable, Accessible, Interoperable, and Reusable) [8] even within centre, researchers are required to have understanding of systems beyond institutional boundaries, policies and collaborations etc. The objectives of RDM teaching and training in these centres is usually to deliver knowledge and skills together with technical expertise in the officially recognised information infrastructure. In these centres participants face diverse range of systems and tasks and form collaboration spontaneously. Quick upskilling in these situations becomes crucial to increase compliance all together with reducing workload of RDM on the researchers' part.

1.2 Technology in research

The development and adoption of Information and communication technologies (ICT) in research are on one hand facilitating research work and its communication. On the other hand, these are raising the bars of expectations and requirements e.g. publication of intermediary results, preprints, FAIR compliance of the stake holders. These expectations and requirements are levying the workload of RDM in addition to the actual research. To relieve the researchers from these additional workloads existing research support systems are being equipped with RDM functions and the new type of RDM systems e.g. Virtual research environments (VRE) are being developed to support the modern research work [9], [10], [11], [12]. These RDM functions

and systems relieve users from technically difficult tasks, workflow based functions e.g. security, encryption in transferring, creation of metadata by means of automation etc. The evolving landscape of research and its supporting RDM systems are also causing a shift of responsibilities and replacing roles some of which through automation and some through introduction of new tools e.g. to seek advice about or search specific material replaced by metadata services, writing commands replaced by drag and drop user interface. The technological development and adoption for imparting training and material are also making shift from conventional teaching and guiding means and methodologies. For example, in person delivery and attendance of courses etc. are being replaced by the recorded multimedia contents and online interactive sessions and materials. New kind of learning platforms are being proposed [13]. Such digitalisation has eased the difficult tasks on one hand on the other hand it raised the importance of information and digital literacy among researchers regardless of their domain of research. Also the push of technical roles more towards IT professionals from librarians seems eminent or the technical education of librarians deemed necessary.

1.3 RDM Education and training

In practice researchers create and manage data during the course of research work. The librarians and data professionals offer them services and support during the research and after the work has been concluded. In this typical settings researchers creating and managing data do not necessarily belong to or come from LIS domain. Therefore, they are offered courses, trainings and workshops to equip for right data management skills and qualifications. Libraries are usually responsible to undertake these teaching and guiding roles. While to support research and offer effective training and related material require understanding of researchers' perspective. But the librarians and data professionals themselves often do not have in-depth experience of state of the art research. To bridge this gap and to make sure the right management of research data various efforts are being made. Some are relating to develop professions, some are relating to define contents, curricula and toolkit alike guides, some are relating to define skillsets and competencies in this domain e.g. [14], [15], [16], [17], [18], [19], [20]. Some are relating to improve the existing professionals by skilling them up/ reskilling and some are focusing on improving education and training of researchers and the trainers themselves through teaching e.g. [21], [22], [23]. The efforts are being made as part of regular academic discourse as well as part of professional education etc. [24], [25], [26].

These teaching practices are broadly categorised in didactic and pedagogic methodologies. Pedagogical methodology is claimed effective in reducing the time required e.g. to develop lab experience e.g. [27], [25]. Initiatives have been taken at different levels to spread awareness, teaching and improve practices of RDM e.g. NFDI[1] at the national level, RDA[2] at the international level.

This paper presents contents and materials, approaches and skills defined to address the challenges of teaching and guiding good RDM which were defined as part of I-Project during the first four years' period starting from early 2020 till the end of 2023. During this time some

1. https://www.nfdi.de/
2. https://rd-alliance.org/

additional relating events were arranged and materials were also served to the participants e.g. by the centrals administration of CRC and libraries of the universities. However, those are not considered in this paper.

The paper is broadly organised under five main themes. The first part introduces the settings i.e. the project itself and methodologies. The second part introduces RDM and its teaching and guiding methodologies. The third part presents the prepared contents and materials, the provision and training together with means of communication and distribution. The fourth part presents the processes and workflows. The fifth part presents the strategies and tactics to enforce the compliance for good RDM.

2 I-Project (Information infrastructure project)

Before the explanation of I-Project the explanation of its parent body TRR277 AMC is deemed necessary to keep in mind. AMC was founded in the beginning of 2020 for twelve years with funds from German research council (DFG). It is a trans regional, also referred to as transregio, collaborative research centre (CRC). It aims to explore Additive Manufacturing (AM) as a novel digital manufacturing technology for the construction industry in interdisciplinary cross-location research projects [28]. It is hosted by Technical University of Munich (TUM) and Technische Universitaet Braunschweig (TUBS). Multiple other institutions situated at distinct locations are also taking part in it. The whole AMC is divided in three focus areas namely Materials and processes, Computational modelling and control and Design and construction. In these focus areas twenty-two interdisciplinary research groups have been formed. All those groups have been dealing with both digital and non-digital research artefacts. The works of some groups are data rich, some practice oriented, some material and some real world object oriented etc. Over 120 members have taken part in it. The total number of projects and participants varied over time.

The research workload of AMC is divided in projects and work packages which are distributed among various teams and institutions regardless of their locations. The participants come from different domains with varying skills and competencies. During the course of investigations participants collaborate in different ways and form interdisciplinary teams depending on the objectives over time. The cooperation and interdependence of work are predefined in the project proposal documents. At times cooperation are also defined during the course of investigations. They use and generate heterogeneous, large amount of data which ranges from surveys, lab experiments, simulations, data models to software code and hardware design as well as real world objects etc. The participants of AMC could be identified as undergrad, graduate/ PhD., post graduate, postdoc., professors and emeritus based on their qualifications. They are from different demographic and geographical backgrounds. However, they are assigned different roles depending on the arrangements and participations.

The Information and infrastructure project of AMC was named as I-Project within AMC. In rest of the document Information infrastructure project will be referred as I-Project and AMC as CRC interchangeably. The I-Project was defined as service project to serve and support all information and data infrastructure aspects of CRC.

DFG emphasises such projects as “An effective, modern and secure information infrastructure is essential to excellent research.” [29].

One aspect of I-Project comprises services facilitation, fostering of internal collaboration and the construction of the collaborative environment to support research projects in handling, storing and sharing data. This includes backup and archive services for long-term storage of research data. Another aspect of this project is an efficient and long-term usability of research data generated, support researchers in RDM activities and demand-oriented targeted transfer of data to university libraries as well as intensive exchange with initiatives e.g. NFDI4Ing[3].

In addition to the provision of technical services and infrastructure, the conception of a research data management plan, conceivement and implementation of a training and qualification program for the fundamentals of reproducible research and data management have also been core components of I-Project.

While carrying out these tasks and processes I-Project is cooperating and working together with the libraries and IT centres i.e. Leibniz Supercomputing Centre (LRZ) and Gauss-IT Centre of both universities. It employed TUM Workbench[4], Globus[5] together with Data Science Storage (DSS)[6] and mediaTUM[7] as official RDM systems and platforms. For details about adoptions and implementations of these systems please refer to [30], [31].

The project was started in early 2020. During this time project also survived through lockdown situations imposed due to Corona related issues and measures.

3 Methodologies

To define topics and contents, strategies and practices, a review of practices and literature produced by both academia and industry was carried out on one hand. On the other hand, official information infrastructure, research work proposed in the funding proposal of the CRC, policies and requirements by the stake holders and communities e.g. DFG were reviewed and analysed. Both of these reviews were then complemented with the review of possible methodologies and technologies.

As a first step fundamentals were determined as a minimal information and practices to be applied for good RDM. In the second step those fundamentals were translated to define an agenda of RDM for CRC. Then based on that agenda teaching and guiding materials and methodologies were defined. Those were further improved while considering the role and qualification of participants.

The training and support materials, concepts and approaches were developed for in person, digital medium as well as for hybrid environment e.g. in lockdown time. The activities were defined for individuals to the group and to the whole CRC while considering small isolated work unit to a complex interdisciplinary task which might be distributed among multiple teams. The cooperation and collaboration of the participants on adhoc bases were also considered into account.

3. https://nfdi4ing.de/
4. https://github.com/eWorkbench/
5. https://www.globus.org/
6. https://doku.lrz.de/data-science-storage-10745685.html
7. https://github.com/mediatum

From each project at least one representative was assigned the responsibilities of all the RDM activities within the project/ group and to communicate and coordinate with the I-Project. Person with this role/ responsibilities was designated as data manager/ steward and rest of the participants were considered as member of the project/ group. The main medium of communication and instruction was English. However, in some cases German language was also used for better understanding of the native German speakers.

In contrast to typical general guidance, training activities and courses which are usually based on theories, general purpose learning and qualification, the guiding and training activities of I-Project were oriented more towards understanding of formal practices, quick adoption and implementation by the participants in their actual and immediate research work. Therefore, the teaching and guidance of I-Project presented in this paper were applied mostly than classic theory learning. And those were subject to update due to change in the requirements, upgradation of systems etc. Those changes were occurring sometimes as part of the plan and sometimes spontaneously caused by others or due to other reasons.

4 Good RDM fundamentals

As a result of literature and practices' review, it was defined that a good RDM will comply with FAIR principles. It will further, if apply, adhere with open data/ open science practices. And, if circumstances permit and possibilities apply, it will support repeatability of research practices and/ or reproducibility of results.

It was further hypothesised that it should not be considered enough in the face of evolving landscape of possibilities of research work and technological development. The rapidly changing and evolving phenomena could also cause unprecedented changes in the requirements and expectations of stake holders and communities. Thus, it should also not be overlooked. For example, it was experienced that DFG released guidelines for research data in 2015. But in 2021 those guidelines were revised by DFG [32]. The newer guidelines were more expressive and detailed than the previous version to comply for research work. RDM activities are also suggested as a mechanism to avoid misconduct and empower researchers against malpractices in research [33].

Therefore, in the fundamentals of RDM, it was also included that the RDM should be resilient and efficient to comply with changing requirements and challenges. May those demands be set forth by funding agencies or other stakeholders e.g. research organising bodies, communities etc. Thus, the fundamentals may also evolve.

5 Agenda of RDM

It was identified that in practice data transit through different user spaces and take place at range of devices e.g. lab sheet to lab terminal to collaborative environment to publically accessible systems and places. At each transition and storage place some data management practices are applied. These are applied e.g. as a policy or voluntarily. It was assumed therefore, that every participant would have some kind of data management understanding depending on one's own exposure to data and related training in the past. However, there would also be a confusion or

misunderstanding between general practices of data management and the state of the art research data management practices and requirements. To deal with such situations and to differentiate between both of these practices the agenda of RDM was defined in two parts, Fundamentals of data management, Research data management. It was defined in a way that the transition to RDM become relevant, incremental and easy. Table [1] presents a short list of topics of both of these parts.

5.1 Fundamentals of data management

As fundamentals, data management was considered that, name the entities consistently, do versioning, identify components and their organisation. Do the documentation e.g. in log, and README type of files. All that should be bundled or packaged in a specific manner so that those are uniquely identifiable within the system and/ or setup.

And, since the digital means are also available, one could employ digital tools and systems e.g. to create and to store the data on need basis.

5.2 Research data management

But the research data management was defined as more than just usual data management. It was defined that in addition to basic data management there are further practices and requirements that need to be taken carefully. For example, for documentation, practices should also be documented by means of metadata, data management plan, layman/ user and technical guides. Provenance and provisioning information should be maintained. It should comply for long term review and use, universal unique ID should be assigned. Data should be attributed e.g. with access and rights. Where necessary anonymization and pseudonymisation should be applied. The tools, systems and setting employed by the researcher should comply with RDM compliant standards/ requirements. To sum up, it was suggested that it is not just a usual data management, it is about a well thought change control agenda.

Based on the defined RDM agenda, teaching and guiding methodologies and materials were defined. Plans were made to deliver those and to enforce practices and compliance accordingly. Improvement methodologies and events were followed respectively. All those are explained as follows.

6 Teaching and guiding methodologies

As a first step policy for research data management was defined. The policy document outlined the principles and rules of research data related conducts [34]. Table [2] presents key topics of RDM policy for CRC. While at the same time efforts were also made to define a custom data management plan (DMP). Usually DMPs consists of questions and the input fields for the responses. In this case DMP was defined together with possible data workflows/ data life cycles, corresponding user spaces, processes/ functions and the required information documentation during the active state of research [35].

Then the process of preparation of teaching and guiding materials was initiated. The material was produced in printable and digital navigable document, presentation slides and multimedia

Fundamentals of data management	Research data management
• Naming convention • Versioning • Identification of independent component and organisation • Documentation – Logging – README/ Docs/ Comments • Packaging/ Bundle • Unique ID • Digital means – Purpose/ use based data processing tools – Purpose/ use based system	• + Fundamentals of data management • Documentation of practices – Metadata – User/ Technical guide – Data management plan (DMP) • Maintenance of provenance and provisioning • Compliance with long term archive compliant format • Data anonymisation and pseudonymisation • Universally unique IDs. – ROR, ORCID, DOI … • Attribution e.g. Licencing • (Standard) procedures/ workflows • Compliant digital means – Compliant tools – Compliant systems * A well thought change control agenda

Table 1: Agenda of RDM

content formats. It was accompanied by supplementary materials. Interactive sessions and evaluation processes in writing, in one to one, in group, in person, online and in hybrid modes were carried out. All that materials and events were enriched with screenshots, screencasts, diagrams, illustrations, animations, glossaries and definitions etc. These included own design and production as well as third party images and contents.

7 Guiding materials

The topics and contents of guiding materials were broadly classified in two groups, Official RDM platform and infrastructure and Research practices. The materials were assigned titles depending on the categories to which they belong to or the purpose they serve for the participants of CRC. The contents were arranged to support a personal learning journey. For example, the material which aims to introduce e.g. system or concept was categorised as introductory material. The material which aimed at the user role in the project or CRC was categorised as user role specific. The material which was prepared to explain and address a particular topic e.g. task to create DMP, tools to log the information, use case to publish data were categorised as topic specific. The other categories of materials were about users' learning phase and pace. Table [3] presents a list of categories of guiding materials.

7.1 Official RDM platform and infrastructure

The printable material centric to the use of official platform and infrastructure was released under the titles of User guide, Quick start guide and Cookbook. As the names imply, Quick start guide

- Purpose/ Motivation
- Target audiences
- Applicability
- Research relevant data and outcomes
- Official RDM platform and infrastructure
- Official Data Management Plan (DMP)
- Information and communication relating to research data and practices
- Use of other tools and systems
- Categories/ Roles and responsibilities of the participants
- Types of research outcomes/ data
- File types and formats
- Long term archiving
- Naming conventions including versioning
- Metadata standards
- Applicability in relation to other policies
- Appendices
  - Data and user migration and profiling
  - Decommissioning of system
  - Use cases of research outcomes and DMPs
  - Workflow for publishing

Table 2: Key topics of RDM policy for CRC

- Introductory
- User role specific
- Topic specific
  - Task/ activity
  - Tool
  - Use case
- Learning phase specific
- Learning pace specific

Table 3: Categories of guiding materials

was offering immediate solution with a list of steps and short explanations while assuming that either reader is already aware or does not require to know supplementary/ detailed information due to some reasons. The purpose was to expedite the on boarding process and reduce the time and efforts. Table [4] presents key topics of Quick start guide.

- Introduction of tools and platform
- Benefits of official platforms and services
- Access to the platform
- Upload and Manage VRE Storage as network drive
- Access and right management
- Event management via appointments/ calendar
- DMP creation and management
- Frequently Asked Questions (FAQs)
- Useful links and contacts
- Glossary

Table 4: Key topics of Quick start guide

The Cookbook was prepared as step by step instructions for known outcomes. It consisted of two main parts, getting started and use cases which followed by some common guidelines. It was offering immediate recipes like solutions of the issues while relieving the users from the overhead of analytical and decision burden caused by multiple tools or possibilities of workflows

etc. Each topic consists of two to three parts i.e. Problem, Solution, Discussion. In some cases, more than one solutions were also proposed for user preferences. Table [5] presents key topics of Cookbook.

- Definition of research data
- Good research data management
- DMP and data lifecycle
- Accessing RDM platform/ system
- Start using RDM platform/ system
- Mandatory elements of the platform/ system
- Naming conventions and best practices
- Organisation of files and folders using storage system
- Getting and storing file/ folder-like data
- Updating and versioning data
- Relate/ refer work done under different project/ file
- Sharing and collaboratively working
- Documenting information relating to the data and activities
- Data management plan (DMP)
- DMP allocation and frequency
- Actions relating to the RDM during the course of a research project
- Use cases
  - Working on a work package
  - Working on a work package collaboratively
  - Working for a goal/ objective e.g. publication
  - Meetings and Workshops
- Hints to identify research data when conducting research and meeting with co-researcher(s)
- Selected code of conducts
- References
- Useful links and contacts

Table 5: Key topics of Cookbook

The User guide was consisting of more comprehensive and detailed contents. There were two variants of user guide based on the roles of participants i.e. one for the managers and one for the members.

Just like usual technical guiding materials, the contents of user guides were started with the expected competencies of the users, then requirements to access and work, and hardware and software requirements. However, the contents were organised broadly in two parts, conceptual part, user scenario and step by step practical part which were then followed by FAQ (Frequently Asked Questions), Useful links, Appendix and Glossary sections. In the conceptual part, theoretical and conceptual foundations were described so that it become logical and easy when users use the systems during their actual work. Those descriptions include the concept of virtual research environment (VRE), entities and functions, user interface design, navigation, access rights, context management and so on so forth.

In the second part, first user scenario was explained and then step by step guides and illustrations were followed. In this part it was explained that how user can employ and make use of offered services and tools to accomplish data related task or tasks related to day to day research activity. This part was consisting of interaction schemes and outcomes, metadata recording, data management plan and communication with external systems etc. Table [6] presents key topics of User guides.

Each part was further divided in chapters based on the themes they serve. Conventions to format text, titles or names of the components, controls and even third party software were used in consistent way. Where images were needed, the important aspects were highlighted using red

borders. Where there was a need to mention sequence or order, borders were also annotated with numbers in ascending order. These guides also included conventions, best practices, sharing and release scenarios using the platform, configuring the function of the platform e.g. for messages and communication etc.

- User competency requirements
- Functional requirements
- Hardware/ Software requirements
- Conceptual/ theoretical foundations
  - Common entities and functions
  - User interface design/ webpages/ forms
  - Navigation and exploration workflows
  - Definition of tools and controls
  - Access rights and privileges
  - Data and context management
  - Data input and persistence procedures
  - Built-in RDM features (e.g. UIDs, backups, communication and messages, metadata)
- Practical part
  - User scenario
  - User interaction schemes and outcomes
  - Metadata
  - Data management plan (DMP)
  - Network and communication protocols
  - Integration and interaction with external systems
- Conventions and best practices
- Documentation possibilities
- Sharing and release scenarios
- Configuring functions and features
- Frequently asked questions (FAQ)
- Access to software, support and services

Table 6: Key topics of User guides

7.2 Research practices

In this group of guiding material, the identification of data based on roles, states and types for example active data versus resource data, how to organise it for example in folder file alike structures were explained. There were situations when larger work packages were needed to be broken down in smaller workloads for e.g. collaborative work, assignment of the tasks, identification. The material to guide in such situations was also produced. Since researchers use different systems and had their own way to carryout the tasks, the roles of systems were also determined and explained. Materials relating to the adoption of standards, DMPs and their filling were created. Last but not least the guide about concluding the research tasks and projects in-line with FAIR and Open practices was also created. Where there was a need of graphical representation to describe and clarify the concepts, the contents usually were consisted of illustrations and animations. Table [7] presents key topics of guiding materials about research practices.

7.3 Multimedia contents

For multimedia contents first the work was carried out to define storyboards. Storyboards were defined based on the themes and topics. A logical sequence of scenes was determined based on those themes and topics. For example, to explain how to get a work package ready in an official RDM platform in one video the storyboard was consisting of scenes, Project basic information, Creating DMP, Storage and folder structures, Accessing guiding materials & help. Each scene

- Data identification based on: -
  - roles, states, types
- Folder/ file organisations and structures
- Identification of atomic/ independent workload
- Collaborative work
  - e.g. task based
- Distribution of larger/ collaborative work packages
- Specific and general rules for decision: -
  - e.g. research data, information necessary for reproducibility, to register data in RDM system, significant change in data to record
- Research and data support systems and their roles
- Adoption of common standards
  - e.g. DataCite, DOI, ORCID…
- Filling DMPs
- Applications of DMP: -
  - atomic workload to the whole CRC
- Data updates e.g. How & When
- Concluding research task/ project in line with FAIR principles and Open Science practices

Table 7: Key topics of guiding materials about research practices

in storyboard was consisted of Title, Action and Dialogue. Then the multimedia contents were created based on those storyboards. The length of the multimedia contents varied depending on the themes or purpose just like printable contents explained in the previous sections. However, they may be categorised as short videos spanning for less than five and ten minutes describing a particular concept, introducing a particular topic, explaining steps to quickly carryout a task etc. and long videos spanning over thirty minutes to more than ninety minutes explaining larger aspects or complete steps to carryout a set of tasks e.g. working on a work package in one composition. Multimedia contents were enriched by screen recording of the actual RDM system, analogy from the real world material and images etc. Table [8] presents key topics of multimedia contents.

- Introduction to CRC, systems and guiding material
- Getting started and introduction to RDM platform
- Introduction to data management in CRC
- Get started with RDM platform
- Create DMP per work package
- Initialising DMP
- Sharing and access with other project
- Using other tools e.g. Globus to upload/ access
- Preparing data
- Update DMP
- Long term archiving
- Publication
- Use case: working on a work package
- Use case: working for publication

Table 8: Key topics of multimedia contents

8 Supplementary materials

During the course of research, above mentioned approaches of guidance and materials are sometimes not sufficient for practical reasons. Therefore, a range of additional material is also required for better understanding, easy adoption, compliance etc. The same requirements applied

in the case of CRC. The key topics and themes for which supplementary materials were provided and the suggestions were made are as follows. Standard reusable contents, practical examples of relevant research data, standard external contents and resources. Just like practical examples, recommendations and hints were provided about tools, best practices and workflows e.g. how the coding, data creation tasks could be carried out including comparison and preferential aspects of updating data through code instead of making change in the original copy of data, scripting and compiled languages, controlled vocabularies to name the data as per its state etc. And when the data is created or acquired how its quality could be assured, how the practices could be digitally represented or documented and archived. Use cases of collaborations and cooperation were further defined and explained. Further tools for RDM were suggested during the later phase e.g. to address the shortcomings of the existing tools, service improvements. Their guiding materials were also prepared and provided. In order to further identify the needs and understanding of the users and improve the offerings of I-Project, a comprehensive survey was also conducted during the course of research work. Reviews, statistical reports based on the evaluations and feedbacks together with suggestions were prepared. Points of considerations and a checklist to conclude research work were also created and provided as supplementary materials. It may be noted that researchers in CRC were dealing with larger number of scenarios. The listed topics and examples do not cover all those scenarios. And the examples were provided in different ways and depending on the needs and situations e.g. some of them were embedded in existing materials like DMP and in survey form as a hint or reference, some were provided as independent material as document, some as part of presentation and some during the discussion and consultation. But, all those were for the actual work and relevant to the ongoing research in the CRC at that time. Table [9] presents a list of key topics of supplementary guiding materials and support.

Next the themes of interactive sessions are being explained.

9 Interactive sessions and workshops

The provision of resources, guiding and supporting materials were always followed by the sessions in which explanations were provided and the users were offered opportunities to raise questions and discuss the offerings and their concerns. The themes of these sessions could broadly be categorised as introductory sessions to introduce e.g. user guide or platform, on boarding and migrations sessions to help the user in adopting new system or features, topic based sessions to help users in addressing a particular issue, accomplishing a task, discuss a workflow or research data examples, how to kind of sessions to help users in an incremental way e.g. using a tool, carrying out a particular task, cooperative sessions to e.g. carry out a task in cooperation with the users or help users while monitoring and supervising the activity from RDM perspective, Q & A sessions to answer the questions of users which user may have in response of e.g. serving a new material, service or survey, preparation and verification sessions to help the users make preparation e.g. data to submit to the RDM system or maintain DMP, review and feedback sessions and more general sessions for example weekly consultation, progress meetings and summer schools. Sometimes, events were also arranged for and to deal with more than one themes. The time span of the events was set at least one hour. However, it also varied from

- Standard reusable contents
  - Metadata templates
  - Simplified metadata templates
  - Folder/ file structures/ packaging schemes
  - Publication templates
  - Labels to improve naming convention
  - Data management plan (DMP)
- Practical relevant research data examples
  - Lab experiment
  - Material mixing
  - Survey
  - Simulation
  - Code
  - Journal/ conference publication
  - Data management plan
- Collections of standard external contents and resources
  - List of licences
  - Tools
  - Metadata standards
- Comprehensive survey
- Recommendation and hints about tools, best practices and workflows
  - Coding: build and packaging, dependency management, compiler, execution environments
  - Data/ metadata creation: collection, selection, retention, transformation, basic units and value ranges
  - Data verification/ quality assurance measures and strategies: integrity, accuracy, completeness, authenticity checks
  - Data analysis, experiments, digital representation
  - Data acquisition, integration, anonymization, pseudonymization, release, archiving
- Use cases of collaboration and cooperation
- Systems and tools for automation
- Guides for additional tools and systems
- Detailed review, evaluation and feedback reports and follow ups
- Considerations/ Checklist to conclude research work

Table 9: Key topics of supplementary guiding materials and support

event to event e.g. presentation of results in quarterly meeting was from 15 to 20 minutes, Q & A session for two hours each, weekly dedicated consultation session for two hours and twice a week. Table [10] presents a list of key topics and themes of interactive sessions and workshops. The occurrences and cycles of these events are explained in the later sections of this paper.

10 Means of communication, distribution and imparting trainings

The provision of guiding materials and imparting training required interactions with users. These requirements were strategically classified in three categories. One was communication which means how to let the users know and respond about the events and materials. Second was distribution of materials itself. The third was interaction and imparting training to the researchers.

For the sake of communication institutional email and official platform based communication services were used. These services were used sometimes also for the exchange of contents and materials itself even during the development phase due to particular reasons e.g. to expedite the responses.

The contents and materials were in both printable document forms as well as watchable multimedia contents. While keeping in view the convenience and to reduce the overhead for the users, the guiding materials were distributed with the function of online play/ stream and read-

• Introduction to the VRE system and concepts • On boarding • User and data migration • VRE for data managers/ RDM • RDM and its basics • RDM tools and systems • How to use (tools specific session) • Software/ tools test session • Workflows e.g. for publication, experiments, collaboration	• Practical research data examples • Do it together session • Weekly consultation session • Issue/ Topic/ Task specific consultation session • Q & A session • Reviews and feedbacks session • Consultations based on the reviews and feedbacks • Quarterly progress meeting • Summer school

Table 10: Key topics and themes of interactive sessions and workshops

ability as well as with download possibilities for offline reading and viewing. The distribution of contents was arranged through the same official platforms which were defined as part of the research information infrastructure by following the defined rules e.g. the naming convention, organisation. However, for multimedia contents third party free online streaming service was also used.

For interactive sessions and imparting training online meeting platform and in person arrangements were made. When I-Project started its functions of RDM guidance, the lockdown measures were already started to emerge in Germany and all around Europe. Therefore, during the lockdown most of the in person events happened in online mode. However, later events took place in person at the location of I-Project, at researchers' place and even at the arranged locations e.g. auditorium of the university for summer school or quarterly meeting etc. and in hybrid mode. These practices remained in place till the end of the first phase of CRC in December 2023.

In summary, institutional email, official platform based communication, online streaming, file hosting, online meeting services and applications were employed together with in person arrangements.

11 Processes and phases to create contents

Due to the facts relating to the complexities of technical resources, research work and participants of CRC, for the creation of guiding material and contents an inclusive strategy was adopted. The process was complemented and improved by the inputs of domain experts, test users and actual recipients from CRC. Domain experts and test users were from the libraries, participants of CRC, fellow researchers and colleagues. The process and phases to create contents were as follows.

The initial draft was created based on the knowledge of the state of the art by research associate or content creator. Then the material was shared via official infrastructure and in some cases using official email services either to the test users, domain experts or both and allocated sufficient time for review and evaluation process. On the receipt of feedbacks, feedback session was

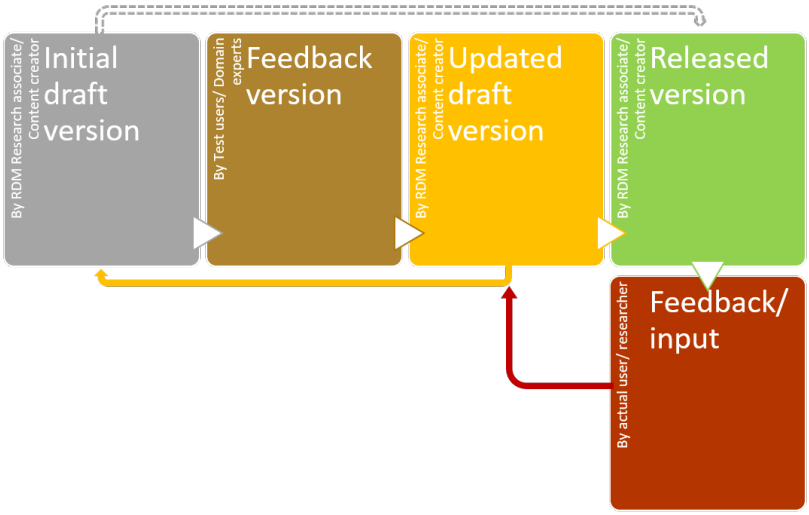

Figure 1: Processes and phases to create contents

organised for one to one discussion and further clarification. Based on the feedback and the results of discussion the draft was updated and if ready to be released then released to the target users. Otherwise again provided for review process. After release, if there was feedback or input by the actual users. A new updated draft was created and the corresponding release process was repeated. The aims to get feedback have been to determine understand ability and clarity, domain relevance, applicability of the described contents and material, the depth of details and examples and conformance to the language of research of the participants. Figure [1] presents processes and phases of content creation.

12 Events to improve guidance and understanding

The provision of resources, guiding material and contents were followed by series of guidance improvement activities which were both regular and randomly occurring events. These series of events used to be initiated after the delivery of materials. After the delivery of materials, interactive sessions, followed by the regular support and reminder events were arranged. The arrangement of these events were sometimes also on need basis either by the users or by the I-Project due to the realisation of necessity or the complexity level. Quarterly meetings or the summer school events after long intervals were also occurred. During this process events to revise already provided material or if there was a need then the creation of new supplementary material were also occurred. The provision of the updated material or supplementary material again followed the same events cycle. Figure [2] presents cycle of events for improving guidance and understanding.

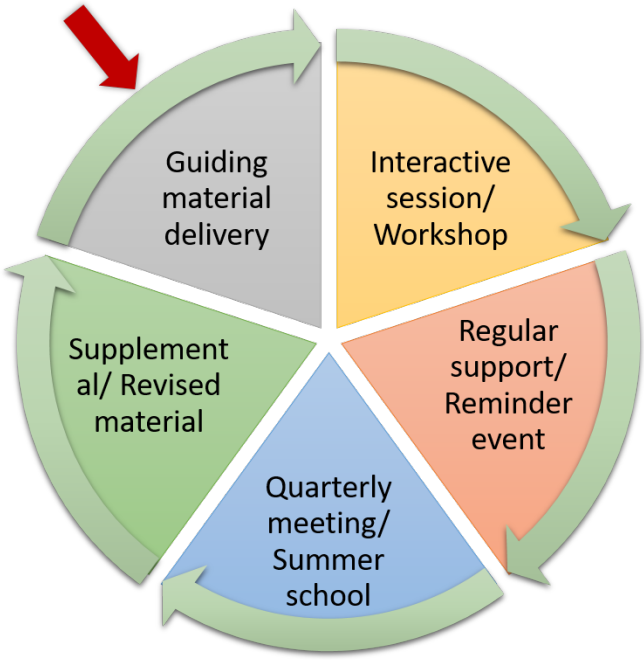

Figure 2: Events to improve guidance and understanding

13 Processes and actions to improve RDM practices

Based on the nature of work and involved data, processes and systems a series of processes and actions were defined and carried out to make the participants further comprehend and improve their RDM practices. This section explains those processes and actions. As a first step data was created or updated by the researchers on their preferred platform and system. Then the data was prepared for the submission to the RDM platform. In these cases, it was assumed that system and platform could be nonofficial but compliance may apply. The data preparation process was followed by the delivery of data to the official RDM platform. Then the evaluation of submitted data was carried out either at the specified time or after a particular interval by the responsible person from I-Project. It is worth mentioning that the purpose of the evaluation was solely about compliance with good RDM practices not the data and data process which are the responsibilities of researcher and research supervisor etc. Based on the evaluation feedback was created. If the data or the required information was missing or for other reason only the reminder should occur, then the reminder was sent. In case there was a need of feedback session then session to discuss the feedback was arranged. And the process went on. Figure [3] presents processes and actions to improve RDM practices.

14 Key compliance strategies

The provision of guiding materials and events were guided by a strategy to make the agenda of RDM easy for adoption, clear to follow and to the success of offerings of I-Project. In this section key strategical aspects to improve the compliance are being listed.

First it was the provision and preparation of the target platform e.g. by already created organ-

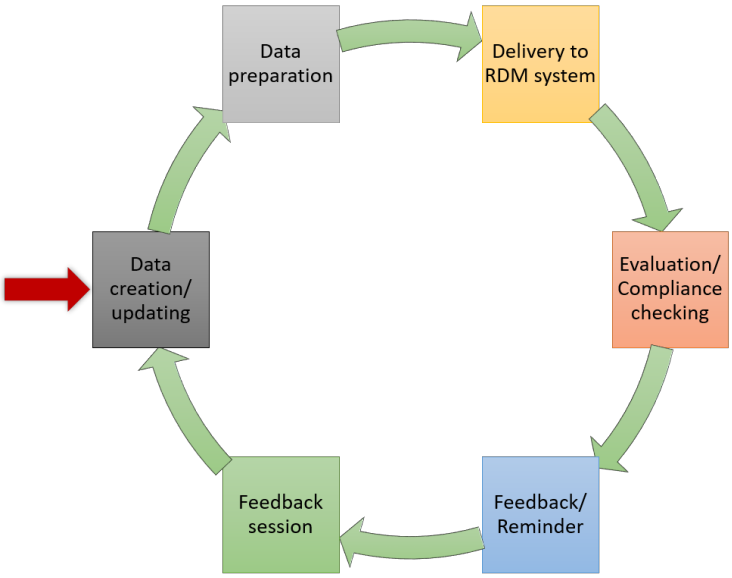

Figure 3: Processes and actions to improve RDM practices

isational structures, access privileges and metadata templates for a smooth and easy kick start. Then the provision of exemplary material followed by interactive sessions, simple and statistical reviews and feedback e.g. based on data audits. On top of that supplementary material to address day to day occurring issues. A comprehensive survey, collaborative work of team of I-Project with the participants of other projects of CRC and continuously improving the guiding material and range of software applications and IT solutions. In case of shortcomings of existing solutions, new solutions were worked out and offered together with the similar guiding materials. Table [11] presents a list of key compliance strategies.

- Provision of initial structures and templates for a smooth and easy kick start
- Examples of best practices
- Detailed interactive sessions
- Interactive do it together sessions
- Q & A rounds
- Detailed reviews and feedbacks
- Interactive review and feedback sessions
- Supplementary materials
- Comprehensive survey about RDM practices
- Collaborative work/ assistance in case of difficulties
- Coordinated support for cooperative research work
- Identification of shortcomings and improvement
- Continuously improving guiding materials and IT solutions
- Mitigation of administrative challenges

Table 11: Key compliance strategies

15 Tactics to enforce compliance

In the cases of the projects of CRC researchers were not only required to have understanding of theoretical and applied knowledge but also demonstrate that knowledge during the actual research activity. In the beginning and even during the course of work it was realised that all

the supports and services would not be sufficient to keep participants engaged and compliant. Tactical approaches to enforce compliance and to remind obligations for RDM were thus deemed necessary. The first formal approach was to define research data policy for CRC in which the obligations were clearly defined and the policy was set to be a binding document for all the participants. The other measures taken during the course of research work were included influence through the board, research supervisor, sending reminders, recognition and appraisal in feedback, presenting evaluation together with statistical data and summaries during collective sessions and meetings. Describing the advantages and the confidence in official platform as well as implications of noncompliance e.g. data loss, consequences of further approval in case of noncompliance etc. Table [12] presents a list of key measures taken to enforce compliance for RDM.

- Obligation through Research data policy
- Advantages and confidence in official platforms
- Influence of research organising board/ committee
- Influence through research supervisor/ supervisory board
- Reminders during collective meeting events, through communication channels before and after the schedule is missed
- Recognition and appraisal in feedbacks
- Presentation of results, summaries and updates during meetings/ collective events
- Presentation of feedback and evaluation reports
- Narrating adverse impacts and implications of non-compliance e.g. data loss, consequences of further approvals
- Follow ups

Table 12: Measures to enforce compliance for RDM

16 Conclusion

The functions of RDM teaching are subject to effectively and efficiently meet the call to support and train the researchers while continuously improving the offerings with retrospectives. In this paper the conceived definition of good RDM and approaches to deliver related practices were presented. The presented approaches and experiences, which may be attributed to didactic and pedagogy methods, were based on the work for the Information infrastructure project which is part of DFG funded interdisciplinary collaborative research centre. The common skills and understanding that every researcher should know before and during the course of research work were listed and presented on one hand. On the other hand, skills and knowledge relating to the common aspects of RDM systems and activities were addressed. Additionally, contents and methodologies for both in person and online as well as for hybrid environments to increase acceptance and understanding were presented.

The skills which relates to the pre-research learning and the practices relating to good RDM were emphasised. The applications and contexts for teaching and guiding were explained. Therefore, such contents could be suggested for teaching to the future researchers and data stewards/ professionals during the regular academic discourse too. The presented contents and strategies may also be adopted for small projects to large scale CRC projects. The prepared material and defined strategies were ready for adoption also in digital learning platforms. The

presented approaches could be beneficial and useful for policy makers, infrastructure developer and providers, academicians, curricula designer, researchers, students, user and community support service providers. Almost all the materials and infrastructural components are accessible and available to use under open source and open access compliant licenses, unless otherwise specified.

It may be noted that the research activities in CRC were more diverse and researchers were dealing with larger number of scenarios. Therefore, it was felt that the examples prepared and demonstrated for research activities were not sufficient to cover all the scenarios. Thus, the work in this direction could be done further and profiles may be constituted for the use of practitioners and trainers etc. It was realised that paying more attention to technical and administrative challenges could improve the offerings and reduce the burden of guiding material and activities. In future, survey and analysis will be carried out to marginalise and evaluate the impact of the presented approaches.

17 Acknowledgements

The work was conducted as part of the Collaborative Research Centre (CRC) ‘Additive Manufacturing in Construction - The Challenge of Large Scale. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project Number 414265976 - TRR 277”.

18 Roles and contributions

Syed Ashfaq Hussain Shah: primary author of the manuscript and responsible to conceive, design and implement the work.

Frank Petzold: principal investigator and the work was done under his supervision.

References

[1] M. Frické, “The knowledge pyramid: The dikw hierarchy,” *Ko knowledge organization*, vol. 46, no. 1, pp. 33–46, 2019, ISSN: 0943-7444. DOI: https://doi.org/10.5771/0943-7444-2019-1-33. [Online]. Available: https://www.nomos-elibrary.de/10.5771/0943-7444-2019-1-33.pdf.

[2] B. Hjørland, “Knowledge organization (ko),” *KO Knowledge Organization*, vol. 43, no. 6, pp. 475–484, 2016, ISSN: 0943-7444. DOI: https://doi.org/10.5771/0943-7444-2016-6-475. [Online]. Available: https://www.nomos-elibrary.de/10.5771/0943-7444-2016-6-475.pdf.

[3] J. Girard and J. Girard, “Defining knowledge management: Toward an applied compendium,” *Online Journal of Applied Knowledge Management*, vol. 3, no. 1, 2015, ISSN: 2325-4688.

[4] J. A. Borghi, S. Abrams, D. Lowenberg, S. Simms, and J. Chodacki, “Support your data: A research data management guide for researchers,” *Research Ideas and Outcomes*, vol. 4, 2018. DOI: 10.3897/rio.4.e26439. [Online]. Available: https://doi.org/10.3897/rio.4.e26439.

[5] M. Baker, "1,500 scientists lift the lid on reproducibility," *Nature*, vol. 533, no. 7604, pp. 452–454, 2016, ISSN: 1476-4687. DOI: 10.1038/533452a. [Online]. Available: https://doi.org/10.1038/533452a.

[6] L. D. Booker, N. Bontis, and A. Serenko, "The relevance of knowledge management and intellectual capital research," *Knowledge and process management*, vol. 15, no. 4, pp. 235–246, 2008, ISSN: 1092-4604.

[7] L. Markauskaite, "Digital media, technologies and scholarship: Some shapes of eresearch in educational inquiry," *The Australian Educational Researcher*, vol. 37, no. 4, pp. 79–101, 2010, ISSN: 2210-5328. DOI: 10.1007/BF03216938. [Online]. Available: https://doi.org/10.1007/BF03216938.

[8] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, *et al.*, "The fair guiding principles for scientific data management and stewardship," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016, ISSN: 2052-4463.

[9] L. Candela, D. Castelli, and P. Pagano, "Virtual research environments: An overview and a research agenda," *Data science journal*, vol. 12, GRDI75–GRDI81, 2013, ISSN: 1683-1470.

[10] M. Assante, L. Candela, D. Castelli, *et al.*, "The gcube system: Delivering virtual research environments as-a-service," *Future Generation Computer Systems*, vol. 95, pp. 445–453, 2019, ISSN: 0167-739X.

[11] D. Bailo, R. Paciello, J. Michalek, *et al.*, "Data management in distributed, federated research infrastructures: The case of epos," *Data Science Journal*, 2024.

[12] G. Sivertsen, "Developing current research information systems (cris) as data sources for studies of research," *Springer handbook of science and technology indicators*, pp. 667–683, 2019, ISSN: 3030025101.

[13] K. B. Read, C. Larson, C. Gillespie, S. Y. Oh, and A. Surkis, "A two-tiered curriculum to improve data management practices for researchers," *PLOS ONE*, vol. 14, no. 5, e0215509, 2019. DOI: 10.1371/journal.pone.0215509. [Online]. Available: https://doi.org/10.1371/journal.pone.0215509.

[14] G. Pryor and M. Donnelly, "Skilling up to do data: Whose role, whose responsibility, whose career?," 2009, ISSN: 1746-8256.

[15] L. Corti and V. Van den Eynden, "Learning to manage and share data: Jump-starting the research methods curriculum," in *The Teaching and Learning of Social Research Methods*. Routledge, 2018, pp. 101–116.

[16] A. Cox, E. Verbaan, and B. Sen, "A spider, an octopus, or an animal just coming into existence? designing a curriculum for librarians to support research data management," *Journal of eScience librarianship*, vol. 3, no. 1, 2014, ISSN: 2161-3974.

[17] JISC, *Research data management toolkit*, Web Page, 2018. [Online]. Available: https://www.jisc.ac.uk/guides/research-data-management-toolkit.

[18] K. A. Briney, H. L. Coates, and A. Goben, "Foundational practices of research data management," 2020.

[19] J. Medina, A. W. Ziaullah, H. Park, *et al.*, "Accelerating the adoption of research data management strategies," *Matter*, vol. 5, no. 11, pp. 3614–3642, 2022, ISSN: 2590-2393.

[20] A. Whyte, D. Green, K. Avanço, *et al.*, "D21 catalogue of open science career profiles, minimum viable skillsets," Report v1.2, 2023. DOI: 10.5281/zenodo.8101903. [Online]. Available: https://doi.org/10.5281/zenodo.8101903.

[21] M. Grootveld and E. Verbakel, "Essentials for data support: Training the front office," *International Journal of Digital Curation*, vol. 10, no. 1, p. 240, 2015, ISSN: 1746-8256.

[22] Generic, 2012. [Online]. Available: http://www.rluk.ac.uk/files/RLUK%20Re-skilling.pdf.

[23] K. Biernacka, K. Helbig, and P. Buchholz, "Adaptable methods for training in research data management," 2021.

[24] N. J. Casson, C. J. Whitfield, H. M. Baulch, S. Mills, R. L. North, and J. J. Venkiteswaran, "A model for training undergraduate students in collaborative science," *Facets*, vol. 3, no. 1, pp. 818–829, 2018, ISSN: 2371-1671.

[25] C. M. Goedhart and J. S. McLaughlin, "Student scientists: Transforming the undergraduate biology lab into a research experience," *The American Biology Teacher*, vol. 78, no. 6, pp. 502–508, 2016, ISSN: 0002-7685.

[26] C. Wiljes and P. Cimiano, "Teaching research data management for students," *Data Science Journal*, vol. 18, pp. 38–38, 2019, ISSN: 1683-1470.

[27] J. S. McLaughlin and M. S. Coyle, "Increasing authenticity inquiry in the cell molecular biology laboratory," *The American Biology Teacher*, vol. 78, no. 6, pp. 492–500, 2016, ISSN: 0002-7685.

[28] AMC-TRR277, *About amc*, Web Page, 2024. [Online]. Available: https://amc-trr277.de/trr-277-mission/about-amc/.

[29] DFG, *Funding information infrastructures for research - a dfg strategy paper*, Government Document, 2018. [Online]. Available: https://www.dfg.de/resource/blob/173276/96ebe993fd3c5d2959d46bef63dc2004/dfg-strategy-paper-funding-info-infrastructures-research-data.pdf.

[30] S. A. Hussain Shah and F. Petzold, *Research data management of large scale projects and a reference model of data life cycle for dynamic dmps*, Conference Paper, 2022. DOI: https://doi.org/10.14459/2022md1686600. [Online]. Available: https://mediatum.ub.tum.de/1686600.

[31] S. A. Hussain Shah and F. Petzold, "Research data management and a system design to semi-automatically complete integrated data management plans [position paper]," *Architecture and Planning Journal (APJ)*, vol. 28, no. 3, p. 40, 2023, ISSN: 2789-8547. DOI: https://doi.org/10.54729/2789-8547.1235.

[32] DFG, *Handling of research data*, Government Document, 2021. [Online]. Available: https://www.dfg.de/resource/blob/174736/92691e48e89bf4ac88c8eb91b8f783b0/forschungsdaten-checkliste-en-data.pdf.

[33] H. Coates, A. Goben, and K. Briney, “What if it didn’t happen: Data management and avoiding research misconduct,” *Journal of eScience Librarianship*, vol. 12, no. 3, 2023, ISSN: 2161-3974.

[34] S. A. Hussain Shah and F. Petzold, *Research data management policy for large crc projects*, Dataset, 2024. DOI: https://doi.org/10.14459/2024mp1734393.

[35] S. A. Hussain Shah and F. Petzold, *Data management plan (dmp) for large crc projects*. Dataset, 2023. DOI: https://doi.org/10.14459/2023mp1709465.