RESEARCH ARTICLE

Matching Data Life Cycle and Research Processes in Engineering Sciences

Tobias Hamann [1]
Michèle Robrecht [2]
Max Leo Wawer [3]
Jonas Maximilian Werheid [1]
Mario Moser [1]
Marcos Alexandre Galdino [1]
Anas Abdelrazeq [1]
Roland Lachmayer [3]
Robert H. Schmitt [1, 2]

1. Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, Aachen.
2. Fraunhofer Institute for Production Technology IPT, Aachen.
3. Institute of Product Development, Leibniz University Hannover, Hannover.

Date Published:
2024-06-15

Reviewer:
Petra Personenname,
Fachhochschule Potsdam

Keywords:
Engineering Research, Research Data Management, Data Lifecycle, FAIR Data, research data management process

Data availability:
Data can be found here: https://doi.org/10.5281/zenodo.11198842

Software availability:

Abstract. Research data management (RDM) has become increasingly significant, focusing on ensuring the most benefit from data creation. This especially applies within the engineering sciences, where many different sources generate big amounts of heterogeneous data. However, integrating RDM into day-to-day work is difficult. Thus, methods need to be defined to effectively and conveniently manage research data with increased complexity and diversity of subdisciplines. This article aims to provide clarity and structure in research processes integrating RDM, considering various models and viewpoints to present a research process for engineering sciences that is inherited from RDM processes. Therefore, interviews and workshops in different formats were used to gather insights about requirements in day-to-day work and research processes. As for further steps, the process will be evaluated on a later stage through a validation survey that will be taken into implementation for the Joint Assistant for Research in Versatile Engineering Sciences (Jarves).

1 Introduction

Since research data management (RDM) can ensure the most benefit from data through quality assurance by curating data at every phase of their life [1], it has been growing in importance for years. It addresses the challenges arising in research: for example globalisation, digitalisation and the need for increased efficiency [2]. Research projects generate an ever-increasing heterogeneous amount of data, often in digital form, demanding a more efficient approach to be handled within the same time. This applies particularly for engineering disciplines, where data management is crucial to enable emerging methods such as machine learning. [3], [4] RDM is one part of research activities and includes the generation, organisation, analysis, selection, storage and sharing of research data to make research and collaboration easier and more efficient. This series of practices encompass the complete research cycle while focusing on the efficient use of resources [5]. The underlying processes have the common goal of keeping data findable, accessible, interoperable, reusable (FAIR) in the long term and independent of individuals [6], [7]. RDM therefore needs to be implementable in a unified yet adaptable way in order to be applicable to engineering researchers. Still, many engineering researchers are unsure about how to apply RDM, which processes to follow, and how to manage their data in day-to-day work [8], [9]. In other words, a process for RDM activities in engineering research-projects needs to be established. To address this challenge, RDM processes have already been proposed, such as the data life cycle (DLC) in several variations [1], [10], [11]. However, these processes cannot be implemented in practice without adjustments [12], do not fit project-oriented research and lack specificity in engineering focus, as presented in section 2.3.

Engineering research activities are often conducted in project structures, as can be proven by literature over the least 15 years [13], [14]. This project-orientated setup of engineering research is reinforced by a rapid increase in project-based allocation of funds by funding organisations [2], [15]. Moreover, project-oriented approach can be found in all types of research across the spectrum from basic research to applied research, as Behlau [16] points out. To address this focus on engineering sciences, it is recommendable to narrow down the scope of such process to research projects, thus further improving its applicability. Hence, the following research question arises:

Can RDM be embedded as a process into engineering research projects?

Therefore, this paper aims to introduce a process for engineering research, that embeds RDM into day-to-day work by taking DLC models into account. Resulting from the listed challenges, the need for a multi-level process that includes both project-oriented and data-oriented perspectives was identified. In addition, the perspective of DLCs' insufficiency for application as well as the other branching activities within the research process need to be considered.

The goal of the proposed process is introducing guidance into the RDM process along the engineering researchers' day-to-day activities through a better understanding of research processes in engineering disciplines [17]. Its structured approach to manage research data along the DLC, from data creation to their handling, aims to facilitate the management of data from different sources in contrast or rather addition to other techniques like, for instance, the usage of metadata schemes. The process will also provide a framework for the development of individual maturity models for defined process areas for quality assurance of RDM in research projects.

To create the proposed process, this paper's structure follows our research activities, starting with a "Related work" analysis. Afterwards, the "Research methodology" and the consequential "Aggregation of the newly proposed research data management process" is described. This process is then examined in the "Discussion" section. The paper ends with a "Conclusion and Outlook". Eventually, a "Glossary" at the end of the paper defines terms within the proposed process.

2 Related work

To get an overview of current processes in engineering in general as well as state of the art RDM processes, a literature review has been conducted. Section 2.1 first presents definitions and the state of the art on the term ”research processes” in the context of engineering sciences. Afterwards, section 2.2 then presents different RDM processes.

2.1 Engineering research processes

Overarching, research is a process that seeks ”knowledge, or formulation of a theory that is driven by inquisitiveness for that which is unknown and useful on a particular aspect so as to make an original contribution to expand the existing knowledge base” [18]. Griem et al. structure this process in a generic way according to an input-processing-output-model (IPO model). Each work package (level below the research process) has an input phase, a processing phase and an output phase. The same applies to each task (level below the work packages) and each tool used to carry out these tasks. The processing step (i.e. the tasks) on the work package level is executed several times and concludes with a result file. The breakdown into smaller levels reduces the complexity of the activities (see figure 1) [19].

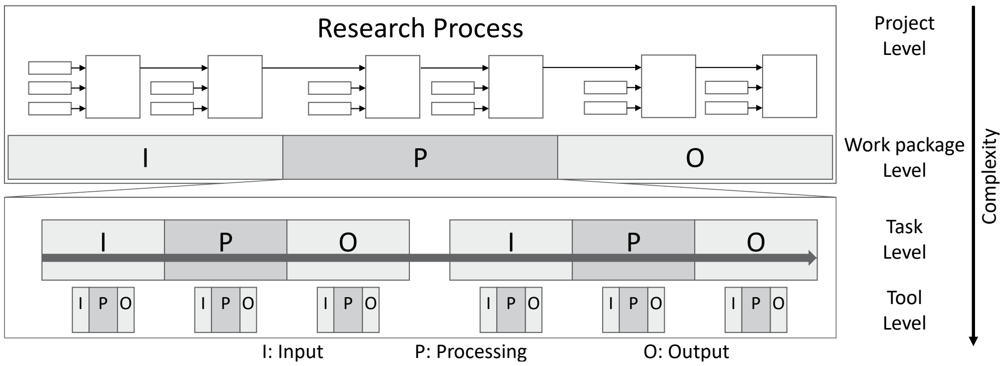

Figure 1: Research process as proposed by Griem et al. [19]

The actual tasks include the generation of hypotheses and their testing, confirmation or refutation through data analysis [18]. In hypothesis testing, there is not one general but a variety of models of research processes, each comprising different perspectives and details. Four exemplary models are first summarized in table 1 and then described.

The focus on research projects, as mentioned in the ”Introduction”, is still the premise for this paper. In this article, the term ”project” refers to the DIN-Definition of projects as “an endeavour that is essentially characterised by the uniqueness of the conditions as a whole” [24]. In addition, the aspects of “relative complexity and relative novelty” of Möller and Dörrenberg's[25] definition are also considered.

Stratmann [20] breaks projects down into four sub-steps: They start with the writing of a research proposal. The project is then executed, on which the funding organization receives a project report. This process clearly ends with project closure.

Tenopir et al. [21] provide a more detailed life cycle than Stratmann [20]: They map a process,

Stratmann [20]	**Tenopir et al.** [21]	**FIZ Karlsruhe** [22]	**Patel** [23]
Research proposal	Ideas	Idea	Research concept
Project execution	Partners	Data	Proposal & data management plan
Research report	Proposal writing	Analysis	Experiment
Project closing	Research process: - Simulate, experiment, observe - Manage the data - Analyse data - Share data	Documentation Search & finding Education & Teaching	Data processing Data analyzing Data interpreting Quality control
	Publication		Data validation

Table 1: Overview of different research process models

starting from the idea to the search for partners, writing the proposal, the research process and publication. The research process itself is an independent cycle. In view of the increasingly data-intensive nature of research, the research process is embedded in a virtual research environment. The focus is on the data generated and used by engineering researchers.

FIZ Karlsruhe [22] does not name the process under consideration ”research” but ”value creation” and expand it to further areas of science and innovation, thus taking a more holistic approach than Tenopir et al. [21]. Its cycle consists of six steps, each of which contains sub-steps. Steps not mentioned in the previous two models are ”Searching & Finding,”, i.e. editing and indexing activities by information facilities, and ”Education & Teaching”. The cycle is split into two parts: one dealing with the implementation of research projects and the other focusing on the subsequent use of the results in different contexts. There is no clear starting point.

Patel’s [23] ”Idealised Scientific Research Activity Model (I2S2)” depicts research activities from a typical experimental project in the physical sciences. The activities in the research process are similar to previous models. Patel adds three further dimensions, which influence each other: administrative, archive and publication activity. For example, a data analysis (research activity) leads to documentation of the analysis (archive activity) and the preparation of supplementary data (publication activity).

The examples of research processes show that there is no general model. The detail and complexity of research processes are be influenced by the methods used in the particular research context as described below. The scientific methodology, as a sum of the methods used, can be divided into four phases in engineering research processes. In the *analysis phase*, researchers build up a deep understanding of a problem that they want to solve with their research, e.g. through observation. They then form a *hypothesis or plan* – an assumption about how the problem might be solved and how they might test that assumption, supported by literature reviews. In *synthesis*, the researchers check their hypothesis or plan and collect information. In *validation*, they formally evaluate with these information whether they have achieved the goal of their research. These steps are repeated until the goal of the research is achieved. [26], [27]

Depending on the type of engineering research activity, researchers use different methods:

- Descriptive or analytical: Descriptive research only describes the state of the art based on given variables. Analytical, on the other hand, can control the variables and identify the causes of phenomena.
- Applied or basic: Applied research deals with a concrete problem in practice in order to solve it. Basic research, on the other hand, deals with generalizations and formulates theories that could be applied in the longer term.
- Quantitative or qualitative: Quantitative research refers to many cases in order to draw representative conclusions. Qualitative focuses on a few or individual cases. [18]

One element in particular underlies the engineering research processes presented: data management. It is important in engineering research due to the extensive and complex nature of the data generated, requiring effective strategies for data collection, organization, storage, and analysis to ensure the application of FAIR principles [6]. Research process models are the subject of the next section.

2.2 Research data management processes

RDM deals with the management of research data along the entire research project. It aims from making the research process as efficient as possible to facilitate cooperation [5]. RDM can also be defined by focusing on the process as a series of practices for dealing with resources [28]. In addition, a definition of research data shall be given:

> “Research data includes measurement data, laboratory values, audiovisual information, texts, survey data, objects from collections or samples that are created, developed or analysed in the course of scientific work. Methodological test procedures such as questionnaires, software and simulations can also represent key results of scientific research and should therefore also be categorised as research data.” [29]

In the follwing, two perspectives on RDM processes are given. Firstly, the commonly used DLCs are presented. Afterwards, alternative concepts are introduced.

2.2.1 Data life cycles

Data management plans (DMPs) and guidelines often refer to the concept of a DLC. DLCs describe different stages of data processing, from is creation or acquisition to its eventual disposal or preservation. It encompasses a closed lifespan of data and involves various processes and activities to manage and extract value from data effectively [1]. This ”approach ensures that all the required stages are identified and planned, and necessary actions [are] implemented, in the correct sequence. This can ensure the maintenance of authenticity, reliability, integrity and usability of digital material” [1]. To give an impression on the different approaches to this topic, two exemplary DLCs are presented below.

The first example is a rather simple model by forschungsdaten.info and is shown in figure 2. It starts with the *planning* activity, followed by *data collection*. Then the data is *processed and analysed*, *shared and published* before the steps *data archiving* and lastly *re-using* occur. [30]

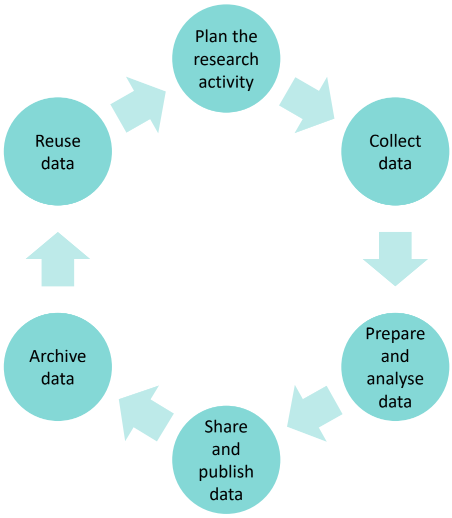

Figure 2: Data life cycle as proposed by forschungsdaten.info [30]

Contrasting the rather simple example presented, there are other more complex life cycles, such as the one proposed by the Digital Curation Centre (DCC), shown in figure 3. The life cycle is structured on several levels: Central to it is the research data. The next level includes activities that are relevant throughout the life cycle. This includes, for example, the management of metadata. [1] The sequential activities are similar to the DLC of forschungsdaten.info [30]. They should be carried out in any case to ensure the curation of the research data, i.e. performing management activities to enable reuse. The last level represents optional activities that may become necessary due to external requirements - for example, migrating data to another format to ensure long-term availability. [1] The DLC focuses on the curation and preservation of research data and was designed to be used by libraries and related institutions rather than in projects.

The two DLCs presented are merely examples. There are many different ones, varying in scope, complexity and visualization like the ones from RDMKit [10] or RWTH Aachen University [11] (for an overview see, for example, [12] or [31]). The forschungsdaten.info's [32] version is limited to the DLC, whereas the DCC [1] version also maps aspects around the DLC itself. In contrast, forschungsdaten.info provides a lot of information that flank the DLC. For example, they compiled domain-specific information [32]. The DCC only gives a concise description of the individual steps. This provides users more room for interpretation in their actual implementation, even though the life cycle looks very specified at first glance. Subsequently, the DCC as an infrastructure institution, published a guide on how to develop RDM services. [33] The representation of the DLC, referred to by RWTH Aachen University, follows a similar approach. As central infrastructure facilities, the library and the IT center of the university offer various services as modular, integrated building blocks to the institutes and chairs, which they can use for individual RDM processes. [34]

The two examples show that a DLC cannot develop its full potential on its own, in the sense of providing a clear roadmap. Only by providing further definitions and guidance, they offer

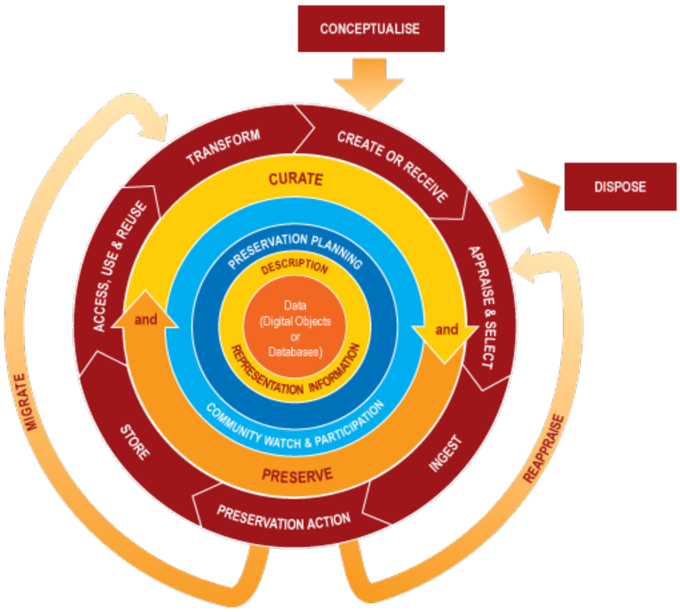

Figure 3: Data life cycle as proposed by Digital Curation Centre [1]

researchers the opportunity to successfully manage their research data.

2.2.2 Other concepts for RDM processes

In contrast to the mentioned DLCs, other models have been developed to address the shortcomings and deficits of the DLC in the context of research projects. For instance, Putnings et al. [35] have developed a different approach for representing RDM as a model. They depart from life cycles as sequential loops and form only loose connections between the building blocks. Although the building blocks are also found in life cycles, they are supplemented by external influences such as legal, ethical or funding policy requirements. Sharing, publishing, modifying, and deleting data can partially be performed. The model divides RDM into two domains - planning and execution - completely breaking down the traditional thinking in cycle paths.

Resorting to a life cycle, yet on another level, Jagerhorn and Chen [36] present a “FAIR Workflow Project” cycle. Staring with grant application and application approval, this cycle then continues with the actual research and finishes with published articles and a “FAIR Workflow Interface”. Yet, it offers a linear approach to research projects, not taking into consideration the often iterative nature of research projects. Similarly, Tripathi and Pandy [37] include “proposal creation”, “project startup” and other steps into a life cycle model, resulting in the same deficits.

The presented examples illustrate the two main issues this article strives to address. Firstly, the definitions of phases and activities vary widely and have very different scopes. Secondly, and arguably more importantly, there is a gap between DLCs and other, more holistic models. DLCs are sufficient, as their name suggests, for supporting the management of data along its life cycle. However, they are not intended to be project management cycles for research data, although they are often and mistakenly used as such. They are a “useful metaphor, but tend to encourage thinking that research processes are highly purposive, uni-directional, serial and occurring in a closed system. Research is often not like this" [12]. Furthermore, other models of research projects often do not include RDM as shown in section 2.1. Putnings et al. [35] address this by distinguishing two meta-phases of planning and execution, and by pointing out the possibility of branching of a research process. The RDM process as proposed by the Alderman Library of the University of Virginia [38] break up the DLC while including backwards loops for reuse, re-purpose and depositing of data. The process described again contains the pre-project activities.

2.3 The gap between data life cycles and engineering research processes

Although the research processes are aligned to the increasingly data-intensive engineering research and take this transformation into account, a research process-oriented consideration of RDM-related requirements is missing for project-oriented research activities in engineering. For example, existing approaches to the presentation of RDM provide a cycle-orientated reuse of data rather than a research process-orientated implementation of RDM requirements. As pointed out in section 2.2.2, data life cycles do not depict the reality of research processes, especially not for engineering sciences. This makes them unsuitable as a basis for project management, although they are often used for this purpose [12]. At the same time, research processes do not sufficiently take into account the dynamic and recurring management of research data. Due to this insufficient integration, RDM is often perceived as a separate and afterthought task to be done after the actual research work.

This illustrates the identified research gap: lacking integration and applicability of RDM in the research process. The challenges mentioned above highlight the need to integrate RDM more effectively into the engineering research process, which are often project-oriented, to reduce workload and improve the applicability of RDM and hence the quality of research data. This can add value to the implementation of RDM specifically for engineering sciences to promote scientific integrity and efficiency in engineering research.

3 Research methodology

To design an RDM process fitting for research in engineering sciences, in terms of research project structure, we pursued a multi-leveled approach. The methodology is based on the design science research approach as proposed by Alturki et al. [39].[1] This approach is sufficient for this project, as its focus is set on the development and evaluation of a new solution to "enable organizations to address important information-related tasks" [40]. Within this approach, engineering researchers' requirements and research processes were collected by focus group interviews and workshops. While in development, the process was presented and discussed on multiple events (see [17], [43]–[46]). This continuous scrutiny allows for fast and broad feedback. This however does not replace an evaluation. The next step is to carry out a validation survey. All results lead to the composition to the result presented in section 5.

1. This approach is feasible and reasonable, as "it focuses on creating and evaluating innovative IT artefacts that enable organizations to address important information-related tasks" [40]. These IT artefacts can contain executable code but may also be just a concept that solves a problem [41]. "Engineering disciplines accept design as a valid and valuable research methodology because the engineering research culture places explicit value on incrementally effective applicable problem solutions" [42].

4 Collection of research processes and requirements

Following, the methodology for the collection of determining factors and requirements are presented:

- Problem-centered interviews
- Workflow workshops: NFDI4Ing Task Area Frank [47]
- Workflow workshops: NFDI4Ing Task Area Base Service S-1 [47]

Firstly, problem-centered interviews were conducted to collect capabilities and hurdles of the implementation of RDM along with corresponding requirements, which formed the basis for a process proposal. In two workshops, this process was tailored towards exemplary research workflows. The two workshop trails have been conducted independent to ensure a broad, heterogeneous and unbiased collection of research workflows. This includes preferably the consultation of as many different projects as possible. This was seen necessary due to the foundational nature of these workflows as they lay out the very groundwork of the process proposed in this paper. An additional set of requirements would be possible to incorporate at a later stage. Meanwhile the foundation of the RDM process is set to be the researchers day-to-day work.

The following sections are each divided into a general introduction to the method, the clarification of the subject of investigation and a description on how participants were selected. Then, the data collection procedure is described before the evaluation and the results are presented.

4.1 Problem-centered interviews

Eight problem-centered interviews were conducted to address the research gap precisely. The aim of these qualitative interviews is to analyze in detail the heterogeneous research processes in engineering applied by different researches. The "first-hand insights [...] on what, how and why" [48] researchers include in their RDM was in the focus of the investigation. Beforehand, a compendium of questions was developed as a guideline for the semi-structured interviews. This was then tested in a pretest interview with the project team's colleagues at the Laboratory for Machine Tools and Production Engineering (WZL) on the RWTH Aachen University. Afterwards, seven interviews were conducted in between February and July 2021. [2]

Subjects of investigation were, on the one hand, a general overview over the status quo of applied RDM in engineering sciences – how it is embedded into the engineering researchers' day-to-day work and which problems arise. This included a inquiry of the acceptance of RDM, used tools and procedures, applied guidelines and standards as well as utilized support and training offers. On the other hand demands and requirements in the context of RDM in engineering were collected. This leads to an overview over the opportunities and challenges of the implementation of RDM in the engineering researchers' work routine and important requirements for this implementation.

The 19 participating engineers originated from Germany, specifically the Karlsruhe Institute of Technology, University Bremen, RWTH Aachen University, TU Berlin, Fraunhofer Institute for Production Technology IPT, Cluster of excellence Internet of Production and a Collaborative

2. More information on the method can be found in Additional information on the problem-centered interviews.

research centre (ger: Sonderforschungsbereich SFB/Transregio) TRR. The interviews followed a defined structure spanning approximately 120 minutes. 48 questions were posed, categorized according to subtopics within RDM. These were designed as a status quo recording structured according to DLC as proposed by forschungsdaten.info [30] (see figure 2). A SWOT analysis was conducted, delving into *strengths*, *weaknesses*, *opportunities* and *threats* related to RDM in the participants‘ work environment.

A total of 292 statements were condensed from all parts of the interviews. Participants could give none, one or multiple statements on each question item. Results are presented in four subtopics:

- **Acceptance of RDM** (63 statements)
- **Tools and approaches** (112 statements)
- **Guidelines and standards** (65 statements)
- **Existing** (17 statements) and **desired offers** (35 statements) for training and support

While six positive statements about **acceptance of RDM** were made, pointing out its relevance and how the exchange with colleagues raises awareness and acceptance, two statements showed a general openness even though the application was denied. 14 statements indicated an immediate negative attitude towards RDM with the most common reason represented in a total of 22 statements: missing awareness for RDM in general, its benefits and best practices. Its expressions range from a general and in parts unconscious lack of awareness over a disunity between benefits and disadvantages and insensitivity to a direct rejection of RDM. Some stated that the topic of RDM just had not reached them yet (four statements), others asked what benefit they had from it (five statements). Two interviewees saw it as a boring duty or as a ’necessary evil’ while one even pointed out that ”the motto of RDM is: Better to do nothing than to do something wrong”.

Additional effort due to RDM was criticised by 15 statements, of which eight pointed out that it hinders or even prevents them from their actual work. A possible solution were brought up by six statements: A light-weight, integrative and (partly) automated RDM process along the engineering researchers’ daily work. The automation aspect was a recurring wish over all subtopics. Furthermore, there are concerns surrounding data sharing and reuse (eleven statements), with publications based on others‘ data being perceived as less prestigious (three statements). Additionally, there is a concern about data sharing could provide a competitive disadvantage, as well as assumptions that own data is too specialized or unusable for external use (three statements). Two statements explicitly demand guidance in the selection and storing of research data. The two last major points raised are the need for guidance (six statements) and the lack of applicability (five statements). Vague concepts, unclear specifications and missing templates lead to a decrease in acceptance.

The most statements regarding **tools and approaches** revolve around the concepts of storing and sharing data (22 statements each), most often in the context of institution internal (19 statements), project-internal (13 statements) or external (seven statements) data sharing. Often, there is ”no uniform structure” but a collection of unconnected solutions, most often based on folder structures on institutional servers or cloud drives. While of significant importance for collaboration, it is an overlying problem that causes uncertainty amongst engineering researchers. Unclear restrictions and guidelines, most of which are not coherent and sometimes even contradict each other, cause confusion and frustration. This applies to projects small, local projects and is amplified if several partners are involved. In total, six statements either demand automation in RDM or claimed to be implementing automation for RDM in their research, namely the automatic creation of folder structures, data recording and partial-automation for DMP creation.

Existing **guidelines and standards**, such as DFG directives and FAIR principles, along with practices like archiving and metadata inclusion, are acknowledged while the interviewees question their compliance. Twelve interviewees gave a statement to existing guidelines in their respective institutions. But the guidelines available ”are very generic” and their compliance is ”not checked”. This is aggravated by the incomplete and unstructured support some of the referred guidelines provide, as they are a loose collection of unconnected and sometimes contradicting specifications and restrictions. However, there is a gap between the theoretical existence of a guideline and its actual utilization, so that these guidelines are treated exactly like that: guidelines. It is no strict set of rules and other than funding organisations requirements there is no basis to enforce them.

Existing offers for training and support in RDM have not been used by many participants due to a lack of availability and time constraints. Most often, the participants only visited basic seminars on RDM (seven statements). The second most important source of RDM knowledge or support is personal research (four statements). Personal exchange, seminars for tools and best practices each only were named once. The participants named as **desired offers** comprehensive and easily applicable tools (five statements) encompassing hands-on best practices and examples (eight statements). They emphasized the importance of increased awareness, acceptance, and tangible value of RDM (one, three and five statements). Additionally, support in workflow management through automation tools (three statements), checklists (two statements), and metadata orientation schemes (three statements) were highlighted. Training sessions raising acceptance and tool utilization as well as best practices were seen as essential for meeting these steps. Lastly, six statements ask for fundamental support in RDM, ranging from definitions of terms on RDM over processes and procedures to step-by-step guidance.

The main requirements emerged from the interviews can be condensed as the following:

- Raise awareness for RDM and incentivate engineering researchers for it
- Reduction of effort in RDM
- Guidance in RDM processes
- Integration of an RDM process into daily work processes
- Streamline and synchronise different valid guidelines for RDM in a research project
- Interconnect different RDM solutions/guidelines
- Automation of RDM processes

As the German engineering research community is recently undergoing a change towards an open access and data sharing culture, we argue that the results are still relevant even though the interviews were conducted three years before this publication. Some early adopters of RDM may have a different priority on the requirements. However, the requirements of engineering researchers will stay unchanged and will rather become a standard over time. Therefore, the results of the interviews are still considered relevant despite originating from 2021.

4.2 Workflow workshops

As another part of the identification of requirements, workshops were conducted. The aim of the workshops, conducted as open interviews, was to capture the research processes of specific research projects the participants were involved in. The interviews were designed to be narrative and unstructured to enable the engineering researchers to describe the research activities from their own experience and perspective. Care was taken to maintain a chronological sequence of activities in order to present the data processing up to the intended research results, embedded in the context of a research project. The participants freely communicated their current research approaches and related processes in detail in an open discussion. While the researchers elaborated on their activities, one of the interviewers documented the described processes, while the other moderated the discussion. The chronological sequence of the process description made it possible to visualise the documented processes during the interviews using a flow chart. In this way, the research processes were described in a research and project-specific manner. The results of the interviews were then analysed and structured. The workshops were conducted from fall of 2021 to spring of 2022. The eight participants were selected from different universities, projects and institutions to form a cross-section of typical engineering projects. They originated mostly from the RWTH Aachen University and Leibniz University Hannover. Engineering researchers from collaborative research centres (ger: Sonderforschungsbereich SFB/Transregio (TRR)) TRR participated within the workshops as well. According to the participants, three kinds of projects have to be considered:

- **Standalone-projects:** performed by a single institute
- **Joint projects:** several participating research institutes
- **Projects with industry-involvement:** small to large consortium containing one or more research institutes and several companies

In total, five + three different engineering researchers were interviewed, which described the research processes in ten + three projects. All of the participants were employed by a research institution and had prior work experience in research activities. Knowledge about ten research projects under the supervision of the aforementioned engineering researchers could be gathered.

Research project kind	Number of projects	Workshop destination
Joint project	6	Aachen
Standalone-project	2	Aachen
Standalone-project with industry involvement	2	Aachen
Joint project	1	Hannover
Standalone-project	2	Hannover

Table 2: Overview of different research process models

In the beginning of the workshops, administrative activities were considered out-of-scope. However, their connection to operational activities - which was the focus - was brought up by all participants. For example, deliverables had to be matched with research activites.

At the workshops in Aachen in connection with the Task Area Frank, no explicit focus was set on RDM, but the “as-is research process” was recorded. Participants were explicitly told that there is no “right” or “wrong”. The research projects derived from these workshops can be split into three levels. Firstly, on the **top-level** there are three phases:*project planning*, *execution* and *completion*. The participants state, that the **top-level** phases happen sequentially, meaning each phase has to be completed before the next phase can start. The **top-level** phases are shown in figure 4 as light blue boxes. Each of the phases contains multiple tasks (white boxes). These tasks A more detailed view on each of the phases can be seen on a **mid-level** perspective. Regarding the mid-level, this paper emphasises on the content-wise execution (i.e. activities containing the generation of knowledge) activities and the publication process. The **low-level** is considered to be the actual task level. Here, the actual methods are conducted and research data is generated.

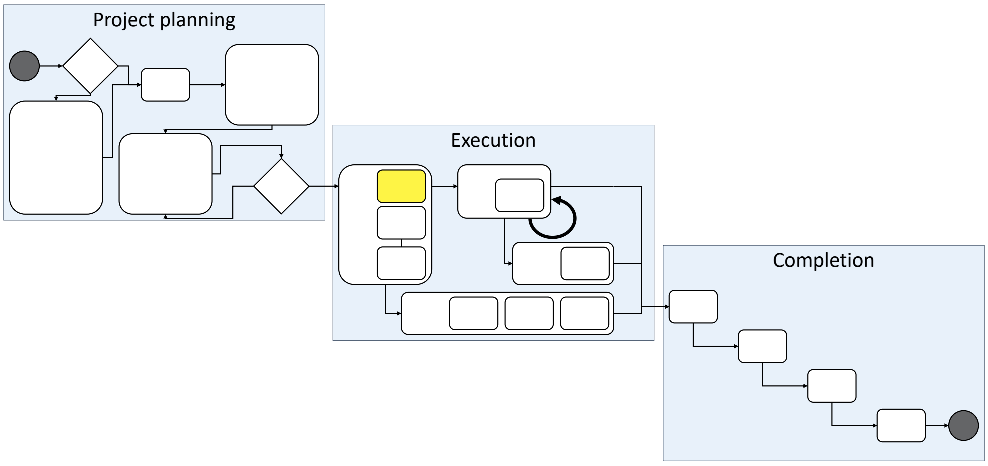

Figure 4: Waterfall like structure of a research project. For a more detailed view, see images 13 to 15 in the Appendix

It becomes apparent, that when conducting the top-level activities, the workflow follows a waterfall approach, which leads to a ”project [...] split into multiple fixed phases, with each phase requiring the analysis and work from the previous phase” [49]. Furthermore, the ”waterfall model assumes that once the initial requirements are set and every goal has been cleared of any ambiguities, there is an unobstructed road which the development team will follow towards finishing the project” [49]. This procedure is exactly resembled by the statements of the participants regarding the project planning and proposal writing. The proposal is considered final for the project execution in terms of initial requirements goals. The same applies for the project completion, as in the execution phase all work packages have to be completed to advance into the next phase. This sequentially structure of the process is essential to waterfall approaches [50]. The same structure is depicted in figure 4 as result of the workshops. Despite the sequential of the **top-level**, the **mid-level** activities can happen simultaneously and in parallel.

As a result of the workshops in Hannover as part of the Task Area Base Services, different research processes were identified among the engineering researchers, depending on their research methods and project structures. Following on from the workshops in Aachen, the aim of the interviews was to record and map research processes and thus identify RDM-related activities in order to integrate them into the research projects as RDM processes. The research processes described could be divided into the **top-level** phases from the workshops described in 4.2. The phases of project planning, execution and completion were described sequentially and could be separated from each other by milestones. Figure 5 shows the research processes abstracted at the top-level with two lanes for the research content level and the data management level in a flow chart.

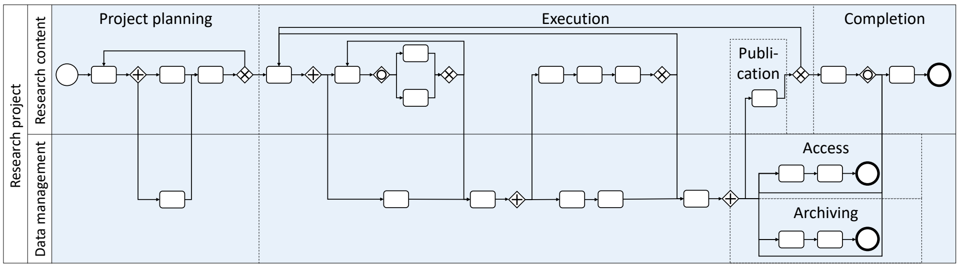

Figure 5: Abstracted flow chart of a research project with data management and research content level

As part of the project planning, the initiated research projects were written down in the form of a proposal. This described the project objective, the sub-objectives to be achieved (usually in the form of defined work packages) and the scope of the project. With regard to RDM, data management plans were part of a funding proposal. After the application phase, if the research project is approved, the content-related execution of the research projects begins with regard to the described schedule. The execution is based on the processing of the defined work packages (**mid-level**) summarised as content-wise execution, whereby these are not only processed sequentially, but also in parallel, depending on the planned process. In this way, the processing of the work packages builds on each other and is partly based on the further processing of results from previous work packages. Within the processing of the individual work packages, a structure of the further planning of the work packages, the data collection (generation or gathering) as a data basis, the content analysis and further processing of the data, through to the research result can be mapped. During this process, the resulting digital objects are stored. The research results are then published in the context of scientific practice. In some cases, the underlying digital objects are also published and archived. The processing of the work packages with the research activities (**low-level**) depends on the research methods and research areas. The processing of the work programme within the defined project duration and the fulfilment of the work programme then leads to the completion of the project. This completion phase includes the documentation of the research results with the writing of a final report, as well as the subsequent archiving and making accessible of the project materials and results, depending on the project framework.

In conclusion, the **top-level** of a research process can be be represented by a flow chart and waterfall project management approach. Within the three included top phases, several steps arise, each of which may be split into sub-steps and tasks.

Nineteen major findings can be pointed out, which were supplemented by a further series of workshops:

1. On the top-level perspective, the workflow resembles the waterfall project management approach.

2. Engineering researchers can relate to the allocation of projects into the work phases *project planning*, *execution* and *completion*.
3. The execution of the project usually takes place in defined work packages or defined sub-goals, in which different research results are achieved.
4. In the content-wise execution, research core data is generated in an iterative way.
5. The content-wise execution consists of several data creation iterations, each of which may depend on none, one or more iterations performed beforehand.
6. During execution, there may be several iterations of data collection in order to collect digital objects that can be processed further.
7. The iterations of the content-wise execution may be terminated if data is found to be insufficient.
8. Completed iterations in the content-wise execution may be considered irrelevant at some point, allowing the deletion of their results.
9. Data is either made accessible and archived during processing after a research result has been achieved and published (e.g. completion of a work package), or at the latest when the project is completed.
10. Data reuse is not considered to be a dedicated step in research projects.
11. In research, data is generated and reused depending on the research approach and research method.
12. Data occurs in different formats and is summarised as digital objects. These are either generated or existing digital objects are gathered.
13. During processing, the collected digital objects are first saved.
14. There may be several iterations of data collection during processing.
15. Research results appear in different forms and are achieved through different aggregations and data processing steps.
16. Only when a research result is achieved the writing of a publication is triggered, together with archiving or access to associated digital objects.
17. The publication process receives the output of either one or many iterations of the content-wise execution as input to transform them into a new form (most often text).
18. There is no new knowledge generated in the publication process.
19. Project management accompanies the research process.

As mentioned before, the results are considered valid despite the fact that the data was gathered from 2021 to 2022. As research processes in engineering will most probably not rapidly change, the findings listed above will mostly keep their validity. The ones most prone to change are findings 10 and 16 as the reuse and publication of data might become more relevant for engineering researchers in the future.

5 Aggregation of the newly proposed research data management process

In this section, the new process for RDM in research projects for engineering sciences is proposed. It aims to better connect the actual research work to RDM practices. As a result of the Collection of research processes and requirements, a combination of waterfall project management and agile approaches are contained in the new process.

The process is based on the research processes collected in section 4.2 and tailored towards the demands of engineering researchers for RDM as communicated by them in section 4.1. This ensures the fulfilment of demands of the researchers questioned that represent the engineering community. Additionally, the development was reinforced by the feedback collected from the community during the conceptualisation. It is then enriched with the knowledge gathered from the Related work. Still, the process has to be refined further in the future with additional evaluation and the current results presented might not be exhaustive.

The result is a process with multiple levels, phases and steps. These are explained below, starting with the overall structure and levels of the process. Then, each level is explained in greater detail. The investigation focused on operational activities instead of administrative ones. Nonetheless, the importance of administrative tasks was brought up frequently in all interviews and workshops. Administrative tasks are therefore included in the proposed RDM process.

5.1 Levels

Three levels of management and research activities are introduced: project (top-level), work package (mid-level) and research data (low-level). As shown in figure 6, the high and the mid-level have a central step that includes one or more passes from the process performed at the level below (grey boxes, 1...n). This structure is similar to the IPO model proposed by Griem et al. [19].

There are two different results from research activities in the form of research data: findings (mid-level) and artefacts (low-level). These two terms shall now be defined in line with the previous results:

> **Finding**: The result of one iteration of the mid-level. A finding may be the result of one work package or the answer to one research question. It was generated by planning what should be found out, compiling one or several artefacts and deriving new knowledge from it.
> *Examples*: Results of an evaluated survey, simulated data from a mathematical model or field data that has been collected via specifically be programmed sensors.
>
> **Artefact**: The result of one iteration of the low-level. An artefact is planned, contextualised and validated research data in its smallest unit. An artefact is a digital object as proposed by Schwardmann [51]. It may be a primary data collection or the implementation of source code. It was generated by planning the artefact, generating the data and performing a check on its validity and, if valid, storing the generated data.
> *Examples*: Raw data of an experiment, source code of an analysis programme, the questionnaire of a survey, the results of a survey, analysed data.

These results are passed to the next higher level and are compiled there. For example, one or several artefacts can be compiled to a finding (see 5.2).

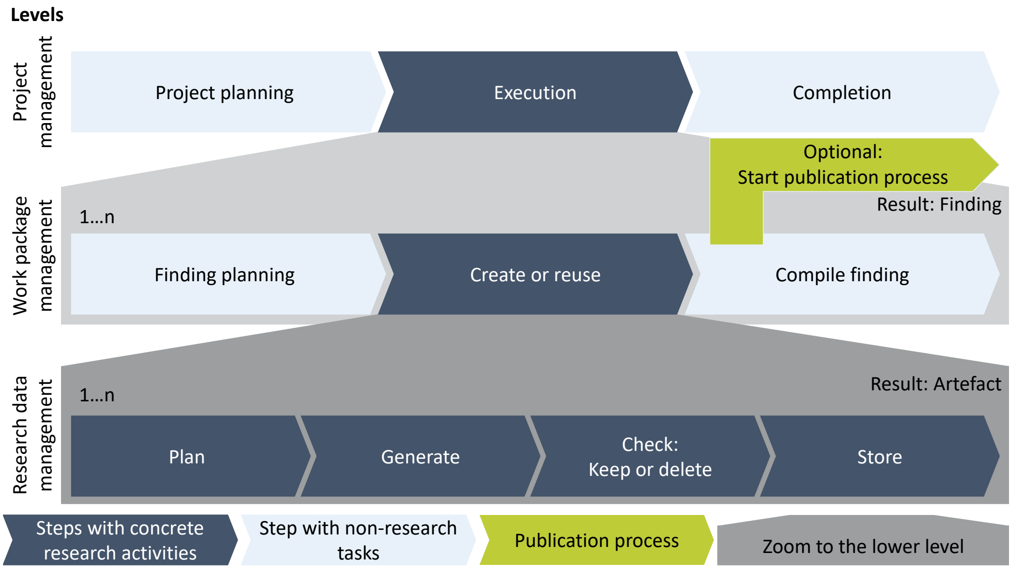

Figure 6: General structure of the proposed RDM process

5.1.1 Top-level: Project management

The top-level is the project management level creates the boundaries for the research project (see figure 7), as "projects are endeavours in which a (more or less) defined result is to be achieved" [25]. For example, funds, resources and deadlines are specified [25]. The three process phases at this level are *project planning*, *execution* and *completion*. The task order shown in figure 7 acts as a guideline and may be adapted if needed.

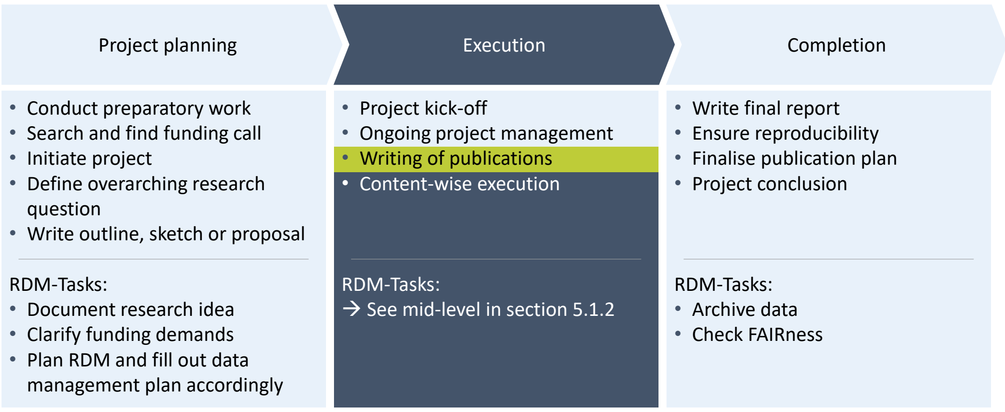

Figure 7: Top-level of the proposed research process: project management

Project planning The project planning starts when a new research idea arises. This idea has to be documented, a funding opportunity and, if necessary, project partners have to be found and a sketch and/or research proposal have to be written. All of this needs to be properly stored as the actual research activity might start much later than when the idea arose. Additionally, several RDM related questions have to be answered, ideally before the research activity starts.

Execution When this step of the top-level is started, the actual research work begins. It is carried out at the lower levels, in the steps at which the artefacts and findings activity are generated. A full description is provided in chapters Mid-level: work package and Low-level: research data management. Administrative tasks in the execution step are the project kick-off and ongoing project management.

Completion The last step is the completion of the research project. This includes the writing of the final report, the ensurance of reproducibility, the completion of the publication plan and project conclusion. A special mention has to be made for the checking of reproducibility and FAIRness of research data as a RDM-Task. This check in the completion step ensures all data is available and stored in a suitable way to start the archiving process. The actual task of ensuring reproducibility and FAIRness is situated in the low level (see section 5.1.3) as it comprises of the FAIRness of the artefacts.

5.1.2 Mid-level: work package

The work package level describes the content-wise execution of the research activity. Here, findings are generated from compiling artefacts, answering (sub) research questions or concluding work packages. As shown in figure 8, firstly the targeted finding is defined by selecting the (sub) research question or work package to be addressed. Then, artefacts are generated or reused to finally compile findings from them.

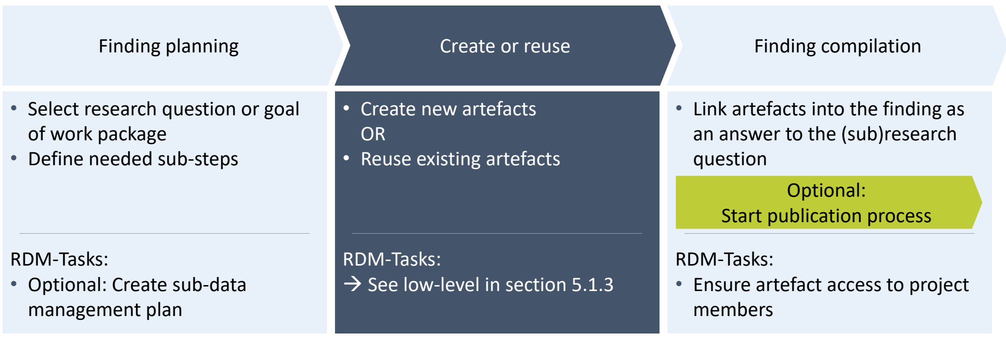

Figure 8: Mid-level of the proposed research process: work package

Finding planning When findings are planned, it is decided which work package or research question should be answered and how it should be answered. Therefore needed data or information are determined and methods for their gathering are selected. Afterwards tasks can be derived from that. Optionally, a sub-data management plan can be created, optimally within a hierarchy of data management plans like offered by RDMO.

Create or Reuse When the needed data or information is known, they can be gathered in the form of artefacts. They are either originally generated, or reused either from own previous works or project-external sources. This is an iterative and agile process and carried out at the low-level. A full description provides section 5.1.3.

Finding compilation After all needed artefacts are generated or gathered, it comes through compiling them to a finding, that will presumably hold the answer to the research question. It then has to be documented, which artefacts were involved in the compiling process. As each artefact contains a documentation of its creation, see 5.1.3, this documentation is already available partly. At this point at the latest, access to all relevant artefacts has to be granted to other project members. As an optional task, a publication can be written when the finding was compiled. A full description of this process provides Publication Process.

5.1.3 Low-level: research data management

On this level, the Core research activities are conducted. Data is either created, gathered, collected, aggregated, analysed or processed in another way. This data provides the basis for the content generation of knowledge in the content-wise execution on the mid-level. When comparing the results from section 4.2, a strong resemblance between the engineering researchers' activities and the DLC becomes apparent. Therefore, an artefact is generated through a process inspired by the data life cycle. Firstly, data generation is planned, then the data is generated or collected. Afterwards it is checked for validity and either deleted or stored for further usage. All of these tasks aim to ensure the Findable, Accessible, Interoperable, Re-usable (FAIR)ness of the generated artefacts by default. This process is presented in figure 9.

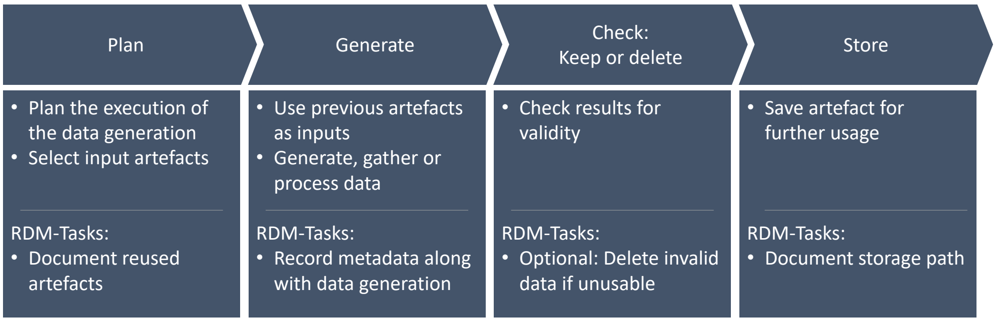

Figure 9: Low-level of the proposed research process: research data management

The proposed process can also be displayed in a new form of the DLC (see figure 10). In addition, the process has an option of canceling the creation of an artefact when it is invalid.

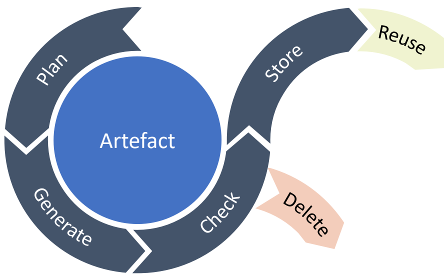

Figure 10: Procedure of artefact generation as a new form of DLC

Plan "Plan" at the low-level describes the concrete planning of the data acquisition task (generating or gathering). This includes planning - and documenting - whether and which previous created artefacts are used to input the new artefact. This step distinguishes from the Project planning and the Finding planning by the level detail. Both of them divide the respective lower level into smaller steps and tasks - this level is the lowest.

Generate When generating data, engineering researchers perform the typical tasks associated with research. This is when someone conducts an experiment, runs a simulation, conducts an interview or research using another approved method. This step can have no, one or multiple inputs in form of previous created artefacts.

Check: Keep or delete After the artefact was created, it is checked for validity. This does not include an analysis to generate new knowledge from it but the verification of correct measurement and data recording. For example, if sensor data is all zeros, the sensor might have been disconnected; rendering the artefact invalid, allowing for its deletion. If, for any reason, the artefact is not valid, it may be deleted if there is no knowledge to be possibly derived. Invalid artefacts may still hold knowledge e.g. on why the artefact is not valid. In that case, the artefact should not be deleted. Artefacts which are used as input for other artefacts must not be deleted to be able to trace the finding integration (see Artefact and finding integration)!

Store If valid, the artefact is saved for further usage. At this stage, it does not need to yet be shared with other project members or external parties. The storage in a save environment with backups is recommended to prevent data loss. If this artefact holds the basis for the answer to a (sub-)research question or the solution to a work package, it is taken one level up to the Mid-level: work package to generate new knowledge from it.

5.2 Artefact and finding integration

Artefacts can receive input in form of one or more other artefacts (project internal and external) in their data generation step. There is no limit to the number of artefacts combined in that way. A combination example is shown in figure 11. The finding that results from these combination answers a (sub)research question and can initiate a publication process.

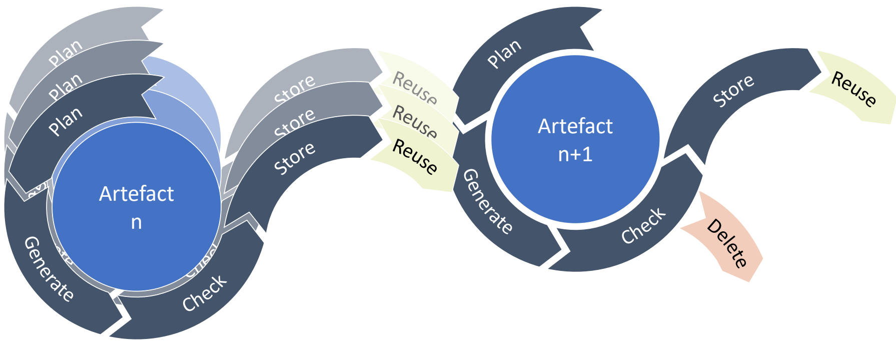

Figure 11: Procedure of artefact combination: Many to one relation

5.3 Publication Process

The term "publication process" here refers to "scholarly communication[, which] is the system through which research and other scholarly writings are created, evaluated for quality, disseminated to the scholarly community, and preserved for future use" [52]. In engineering research, a publication process is usually initiated by a finding compilation (ref. 5.1.2). It can therefore be carried out at different points in the research process, several times. Although it has synergies with research and RDM tasks, it is a standalone process and its execution is independent of other activities. Figure 12 shows the tasks (cf. [53], [54]) in a publication process and how they are linked to the RDM.

After the topic of the publication has been determined by a finding compilation, preparatory activities such as the selection of a suitable publication medium are carried out. Other steps in the research process have an influence on these decisions, e.g. requirements of funding organizations. The editing step involves transferring the knowledge generated by the artefacts and findings into text. In addition to the text, the artefacts and findings described in the publication are presented in further forms, e.g. plots or source code. As described in Artefact and finding integration, the result of a work package is a final finding; the text usually only refers to this. The contents of the previous data generation steps – well documented by RDM activities – are therefore only described in text form as part of the methodology. In order to meet the requirements of the RDM, the artefacts must be published and referenced as part of the publication process in accordance with the project regulations (see section Project planning). In this way, the data processing steps for achieving the research results can be reconstructed by the artefact compilation and associated artefacts can be reused. Therefore the artefacts must be prepared within the framework of the FAIR Data Principles. This makes it easier for the peers to carry out a review after submission to the publication medium in order to check the quality of the text. After publication it is possible to disseminate only the research data, only the text content or all together as a package, which allows particularly target group-oriented communication for dissemination.

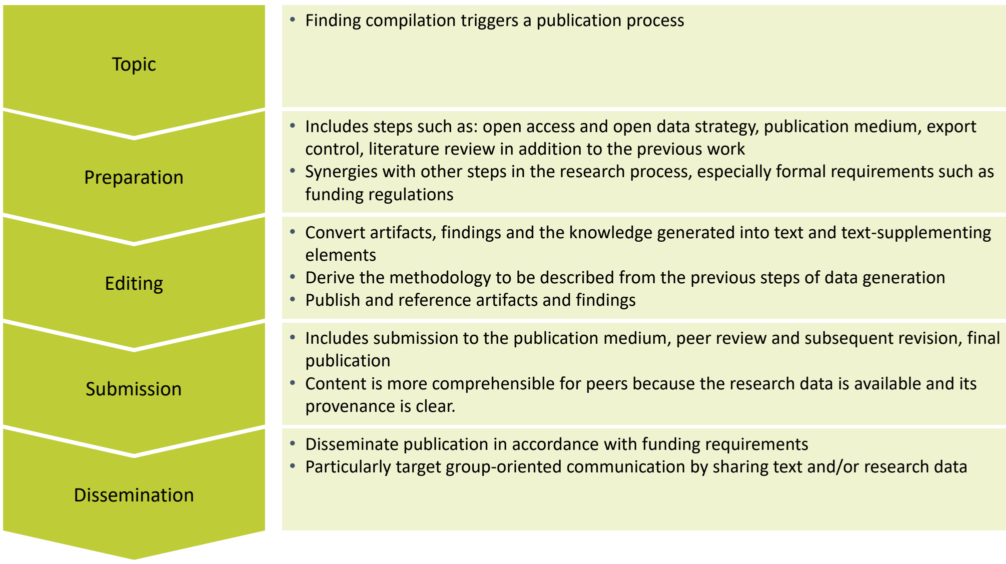

Figure 12: The publication process

6 Discussion

In this article, we introduced a process for RDM in engineering research. This process is embedded into the engineers research activities. Hence, we consider the research question, if RDM can be embedded as a process into engineering research projects, answered positively. Furthermore, the research gap found, that RDM should fit the research process and be integrated into it rather than be added as an afterthought, was addressed. As we used the requirements formulated by the researchers as a basis for the development of our process, they are manifested within the process fundamentally. Still, more requirements might to be formulated by other engineering researchers and therefore further adjustments might need to be made in the future. The validation, in which such emergence of new requirements is mostly expected, was not yet performed.

The presented process (see figures 6 to 10) was presented in several events (see [17], [43]–[45]) and was quite well received. We collected feedback from those events and included it into the development up to this point. Yet, a formal evaluation has to be conducted. For this purpose, a wide spread and statistically reliable validation of the proposed process is planned.

Additionally, the limitation to only project-based research is a major restriction. While the decision was well supported by literature, it is difficult to estimate to which extend research is conducted in projects or in other forms like e.g. research supported by basic funding [2], [13]–[16], [47]. To strengthen this point of view, a question about the type of research (project or non-project related) will be included a validation survey following this publication.

It also has to be considered, that the activities and RDM-tasks described and presented in figures 7 to 9 will most certainly be incomplete. Due to the effortful collection of requirements and processes, rather small groups of people were interviewed. Hence, new activities and tasks might arise within the further evaluation of the process.

While the process is meant to be adaptive to different research methodologies used on the **mid-level** and methods applied on the **low-level**, the overall structure of the process is not adaptable. This is due to the nature of research in engineering sciences: Firstly a research question is formulated, then suitable methods are chosen before data is collected. Afterwards new knowledge is derived from the analysed data. This structure has a logical order which should not be reorganised. Hence, major revisions of the process like new levels or an overall new structure should be avoided while minor adjustments are still possible, like, for instance, adding new activities and tasks.

7 Conclusion and Outlook

In this article, a new process for RDM in engineering sciences was presented. The aim was to tailor it towards the needs of engineering researchers and make it possible to integrate it into the everyday work in engineering research. For this aim, requirements were gathered as well as typical research processes from interviews and workshops.

The result is a process with three levels and a differentiated structure. It contains both a waterfall project management perspective on the **top-level** as well as iterative approaches on the **mid-level** and **low-level**. The latter resembles a DLC in it semi-circular design, integrating the often cited DLC model into the process.

In a future work, the applicability of the process will be evaluated. If valid, the model might fill an important gap in research, connecting existing RDM models with the actual reality of research. A transfer to other disciplines than engineering might also be possible.

Eventually, the evaluated process will be implemented into Jarves, the Joint Assistant for Research in Versatile Engineering Sciences [3]. This digital data steward aims to guide engineering researchers in their RDM throughout their research project. In addition to the process presented in this article, it will feature the management of RDM guidelines, a decision support system, linkage of training materials at the point of need and a partial automation of RDM-related tasks, e.g. in the synchronisation of data between different RDM tools.

8 Appendix

To further elaborate on certain topics and give corresponding examples if needed, the appendix contains tree major parts. Firstly additional information on the workflow workshops is given in section 8.2. Additionally, an exemplary project's fit into the proposed process is presented in section 8.4. Lastly, a Glossary contains the most important terms of this article.

8.1 Additional information on the problem-centered interviews

The interviews were originally designed to be focus groups. The method was planned to be based on Gail and Vetter's approach [55]. However, the groups were not always reaching the number of participants needed for focus groups, which is set to five to eight people [56], as the number of participants varied between one and six. A greater number of participants did not prove to be efficient for the process, as the interviews were conducted virtually due to the Covid-19 pandemic.

Therefore, the interviews were instead considered to be problem-centered interviews rather than focus groups [48], [57]–[59]. Problem-centered interviews are applicable in this context, as the method should provide "first-hand insights into" [48] the researchers RDM practices, showing "what, how and why [...] actions, appraisals and opinions" [48] are present among engineering researchers. Both methods do not differ from one another in terms of preparation and evaluation and inherit similar methodological premises. Also, both methods follow a semi-structured approach [55], [60]. In this case, the structure was provided by an interview guideline. The main difference lays in the absence of controversial opinions within the groups of interviewees. Yet, both methods are based on the principles of qualitative research to encourage the interviewees to share their actual point of view [56], [58], [60]. For the evaluation, the changed format does not result in a different way of analysing qualitative answers. Main statements are summarized in a methodological controlled way as proposed by Mayring [61].

3. https://jarves.nfdi4ing.de/

8.2 Additional information on the workflow workshops: TA Frank

To give further insight on the procedure and derivation of processes, this section of the appendix describes the workflow workshops of TA Frank in greater detail.

8.2.1 Description of the workshop programme

Firstly, the interviewees were introduced with a short slide deck, explaining the procedure and goals of the workshop as well as the planned usage of the gathered information. As the interviewees had previous knowledge about the NFDI4Ing and the interviewer's involvement in the context of RDM, an explicit disclaimer was given, declaring that there is no explicit focus on RDM to avoid bias. The engineering researchers were then asked to describe their usual research process from the beginning to the completion of a research project. This included the naming of possible entry points as well as their definition of the point of time when a project is completed. They were then asked to describe their process between these two points. The documentation of the processes was explicitly not based on any process model or method of process modelling so that there was no risk of a bias of the interviewees.

8.2.2 In depth: Top-level perspective

To illustrate the top-level perspective, the following describes the interviewees perspective. These results have already been condensed and homogenised in their wording. On this level, the interviewees mentioned the most administrative tasks in their research. Therefore, administrative activities were taken into account, although being originally out-of-scope. These tasks are closely related to research, even if they do not create new knowledge, but rather support its generation.

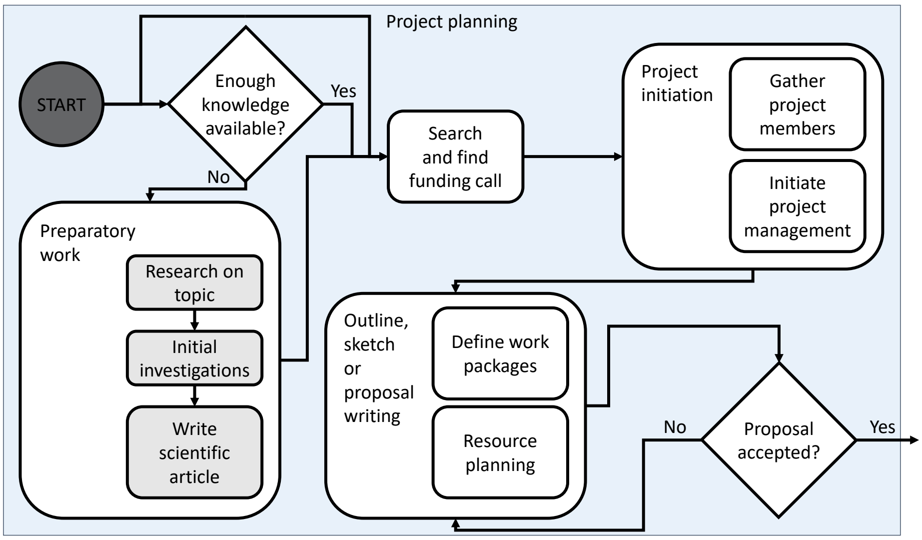

Figure 13: Typical activities and decisions in the project planning phase

Condensing the interviewees answers, the *project planning* of a research project contains all the planning activities depicted in figure 13. If not enough prior knowledge of the topic is available at the institute, preparatory work can be performed to enhance chances of proposal acceptance: By conducting an in depth research on the state of the art, performing initial investigations like feasibility checks or the writing of a scientific article. This may become obsolete, if enough expertise is already at hand. Most often, the project planning includes the search for a suitable funding call or starts with the finding of one. An initial project management for the proposal creation is then set up and project members, both internal and if needed external, are gathered. Then a research proposal is written, which in turn needs research and sometimes pre-studies. Additionally, the proposal defines work packages and allocates resources. Lastly, it has to be mentioned that some funding proposals are a staged procedure with a sketch and a proposal as two separate documents, which is depicted in figure 13 as an unaccepted proposal because even if the sketch is accepted a proposal has to be written and accepted.

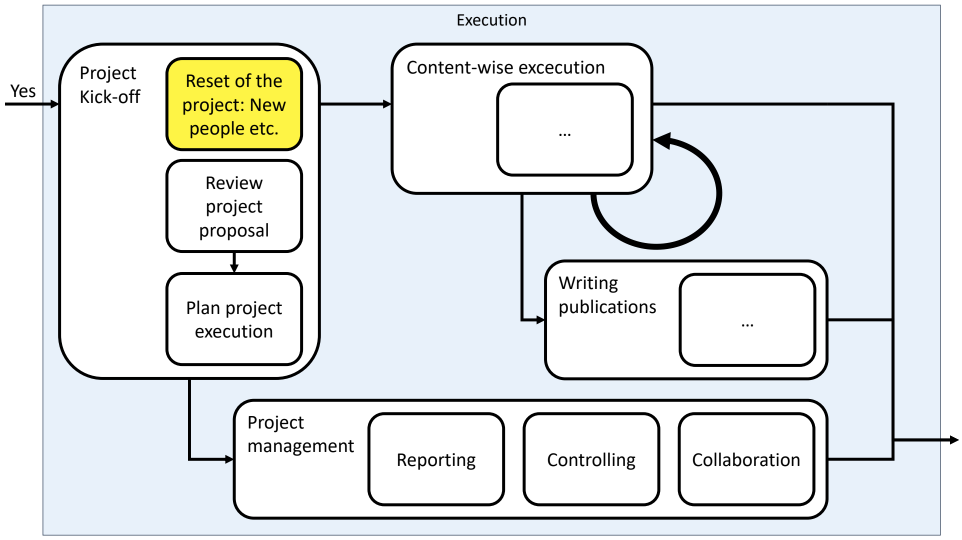

Figure 14: Typical activities and decisions in the execution phase

The *execution* has four major components (c.f. figure 14): The project kick-off, the content-wise execution, writing of publications and the project management. As the interviewees described, each of these steps is highly different from the other.

The project kick-off is depicted as an event at the start of a research project, which may even spread over the course of several weeks. Due to the fact that it often happens months or years after the proposal has been handed in, the project kick-off is a reset of the previous work on the project. The interviewees pointed out, that the actual project members are often not the ones who have written the proposal, because those people have already left the institution. This causes that the knowledge of the project's original idea is partially lost, as the underlying thoughts can hardly be represented within a research proposal. As a result, the project kick-off has to contain a review of the proposal to identify needed work packages and plan their execution.

The content-wise execution is the actual research work performed in the project over the course of several months to years. Here, data is collected which generates new knowledge. This data shall therefore be disclaimed as research core data. In contrast, information such as involved and responsible members, allocation of resources or discipline can be described as research project data. The content-wise execution contains many individual and highly context related tasks of data generation, preparation, analysis and storage. This step is explained in detail later in this section and is not portrayed in detail in figure 14.

When it comes to the writing of publications, the workshop participants divided. Some stated a direct relation to the content-wise execution. Others argued that the publication process does not generate new knowledge, but rather transforms existing knowledge from the research core data generation into a text. So publications have to be differentiated from the content-wise execution, while founding on it. Again, the detailed processes are not depicted in figure 14.

The fourth part is project management. It is relevant in the complete *execution* phase as it relies on the project kick-off's results and controls the content-wise execution. Project and financial reports, the coordination of tasks and collaboration aspects are all relevant in this unit.

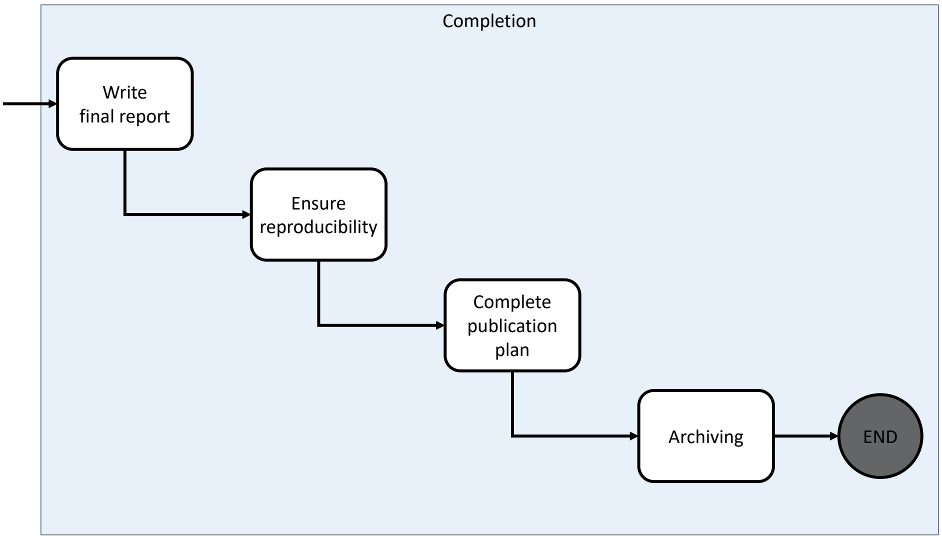

Figure 15: Typical activities and decisions in the completion phase

Lastly, the *completion* phase (see figure 15) contains several steps to finish a project. The mandatory writing of a final report is usually the first of the steps, but the order in which further steps are performed may vary. Figure 15 shows a possible work order. The reproducibility of data should be ensured, meaning that all needed files and data are collected and are transformed to long term storable formats. The interviewees described a publication plan that has been laid out in the beginning of the project and contains information on data, software and papers to be published during or after the research activity. This plan has to be completed. Eventually, the data, source code and publications as well as all other documents are archived, as of now most often on institute hard drives as the participants stated.

8.2.3 In depth: Content-wise execution

The content-wise execution will now be explained in greater detail, as it is the most complex step. Figure 16 depicts the main question arising in this step: What kind of data that shall be created? Interviewees named several different data creation methods, for example experiments, data preparation and analysis, simulations, programming, surveys, 3D-modelling and printing, prototype construction, process analysis and others. In figure 16, these are clustered by the kind of data these methods create. Adding to the complexity of the step, it also contains iterations as the data created in each cycle of the iteration is fed back into a new cycle (see blue arrow in figure 16).

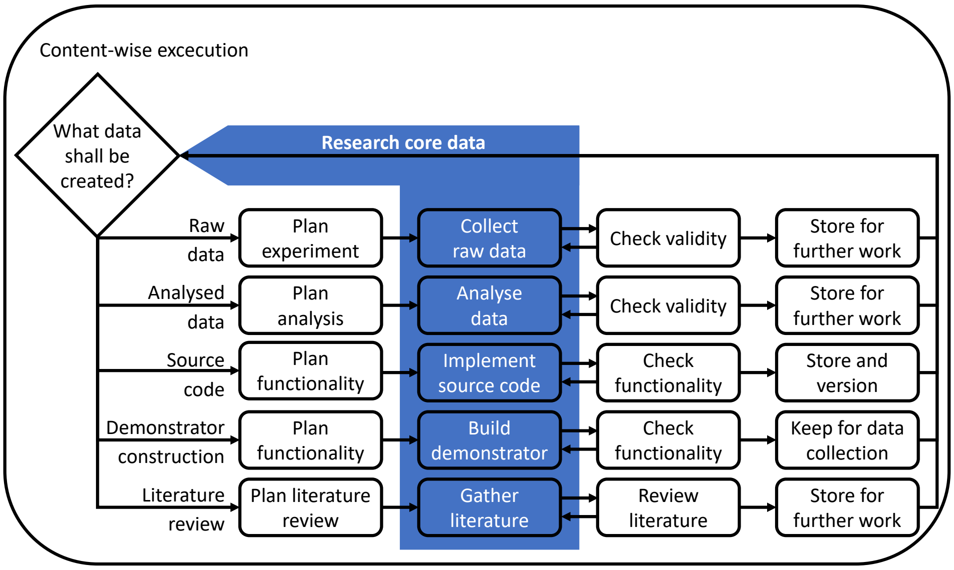

Figure 16: Procedure of research core data generation in the content-wise execution

The term raw data represents any recorded or created data ranging from experiment data to interview answers. The collection of raw data may have other data (both virtual and physical) as input, e.g. a questionnaire or experimentation rig. The main distinction between raw data and analysed data can be seen in the knowledge generation. While raw data does not generate any new knowledge, analysed data does.

Raw data collection is firstly planned in terms of goals, methods and qualitative criteria. It is then collected and checked for validity. The validity check does not extract any new knowledge rather than a review of the plausibility and reasonability of the data. If not valid, data can be deleted as it has no use and the data collection has to be repeated or found infeasible for the targeted results. If valid, the data collected is stored for further processing in the next iteration, as shown with the blue arrow in figure 17.

In contrast, analysed data contains any data that is processed in any way that new information can be derived from it. This might be due to visualisation of raw/already analysed data or as data is combined, e.g. when simulating new data with a simulation model and measured input data. After planning the analysis, its conduction involves the usage of data to be analysed as input and optionally more input factors like, for example, analysis source code.

A different perspective has to be put on source code. Used for data collection, data analysis or data manipulation along with many other utilisation possibilities, the term “source code” describes any code written and implemented in a research project. Firstly, its functionality is planed. Then a most often agile feedback loop is used to move between implementation and functionality checking. When ready for application, the code can be stored and versioned, meaning that changes to the code are allowed and accounted for.

When is comes to physical objects, the process of demonstrator construction was brought up by the interviewees. These are firstly planned and then built. If usable, they are stored (in a physical form) for further usage or (raw) data collection.

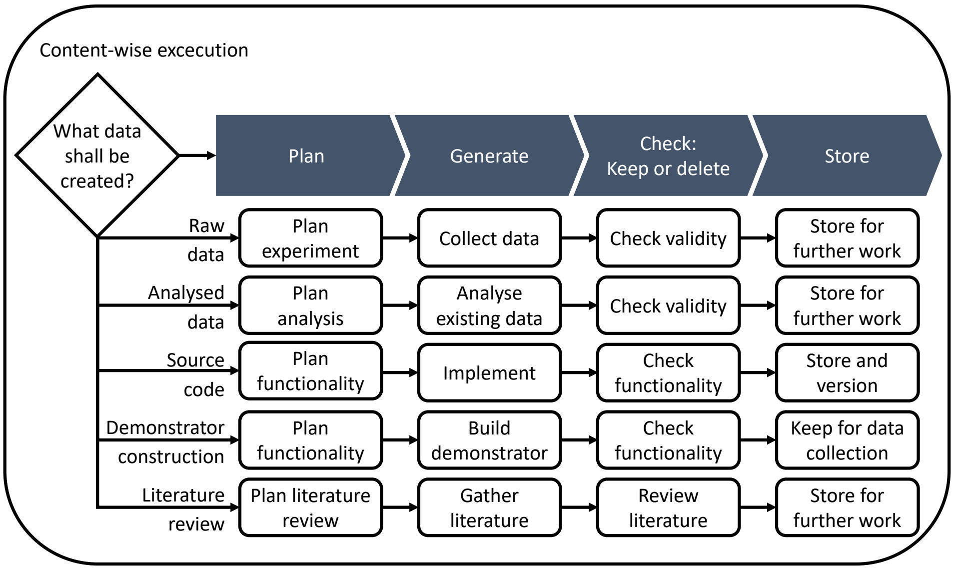

Figure 17: Analogy to a DLC in the content-wise execution step

Lastly, literature reviews are described by the workshop participants. While the collection of literature is the main aspect of this reuse of existing data stated by the interviewees, it should not be considered the only purpose of this step. According to the interviewees, any data gathered not originating from the research project itself can be clustered in this category. This includes reused data or source code as well. This reuse is planned and the external data is then gathered and reviewed. If sufficient, it is stored for further usage, else more data has to be gathered own data have to be created.

As stated before, the content-wise execution is an iteration over several data and knowledge generation methods. An example of the relation of the content-wise execution and the results generated can be seen in figure 18. Here literature is gathered to generate a interview guideline, which is used to conduct interviews. Simultaneously, software packages are gathered and an analysis code is written. The information gained by the interviews is lastly passed into the code and analysed, which creates the research result.

As stated by all interviewees, the content-wise execution is an iterative process. As a result, cancellations have to be accounted both within the content-wise execution in the check/review step (see figure 16) as well as between individual iterations.

8.3 Additional information on the Workflow workshops: Base Service S-1

To give further insight on the procedure and derivation of processes, this section of the appendix describes the workflow workshops of Base Service S-1 in greater detail.

8.3.1 Description of the workshop programme

Firstly, the interviewees were introduced with a short slide deck, explaining the procedure and goals of the workshop as well as the planned usage of the gathered information. In this workshop, a focus was placed on the description of research activities and associated RDM tasks, as the

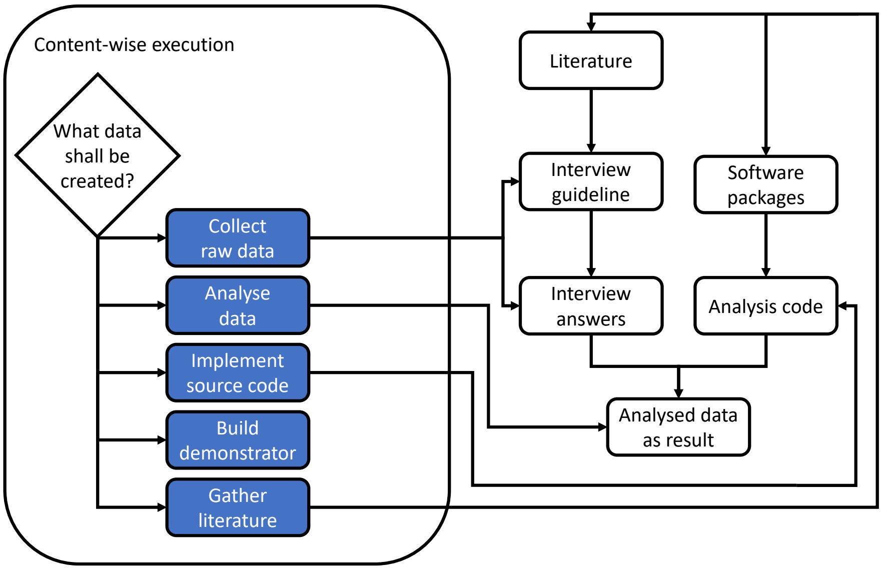

Figure 18: Example on how the content-wise execution compiles different data generation methods into results

interviewees already had prior knowledge of the NFDI4Ing and the interviewer's involvement in the context of RDM. The engineering researchers were then asked to describe their usual research process from the beginning, with the writing of a research proposal, to the completion of a research project. Milestones and processes were described and documented within these points. The business process modelling notation was then selected for the process representation and visualised as a flow chart. Two levels with two lanes were defined to represent the research activities (content processing) and the RDM activities (data management), see figures 19, 20 and 21.

In the following, the identified research process is described from the interviewees' descriptions. The contents have already been compiled and standardised in a generally applicable description. At this level, basic steps and activities are considered that are carried out by the engineering researchers in research projects without affecting the content of the work. The entire research process within a research project can be divided into three major process stages with defined milestones. The project planning activities start with the definition of the research topic and research objectives. This leads to the preparation of the proposal, whereby the RDM is also planned and documented in the form of a data management plan. The resulting application is then submitted to the relevant funding organisation. If this is accepted, the first milestone is reached with the funding of the research project, which also marks the end of the project planning phase.

This is followed by the start of the research project, which marks the start of the content-related work, the execution. The work packages are processed following the project plan, based on the research proposal. These are processed one after the other, sometimes building on each other. The work packages are often described with the results to be achieved. Based on the work packages, data foundations are created as part of the data collection. The implementation of the work packages is planned by defining the research method, the data collection method,

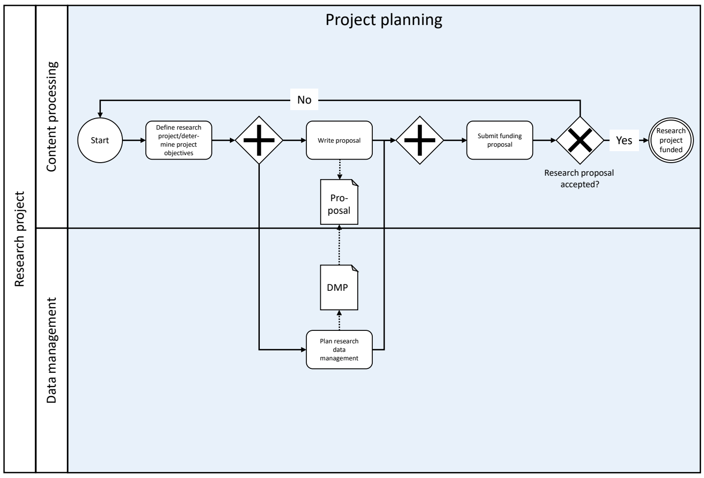

Figure 19: Flow chart extract of the project planning phase

the data collection procedure, and the further processing within the scope of the work package. The collection is then planned based on the work packages concerning the research results to be achieved. The ongoing data-related documentation also starts at this stage. Once the data collection has been planned, it is carried out and new data is generated or existing data is gathered and utilized. If the data collection proceeded as planned and can be processed further, it is saved together with the documentation of the planning and implementation. If the data cannot be processed further due to insufficient quality, the data collection must be planned and carried out again. If the data can be processed further, it is filtered and aggregated. Based on this, the data is analysed to answer one or more defined research questions. These data analysis steps are documented for traceability and the underlying data processing is versioned. The research data generated from planning to analysis as part of context wise execution can involve different types and numbers of research core data. If the research result corresponds to the expected results of the work package, the data will be saved. The underlying data of the research result is then merged with the associated data documentation. The database then leads to the next milestone of the secured research result.

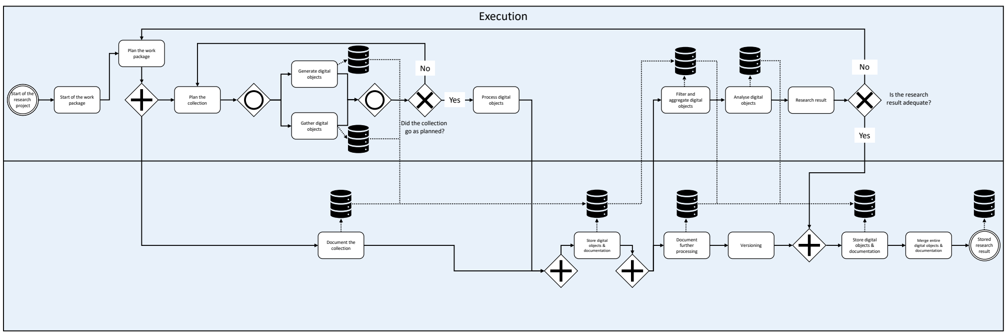

Figure 20: Flow chart extract of the execution phase

When the research result is achieved, decisions are made about the publication of the content and when the content work is finalised, the completion phase is initiated, as shown in figure . a decision is made whether it should be published as part of a text publication; at the same time, a decision is also made whether the associated database should be made accessible and archived. Archiving, and access require the selection of data and the securing of file formats to ensure interoperability and reusability. Additional metadata for documentation and the selection of an archiving service or an access platform (repository) are also defined. If the data is made accessible in this way, it can be cited directly as part of the text publication. The publication is then submitted to the chosen scientific publisher.

If further work packages need to be processed as part of the project, the process starts again from the beginning of the execution phase. Once all work packages have been successfully completed, the content of the project can be finalised and the project can be closed. Depending on the requirements, a project report is written. It is also checked whether all data has been properly archived or whether data still needs to be published. Once this has been done, the project can be finalised.

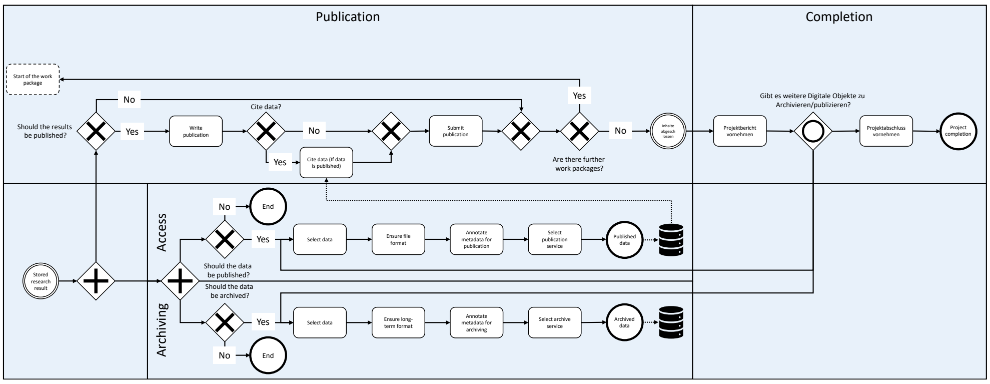

Figure 21: Flow chart extract of the publication and completion phase

8.4 Exemplary explanation of the proposed process

As previously explained, the research project that provided the context for this article is the NFDI4Ing [47]. The project also follows the structure of the proposed process. However, the extreme size of the project would over-complicate the following example. Therefor only a fraction of the project is displayed. This fraction is the the digital RDM-assistant Jarves[4], which is developed within NFDI4Ing. While Jarves and NFDI4Ing both follow the waterfall-based approach on the 5.1.1 as proposed in figure 4, the generation and especially linking of artefacts has to be explained in greater detail. To illustrate this process, firstly the linkage of artefacts is displayed before the meaning of this linkage is set into context of findings. Lastly the view is broadened to the whole project of Jarves, depicting an exemplary structure of artefacts, findings and their connections.

Figure 22 shows the exemplary artefact structure of the research behind this paper. In this depiction, "Finding 1" is the conceptual model as it is up until this section. Artefacts can have multiple previous artefacts as input, independent of the time of their creation and independently from prior usage as input. Therefore, artefacts have a many to many relationship with their predecessors and successors.

4. https://jarves.nfdi4ing.de/

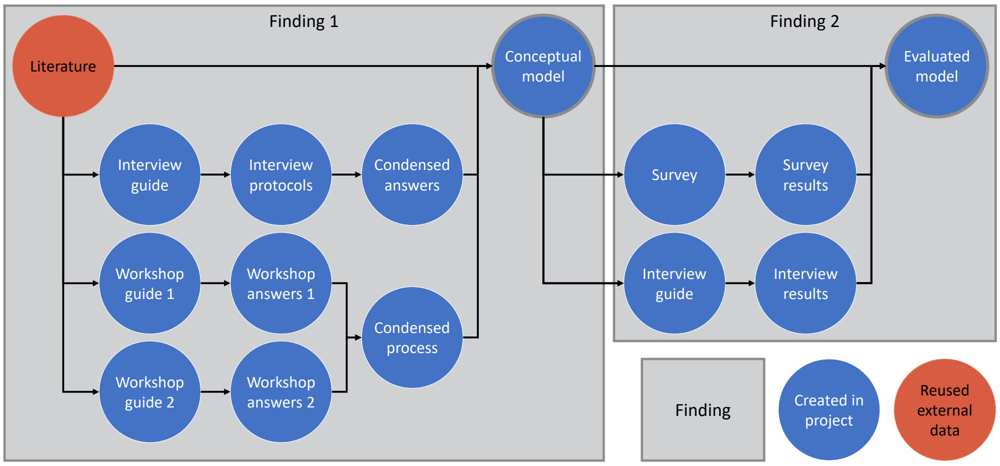

Figure 22: An example of artefact combination

On the Mid-level: work package, findings refer to (sub-)research questions. This is shown in figure 23. Optionally, a finding can result in a publication, e.g. the one you are reading now.

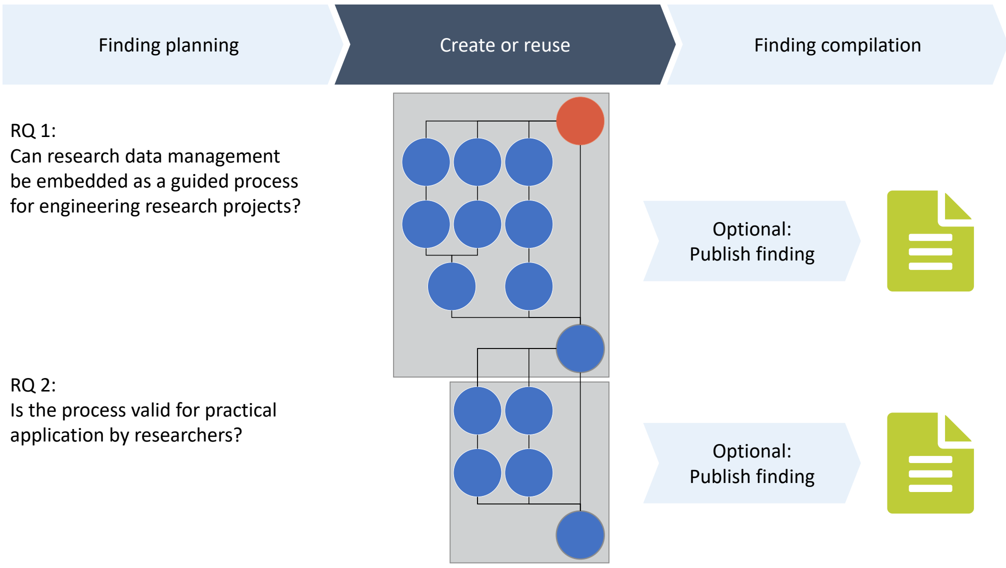

Figure 23: An example of finding combination

Furthermore, findings can be integrated further as the perspective is set on the complete project. For instance, this paper is part of the goals of NFDI4Ing's Task Area Frank. There, Jarves, an assistant for RDM, is being developed, that eventually will integrate the final process model. An exemplary and simplified view on the project structure of Jarves is given in figure 24. There, the integration of artefacts to findings and findings to the whole research project is depicted. Furthermore, the results serve as a design level for the architecture of RDM maturity models in the NFDI4Ing Task Area ”quality assurance in RDM processes and metrics for FAIR data”.

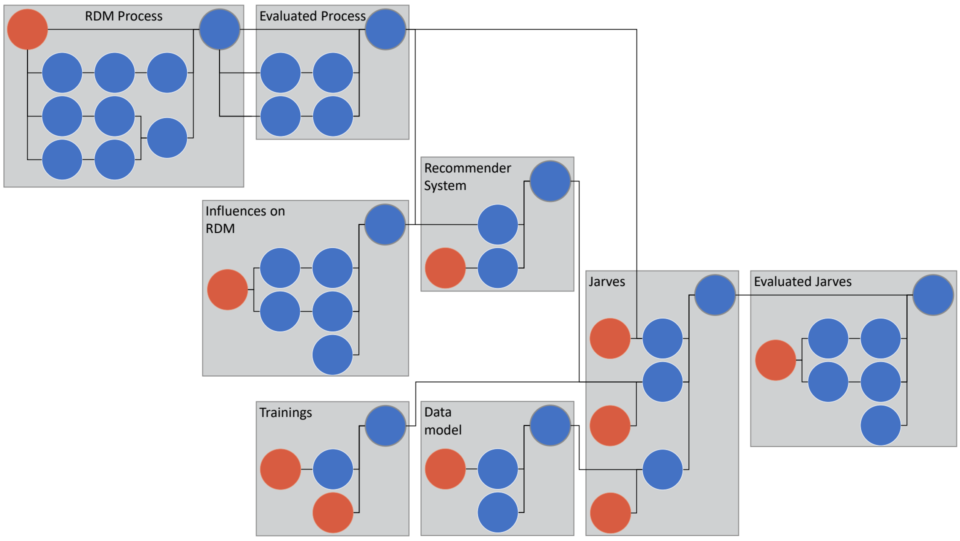

Figure 24: An example of the research project Jarves

8.5 Glossary

Analysed data	Any data that is processed in any way that allows for new information to be derived from it. **Low-Level. Iterative.**
Artefact	The result of one iteration of the low-level. An artefact is planned, contextualised and validated research data in its smallest unit. It may be a primary data collection or the implementation of source code. It was generated by planning the artefact, generating the data and performing a check on its validity and, if valid, storing the generated data. **Low-Level. Iterative.**
Completion	The last of the three research project phases. Can only be started if the execution is finished. **Top-Level. Waterfall.**
Content-wise execution	All activities on the mid-level containing the generation of knowledge. **Mid-Level. Iterative.**
Core research activities	All activities on the low-level containing the generation of research core data. **Low-Level. Iterative.**
Data/Research data	“Research data includes measurement data, laboratory values, audiovisual information, texts, survey data, objects from collections or samples that are created, developed or analysed in the course of scientific work. Methodological test procedures such as questionnaires, software and simulations can also represent key results of scientific research and should therefore also be categorised as research data” [29]. **Low-Level. Iterative.**

Execution	The second of the three research project phases. Can only be started if the project planning is finished. **Top-Level. Waterfall.**
Finding	The result of one iteration of the mid-level. It may be the result of one work package or the answer to one research question. It was generated by planning what should be found out, compiling one or several artefacts and deriving new knowledge from this process. **Mid-Level. Iterative.**
Research data management (RDM)	The handling of research data (collection, organization, storage, and documentation) during and after a research process.
Project	Unique activity with a defined outcome and predetermined resources in terms of costs and time [25].
Project-oriented research	Research with a planned and defined outcome and predetermined resources in terms of costs and time (c.f. [25]).
Project planning	The first of the three research project phases. The research project starts with this phase. **Top-Level. Waterfall.**
Raw data	Any data recorded or created, but not yet analysed. Does not generate any new knowledge. **Low-Level. Iterative.**
Research core data	The research data collected in a research project, which generates or is used to generate new knowledge. **Low-Level. Iterative.**
Research project	Project to investigate a specific research question.
Research project data	The data used to describe the information about a research project without the creation of new knowledge.

9 Acknowledgements

The authors would like to thank the Federal Government and the Heads of Government of the Länder, as well as the Joint Science Conference (GWK), for their funding and support within the framework of the NFDI4Ing consortium. Funded by the German Research Foundation (DFG) - project number 442146713.

10 Roles and contributions

Tobias Hamann: Idea, Conceptualization, Methodology, Writing

Michèle Robrecht: Conceptualization, Methodology, Writing

Max Leo Wawer: Conceptualization, Methodology, Writing

Jonas Maximilian Werheid: Conceptualization, Methodology, Writing

Mario Moser: Conceptualization, Review

Marcos Alexandre Galdino: Writing - Review

Anas Abdelrazeq: Writing - Review

Roland Lachmayer: Writing - Review, Supervision, Funding acquisition

Robert H. Schmitt: Writing - Review, Supervision, Funding acquisition

References

[1] S. Higgins, "The DCC Curation Lifecycle Model," *International Journal of Digital Curation*, vol. 3, no. 1, pp. 134–140, 2008. DOI: 10.2218/ijdc.v3i1.48.

[2] W. Krull and A. Tepperwien, "Neue Herausforderungen für die öffentliche und private Forschungsförderung," in *Handbuch Wissenschaftspolitik*, ser. Springer Reference Sozialwissenschaften, D. Simon, A. Knie, S. Hornbostel, and K. Zimmermann, Eds., Wiesbaden: Springer, 2016, pp. 447–463, ISBN: 978-3-658-05454-0. DOI: 10.1007/978-3-658-05455-7{\textunderscore}17.

[3] D. Williams and H. Tang, "Data quality management for industry 4.0: A survey," *Software Quality Professional*, vol. 22, no. 2, pp. 26–35, 2020.

[4] Y. Roh, G. Heo, and S. E. Whang, "A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 4, pp. 1328–1347, 2021, ISSN: 1041-4347. DOI: 10.1109/TKDE.2019.2946162.

[5] T. Koltay, *Research Data Management and Data Literacies* (Chandos Information Professional Series). Amsterdam: Elsevier, 2022, ISBN: 9780128244753. DOI: 10.1016/C2020-0-02068-2.

[6] M. D. Wilkinson, M. Dumontier, I. J. J. Aalbersberg, *et al.*, "Comment: The FAIR Guiding Principles for scientific data management and stewardship," *Scientific data*, vol. 3, 2016. DOI: 10.1038/sdata.2016.18.

[7] *Wissenschaftsbereiche: Spezifisches FDM aus den Fächern und für die Fächer*, 2023. [Online]. Available: https://forschungsdaten.info/wissenschaftsbereiche/ (visited on 01/12/2024).

[8] A. Wuchner, M. Robrecht, P. Kehl, and H. R. Schmitt, "Challenges in publishing research data – a Fraunhofer Case Study: To be published.," *ing.grid*, 2023. [Online]. Available: https://preprints.inggrid.org/repository/view/20/ (visited on 02/27/2024).

[9] D. Ortloff, S. Anger, and M. Schellenberger, "An Empirical Study of the State of Research Data Management in the Semiconductor Manufacturing Industry: An analysis of industry and research institutions in the iDev4.0 project: To be published.," *ing.grid*, 2023. [Online]. Available: https://preprints.inggrid.org/repository/view/19/ (visited on 02/27/2024).

[10] *Your domain*, 2023. [Online]. Available: https://rdmkit.elixir-europe.org/your_domain (visited on 06/22/2023).

[11] M. Politze, *A reference architecture and implementation enabling data protection in distributed elearning and escience processes: Eine Referenzarchitektur und Implementierung unter Berücksichtigung des Datenschutzes in verteilten eLearning und eScience Prozessen* (Ergebnisse aus der Informatik). Aachen: Apprimus, 2019, vol. 10. DOI: 10.18154/RWTH-2019-11476.

[12] A. M. Cox and W. W. T. Tam, "A critical analysis of lifecycle models of the research process and research data management," *Aslib Journal of Information Management*, vol. 70, no. 2, pp. 142–157, 2018, ISSN: 2050-3806. DOI: 10.1108/AJIM-11-2017-0251.

[13] H. P. Bonn, "Die Vernetzung zwischen Hochschulforschung und Industrie: Facetten und Vorteile für den Wirtschaftsstandort," in *Zukunft Ingenieurwissenschaften – Zukunft Deutschland*, M. Nagl, H.-J. Bargstädt, M. Hoffmann, and N. Müller, Eds., Berlin and Heidelberg: Springer, 2009, pp. 111–120, ISBN: 978-3-540-89608-1. DOI: 10.1007/978-3-540-89609-8{\textunderscore}12.

[14] E. Berning and S. Falk, "Das Promotionswesen im Umbruch," *Beiträge zur Hochschulforschung*, vol. 27, no. 1, pp. 48–73, 2005. [Online]. Available: https://www.bzh.bayern.de/fileadmin/news_import/1-2005-berning-falk.pdf (visited on 02/27/2024).

[15] N. Winterhager, *Drittmittelwettbewerb im universitären Forschungssektor* (Organization & Public Management). Wiesbaden: Springer, 2015, ISBN: 9783658071882. DOI: 10.1007/978-3-658-07188-2.

[16] L. Behlau, *Forschungsmanagement: Ein praktischer Leitfaden*. Berlin and Boston: De Gruyter, 2017, ISBN: 9783110517828. DOI: 10.1515/9783110517828.

[17] T. Hamann and M. Moser, *Workflows ingenieurwissenschaftlicher Forschungsabläufe: Auszug des aktuellen Forschungsstands*, 2022. DOI: 10.5281/zenodo.7125571.

[18] D. Deb, R. Dey, and V. E. Balas, *Engineering Research Methodology: A Practical Insight for Researchers* (Intelligent Systems Reference Library). Singapore: Springer Nature Singapore, 2019, vol. 153, ISBN: 978-981-13-2946-3. DOI: 10.1007/978-981-13-2947-0.

[19] L. Griem, P. Zschumme, M. Laqua, *et al.*, "KadiStudio: FAIR Modelling of Scientific Research Processes," *Data Science Journal*, vol. 21, 2022. DOI: 10.5334/dsj-2022-016.

[20] F. Stratmann, "Prozessorientierung in Hochschulen – mehr als Tools und Referenzmodelle," in *Prozessorientierung in Hochschulen – mehr als Tools und Referenzmodelle*, ser. HIS: Forum Hochschule, P. Altvater, M. Hamschmidt, and F. Stratmann, Eds., 2011, pp. 1–24. [Online]. Available: https://his-he.de/fileadmin/user_upload/Publikationen/Forum_Hochschulentwicklung/fh-201112.pdf (visited on 06/12/2023).

[21] C. Tenopir, S. Allard, K. Douglass, *et al.*, "Data Sharing by Scientists: Practices and Perceptions," *PloS ONE*, vol. 6, no. 6, 2011. DOI: 10.1371/journal.pone.0021101.

[22] *Mission statement*, 2022. [Online]. Available: https://www.fiz-karlsruhe.de/sites/default/files/FIZ/Dokumente/Broschueren/fiz-leitbild-broschuere-en-20221115.pdf (visited on 06/13/2023).

[23] M. Patel, *I2S2 Idealised Scientific Research Activity Lifecycle Model*, 2011. [Online]. Available: https://purehost.bath.ac.uk/ws/portalfiles/portal/11246894/I2S2_ResearchActivityLifecycleModel_110407.pdf (visited on 06/02/2023).

[24] Deutsches Institut für Normung e. V., *Projektmanagement - Projektmanagementsysteme - Teil 5: Begriffe*, Berlin, 2009-01.

[25] T. Möller and F. Dörrenberg, *Projektmanagement* (WiSorium - Wirtschafts- und Sozialwissenschaftliches Repetitorium). München: Oldenbourg, 2003, ISBN: 978-3-486-27332-8.

[26] P. Bock and B. Scheibe, *Getting It Right: R&D Methods for Science and Engineering*. San Diego: Elsevier, 2001, ISBN: 9780121088521. DOI: 10.1016/B978-0-12-108852-1.X5000-5.

[27] H. Tang, *Engineering Research: Design, Methods, and Publication*. Hoboken: John Wiley & Sons, 2021, ISBN: 978-1-119-62453-0. DOI: 10.1002/9781119624547.

[28] T. Willaert, J. Cottyn, U. Kenens, T. Vandendriessche, D. Verbeke, and R. Wyns, “Research Data Management and the Evolutions of Scholarship: Policy, Infrastructure and Data Literacy at KU Leuven,” *LIBER Quarterly*, vol. 29, no. 1, pp. 1–19, 2019. DOI: 10.18352/lq.10272.

[29] *Leitlinien zum Umgang mit Forschungsdaten*, 2015. [Online]. Available: https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/leitlinien_forschungsdaten.pdf (visited on 06/02/2023).

[30] *Der Datenlebenszyklus: Stationen des Forschungsdatenmanagements*, 2023. [Online]. Available: https://forschungsdaten.info/themen/informieren-und-planen/datenlebenszyklus/ (visited on 06/02/2023).

[31] S. I. H. Shah, V. Peristeras, and I. Magnisalis, “DaLiF: a data lifecycle framework for data-driven governments,” *Journal of Big Data*, vol. 8, no. 1, 2021. DOI: 10.1186/s40537-021-00481-3.

[32] *Was ist Forschungsdatenmanagement? Aus kleinem Anfang entspringen alle Dinge*, 2023. [Online]. Available: https://forschungsdaten.info/themen/informieren-und-planen/was-ist-forschungsdatenmanagement/ (visited on 02/27/2024).

[33] S. Jones, G. Pryor, and A. Whyte, *How to Develop Research Data Management Services - a guide for HEIs*, Edinurgh, 2013. [Online]. Available: https://www.dcc.ac.uk/sites/default/files/documents/publications/How-to-develop-RDM-services_finalMay2013rev.pdf (visited on 06/22/2023).

[34] D. Schmitz and M. Politze, “Forschungsdaten managen – Bausteine für eine dezentrale, forschungsnahe Unterstützung,” *o-bib*, vol. 5, no. 3, pp. 76–91, 2018. DOI: 10.5282/O-BIB/2018H3S76-91.

[35] M. Putnings, “Datenmanagement,” in *Praxishandbuch Forschungsdatenmanagement*, M. Putnings, H. Neuroth, and J. Neumann, Eds., Berlin/Boston: De Gruyter Saur, 2021, pp. 297–302, ISBN: 9783110657807. DOI: 10.1515/9783110657807-017.

[36] M. Jagerhorn and X. Chen, “DataCite and FAIR Workflows,” *Procedia Computer Science*, vol. 211, pp. 83–92, 2022, ISSN: 18770509. DOI: 10.1016/j.procs.2022.10.179.

[37] D. P. Tripathi and S. R. Pandy, “Developing a Conceptual Framework of Research Data Management for Higher Educational Institutions,” in *IEEE 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services [ETTLIS 2018]*, S. Kataria, J. P. Anbu K., R. Gartner, and G. Sandhu, Eds., Piscataway, NJ: IEEE, 2018, pp. 105–110, ISBN: 978-1-5386-0828-9. DOI: 10.1109/ETTLIS.2018.8485193.

[38] *Research Data Management: Research Data Lifecycle*, 2024. [Online]. Available: https://guides.lib.virginia.edu/c.php?g=515290&p=3522215 (visited on 02/01/2024).

[39] A. Alturki, G. G. Gable, and W. Bandara, “A Design Science Research Roadmap,” in *Service-Oriented Perspectives in Design Science Research*, ser. Lecture Notes in Computer Science, H. Jain, A. P. Sinha, and P. Vitharana, Eds., Berlin and Heidelberg: Springer, 2011, pp. 107–123, ISBN: 978-3-642-20632-0. DOI: 10.1007/978-3-642-20633-7{\textunderscore}8.

[40] A. R. Hevner, S. T. March, J. Park, and S. Ram, “Design Science in Information Systems Research,” *MIS Quarterly*, vol. 28, no. 1, pp. 75–105, 2004, ISSN: 02767783. DOI: 10.2307/25148625.

[41] R. Baskerville, “What design science is not,” *European Journal of Information Systems*, vol. 17, no. 5, pp. 441–443, 2008, ISSN: 0960-085X. DOI: 10.1057/ejis.2008.45.

[42] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A Design Science Research Methodology for Information Systems Research,” *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, 2007, ISSN: 0742-1222. DOI: 10.2753/MIS0742-1222240302.

[43] T. Hamann, *Jarves: Der digitale FDM-Assistent*, Aachen, 2023. DOI: 10.5281/ZENODO.7642481.

[44] T. Hamann, *Jarves: The digital RDM assistant*, Gothenburg, 2023. DOI: 10.5281/ZENODO.7715683.

[45] T. Hamann and J. Werheid, *Jarves: The Digital Data Steward for Engineering Science Research*, Online, 2023. DOI: 10.5281/ZENODO.8378834.

[46] L. Schreyer, M. Cyra, and L. C. Bossert, *FDM-Werkstatt: Into the RDM Toolbox*, 2024. [Online]. Available: https://fdm-nrw.coscine.de/#/FDM-Werkstatt?id=program (visited on 05/06/2024).

[47] R. H. Schmitt, V. Anthofer, S. Auer, *et al.*, *NFDI4Ing - the National Research Data Infrastructure for Engineering Sciences: Excerpt from the Funding Proposal submitted in October, 2019, to the German Research Foundation (DFG)*, 2020. DOI: 10.5281/zenodo.4015201. (visited on 04/13/2023).

[48] A. Witzel and H. Reiter, *The problem-centred interview*. Los Angeles: SAGE, 2012, ISBN: 9781446291597.

[49] B.-A. Andrei, A.-C. Casu-Pop, S.-C. Gheorghe, and C.-A. Boiangiu, “A study on using waterfall and agile methods in software project management,” *Journal of Information Systems & Operations Management*, vol. 8, no. 2, pp. 125–135, 2019.

[50] B. Unhelkar, *The Art of Agile Practice: A Composite Approach for Projects and Organizations* (Advanced and Emerging Communications Technologies Series). Boca Raton: Taylor & Francis, 2013.

[51] U. Schwardmann, "Digital Objects – FAIR Digital Objects: Which Services Are Required?" *Data Science Journal*, vol. 19, 2020. DOI: 10.5334/dsj-2020-015.

[52] *Principles and Strategies for the Reform of Scholarly Communication 1*, 2003. [Online]. Available: https://www.ala.org/acrl/publications/whitepapers/principlesstrategies (visited on 02/26/2024).

[53] C. Ascheron, *Wissenschaftliches Publizieren und Präsentieren: Ein Praxisleitfaden mit Hinweisen zur Promotion und Karriereplanung*. Berlin: Springer, 2019. DOI: 10.1007/978-3-662-58053-0.

[54] L. Schilhan and C. Kaier, Eds., *Publikationsberatung an Universitäten: Ein Praxisleitfaden zum Aufbau publikationsunterstützender Services*. Bielefeld: transcript, 2020, ISBN: 9783839450727. DOI: 10.14361/9783839450727.

[55] F. Gail and M. Vetter, "Systematische Zielgruppenbefragung - Methode und Ergebnisse von Fokusgruppen-Interviews durch ZB MED," *Informationspraxis*, vol. 2, no. 2, 2016. DOI: 10.11588/ip.2016.2.30984.

[56] R. A. Krueger and M. A. Casey, *Focus groups: A practical guide for applied research*, 5th edition. Los Angeles: SAGE, 2014, ISBN: 9781483365244.

[57] L. C. Murray, "Book Review: The problem-centred interview," *Journal of Mixed Methods Research*, vol. 10, no. 1, pp. 112–113, 2016, ISSN: 1558-6898. DOI: 10.1177/1558689815577032.

[58] U. Flick, *An Introduction to Qualitative Research*, 4th ed. London: Sage Publications, 2009, ISBN: 9781446205624.

[59] E. Scheibelhofer, "A Reflection Upon Interpretive Research Techniques: The Problem-Centred Interview as a Method for Biographic Research," in *Narrative, Memory & Everyday Life*, N. Kelly, C. Horrocks, K. Milnes, B. Roberts, and D. Robinson, Eds., Huddersfield: University of Huddersfield, 2005, pp. 19–32, ISBN: 9781862180758. [Online]. Available: http://eprints.hud.ac.uk/id/eprint/4923/.

[60] S. Döringer, "'The problem-centred expert interview'. Combining qualitative interviewing approaches for investigating implicit expert knowledge," *International Journal of Social Research Methodology*, vol. 24, no. 3, pp. 265–278, 2021, ISSN: 1364-5579. DOI: 10.1080/13645579.2020.1766777.

[61] P. Mayring, "Qualitative Content Analysis," *Forum Qualitative Sozialforschung*, vol. 1, no. 2, 2000. DOI: 10.17169/fqs-1.2.1089.