


# A survey on the dissemination and usage of research data management and related tools in German mechanical and industrial engineering sciences

Tobias Hamann <sup>1</sup>  
Amelie Metzmacher <sup>1</sup>  
Patrick Mund <sup>1</sup>  
Marcos Alexandre Galdino <sup>1</sup>  
Anas Abdelrazeq <sup>1</sup>  
Robert Schmitt <sup>1, 2</sup>

1. Laboratory for Machine Tools and Production Engineering (WZL) | Chair of Production Metrology and Quality Management & Institute for Information Management in Mechanical Engineering (Prof. Robert Schmitt), RWTH Aachen University, Aachen.  
2. Fraunhofer Institute for Production Technology (IPT), Aachen.



**Date Submitted:**  
2023-01-01

**Licenses:**  
This article is licensed under: 

**Keywords:**  
Research data management, RDM, Survey, Dissemination, Usage of Research Data Management

**Data availability:**  
Data can be found here:  
<https://doi.org/10.5281/zenodo.7645548>

**Software availability:**  
Software can be found here:  
<https://doi.org/10.5281/zenodo.7645548>

**Corresponding Author:**  
Tobias Hamann  
[t.hamann@wzl-mq.rwth-aachen.de](mailto:t.hamann@wzl-mq.rwth-aachen.de)

**Abstract.** As the amount of collected and analysed data increases, a need for data management arises to ensure its usability. This also applies in research. This challenge can be addressed by Research Data Management (RDM), which focuses on the reusability of data. To answer the research question “What is the current status of RDM the field of mechanical and industrial engineering in German Engineering Sciences?”, an exploratory survey was conducted over the last quarter of 2020. This survey depicts the status quo of RDM in engineering sciences in Germany, as well as challenges and improvement chances. 168 (n=168) researchers from the engineering sciences in Germany provided their view via a questionnaire that contains 39 questions. The results give insights on the interviewees’ knowledge and perceived relevance of RDM in their daily research activities. Key findings are the need of researchers in mechanical and industrial engineering, especially guidance and support regarding RDM. Additionally, some main reasons against RDM (missing knowledge and additional effort) could be identified. The survey closed with questions regarding reasons that could prevent researchers from adapting sustainable RDM. This paper provides a qualitative overview over the current RDM in engineering and pointing out possible strategies to foster it, namely the integration of guidance and education for RDM. Along the paper we publish the collected data set to enable further analysis and reuse.

## 1 Introduction

As the amount of data has been growing for years [1]–[3], the effort required to manage this data increases. Adding to the sheer amount of data, the requirements of data processing and data reuse further raise the effort in data management. Especially in the context of engineering

5 and industry 4.0 data has to be managed to facilitate the application of related methods as, for  
6 example, machine learning [4], [5]. This is not only relevant for industrial applications but also  
7 related research performed in engineering sciences. The interest in data collected or generated in  
8 the context such research projects is raising as well [6]. Data can be reused to enhance the own  
9 research or validate existing results. Therefore, research data management (RDM) is becoming  
10 more and more important in many research areas, including engineering. As a result, research  
11 data management is introduced to engineering researchers. This applies not only for engineering  
12 sciences in general but also mechanical and industrial engineering in particular. To facilitate the  
13 process and cultural change in engineering sciences, the current status of RDM first has to be  
14 recorded before requirements are scouted and solutions are developed. To start this process for  
15 mechanical and industrial engineering, the following research question has to be answered:

16 ***What is the current status of RDM the field of mechanical and industrial engineering in  
German Engineering Sciences?***

17 As soon as this question is answered, it will become clearer, in which contexts RDM is already  
18 applied successfully in the field of mechanical and industrial engineering in German Engineering  
19 Sciences and in which areas more support is needed. After that, conclusions can be drawn,  
20 deriving reasons against the application of RDM and possibilities how RDM can be improved to  
21 fit the needs and demands of researchers better.

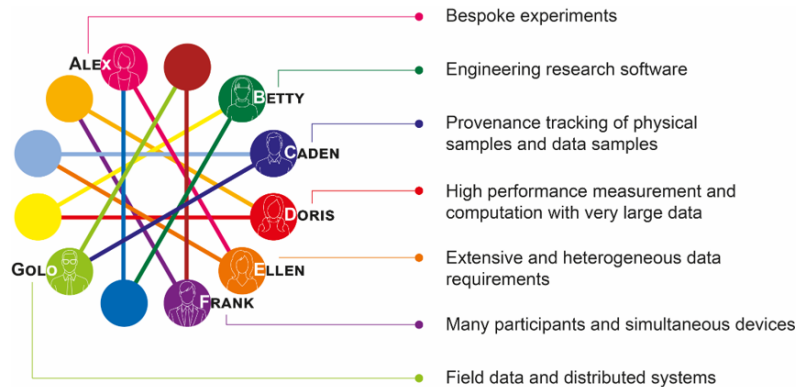
22 Therefore, an explorative qualitative survey has been deployed, which asked researchers about  
23 the use of RDM in the context of their activities. The survey could sketch out the status of RDM  
24 in engineering. Key findings are the knowledge and usage of RDM tools and support structures  
25 as well as possible reasons for researchers to not integrate or apply RDM in their research.

26 To establish a framework delineating the terms of RDM, it is imperative to commence with a pre-  
27 cise definition of RDM. "Research data management encompasses the processes of transforming,  
28 selecting and storing research data with the common goal of keeping it accessible, reusable and  
29 verifiable in the long term and independent of individuals" [7] while research data is "(digital)  
30 data generated during scientific activity (e.g. through measurements, surveys, source work)" [8].

31 Furthermore, the context of this survey shall be clarified. Within the framework of the NFDI4Ing  
32 consortium, the use and management of research data is to be disseminated and improved. In  
33 order to achieve the required improvement, so-called Archetypes and community clusters were  
34 used to categorise the research landscape in engineering. These Archetypes cover common fields  
35 of research methodologies (e.g. working with experimental or field data, using code or working  
36 with material samples). A researcher can relate to more than one Archetype in a fluent way. The  
37 community clusters separate the researchers thematically into the five DFG classifications of the  
38 engineering sciences that were valid when NFDI4Ing was founded [9].

39 This survey was prepared and conducted within the NFDI4Ing's Archetype Frank. Frank's  
40 methodology revolves around the concept of many participants (either as researchers or observed  
41 individuals), both human and artificial [9]. The key objective of Archetype Frank is the facil-  
42 itation of RDM in environments with these participants, dealing with the variety of involved  
43 engineering disciplines, considering heterogeneous data sources and their synchronisation as  
44 well as taking into account the collaborative aspects of working with many participants. Potential  
45 users have a background that "is mostly informed by production engineering, industrial engi-

46 neering, ergonomics, business engineering, product design and mechanical design, automation  
 47 engineering, process engineering, civil engineering and transportation science.” [9]. To better  
 48 compare Archetype Frank to other archetypes, all of them and their core characteristics are  
 49 depicted in figure 1.



**Figure 1:** The NFDI4Ing's Archetypes (c.f.[9])

50 To facilitate the application of RDM, the needs of researchers should be met. To identify  
 51 such needs, it is necessary to conduct interviews and surveys among a broad cross-section of  
 52 researchers, who identify with Archetype Frank or work in similar environments [9]. In addition,  
 53 Archetype Frank has a strong overlap with production engineering and mechanical engineering  
 54 as stated above, which leads to a partial representation of the NFDI4Ing's CC41 "Mechanical  
 55 and industrial engineering (CC41)" [9] as well. Therefore, the survey is specifically focused on  
 56 mechanical and industrial engineering.

57 While there are some publications on the status quo of RDM in general, there is not yet a survey  
 58 on RDM in engineering sciences with a broad approach in Germany. Therefore, this survey aims  
 59 to penetrate the circle of potential RDM users in engineering, specifically Archetype Frank in an  
 60 explorative manner. The survey is intended to give Archetype Frank an overview of the status  
 61 quo and to enable it to ask more specific questions, for example in interviews or further surveys.

## 62 2 Related work

63 To screen the papers addressing similar questions on the status quo of RDM, a literature review  
 64 has been performed. This literature review aims to get an overview over similar approaches in  
 65 the context of RDM. While the focus is set on engineering, other disciplines are also considered  
 66 whenever they offer an adequate perspective on the topic of this paper.

### 67 2.1 Procedure of the literature review

68 The literature review was performed on the platforms ScienceDirect, Web of Science and IEEE  
 69 Xplore. The review was last updated in November 2023. Only results newer than the original  
 70 FAIR Principles [10] were considered relevant, causing results to not date back further than  
 71 2016. To perform the review in context of the research question, a search string was compiled  
 72 based on the terms *research data management* and *engineering* and *survey* or synonymous terms,

73 namely *analysis, audit, check* or *inquiry*. The resulting search string is explicitly formulated not  
 74 as strict towards mechanical and industrial engineering as the research question to find results  
 75 from neighbouring fields. The search string was used in three search engines listed in table 1.  
 76 Afterwards, the results of the search engines were filtered for their Year (see table 1). Lastly, the  
 77 resulting papers were exported in the .ris format along with their abstracts.

Search Engine	Last Searched	Filters Used	Results
ScienceDirect	08.11.2023	Year: 2016 or newer	269
Web of Science	08.11.2023	Year: 2016 or newer	62
IEEE Xplore	08.11.2023	Year: 2016 or newer	8
		Sum:	339
		Included duplicates:	121
		Unique records found:	218

**Table 1:** Used search engines, filters and results for the literature review

78 The .ris files were imported to the PICO Portal to screen the collected papers for their relevance  
 79 based on their abstracts. For this screening, exclusion criteria were formulated. These are listed  
 80 in table 2. Any papers matching the exclusion criteria (n=194) as well as any duplicates (n=121)  
 81 were removed from the review process. It has to be mentioned that the full text of Todorova et al.  
 82 about "Comparative Findings from Data Literacy Survey in Three Bulgarian Universities" [11]  
 83 was not accessible at the time this paper was written and is therefore not included. A complete  
 84 flow diagram of the process is depicted in 13 in the Appendix in subsection 8.1.1.

Criteria Number	Exclusion Criteria
1.	Not related to research data management
2.	Not a survey or interview or similar data collection
3.	Not related to engineering sciences
4.	Not containing information on the current status of RDM usage/application

**Table 2:** Exclusion criteria for the literature review

85 The resulting 23 papers were screened a second time, but based on their full texts. Again the  
 86 exclusion criteria from figure 2 were used, resulting in two excluded records for the first criteria,  
 87 four for the second, seven for the third and three for the fourth criteria. One of the found  
 88 papers ([12]) summarised another paper ([13]) which was directly cited instead on relying on the  
 89 summary. Lastly, six papers have been chosen by the full text review.

90 In addition to the systematic literature review, other sources of literature have been considered  
 91 as well. The journals ing.grid and BausteineFDM have also been consulted to identify papers  
 92 that are relevant but are not listed on the aforementioned platforms. Also, Zenodo as an catch-all  
 93 repository has been consulted. BausteineFDM contained one more paper relevant in this context  
 94 while in ing.grid's preprint server, two additional papers could be found. Zenodo included three  
 95 additional relevant publications. These six papers are also included in this review.

96 **2.2 Results of the literature review**

97 In table 3, the results of the literature review are shown, sorted by their most common statements  
 98 on the status quo of RDM. No literature found contains direct information on the status quo of  
 99 RDM in mechanical and industrial engineering. Yet, insights into current RDM practices and  
 100 issues are granted, not only for engineering but also neighbouring fields.

Record	Year	Methods	Subject Area	Region	Findings	Need for awareness	Need for training	Need for instructions	Need for resources	Need for RDM tools	Need for support	Need for incentives
[14]	2017	Survey	Information Systems	Germany		●	◐	●	●	○	◐	◐
[15]	2019	Survey	Science and technology	Europe		○	●	●	○	●	●	◐
[16]	2022	Survey	Generic	Canada		●	◐	◐	●	○	◐	◐
[17]	2021	Case study	Engineering	Europe		●	○	○	●	◐	◐	◐
[18]	2017	Survey	Neuroimaging	USA		●	●	●	●	○	○	○
[13]	2021	Survey	Engineering	Egypt, Jordan, Saudi Arabia		●	●	◐	○	●	○	○
[19]	2023	Survey	Physics	Germany		●	◐	◐	○	●	○	○
[20]	2023	Case study	Engineering	Germany		●	○	●	○	○	●	○
[21]	2017	Survey	Social science, STEM <sup>1</sup> , Humanities	UK, France, Turkey, Iceland		●	●	○	○	○	◐	○
[22]	2023	Expert interview	Engineering	Germany		~	●	●	○	○	○	○
[23]	2019	Literature review	Generic	Global		○	●	○	○	○	○	○
[24]	2021	Survey	Engineering	Canada		◐	◐	○	○	○	○	○
[25]	2017	Survey	Generic	Slovenia		~	○	○	○	○	○	○

~: Contradiction      ○: Not mentioned      ◐: Somewhat supported  
 ◑: Mildly supported      ◑: Strongly supported      ●: Fully supported

**Table 3:** Results of the literature review

101 For instance, the most prominent topic in literature are the need for awareness amongst researchers.  
 102 This seems to be a global and cross-disciplinary problem, as it is mentioned in almost every

1. science, technology, engineering and mathematics

103 record found by the literature review. Still, two records state, RDM awareness is not a problem.  
104 One of them is based on "spotlight investigation" [22] based on expert interviews, which might  
105 cause a bias on the results. The other is a RDM survey from Slovenia with no specific focus on a  
106 research area. [25] Similarly, the need for training and instructions are often mentioned.

107 While the need for resources is less often mentioned than the aforementioned aspects, the records  
108 that empathised on this aspect point out the importance of the effort connected to the application  
109 of RDM. This also, to some extent, is mentioned by the papers referring to the need for (specific)  
110 RDM tools, as these facilitate the application of RDM. However, this seemingly stands into  
111 contrast to the fewer mentions of need for support, which indicates that the effort of RDM can  
112 not be outsourced but has to be applied in the very context of a specific project. This is also  
113 supported by the many mentions of need for training.

114 One last interesting investigation has to be made. While the need for awareness is the most  
115 mentioned aspect, the need for incentives is the least mentioned one. This leads to the conclusion,  
116 that the intrinsic motivation for RDM is more important than external factors enforcing it.  
117 Researchers should be aware that and why their RDM is important not only to themselves but to  
118 others as well.

119 All publications presented either include RDM (in engineering) in a broader (e.g. nation wide)  
120 survey like [25] and [21] or refer to certain use cases or projects like [17] or [20]. The focus on  
121 RDM in Germany can only be found in related fields like IT sciences [14] or physics [19] or are  
122 not part of a survey but a case study [20] or a "spotlight investigation" [22]. While the presented  
123 literature does not fully match the scope, it still offers insights on related fields of the research  
124 question. All relevant findings are discussed and compared to the results of this article in section  
125 5.

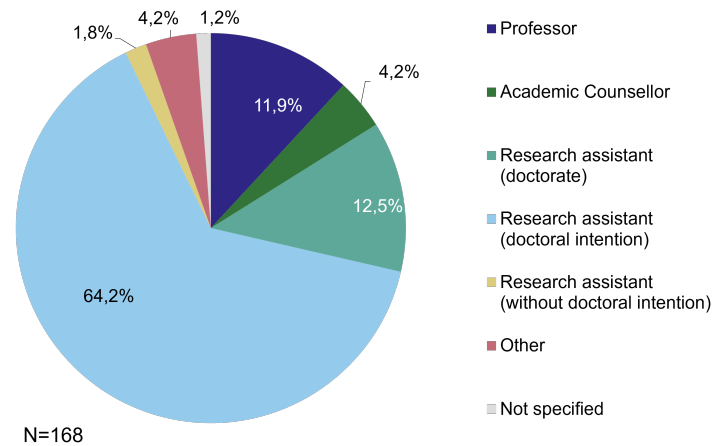
### 126 **3 Methodology**

127 This chapter introduces the methodology of the conducted survey. Firstly, the interviewees and  
128 the approach are discussed, followed by the surveys structure and the categories of questions  
129 contained. As a result both the interviewees and the questions are clarified before the results are  
130 discussed in chapter 4. The survey was implemented within the online tool [soscisurvey.de](https://www.soscisurvey.de).  
131 The results have been collected within soscisurvey and were then exported to .csv files for further  
132 analysis in python. The code used was documented in a Jupyter Notebook and uploaded together  
133 with the .csv files. The code written in python generated images of which the most important  
134 ones were chosen and recreated in PowerPoint to give them an appropriate finish. The keys  
135 given in a figured caption refer to the keys used on the supplemental files.

#### 136 **3.1 Interviewees and Approach**

137 The survey took place from October to December 2020. 168 researchers were interviewed,  
138 most of which are employed as research assistant seeking a doctoral degree (64%) (see figure  
139 2). The distribution of participants in the survey is slightly (ca. 4%) shifted towards more  
140 research assistants and less professors in comparison to the average in the field of German  
141 engineering sciences under consideration. [26] Based on the most recent data available for the

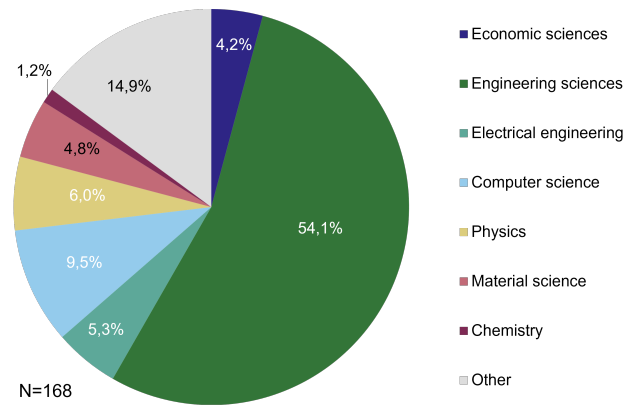
142 distribution of scientific staffing in Germany from 2021, this means about 0.3% of German  
 143 engineering researchers in total (total population: 56,332) and about 0.8% of the specialist areas  
 144 “General engineering sciences, industrial engineering with an engineering focus and mechanical  
 145 engineering/process engineering” [26] (total population: 20,355) were reached with the survey.



**Figure 2:** Distribution of answers on the question: In which position are you employed? (Key:AM03)

146 As the research question is focussed on mechanical and industrial engineering in German En-  
 147 gineering Sciences, the target group of survey participants was chosen accordingly. Hence,  
 148 the surveyed researchers are composed of members of the “Scientific Society for Production  
 149 Engineering” (“Wissenschaftliche Gesellschaft für Produktionstechnik”, in short WGP), the  
 150 “Scientific Society for Product Development” (“Wissenschaftliche Gesellschaft für Produk-  
 151 tentwicklung”, in short WiGeP) and researchers from the RWTH Aachen Cluster of Excellence  
 152 “Internet of Production” (IoP) as well as members of the “Fraunhofer-Verbund Produktion”.  
 153 These consortia stand for “Cutting-edge research [...] in the area of basic research as well as  
 154 applied and industrial research” [27] with a “close collaboration with economy and science” [28]  
 155 as well as a strong focus on “application-oriented research” [29]. The IoP states a “balanced  
 156 composition of participating researchers from five faculties at RWTH Aachen University and six  
 157 non-university research institutions” on their website [30]. E-mails were distributed to leading  
 158 entities of these scientific groups who agreed to further disseminate the survey amongst their  
 159 employees, asking for participation in the survey. As a result of this dissemination method, the  
 160 exact number of researchers who received the survey is unclear, meaning the response rate can  
 161 not be calculated but estimated. The WGP has about 2,000 members, the WiGeP has circa 1,200  
 162 while the IoP unites about 600 researchers and the Fraunhofer-Verbund Produktion consists of  
 163 about 3,000 employees. [27], [28], [30], [31] With this estimation about 6800 researchers were  
 164 contacted, resulting in an estimated response rate of 2.5%. About 43% of the respondents work  
 165 at the RWTH, while circa 39% originated from universities all over Germany. Lastly, 18% are  
 166 employees of Fraunhofer institutes.

167 All of the listed organisations are focused on engineering, particularly in mechanical engineering  
 168 and production technology. However mechanical engineering often involves interdisciplinary  
 169 approaches. Thus, plenty of subject areas are represented within the interviewees. As a result,  
 170 the survey represents not only Archetype Frank but also gives insights into Community Cluster  
 171 41 (CC41). Figure 3 depicts the subject areas of the interviewees. More than half of the surveyed



**Figure 3:** Distribution of answers on the question: What is your field of research? (Key: AM05)

172 researchers are from the subject area of mechanical engineering. The other half is a wide mix  
 173 of different subject areas. While some more are in the scope of mechanical engineering and  
 174 production technology than others, all of them are researching within the context of production  
 175 technology.

### 176 3.2 Survey Structure and Questions

177 The survey consists of 39 questions with 216 question items on 14 pages. Only fully filled  
 178 questionnaires were considered within the evaluation of the survey. Participants spend in average  
 179 14 minutes to fill the survey. The survey started with a demographic inquiry of the respondents'  
 180 data to validate the fit of the respondents. This is followed by an exploratory self-assessment,  
 181 which contains three introductory questions to the overall usage and knowledge of RDM.

Category	Number of ques- tions	Number of question items
Demographic data	7	7
Explorative questions	12	15
General RDM questions (FAIR , DMP, DLC)	3	3
Data life cycle	6	27
Tools	1	116
File formats	1	39
Specifications and support structures	8	8
Acceptance aspects (free text)	1	1
Sum	39	216

**Table 4:** Summary of the topics and their corresponding number of question items within the survey

182 Interviewees were questioned if they are aware of the FAIR principles [10] for research data, if  
 183 they (or a third party, if applicable) create a data management plan and if they base their research  
 184 on the data life cycle. The self-assessment is followed by detailed questions of how research  
 185 projects carried out along the data life cycle as proposed by forschungsdaten.info [32]. One  
 186 question for RDM-tools and one for file formats used hold the majority of question items, as the  
 187 usage of many tools and formats were queried. The questionnaire is rounded off by the question  
 188 about the RDM-specifications and -support available to the respondents. The opportunity to add



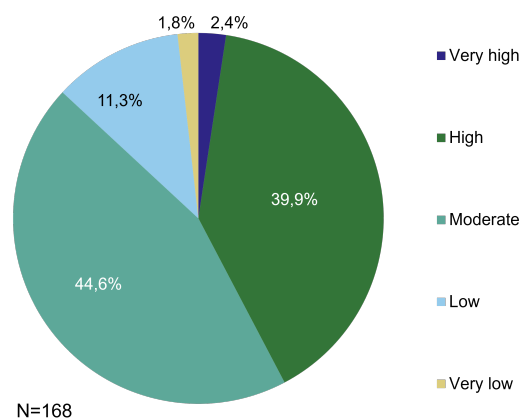
189 further comments via free text is given to the respondents throughout the survey, which only  
190 counts as a separate question when it is not an opportunity for further explanation or extension of  
191 answers to an existing question and is otherwise included in the number of question items rather  
192 than the questions number. The structure of the questionnaire with question categories of the  
193 survey and the corresponding numbers of questions and question items contained can be found  
194 in table 4. Free text answers are included within the numbers of questions stated in the table.

## 195 4 Results

196 After validating the fit of the respondents background in terms of discipline and employment,  
197 the actual evaluation of the survey results follows. This chapter is based on the structure of the  
198 survey mentioned in chapter 3.2 and is subdivided accordingly. As the sample size of the survey  
199 is quite small in comparison to the base population, the findings of this article are formulated  
200 as hypotheses rather than facts. Hence, these hypotheses can be compared to similar works  
201 as presented in 2 and be referenced by future works to validate or debunk them. Additionally,  
202 while only completely filled surveys were evaluated, respondents were able to refuse a definitive  
203 answer with a “Not specified” answer.

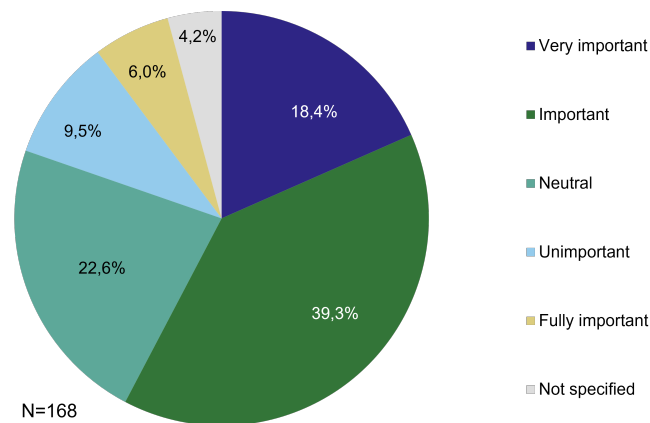
### 204 4.1 RDM Knowledge and Perceived Relevance of RDM

205 The first set of non-demographic questions aims at providing a rough assessment of the respon-  
206 dents knowledge on RDM in general. Regarding research data handling, more than half of the  
207 respondents stated that their knowledge was moderate or lower. Only 42% stated that they had a  
208 high or very high level of knowledge regarding the handling of research data (see figure 4). At  
209 the same time, over 57% of respondents rate RDM as important or very important. Only about  
210 15% perceive RDM as unimportant or completely unimportant (see figure 5).



**Figure 4:** Distribution of answers on the question: How high do you rate your own knowledge of handling research data? (Key: EF01)

211 When comparing those two statements above, there seems to be a gap between the group of  
212 researchers with (very) high RDM knowledge and a (very) high perceived importance of RDM.  
213 There are 14% less researchers who have a RDM-knowledge specified as high or above than  
214 there are researchers who perceive RDM as at least important. This leads to the first hypothesis

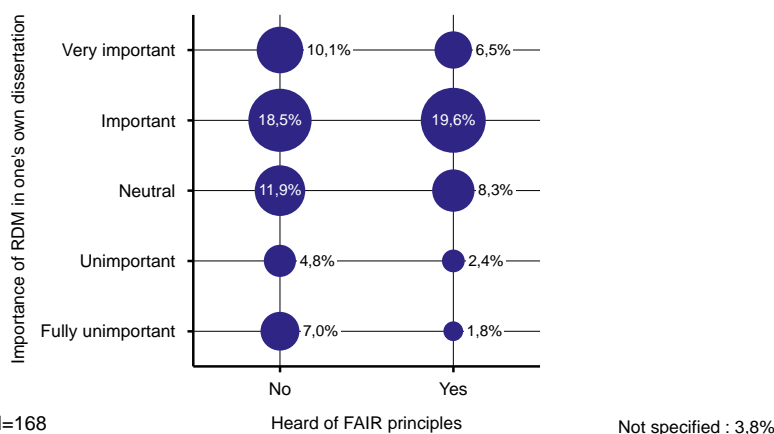


**Figure 5:** Distribution of answers on the question: How important is/was research data management to you in your personal dissertation project? (Key:EF02)

215 of this paper, that there is a gap in knowledge of researchers. Additionally, missing knowledge  
 216 may also lead researchers into perceiving RDM less important, potentially widening the gap.

217 1. *There is a need for RDM knowledge among researchers in the engineering sciences,*  
 218 *specifically for researchers of the Archetype Frank respectively amongst researchers in*  
 219 *the field of mechanical engineering and production technology (CC41).*

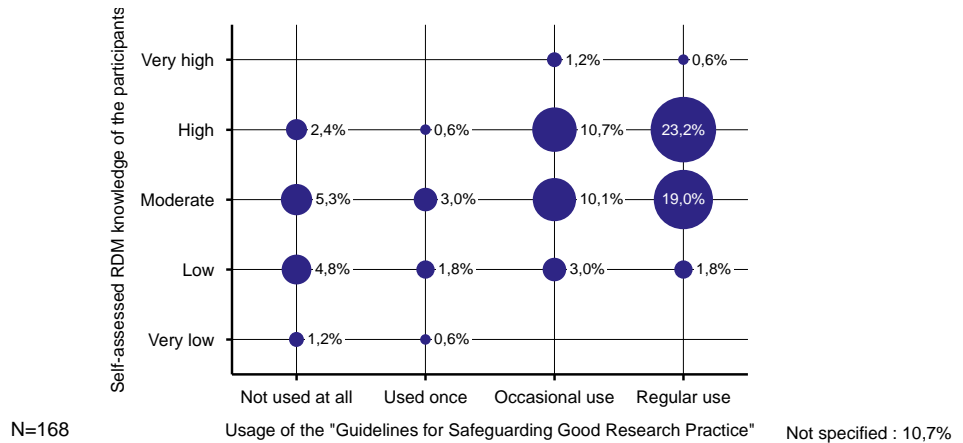
220 To better understand the relevance and reliability of the self-accessed RDM knowledge, the  
 221 following question was asked: "Have you ever heard of the FAIR principles (Findable, Accessible,  
 222 Interoperable, Reusable) [10] for research data?". The responses are mapped on the answers  
 223 from figure 5 and shown below in figure 6. It becomes apparent, that the relevance of RDM in  
 224 ones own dissertation and knowledge about the FAIR principles are somewhat correlated, yet it  
 225 is unclear which is caused by what.



**Figure 6:** Importance of RDM in one's own dissertation in dependency of the share of respondents who have heard about the FAIR principles (Findable, Accessible, Interoperable, Reusable) [10] for research data (Keys: EF02 over DL01\_01)

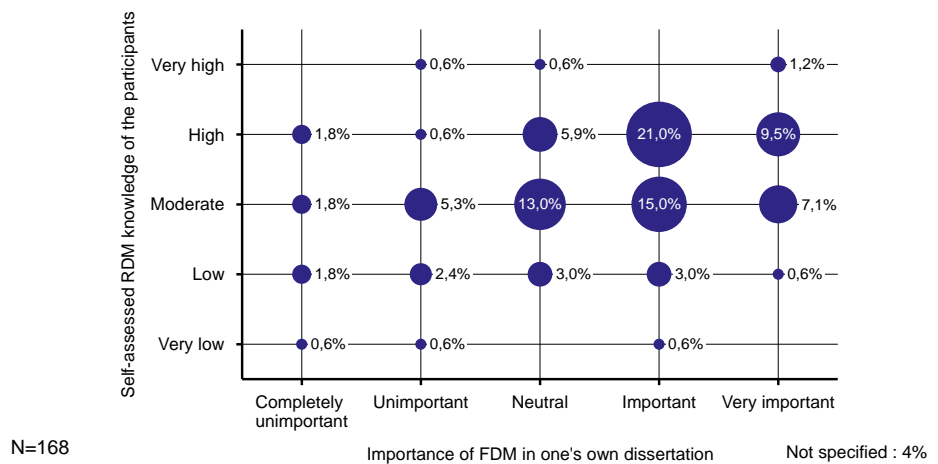
226 The survey also asked for the usage of the Code of Conduct of the "Guidelines for Safeguarding  
 227 Good Research Practice" published by the DFG [33]. These have already been applied several

228 times by almost three quarters of all respondents (see figure 7), however this does not lead to  
 229 a consistently high level of knowledge regarding research data management. The correlation  
 230 coefficient between these factors is 29%, which does indicate a mild correlation. Generally  
 231 speaking, the correlation coefficient measures how close two values are linearly dependant [34].  
 232 As the correlation coefficient is positive, this indicates an increase in RDM-related knowledge  
 233 when a person regularly uses the DFG guidelines. This effect can also be seen in figure 7.



**Figure 7:** Perceived relevance of RDM among the participants in dependency of the usage of the Code of Conduct of the "Guidelines for Safeguarding Good Research Practice" by DFG (Keys: EF01 over EF03)

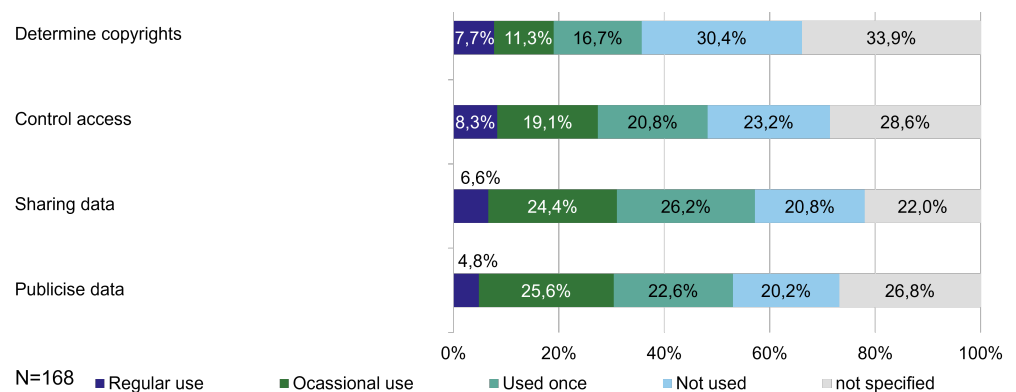
234 A similar effect, can be seen between the perceived relevance of RDM in the interviewees own  
 235 dissertations and the knowledge about RDM (see figure 8). Here, the correlation coefficient  
 236 amounts to 33%, indicating a mild positive correlation, meaning that the more important RDM  
 237 is perceived in context of the one's own dissertation, the more one knows about RDM [34].



**Figure 8:** Perceived relevance of RDM among the participants in dependency of the perceived relevance of RDM in the researchers own dissertation (Keys: EF01 over EF02)

## 238 4.2 Application of RDM Related Tasks

239 While 58% (see figure 8) claim to find RDM important in their own dissertation, the self-assessed  
 240 knowledge amongst the interviewees is mostly moderate to very low. Moreover, the claim of  
 241 regular use of the "Guidelines for Safeguarding Good Research Practice" is questioned by the  
 242 answers of the interviewees in the later questions of the survey. For example: The Guidelines  
 243 state that "Researchers decide autonomously [...] whether, how and where to disseminate their  
 244 results." This includes the process of determining copyrights and the control of access, which  
 245 is especially important when handling data that is not shared due to reasons such as secrecy  
 246 or of patent applications. In that case, a decision has to be made to control the access to only  
 247 those who are allowed to access such data. However, less than 10% of the interviewees regularly  
 248 determine copyrights, control access or share their data (see figure 9).



**Figure 9:** Activities from the sharing phase - Distribution of answers on the question: Please indicate whether and to what extent you use the individual steps of the data life cycle. (Keys: DL02\_15 to DL02\_18)

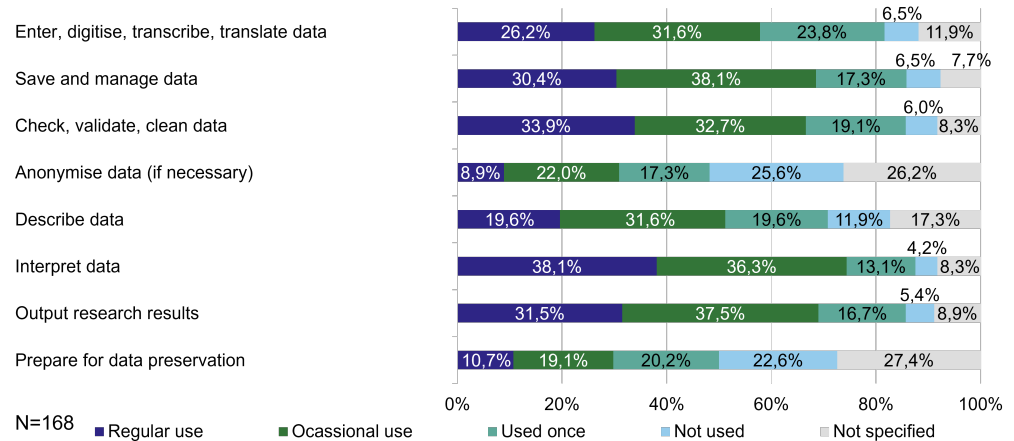
249 Even less make their data publicly available (<5%). To set this into perspective, 44% of the  
 250 surveyed researchers claimed to regularly use the DFG's "Guidelines for Safeguarding Good  
 251 Research Practice" [33]. In other words, only about one in nine researchers who regularly use  
 252 this guideline "make all results available as part of scientific/academic discourse", although  
 253 research data should be included "where possible and reasonable" [33] as proposed by the DFG.

254 Similar low rates of regular application of research data management tasks can be observed  
 255 throughout various steps of the data life cycle. This leads the following hypothesis:

256 2. *While the use of Guidelines like the "Guidelines for Safeguarding Good Research Practice"*  
 257 *tend to improve the self assessed RDM knowledge among the interviewees (see figure 7),*  
 258 *it does not necessarily imply the application of RDM connected tasks.*

259 The only step of the data life cycle that has a high rate of regularly performed tasks is the "prepare  
 260 and analyse data" phase, as shown in figure 10. The highest rated task is "Interpret data", which  
 261 scores a 38% regular application rate. An additional 36% occasional application rate is adding up  
 262 to 74% of the researchers who at least occasionally interpret their data on their own. Taking into  
 263 consideration that 16% of the interviewees are professors or academic councillors, this initially  
 264 rather low rate of data interpretation among researchers becomes clearer.

265 This leads to the next hypothesis put forward in this article:



**Figure 10:** Activities from the prepare and analyse data phase - Distribution of answers on the question: Please indicate whether and to what extent you use the individual steps of the data life cycle. (Keys: DL02\_07 to DL02\_14)

266 3. RDM-related tasks that are not directly part of the everyday research activity (like deter-  
 267 mining copyrights) are much less likely to be carried out than those who are mandatory to  
 268 receive results from data, such as transcribing, preparing, interpreting or validating data.

#### 269 4.3 Data Sharing with Third Parties

270 Another set of questions asked about the willingness to share research data with third parties and  
 271 the reuse of third party research data. This set of questions however seems to be inappropriately  
 272 specified, as the results are inconsistent. One participant gave feedback on this topic:

273 *"The questions [regarding sharing research data with third parties] are flawed, as the attitude*  
 274 *towards any third party is different than within the institute or a network."*

275 Anticipating focus group interviews that took place months after the survey with different  
 276 participants, it can be said that this definition of "third parties" harshly varies in the understanding  
 277 of researchers. The questions in this survey aimed towards the interpretation of third parties as  
 278 "not related to the research project in any way". This however seems to be misinterpreted by  
 279 some of the participants. The questions that asked for the data life cycle, specifically the ones for  
 280 the sharing data phase, show that 57% shared data at least once, which was shown in figure 9.

281 When asked for the actual possibility for third parties to access one's own research data, this  
 282 value raises to 65%. This can be explained in two ways:

283 1. The additional 8% of interviewees did not specify an answer in the corresponding question  
 284 set at the data life cycle section of the survey.

285 2. The surveyed researchers interpreted the expression "third party" as "involved in the actual  
 286 research project, but not part of the own institute".

287 It is unclear which of the two applies in this case. It has to be noted that, although the expression  
 288 "third parties" is used in the "Guidelines for Safeguarding Good Research Practice", it is never  
 289 specified in the document itself [35].

#### 290 4.4 Usage of RDM Tools and Services

291 The next part focused on tools and services. A distinction is made between usage and awareness  
 292 of tools. The term usage refers to the following options: "regular use" and "occasional use".  
 293 Awareness means the tool is either "known by name" or has at least a "one-time use". Respectively,  
 294 unawareness refers to the option "unknown". A "not specified" option was given as well.

295 More than 70% of all responses are "unknown". A further 19% are assigned to the answer option  
 296 "not specified". It has to be noted that this distribution also applies if only the answers of those  
 297 are taken into consideration, who have stated to have a high or very high self accessed RDM  
 298 knowledge. In this case, 69% answered "unknown" and 20% answered "not specified" or did  
 299 not answer the question at all. In general, the answers of the respondents are strongly polarised.  
 300 A few tools stand out due to regular use, while others are almost completely unknown.

301 Literally the most prominent example is Git, with 72% awareness among respondents. Almost  
 302 30% use the tool regularly and 25% occasionally. 7% have used Git at least once and 10%  
 303 are familiar with it by name. No other tool has a similar level of awareness and use among  
 304 researchers. Although mySQL is better known than Git (78%), it is used much less frequently  
 305 (regularly 12% and occasionally 22%) and is limited to one-time use (28%).

306 An overview of awareness ("known by name" and all mentions of useage) and usage (sum of  
 307 the mentions of "occasional" and "regular use") is given in table 5, sorted by the proportion of  
 308 respondents who state multiple uses. Due to the large number of tools surveyed, only those used  
 309 more than once by at least 5% of the respondents are mentioned below for the sake of clarity.

Tool/Service	Category	Awareness [%]	Usage [%] ▽
Git	Data organisation	72	55
mySQL	Databases and repositories	78	34
DOI Citation Formatter	Citation	45	30
KeePass	Password help	44	26
TIB PID Competence Centre	Persistent identifiers	35	22
Microsoft Project	Collaborative work	64	20
NoSQL	Databases and repositories	42	14
TortoiseSVN	Data organisation	34	14
TortoiseGit	Data organisation	32	11
PostgreSQL	Databases and repositories	29	8
Google Dataset Search	Find research data	32	8.3
STD-DOI	Citation	17	8.3
Apache Subversion	Data organisation	23	7.7

**Table 5:** Awareness and use of tools among researchers sorted by use among respondents

310 As shown in table 5, of the 90 tools and services surveyed, only 13 have been used more than once  
 311 by at least 5% of the respondents. Seven of those 13 come from the field of software development,  
 312 i.e., they are directly or indirectly related to programming. Those can be recognised by the  
 313 categories "Data organisation" and "Databases and repositories". The remaining six tools/services  
 314 are two tools for citation (DOI Citation Formatter and STD-DOI), one for persistent identifiers  
 315 (TIB PID Competence Centre), one for finding research data (Google Dataset Search), a password  
 316 organiser (KeePass) and a tool for collaborative working (Microsoft Project).

Tool/Service	Category	Awareness [%]	Usage [%] ▽
Microsoft Project	Collaborative work	88	45
mySQL	Databases and repositories	69	43
Git	Data organisation	40	31
KeePass	Password help	31	22
NoSQL	Databases and repositories	48	20
TortoiseGit	Data organisation	34	20
DOI Citation Formatter	Citation	30	20
TortoiseSVN	Data organisation	33	19
Google Dataset Search	Find research data	36	18
TIB PID Competence Centre	Persistent identifiers	26	15
PostgreSQL	Databases and repositories	37	13
Apache Subversion	Data organisation	26	9
STD-DOI	Citation	15	9.7
GNU Arch	Data organisation	30	5.6

**Table 6:** Awareness and use of tools among researchers who have stated to have a high or very high self accessed RDM knowledge sorted by use among respondents

317 As shown in table 6, a similar distribution can be observed when only reviewing the answers of  
 318 researchers who have stated to have a high or very high self accessed RDM knowledge. Here, 14  
 319 have been used more than once by at least 5% of the respondents. The same focus on software  
 320 development becomes apparent with eight of the 14 listed tools related to this area.

321 The majority of the best-known or most-used tools have in common that they offer solutions to  
 322 researchers' everyday problems (compare finding 3). For example, the versioning tool Git offers  
 323 a possibility to version source code, which can hardly be kept manageable without versioning.  
 324 The added value of Git is known and is also passed on to other researchers, at least in groups that  
 325 regularly deal with source code. The immediate applicability is what separates those best-known  
 326 and most-used tools from especially the less-used RDM tools.

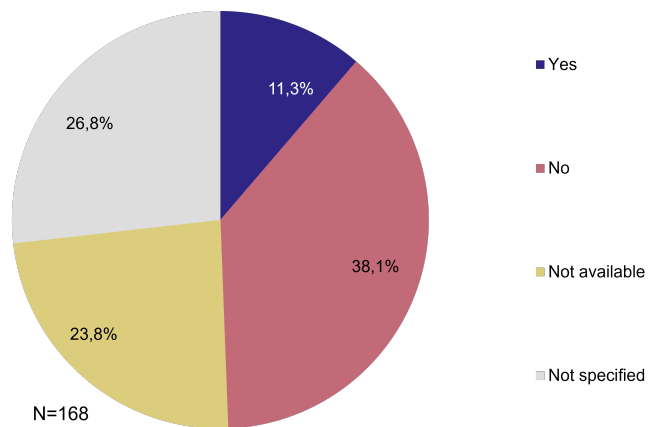
327 Such RDM tools that should mainly accompany the research process, are virtually unknown and  
 328 unused. The majority of respondents thus lacks knowledge about suitable programs, supporting  
 329 tools or services in the context of RDM. Therefore, such programs, tools or services are not used  
 330 by the majority of respondents, which is another hypothesis in this article:

331 4. *Researchers lack awareness about existing solutions for RDM specific problems and*  
 332 *therefore the knowledge and ability to use those solutions.*

#### 333 4.5 Specifications and Support Structures

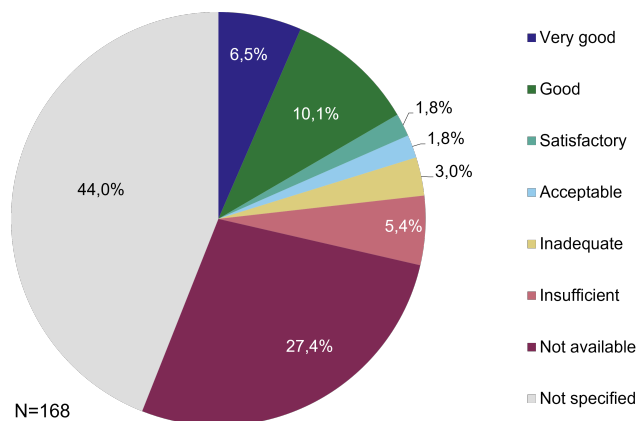
334 The last question set is directed at the requirements and support structures for RDM that are  
 335 specified or offered by the respondents' respective institution. Those include, but are not limited  
 336 to, RDM-Teams at universities, available tools for RDM or specific support at institutions. The  
 337 exact question was "Is there support within your organisation in the area of RDM?".

338 Shown in figure 11 are the responses of researchers asked if they use offered support structures  
 339 at their organisation. Only about one tenth of the surveyed researchers have used offered support  
 340 structures while almost a quarter states there was no support available at their institution. The



**Figure 11:** Distribution of answers on the question: Do you use the existing support services? (Key:FO06)

341 survey did not include any questions regarding why support structures are not used by researchers.



**Figure 12:** Distribution of answers on the question: How good do you rate the support? (Key:FO06)

342 However, there might be two reasons for this. Firstly, support structures are available but not  
 343 known, which is relevant only for the 23% of researchers who claim that there are none. Secondly,  
 344 the benefit of such structures is not perceived as important enough to be worth the expense.  
 345 One third of researchers who know about support structures do not use them despite having the  
 346 opportunity to do so. This, in turn, might be a result of either insufficient support structures (may  
 347 it be in terms of offered service, format or content) or lack of knowledge about how and why  
 348 such structures could improve the interviewees RDM. The survey also asked for an evaluation  
 349 of the offered support structures with the results being shown in figure 12.

350 Combining the data basis from figure 12 with figure 11, there are several groups of researchers  
 351 to be identified, clustered by their access to and their usage of RDM support, shown in table 7.

352 **4.6 Further Open Questions**

353 In further open questions, respondents were given the opportunity to mention possible reasons  
 354 that might prevent researchers from RDM in the form of free text answers. Most interesting are



Group of researchers who have...	Respondents [%]
... access to RDM support and use it.	11.3
... access to RDM support and do not use it.	17.9
... no access to RDM support, but would like to use it.	23.8
... no access to RDM support and do not criticise its absence.	23.8
... not specified it.	27.6

**Table 7:** Groups of researchers clustered by their access to and their usage of RDM support structures

355 the answers on the question "What reasons could prevent researchers from sustainable RDM?",  
 356 which 39 of the 168 interviewees (23%) answered. A detailed list of quotes of the respondents  
 357 can be found in the [Appendix](#). The effort or workload for the establishment and operation of  
 358 RDM is with 16 mentions the most recognisable reason against proper RDM usage. Likewise,  
 359 the lack of clear standards or guidelines for RDM is cited twelve times, closely followed by the  
 360 lack of awareness of RDM among researchers (nine mentions). This last statement is specified:  
 361 RDM is primarily perceived as an additional expense, there is no incentive to use it and no  
 362 necessity for RDM is seen. The lack of necessity is justified by the time-limited nature of projects  
 363 and their isolation in research environments. Other reasons against RDM application are a lack  
 364 of knowledge (seven mentions), the concern of data misuse or data usage without permission or  
 365 citation (six mentions) and problems with missing or complicated support structures, which five  
 366 interviewees mentioned.

367 The feeling that the own data can only be used for the own projects prevails for many. Contrarily,  
 368 others who consider their data to be usable, fear data misuse. In this case the protection of the  
 369 own research is seen as more important than a provision of data within the framework of RDM.  
 370 This is expressed, for example, in the following quote from one of the respondents:

371 "Real data, e.g. from production, is not easy to obtain. Those who have such data  
 372 sets have an advantage. Therefore, data is not shared, although it would be useful to  
 373 promote scientific progress and test results for reproducibility."

374 Many of the interviewees' statements can be condensed into the following statement (adapted in  
 375 wording for the purpose of anonymisation), which was formulated by an interviewee:

376 " Besides the most obvious reason - lack of knowledge - I think [RDM] just meets  
 377 [ignorance] by and large. One Example: For [research] I have collected publicly  
 378 available data. Of course I maintain and cherish my data and go through large  
 379 parts of the data life cycle, but for that I don't need thousands of tools that nobody  
 380 else [in my organisation] uses. It is also likely that others will not (be able to)  
 381 continue to use this data - which is why it makes sense to maintain it sustainably. It  
 382 is similar with research projects. The more isolated and smaller the project is, the  
 383 less sense there really is in elaborate management [...]. This is not only true for the  
 384 data. Furthermore, it is unfortunately inherent in the research system that I could  
 385 suffer great professional damage if I give out my data beyond a certain level. In  
 386 applied research projects the situation is certainly different, but even here I need (at  
 387 least initially) a more or less exclusive use of data so that I can firstly secure my

388 livelihood. Furthermore, there are often confidentiality clauses that do not allow me  
389 to pass on the data.”

390 The free-text answers were used to formulate more hypothesis, as they allowed for deeper  
391 insights, especially when considering reasons against RDM. However, it needs to be mentioned  
392 again that those only originate from 39 of the 168 interviewees (23%) of the interviewees, which  
393 further diminishes the sample size.

- 394 5. *The interviewees see the effort of RDM in terms of initialisation, familiarisation with it*  
395 *and everyday work as a reason that prevents researchers from sustainable RDM.*
- 396 6. *The interviewees name the lack of clear guidance through the RDM process like guidelines,*  
397 *standards or processes as a reason that prevents researchers from sustainable RDM.*
- 398 7. *The interviewees perceive that RDM as a topic does not receive enough awareness yet,*  
399 *which is a reason that prevents researchers from sustainable RDM.*
- 400 8. *The interviewees see a lack of knowledge among themselves and other researchers, which*  
401 *is a reason that prevents researchers from sustainable RDM.*
- 402 9. *The interviewees consider the risk of data misuse and data usage without citation or*  
403 *permission as a reason that prevents researchers from sustainable RDM.*
- 404 10. *The interviewees see the lack or quality of support structures as a reason that prevents*  
405 *researchers from sustainable RDM.*

406 The acceptance of the reuse of data among the respondents is limited. Thus, the “not-invented-  
407 here syndrome” [36] is cited by the respondents. This effect describes the rejection of ideas  
408 and inventions not founded in one’s own institution for reasons other than monetary ones. For  
409 example, openly available data might not be reused because it is not trusted as it is of other origin  
410 as the own institution. As a result, the subsequent use of existing data is omitted and additional  
411 work is done, since data must be collected by the institution itself [36].

## 412 5 Discussion

413 Within this paper ten hypotheses could be drawn, derived from the data of the survey results.  
414 While these ten hypotheses do only provide a qualitative approach to the topic of RDM usage  
415 and application on a small sample size of 168 compared to the population of 20,355 in the  
416 research field under consideration, the survey still provided indications regarding main issues in  
417 the context of RDM and opened the possibility to derive potential measures. In the following the  
418 findings and conclusions drawn from the small sample size are presented.

### 419 5.1 Findings

420 The survey indicates that the knowledge, awareness and usage of RDM has to be fostered  
421 to enhance the management and therefore FAIRness [10] of research data. To achieve this,  
422 researchers firstly need to know what to do when starting managing research data (see hypotheses  
423 4., 5. & 8.). An appropriate approach needs to be handed to them with a clear entry point and  
424 a structured and adaptable process needs to be defined (see hypothesis 6.). When questions  
425 occur, those have to be answered right away (see hypotheses 5. & 10.). Also, training materials  
426 to the very topic of the question have to be provided and suitable tools have to be introduced

427 (see hypotheses 1. & 4.). Those materials should be light-weight and focused on applicability.  
428 Light-weight in this context means that provided information should only focus on the very  
429 specific problem of the researcher. A huge amount of additional and inapplicable instructions  
430 will compromise the will of researchers to use RDM and cause frustration. The process of RDM  
431 has to be embedded within everyday research (see hypothesis 3).

432 Incentivation for RDM usage needs to be provided as the requirements of, for example the DFG,  
433 are not sufficient to enhance the application of RDM (see hypothesis 2.). Also, the awareness for  
434 RDM has to be broadened (see hypothesis 7.). Suitable measures could be the requirements of  
435 RDM in connection with dissertations or bachelor/master theses.

436 Opposing to the incentivation is the fear of data misuse or missing citations of the own work (see  
437 hypothesis 9.). This could be addressed by the possibility of storing data in closed repositories and  
438 clear instructions of how data can be made publicly available in a way that it is unambiguously  
439 recognisable who the author is and to whom the data belongs. Access management and licensing  
440 has therefore to be taken into consideration, granting the possibility of a controlled reuse of data.

## 441 5.2 Comparison of hypotheses and related work

442 To conclude this discussion, a comparison of hypotheses to findings of the literature review  
443 shall be given, ordered by the number of hypotheses listed above. This comparison is drawn to  
444 different disciplines and countries than the scope of this survey. Yet there are some similarities  
445 and common challenges that form a reoccurring pattern in the nature of RDM. In the following,  
446 each hypothesis is referenced by its number and a short hand at the corresponding paragraph.

447 **1: Researchers are missing RDM knowledge** For instance, hypothesis 1 is supported by  
448 several papers. The "lack of trainers in RDM practices" [15], "lack of knowledge/training" [18],  
449 a lack of "data sharing skills" [23], or the need of training as stated by Elsayed and Saleh [13] is  
450 represented in many papers. The only contradiction found in literature by Costanzo et al. states  
451 that "Lack of RDM Knowledge [is a] low barrier" [16].

452 **2: Guidelines do not equal good RDM** Costanzo and Cooper support hypothesis 2, describing  
453 the "lack of institutional understanding and awareness of [...] expectations" [16]. Wilms et  
454 al. state, a "requirement to comply with possible guidelines" [14] is not enough incentive for  
455 researchers to adhere to good RDM practices.

456 **3: RDM is only done when necessary for results** The third hypothesis is not supported by any  
457 findings in the literature. Therefore, this hypothesis could benefit from a revision in the future.  
458 However, Palsdottir states that RDM "is not a normal practice" in the researchers work [21].  
459 Still, the reasons for the usage of tools should be clarified. The hypothesis can not be supported  
460 by literature but is still a finding of this paper.

461 **4: Specific solutions for RDM are unknown** While Björnmalm et al. see the problem in too  
462 many generic and yet too few specific RDM tools [15], Israel et al. state that "respondents  
463 continue to rely on [...] paper laboratory notebooks" [19] instead of electronic laboratory  
464 notebooks. While there are many tools available for RDM activities both generic and specific

465 [19], the "lack of knowledge" [18] about these tools can be seen as the actual challenge RDM is  
466 facing in this context. This also supports hypothesis 4.

467 **5: Effort is a hindrance for RDM** Hypothesis 5 is also represented within literature. RDM is  
468 seen as "a significant burden" [17] as "the amount of time it takes" [18] is a "perceived increased  
469 workload" [14], opposing a "lack of resources (time, budget, personnel etc.)" [16].

470 **6: Uncertainty is a hindrance for RDM** Connected to the effort required for RDM, the lack  
471 of guidance (hypothesis 6) is found both in the answers of this survey as well as the literature.  
472 Björnmalm et al. found a lack of "specific instructions (or links to relevant guidelines)" [15],  
473 which is supported by Costanzo et al. regarding the "lack of institutional understanding and  
474 awareness of [...] expectations" [16] as well as the findings of Borghi and Van Gulick that there is  
475 missing guidance through "lack of best practices" [18]. The "large number of tools and methods"  
476 [19] and "complexity in data structures [,] formats [and] documentation" [19] is a challenge yet  
477 to be faced. As "processes are not yet clearly defined, let alone standardised" [20] "researchers  
478 needed assistance" [20] in RDM, which is also supported by [21]. Additionally, "establishing  
479 [...] guidelines" can improve RDM [22].

480 **7: Awareness for RDM is low** Many papers also address hypothesis 7, however some support  
481 it while others oppose it. While Björnmalm et al. see "too few incentives for researchers that  
482 reward and incentivise implementation of RDM practices into everyday workflow" [15], Wilms  
483 et al. see that the "overall acceptance of RDM policies is low" [14]. According to Austin et al.  
484 there is a "need to demonstrate to researchers the value of data management" [17]. Similarly,  
485 Borghi and Van Gulick point out that the importance of RDM is not commonly known [18]. These  
486 four statements support hypothesis 7. Israel et al. point out that "making data FAIR needs to start  
487 most importantly, awareness" [19], also supporting hypothesis 4 to some extent. However, Vilar  
488 and Zabukovec oppose these theories, stating that researchers are rather convinced by RDM [25].  
489 Ortloff et al. also argue in their spotlight investigation that "most of the partners are strongly  
490 aware of the benefits provided" [22] by RDM. The incentivisation of RDM, as for example brought  
491 up by Borghi and Van Gulick, has to be addressed by funding organisations, universities and  
492 institutions. However, it is not part of this paper, as the focus lies on the researchers perspective  
493 on RDM. Still, the topic of incentives has to be considered from all sides, from making funding  
494 dependent on concrete RDM practices to the demanded RDM in the context of a dissertation.

495 **8: Missing knowledge hinders RDM's application** While hypothesis 8 is not directly supported  
496 or opposed by literature, it is to some extent a consequence from hypotheses 1 and 4. Palsdottir  
497 states the "limited knowledge" and that RDM "is not a normal practice" as well as an "urgent need  
498 to increase the researcher's knowledge and understanding of the importance of data management"  
499 [21]. However, it can neither be contradicted nor be proven that the lack of knowledge hinders  
500 the application of RDM. The lack of knowledge has been stated several times, both in this survey  
501 and the literature. A plausible outcome might be the hindering of (sustainable) RDM.

502 **9: Researchers fear data sharing** The ninth hypothesis is addressed by five papers. Austin et  
503 al. state that more than half of the involved partners in the projects rejected data sharing [17].

504 This is mostly based on the "concerns regarding IP protection" [22] respectively "intellectual  
505 property rights" [24] and the "fear of losing control" [14]. The "partner's consent for publication  
506 was the biggest hurdle" [20].

507 **10: RDM support is insufficient** Lastly, hypothesis 10 is supported by some papers. Elsayed  
508 and Saleh see a need for support [13] as well as [21], while Björnmalm et al. see a lack of  
509 "support at a faculty level" [15], similar to the "lack of availability of support materials" [16]  
510 stated by Costanzo et al. Wuchner et al. also see a need for support, but on a more immediate  
511 level. While the aforementioned papers focus on generic support, Wuchner et al. see a direct  
512 assistance needed for "data publications – especially FAIR ones [because they are] are a major  
513 challenge for researchers" [20]. This last statement excluded, all papers revolve around the lack  
514 of support, which is partially true, but might also be a consequence of the lack of knowledge and  
515 awareness, as stated in hypotheses 1, 4 and 8.

## 516 **6 Limitations**

517 Although the hypotheses formulated in this article are mostly supported by literature, the survey  
518 has limitations. Firstly, the sample size of 168 respondents is rather small, which is caused by the  
519 low response rate of 2.5%. About 0.8% of the population of researchers in German mechanical  
520 and industrial engineering sciences were reached. Secondly, the response rate of the researchers  
521 located at the RWTH Aachen University was significantly higher, resulting in a strong bias  
522 of respondents as 43% of them work at the RWTH. Hence, the survey should be seen as an  
523 exploratory assessment rather than a statistically valid and quantitative analysis.

524 As a consequence, the free-text statements of respondents are even less reliable. Only 39 of  
525 the 168 interviewees (23%) took the opportunity to communicate their reasons against the  
526 application of RDM, which further diminishes the sample size. Yet, these statements were the  
527 ones giving most insights in the problems of the surveyed researchers in the context of RDM.  
528 Further investigation on how to incentivise researchers for RDM is required. In consideration  
529 of the literature presented, the needed incentives need to originate from the researchers own  
530 intrinsic motivation, demanding for more awareness within them.

531 Both the small sample size of the survey and the low answer rate of some questions in the survey  
532 point towards the need of a shorter survey. This would cause the participation time to be not as  
533 long, meaning more researchers are likely to fill out the survey. Again, the current state of the  
534 survey does not allow for more insights derived from the responses.

## 535 **7 Summary and Outlook**

536 This paper has shown the results of a survey that took place from October to December 2020.  
537 With 168 researchers, a rather small sample size was interviewed and the results were derived  
538 from their answers to the 39 questions within the survey. Main topics of the survey as well as  
539 (sub)sections within this paper were "RDM Knowledge and Perceived Relevance of RDM",  
540 "Application of RDM Related Tasks", "Specifications and Support Structures" and responses to  
541 "Further Open Questions". The survey aimed to answer the following research question:

542 ***What is the current status of RDM the field of mechanical and industrial engineering in***  
543 ***German Engineering Sciences?***

543 This question was answered in the form of hypotheses, as the sample size is considered too  
544 small to state in depth statistical analysis with a sufficient confidence interval. These hypotheses  
545 indicate the current status of mechanical and industrial engineering in German Engineering  
546 Sciences. The hypotheses can also be summarised: Researchers in engineering sciences are in  
547 need for guidance and support regarding RDM in their everyday research. This results from  
548 the main reasons against RDM, namely missing knowledge about guidelines, tools and support  
549 in RDM as well as the additional effort connected. Guidance should be provided in form of  
550 use case related processes that integrate into everyday research and support researchers with  
551 knowledge and tool support when needed.

552 Although the survey took place in 2020, the results are still considered relevant, as the cultural  
553 change towards RDM and open science the German engineering sciences are currently undergoing  
554 are yet to be finished. Furthermore, new researchers entering the German engineering science,  
555 may it be by migration, by graduation or change of career, will face the same cultural change by  
556 them selves, meaning they presumably have the same attitude towards RDM as the interviewees  
557 of the survey. However, as RDM enters curricula and is adapted by more and more researchers  
558 of German engineering sciences, the relevance of the current status will vanish.

559 Hence, future research on the same topic will be able to document the ongoing cultural change  
560 and its success or failure. Additionally, further research on RDM requirements of researchers,  
561 integration of RDM into everyday research, general feasibility and practices resulting would  
562 support the application of RDM, eventually leading to a broad adoption in the engineering  
563 community. The applicability and usability of RDM should be fostered to facilitate the needed  
564 cultural change in engineering sciences.

565 Additionally, the authors would like to point out that a complete statistical analysis of the linked  
566 data might result in further findings, especially if the data is combined with similar data of other  
567 sources. As a standalone, the linked data could have a too small sample size for a complete  
568 statistical analysis. The linked data is specifically intended to be reused.

## 569 8 Appendix

570 The appendix holds both more information on the [Related work](#) and the reasons against RDM  
571 brought up by the participants of the survey.

### 572 8.1 Additional info on the review of related work

573 In the following, firstly the review process is displayed before an explanation on every included  
574 record and their contents.

#### 575 8.1.1 Review process

576 A complete diagram of the review process with references to exclusion reasons is depicted in  
577 [figure 13](#).

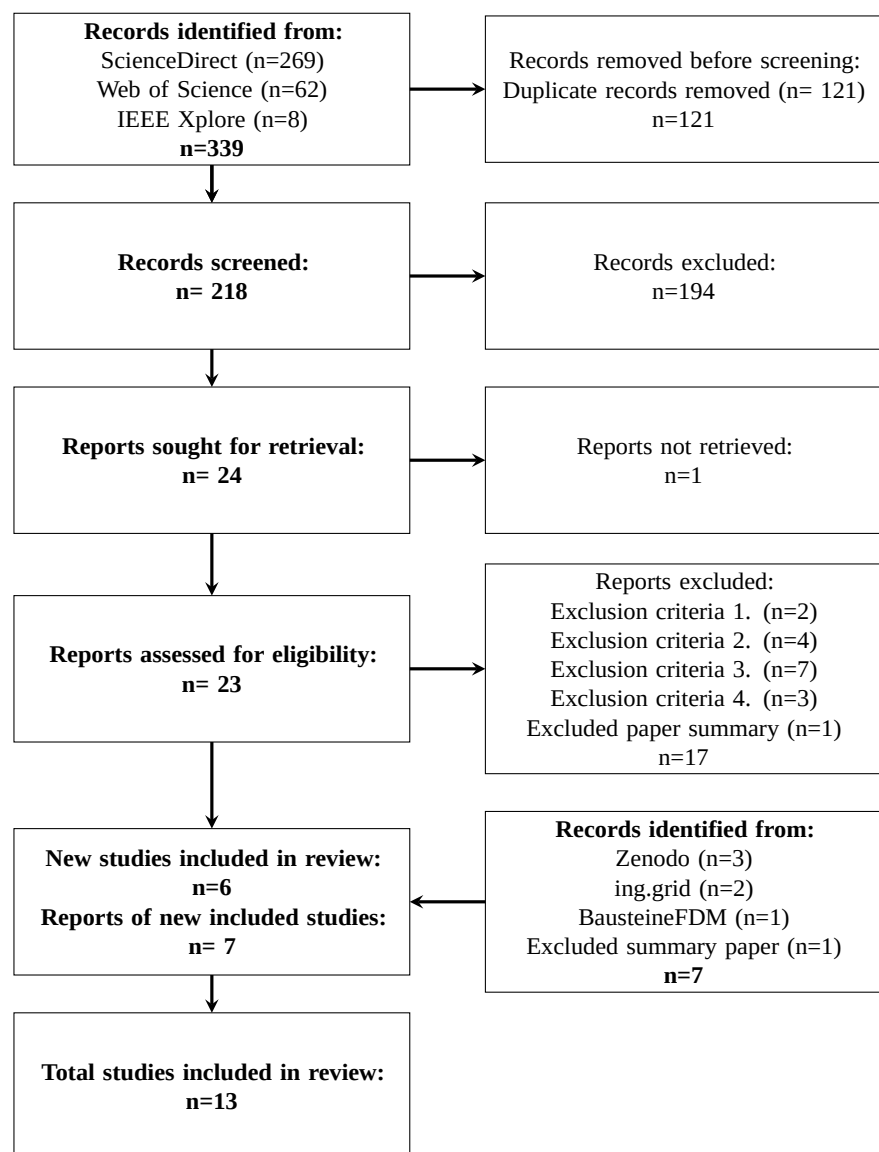


Figure 13: PRISMA 2020 Flow Diagram, c.f. [37]

**578 8.1.2 Information of related work by authors**

579 In the following, each record included from the literature review is presented.

580 Wilms et al. present "a quantitative study of the factors affecting researcher's intention to comply  
581 with guidelines on handling research data" [14]. A total of 111 researchers from the discipline  
582 of information systems in Germany responded to the survey. While the subject of information  
583 systems is part of the IT sciences, it is still considered technical enough for this paper. They point  
584 out that the "overall acceptance of RDM policies is low" [14], that "90 % of the participants  
585 indicate that they do not use institutional or national standards" [14] for research data management  
586 and that "a large part of respondents claimed not to practise RDM" [14]. The "requirement to  
587 comply with possible guidelines is clearly not sufficient to convince researchers to change their  
588 current inadequate data management strategies" [14]. On the one hand, uncertainty is listed as  
589 one possible explanation, as it results from the fear of losing control over the own data, on the  
590 other hand "uncertainty can prevent people from choosing an option even if they evaluate it as  
591 more beneficial" [14]. Another reason for the lack of RDM usage is the "perceived increased  
592 workload" [14]. A possible solution might be the provision of technologies to support RDM and  
593 "convince them that no additional technical effort is required" [14].

594 Björnmalm et al. conducted a survey on institutional level on which 21 universities of science  
595 and technology united within CESAER participated. They see the challenges of RDM in the lack  
596 of "specific instructions (or links to relevant guidelines)" [15] of RDM policies and "support at a  
597 faculty level" [15] and in the lack of "lack of trainers in RDM practices" [15]. It is concluded that  
598 there are on the one hand too many generic RDM tools, but on the other hand yet too few specific  
599 ones. Also, the missing "incentives for researchers that reward and incentivise implementation  
600 of RDM practices into everyday workflow" [15] are criticised. One of the recommendations they  
601 draw from their survey are the introduction of discipline-specific workflows, that "should provide  
602 information tailored to science and technology disciplines, e.g. data infrastructures available for  
603 the different types of data produced, different tools for documentation, implications of producing  
604 data following the FAIR principles, and when and how to publish their research data. In essence,  
605 help researchers make better sense of high-level (university-wide) requirements" [15]. Another  
606 recommendation is, to utilise "solutions with open APIs to facilitate the integration of relevant  
607 tools and software and to safeguard long-term function" [15].

608 A presentation of Costanzo et al. on IASSIST 2023 contained the results of two surveys from  
609 2019 and 2022. The focus was laid on the application of the "Tri-Agency RDM Policy" [16], that  
610 states "to support Canadian research excellence by promoting sound RDM and data stewardship  
611 practices" [16]. Main institutions representing the "Tri-Agency RDM Policy" are the Canadian  
612 Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council  
613 of Canada (NSERC), and the Social Sciences and Humanities Research Council of Canada  
614 (SSHRC) [16]. Main barriers for the proper application of RDM are the "lack of resources (time,  
615 budget, personnel etc.) [,] lack of institutional understanding and awareness of the Tri-Agency  
616 expectations [and] lack of availability of support materials" [16].

617 Austin et al. reviewed ten engineering research projects that have been conducted as Open  
618 Research Data pilots at the Horizon 2020 research programme. While the paper sets a focus on  
619 avantgarde projects that specifically aim for the application of RDM, the findings for engineering



620 sciences still offer a value for this paper. The "need to demonstrate to researchers the value of  
621 data management" [17] is clearly stated to point out the need for a change in research culture.  
622 More than half of the involved partners rejected data sharing. Another challenge is the effort of  
623 RDM, as "data gathering tasks will remain a significant burden [...] until [...] data technologies  
624 (i.e. interoperability standards) required for seamless data exchange and aggregation" [17] have  
625 been developed. While possible solutions are also discussed, the presented challenges in the  
626 presented projects can be expected to occur in most research projects in engineering sciences.

627 While their paper is set in neuroimaging, Borghi and Van Gulick point out the current challenges  
628 of RDM in their field. They figure that the researchers "ubiquity indicates that there is not an  
629 optimal amount of communication about the importance of RDM even within individual research  
630 groups or projects" [18]. Additionally, they point out limitations of RDM and reasons against  
631 data sharing. Limiting factors are "the amount of time it takes [... with at least] 69.60%[, a] lack  
632 of best practices [... with at least] 43.20%[, the] lack of incentives [... with at least] 32.18% [and  
633 the] lack of knowledge/training [... with at least] 32.80%" [18]. The main reason against data  
634 sharing is the fear of use of not yet analysed/sensitive data, with 50% respectively 30%. [18]

635 When taking a look at life sciences and engineering in the universities in Egypt, Jordan and  
636 Saudi Arabia, Elsayed and Saleh [13] found, that "42% [of researchers are] unfamiliar with data  
637 management plans" [13] and "more than half [... have] no data management plan". They state,  
638 that "despite researchers' recognition of the importance of data sharing, they lacked the capability  
639 to actually share data" [13] and that "the practice of depositing data in open data repositories  
640 was not prevalent" [13]. "56% indicated that they needed training in RDM" [13].

641 From March to May of 2020, Israel et al. "conducted an online survey among research physicists  
642 in Germany [...] to determine the status of their RDM and the resulting agenda for an NFDI  
643 consortium" [19]. While the focus lies on physicists, it has a very similar scope to this papers goal  
644 in performing a broad survey on the status quo of RDM. 237 complete answers from universities  
645 all over Germany could be collected via the survey. This survey was also conducted in the  
646 context of the German National Research Data Infrastructure (NFDI) initiative. Their findings  
647 point out that "documentation of research activities is not as seamlessly digitized" [19], for  
648 instance instead of electronic laboratory notebooks (ELNs), paper laboratory notebooks are still  
649 being used. The main challenges of RDM are stated as the "complexity in data structures and  
650 formats (69% approval), the large number of tools and methods (61% approval), complexity of  
651 documentation (59% approval), and confusion about underdeveloped metadata standards (50%  
652 approval)" [19]. Their most important conclusion in the context of this paper is the following:  
653 "The 2020 survey on RDM in physics has shown that making data FAIR needs to start at the  
654 foundational level of terminology, file formats and, most importantly, awareness." [19]. Physics  
655 sciences in Germany do "not live up to the standards of RDM best practices" [19].

656 Wuchner et al. present a case study with no broad survey. Still, there are findings specifically  
657 relevant for engineering sciences. They point out the lack of clearly defined or even standardised  
658 processes. Additionally it is stated, that "for the researcher, obtaining the project partner's consent  
659 for publication was the biggest hurdle" [20], reinforcing the statement of Ortloff et al. [22] about  
660 concerns regarding intellectual property protection. If researchers are introduced to new tasks,  
661 assistance is needed, for example, in the case study "the researcher needed assistance in the

662 publication process, especially since it was his first” [20]. There is a ”need for experts to assist  
663 researchers with data publications and overall research data management” [20], last but not least  
664 because ”data publications – especially FAIR ones – are a major challenge for researchers” [20].

665 A similar survey has been conducted in Iceland by Palsdottir in 2017. Out of the 139 respondents  
666 about 39% originated from sciences, containing engineering sciences [21]. It was found that  
667 ”the researchers had limited knowledge about the procedures of data management [, ...] it is not  
668 a normal practice in their research work” [21] and ”that there is an urgent need to increase the  
669 researcher’s knowledge and understanding of the importance of data management [...], as well  
670 as to provide them with the resources and training that enables them to make effective [...] use  
671 of data management methods” [21]. It is concluded that information specialists are needed to  
672 assist in the design of RDM services to support researchers in their data management [21].

673 In contrast, Ortloff et al. [22] point out that the ”interviewed partners are aware of the Open  
674 Access requirements and the FAIR principles” [22] and that ”most of the partners are strongly  
675 aware of the benefits provided by extended data usage and the respective demands” [22]. While  
676 they conclude that ”there are concerns regarding IP protection and data security” they also state  
677 that ”establishing proper templates, guidelines, and training for data collection, analysis, and  
678 sharing” can improve RDM practices. A cultural shift is seen as urgently needed in many of the  
679 interviewed organisations [22]. These conclusions are drawn from a ”spotlight investigation”  
680 [22] based on expert interviews, not a wide range of researchers from engineering.

681 A presentation by Melissa Cheung at IASSIST May 2021 points out restrictions on data sharing  
682 in engineering. Again, the concern about ”intellectual property rights (24%)” [24] is listed as  
683 very important, second to the ”Need to publish before sharing (50%)” [24].

684 Chawinga et al. describe motivational factors as well as challenges listed in 105 papers. While  
685 the motivational factors shall not be discussed here, the challenges of RDM need to be taken into  
686 consideration although the focus of Chawinga et al. is set on funding and institutional matters,  
687 they still point out that 92% of papers list the data sharing skills as an issue for RDM [23].

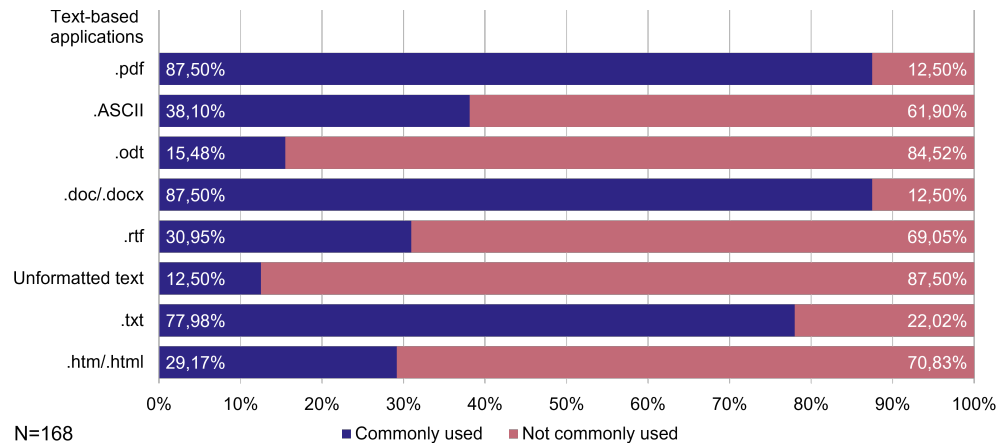
688 In 2021, Polona Vilar and Vlasta Zabukovec conducted an online survey on research data  
689 management in Slovenian science, including engineering sciences [25]. They differentiate  
690 between the perception and the behaviour of researcher to point out groups of researchers based  
691 on their discipline. They state that researchers from the engineering sciences perceive RDM as  
692 unproblematic and are rather convinced by it. In terms of behaviour, engineering researchers  
693 show a considerable spread in their answers. Some do not utilise metadata and follow no file-  
694 naming conventions/standards, while others often use file-naming conventions/standards along  
695 with version-control systems and are experienced with public-domain data.

## 696 8.2 Further information on results

697 Below, additional results of the survey are presented that do not directly contribute to the  
698 answering of the research question but may be beneficial for further research on different aspects  
699 of RDM.

### 700 8.2.1 Usage of File Formats

701 The survey also asked about the frequently used file formats. 31 file formats as well as oppor-  
 702 tunities for free text answers were given. The interviewees could choose whether or not they  
 703 use that file format. File formats cover the MS Office family, PDF and common image and  
 704 video formats as well as formats for quantitative data and text-based formats. The later ones also  
 705 contain file formats for source code such as .py or .cpp.



**Figure 14:** Common usage of text-based file formats among interviewees - Distribution of answers on the question: In the following, we ask you to mark the file formats you use frequently or to add further formats. (Keys: D101\_12 to D101\_18 and D101\_23)

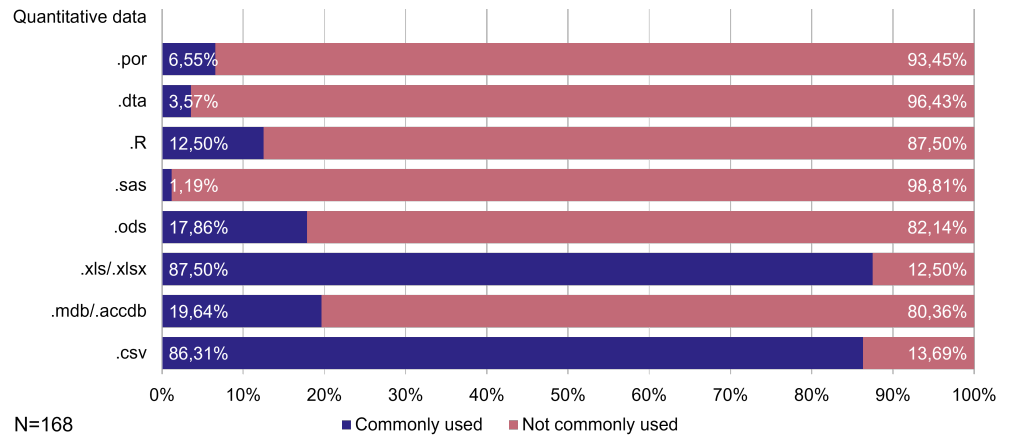
706 When reviewing the results for file formats in text-based applications, a strong distinction between  
 707 commonly used and not commonly used formats is possible (see figure 14). MS Word files (.doc  
 708 or .docx), just like PDF documents, are frequently used by 87% of the respondents. With 78%,  
 709 .txt is the most frequently used format for unformatted text. Other file formats are commonly  
 710 used by a minority of the interviewees as shown in figure 14.

711 MS Excel files (.xls or .xlsx) are used by 87% of the respondents (see figure 15). Close behind  
 712 (86%) is .csv, another file format usable in Excel. Again, other file formats are much less  
 713 commonly used than the aforementioned, making the distinction between commonly used file  
 714 formats and not commonly used file formats very unambiguous.

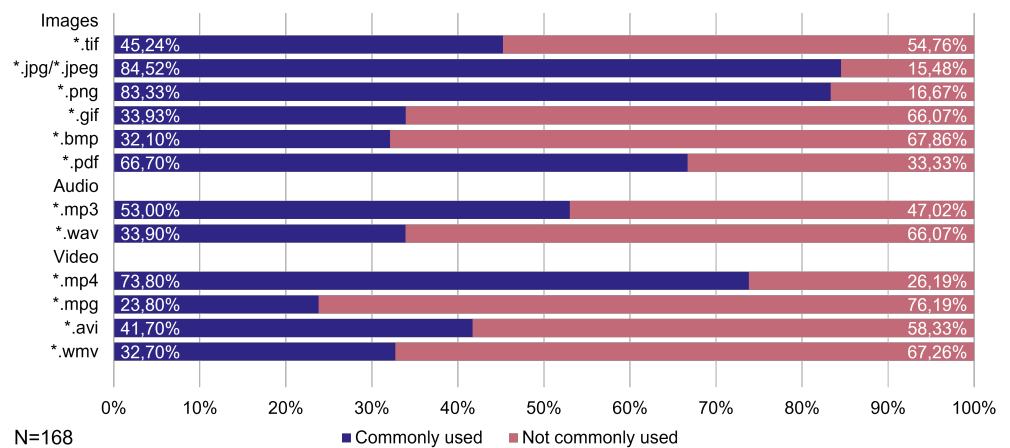
715 For media files (image, audio and video files), the spread in the answers given is not nearly as  
 716 pronounced as for example in quantitative data. However the aforementioned formats .jpg/.jpeg,  
 717 .png, .mp3 and .mp4 are predominant for their respective category (see figure 16).

718 The commonality of the aforementioned formats is their general widespread use, familiarity  
 719 and the resulting usability. All these can be used on a standard Windows PC with MS Office  
 720 installed, without the need for further installations. The latter is a factor not to be neglected. On  
 721 the one hand an installation of further programmes may have to be carried out by corresponding  
 722 IT departments, which is associated with personnel and time expenditure. On the other hand,  
 723 depending on the file format, there are licence fees for associated programmes. The latter becomes  
 724 more important if there are free or already available alternatives in the work environment.

725 This relation is expressed most strongly in the processing of quantitative data, e.g. table-based



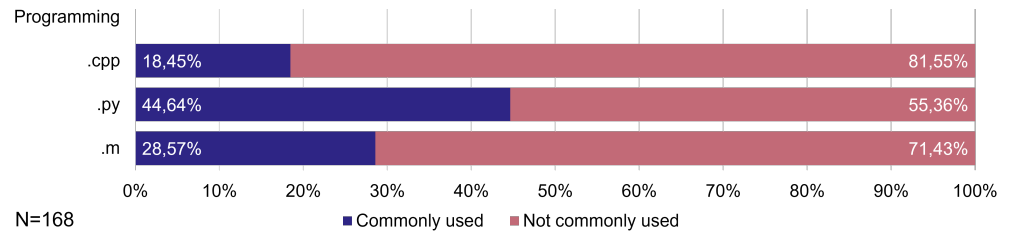
**Figure 15:** Common usage of file formats for quantitative data among interviewees - Distribution of answers on the question: In the following, we ask you to mark the file formats you use frequently or to add further formats. (Keys: D101\_02 to D101\_09)



**Figure 16:** Common usage of media file formats among interviewees - Distribution of answers on the question: In the following, we ask you to mark the file formats you use frequently or to add further formats. (Keys: D201\_25 to D201\_30, D201\_33, D201\_34 and D201\_37 to D201\_40)

726 evaluation of data through Excel. MS Office, including Excel, is one of the standard installations  
 727 on Windows PCs, as already mentioned above. Therefore, the use of .csv, .xls and .xlsx files is  
 728 possible on the majority of Windows PCs; these formats are used by 87% of the respondents. In  
 729 contrast, the use of the .por format, which was developed by IBM for the statistical programme  
 730 SPSS and is only used by 6% of respondents, is only possible in this very programme [38]. For  
 731 other formats in the field of quantitative data, the usage rates are hardly higher and formats usable  
 732 with Excel seem to be the only option. In contrast, only 15% of respondents use the .odt format,  
 733 although this can also be opened and edited in licence-free and openly available programmes.

734 The usage of file formats is primarily based on programmes and tools available and the usability  
 735 of the formats. The usability is partly dependent on the availability of programmes or their  
 736 corresponding licences. It is unclear why specific programming languages and file formats (see  
 737 figure 17) are used in software development. The reasons for or against an approach are not part  
 738 of the survey, as researchers should be supported in everyday research and not forced into new



**Figure 17:** Common usage of file formats used in programming among interviewees - Distribution of answers on the question: In the following, we ask you to mark the file formats you use frequently or to add further formats. (Keys: D201\_19 to D201\_21)

739 directions. The collected knowledge about the used file formats used does not provide any direct  
 740 recommendations for action to advance RDM. It rather shows the heterogeneous file formats  
 741 that need to be taken into account when working with research data.

### 742 8.2.2 Further reasons against RDM

743 Interviewees were asked "What reasons could prevent researchers from sustainable research data  
 744 management?". Their answers on this questions can be found below. The statements are split up  
 745 into the following categories:

- 746 • Effort
- 747 • Guidelines and Standards
- 748 • General Acceptance, Discipline and Awareness of RDM
- 749 • RDM Knowledge
- 750 • Data Misuse and Permissions
- 751 • Support Structures
- 752 • Longer Statements

753 Some statements contained content that would fit into multiple of these categories. Such state-  
 754 ments were split into two or more parts and listed in the corresponding category if the meaning  
 755 was untouched by the split. If a concrete distinction between two parts cannot be made within  
 756 one statement, the quote will be listed in multiple categories.

757 **Effort** One of the main concerns of the interviewed researchers is the effort connected to RDM.  
 758 16 of the 39 free-text answers mentioned the effort or time expenditure as a reason to not manage  
 759 research data.

- 760 • "Time-limited projects that one works on alone. Sustainable and systematic data storage  
 761 usually **only additional effort**."
- 762 • "Time required for upkeep"
- 763 • "Much too elaborate, no predefined structures. Clear specifications must be applicable  
 764 and clear"

- 765 • "Time expenditure"
- 766 • "Effort"
- 767 • "**Effort during set-up**"
- 768 • "Lack of time"
- 769 • "Effort and time"
- 770 • "**Additional effort is considered too high - regardless of the desire for implementation.**
- 771 **Familiarisation with formats is too time-consuming, as step-by-step introduction along**
- 772 **the daily work routine is not available."**
- 773 • "Too much effort"
- 774 • "**High organisational and training costs** with low capacities"
- 775 • "**Too complicated**, no infrastructure, no advice, no support, importance is not rewarded"
- 776 • "**Increased documentation effort**, restrictions in the use of file formats and systems for
- 777 data storage"
- 778 • "lack of processes - lack of contact persons - **time expenditure** / "inertia" → initially
- 779 no direct benefit for the person who has to do RDM - lack of IT infrastructure - lack of
- 780 know-how regarding data migration, data security, data representation, etc."
- 781 • "**Sustainable RDM takes time** and goes beyond use in own promotion - joint effort needed."
- 782 • "Ignorance and carelessness, **additional effort if there are no clear rules** from the begin-
- 783 ning"
- 784 • "Extensive/varied software to support - lack of standardisation? - Lack of knowledge? -
- 785 **High effort in the life cycle (pre-planning, ..., archiving)"**
  
- 786 **Guidelines and Standards** The following twelve quotes make statements about guidelines and
- 787 standards not being sufficient or too ambiguous.
- 788 • "Lack of awareness, **no existing or communicated guidelines**"
- 789 • "**Ambiguities** in the specifications"
- 790 • "Ignorance and carelessness, **additional effort if there are no clear rules** from the begin-
- 791 ning"
- 792 • "Much too elaborate, **no predefined structures. Clear specifications must be applicable**
- 793 **and clear**"
- 794 • "The **lack of time** to deal with new formats/tools and to carry out extensive data prepara-
- 795 tion."
- 796 • "Missing or unclear specifications."
- 797 • "Researchers are **not aware of what proper research data management should look like.**"

- 798 • **"No information culture** regarding RDM exists. Framework conditions are completely  
799 unknown"
- 800 • **"Lack of knowledge. Non-existent guidelines in the organisation"**
- 801 • **"Too complicated, no infrastructure, no advice, no support, importance is not rewarded"**
- 802 • **"lack of processes** - lack of contact persons - time expenditure / "inertia" → initially  
803 no direct benefit for the person who has to do RDM - lack of IT infrastructure - lack of  
804 know-how regarding data migration, data security, data representation, etc."
- 805 • **"Extensive/varied software to support - lack of standardisation? - Lack of knowledge? -**  
806 **High effort in the life cycle (pre-planning, ..., archiving)"**
- 807 **General Acceptance, Discipline and Awareness of RDM** Nine researchers referred to general  
808 acceptance of RDM as well as discipline and awareness issues.
- 809 • **"Own evaluations paired with expertise"**
- 810 • **"Lack of awareness. Silo thinking"**
- 811 • **"No sense of necessity"**
- 812 • **"Negligence, workload, ignorance, too much variety of options"**
- 813 • **"Benefits not always easily recognisable for others"**
- 814 • **"Meaning-making. Knowledge of the tools"**
- 815 • **"No more recognisable added value in relation to the effort involved in familiarisation**  
816 **when it also works with self-structured Excel files."**
- 817 • **"In my opinion, it is much more important that the generated data can also be reproduced**  
818 **by third parties. Therefore, for me, providing the code in conjunction with a sandbox**  
819 **environment is much more important than the data itself."**
- 820 • **"Agreement on duration of employment/project duration. A large part of the data is only**  
821 **generated towards the end of the project duration/employment contract period, as the**  
822 **experimental facilities must first be set up and put into operation. And: Lack of state**  
823 **positions/permanent positions and high additional workload due to teaching/relocation"**
- 824 **RDM Knowledge** Seven quotes addressing RDM knowledge issues are listed below.
- 825 • **"Too little own expertise and too much effort for familiarisation. Offers and tools not**  
826 **sufficiently known. Especially the technological progress: Often standard software from**  
827 **10 years ago no longer runs on new operating systems, media for persistent storage lose**  
828 **their functionality in the medium term, necessary software and the knowledge to use this**  
829 **software could no longer be available after a few years."**
- 830 • **"There are many tools but too little experience to choose the appropriate ones."**
- 831 • **"Excessive number of tools. No clear place to save."**

- 832 • "No information culture regarding RDM exists. Framework conditions are completely  
833 unknown"
- 834 • "Lack of knowledge. Non-existent guidelines in the organisation"
- 835 • "Extensive/varied software to support - lack of standardisation? - Lack of knowledge? -  
836 High effort in the life cycle (pre-planning, ..., archiving)"
- 837 • "lack of processes - lack of contact persons - time expenditure / "inertia" -> initially  
838 no direct benefit for the person who has to do RDM - lack of IT infrastructure - **lack of**  
839 **know-how regarding data migration, data security, data representation, etc.**"

840 **Data Misuse and Permissions** Another concern of researchers is the fear of data misuse or  
841 data usage without permission or citation, mentioned six times.

- 842 • "**Protection of own research, as not everything has been published yet**"
- 843 • "**Fear of data misuse (publication without naming the source or similar)**"
- 844 • "**Fear for data sovereignty**"
- 845 • "Data loss, violation of DFG rules"
- 846 • "Fear that third parties could overtake you in your own research. Worry that one's own  
847 data has not been collected or analysed cleanly enough. (But hey, others only boil with  
848 water, too)"
- 849 • "Real data, e.g. from production, is not easy to obtain. Those who have such data sets  
850 have an advantage. Therefore, data is not shared, although it would make sense to do so  
851 in order to promote scientific progress and check results for reproducibility."

852 **Support Structures** Last but not least, five of the quotes contain comments on support structures  
853 etc. and what reasons against RDM are connected to those.

- 854 • "There is little support [at my institute]. **Training and education on tools and possibilities**  
855 **would be particularly useful, as would an institute-wide standard. Solutions for individual**  
856 **projects are currently failing due to the IT department and the administration. (Topic**  
857 **licences, accesses, installations)"**
- 858 • "Much too elaborate, **no predefined structures. Clear guidelines must be applicable and**  
859 **clear**"
- 860 • "**Non-existent or impractical to use infrastructure.**"
- 861 • "Too complicated, no infrastructure, no advice, **no support, importance is not rewarded**"
- 862 • "lack of processes - lack of contact persons - time expenditure / "inertia" -> initially  
863 no direct benefit for the person who has to do RDM - **lack of IT infrastructure** - lack of  
864 know-how regarding data migration, data security, data representation, etc."

865 **Longer Statements** As wrap up, two rather long statements that address multiple of the topics  
866 listed above may be cited:



867 *”Lack of tool support. Unclear what ”research data” comprises. The DFG defi-*  
868 *inition is very broad and thus not very clear. Classically, it was measurement and*  
869 *observation data, interview data and the like. In the meantime - and this is also well*  
870 *reflected in some of the questions in this survey - the term encompasses practically*  
871 *every piece of information that a researcher comes across in his or her life. But*  
872 *this is difficult because everyone (if one takes the principle of assignability of ideas*  
873 *strictly seriously) would have to keep a complete documentation of all conversations,*  
874 *impressions, experiences in the professional and private environment because it*  
875 *cannot be ruled out that a remark made by a third party during small talk, remem-*  
876 *bered by chance weeks later, provides the decisive push to get ahead with a problem*  
877 *in a completely different context. Lack of awareness - It is now common knowledge*  
878 *that primary data must be kept secure. What primary data is is more of a question,*  
879 *especially in disciplines that are more constructive and less observational/measur-*  
880 *ing. Not only in data management, but also there: ”Not invented here” syndrome*  
881 *(especially in software-heavy projects a widespread nuisance, partly forced by too*  
882 *tight copyright / too tight patent protection).”*

883 *”Apart from the most obvious reason - lack of knowledge - I believe that it simply*  
884 *encounters a lot of irrelevance in various fields on the whole. Ex: I collected publicly*  
885 *available data for my dissertation. Of course I maintain and care for my data and*  
886 *go through large parts of the data life cycle, but for that I don’t need thousands of*  
887 *tools that no one else at the [institute] uses. Also, others will probably not (be able*  
888 *to) continue to use this data - this also results in the meaninglessness of sustainable*  
889 *maintenance. It is similar to research projects. The more isolated and smaller the*  
890 *project, the less sense there really is in complex management around it. This does*  
891 *not only apply to the data. Moreover, it is unfortunately inherent in the research*  
892 *system that I could suffer great professional damage if I give out my data beyond a*  
893 *certain level. In applied research projects the situation is certainly different, but*  
894 *here, too, I need (at least initially) a more or less exclusive use of data so that I can*  
895 *initially secure my livelihood. Furthermore, there are often confidentiality clauses*  
896 *that do not allow me to pass on the data.”*

## 897 **9 Acknowledgements**

898 The authors would like to thank the Federal Government and the Heads of Government of the  
899 Länder, as well as the Joint Science Conference (GWK), for their funding and support within the  
900 framework of the NFDI4Ing consortium. Funded by the German Research Foundation (DFG) -  
901 project number 442146713 [35].

## 902 **10 Roles and contributions**

903 **Tobias Hamann:** Conceptualization and Methodology of the survey evaluation, Writing

904 **Amelie Metzmacher:** Conceptualization, Methodology and Execution of the survey

905 **Patrick Mund:** Conceptualization and Methodology of the survey

906 **Marcos Alexandre Galdino:** Writing - Review

907 **Anas Abdelrazeq:** Writing - Review

908 **Robert Schmitt:** Idea, Supervision, Funding acquisition

## 909 References

- 910 [1] Statista IDC. “Volume of data/information created, captured, copied, and consumed  
911 worldwide from 2010 to 2020, with forecasts from 2021 to 2025 (in zettabytes).” (2021),  
912 [Online]. Available: [https://www.statista.com/statistics/871513/worldwide  
913 -data-created/](https://www.statista.com/statistics/871513/worldwide-data-created/) (visited on 03/15/2023).
- 914 [2] G. Bell, T. Hey, and A. Szalay, “Computer science. Beyond the data deluge,” eng, *Science*  
915 (*New York, N.Y.*), vol. 323, no. 5919, pp. 1297–1298, 2009, Journal Article. DOI: [10.112  
916 6/science.1170411](https://doi.org/10.1126/science.1170411). eprint: [19265007](https://arxiv.org/abs/19265007).
- 917 [3] T. Hey and A. Trefethen, *The Data Deluge: An e-Science Perspective*, Mar. 11, 2003. DOI:  
918 [10.1002/0470867167.ch36](https://doi.org/10.1002/0470867167.ch36).
- 919 [4] D. Williams and H. Tang, *Data quality management for industry 4.0: A survey*. 2020. [On-  
920 line]. Available: [https://search.proquest.com/openview/f7aaf15f64ccb03285  
921 2e4edbd1ad16c9/1?pq-origsite=gscholar&cbl=25782](https://search.proquest.com/openview/f7aaf15f64ccb032852e4edbd1ad16c9/1?pq-origsite=gscholar&cbl=25782) (visited on 03/15/2023).
- 922 [5] Y. Roh, G. Heo, and S. E. Whang, “A survey on data collection for machine learning: A big  
923 data - ai integration perspective,” *IEEE Transactions on Knowledge and Data Engineering*,  
924 vol. 33, no. 4, pp. 1328–1347, 2021, ISSN: 1041-4347. DOI: [10.1109/TKDE.2019.294  
925 6162](https://doi.org/10.1109/TKDE.2019.2946162).
- 926 [6] J.-M. Rodriguez, “Aktuelle Softwareplattformen für Forschungsdatenrepositorien auf  
927 dem Prüfstand,” *ABI Technik*, vol. 39, no. 1, pp. 23–25, 2019, ISSN: 0720-6763. DOI:  
928 [10.1515/abitech-2019-1004](https://doi.org/10.1515/abitech-2019-1004).
- 929 [7] E. Böker. “Was ist Forschungsdatenmanagement? | Informieren und Planen | Themen  
930 | Forschungsdaten und Forschungsdatenmanagement.” de. (2023), [Online]. Available:  
931 [https://forschungsdaten.info/themen/informieren-und-planen/was-ist-  
932 forschungsdatenmanagement/](https://forschungsdaten.info/themen/informieren-und-planen/was-ist-forschungsdatenmanagement/) (visited on 02/15/2023).
- 933 [8] forschungsdaten.info. “Glossar | Praxis kompakt | Forschungsdaten und Forschungsdaten-  
934 management.” de. (2023), [Online]. Available: [https://forschungsdaten.info/pr  
935 axis-kompakt/glossar/#c269821](https://forschungsdaten.info/praxis-kompakt/glossar/#c269821) (visited on 02/15/2023).
- 936 [9] R. H. Schmitt, V. Anthofer, S. Auer, *et al.*, *Nfdi4ing - the national research data infras-  
937 tructure for engineering sciences*, 2020. DOI: [10.5281/ZENODO.4015201](https://doi.org/10.5281/ZENODO.4015201).
- 938 [10] M. D. Wilkinson, M. Dumontier, I. J. J. Aalbersberg, *et al.*, “The fair guiding principles  
939 for scientific data management and stewardship,” eng, *Scientific Data*, vol. 3, no. 1,  
940 p. 160 018, 2016, Journal Article, ISSN: 2052-4463. DOI: [10.1038/sdata.2016.18  
941 eprint: 26978244](https://doi.org/10.1038/sdata.2016.18).

- 942 [11] T. Todorova, M. Garvanova, M. Peteva, and V. Avramova, “Comparative findings from data  
943 literacy survey in three bulgarian universities,” in *EDULEARN19 Proceedings*, L. Gómez  
944 Chova, A. López Martínez, and I. Candel Torres, Eds., ser. EDULEARN Proceedings,  
945 IATED, 2019, pp. 932–940. DOI: [10.21125/edulearn.2019.0304](https://doi.org/10.21125/edulearn.2019.0304).
- 946 [12] J. Kaari, “Researchers at arab universities hold positive views on research data management  
947 and data sharing,” *Evidence Based Library and Information Practice*, vol. 15, no. 2,  
948 pp. 168–170, 2020. DOI: [10.18438/eblip29746](https://doi.org/10.18438/eblip29746).
- 949 [13] A. M. Elsayed and E. I. Saleh, “Research data management and sharing among researchers  
950 in arab universities: An exploratory study,” *IFLA Journal*, vol. 44, no. 4, pp. 281–299,  
951 2018, ISSN: 0340-0352. DOI: [10.1177/0340035218785196](https://doi.org/10.1177/0340035218785196).
- 952 [14] K. L. Wilms, S. Stieglitz, B. Ross, and C. Meske, “A value-based perspective on supporting  
953 and hindering factors for research data management,” *International Journal of Information*  
954 *Management*, vol. 54, p. 102 174, 2020, ISSN: 02684012. DOI: [10.1016/j.ijinfomgt](https://doi.org/10.1016/j.ijinfomgt.2020.102174)  
955 [.2020.102174](https://doi.org/10.1016/j.ijinfomgt.2020.102174).
- 956 [15] Mattias Björnmalm, Federica Cappelluti, Alastair Dunning, *et al.*, *Advancing research data*  
957 *management in universities of science and technology*, 2020. DOI: [10.5281/ZENODO.36](https://doi.org/10.5281/ZENODO.3665372)  
958 [65372](https://doi.org/10.5281/ZENODO.3665372).
- 959 [16] L. Costanzo and A. Cooper, *Setting the foundations for stronger partnerships and col-*  
960 *laborations for developing institutional rdm strategies in canada*, 2023. DOI: [10.5281](https://doi.org/10.5281/ZENODO.8010869)  
961 [/ZENODO.8010869](https://doi.org/10.5281/ZENODO.8010869).
- 962 [17] T. Austin, K. Bei, T. Efthymiadis, and E. Koumoulos, “Lessons learnt from engineering  
963 science projects participating in the horizon 2020 open research data pilot,” *Data*, vol. 6,  
964 no. 9, p. 96, 2021. DOI: [10.3390/data6090096](https://doi.org/10.3390/data6090096).
- 965 [18] J. A. Borghi and A. E. van Gulick, “Data management and sharing in neuroimaging:  
966 Practices and perceptions of mri researchers,” *PloS one*, vol. 13, no. 7, e0200562, 2018.  
967 DOI: [10.1371/journal.pone.0200562](https://doi.org/10.1371/journal.pone.0200562).
- 968 [19] H. Israel and M. M. Becker, “What does data stewardship mean in physics?,” 2023. DOI:  
969 [10.17192/bfdm.2023.2.8570](https://doi.org/10.17192/bfdm.2023.2.8570).
- 970 [20] A. Wuchner, M. Robrecht, P. Kehl, and H. R. Schmitt, “Challenges in publishing research  
971 data – a fraunhofer case study: Ing.grid preprint,” 2023. [Online]. Available: [https://p](https://preprints.inggrid.org/repository/view/20/)  
972 [reprints.inggrid.org/repository/view/20/](https://preprints.inggrid.org/repository/view/20/) (visited on 10/31/2023).
- 973 [21] A. Palsdottir, “Data literacy and management of research data – a prerequisite for the  
974 sharing of research data,” *Aslib Journal of Information Management*, vol. 73, no. 2,  
975 pp. 322–341, 2021, ISSN: 2050-3806. DOI: [10.1108/AJIM-04-2020-0110](https://doi.org/10.1108/AJIM-04-2020-0110).
- 976 [22] D. Ortloff, S. Anger, and M. Schellenberger, “An empirical study of the state of research  
977 data management in the semiconductor manufacturing industry: An analysis of industry  
978 and research institutions in the idev4.0 project: Ing.grid preprint,” vol. 2023, 2023. [Online].  
979 Available: <https://preprints.inggrid.org/repository/view/19/> (visited on  
980 10/20/2023).

- 981 [23] W. D. Chawinga and S. Zinn, “Global perspectives of research data sharing: A systematic  
982 literature review,” *Library & Information Science Research*, vol. 41, no. 2, pp. 109–122,  
983 2019, ISSN: 07408188. DOI: [10.1016/j.lisr.2019.04.004](https://doi.org/10.1016/j.lisr.2019.04.004).
- 984 [24] M. Cheung, *Data culture in canada: Perceptions and practice across the disciplines*, 2021.  
985 DOI: [10.5281/ZENODO.6760995](https://doi.org/10.5281/ZENODO.6760995).
- 986 [25] P. Vilar and V. Zabukovec, “Research data management and research data literacy in  
987 slovenian science,” *Journal of Documentation*, vol. 75, no. 1, pp. 24–43, 2019, ISSN:  
988 0022-0418. DOI: [10.1108/JD-03-2018-0042](https://doi.org/10.1108/JD-03-2018-0042).
- 989 [26] Statistisches Bundesamt, *Bildung und kultur, Personal an hochschulen*, Fachserie 11 Reihe  
990 4.4, Statistisches Bundesamt, 2022. [Online]. Available: <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Publikationen/Downloads-Hochschulen/personal-hochschulen-2110440217004.pdf> (visited on 06/24/2024).
- 994 [27] WGP. “Über uns - wgp.” (2022), [Online]. Available: <https://wgp.de/de/ueber-uns/>  
995 (visited on 01/02/2023).
- 996 [28] WiGeP. “Über uns – WiGeP – Wissenschaftliche Gesellschaft für Produktentwicklung.”  
997 (2022), [Online]. Available: <https://wigepe.de/wigepe/#wigepe> (visited on 01/02/2023).
- 998 [29] Fraunhofer. “Fraunhofer-Verbund Produktion.” Copyright: Copyright. (2022), [Online].  
999 Available: <https://www.produktion.fraunhofer.de/> (visited on 01/02/2023).
- 1000 [30] E. Internet of Production. “”The World becomes a Lab” - die Vision des IoP - RWTH  
1001 AACHEN UNIVERSITY Exzellenzcluster Internet of Production - Deutsch.” (2022),  
1002 [Online]. Available: <https://www.iop.rwth-aachen.de/cms/Produktionstechnik/Das-Exzellenzcluster/~rgqb/Leitbild/>  
1003 (visited on 01/02/2023).
- 1004 [31] N. Schmidtke, *Anzahl der mitarbeitenden im fraunhofer-verbund produktion*, T. Hamann,  
1005 collab., Telefonat, Jun. 28, 2024.
- 1006 [32] forschungsdaten.info. “Datenlebenszyklus | Informieren und Planen | Themen | Forschungs-  
1007 daten und Forschungsdatenmanagement.” de. (2022), [Online]. Available: <https://forschungsdaten.info/themen/informieren-und-planen/datenlebenszyklus/>  
1008 (visited on 11/29/2022).
- 1010 [33] Deutsche Forschungsgemeinschaft, “Guidelines for Safeguarding Good Research Practice.  
1011 Code of Conduct,” de, 2022. DOI: [10.5281/ZENODO.6472827](https://doi.org/10.5281/ZENODO.6472827).
- 1012 [34] L. Fahrmeir, C. Heumann, R. Künstler, I. Pigeot, and G. Tutz, *Statistik, Der Weg zur*  
1013 *Datenanalyse* (Springer-Lehrbuch), ger, 8., überarbeitete und ergänzte Auflage. Berlin and  
1014 Heidelberg: Springer Spektrum, 2016, 581 pp., Fahrmeir, Ludwig (VerfasserIn) Heumann,  
1015 Christian (VerfasserIn) Künstler, Rita (VerfasserIn) Pigeot, Iris (VerfasserIn) Tutz, Gerhard  
1016 (VerfasserIn), ISBN: 978-3-662-50371-3. DOI: [10.1007/978-3-662-50372-0](https://doi.org/10.1007/978-3-662-50372-0).
- 1017 [35] DFG. “Dfg - gepris - nfdi4ing – national research data infrastructure for engineering  
1018 services.” (2022), [Online]. Available: <https://gepris.dfg.de/gepris/projekt/442146713?language=en>  
1019 (visited on 11/29/2022).

- 1020 [36] H. Mehrwald, *Das “Not Invented Here”-Syndrom in Forschung und Entwicklung*. Wies-  
1021 baden: Deutscher Universitätsverlag, 1999, ISBN: 978-3-8244-0483-4. DOI: [10.1007/9](https://doi.org/10.1007/978-3-663-08337-5)  
1022 [78-3-663-08337-5](https://doi.org/10.1007/978-3-663-08337-5).
- 1023 [37] M. J. Page, J. E. McKenzie, P. M. Bossuyt, *et al.*, “The prisma 2020 statement: An updated  
1024 guideline for reporting systematic reviews,” eng, *BMJ (Clinical research ed.)*, vol. 372,  
1025 n71, 2021. DOI: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71). eprint: [33782057](https://eprints.bmj.com/eprint/33782057).
- 1026 [38] fileinfo.com. “Spss portable file.” (2011), [Online]. Available: [https://fileinfo.com](https://fileinfo.com/extension/por)  
1027 [/extension/por](https://fileinfo.com/extension/por) (visited on 01/19/2023).