Received: 25 August 2020 | Revised: 20 April 2021 | Accepted: 29 April 2021
DOI: 10.1002/asi.24492

RESEARCH ARTICLE

JASIST WILEY

Between administration and research: Understanding data management practices in an institutional context

Stefan Reichmann[1] | **Thomas Klebel**[2] | **Ilire Hasani-Mavriqi**[3] | **Tony Ross-Hellauer**[1,2]

[1]Open and Reproducible Research Group, Institute of Interactive Systems and Data Science, Graz University of Technology, Graz, 8010, Austria

[2]Know Center GmbH, Graz, 8010, Austria

[3]RDM Team, Institute of Interactive Systems and Data Science, Graz University of Technology, Graz, Austria

Correspondence
Stefan Reichmann, Open and Reproducible Research Group, Institute of Interactive Systems and Data Science, Graz University of Technology, Inffeldgasse 16c/I, 8010 Graz, Austria.
Email: stefan.reichmann@tugraz.at

Funding information
Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie, Grant/Award Number: FAIR Data Austria

Abstract

Research Data Management (RDM) promises to make research outputs more transparent, findable, and reproducible. Strategies to streamline data management across disciplines are of key importance. This paper presents results of an institutional survey ($N = 258$) at a medium-sized Austrian university with a STEM focus, supplemented with interviews ($N = 18$), to give an overview of the state-of-play of RDM practices across faculties and disciplinary contexts. RDM services are on the rise but remain somewhat behind leading countries like the Netherlands and UK, showing only the beginnings of a culture attuned to RDM. There is considerable variation between faculties and institutes with respect to data amounts, complexity of data sets, data collection and analysis, and data archiving. Data sharing practices within fields tend to be inconsistent. RDM is predominantly regarded as an administrative task, to the detriment of considerations of good research practice. Problems with RDM fall in two categories: Generic problems transcend specific research interests, infrastructures, and departments while discipline-specific problems need a more targeted approach. The paper extends the state-of-the-art on RDM practices by combining in-depth qualitative material with quantified, detailed data about RDM practices and needs. The findings should be of interest to any comparable research institution with a similar agenda.

1 | INTRODUCTION: DATA MANAGEMENT AS THE NEW FRONTIER IN RESEARCH ADMINISTRATION

Recent years have seen rapid growth in the adoption of Research Data Management (RDM) policies across research institutions. Accompanying these developments, there is now a quickly growing literature studying data management and sharing (Fecher, Friesike, & Hebing, 2015; Tenopir et al., 2011, 2015; Unal, Chowdhury, Kurbanoglu, & Boustany, 2019), uptake (Fuhr, 2019; Kalichman, Sweet, & Plemmons, 2015; Read, Larson, Gillespie, & So Young, 2019; Vilar & Zabukovec, 2019), regulation (Grant, 2015; Higman & Pinfield, 2015), and infrastructures (Amorim, Castro, da Silva, & Ribeiro, 2017; Bugaje & Chowdhury, 2018; Cox, Kennan, Lyon, & Pinfield, 2017; Knight, 2015) in order to understand the extent to which researchers need support in handling increasing amounts of research data (Awre, Baxter, Clifford, & Colclough, 2015; Van Tuyl & Michalek, 2015). Although data management is left to individual researchers in some cases (Van Tuyl & Michalek, 2015), research institutions are increasingly expected to implement policies governing, and integrated service to support, RDM (Higman & Pinfield, 2015). Assessing researchers' needs indirectly through surveys is a common strategy employed in developing services or policies (Akers & Doty, 2013). Following (Mancilla et al., 2019), we assumed that in order to (positively) influence RDM practices, we first need to understand current practices and needs. Three main strategies were adopted: (a) semistructured interviews with researchers across all faculties, (b) meetings with deans of faculties to explain the aims and approaches of the project and assemble an RDM policy working group, and (c) a survey about RDM practices. While the interviews and faculty meetings helped to build trust and develop in-depth understanding of RDM practices, the survey was designed to help in developing robust project goals. In informal discussions with deans as well as in-depth interviews with researchers, we have been struck by the recurring appearance of a number of issues: (a) "policy" was received badly, eliciting associations of bureaucracy and additional workloads for researchers, (b) problems of interoperability/reproducibility between generations of researchers (i.e., fluctuation of staff as one source of RDM problems), (c) costs and effort associated with RDM (generally to be paid for by departments). This paper focuses on understanding the breadth of RDM practices and needs. It presents the most salient findings of the faculty survey, in addition to providing context information from semi-structured interviews to address the following broad research questions:

1. What are faculty-specific RDM practices and needs?
2. What are the most salient dimensions of variation in RDM practices?
3. How and to what extent do the seven faculties differ in their RDM needs?
4. To what extent can RDM needs be generalized across faculties?

2 | BACKGROUND: WHAT DO WE KNOW ABOUT RDM?

RDM has received increasing attention from researchers as well as research funders and administrators in recent years. Commitment to the FAIR principles (to make data Findable, Accessible, Interoperable, and Reusable) (Wilkinson et al., 2016) is now a prerequisite for the acquisition of research funding, for example, from the - European Commission. Data Sharing has been associated with increased citation rates (Piwowar, Day, & Fridsma, 2007), economic growth (Tennant, Jacques, & Collister, 2016), increased transparency (Gilmore, Diaz, Wyble, & Yarkoni, 2017) and reproducibility (Toelch & Ostwald, 2018), and better and more efficient science (Leonelli, Spichtinger, & Prainsack, 2015). Previous work on RDM practices is built around a variety of approaches (Van Tuyl & Michalek, 2015), most of which rely on self-reports (Perrier, Erik, Patricia Ayala, & Dearborn, 2017). We found institution-wide surveys (Akers & Doty, 2013; Mancilla et al., 2019; Whitmire, Boock, & Sutton, 2015), national surveys across institutions (Aydinoglu, Dogan, & Taskin, 2017), surveys focusing on specific sub-groups within the university (Steinhart, Chen, Arguillas, Dietrich, & Kramer, 2012) or on specific disciplines (Geskin et al., 2015; Hsu, Martin, McElroy, & Litwin-Miller, 2015) as well as in-depth interviews or a combination of both (Anderson, Sally Lee, Scott Brockenbrough, & Minie, 2007; Myneni et al., 2016). The bulk of work on RDM practices relies on qualitative (survey and interviews) methods due to the difficulties associated with observing data practices, for example, in a laboratory situation. The downside of surveys is that they struggle to confer context information important to respondents but unthought of by investigators. Through triangulation of a survey and interviews, we sought to produce a fuller picture of the issues and needs with respect to RDM.

Research aims and overall research culture are important variables in RDM (Borgman, 2012). However, these and other context factors are oftentimes considered only to the extent that departmental/disciplinary differences in RDM are made the object of case studies. Problems surrounding RDM are predominantly conceived in technical terms (i.e., to the extent that they can be made the target of interventions). Kurata, Matsubayashi, and Mine (2017) lament that too few studies concern themselves with a definition of "research data," which effectively renders comparing results across cases difficult. Lack of appreciation of these dimensions is due, in part, to the exceeding difficulties in accounting for cultural variations with surveys/interviews as opposed to extensive field work.[1] Notable exceptions are the works of Christine Borgman (2010, 2015) and Sabina Leonelli (2016). Based on extensive ethnographic studies of data interoperability,[2] Borgman (2012) stresses the importance of context and shows how the characteristics of data change with the research questions/methods. Leonelli studied the work of data curators in biology needed to render data shareable across laboratories and research contexts (Leonelli, 2016). In addition, Borgman (2012) maintains that "data" is more of an umbrella term covering a broad range of meanings and referring to a diverse range of objects, and that correspondingly, "no social framework for data exists that is comparable to that for publishing" (Borgman, 2010, p. xviii).

The bulk of work on RDM practices relates to the sciences (Perrier et al., 2017; but see Schöpfel & Prost, 2016). Many studies find a lack of awareness of RDM topics (Chen & Ming, 2017; Mancilla et al., 2019), and of technical skills (Aydinoglu et al., 2017) and institutional guidelines (Chigwada, Chiparausha, & Kasiroori, 2017; Cox & Williamson, 2015), and in general a huge variability in data management practices, infrastructures, and approaches within and between disciplines (Akers & Doty, 2013; Borghi & Van Gulick, 2018), for example, with respect to the kinds and amounts of data collected (Borghi & Van Gulick, 2018; Cox & Williamson, 2015; Whitmire et al., 2015). "Big Science" projects, for example, in astronomy, genomics, or high-energy physics, collect more sophisticated kinds of data (well curated, with a high degree of automation) than "Little Science" disciplines. The latter account for the bulk of scientific data collection, but the resulting data sets are seldom preserved or shared (Wallis, Rolando, & Borgman, 2013). Big Science projects exhibit a high degree of consensus and international collaboration, usually make use of large-scale infrastructures, and are data- and computation-intensive, whereas Little Science is small-scale, with heterogeneous methods/data, and requiring less infrastructure (Borgman, 2010). Consequently, researchers in Big Science are most familiar with RDM practices and are most likely to share their data outside of their groups (Akers & Doty, 2013). Departments tend to give priority to using their own data storage infrastructures over institutional (university-wide) infrastructures (Steinhart et al., 2012; Whitmire et al., 2015).

Combining survey and in-depth interviews, we describe data management practices and needs of researchers at an internationally recognized technical (research and teaching) university which has recently embarked upon a project to introduce RDM policies and services and to create workable data management policies and training for the university as a whole as well as for each of its seven faculties. The novelty of the approach has been to involve all stakeholders (faculties, library, research funding administrators) in a researcher-led, iterative process of knowledge-gathering and co-creation to ensure that researchers' practices and needs are reflected in the high-level policies to begin with.

One important downside of empirical work on RDM practices is that, as has been said above, it seems to be based on an intuitive understanding of data (Kurata et al., 2017) and disciplines. Researchers in our sample predominantly regard RDM as an administrative task. However, focusing on RDM exclusively as a top-down, policy-driven administrative activity, rather than as an organic and integral part of research which must take account of the very diverse demands of specific research contexts, runs the risk of reproducing the administrative structure of the institutions studied and remaining blind to variations in data practices that transcend traditional categories ("discipline," "research field," etc.) (Leonelli, 2020, p. 5). By framing it as a research activity, we can see how RDM transcends boundaries of disciplines. We extend the state-of-the art literature on RDM practices by combining in-depth qualitative material on RDM and data practices with a survey of research staff, yielding quantified, detailed data about RDM practices and needs. The findings should be of interest to any comparable research institution with a similar agenda.

3 | APPROACH: TRIANGULATION OF SURVEY AND INTERVIEWS

3.1 | Rationale

We combined large-scale quantitative survey data with in-depth interviews. While surveys allow for large populations to answer a small number of standardized questions but usually lack breadth of coverage, interviews target a significantly smaller group to answer questions in depth but usually do not allow for broad coverage. Combining the two methods allows for their strengths to be meaningfully and mutually enhanced, both at the level of developing the research topic (initial interviews and faculty visits helped to gauge the state-of-play) and at the level of developing the survey questions.

All data were collected at a university with a STEM focus that features 1,784 researchers (2019) in 97 institutes across seven faculties, 5 (predominantly) engineering, one (predominantly) scientific, and one (Architecture) (predominantly) humanities. It offers 18 undergraduate degrees, 33 graduate degrees, as well as 2 doctoral degrees. In 2019, 13,566 students were enrolled. The university has an annual budget of EUR 244 million (EUR 79.2 million thereof is third-party-funding) and an annual research output of approximately 2,280 publications (2019).[3] In 2018, a programme coordinated by the Vice rector for digitization and change management was launched with the aim of streamlining digitization across research, teaching, administration, and third mission. RDM is part of this broader agenda and includes the development of a repository for research data as well as training in RDM.

3.2 | Survey instrument, questions, and sample

Based on preliminary interview findings and individual stakeholder consultation, a survey was designed to understand how research data size, data handling, and research styles influence RDM practices. We used LimeSurvey to administer the questionnaire which was sent out via email to all members of scientific staff between September and October 2019. The survey was kept open for 5 weeks. Two reminders were sent, after 2 weeks and before the end of the survey. Consultations were held with responsible bodies at the university to ensure that established protocols were met. All aspects of this research involving human subjects were developed in line with data protection policies. The survey consisted of 41 questions in five sections:

- Data types (8 questions): typical types of data generated, data formats, other research outputs.
- Data quantity (4 questions): typical data amounts per year (per project, per year, required storage space).
- Data handling routines (9 questions): practices of data handling, sharing, reuse, and use of repositories.
- Obstacles to research data management (17 questions): researchers' experiences with RDM including barriers.
- Demographics (3 questions): faculty, position, and role of the respondent (in line with data protection policies).

The survey items were adapted from a survey of ERC grantees' use of Open Science practices (PPMI 2018), as well as from preliminary insights from faculty visits and interviews. The survey went through several rounds of refinement, including pretesting via 10 faculty representatives. In line with policies at the university, the survey was distributed via a mailing list containing all scientific staff active at the time of the survey. This mailing list defined the following target group:

- Assistant/Associate/Contract/Full Professors
- Contract/Senior Lecturers
- Contract/Senior/Project-Senior Scientists
- University Assistants (PhD/PostDoc)
- University-Project Assistants (PhD/PostDoc)

3.3 | Interviews (approach, participants, analysis)

Eighteen semistructured interviews were conducted with respondents occupying diverse roles within their departments after 37 researchers were contacted between April and July 2019 (purposive sampling). The interviews were designed to aid in developing the survey questions and to acquire contextual information on RDM practices from a subset of the population. Informants were selected in a - three-stage process. Faculty deans were approached as gatekeepers. Project aims were introduced, and faculty heads were asked to support the research, for example, by providing contact details of possible informants. In addition, informants were selected using public staff lists. Departments within the faculties were selected with a view towards relative balance between research aims (e.g., between scientific and engineering fields) and interest in data management. At a third stage, potential informants were selected within the departments. The first part of each interview was devoted to understanding the respondents' research field and specific research aims in depth, how research aims relate to specific conceptions of the phenomena in question (Bogen & Woodward, 1988). Respondents were asked to describe their research work and to elaborate on the role of data in their research. All interviews were conducted in German and lasted between 45 and 90 min, yielding ~700 min of recorded material. All interviews were recorded and fully transcribed. Interview transcripts were coded in two rounds using the qualitative analysis software package RQDA (Huang, 2016) and making extensive use of theoretical memos (Timmermans & Tavory, 2012). For the purposes of presentation, individual quotations were translated into English by the corresponding author.

4 | RESULTS

4.1 | Survey response rate and respondent demographics

The survey was sent out to 1,784 scientific staff members in September 2019. A total of 498 respondents started the survey and 258 completed the questionnaire (response rate: 14.5%). Only these were included in the analysis.[4] Respondents came from all seven faculties and from all academic positions (Table 1). Ninety-three respondents identified themselves as (full, associate or assistant) professors, 9 self-identified as lecturers, 1 Senior Scientist, 40 PostDocs, 43 researchers, 93 PhD candidates, 1 university assistant, and 7 students. The categories are not exclusive. Response rate (Figure 1) was highest at the Faculty for Mathematics, Physics and Geodesy (23.3%), followed by Electrical and Information engineering (16.8%) and Computer Science and Biomedical Engineering (14.6%). However, responses by faculty do not differ markedly in absolute numbers, the Faculty for Architecture being the only exception, where only eight people

TABLE 1 Education of respondents (highest completed degree) ($N = 258$)

Faculty	Highest completed degree Doctorate/PhD	Master's degree (MSc., MA, Dipl. Ing., Mag.)	Bachelor's degree	Other
Architecture	4	4	0	0
Civil Engineering Sciences	13	15	1	0
Computer Science and Biomedical Engineering	27	11	4	0
Electrical and Information Engineering	20	21	0	0
Mathematics, Physics, and Geodesy	34	17	0	0
Mechanical Engineering and Economic Sciences	22	19	1	2
Technical Chemistry, Chemical and Process Engineering, Biotechnology	34	9	0	0
Total	154	96	6	2

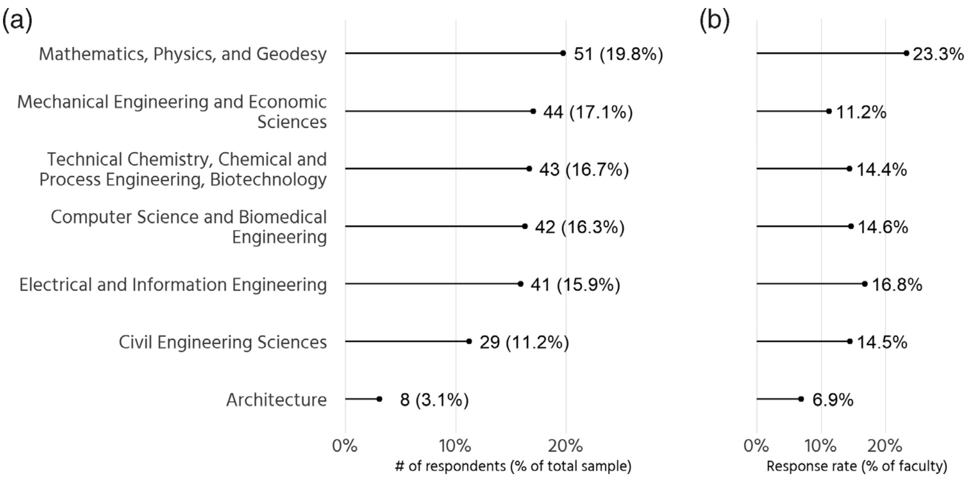

FIGURE 1 Respondents per Faculty. (A) Absolute and relative number of respondents per faculty. (B) Response rate per faculty (calculated as share of the total number of scientific staff)

(6.9% faculty response rate) responded.[5] In total, 154 respondents (59%) held a doctorate, and 96 (37%) held a master's degree or equivalent (Table 1). The following sections present descriptive results on data production and data storage as well as opinions on data sharing and obstacles to RDM. Approval and rejection have been operationalized by combining the two categories on either point of the Likert scales used ("*Strongly agree/tend to agree*" and "*Tend to disagree/disagree*," respectively).

4.2 | Data production

Data amount was operationalized as "data collected in a (typical) research project" and as "data collected per year" since there are no official numbers on how much research data is produced at the university (Figure 2). Data sets come in many different sizes. We found no systematic variation between faculties ($\chi^2 = 32.9$, $p = .325$). With the exception of architecture, faculties are similar with respect to types of datasets commonly used.[6] 69.7% of researchers at the Faculty for Technical Chemistry work with datasets below 10 GB, while only 9.4% work with datasets >100 GB. Electrical and Information engineers work with datasets across the entire domain specified, both very small and very large. At the Faculty for Computer Science and Biomedical Engineering the degree of variation is most salient. Researchers work with a wide variety of data amounts, from very small (<1 MB, 2.4%) to very large (>100 GB, 16.6%). The Faculty for Mathematics, Physics, and Geodesy is striking in this respect, as more than a quarter (27.4%) of respondents work with datasets >100 GB. The fraction of researchers working with datasets >500 GB is also largest here (19.6%). Out of the seven faculties, only three feature researchers who do not work with datasets in the <1 MB category. Mathematics, Physics, and Geodesy is featured

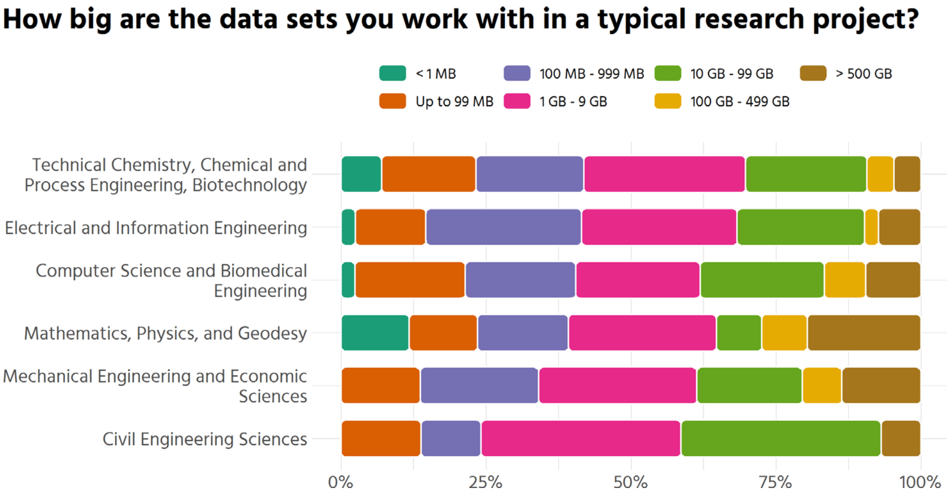

FIGURE 2 Data sizes in a typical research project by faculty ($n = 250$). This figure depicts the (estimated) average data sizes researchers work with in typical research projects. This includes all data used as input to or output of research

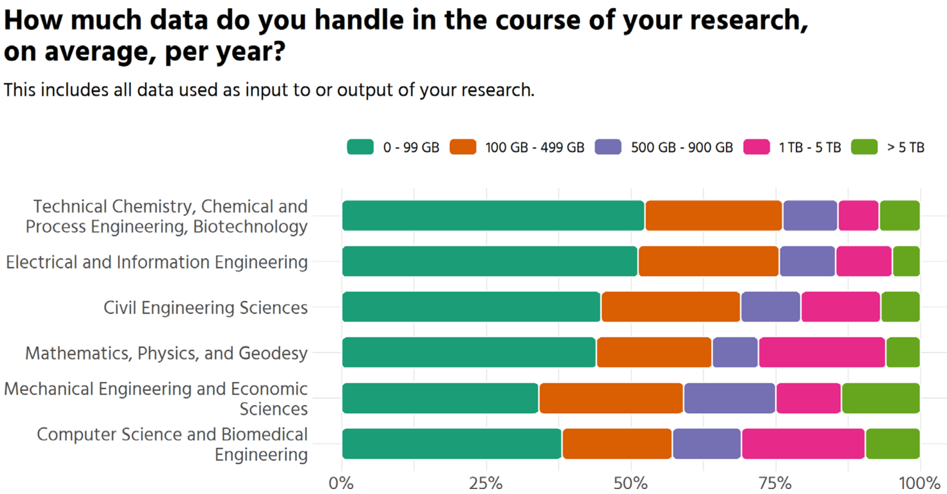

FIGURE 3 Data per year by faculty ($n = 248$). How much data researchers handle in the course of their research, on overage, per year. This includes all data used as input to or output of research

in this category (23.6% use datasets below 99 MB). For Mechanical Engineering and Economic Sciences no respondents use datasets below 1 MB, with 13.6% of respondents' datasets up to 99 MB, 34% up to 999 MB, and 60.4% up to 9 GB. In total, 18.2% of Mechanical Engineers use datasets above 10 GB but below 99 GB, and 20.4% work with datasets above 100 GB. In the Civil Engineering Faculty, the 100–499 GB category remains empty; 92% of researchers use datasets of up to 99 GB, and 6.9% use datasets above 500 GB.

Researchers were asked to estimate the amount of data handled per year (Figure 3). In total, 68.6% of researchers handle less than 500 GB of data per year, the remaining 37.4% handle more than 500 GB. This percentage is roughly the same for Technical Chemistry and Process Engineering and Electrical and Information Engineering where 76.2% of researchers handle a total averaging below 500 GB of data. The percentage of researchers who handle data sets >5 TB is small across all faculties, with the highest percentage (13.6%) found in Mechanical Engineering.

4.3 | Data storage

Researchers were asked to indicate how much of their data needs to be stored in the short (<3 years) and long term (>3 years). Three years were used because this period best represents the duration of many third-party funded research projects. "Short-term storage" was taken to imply that data may be deleted after 3 years. Only a tiny fraction of respondents (2%) needs to store more than 100 TB of data in the long term. Just under 25% of respondents have less than 10 GB of data needing to be stored in the short term. 25% of respondents hold between 10 and 99 GB needing to be stored for less than 3 years. The fraction of respondents needing to store more than 10 TB is small, both for short- and long-term storage.

For "long-term storage" and faculty, the need for long-term storage varies systematically by faculty ($\chi^2 = 41.2$, $p = .022$). With the exception of Civil Engineering (6.9%) no respondents said they need to archive >100 TB of data (Figure 4). Electrical Engineering (52.5%) and Technical Chemistry (40.5%) archive data sets between 10 and 99 GB, more than any other faculty in this category. The size of typical data sets is related to data archiving needs (long-term storage) ($N = 238$; $\chi^2 = 90$, $p < .001$). The frequencies of "Long-term Data Storage" are presented in Figure 4.

"Short-term storage" refers to data storage for the duration of a research project, that is, so long as researchers can be expected to work actively with that data (while it is clearly possible to revisit data after a research project ends, this scenario seems less likely). Short-term storage needs are associated with the amounts of data researchers work with in typical research projects (Figure 5).

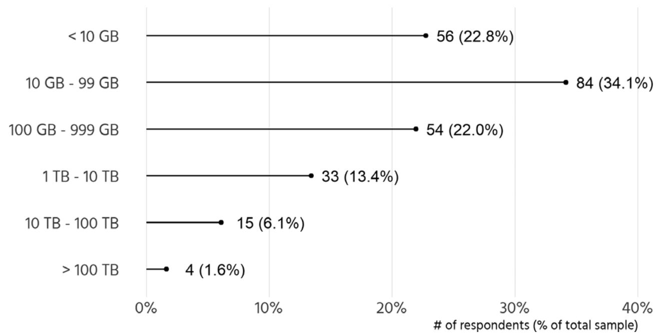

FIGURE 4 Long-term data storage: Average size of data sets respondents need to archive (store in the long term) – percentage of respondents ($N = 258$)

Researchers were asked to indicate how they store data in the short term, bearing in mind that multiple storage devices may be used. There are huge differences between using local hard drives as opposed to using cloud services, whether hosted by the university or by a third party. Seventy-six percent of respondents store data on local hard drives during a project, compared to 27% who store data on cloud services provided by the university. 12.5% use third-party cloud services to store data. 68.4% store their data on institutional servers.

4.4 | Data sharing

Pressure to release research data for the use by others comes from various sources (Borgman, 2012, p. 1066): funder mandates, policy bodies, journal publishers, and researchers themselves. There exist a variety of rationales for sharing research data (Borgman, 2012, p. 1067): Reproduction/verification of research, availability of publicly-funded research, enabling others to ask new questions, and advancement of research. We found that when discussing data sharing practices, interviewees clearly distinguished between publicly and privately funded research. In general, respondents support data sharing: 90% agree that "sharing data enables better research" ($N = 232$) (Figure 6). 82.2% agree that "sharing data with peers is encouraged" in their discipline (Figure 7).

Despite the high standing of data sharing, 42.2% said that "sharing research data is not a priority" ($N = 223$)—perhaps reflective of the fact that 48.1% of respondents ($N = 212$) said that "sharing data is mandatory for most important journals" (Figure 8). Conversely, 90% of respondents agree with the view that "sharing data

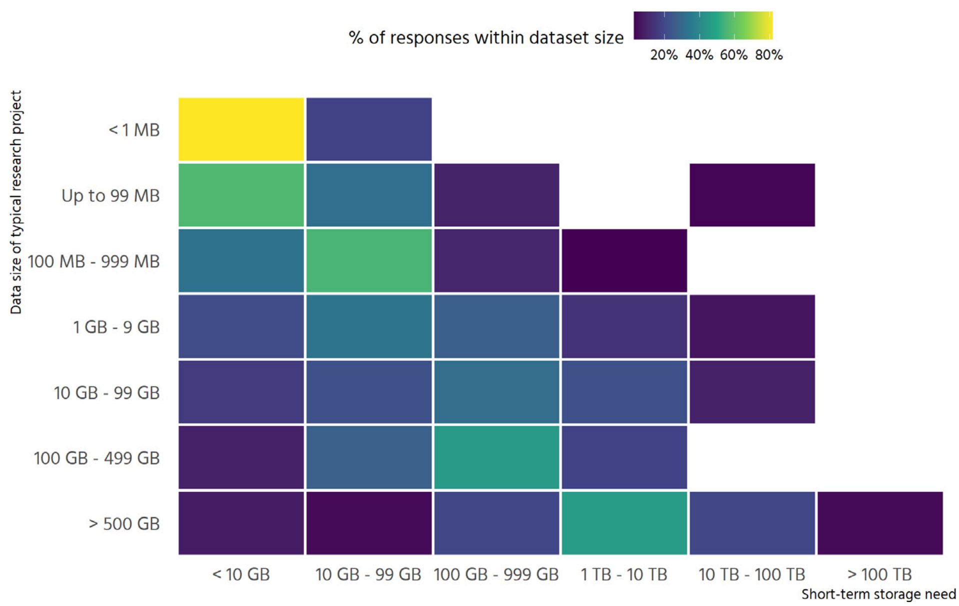

FIGURE 5 Short-term data storage by data size: How much data needs to be stored in the short term displayed by typical data set ($N = 233$)

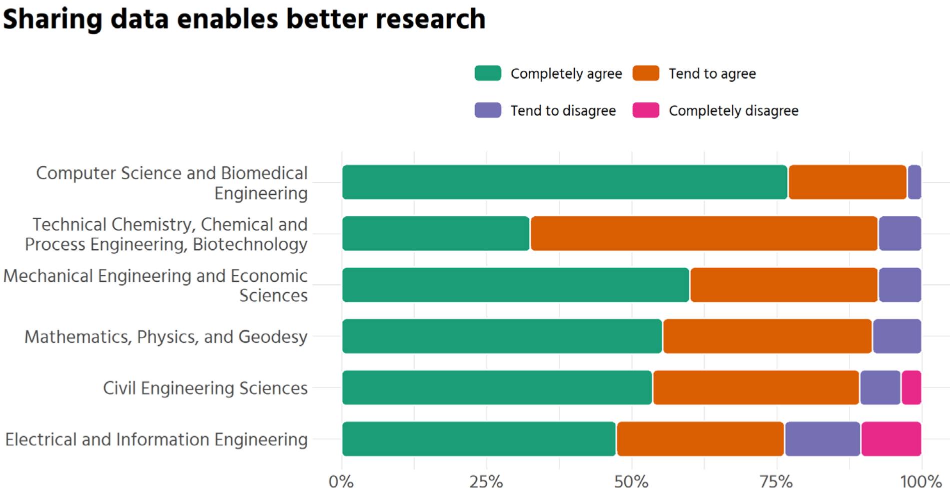

FIGURE 6 Agreement/disagreement with the statement "Sharing data enables better research" ($N = 232$)

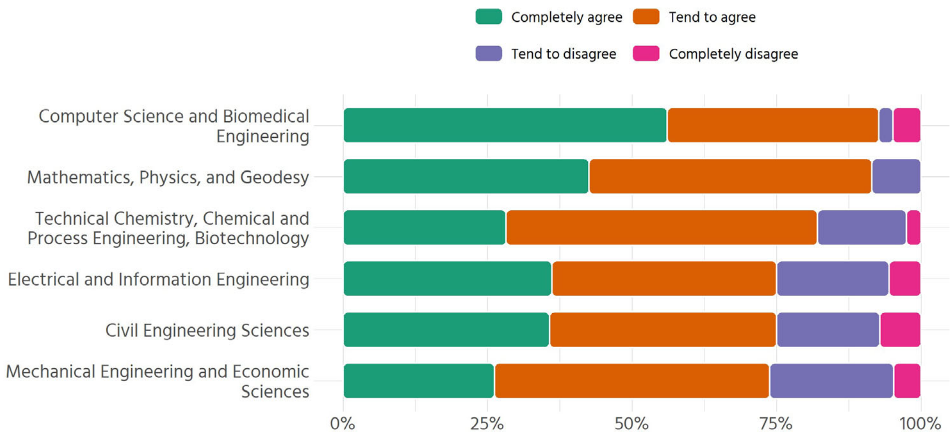

FIGURE 7 Agreement/disagreement with the statement "Within my discipline, data sharing is encouraged" ($N = 233$)

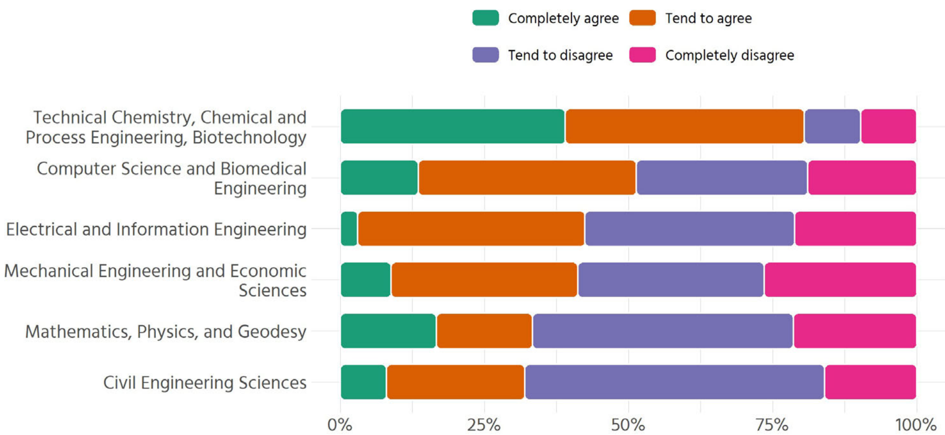

FIGURE 8 "When you think about your discipline, would you say that sharing data is mandatory for most important journals?" ($N = 212$)

enables better research" ($N = 232$, $\chi^2 = 36.5$, $p = .001$ for "Faculty"). Establishing a connection between disciplines' endorsement of data sharing ("sharing research data with peers is encouraged") and respondents' faculty was unsuccessful ($\chi^2 = 19.56$, $p = .190$). Sharing research data is challenging not just in a technical sense, but also as a matter of organization. Data sharing does not seem to be common, but there are considerable differences between faculties. While 27.1% ($N = 240$) of respondents share data as Data S1, Supporting Information to a publication, 28.9% ($N = 239$) of respondents share data to institutional repositories. This is striking in terms of governance.[7]

More effort may need to be invested in providing institutional data storage services or making them more attractive:

> Well, there you'd have to [give rights to] external people. I considered doing this over SharePoint [collaboration software], but I have to admit [...] I'm not very fond of it, they're not very logical to handle, I don't like them. I keep having [this problem] with [company], with whom we do a lot of projects and where everything is done via share points, where we upload and download stuff, but it's not really like a server that can be searched with an explorer. (Civil Engineering Sciences)

Noninstitutional repositories are less widespread (21.3%). More than a quarter (27.1%) of respondents rarely or never share their data as an appendix to a publication. Standalone data publications are the least popular way to share data for researchers (10%). The Life Sciences are particularly interesting with respect to data sharing because there, researchers almost unanimously agree that sharing enables better research (92.7%, $N = 232$). The reason for high rates of assent is that researchers see how sharing has benefited their field:

> Well, I believe the community realized that it will profit [from making research data available] [...] communities differ and this one is very communicative I'd say. (Technical Chemistry, Chemical and Process Engineering, Biotechnology)

Not everyone sees value in making data available. Whether a practice is seen as conducive to the common good is not simply a matter of demonstrating its efficacy to a research community. On the other hand, researchers acknowledge that publicly funded research should be publicly available. However, there is a discrepancy between a global desire for accessibility and the desire to make one's own data accessible. The motivations for sharing data are fraught with researchers' and institutions' vested interests. One important aspect of these vested interests pertains to secrecy, a topic of fundamental importance at an institution that depends on industry funding:

> Well, they're in place, agreements to maintain secrecy that you sign as project lead, they are valid for all employees. And employees are aware of that. (Mechanical Engineering and Economic Sciences)

Respondents are ambivalent about data sharing, aware of their contractual obligations (per agreement, all data produced by employees belong to the university), and of the potential competitive advantage associated with keeping data secret. Where researchers compete on the same turf, there is a clear incentive to withhold data. Anderson et al. (2007) point out that even within communities, data sharing can be challenging due to the highly sophisticated, context-dependent nature of the research (see also Borgman, 2012, 2015). The value of sharing data is frequently not seen in its accessibility to the general public. Rather, data sharing is seen as forcing researchers to have proper data management structures enabling reuse and data security.

4.5 | Data reuse

Research data is being produced in new and innovative ways (Hine, 2016), but data reuse is not yet routine (Borgman, 2012; Higman & Pinfield, 2015; Wallis et al., 2013). Our survey instrument included two questions on data reuse: "During a project, how frequently do you/does your group reuse data from third parties?" and "When using third-party data, these are usually obtained through direct contact with data producers, data repositories, or supplementary files linked to publications?" Both items need a bit of explanation. Frequency of data reuse can be interpreted as exclusive, in the sense that the more researchers rely on data produced by others, the less data they generate themselves. Second, data reuse does not extend exclusively to researchers outside one's group/department. Data might be obtained from colleagues in one's group/department. Many interviewees regard data generation as the core business of research. Our results suggest that data reuse is increasingly prevalent: While 12.2% of respondents reuse third-party data always or most of the time, 35.8% sometimes reuse third-party data. However, almost half of all respondents (51.8%) rarely or never reuse third-party data (Figure 9). The proportion of researchers who reuse data from third parties at least "sometimes" is highest at the Faculty for Computer Science and Biomedical Engineering (65.9%). Since the faculty is heterogeneous, combining Life Sciences, Computer Science, and Engineering fields, the results are not uniform.

In Mechanical Engineering and Economic Sciences, 56.4% reuse data from third parties at least "sometimes." A significant proportion (20.5%) never reuse data. At the Faculty for Technical Chemistry, Process Engineering, and Biotechnology, only 17.1% of respondents reuse data from third parties. Conversely, 36.6% of respondents never reuse data from third parties. As for respondents from the Faculty for Mathematics, Physics, and Geodesy, only 12.2% frequently reuse data. The same is true for Civil Engineering Sciences, where even fewer respondents (11.5%) do. 53.8% "rarely" or "never" reuse data from third parties.

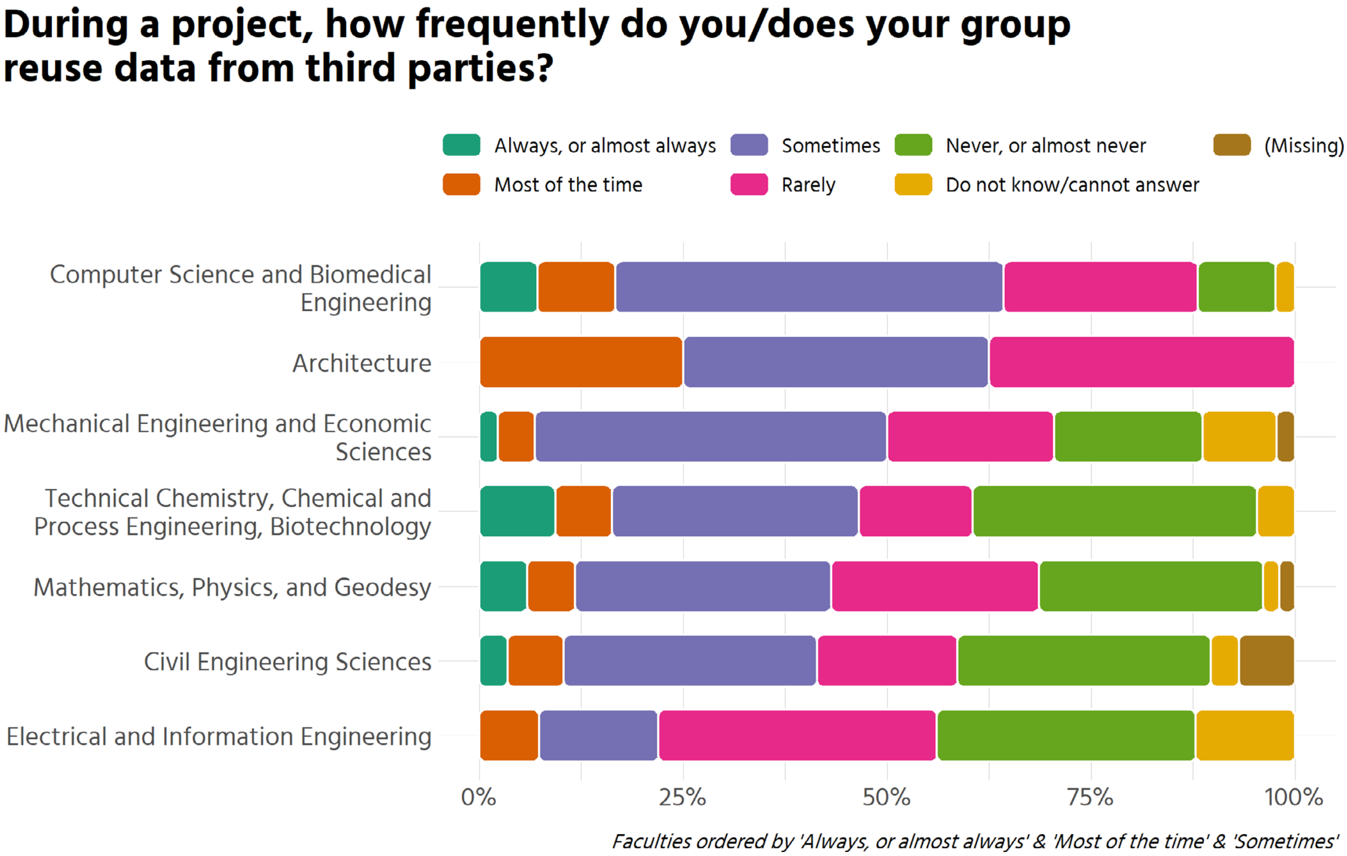

FIGURE 9 "During a project, how frequently do you/does your group reuse data from third parties?"

Within Electrical and Information Engineering, 8.3% of respondents reuse data at least most of the time, whereas 36.1% chose the "Never"-category:

> It [data reuse] happens, but very, very rarely. You ask someone [for their data] and most of the time this question forms the context. (Electrical Engineering)

The following excerpt contains possible reasons:

> You can hit people [with the sheer amount of] data. We can produce finite elements-data that don't mean anything, just data production. The scientific value is what we learn, a finished prototype or machine we built. Or, we did 37 measurement series. Then there are five measurement series that demonstrate what we actually wanted. And we don't need the rest. [...] Those are discarded. Because they pull focus from what is essential. [...] The result, a finished prototype design, or an illustration or model of a phenomenon. (Electrical Engineering)

While electrical engineers produce large amounts of data, the outcome of analysis is, in most cases, a prototype or model, not a figure. Accordingly, data are not valued in and of themselves. Thylstrup (2019) suggested to understand the culture of traceability surrounding Big Data in terms of waste. These excerpts suggest that there is another explanation: While some fields regard data as potential future evidence (Leonelli, 2016), others regard data as a means to an end; once the end (e.g., a prototype) is achieved, the data become a mere milestone on the way. These excerpts therefore testify to a bundle of fields ill-attuned to the benefits of data reuse.

4.6 | Sources of secondary data

Respondents who said they reuse third-party data were asked to choose between three possible (nonexclusive) sources of data for reuse:

- Data repositories or data platforms (discipline-specific, institutional, or otherwise).
- Data producers who would be able to provide data directly upon request.

- Data found through links in publications/as supplements to publications.

These three options were considered exhaustive in the sense that research data for specific studies can either be obtained by asking data producers directly, by following links in publications, or by searching dedicated repositories. At the Faculty for Computer Science and Biomedical Engineering, 44.2% reuse data from repositories ("Always" and "Most of the time"). The proportion of respondents who obtain data through repositories is second largest among respondents from the Faculty for Civil Engineering and Economic Sciences (25.1%). At the Faculty for Technical Chemistry, the proportion was 8%.

In total, 54.9% of researchers reuse data obtained from other researchers ("Always" and "Most of the time"). At the Faculty for Mechanical Engineering and Economic Sciences, 64.5% of respondents said they obtain data from data producers. The figure is lower at the Computer Science and Biomedical Engineering Faculty (24.3%). At the Faculty for Civil Engineering Sciences, 68.8% said they obtain data from data producers ("Always" and "Most of the time." At the Faculty for Technical Chemistry, 57.7% said they do so, in addition to 54.9% of respondents at the Faculty for Electrical and Information Engineering.

A χ^2-test was used to determine whether respondent faculty is connected to aspects of data reuse. We found that reuse of data obtained through publication supplements is more prevalent in some faculties than others ($N = 166$, $\chi^2 = 34.2$, $p = .025$). 31.3% of respondents reuse data from supplementary files ("Always" and "Most of the time"). The proportion of researchers who reuse data obtained through publication supplements is highest in the Faculty for Technical Chemistry, Process Engineering, and Biotechnology (42.3%). The figure is lower in the Faculty for Computer Science and Biomedical Engineering (41.7%). In the Faculty for Mathematics, Physics, and Geodesy, 27.3% of respondents reuse data from publication supplements. At the Faculty for Mechanical Engineering and Economic Sciences, this figure is 32.1%, with 26.7% at the Faculty for Civil Engineering Sciences, and 9% at the Faculty for Electrical and Information Engineering.

4.7 | Needs analysis

4.7.1 | Workload and lack of time

Emergent data management topics in the interviews included reproducibility and data security. Specifically, researchers were asked what they perceived to be obstacles to RDM. Time and effort required present a big obstacle to researchers (Anderson et al., 2007); Tenopir et al. (2011) found that lack of time was the top reason for not sharing data. Seventy percent regard increased time and effort associated with RDM as an obstacle at least "to some extent." These numbers are somewhat in tension with respondents' reservations that were found in the interviews and faculty visits. There, many said that the large amount of time it takes to produce data sets constitutes an incentive (rather than an obstacle) to preserve and curate the data:

> Especially the analysis needs a lot of time, many, many working hours and you never know whether you'll need them [the data] again some time, but it would be terrible if you wanted to look up something for a publication and only had access to the raw data, because then you'd have to start over. A month's worth of analysis, I reckon, which is why we want to store them. (Technical Chemistry, Chemical and Process Engineering, Biotechnology)

Respondents are ambivalent with respect to time. Where researchers recognize the benefit of sharing/reusing data, increased time requirements do not seem to be an issue. Respondents who believe that "sharing data enables better research" tend not to regard increased time associated with RDM as a problem. For example, a cumulated 58.4% of those who completely agree that "sharing data enables better research" do not believe that increased time is an obstacle to RDM.

4.7.2 | Employment contracts and turnover of the workforce

There is a recurring narrative that was encountered across almost all interviews: High turnover of scientific staff, even though desired by management, has adverse effects such as knowledge loss (Interview field notes, Architecture). This situation is pressing for all faculties as everywhere, fixed-term contract researchers (e.g., PhD candidates) shoulder the bulk of data collection. For department heads/PIs in the sample, high turnover was desirable because new staff is believed to bring in new ideas (Bennion & Locke, 2010). Staff turnover is a social problem recognized by respondents as requiring organizational solutions:

> We have a large quota of doctoral students [...] two theses per year, which means that I have a lot of turnover of assistants and project assistants, which makes research data management a very pressing problem. (Mechanical Engineering)

An important aspect of streamlining data management is documentation and communication between domain experts, data managers as well as their successors (see, e.g., Anderson et al., 2007). What is needed is "a concerted, focussed effort to build [...] a database with corresponding data entry masks, [...] and someone watching over people typing in their things correctly (Technical Chemistry) as '[staff turnover] is quite annoying, really, and my only way around that is good documentation and, if possible, to have people's contracts overlap' (Mechanical Engineering)." Further, "it is not at all easy to pass on the know-how from one PhD candidate to the next, because when one leaves you have a 6-month-break until the next one starts" (Electrical Engineering).

Staff turnover is an important but often overlooked dimension of RDM since it threatens to take away data control from researchers. Those responsible for data management have a vested interest in RDM since they are favored within the current distribution of responsibilities. The switch to a data stewards-model where, for example, each faculty employs RDM specialists to support researchers would take power away from them. At present, different coping mechanisms are in place, for example, to offer specific RDM training (based on the routines of the respective research group). Managing employment structures and managing data are thus two sides of the same coin:

> That's the same situation, which is why we attempt to, in my group we have certain standards regarding project structure where we mandate data structures for our projects, where it is stored, where code is stored, where data are stored, where results are stored. (Mechanical Engineering)

4.7.3 | Industrial research: Dealing with commercially sensitive data

Commercially sensitive data pose multiple problems. RDM provisions might conflict with existing Intellectual Property Rights (IPR) regulations, contract research agreements, or with the university's interests in patenting and commercialization of ideas. These issues are not relevant to all researchers, as not everyone is involved in industry projects. However, from meetings with deans we learned that around one third of third-party funding is provided by industry (a figure backed up by the university's official numbers). Industry projects come with individual data sharing agreements (often overseen by IPR/patenting experts at research services) which often mandate that industry partners retain control over research data, even while (some of) the project results are published in academic journals. In total, 48.8% of respondents ($N = 201$) said that dealing with confidentiality issues was an obstacle to RDM and sharing (cumulative of "to a large extent" and "to a very large extent"). Fewer (35.2%) are worried about their shared data being used commercially ($N = 210$). Industry projects thus have various repercussions for data sharing. One not-so-obvious consequence concerns the question whether findings of applied projects are publishable:

> The entire civil engineering faculty [...] is not so active when it comes to publishing in high [impact] journals, that is not considered important at our faculty. This is a disadvantage, in my opinion, but we put a lot [of effort] into contract research which is difficult to exploit for good journals afterwards [...] field trials are impossible to publish, there's nothing there. (Civil Engineering)

Another consequence has to do with ownership over the data:

> Of course, there are extreme cases, if you say, this is completely industry-financed, then results belong to industry, but even then, there's the question, once inventions are involved, they belong to the inventor. (Civil Engineering)

Industry data thus pose multiple obstacles for RDM, for example, with respect to external standards, embargoes, and relevance of industry data to basic research.

4.7.4 | Lack of standards and skills

Standards and skills are an issue when it comes to RDM. 46.9% of respondents ($N = 209$) regard missing standards as an obstacle to data management at least to some extent. Unclear responsibilities are a pressing issue across all faculties. Many fields lack common standards, for example, with respect to metadata, making it difficult for researchers to know what to do. Without shared norms, acquiring the respective RDM skills is also difficult. Indeed, respondents regard both a lack of skills ($N = 212$) and a lack of knowledge about data repositories ($N = 214$) as problematic (36.3% resp. 35.1%). The provision of training modules would therefore be a promising move. 31.7% of respondents ($N = 205$) think that possible misinterpretation poses an obstacle to data sharing. Here, further analysis of reasons is required. In most fields, there is currently no reward system for making data available. 28.1% of respondents ($N = 185$) regard a lack of recognition for data management and sharing activities as an obstacle to data sharing. Time and effort needed to curate and share data make lack of recognition problematic. While this is in principle attainable at an institutional level, it is markedly less clear how disciplinary reward systems could be influenced in such a way. Depending on the field, other problems arise with respect to data management and sharing, giving the issue aspects of a "wicked problem" (Awre et al., 2015). Where data collection depends on the availability of specific infrastructure, the problem of data sharing extends to communicating the parameters of experimental setups. As shown in seminal work by Collins (1985), the knowledge necessary for successfully replicating experiments is tacit to a large degree (see also Leonelli, 2016):

> A measurement setup is not like, I take my thermometer to check the temperature of the water in my water heater. Rather, you have a lot of sensors, and until you get the drive started. And it's not that easy to pass on the know-how from one doctoral student to the next, because when one leaves, you have a six-month-gap, that's the problem. (Electrical Engineering)

Respondents frequently lament lack of standards for communicating experiment parameters while acknowledging that RDM is time-consuming even with standards in place:

> Well for results it has to be clear that both the author as well as the supervisor are able to say after two years, those are the associated raw data. Whether they are directly linked or not, or whether there is a file somewhere that says, 'data for paper' or 'data for measurement' [...] I haven't asked anybody but I think as a researcher, you want to be able to prove the stuff you publish, if someone asks. (Electrical Engineering)

The above excerpts do not have clear implications as to where respective standards should come from. At the moment, the situation is difficult to pin down in terms of RDM principles.

5 | DISCUSSION: AMBIVALENCES OF RDM

5.1 | RDM within faculties: Commonalities and differences

While there was almost universal agreement about benefits of data sharing, this commitment is not always reflected in reported data sharing and reuse practices. Data reuse seems to be increasingly prevalent, although more in some faculties (e.g., Technical Chemistry) than others (Electrical Engineering). Data amounts and archiving needs vary across faculties, with a notable discrepancy between data amounts and storage needs in some cases (e.g., Electrical Engineering, Civil Engineering). This suggests that a significant proportion of data does not warrant archiving/sharing in the view of the respective disciplines. While some fields clearly depend on shared data and have developed a culture of sharing (e.g., the Life Sciences), others (e.g., Civil and Electrical Engineering) rarely make use of secondary data and do not share data as readily, even though large data amounts are produced. This discrepancy may be attributed to the purpose of data creation in a given field. While engineering data are predominantly produced for a specific model/prototype, basic sciences data are of a more general nature, thought to back up general knowledge claims about a given phenomenon (see, e.g., Leonelli, 2016).

5.2 | Resources, responsibility, and vested interests

Many institutions have started to develop RDM policies as a way of dealing with fluctuation (e.g., TU Delft[8]), and indeed, our findings suggest that researchers tend to emphasize the administrative aspects of RDM over potential epistemic advantages of data curation (Tenopir et al., 2011). This interpretation, in addition to a lack of standards, recognition, and skills, adds to a certain reluctance to share data. RDM is frequently interpreted as a bundle of technical solutions for what are ultimately organizational problems: Respondents lament that the person responsible for managing datasets is on a fixed-term contract which renders streamlining data management difficult, even within a department or research group. Fluctuation accounts for various challenges associated with data storage and retention of knowledge/assets at the university, since the bulk of data is produced during PhD-projects that end after 3 years in most cases, after which time PhD-candidates leave. Respondents interpret RDM as a coping strategy that threatens to decouple researchers from their data. This is one cause of resistance, but certainly not the only one, as data are considered a competitive advantage even by researchers with unlimited contracts. As a consequence, policies mandating RDM are perceived as disproportionately benefitting the organization since RDM entails taking knowledge and responsibilities away from the primary data producers. However, there might be other aspects of data sharing, such as ensuring reproducibility of results, that are not captured by this perspective.

5.3 | The double nature of RDM as research and administration

RDM thus leads a double life among researchers: both as an acknowledged aspect of research as well as a menial administrative task. Open Science holds the promise to make research more transparent, accountable, and accessible (McKiernan, Bourne, Titus Brown, & Buck, 2016). Data Sharing as one pillar of Open Science (Anagnostou et al., 2015; Andreoli-Versbach & Mueller-Langer, 2014; Fecher & Friesike, 2014) has been associated with increased citation rates (Piwowar et al., 2007), economic growth (Tennant et al., 2016), increased transparency (Gilmore et al., 2017) and reproducibility (Toelch & Ostwald, 2018), and more efficient science (Leonelli et al., 2015). Our findings suggest that RDM is more readily interpreted as an administrative task than as a part of good research practice. Even though we found that data reuse is increasingly common, data management and curation are not. Researchers are expected to produce their own data as part of their qualification, and this is clearly reflected in data reuse habits.

5.4 | Industry data: External funding is no reason not to share data

Data from third-party funded projects are among the least shareable data types (Banal-Estañol, Jofre-Bonet, & Lawson, 2015). However, in the case of industry-funded projects, data sharing tends to be governed by extensive agreements to account for industry partners' need for profit and researchers' need to publish (Myneni et al., 2016). We found no systematic variation with respect to data sharing between industry-heavy disciplines and others, but some evidence that even in fields that rely heavily on third-party funding and do not tend to share their data widely, there seems to be no causality between the two. Rather, for disciplines like these (e.g., Electrical Engineering) data do not play the same epistemic role as for basic sciences because the epistemic weight lies with the respective prototype. In Sabina Leonelli's (2020, p. 8) terminology, prototypes serve to terminate data journeys at the outset.

6 | CONCLUSION AND PRACTICAL IMPLICATIONS: WHERE TO MOVE NEXT

We asked the following questions: What are faculty-specific research data management (RDM) practices and needs of scientists and engineers? What are the most salient dimensions of variation in RDM practices? How and to what extent do the seven faculties differ in their RDM needs? To what extent can RDM needs be generalized? Our results show only the beginnings of a culture attuned to RDM. In terms of RDM governance, our findings suggest that policies should be as (discipline)specific as possible without compromising invariant principles (such as the FAIR principles). These policies need to be flexible to accommodate the advances in data sharing practices and technologies. The interview results suggest a need to build more awareness of RDM as part of good research practice (instead of labeling RDM an administrative task).

The most surprising result was the ambivalence surrounding RDM. While RDM practices are clearly on the rise, the researchers we interviewed regard RDM as an administrative task, not as an aspect of good research practice. With respect to data practices, engineering disciplines and basic sciences seem to be fundamentally different. While engineers work with large amounts of data, these are usually specific to a given model or prototype, and therefore rarely shared. Large variability was found with respect to data storage needs. Even within faculties, departments and research groups study vastly different subject matters and collect diverse research data, rendering one-size-fits-all solutions to data storage/management difficult. Faculty-specific differences extend to the significance accorded to data management and to efforts invested in data gathering/processing. Whether a field makes use of secondary data, and whether data sharing is valued or not intimately depends on the specificities of the community, in terms of research aims, methodologies, and tools. These tend to transcend traditional administrative categories. Accordingly, we conclude that given the need for RDM governance, policies should take into account disciplinary differences.[9] Data governance needs to accommodate to data practices which do not always conform to disciplines. Attempting to press data management provisions into predefined administrative boxes risks overlooking the specificities of data practices.

7 | LIMITATIONS

The study provides a snapshot of one institution in one country and focuses exclusively on STEM fields. While we do not expect large national differences, results might be radically different for the social sciences and humanities. The survey did not ask about time spent on RDM. The largely exploratory study was only able to focus on differences and similarities between science and engineering fields. In addition, university policy prohibited asking for any demographic information beyond faculty and completed degree. Consequently, gender is absent from the analysis. The discrepancy between RDM as an administrative task and RDM as part of research practice renders choosing faculties as the unit of analysis problematic. The data clearly show that while RDM is predominantly interpreted as an administrative task, it has a much more prominent role to play in research practice. This, however, is a genuinely novel finding that was not apparent when planning the study.

ORCID

Stefan Reichmann https://orcid.org/0000-0003-1544-5064

ENDNOTES

[1] Though it must be noted that this is changing rapidly. See, for example, the excellent collection of work on "data journeys" in Leonelli and Tempini (2020). Data journeys is a framework to accommodate the situated and mutable nature of data (Leonelli, 2020).

[2] Maintaining data interoperability is widely recognized in data science as a crucial condition for data sharing, as one of the four crucial aspects of FAIR data (findability, accessibility, interoperability, reusability) (see Wilkinson et al., 2016).

[3] Official numbers record 1762 members of scientific staff in 2019, which defines a slightly smaller sample. The survey was based on data from the HR department which is more accurate for the time the survey was administered. Of those 1,762 staff members, 1,040 were third-party funded.

[4] The full data set is available at: https://doi.org/10.5281/zenodo.4701213.

[5] In terms of considering the views of architects, this has proved problematic, which is why respondents from this faculty have been excluded from all computations where "faculty" was used as an explanatory variable. The cases from architecture have been included wherever the data were considered globally (not by faculty).

[6] As indicated above (FN 5), Architecture has been excluded from the computations here.

[7] When this paper was first submitted, the university had only just ratified an RDM policy.

[8] https://d1rkab7tlqy5f1.cloudfront.net/Library/Themaportalen/RDM/researchdata-framework-policy.pdf.

[9] In some instances, this is already being implemented, for example, in the form of faculty-specific implementation strategies. For a recent example see https://www.tudelft.nl/en/library/current-topics/research-data-management/r/policies/tu-delft-faculty-policies/.

REFERENCES

Akers, K. G., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation*, *8*(2), 5–26.

Amorim, R. C., Castro, J. A., da Silva, J. R., & Ribeiro, C. (2017). A comparison of research data management platforms: Architecture, flexible metadata and interoperability. *Universal Access in the Information Society*, *16*(4), 851–862.

Anagnostou, P., Capocasa, M., Milia, N., Sanna, E., Battaggia, C., Luzi, D., & Bisol, G. D. (2015). When data sharing gets close to 100%: What human paleogenetics can teach the open science movement. *PLoS One*, *10*(3), e0121409.

Anderson, N. R., Sally Lee, E., Scott Brockenbrough, J., & Minie, M. E. (2007). Issues in biomedical research data management and analysis: Needs and barriers. *Journal of the American Medical Informatics Association*, *14*(4), 478–488.

Andreoli-Versbach, P., & Mueller-Langer, F. (2014). Open access to data: An ideal professed but not practised. *Research Policy*, *43*(9), 1621–1633.

Awre, C., Baxter, J., Clifford, B., & Colclough, J. (2015). Research data management as a "wicked problem". *Library Review*, *64*(4/5), 356–371.

Aydinoglu, A. U., Dogan, G., & Taskin, Z. (2017). Research data management in Turkey: Perceptions and practices. *Library Hi Tech*, *35*(2), 271–289.

Banal-Estañol, A., Jofre-Bonet, M., & Lawson, C. (2015). The double-edged sword of industry collaboration: Evidence from engineering academics in the UK. *Research Policy*, *44*(6), 1160–1175.

Bennion, A., & Locke, W. (2010). The early career paths and employment conditions of the academic profession in 17 countries. *European Review*, *18*, 7–33. https://doi.org/10.1017/S1062798709990299

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *Philosophical Review*, *97*(3), 303–352.

Borghi, J. A., & Van Gulick, A. E. (2018). Data management and sharing in neuroimaging: Practices and perceptions of MRI researchers. *PLoS One*, *13*(7), e0200562.

Borgman, C. L. (2010). *Scholarship in the digital age: Information, infrastructure, and the internet* (1st ed.). MIT Press.

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, *63*(6), 1059–1078.

Borgman, C. L. (2015). *Big data, little data, no data: scholarship in the networked world*. MIT Press.

Bugaje, M., & Chowdhury, G. (2018). Identifying design requirements of a user-centered research data management system. In *Lecture notes in computer science, information systems and applications, incl. internet/web, and HCI* (Vol. 11279, pp. 335–347). Cham: Springer. https://link.springer.com/chapter/10.1007/978-3-030-04257-8_35

Chen, X., & Ming, W. (2017). Survey on the needs for chemistry research data management and sharing. *Journal of Academic Librarianship*, *43*(4), 346–353.

Chigwada, J., Chiparausha, B., & Kasiroori, J. (2017). Research data management in research institutions in Zimbabwe. *Data Science Journal*, *16*, 1–31. http://dx.doi.org/10.5334/dsj-2017-031.

Collins, Harry M. (1985). Changing order. In *Replication and induction in scientific practice* (1st ed.). SAGE.

Cox, A., & Williamson, L. (2015). The 2014 DAF survey at the University of Sheffield. *International Journal of Digital Curation*, *10*(1), 210–229.

Cox, A. M., Kennan, M. A., Lyon, L., & Pinfield, S. (2017). Developments in research data management in academic libraries: Towards an understanding of research data service maturity. *Journal of the Association for Information Science and Technology*, *68*(9), 2182–2200.

Fecher, Benedikt, & Friesike, Sascha (2014). Open science: One term, five schools of thought. In *Opening science: The evolving* *guide on how the internet is changing research, collaboration and scholarly publishing* (pp. 17–48). Springer.

Fecher, B., Friesike, S., & Hebing, M. (2015). What drives academic data sharing? *PLoS One*, *10*(2), e0118053.

Fuhr, J. (2019). '"How do I do that?" A literature review of research data management skill gaps of Canadian health sciences information professionals. *Journal of the Canadian Health Libraries Association*, *40*(2), 51–60.

Geskin, A., Legowski, E., Chakka, A., Chandran, U. R., Michael Barmada, M., LaFramboise, W. A., Berg, J., & Jacobson, R. S. (2015). Needs assessment for research use of high-throughput sequencing at a large Academic Medical Center. *PLoS One*, *10*(6), e0131166.

Gilmore, R. O., Diaz, M. T., Wyble, B. A., & Yarkoni, T. (2017). Progress toward openness, transparency, and reproducibility in cognitive neuroscience. *Annals of the New York Academy of Sciences*, *1396*(1), 5–18.

Grant, R. (2015). Identifying HSS research data for preservation: A snapshot of current policy and guidelines. *New Review of Information Networking*, *20*(1), 97–103.

Higman, R., & Pinfield, S. (2015). Research data management and openness. *Program: Electronic Library and Information Systems*, *49*(4), 364–381.

Hine, C. (2016). Databases as scientific instruments and their role in the ordering of scientific work. *Social Studies of Science*, *36* (2), 269–298.

Hsu, L., Martin, R. L., McElroy, B., & Litwin-Miller, K. (2015). Data management, sharing, and reuse in experimental geomorphology: Challenges, strategies, and scientific opportunities. *Geomorphology*, *244*, 180–189.

Huang, R. (2016). *RQDA: R-based qualitative data analysis. R package version 0.2-8*. Retrieved from http://rqda.r-forge.r-project.org/

Kalichman, M., Sweet, M., & Plemmons, D. (2015). Standards of scientific conduct: Disciplinary differences. *Science and Engineering Ethics*, *21*(5), 1085–1093.

Knight, G. (2015). Building a research data management service for the London School of Hygiene & Tropical Medicine. *Program Electronic Library & Information Systems*, *49*(4), 424–439.

Kurata, K., Matsubayashi, M., & Mine, S. (2017). Identifying the complex position of research data and data sharing among researchers in natural science. *SAGE Open*, *7*(3), 2158244017717301.

Leonelli, S. (2016). *Data-centric biology. A philosophical study*. The University of Chicago Press.

Leonelli, S. (2020). Learning from data journeys. In S. Leonelli & N. Tempini (Eds.), *Data journeys in the sciences* (pp. 1–24). Springer.

Leonelli, S., Spichtinger, D., & Prainsack, B. (2015). Sticks and carrots: Encouraging open science at its source. *Geo*, *2*(1), 12–16.

Mancilla, H. A., Teperek, M., van Dijck, J., den Heijer, K., Eggermont, R., Plomp, E., van der Velden, Y. T., & Kurapati, S. (2019). On a quest for cultural change – surveying research data management practices at Delft University of Technology. *Liber Quarterly*, *29*(1), 1–27.

McKiernan, E. C., Bourne, P. E., Titus Brown, C., & Buck, S. (2016). How open science helps researchers succeed. *eLife*, *5*, e16800.

Myneni, S., Patel, V. L., Steven Bova, G., Wang, J., Ackerman, C. F., Berlinicke, C. A., Chen, S. H., Lindvall, M., & Zack, D. J. (2016). Resolving complex research data management issues in biomedical laboratories: qualitative study of an industry–academia collaboration. *Computer Methods and Programs in Biomedicine*, *126*, 160–170.

Perrier, L., Erik, B., Patricia Ayala, A., & Dearborn, D. (2017). Research data management in academic institutions: A scoping review. *PLoS One*, *12*(5), e0178261.

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS One*, *2*(3), e308.

Read, K. B., Larson, C., Gillespie, C., & So Young, O. (2019). A two-tiered curriculum to improve data management practices for researchers. *PLoS One*, *14*(5), e0215509.

Schöpfel, J., & Prost, H. (2016). Research data management in social sciences and humanities: A survey at the University of Lille (France). *Libreas: Library Ideas*, *29*, 98–112.

Steinhart, G., Chen, E., Arguillas, F., Dietrich, D., & Kramer, S. (2012). Prepared to plan? A snapshot of researcher readiness to address data management planning requirements. *Journal of eScience Librarianship*, *1*(2), 63–78.

Tennant, J., Waldner, F., Jacques, D. C., Masuzzo, P., Collister, L. B., & Hartgerink, C. J. (2016). The academic, economic and societal impacts of open access: An evidence-based review. In (Ed.), *F1000Research* (Vol. 5, p. 632). https://f1000research.com/articles/5-632.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS One*, *6*(6), e21101.

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS One*, *10*(8), e0134826.

Thylstrup, N. B. (2019). Data out of place: Toxic traces and the politics of recycling. *Big Data & Society*, *6*(2), 2053951719875479.

Timmermans, S., & Tavory, I. (2012). Theory construction in qualitative research: From grounded theory to abductive analysis. *Sociological Theory*, *30*, 167–186.

Toelch, U., & Ostwald, D. (2018). Digital open science—Teaching digital tools for reproducible and transparent research. *PLoS Biology*, *16*(7), e2006022.

Unal, Y., Chowdhury, G., Kurbanoglu, S., Boustany, J., & Walton, G. (2019). Research data management and data sharing behaviour of university researchers. *Information Research*, *24*(1). https://e-space.mmu.ac.uk/623415/3/ISIC%202018_published-1.pdf.

Van Tuyl, S., Michalek, G. (2015). Assessing Research Data Management Practices of Faculty at Carnegie Mellon University. *Journal of Librarianship and Scholarly Communication*, *3*(3), 1258. http://dx.doi.org/10.7710/2162-3309.1258.

Vilar, P., & Zabukovec, V. (2019). Research data management and research data literacy in Slovenian science. *Journal of Documentation*, *75*(1), 24–43.

Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS One*, *8*(7), e67332.

Whitmire, A. L., Boock, M., & Sutton, S. C. (2015). Variability in academic research data management practices. *Program: electronic library and information systems*, *49*(4), 382–407. http://dx.doi.org/10.1108/prog-02-2015-0017.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Reichmann, S., Klebel, T., Hasani-Mavriqi, I., Ross-Hellauer, T. (2021). Between administration and research: Understanding data management practices in an institutional context. *Journal of the Association for Information Science and Technology*, *72*(11), 1415–1431. https://doi.org/10.1002/asi.24492