Challenges in publishing research data – a Fraunhofer Case Study

Andrea Wuchner [ID] [1]
Michèle Robrecht [ID] [2]
Pierre E. Kehl [ID] [2]
Robert H. Schmitt [ID] [2, 3]

1. Fraunhofer Information Center for Planning and Building IRB, Stuttgart.
2. Fraunhofer Institute for Production Technology IPT, Aachen.
3. WZL | RWTH Aachen University, Aachen.

Abstract. Sharing of research data is becoming more and more established as part of the scientific process, triggered by corresponding requirements of research funders. A large number of subject-specific and institutional research data repositories were created as publication agents for the research data. Nevertheless, the publication processes are not yet established and need to find a way to best practice. The aim of this paper is to work out what challenges currently exist in the publication of research data and how these can be overcome. Answers to the research question are provided by a case study of research data publication with the participation of industrial partners in the institutional research data repository of the Fraunhofer-Gesellschaft "Fordatis". The publication process is described from the perspective of the researcher, the data curator and the repository operator. In summary, the challenges can be overcome primarily by the division of labor and communication.

1 Introduction

"Open Science", "Open Access", "Open Code" and similar buzzwords have been on everyone's lips in the scientific community for some years now. There are a multitude of definitions, opinions and views on them that result in a "vivid discourse [...] whose lowest common denominator is perhaps that science in the near future somehow needs to open up more." [1] Open Science also includes "Open Data", i.e. making research data available. This is necessary for a democratic science in which everyone has a right to knowledge, especially if its generation was financed by taxpayers' money. [1] The Alliance of Science Organizations in Germany understands the term "research data" to mean data generated in the course of scientific projects, e.g., through digitization, source research, experiments, measurements, surveys or interviews. [2] In the following, this term includes all types of data that are generated or reused to gain scientific knowledge. This includes, for example, images, video recordings, audio recordings, visualizations or models, texts, tables, databases, programs and device-specific data.

However, sharing research data is not yet a mandatory part of the research process; the reasons

for this are unclear so far. [1] Therefore, the aim of this paper is to investigate what challenges currently exist in publishing research data and how they can be overcome. For this, a review of the state-of-the-art on data publications establishes the theoretical assumptions in the following. For this purpose, both research in general and the processes in the Fraunhofer-Gesellschaft are considered, with own experiences supplementing literature sources. These processes are the object of application of the case study, which is intended to show exemplary challenges in publishing research data. The description of the methodology and the prerequisites is followed by a presentation of the results from the perspectives of the three main participants in the publishing process. In the last section, a discussion of the results leads to a conclusion, which is supplemented by an outlook.

2 Data publications in research and at Fraunhofer

The process of publishing research data is only a small part of the life cycle of research data. The following section therefore places it in an overarching context at the outset.

2.1 Data publications as part of the research data life cycle

A research dataset passes through several phases during the research process, which can be portrayed by a data life cycle. These cycles exist in different variants and at different levels of detail, for example according to the model of the Harvard Business School [3] or that of Princeton University [4]. This paper is based on the data life cycle of the UK Data Archive [5] (see Figure 1). All versions of this life cycle have in common that the publication of research data has a fixed place. After all, this is the only way to fully exploit the value that lies in the research data. This benefit accrues to several research stakeholders: [6]

- Researchers: Researchers gain visibility through their data publications and have the opportunity to gain reputation through citations. They rise in perception as potential project partners and can qualify themselves for senior positions.

- Research organizations: Research organizations also gain visibility and reputation through data publications, for example by making high-quality research datasets available in their own repositories.

- Disciplines: Sharing research data makes it possible to collectively build large repositories within a discipline. Researchers can consider them from a new perspective, which makes research more efficient because it avoids duplication of data collection. The analysis of research data fundamentally makes findings possible that cannot be generated with analog research objects – a potential that sharing also makes accessible to researchers with limited financial resources.

- Industry: With the increasing complexity of industrial processes, constant price pressure and the demand for sustainable production (especially in the context of Industry 4.0), the use of data-driven solutions such as Artificial Intelligence (AI) is growing. To be able to deploy these in a targeted and rapid manner, industrial companies need data. Shared research data enables them to easily transfer research results to their own production and to use AI without the additional expense of their own measurement setups.

- Society: Research activities become more transparent by making research data available, which strengthens the trust and acceptance of citizens in research. In addition, society as a whole benefits from the scientific progress that all the aforementioned advantages of sharing research data make possible.

How the process of publishing research data looks in detail and which persons are involved is described in the next section.

2.2 Process participants and their tasks during data publication

Various steps and decisions are necessary for sharing research data (see Figure 1). They are not necessarily performed at this stage of the research data life cycle, but for example already when writing a data management plan. Moreover, they do not have a specific order, only the start and end points are fixed: the selection of research data and the ingest process.
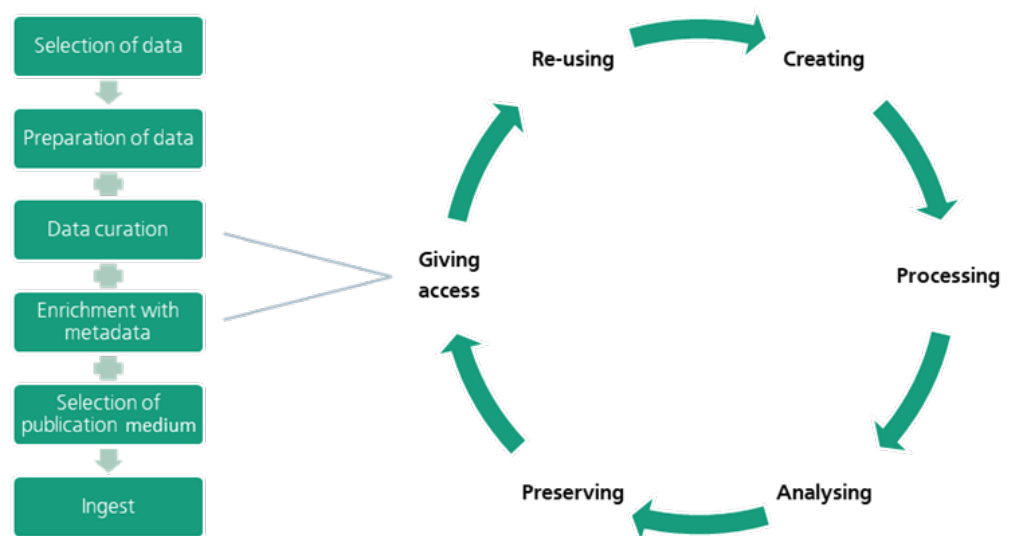

Figure 1: Work steps for data publications in the data life cycle

Researchers collect and analyze research data at the beginning of the data life cycle as part of research or industry projects. Many of these projects are carried out either bilaterally or in a consortium with other research and industry partners. Industry partners enable a strong application focus to ensure broad use of the results as quickly as possible. However, as a result, the results and thus also the research data resulting from such a project are not the sole property of the researchers, but also of the industry partners who participated in the collection of the data.

The first step in the process of publishing research data is the selection of a suitable sample from the set of all research data of a project that is to be published. This should contain research data that are available at an aggregation level suitable for subsequent use and that are meaningful. Legal framework should be clear and publication rights should be available. The interests of possible industry partners must be taken into account. It is often difficult to obtain their consent to a data publication because they want to exploit the value of the research data themselves. It is therefore all the more important for all project participants to reach agreements at the beginning

on how they want to handle the jointly created research data. Finally, if necessary, an export control must be carried out in order to exclude the possibility of dual use.

Once a suitable sample has been selected, the research data must be prepared for publication. This can include, for example, transferring the data into another, open format to facilitate subsequent use by other persons. In addition, the data sets themselves must be checked for completeness and correctness with regard to labels and values.

A data curator[1] can check the research data before publication. This check usually includes formal criteria such as data formats, variable labels, missing values, data format as well as completeness. Automatic check routines can be used in combination with manual checks Checking the quality of the content is the responsibility of the researchers. [7]

In order to comply with the FAIR principles [8] (see section 2.3 "FAIR principles as a quality feature for data publications"), describing research data with metadata is essential. If possible, researchers document them when creating the research data, so that they only have to add a few more for publication, depending on the requirements of the publication medium, and no information is lost during the research process. Ideally, metadata schemas are used for this purpose; generic or discipline-specific ones are possible, depending on the context. Since the capture in softwares for data management is not yet mature, open and widely used data formats such as TXT or CSV are recommended for the documentation of metadata, if possible. This task can only be performed by the researchers themselves, since external persons cannot know what information is relevant for the metadata.

A suitable publication medium must also be selected for the publication of research data. Basically, three different paths have become established: [9]

- Supplements to articles: The research data related to the findings in a scientific article is published as an appendix. This is the oldest way.

- Data journals: Data journals are scientific journals that publish research data, even if the researcher has not yet derived research results from them. The journals contain descriptive texts about the data publications. The actual research data is stored in a repository.

- Research data repositories: These "system[s] for storing, making available, and re-using digital data from and for science" [10] exist in discipline-specific, institutional, and generic forms. Publishing research data in a repository has the advantage of achieving the FAIR principles [8].

Publishing in a repository is preferable to the other two types of publication with respect to FAIR principles.[2] Discipline-specific repositories are more recommended than institutional or

---

1. A data curator is a person who supports researchers in research data management and ensures the formal quality of research data. To be distinguished from this are the more broadly based data stewards. In organizations, these are responsible for the research data management of a department. There is often a separate job profile for this role, which may also include the activities of a data curator.
2. The "Re3Data.org" directory (https://www.re3data.org/) provides a searchable selection of potential repositories.

generic ones, as they are most likely to reach the target audience of data publication and use a metadata schema adapted to the discipline. Librarians and other publication supporting positions can assist in selecting an appropriate publication medium. Some research libraries also contain or are part of central FDM offices. They provide advice and infrastructure (such as repositories), which can sometimes give them a dual role as data curators.[3]

After the researchers have selected a suitable sample and prepared its publication, the ingest completes the process. This is understood to be the input and upload process in a repository. This usually consists of entering the metadata (manually or automated via an interface) and uploading the files containing the research data. In most cases, researchers also grant a license in this process, which defines the conditions for subsequent use of the research data. In addition, the data publications are mostly given persistent identifiers.

Many research funders have recognized the potential of publishing research data and take research data management into account in their funding guidelines. At the international level, for example, the publication of research data is mandatory for projects in the European Union's Horizon Europe framework program. [11] In Germany, the Deutsche Forschungsgemeinschaft calls for data sharing in its guidelines on research data management. [12] The German Federal Ministry of Education and Research sets individual requirements for the publication of research data in its calls for proposals. It is also committed to promoting data infrastructures, data-based innovations and data competencies with its "Aktionsplan Forschungsdaten" [13].

A recurring aspect of publishing research data are the FAIR principles, which are the topic of the next section.

2.3 FAIR principles as a quality feature for data publications

In order to fully realize the value proposition for research data (see section 2.1 "Data publications as part of the research data life cycle"), it is important that the respective data publication complies with the FAIR principles. The abbreviation stands for [8]:

- Findable

- Accessible

- Interoperable

- Reuseable

Sub-criteria are assigned to each of these points. There are two prerequisites for fulfilling the principles: First, the publication medium must be technically designed to enable FAIR publication, which basically only research data repositories can fulfill. This includes extensive metadata fields, an open protocol such as HTTPS, and a powerful search index that allows downstream researchers to find the data publications again. Second, the uploading researcher must take full

---

3. For example: https://www.eresearch.uni-goettingen.de/ or https://www.tib.eu/de/publizieren-archivieren/publikationstypen/forschungsdaten

154 advantage of the repository's capabilities and provide, for example, accurate and comprehensive
155 metadata. He or she must upload the research data in open and durable formats.

156

157 Table 1 shows an overview of the FAIR principles [8] and which ones have to be fulfilled by
158 the repository or the researchers:

159

Findable	Who?
Metadata and data are given a unique, long-term identifier	Repository
Data is described by rich metadata	Repository/Researcher
Metadata and data are indexed in a searchable resource	Repository/Researcher
The metadata specifies the unique identifier	Repository
Accessible	**Who?**
Metadata and data are accessible via a unique identifierusing a standardized protocol (e.g. http)	Repository
The protocol is open, free and universally applicable	Repository
The protocol allows authorization and authentication	Repository
Metadata is still accessible even if the data can no longer be accessed	Repository
Interoperable	**Who?**
For representation, data and metadata use formal and common languages	Researcher/Repository
Controlled vocabularies are used for the metadata and data, in accordance with the FAIR principles of	Repository
Metadata and data have qualified references to other resources, e.g., papers	Repository/Researcher
Re-Usable	**Who?**
Metadata and data contain diverse and accurate attributes	Researchers
Metadata and data are provided with a unique usage license	Researcher/Repository
Metadata and data are linked to their creation history	Researcher/Repository
Metadata and data obey the standards of the respective subject community	Researcher/Repository

Table 1: FAIR principles and responsibilities

160 The following section describes the prerequisites Fraunhofer creates for data publications.

2.4 Data publications at Fraunhofer

162 The Fraunhofer-Gesellschaft began setting up a research data infrastructure in 2016. The head-
163 quarters of the Fraunhofer-Gesellschaft's mandate was implemented by the department "Re-
164 search Services & Open Science" at the Fraunhofer Information Center for Planning and Build-
165 ing IRB. Its staff consists of software developers, specialists in media and information services
166 and information scientists. Together they realized central services for training and support as

167 well as the institutional research data repository "Fordatis". [14] The repository is based on the
168 software DSpace 6.3 JSPUI and was launched in September 2019. Currently, 113 datasets are
169 published in Fordatis. [15] A research data policy does not yet exist for Fraunhofer.
170

171 The ingest process at Fordatis is composed of a three-stage workflow (see Figure 2). The pro-
172 cess is started by the researchers. They enter the metadata using a set of rules and upload the
173 research dataset to the repository. They then send the Fordatis record to the data curator.
174

175 Half of the Fraunhofer institutes appointed a staff member as data curator[4]. Often, this role is
176 located in the institutes' libraries or in the scientific information management[5]. The data curator
177 checks the metadata using the set of rules. It is particularly important that the entries are in the
178 correct form in the correct field. The data curator can edit the Fordatis record, send it back to the
179 researcher in case of queries or forward it for final approval. Here, the data curator corresponds
180 to the prototype "data stewards as coordinators". This is characterized by a relatively narrow
181 range of tasks, a central position, high service orientation, interdisciplinary basic knowledge of
182 research data management, disciplinary expertise that is not necessary, and little IT/program-
183 ming knowledge. [16]
184

185 When the data curator forwards the Fordatis record to the repository operators (the department
186 "Research Services & Open Science" of Fraunhofer IRB), they check the metadata the last time.
187 They also conclude a deposit license with the submitting researchers, which transfers to them
188 all necessary rights of use in accordance with German copyright law.Finally, they release the
189 Fordatis record. At this point, the data publication receives a Digital Object Identifier (DOI) and
190 is discoverable and accessible worldwide via its metadata. It fulfills almost all FAIR principles.
191 [15]
192

193 These descriptions of the state-of-the-art in science and at Fraunhofer represent the basic as-
194 sumptions about processes and participants in the case study. More details on the method used
195 are given in the following section.

3   Methodology of the case study

197 In order to answer the research question, a method was chosen that is based on a single case
198 study. Case studies examine currently occurring phenomena, taking into account the associated
199 context. Their goal is to make the relationship between this phenomenon and its circumstances
200 mappable. There are "many more variables of interest than data points" [17], so the results tend
201 to be qualitative and explorative in nature.
202

---

4. As of April 2023. If no data curator has been appointed at the respective Fraunhofer Institute, the process skips this
step, so that the Fordatis record reaches the repository operators immediately.
5. "Scientific information management" is a functional term used by Fraunhofer to refer to all services provided in the
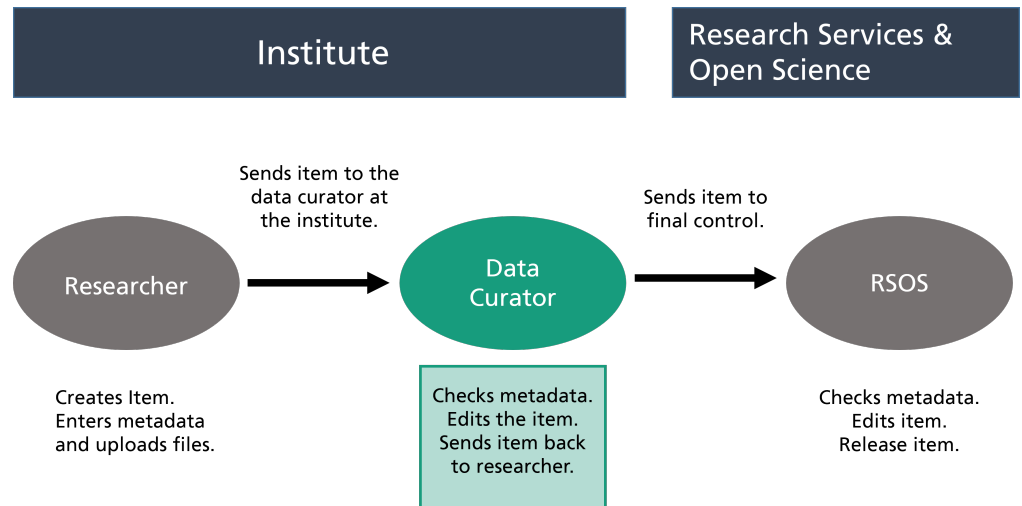area of libraries and publication support.

Figure 2: Work steps for data publications in the data life cycle

Here, a single case study was carried out, i.e. only one case was considered. This case is a typical case in order to be able to pay particular attention to relevant circumstances. [17] The case involves a researcher who is involved in a research project. He generated research data together with an industry partner and published his results in a paper based on them. The research data initially remained under NDA, as it was owned equally by the researcher and the industry partner. To investigate this case, which is typical for applied research, the researcher received permission from the industry partner to upload the research data to a repository. The researcher works as a research associate at the Fraunhofer Institute for Production Technology IPT. He had not previously published a research dataset or had any contact with a repository.

In order to be able to look at the entire process, other people besides the researcher are involved in the case study: The data curator works in scientific information management at the Fraunhofer IPT and advises staff on research data management within this context. She has a central position in the institute and not in a department, which means that her range of tasks is broad. Due to her position, her service orientation is very pronounced. Her basic knowledge of research data management relates primarily to engineering, although knowledge of other subjects is also present. Disciplinary subject knowledge is not strong, although it would be useful in consulting. IT and programming skills, on the other hand, have not been needed at all. Thus, the data curator most closely matches the prototype of an "FDM consulting data steward." [16] The data release was carried out by library staff in the "Research Services & Open Science" department at Fraunhofer IRB, which is also the point of contact for questions about the repository. Knowledge of metadata and metadata standards is particularly relevant for data release. An important tool in this context is the Fordatis set of rules. It contains specifications about which metadata must be entered in which form in which field in the input mask.

Case studies can use "six sources of evidence" [17]:

- Documentation

- Archival records

- Interviews

- Direct observations

- Participant-observation

- Physical artifacts

The self-observation of the persons involved corresponds most closely to participant-observation. One potential of this approach is that it can provide unique insights into the process of publishing research data from an inside perspective. [17] Likewise, the profiles of the people involved offer the advantage that they could contribute a lot of background knowledge and also pay attention to details. On the other hand, this can skew the results because the individuals could not start the case study "tabula rasa". [17] They not only observe the research field, but are already a part of it before the case study.

The three persons involved in the case study observed the process of publishing research data and documented their steps in a joint protocol. This was the basis for identifying the current problems presented in the following.

4 The data publication process from three points of view

The following sections summarize the protocolled process of publishing research data in the Fordatis research data repository from the perspectives of the people involved.

4.1 Data publication from the researcher's perspective

For the researcher, the motivation for data publication lays in its benefits: The disclosure of the measurement data underlines the credibility and reproducibility of the research results presented in the paper. It also promotes exchange and possible cooperation with other researchers.

The research data set discussed in this case study was collected jointly by several partners within a consortium in a research project. For this reason, they could not be assigned as research results to just one consortium member, which is why the researcher had to obtain the relevant consents for publication in the "selection of data" step. In this process, the consortium agreement regulated ownership claims and necessary queries. Accordingly, the researcher had to enter into dialogue with the partner concerned, understand the legal situation and present the added value of data publication in order to obtain approval. Since the research dataset contains sensitive content of the partner, they agreed to the publishing it only under the condition that the data publication would be embargoed and only made freely available after three years.

The measurement data existed as packet capture (PCAP) documents and were accordingly only available in a subject-specific form. In the step "preparation of data", the researcher had to convert them into a more open and widely used format for easier subsequent use of the research data.

To do this, he chose XML and CSV - formats that Fordatis supports. Researchers must decide on a format depending on the publication medium, as not all publishing platforms support the same ones.

In the step "enrichment with metadata" (which overlapped with the ingest process), the researcher described the measurement data with metadata. The metadata enables subsequent users of the data publication to understand the measurement setup, the background of the measurement series, the research dataset and a possible analysis. Fordatis asks for various metadata in the input mask. These can be divided into four groups:

- General information such as the name of the research dataset as well as participating researchers and institutes.

- Content description such as freely selectable keywords, a summary of the research data and a technical description. The measurement data in the case study are the basis for a paper [18] which describes both the measurement setup and the analysis in great detail. The researcher used this as a template and linked it directly in the Fordatis record.

- Project information of the associated research project.

- Information on the initial publication of the research data.

The researcher set a three-year embargo period in the further ingest process and uploaded the measurement data. However, the final release of the Fordatis record initially failed due to the lack of a deposit license. Uploading researchers must agree to this for each individual data publication. Fordatis generates the license documents automatically, the researchers have to download them, sign them and send them by e-mail to the repository operators. They manually review the document and deposit it in the system. However, the repository generates this document only at a single point in the ingest process, which the researcher missed when he submitted it. Thus, he had to request the deposit license from the repository operators together with the data curator. After the final release, the researcher received an email that his "Submission Approved and Archived" and the DOI was now functional. [19]

A major challenge for the researcher in carrying out the data publication was obtaining the consent of all parties involved. Industry partners in particular harbor a certain skepticism about publishing research data, as this can contain unique know-how about the technology or process used. The researcher had to convince the partner accordingly with the added values of data publication, which proved difficult and was only possible in combination with an embargo period.

The researcher is an expert in his field of research. Accordingly, the contents, representations and information in the research data are comprehensible to him. He also works with special software that can read and display the data formats used. For the researcher, it was therefore a further challenge to present the research data in a generally understandable way and to convert the specific formats into generally usable ones without risking data loss.

Another hurdle was the technical implementation of Fordatis. Despite a detailed input mask, it was not apparent to the researcher what the process was for creating and submitting the deposit license. Without a data curator, the process would have been more time consuming. In addition, the data publication is no longer accessible to the researcher himself, as he is also prevented from downloading by the embargo period. However, for a subsequent control of the correctness of the released files, an insight would be desirable.

The fact that this data publication was the researcher's first and the lack of experience made the process more laborious also presented a challenge. However, support services and tools were able to compensate for this: The clear prompts in the Fordatis guidelines easily guided the researcher through the ingest process, ensuring that they were adhering to a standard. The explanations of each metadata field were also very helpful. In addition, the support provided by the data curator was important. This began with the selection of a suitable publication medium and accompanied almost all steps in the publication process. Particularly helpful was the assistance in filling out the input masks in Fordatis, especially in the general process of publishing research data, in licensing issues, and in checking the readability of the Fordatis record after enrichment with metadata.

It is already clear here that the data curator took on an important function in the process. The next section elaborates on this.

4.2 Data publication from the data curator's point of view

In the case study, the data curator mainly had a supporting role: The researcher decided to publish the measurement data. At this point, the data curator became active and explained the general process of publishing research data. She also explained what requirements the repository had for the research data. So she supported in the steps "preparation of data", "data curation" and "selection of publication medium".

The researcher finally created a Fordatis record in the ingest process. The initial metadata entry was done jointly with the data curator, who assisted with knowledge gained through Fraunhofer-internal training as well as the rule set. Researcher and data curator discussed each metadata field and its properties and filled in those that were possible right away. Notably, the final selection of a license remained open at this point and drew further research from the data curator. The researcher added any remaining open metadata later and finally sent the database entry to the data curator for review. At this point, the ingest process overlapped with the "enrichment with metadata" step, as the entry also required metadata that the researcher had not previously assigned.

In the next part of the ingest process, the data curator reviewed the Fordatis record for completeness and correctness. She added optional metadata such as the affiliations of the researchers, a proposal for the German translation of the metadata, and a classification in the Dewey Decimal Classification. She also chose the license that was most appropriate based on her research.

Most of these changes required the researcher's review and approval as a specialist; the data curator also did not have the necessary permissions for one, a technical error. For these reasons, she sent the Fordatis record back to the researcher with appropriate comments. He revised the dataset according to the comments and sent it back to the data curator for final submission.

The challenges for the data curator were primarily that she was not trained in engineering. Thus, she and the researcher had to establish a common basis of understanding at the beginning of their project - also in reverse, since the researcher did not know in detail about the process for publishing research data. Proposals by the data curator had to be checked by the researcher for technical correctness and plausibility, which took time and resources on both sides.

Another challenge was the selection of licenses. Since there are few generally applicable recommendations for licensing research data, the possibilities had to be explored and assessed by the data curator. However, even after thorough investigation, there are ultimately residual uncertainties, as the data curator is not juridical trained.

The assignment of metadata presented a final challenge. Since there are still few standards for metadata for research data (especially in the engineering sciences), the question arose as to what information should accompany the data publication and, in particular, what was needed to fulfill the FAIR principles. Here, again, the researcher was primarily in the debt of providing input as a specialist. The data curator could only contribute to readability and formal correctness by checking the metadata. Questions about entering metadata in the Fordatis mask in particular were mostly answered by internal help materials.

Nevertheless, incorrect or incomplete metadata would have been intercepted by the repository operator, who could have returned the Fordatis record to the institute with appropriate comments after final submission. This final quality control is described in more detail in the next section.

4.3 Data publication from the repository operators' point of view

The submission of the measurement data in this case study proceeded successfully in Fordatis. The researcher and the data curator filled in the metadata fields completely, especially mandatory fields. They did not need extensive support from the repository operators; the data curator only approached them with a question about license selection.

The only challenge was that at the end of the ingest process, the deposit license was automatically generated but not retrieved in the first attempt. The repository operators therefore had to send the Fordatis record back to the researcher again to restart this process.

The following section summarizes and discusses the main findings and provides an outlook.

5   Discussion and Outlook

The individual case study conducted makes clear that challenges exist for everyone involved in the process of publishing research data: The ingest process, as outlined in Section 2.4 "Data publications at Fraunhofer", could not be carried out the described way because there was a need for coordination beyond it. The intensity of communication between all parties involved was not foreseen. The data curator played an important role in this process. She served as the first point of contact for the researcher and thus as a multiplier for processes in research data management and data publications. She thus became the link between the researcher and the repository operators, mediating between them and facilitating the work of both sides. Especially within the institute, close collaboration between the data curator and the researcher is therefore rewarding and would be even more profitable if the data curator had expertise in the field. In this case study, the data curator took on more tasks than his or her role in the Fordatis ingest process actually envisaged. This is an exemplary extension of the profile of a Fraunhofer data curator. Ongoing training is essential for them, as research data management and data publishing are sub-ject areas that are constantly evolving. Special attention should be paid to discipline-specific know-how if the persons to be advised are limited to one discipline.

For the researcher, obtaining the project partner's consent for publication was the biggest hurdle. Although a contract defined ownership rights and coordination obligations, it did not include data publication. Such projects should be planned as early as possible in the application phase to save time and resources. Furthermore, the researcher needed assistance in the publication process, especially since it was his first. Aids such as documents were not sufficient in this regard; for more, personal support and advice from the data curator was needed. As a result, there is a need for experts to assist researchers with data publications and overall research data management. As the state-of-the-art has shown, the processes are not yet clearly defined, let alone standardized. Thus, data publications – especially FAIR ones – are a major challenge for researchers.

Fordatis functions largely without problems as an institutional repository. It allows the descrip-tion of research data with metadata according to the Dublin Core Standard. However, the col-lection of metadata only starts at the ingest process; the researcher is responsible for the earlier metadata collection. The repository assigns a DOI as a persistent identifier, which in no small part helps to ensure that most of the FAIR principles are met. The ingest process can be used intuitively by researchers. A set of rules and explanatory texts for each metadata field are avail-able to assist. The help documents on licensing issues could be expanded in the future, as they explain possible licenses in an understandable way, but do not provide any decision sup-port. However, the automatic generation of the deposit license is problematic, as it cannot be retrieved independently by the researcher after closing the browser tab. Repository operators must then generate the license document manually. There is no other way to implement the process in a technically and legally secure manner with the DSpace version currently in use.

The limitations of the case study lie primarily in its institutional perspective. Looking at other

cases is essential for uncovering further challenges and their solution. Other institutional as well as generic repositories should be considered, as well as the other two publication channels (supplements to articles and data journals), since they offer researchers a way to publish research data just like repositories do.

During the data publication process, several pitfalls were identified for the researcher, which he had to and could solve together with the data curator. In conclusion, the key to a process as smooth as possible is division of labor and communication, and sharing best practices would be worthwhile.

6 Acknowledgements

7 Roles and contributions

Andrea Wuchner: Conceptualization, Writing Original Draft, Writing Review and Editing

Michèle Robrecht: Conceptualization, Writing Original Draft, Writing Review and Editing, Data Curator

Pierre E. Kehl: Conceptualization, Writing Original Draft, Writing Review and Editing

Robert H. Schmitt: Funding Acquisition

References

[1] B. Fecher and S. Friesike, "Open science: One term, five schools of thought," in *Opening Science*, S. Bartling and S. Friesike, Eds., Cham: Springer, 2014, pp. 17–47. DOI: 10.1007/978-3-319-00026-8_2.

[2] M. Franke, S. Heinzel, R. Mauer, *et al.*, *Positionspapier „research data at your fingertips" der arbeitsgruppe forschungsdaten*, Potsdam, 2015. DOI: 10.2312/allianzfd.001.

[3] T. Stobierski, *8 steps in the data life cycle*, 2021. [Online]. Available: `https://online.hbs.edu/blog/post/data-life-cycle` (visited on 07/04/2023).

[4] *Research lifecycle guide*, 2023. [Online]. Available: `https://researchdata.princeton.edu/research-lifecycle-guide/research-lifecycle-guide` (visited on 07/04/2023).

[5] V. van den Eynden, *Data life cycle & data management planning*, 2013. [Online]. Available: `http://dam.ukdataservice.ac.uk/media/187718/dmplanningdm24apr2013.pdf` (visited on 07/04/2023).

[6] T. Poisot, R. Mounce, and D. Gravel, "Moving toward a sustainable ecological science: Don't let data go to waste!" *Ideas in Ecology and Evolution*, vol. 6, no. 2, pp. 11–19, 2013. DOI: `10.4033/iee.2013.6b.14.f`.

[7] F. Cremer, M. Fräßdorf, J. Neumann, *et al.*, "Institutionelle workflows zum forschungsdatenmanagement: Bestandsaufnahme und lösungsvorschläge aus der leibniz-gemeinschaft," *Bausteine Forschungsdatenmanagement*, no. 3, pp. 142–150, 2021. DOI: `10.17192/bfdm.2021.3.8346`.

[8] M. D. Wilkinson, M. Dumontier, I. J. J. Aalbersberg, *et al.*, "Comment: The fair guiding principles for scientific data management and stewardship," *Scientific data*, vol. 3, 2016. DOI: `10.1038/sdata.2016.18`.

[9] *Daten publizieren: Vorteile und möglichkeiten des data-sharing*, 2023. [Online]. Available: `https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/daten-publizieren/` (visited on 06/26/2023).

[10] *Repositorien: Daten professionell zur verfügung stellen*, 2023. [Online]. Available: `https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/repositorien/` (visited on 06/26/2023).

[11] *Horizon europe (horizon): Model grant agreement. unit grants. (he unit mga — multi & mono)*, 2021. [Online]. Available: `https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/agr-contr/unit-mga_he_en.pdf` (visited on 06/26/2023).

[12] *Leitlinien zum umgang mit forschungsdaten*, 2015. [Online]. Available: `https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/leitlinien_forschungsdaten.pdf` (visited on 06/26/2023).

[13] *Bmbf-aktionsplan forschungsdaten: Impulse für eine kultur der datenbereitstellung und weiterverwendung in bildung, wissenschaft und forschung*, 2020. [Online]. Available: `https://www.bmbf.de/bmbf/shareddocs/downloads/files/163_20_faktenblatt_aktionsplan_4.pdf?__blob=publicationFile&v=2` (visited on 06/26/2023).

[14] *Fordatis - forschungsdaten-repositorium der fraunhofer-gesellschaft*, 2023. [Online]. Available: `https://fordatis.fraunhofer.de/` (visited on 06/26/2023).

[15] O. Beyan, A. Wuchner, D. Eisengräber-Pabst, C. Quix, C. Zaschke, and O. Schumacher, *Research data in the fraunhofer digital project : Creating a fair research data infrastructure and culture*, 2018. DOI: `10.5281/zenodo.1441011`.

[16] E. Seidlmayer, F. Hoffmann, J. Dierkes, B. Lindstädt, R. Depping, and K. U. Förstner, *Forschung unterstützen: Empfehlungen für data stewardship an akademischen forschungsinstitutionen: Ergebnisse des projektes datastew. (stand: 02.02.2023)*, Köln, 2023. DOI: `10.4126/FRL01-006441397`.

[17] R. K. Yin, *Case study research and applications: Design and methods*, Sixth edition. Thousand Oaks, California: SAGE, 2018.

[18] P. Kehl, J. Ansari, M. H. Jafari, *et al.*, "Prototype of 5g integrated with tsn for edge-controlled mobile robotics," *Electronics*, vol. 11, no. 11, 2022. DOI: `10.3390/electronics11111666`.

[19] P. Kehl and J. Ansari, *Prototype of 5g integrated with tsn: Dataset,* 2023. DOI: `10.24406/fordatis/208`.