ing.grid preprints

RESEARCH ARTICLE

Pragmatic Research Data Management for Heterogenous Sensor Data

Bridging the Gap Between Best and Current Practices

Matthias Bodenbenner [1]
Tobias Hamann [1]
Mario Moser [1]
Mark P. Sanders [1]
Dominik Wolfschläger [1]
Robert H. Schmitt [1, 2]

1. Intelligence in Quality Sensing, WZL | RWTH Aachen University, Aachen.
2. Fraunhofer Institute for Production Technology (IPT), Aachen.

Date Submitted:
2025-09-09

Keywords:
bespoke experiments, data reuse, interoperability, sensor data, machine tools, sensor interfacing language

Data availability:

Software availability:

Corresponding Author:
Matthias Bodenbenner
matthias.bodenbenner@wzl-iqs.rwth-aachen.de

Abstract.
Research Data Management (RDM) is essential to facilitate open, effective, and accountable research. However, embedding RDM in researchers' workflows is mainly hampered by two aspects: lack of incentives for RDM adoption, and limited knowledge about tools and services supporting RDM. We address this issue by implementing a standardized data description model and describe how we applied effective RDM practices to our research activity, *Virtual Climatization*. We demonstrate how implementing RDM actions can increase efficiency and effectiveness in the long run. This article showcases the success of our measures, including open-access publications, software artifacts, and reusability of data and services by other projects and researchers. We also highlight the importance of incentives, tools, and knowledge management for widespread RDM adoption. This paper provides insights into how effective RDM can enhance research continuity and depth, making it a landmark study in the engineering domain.

1 Introduction and Motivation

Research Data Management (RDM) is crucial to facilitate successful, transparent, and sustainable research practices. The importance of widely shared (intermediate) research results is nowadays often demanded by funding bodies [1]. These results include written publications but also datasets and open-source software artifacts. Such demands impose an increased overhead effort associated with data management processes and requirements [2]. Funding often does not compensate for this additional burden, making it difficult for researchers to justify investing their time and resources into RDM activities [3]. As a result, the benefits of proper data management practices are not always evident to researchers, leading to a reluctance to adopt them. Moreover, establishing incentives by structural measures, such as rewarding researchers for their contributions to good RDM, potentially take years to be fully realized [4], [5].

These deficits necessitate the development of conceptual guidelines, such as the *FAIR Guiding Principles* [6], and implementation of technical solutions that support researchers and automate the RDM process [7].There are plenty of national and international initiatives [4], [8], [9], [10], [11], [12], [13] already working on guidelines and supportive tools. In recent years, these initiatives have led to significant developments in this field, resulting in the emergence of various approaches and frameworks providing structure and support for researchers in their project-related RDM activities [14], [15], [16]. However, using and integrating these approaches and frameworks into everyday scientific life is not trivial and is still associated with high effort. Thus, to pave the way for successful RDM shortly, it is imperative to lower the barriers and reduce the effort required for researchers to engage in effective data management [4], [5]

Contribution. We recognize this need and aim to contribute to the advancement of RDM practices by establishing a technical infrastructure capable of automating key data management steps across various research projects. In this article, we present the motivations, challenges, and our approach in advancing RDM through technical solutions and application of RDM. By providing researchers with user-friendly tools and standardized data models, which can be readily applied in diverse research contexts, we seek to empower researchers to engage in efficient and effective RDM in practice. We will highlight the benefits of good RDM by presenting the research activity *Virtual Climatization* being a beacon for careful research data management and (internal) data reuse over several years by multiple researchers in individual research projects and dissertations. By realizing this technical infrastructure, we aim to bridge the gap between the current state of RDM and its desired future, where data management becomes an integral part of research workflows and yields tangible benefits for researchers and the wider scientific community.

Article Structure. The remainder of this article invests this contribution in detail as follows: First, we underline the need for RDM by introducing the research activity *Virtual Climatization*. After that, the individual steps of the RWTH's research data life-cycle (DLC) are tackled constituting the conceptual frame for all our RDM activities, see Figure 1. Each step of the DLC is discussed in a dedicated chapter, which is split into two sections. First, the phase-specific theory, state-of-the-art, and best practice suggestions are briefly reviewed. Second, our execution of the step-specific research actions for the presented research activity is explained. The paper is finalized by a discussion on the results in chapter 9 and a conclusion in chapter 10.

2 *Virtual Climatization* – driven by RDM in engineering

Virtual Climatization is a research activity in the realm of machine tools at WZL of RWTH Aachen University, where the need for RDM became apparent and which we introduce in this section. Machining workpieces on machine tools is an inherently energy-intensive process. To achieve ever-increasing tolerance requirements, thermal disturbances on machine tools and workpieces and resulting thermo-elastic deformations need to be addressed. Instead of cooling

Figure 1: The research data life-cycle (DLC) of the RWTH as defined by Politze [17]. Each of the six steps covers distinctive and explicit actions to be undertaken in research data management to achieve successful, transparent, and sustainable research.

whole shopfloor environments and machines, compensating the deviation induced by temperature changes by controlled process adaption can potentially reduce the carbon footprint of machining processes. Hence, the term *"Virtual Climatization"* was coined to signify model-based error prediction over energy-intensive error prevention [18].

To train, test, or validate such models, experimentally acquired datasets as described in Section 5.2 are a base requirement. Equipping a machine tool with additional sensors, using a laser tracker, and running a machine tool continuously for weeks is labor and hardware-intensive and, hence, expensive. Thus, such datasets cannot easily be reacquired and should be (re-)used as efficiently and effectively as possible.In the experimental setup involving multiple actors and sensors, managing specific data outputs with distinct units and uncertainties through individual APIs poses a challenge. Ensuring consistent data access and interpretation is complex, which can be mitigated by standardizing data formats for streamlined handling. Once the data is accessed, the focus shifts to data transfer. Data should be made available directly for supervising experiments or other uses and should be stored persistently. To facilitate such experiments, the Sensor Interfacing Language (SOIL), a domain-specific language (DSL) for interoperable data models, has been developed to simplify integrating new sensors in a data infrastructure with automated, time series database storage [19], [20], [21]. The development, impact, and contribution of SOIL is explained in detail in the next chapters.

3 DLC - Step 1: Planning

Starting into the data life-cycle, the first step is the "planning" phase. In this chapter, we consolidate multiple solutions that support the creation of Data Management Plans (DMPs) and consider which tool, together with which questionnaire, is sufficient and well-suited for our research activity. Before deciding for Research Data Management Organizer (RDMO) as DMP tool, we give a recap on the project history and why DMPs have not always been a part of the research activity. Because *Virtual Climatization* is funded by German Research Foundation (DFG), we decided for the DFG-based NFDI4Ing questionnaire in RDMO.

3.1 Best practice

For planning a research project or activity, the current state of the art is the usage of a Data Management Plan. DMPs are not only a requirement of funding organizations but a powerful source of information within and beyond research projects [22]. On top of supporting RDM, DMPs directly facilitate the execution of the research. By describing the goals and methods of data acquisition and analysis, DMPs offer a valuable resort to find needed information. A DMP also helps conclude research activities by giving an overview of the tasks performed [22].

The contents of DMPs can vary harshly from one another [23], [24]. Typically, all DMPs contain administrative information like the project name, data originator, and funding program along with a project description [23], [25]. However, the core is the information about the research data. For example, the DMP should contain information on how the data is produced, what types and formats are used, and an estimation of the expected data volume [23], [25]. As the ”*ultimate goal of FAIR*” [26], the reuse of data has to be taken into consideration. On the one hand, reuse of existing data should be planned, but on the other hand, the FAIRness of data [6] produced later has to be laid out. How, where, when, and what data will be archived and shared? What licenses will be applied? Who will be responsible for the long-term data? What legal aspects and costs have to be considered? [25]

DMPs reveal maximum benefit if they are not treated as once and initially written and then fixed but rather as a living document, which shall be updated throughout the research activity [5], [27]. However, this continuous curating of a DMP can be time-consuming and is often perceived as a hurdle and additional effort [27], [28]. Therefore, several tools are available supporting researchers in this task with automation mechanisms and predefined questionnaries, such as DMPonline [29], DMPTool [30], DS-Wizard [31], and DMP OPIDoR [32]. Among the available tools, the National Research Data Infrastructure for the Engineering Sciences (NFDI4Ing) suggests using the questionnaries of the Research Data Management Organizer (RDMO) for Data Management (DM) planning [33]. Moreover, the RDMO tool allows one to share the work and collaborate with researchers from different institutes and is therefore used in our research.

3.2 Planning in practice

The origin of the experiments on the use-case of *Virtual Climatization* dates back to around 2015, when RDM was not in focus in engineering science as today, whereas it never was less important. Thus, planning of the research data management has been initially conducted less formalized and needed to be done without the support of proper tools, such as RDMO. To align the research on *Virtual Climatization* with state-of-the-art best-practice suggestions, the research data management planning has been done most recently using (newly) available tools. In doing so, the actions undertaken in the last years have been recapitulated, decisions and measures have been critically reflected, and the frame for future research activities has been set.

The success of the research activity *Virtual Climatization* can be significantly attributed to the implicit application of research data management principles. One example of this is the development of a standardized data description model. Retrospectively, choosing an existing model would have supported interoperability, but it failed due to the lack of appropriate models. Initially, the development of the data model and infrastructure was driven by the need for standard formats and interfaces, which resulted in the usage of JSON for the structured data representation. However, the description model has been newly developed and, thus, is now only partially compatible with widely used models, underlining the consideration of interoperability already in the planning phase. Nevertheless, the resulting *Sensor Interfacing Language* has become an influential contribution on its own due to its abstraction power and extensive reuse within the institute, such as by the *Digital Twin Pipeline* [34].

The publication of source code and datasets was not explicitly foreseen from the beginning. Although the usage of open-source libraries for implementing software artifacts has always been one of the top premises, not all software implemented can be easily published now. One reason is joint work on some parts of the software with foreign researchers where concerns of copyright and licensing for publishing have not been properly targeted. As a result, this source code cannot be published. Incidents like these motivated us to carefully plan future research activities of *Virtual Climatization* and prevent such issues by creating a proper Data Management Plan.

According to the best practices, we are using RDMO for the creation and maintenance of our DMP. As the DFG is funding the related projects of this paper, the NFDI4Ing template questionnaire for the DMP was chosen. This template is a questionnaire corresponding to the DFG checklist for handling research data [35], [36], which in turn is based on the ”practical guide to the international alignment of research data management”, originating to 2019 [10]. The possibility to add different datasets and describe them individually is important in this research activity and engineering in general. This way, it can clearly distinguish between certain parts of the research activity and related data. The DMP of our research activity has been published. [37].

4 DLC - Step 2: Production

Research data production is inherently diverse and individual, and due to the high heterogeneity of acquired research data, it imposes standardization challenges. We established a standardized data description model based on a few best practice suggestions. We implemented a toolchain for automating data production, reducing effort in subsequent phases, like analysis and storage.

4.1 Best Practice

Among all the steps of the research data life-cycle, the production and analysis phases are the least bespoke in the literature, as these two phases are the most individual regarding the tasks conducted. The way production is implemented and executed is mainly shaped by the executed experiments and the involved tools and devices. As these implemented tools, devices, and methods are highly heterogeneous, the experiments are very specific. Therefore, a generalized and standardized way of handling data production is hard to define. Nevertheless, there are general suggestions for structuring the data production phase, which positively impact the outcome and the reusability of the acquired data.

First, it is vital to target the demand for data description and collection of metadata as soon as possible [38]. This diminishes the effort for data harmonization and documentation in later life-cycle phases like analysis and archiving [39]. If the data is analyzed (or archived) after a long interval of weeks, months, or even years, even the researchers who originally acquired the data might have difficulties remembering all relevant details. This reduces the reusability of data significantly. Second, the curation of the data and metadata should already start in parallel to data production [39]. Third, standardizing data models at the very beginning allows for automating tasks in the later phases. For instance, if data is automatically transferred to a standardized, interoperable model and format, this can directly be used for archiving and re-using the data. Analogously, data analysis algorithms can be implemented against the standard data model, so the data parser must only be implemented once in advance. Moreover, the standardized modeling allows for the implementation and usage of middleware to check data quality automatically.

4.2 Production in practice

Within the research activity *Virtual Climatization*, the tools and software were developed to reduce the effort for the subsequent analysis, archiving, and reuse. Moreover, the additional effort for changing the experimental setup and replacing measuring equipment should be minimized. The success of this idea has been achieved by employing five fundamental rules and approaches aligned with the *FAIR Guiding Principles* [6]:

1. Definition of a minimal, standardized meta-model for all measurement data and the interfaces of the measurement devices: The SOIL as proposed by Bodenbenner et al. [20].
2. Usage of open and established standards for data transmission, like MQTT and HTTP as protocol and JSON as data format.
3. Reuse of open-source software tools and libraries wherever possible, such as RabbitMQ [40], InfluxDB [41] or Grafana [42].
4. In case of self-developed software artifacts, minimizing the reuse effort for others by providing them as ready-to-use libraries [43], [44].
5. For data and software code, open and interoperable file formats have been chosen.

4.2.1 Standardized data representation via SOIL

The development of the standardized data model started in 2018 [45] The SOIL has been developed on top of this meta-model [20], [21]. SOIL provides a meta-model for describing the capabilities of measuring devices, which simultaneously specifies the data structure for measurement data and device-specific meta-information, such as configuration or calibration. SOIL enforces vital metadata such as the physical unit, data type, name, description, timestamp, and measurement uncertainty for all measurements sent via the measuring device interface, see Figure 2. The simple meta-model can also be mapped to other standards and models, which increases interoperability. The integration of semantic definitions is currently being developed and will allow seamless integration with archiving tools (such as Coscine [46]) in the future.

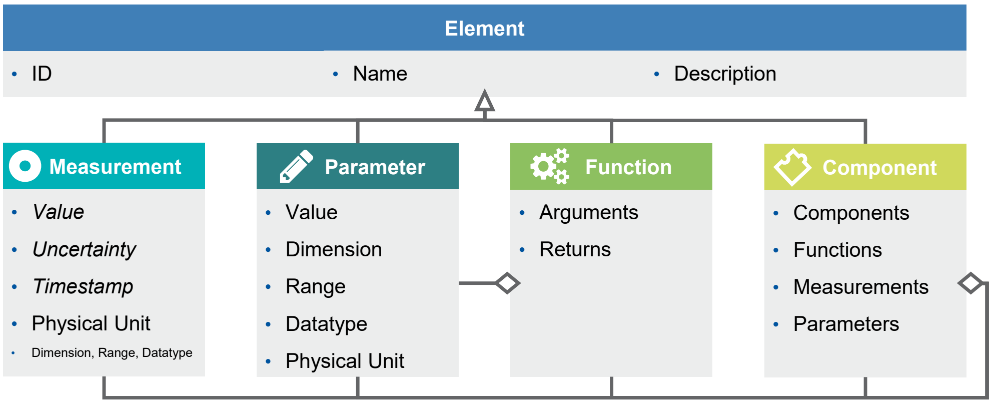

Figure 2: Simplified UML diagram of the meta-model of the SOIL showing the four types of elements of the language and the attributes of each data element, which define the meta information stored for each data point.

Using the SOIL, the definition of the data structure of a measuring device goes hand in hand with the development of the communication interface for that device. The generalized meta-model is used to define a device-specific model via a Web UI from which an implementation template is generated using model-based software development. Moreover, the database structure can be derived automatically from the model, reducing new devices' integration time. The data structures for all measurement devices of the use-case *Virtual Climatization* have been specified using the SOIL. After that, all other parts of the pipeline work out of the box, as described in the next section.

4.2.2 Light-weight data production pipelines

Following the rules 2-4 presented above, we implemented a lightweight pipeline of open-source tools for acquiring, transmitting, and storing the measurement data of experiments. The pipeline currently consists of five components:

- **Measurement devices** measuring, digitizing, and offering the data to consumer clients. Devices are equipped with a REST-API via an HTTP interface and an MQTT interface for publishing data.
- **RabbitMQ** [40] for conveying the data from publisher to subscribers using the MQTT protocol. Data acquisition and consumption decoupling allows clients to perform additional analysis or processing steps easily.
- **InfluxDB** [41] is a time-series database for persistently storing the measurement data.
- **MQTT2Influx-Bridge** [47] is a self-developed Python software deriving the database and table structure from incoming an MQTT message based on its topic and message body. The latter is standardized using the meta-model of the SOIL for all measurement data.

- **Grafana** [42] serves as Web-UI for low-barrier visualization of the data stored in the InfluxDB.

Figure 3 illustrates the complete data infrastructure built upon these components. Available open-source tools have been used for transmission, storage, and visualization. Different libraries and software packages have been developed and published to seamlessly integrate the SOIL-conformed data models. Various libraries have been implemented in C++ and Python to distribute the measurements, which are used to wrap the manufacturer-specific APIs for the laser tracker, the machine tool, and the temperature sensors, and transform the proprietary data models into SOIL-compliant ones. The libraries include fully implemented HTTP and MQTT interfaces. Hence, researchers only need to deal with internal device logic. Modularization is used, so components like the MQTT-Publisher can be replaced with a different library or protocol for Publish/Subscribe without touching other components like the HTTP Server.

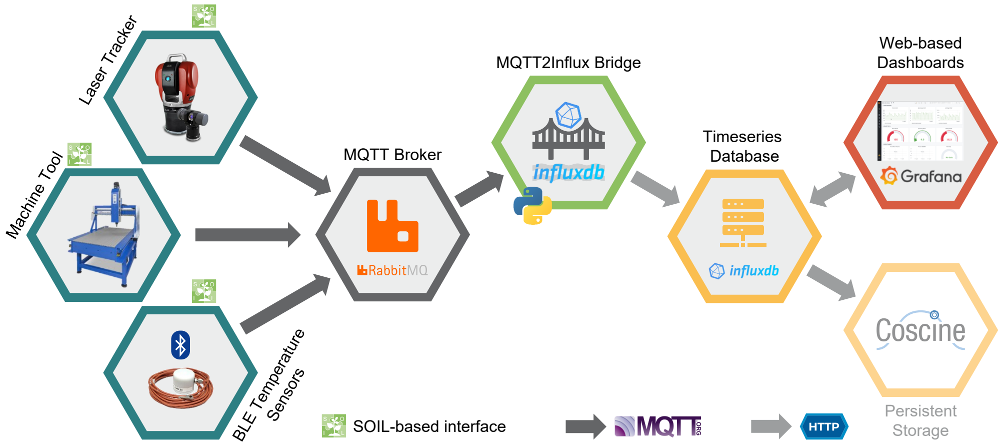

Figure 3: Infrastructure for automated data collection, analysis, storage, and access developed and used in the research activity *Virtual Climatization*. Based on this infrastructure, the thermal machine tool error dataset (see Section 5.2) is collected with the three devices depicted in the left part of the figure. The colors indicate the role of the individual components according to steps of the research data life-cycle, see Figure 1. An automated data upload to Coscine is planned.

To automatically configure the ingestion process of the acquired data into the InfluxDB, a small middleware called *MQTT2Influx-Bridge*, further only called *bridge*, has been implemented [47]. The bridge subscribes to configured MQTT-topics of the measuring devices and automatically transforms SOIL-compliant messages into a representation that can be stored in an InfluxDB database. The database and table structures are derived from the metadata contained in the message and MQTT-topic of the message. As a positive side effect, the data is inherently checked for completeness and conformity with the meta-model so that only well-documented data is persisted.

Whereas the software of the measuring devices is executed on edge devices, the MQTT broker, MQTT2Influx-Bridge, and UI are hosted on virtual machines as containerized services, which expose the tools publicly within the network of the RWTH. According to the best practice for research software management, the deployment has been automated using the CI/CD functionality of GitLab and GitLab-Runners. This enables rapid deployment in similar environments and easy re-deployment in case the configurations are updated.

5 DLC - Step 3: Analysis

To make full use of information gathered in the production stage, data is analyzed for patterns, trends, relationships, or possible conclusions in the third step of the research data life-cycle. Analogously with data production, this step is highly individual, so widely applicable best practices explicitly addressing analysis are sparse. Following the general principles of good scientific practice, our data analysis and process of gaining scientific knowledge have been documented to be as reproducible and verifiable as possible.

5.1 Best Practice

The analysis produces new (aggregated) data, so the two steps *production* and *analysis* are highly coupled and can not always be separated. Thus, most best practice suggestions for production also address analysis.

While there are numerous ways to analyze data using various tools, key requirements include that analysis steps should be reproducible and verifiable [4]. With new incoming data, analyses can easily expand and be verified. All steps used in the analysis process are comprehensible and stored for future reference. Interim results and algorithms used during the analysis process are also important to store for transparency. While not applicable for all analyses, the FACT principles (Fairness, Accuracy, Confidentiality, and Transparency) are an additional guideline [48], [49].

The data analysis is always executed with the subject to formulate or answer research questions, which differ at least slightly and imply the execution of different analysis methods. Thus, best practices for the analysis step are typically context and domain-specific. Schmitt et al. [4] identified seven different archetypes of researchers with respective research methods for the engineering domain. Nevertheless, general best practices and guidelines for responsible research, like comprehensive and extensive documentation, usage of open and interoperable data formats, etc., also apply to all activities in the analysis phase.

5.2 Analysis in practice

Within the research activity *Virtual Climatization*, all analysis steps are carried out on the data stored in an InfluxDB (see Sections 4 and 3). An exemplary dataset contains experimental data used to derive a thermal machine tool error model. It includes data sources that capture relevant information about the machine tool's thermal behavior and its corresponding error (see Figure 4). The dataset has been acquired by moving a machine tool with several fixed paths a hundred times for each run [50]. By storing the measurement data in the SOIL-compliant data format, cf. Section 4.2.1, in the database and the nominal values as GCode we use domain-specific interoperable standards. In detail, the dataset contains the following three main types of data points (see Figure 4):

Machine Tool Position The axes positions (X,Y,Z) are output in real-time. The GCode defines the nominal machine trajectory and contains 924 standstill points.

Laser Tracker Laser trackers and Spherically Mounted Retroreflectors (SMRs) are precise measurement tools used to track the actual position of machine tools. By attaching an SMR to the machine tool's Tool Centre Point (TCP) and using a laser tracker, a highly accurate measurement of the actual stand still position of the machine tool is obtained. This information provides ground truth data for training of the model and validation of the model's predictions. The laser tracker measurements in the TCP serve as a reference for quantifying the thermal errors present in the machine tool.

Temperature Data Temperature sensors placed strategically within the machine tool provide real-time temperature measurements at different locations, e.g., of the machine bed, all axes, and the machine's frame.

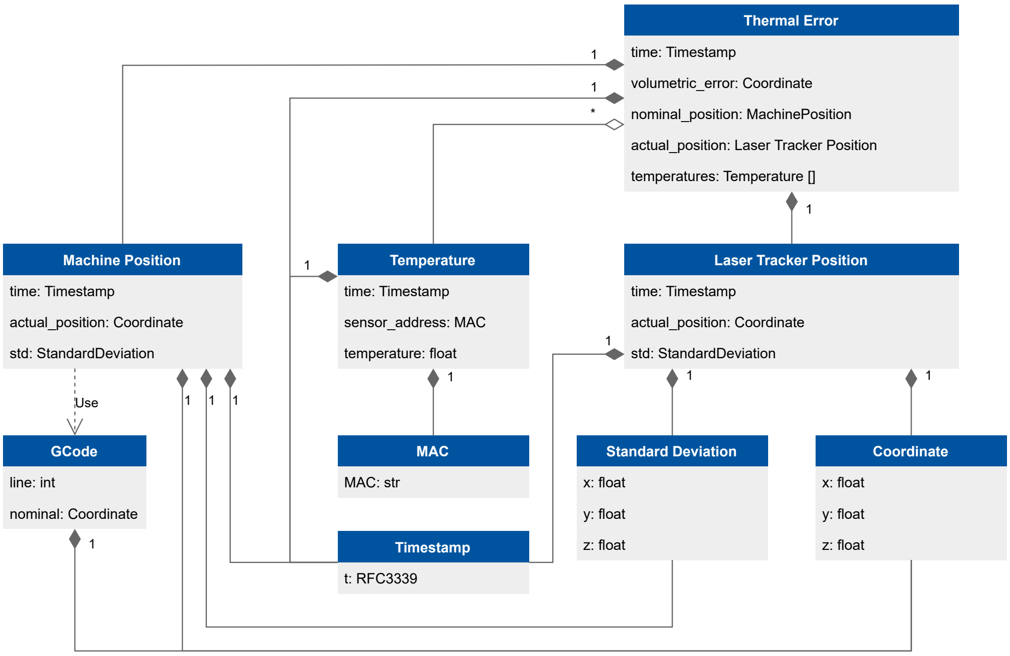

Figure 4: *Virtual Climatization* dataset details represented as UML [51]. Core inputs for the analysis obtained in the experiments are the nominal *Machine Position*, the measured *Laser Tracker Position*, and the *Temperature* data. The result of the data analysis is the *Thermal Error* of the investigated machine tool.

Figure 5 shows the respective analysis steps. For all steps, intermediary results are stored (temporarily). First, in machine position data and laser tracker data, the standstill points need to be found, and the standstill points need to be matched to nominal points. Because the measurements of the laser tracker are initially in the local coordinate system of the tracker, the laser tracker points must be transformed to machine coordinate systems using a rigid fit algorithm [52], [53]. Temperature data is not necessarily recorded in the same interval and at the same point in time. Hence, temperature data for all sensors is interpolated for each position. All runs of the GCode are separated and numbered. The first run and the respective mean temperatures at all sensors are references. Thus, the overall volumetric error at the reference state can be separated into a static part and a transient part (depending on the temperature difference relative to the reference state). [50]

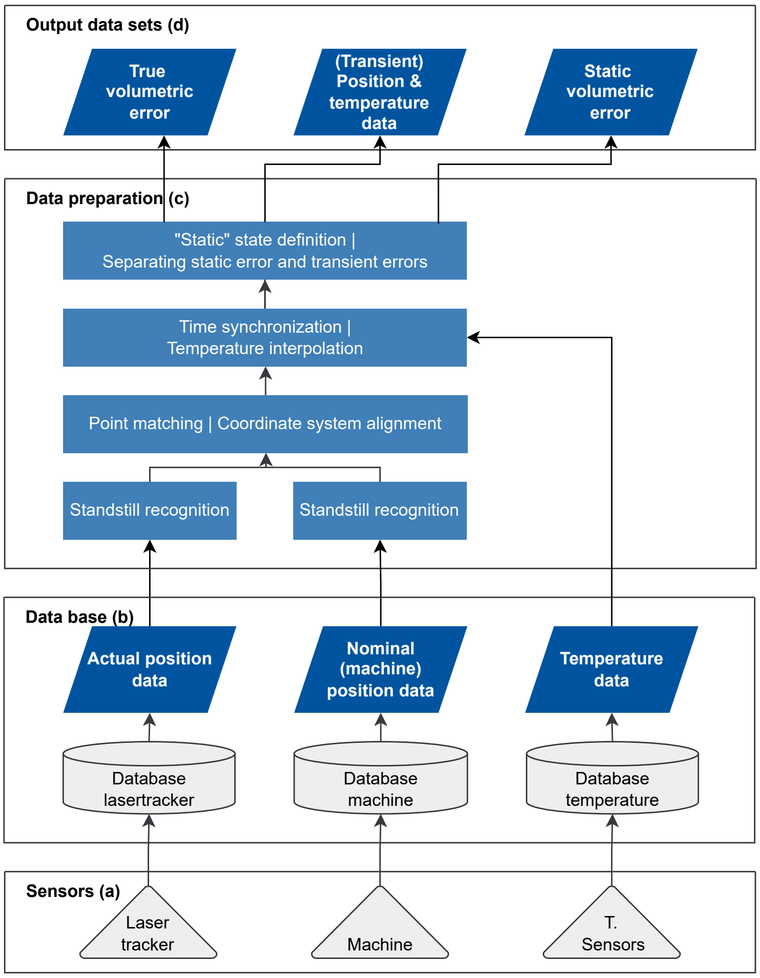

Figure 5: Data set analysis steps for a *Virtual Climatization* dataset summarized [51]. **(a)** Recorded and stored raw data **(b)** is combined and analyzed **(c)** using machine learning models to finally yield the thermal error at machine position and temperature state **(d)**. Instead of training and developing machine learning models, existing models and methods can be validated with this pipeline.

For implementation of these individual analysis steps we reused open-source libraries for data handling and machine learning tasks. Widely applied open-source libraries supported by a large community increase the reusability, reliability, and efficiency of the implemented software artifacts, as they are usually well-documented, optimized, and tested. The raw dataset is saved from the database to RWTH's GitLab instance using Git LFS. Analyses are conducted in Python using Luigi [54] as a pipeline tool and the pipeline source code is stored in a repository. Intermediate results are stored locally on the machine running the pipeline. The resulting dataset is again saved to a GitLab repository using LFS and then manually transferred to Coscine. In the future, data should be stored in Coscine [55] directly instead.

6 DLC - Step 4: Storage

After and alongside with the production and analysis of the data, the next step addresses the storage of the acquired and generated data, including metadata enrichment. In this research activity, the archiving aspect has been tackled later in time. Beforehand, data storage was limited to daily backups and manual saving of data to institutional servers. As a result, the focus of this chapter lies in the archiving aspect of storage and long-term availability of data and software and its alignment with the scientific best practice.

6.1 Best Practice

The archiving of collected and analyzed data is crucial to make the data *FAIR* for the long term. Therefore, archiving covers multiple aspects of research data management: the resilience to technical failures of storage systems, the persistent and long-term storage of the data, and preparation for data access and reuse. To deal with possible failure of components, e.g. a PC malfunction or similar incidents, sufficient backup needs to be applied to the data and source code.

Additionally, data needs to be archived in a way that makes it available and reusable. Otherwise, the archived data would only generate costs, not benefits. To achieve this, the aforementioned collection of metadata and links to it are required as required by the *FAIR Guiding Principles*. This should take place ideally during the production and analysis phases already [56]. However, this practice is not always integrated into researchers' everyday work, eventually causing the loss of important and valuable metadata [46]. The result of the metadata description *should* be that all collaborating researchers can find and reuse data within the research project or activity.

Archiving is not only key for long-term availability and reusability but is also often mandated by funding organizations and universities. In the *Code of Conduct of the Guidelines for Safeguarding Good Research Practice* archiving of data is specified, expanding this demand. It states that research data shall be archived "for a period of ten years at the institution where the data were produced or in cross-location repositories" [57]. However, there are diverging recommendations. While some state that a period of five to ten years is sufficient [58], others claim that datasets have been reused which date back ten to fifty years [59].

Using an institutional repository for data storage depends on the specific research project, as institutions have their own or external archiving solutions. Notably, 65.2% of repositories on re3data are discipline-specific, while only 24.5% are institutional [60]. Institutional repositories must ensure data and metadata storage for ten years. Alternatively, using a broader repository can reach a wider audience than local ones. Currently, there are no domain-specific repositories for mechanical engineering or production technology in search engines like FAIRsharing or re3data. FAIRsharing lists six repositories for mechanical engineering and one for production engineering [61], [62], none being domain-specific but rather general "catch-all" repositories. re3data lists six acoustics-related repositories [63]. Thus, a generic repository would be necessary for engineering data (cf. Section 7).

As a conclusion to the best practice, the key elements of best practice archiving are a minimum storage time of the research data for ten years in a FAIR way explicitly including the usage of open and standardized data formats. The data does not need to be publicly accessible by default, but researchers should be able to find out that there is existing data, enabling possible reusers to request access. The repository in which the data is stored can be either a repository located at their university or an external one.

6.2 Storage in practice

During the active research phases - production and analysis - of the activity *Virtual Climatization* we store the measurement data in an InfluxDB time-series database, hosted at a server of the institute, cf. Section 4.2. The stored data is easily accessible, browsable, and retrievable for researchers at our institute via a Web UI. The server is registered in the backup systems of RWTH, such that the data is stored in the backup system once per day. The source code was stored in the RWTH's GitLab repository, which in itself is already backed up.

According to the best practice suggestions of the DFG [36] and the RWTH [64], our data need to be archived for a minimum of ten years in a repository located at our university or another fitting repository, cf. Section 6.1. We used a hybrid approach, which addresses the two aspects, long-term persistence, and accessibility, independently from one another. First, in this section, the on-premise solution of the RWTH is portrayed to achieve long-term persistence. Later, in Section 7.2 we will address a more widespread approach, following open access standards.

We archived our data of the research activity *Virtual Climatization* with Coscine, the on-premise archiving toll of RWTH. Coscine implements a storage and archiving solution by closely following the FAIR principles [6], [46]. The stored data must must be enriched with metadata, which can be customized using the *Metadata Profile Service*, formerly called *Application Profile Generator* [65]. The tool aims to integrate FAIR handling of research data into researchers' daily routines. The creation of application profiles has been partially determined in the planning and has been supplemented by additional metadata schemas for specific datasets. For example, our code on GitLab is equipped with metadata via the ”Radar” application profile [66]. Coscine offers a broad application programming interface (API) that have been used for (automatic) data and metadata uploads, reducing effort and enhancing usability. The data is automatically linked to the attached data with partial enforcement of metadata usage. Thus, our metadata is searchable resulting in easy findability. Due to Coscine and the settings we chose our research data is publicly findable. However, due to the restrictions of Coscine is not openly accessible. This drawback will hence be addressed in the following chapter 7.

7 DLC - Step 5: Access

While Archiving of data and software is an essential key to long-term availability, it does not necessarily include a broad distribution of the archived files. The possible usage and accessibility are therefore determined in the ”Access” phase of the data life-cycle. Within our interpretation of the last step ”Re-Use”, the licensing of published data is also included in the ”Access” phase.

7.1 Best Practice

The last step of the data life-cycle, which can be actively sculpted by the authors and data owners, is the access step. By definition of the Budapest Open Access Declaration in 2002,

”open access” is literature that is freely and publicly available on the internet and has only very limited restrictions regarding re-use. The underlying idea is that open access removes barriers to literature and accelerates research, as well as fosters education across the whole society. [67]

While the concept of open access is not new, the general attitude towards it has shifted. Beginning with the Budapest Open Access Declaration [67], open access spread, affecting research. While it was critically reviewed, whether it could fulfill the demands of scientific publication standards [68], it is now a globally accepted standard and best practice, fostered and driven by initiatives and funding organizations. [69] For instance, the German Federal Ministry of Education and Research (BMBF) has proclaimed the following statement, ”Open access is to become the standard for scientific publishing in Germany. The idea of open access to scientific publications was developed by the scientific community and is recognized there” [70]. In addition to that, the RDM guidelines of RWTH state, that the data must always stay freely available for scientific re-use, even if publication and usage rights might be transferred to third parties [64].

The current best practice for licensing published data is CC-BY 4.0. This results from the open access policy of Creative Commons (CC), which aims to be ”free, simple, and standardized [in a] way to grant copyright permissions for creative and academic works; ensure proper attribution; and allow others to copy, distribute, and make use of those works” [71]. However, this license it is not recommended for software [72], but other different kinds of dedicated software licenses exist. The more open standard are a reciprocal license (often referred to as ”copyleft”), which highly influences the remaining copyright of the creator. The alternative to reciprocal licenses is permissive licenses, such as the MIT license [73]. Not only do these ensure more rights for creators of the software, but permissive licenses also offer higher compatibility with other licenses and better interoperability of the licensed software. [74]

7.2 Access in practice

Following these best practice guidelines, two aspects must be covered to guarantee access to all artifacts of the research activity *Virtual Climatization*: First, an appropriate repository must be chosen that supports open access. Second, all published artifacts must be properly licensed to allow for re-use in the next phase of the research data life-cycle.

The RWTH suggests to use the RWTH Publications Service of the University Library [75], [76], which allows for open access and to connect a Coscine project archiving the research data to the record on RWTH publications, cf. 6.2. This connection between the two records further enhances the research data's findability [77]. RWTH's Open-Access-Policy follows best practices in research data management. Therefore, we are following the gold standard of open access in this research activity, while also using the infrastructure provided by the RWTH as far as possible. This implies this paper will also be added to RWTH Publications, allowing for findability and making it openly accessible. In addition, the previously mentioned connection of the RWTH Publications record to the Coscine project will be registered. However, using Coscine the open accessibility is partially limited (cf. Section 9).

Thus, using RWTH Publications for giving access to the publication and using Coscine for giving access to the data, we achieved open access for the publication but not for the data. For an immediate solution, we uploaded the data to a different repository, which follows the standards of open access. As there is currently no suitable discipline-specific repository for mechanical engineering or production technology [61], [62], [63], we published our data via Zenodo. Zenodo allows for open access along with good findability due to the usage of Digital Object Identifiers (DOIs). However, because Zenodo is restricted to one license per record, we store the source code and the data in two different records to choose individual appropriate licenses for both. This also allows for updates on the software without triggering a new version for the related data.

We support widespread open access using a CC-BY 4.0 license for the acquired data. For the software, we use the MIT License, which is easily applicable and understandable [73]. This license has a much shorter legal text than for example the GNU General Public License v3.0 or later [78], [79]. As a result, it comes with fewer restrictions while allowing for almost unrestricted reuse of the licensed software. Additionally, it does not exclude the change to a different license at a later point [80]. Following the recommendations of [73] and [80], we consider the MIT license to be the best practice for all software of the research activity *Virtual Climatization*.

Our resulting access solution follows the structure as shown in Figure 6. We consider this procedure as best practice to fulfill the requirements and demands of the DFG and the RWTH and, most importantly, to ensure the FAIRness of the data and software. While we might be overshooting the target with this approach, it safeguards good scientific practice and facilitates the reuse of our research results in the last phase of the research data life-cycle.

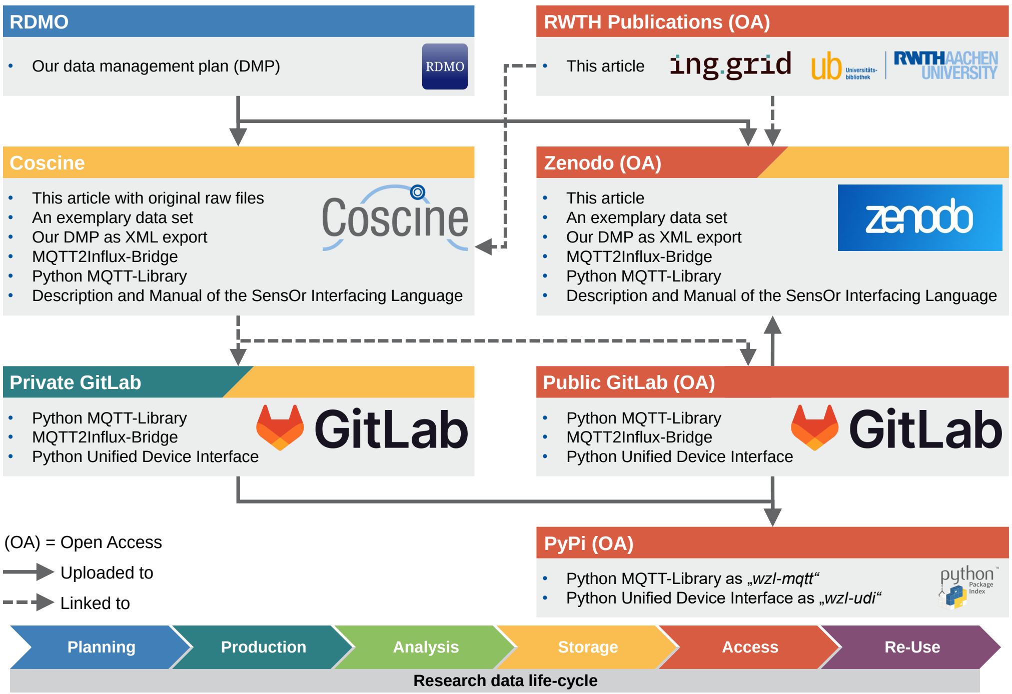

Figure 6: Visualisation of the complete archiving and access structure showing which artifact is stored and published via which platform. As shown at the bottom, the coloring of the different platforms picks up the coloring of the research life-cycle, cf. Figure 1, and indicates which platform is primarily used in which phase. The arrows indicate the links between the different platforms.

8 DLC - Step 6: Re-Use

Two perspectives must be distinguished in step six of the DLC. First, the perspective of the data creator and publisher who aims to enable reuse which mainly results from sticking to best practice suggestions in the previous phases of RDM. The second perspective covers the actual reuse and concerns the code of conduct when reusing foreign (or own) research data. Moreover, the reuse of research data has two forms, internal reuse, by the original researchers or within the institute, and external reuse by third parties and foreign researchers outside of the institute. While Section 8.1, *Best Practice*, focuses on the second perspective, reusing research data, Section 8.2, *Reuse in practice*, discusses the benefits of internal reuse and summarises the RDM achievements of *Virtual Climatization*.

8.1 Best Practice

The reuse of data and source code is the key target of all aforementioned steps. Without the reuse of at least parts of the data and source code generated in the research activity, all RDM efforts are to some extent pointless. Therefore we would like to encourage anyone to reuse our data and software. This will require a cultural change. The ”Not Invented Here”-Syndrome [81] might still cause researchers - or supervisors - to acknowledge research data originating from anywhere else than their institute or project. This attitude is about to vanish but persists within many. A cultural change towards data reuse is imminent and is ongoing for every researcher. To facilitate the reuse of data and allow everyone to easily and without risk use our data, we empathized on open access as best practice for access in chapter 7. In this section, we will therefore explain how to reuse our data and source code in your project.

As the data is licensed under a CC-BY 4.0 license [82], [83], the reuse of data can be realized almost without restrictions. Anyone is allowed to copy, redistribute, transform, and extend the data. There are only two conditions to this:

1. **Attribution:**
   You have to give credit to us with indication, if changes have been applied to the data or not as well as a license notice of the data.
   Do: Cite the Zenodo repositories of the data you are reusing via their DOI. Licenses are included there.
   Don't: Present the data as your own.

2. **Do not apply additional restrictions:**
   You may not try to limit the accessibility to the original data.
   Do: Use the data as you like and publish your results under any license you like, but give credit to the original dataset as mentioned in 1.
   Don't: Try to restrict data accessibility by not providing the link to the CC-BY 4.0 licensed dataset. (Would also violate the first condition.)

To reuse our source code, the reuser has to take into consideration the MIT license, which is even less restrictive than the CC-BY 4.0 license. [78] Everyone can ”use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the” [78] source code. The only condition is to include the license notice in your code if you reuse ours. While it is not obligated to do so, we highly suggest also citing the Zenodo repository of our source code via its DOI.

8.2 Re-Use in practice

To facilitate internal reuse, the developed software artifacts are managed via GitLab instance of the RWTH. We created a group in Gitlab and all colleagues of the department are members of this group. Based on this centralized and cooperatively used software management and versioning system, it is ensured that all colleagues can easily access and reuse the developed software. Furthermore, software with high reuse potential has been published open-source and licensed under the terms and conditions of the MIT License. To reduce the reusability effort our Python source code is packed as a library and published to PyPI making the libraries very easy to install[43], [44]. Within the department, the MQTT library is reused extensively, as it provides a simple and minimal interface. The solution is tailored to our infrastructure so that the configuration effort to connect to our MQTT broker is low. Accompanied by comprehensive documentation and sample scripts demonstrating the usage of the library, the source code is used by most colleagues in the department for their measurements and experiments.

Moreover, the meta-model of the SOIL has been adapted by colleagues for their research. One example is the *Digital Twin Pipeline* for definition and generation of communication interfaces for flexible assembly systems [34].

The developed tool stack, presented in Section 4.2.2 and Figure 3, has been used for all experiments related to the research activity *Virtual Climatization* in the last few years. The measurement data produced, analyzed, and stored using the tool stack form the basis of one successfully finished [50] and two ongoing doctoral theses. Moreover, the establishment of the tool stack, in particular, the standardization of the (meta) data models and communication interfaces is the subject of two other doctoral theses of which one is already finished [84]. This underlines that RDM and in particular reuse of prior results is the foundation for successful and responsible research.

9 Discussion

To reconsider all RDM steps and their results as well as shortcomings, this chapter is divided into three parts. First, we describe our successes and the benefits generated with our RDM and how we strived to make our data and source code reusable. Secondly, we cut down our achievements by listing the caveats and problems we see in our RDM. Lastly, we discuss the RDM beyond our project, where future improvements of tools could help researchers to implement RDM in their work.

9.1 Review of our achievements based on RDM

The implementation of RDM has significantly advanced our research on *Virtual Climatization* by standardizing data descriptions through the meta-model of SOIL, facilitating the use of common algorithms and enabling quick creation of interoperable measurement interfaces. This standardization enhances archiving capabilities with tools like Coscine, allowing for effective cross-checks across experiments. By managing software artifacts in RWTH's GitLab and publishing open-source code, we have promoted accessibility and collaboration among colleagues. While our RDM practices have resulted in well-documented and reusable data and software, we recognize areas for improvement, such as better integration with existing ontologies to enhance interoperability. Additionally, early adoption of RDMO and Coscine could have streamlined archiving efforts, highlighting the need for dedicated data stewardship to manage datasets effectively after authors leave the institute.

9.2 Remarks on current best practices and used tools

To access this paper's full potential, we would like to not only take a look at our work and how it could have been improved but would also like to take this opportunity to reflect on the external factors, which influenced the research data management in this research activity

The RWTH data life-cycle [17] has been useful for structuring research data management but should not be viewed as a strict roadmap for an entire research project. It focuses primarily on handling data at various stages — planning, creation, analysis, storage, publication, and reuse — rather than guiding an entire project from start to finish. Continuous access and storage options like Coscine should be utilized throughout the project life-cycle for backup and collaborative purposes.

The current RDMO questionnaire of NFDI4Ing [24], [35] based on the DFG checklist [36] is seen as counterintuitive due to its structure, which mixes overarching project-related questions with dataset-specific ones. A restructuring could separate these categories to improve usability and readability by addressing overarching questions first before diving into dataset-specific details. This feedback stems from applying the questionnaire in complex projects like *Virtual Climatization*, with discussions initiated to enhance service usability without altering core questions aligned with DFG's checklist requirements.

Currently, while metadata about projects is publicly findable in Coscine, accessing actual files requires an additional request process. Enabling open access to files directly through Coscine would significantly improve its utility for researchers by streamlining accessibility and enhancing collaborative capabilities throughout the research process. A possible workaround could be the establishment of a data steward position for each institute as a long-term employment. The data steward owns all the research data acquired in all research activities within the institute. This data steward would be of great benefit for institutes which are not going to be discussed here.

10 Conclusion

Successful, transparent, and comprehensible research requires responsible Research Data Management (RDM), as RDM is one of the foundations of good research, facilitating the sharing of reusable and interpretable research data beyond textual publications. Whereas the need for purposeful RDM is obvious and widely expressed, the realization is oftentimes lacking due to missing incentives, services, and tools that foster or support RDM in practice. To overcome these shortcomings and illustrate the benefits of implemented RDM, we described the research activity *Virtual Climatization* and the important role of RDM for its successful accomplishment. We explained and discussed RDM founded on the research data life-cycle, the current best practices for the individual phases of the life-cycle the implementation of the phases in practice.

Research data management planning addressed all relevant points in the context of this activity. In the production and analysis phase, a central building block of *Virtual Climatization* is a standardized data description based on the SOIL developed at the institute. Based on these interoperable data models the integration of new equipment, and (re-)configuration of experiments and setups require much less effort as data acquisition and storage is widely automated. Moreover, the long-term availability of data and source code generated has been considered. The archiving as well as access have been structured in a manner that ensures availability for a minimum of ten years from now on along with complete open access and findability. Lastly, a strong focus was set on the reusability of the data and source code.

The strategies applied and the activity undertaken by RDM had and still have a large impact on the successful research at our institute. The data description model of the SOIL has also been adopted by colleagues and in other research projects, which underlines its usefulness and the positive impact of providing adequate services and tools for RDM. Nevertheless, some decisions we made years ago (before the wide-spreading of formalized RDM), turned out not being always fully aligned with guidelines and best practices published most recently. We started to adapt these and described how we are aiming to continue our work in the future. Among others, these cover alignment of the meta-model of the SOIL with other state-of-the-art data models and the integration of semantic modeling techniques.

Consequently, we aim to apply RDM and all its steps for current and future projects from the very beginning. Doing so, we strive to provide FAIR research data of all our activities, which ultimately leads to *FAIR research* maximizing its impact. As a step towards this aim, we plan to integrate the connection to Coscine into our data acquisition and processing pipelines to seamlessly integrate archiving into research workflows. Furthermore, lifting our data structures to fully semantic data models will contribute to the FAIRness of our research data. Moreover, we plan to conduct empirical studies on the effectiveness and applicability of our approaches.

We also used several tools and services for RDM, such as RDMO, Coscine, Zenodo, or RWTH Publications. While doing so, we have also considered the usability of the tools and services. Along with the benefits of the used tools, we also discussed potential improvements to these services as well as overarching issues, which could be addressed in the future to enhance RDM. Most importantly, the usability of the services has to be improved and current stand-alone solutions for specific RDM problems have to be connected to a network of solutions that share data amongst them. This will improve the applicability of RDM due to amortization, reducing effort and increasing productivity. In addition, future research and development should include the integration of individual tools into an interconnected toolchain. This would reduce the burden on researchers, who currently have to provide a lot of information redundantly and individually for different tools. The development of interoperable interfaces and APIs for automatic data exchange between different tools such as RDMO, Coscine, RWTH Publications, and Zenodo would be useful.

As a final thought, we would like to empathize the reuse of data. Everyone has to be part of this cultural change. The reuse of our data and source code is the main goal, overarching motivation, and reason for this paper.

11 Roles and contributions

Matthias Bodenbenner: Conceptualization, Methodology, Software, Writing - original draft

Tobias Hamann: Conceptualization, Methodology, Writing - original draft

Mario Moser: Conceptualization, Methodology, Writing - original draft

Mark P. Sanders: Conceptualization, Methodology, Software, Writing - original draft

Dominik Wolfschläger: Writing - review & editing, Supervision, Project administration

Robert H. Schmitt: Writing - review & editing, Supervision, Project administration, Funding acquisition

References

[1] R. Bryant, B. Lavoie, C. Malpas, Rebecca Bryant, Brian Lavoie, and Constance Malpas, *A tour of the Research Data Management (RDM) service space // The Realities of Research Data Management Part One: A Tour of the Research Data Management (RDM) Service Space: The realities of Research Data Management, Part 1*. Dublin, Ohio: OCLC Research, 2017. DOI: 10.25333/C3PG8J. [Online]. Available: https://apo.org.au/sites/default/files/resource-files/2017-03/apo-nid112471.pdf.

[2] S. S. Feger, P. W. Wozniak, L. Lischke, and A. Schmidt, "'yes, i comply!'" *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW2, pp. 1–26, 2020. DOI: 10.1145/3415212.

[3] T. Grynoch, "Implementing research data management services in a canadian context," *Dalhousie Journal of Interdisciplinary Management*, vol. 12, no. 1, 2016, ISSN: 1923-6530. DOI: 10.5931/djim.v12.i1.6458. [Online]. Available: https://ojs.library.dal.ca/djim/article/view/6458/5824.

[4] R. H. Schmitt et al., *Nfdi4ing - the national research data infrastructure for engineering sciences*, 2020. DOI: 10.5281/ZENODO.4015201.

[5] M. Putnings, H. Neuroth, and J. Neumann, *Praxishandbuch Forschungsdatenmanagement*. De Gruyter, 2021, ISBN: 9783110657807. DOI: 10.1515/9783110657807.

[6] M. D. Wilkinson et al., “The fair guiding principles for scientific data management and stewardship,” *Scientific Data*, vol. 3, no. 1, p. 160 018, 2016, ISSN: 2052-4463. DOI: 10.1038/sdata.2016.18.

[7] J. Schöpfel and O. Azeroual, “Rewarding research data management,” in *Companion Proceedings of the Web Conference 2021*, J. Leskovec, Ed., ser. ACM Digital Library, New York,NY,United States: Association for Computing Machinery, 2021, pp. 446–450, ISBN: 9781450383134. DOI: 10.1145/3442442.3451367.

[8] N. Hartl, E. Wössner, and Y. Sure-Vetter, “Nationale forschungsdateninfrastruktur (nfdi),” *Informatik Spektrum*, vol. 44, no. 5, pp. 370–373, 2021, ISSN: 0170-6012. DOI: 10.1007/s00287-021-01392-6.

[9] Shaping Europe’s digital future, *European open science cloud (eosc)*, 2023. Accessed: Jul. 28, 2023. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/open-science-cloud.

[10] Science Europe, *Practical guide to the international alignment of research data management - extended edition*, 2021. DOI: 10.5281/zenodo.4915862. Accessed: Aug. 24, 2023. [Online]. Available: https://doi.org/10.5281/zenodo.4915862.

[11] R. Hanisch, *Nist research data framework (rdaf)*, Gaithersburg, MD, 2023. DOI: 10.6028/NIST.SP.1500-18r1.

[12] ARDC, *About the ardc*, 2022. Accessed: Aug. 24, 2023. [Online]. Available: https://ardc.edu.au/about-us/.

[13] OECD ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *Oecd principles and guidelines for access to research data from public funding*, 2007. Accessed: Aug. 24, 2023. [Online]. Available: http://www.oecd.org/science/sci-tech/38500813.pdf.

[14] S. Nosthoff and J. Baur, *Fdm erklärt – was ist eigentlich ein daten-lebenszyklus? - forschungsdaten – aktuelles und wissenswertes*, 2021. Accessed: Aug. 21, 2023. [Online]. Available: https://blog.rwth-aachen.de/forschungsdaten/2021/04/16/fdm-erklaert-was-ist-eigentlich-ein-daten-lebenszyklus/.

[15] D. Schmitz and M. Politze, “Forschungsdaten managen – bausteine für eine dezentrale, forschungsnahe unterstützung: 76-91 seiten / o-bib. das offene bibliotheksjournal / herausgeber vdb, bd. 5 nr. 3 (2018),” *Das Offene Bibliotheksjournal*, 2018. DOI: 10.5282/o-bib/2018H3S76-91.

[16] S. Higgins, “The dcc curation lifecycle model,” *International Journal of Digital Curation*, vol. 3, no. 1, pp. 134–140, 2008. DOI: 10.2218/ijdc.v3i1.48.

[17] M. Politze, “A reference architecture and implementation enabling data protection in distributed elearning and escience processes,” Dissertation, Rheinisch-Westfälische Technische Hochschule Aachen and Apprimus Verlag, 2019.

[18] D. Emonts, M. P. Sanders, P. Dahlem, M. Bodenbenner, B. Montavon, and H. R. Schmitt, “Virtuelle klimatisierung/virtual climatization,” *wt Werkstattstechnik online*, vol. 111, no. 11-12, pp. 887–892, 2021, ISSN: 1436-4980. DOI: 10.37544/1436-4980-2021-11-12-117.

[19] M. Bodenbenner, “Sensor interfacing language for FAIR sensor services in manufacturing,” Dissertation, RWTH Aachen University, Aachen, 2025. DOI: 10.18154/RWTH-2025-03359.

[20] M. Bodenbenner, M. P. Sanders, B. Montavon, and R. H. Schmitt, “Domain-specific language for sensors in the internet of production,” in *Production at the leading edge of technology*, ser. Lecture Notes in Production Engineering, B.-A. Behrens, A. Brosius, W. Hintze, S. Ihlenfeldt, and J. P. Wulfsberg, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2021, pp. 448–456, ISBN: 978-3-662-62137-0. DOI: 10.1007/978-3-662-62138-7{\textunderscore}45.

[21] M. Bodenbenner, *Description and manual of the sensor interfacing language*, 2023. DOI: 10.5281/ZENODO.7757250.

[22] W. K. Michener, “Ten simple rules for creating a good data management plan,” *PLoS computational biology*, vol. 11, no. 10, e1004525, 2015. DOI: 10.1371/journal.pcbi.1004525.

[23] RDMO, *Github - rdmorganiser/rdmo-catalog*, Aug. 2023. Accessed: Aug. 24, 2023. [Online]. Available: https://github.com/rdmorganiser/rdmo-catalog/tree/master.

[24] GitHub, *Rdmo-catalog/shared/nfdi4ing/nfdi4ing_template at master · rdmorganiser/rdmo-catalog: Commit: Https://github.com/rdmorganiser/rdmo-catalog/commit/42cab05850bff7ae396356* 2023. Accessed: Jul. 31, 2023. [Online]. Available: https://github.com/rdmorganiser/rdmo-catalog/blob/master/shared/nfdi4ing/NFDI4Ing_template/NFDI4Ing_Vorlage.xml.

[25] TU9-FDM, *Datenmanagementplan*, 2020. DOI: 10.5281/zenodo.3974562.

[26] A. Jacobsen et al., “Fair principles: Interpretations and implementation considerations,” *Data Intelligence*, vol. 2, no. 1-2, pp. 10–29, 2020. DOI: 10.1162/dint{\textunderscore}r{\textunderscore}00024.

[27] T. Miksa, S. Oblasser, and A. Rauber, “Automating research data management using machine-actionable data management plans,” *ACM Transactions on Management Information Systems*, vol. 13, no. 2, 2021, ISSN: 2158-656X. DOI: 10.1145/3490396.

[28] S. S. Feger, P. W. Wozniak, L. Lischke, and A. Schmidt, “’yes, i comply!’” *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW2, pp. 1–26, 2020. DOI: 10.1145/3415212.

[29] DCC-Digital Curation Centre, *Dmponline*, 2023. Accessed: Jul. 31, 2023. [Online]. Available: https://dmponline.dcc.ac.uk/.

[30] California Digital Library, *Dmptool*, 2023. Accessed: Jul. 31, 2023. [Online]. Available: https://dmptool.org/.

[31] Data Stewardship Wizard, *Data stewardship wizard*, 2023. Accessed: Jul. 31, 2023. [Online]. Available: https://ds-wizard.org/.

[32] Inist-CNRS, *Dmp opidor*, 2023. Accessed: Jul. 31, 2023. [Online]. Available: https://dmp.opidor.fr/.

[33] I. Mozgova et al., *Product life cycle oriented data management planning with rdmo at the example of research field data*, 2022. DOI: 10.11588/HEIBOOKS.979.C13722.

[34] A. Göppert, L. Grahn, J. Rachner, D. Grunert, S. Hort, and R. H. Schmitt, "Pipeline for ontology-based modeling and automated deployment of digital twins for planning and control of manufacturing systems," *Journal of Intelligent Manufacturing*, vol. 34, no. 5, pp. 2133–2152, 2023, ISSN: 1572-8145. DOI: 10.1007/s10845-021-01860-6. [Online]. Available: https://link.springer.com/article/10.1007/s10845-021-01860-6.

[35] Technische Universität Darmstadt, *Rdmo nfdi4ing - fragenkataloge*, 2023. Accessed: Jul. 31, 2023. [Online]. Available: https://rdmo.nfdi4ing.de/projects/1252/update/catalog/.

[36] DFG, *Umgang mit forschungsdaten: Checkliste für antragstellende zur planung und zur beschreibung des umgangs mit forschungsdaten in forschungsvorhaben*, Dec. 2021. Accessed: Jul. 28, 2023. [Online]. Available: https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf.

[37] M. Bodenbenner, M. P. Sanders, and T. Hamann, *Research data management plan for research activity "virtual climatization"*, 2023. DOI: 10.5281/zenodo.8247455.

[38] J. Dierkes, "4.1 planung, beschreibung und dokumentation von forschungsdaten," in *Praxishandbuch Forschungsdatenmanagement*, ser. De Gruyter Praxishandbuch, M. Putnings, H. Neuroth, and J. Neumann, Eds., Berlin: De Gruyter Saur, 2021, pp. 303–326, ISBN: 9783110657807. DOI: 10.1515/9783110657807-018.

[39] L. Pouchard, "Revisiting the data lifecycle with big data curation," *International Journal of Digital Curation*, vol. 10, no. 2, pp. 176–192, 2015. DOI: 10.2218/ijdc.v10i2.342.

[40] VMware Inc., *Rabbitmq: Easy to use, flexible messaging and streaming*, 2023. Accessed: Aug. 28, 2023. [Online]. Available: https://www.rabbitmq.com/.

[41] InfluxData Inc., *Influxdb | real-time insights at any scale*, 2023. Accessed: Aug. 28, 2023. [Online]. Available: https://www.influxdata.com/influxdb/.

[42] Grafana Labs, *Grafana | query, visualize, alerting observability platform*, 2023. Accessed: Aug. 28, 2023. [Online]. Available: https://grafana.com/grafana/.

[43] M. Bodenbenner, *Wzl-udi | python unified device interface*, PyPI, Ed., 2023. Accessed: Aug. 28, 2023. [Online]. Available: https://pypi.org/project/wzl-udi/.

[44] M. Bodenbenner, *Wzl-mqtt*, PyPI, Ed., 2023. Accessed: Aug. 28, 2023. [Online]. Available: https://pypi.org/project/wzl-mqtt/.

[45] B. L. Montavon, J. P. Dahlem, and R. H. Schmitt, "Fast machine tool calibration using a single laser tracker," in *Laser metrology and machine performance XIII : 13th International Conference and Exhibition on Laser Metrology, Machine Tool, CMM & Robotic Performance : Lamdamap 2019 / organised by: euspen ; edited by Professor Liam Blunt & Dr. Wolfgang Knapp*, Cranfield University Campus, Bedfordshire, UK: euspen, 2019, pp. 203–213. [Online]. Available: https://publications.rwth-aachen.de/record/762794.

[46] I. Lang, *Coscine – fair im datenmeer*, 2022. DOI: 10.5281/ZENODO.7323569.

[47] M. P. Sanders and M. Bodenbenner, *Mqtt2influx-bridge*, 2023. DOI: 10.5281/zenodo.8247483.

[48] Responsible Data Science Initiative, *Responsible data science - fairness - accuracy - confidentiality - transparency*, 2016. Accessed: Aug. 31, 2023. [Online]. Available: https://redasci.org/.

[49] W. M. P. van der Aalst, M. Bichler, and A. Heinzl, "Responsible data science," *Business & Information Systems Engineering*, vol. 59, no. 5, pp. 311–313, 2017, ISSN: 1867-0202. DOI: 10.1007/s12599-017-0487-z.

[50] J. P. Dahlem, "Hybrid modeling of transient volumetric machine tool errors for virtual climatization," Ph.D. dissertation, RWTH Aachen University, 2023. DOI: 10.18154/RWTH-2023-01620.

[51] M. P. Sanders, P. Dahlem, M. Bodenbenner, and R. H. Schmitt, *Research data for research activity "virtual climatization"*, 2023. DOI: 10.5281/ZENODO.8304983.

[52] P. H. Schönemann, "A generalized solution of the orthogonal procrustes problem," *Psychometrika*, vol. 31, no. 1, pp. 1–10, 1966, ISSN: 0033-3123. DOI: 10.1007/BF02289451.

[53] W. Kabsch, "A solution for the best rotation to relate two sets of vectors," *Acta Crystallographica Section A*, vol. 32, no. 5, pp. 922–923, 1976, ISSN: 0567-7394. DOI: 10.1107/S0567739476001873.

[54] Spotify AB, *Luigi*. Accessed: Aug. 31, 2023. [Online]. Available: https://github.com/spotify/luigi.

[55] I. Lang and M. Politze, *Coscine*, Aachen, 2023. Accessed: Jul. 31, 2023. [Online]. Available: https://coscine.rwth-aachen.de/.

[56] B. P. A. Heinrichs, *Automatic extraction of descriptive metadata to promote the usage of rdm tools*, 2022. DOI: 10.18154/RWTH-2022-10071.

[57] Deutsche Forschungsgemeinschaft, *Guidelines for safeguarding good research practice. code of conduct*, 2022. DOI: 10.5281/ZENODO.6472827.

[58] A. Chasapi, F. D'Anna, B. Bösl, and S. Fuchs, *Rdmkit: Preserving*, Aug. 2023. Accessed: Aug. 24, 2023. [Online]. Available: https://rdmkit.elixir-europe.org/preserving.

[59] M. M. Koopman and K. de Jager, "Archiving south african digital research data: How ready are we?" *South African Journal of Science*, vol. 112, no. 7/8, p. 7, 2016. DOI: 10.17159/sajs.2016/20150316.

[60] re3data.org, *Statistics | re3data.org*, Aug. 2023. Accessed: Aug. 24, 2023. [Online]. Available: https://www.re3data.org/metrics.

[61] FAIRsharing, *Fairsharing | search mechanical engineering repositories*, Jul. 2023. Accessed: Aug. 1, 2023. [Online]. Available: https://fairsharing.org/search?recordType=repository&page=1&subjects=mechanical%2520engineering.

[62] FAIRsharing, *Fairsharing | search production technology repositories*, Jul. 2023. Accessed: Aug. 1, 2023. [Online]. Available: https://fairsharing.org/search?recordType=repository&page=1&subjects=production%2520technology.

[63] GFZ German Research Centre For Geosciences et al., *Registry of research data repositories*, 2013. DOI: 10.17616/R3D. [Online]. Available: https://www.re3data.org/search?query=&subjects%5B%5D=40204%20Acoustics&subjects%5B%5D=4%20Engineering%20Sciences.

[64] R. A. University, *Leitlinie zum forschungsdaten-management an der rwth aachen - rwth aachen university - deutsch*, 2016. Accessed: Jan. 22, 2023. [Online]. Available: https://www.rwth-aachen.de/cms/root/Forschung/Forschungsdatenmanagement/~ncfw/Leitlinie-zum-Forschungsdatenmanagement/.

[65] M. Grönewald et al., *Mit aims zu einem metadatenmanagement 4.0 : Faire forschungsdaten benötigen interoperable metadaten*, 2022. DOI: 10.11588/HEIBOOKS.979.C13721.

[66] H. Führ, *Radar application profile*, Oct. 2022. Accessed: Aug. 23, 2023. [Online]. Available: https://purl.org/coscine/ap/radar/.

[67] L. Chan, D. Cuplinskas, M. Eisen, F. Friend, and Y. Genova, *Budapest open access declaration: Budapest open access initiative*, Budapest, Hungary, Feb. 2002. Accessed: Aug. 9, 2023. [Online]. Available: https://www.budapestopenaccessinitiative.org/read/.

[68] U. Müller, *Standards Und Best Practices Im Kontext Von Open Access*. Zenodo, 2017. DOI: 10.5281/zenodo.804529.

[69] Janet Catterall and Virginia Barbour, *Open access initiatives by research active institutions in australia and aotearoa new zealand: A snapshot of the landscape in 2022*, 2023. DOI: 10.5281/zenodo.8081167.

[70] BUNDESMINISTERIUMS FÜR BILDUNG UND FORSCHUNG, *Open access in deutschland: Die strategie des bundesministeriums für bildung und forschung*, 2018. Accessed: Aug. 31, 2023. [Online]. Available: https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/1/24102_Open_Access_in_Deutschland.pdf?__blob=publicationFile&v=6.

[71] Creative Commons, *What we do - creative commons*, Apr. 2023. Accessed: Aug. 11, 2023. [Online]. Available: https://creativecommons.org/about/.

[72] Creative Commons, *Frequently asked questions - creative commons*, Jul. 2023. Accessed: Aug. 11, 2023. [Online]. Available: https://creativecommons.org/faq/#can-i-apply-a-creative-commons-license-to-software.

[73] Choose a License, *Choose an open source license*, Jul. 2023. Accessed: Aug. 11, 2023. [Online]. Available: https://choosealicense.com/.

[74] Joinup, *Licence compatibility, permissivity, reciprocity and interoperability*, Aug. 2023. Accessed: Aug. 11, 2023. [Online]. Available: https://joinup.ec.europa.eu/collection/eupl/licence-compatibility-permissivity-reciprocity-and-interoperability.

[75] Universitätsbibliothek RWTH Aachen, *Rwth publications*, Aug. 2023. Accessed: Aug. 8, 2023. [Online]. Available: https://publications.rwth-aachen.de/?ln=de.

[76] R. A. University, *Open-access-policy der rwth aachen - rwth aachen university - deutsch*, 2019. Accessed: Aug. 8, 2023. [Online]. Available: https://www.rwth-aachen.de/cms/root/Die-RWTH/Profil/~lkcc/Open-Access-Policy-der-RWTH-Aachen/.

[77] R. A. University, *Archiving - documentation | coscine*, Jul. 2023. Accessed: Aug. 9, 2023. [Online]. Available: https://docs.coscine.de/en/archiving/#5-optional-linking-with-rwth-publications.

[78] MIT, *Mit license | software package data exchange (spdx)*, Jun. 2023. Accessed: Aug. 11, 2023. [Online]. Available: https://spdx.org/licenses/MIT.html.

[79] Free Software Foundation, Inc., *Gnu general public license v3.0 or later | software package data exchange (spdx)*, Jun. 2023. Accessed: Aug. 11, 2023. [Online]. Available: https://spdx.org/licenses/GPL-3.0-or-later.html.

[80] Open Source Guides, *The legal side of open source*, 2023. Accessed: Aug. 11, 2023. [Online]. Available: https://opensource.guide/legal/.

[81] H. Mehrwald, *Das 'Not Invented Here,-Syndrom in Forschung und Entwicklung*. Wiesbaden: Deutscher Universitätsverlag, 1999, ISBN: 978-3-8244-0483-4. DOI: 10.1007/978-3-663-08337-5.

[82] Creative Commons, *Attribution 4.0 international (cc by 4.0) legal code: Creative commons - attribution 4.0 international - cc by 4.0*, Aug. 2023. Accessed: Aug. 18, 2023. [Online]. Available: https://creativecommons.org/licenses/by/4.0/.

[83] Creative Commons, *Attribution 4.0 international (cc by 4.0) deed: Creative commons - attribution 4.0 international - cc by 4.0*, Feb. 2023. Accessed: Aug. 18, 2023. [Online]. Available: https://creativecommons.org/licenses/by/4.0/legalcode.

[84] B. L. Montavon, R. H. Schmitt, and M. Heizmann, "Virtual reference frame based on distributed large-scale metrology providing coordinates as a service: Lehrstuhl für fertigungsmesstechnik und qualitätsmanagement / werkzeugmaschinenlabor wzl der rwth aachen," Ph.D. dissertation, Apprimus Verlag and Dissertation, RWTH Aachen University, 2021, 2021. DOI: 10.18154/RWTH-2021-10238. [Online]. Available: https://publications.rwth-aachen.de/record/835033.