RESEARCH ARTICLE

Data-Producing Methods in CRC 985: Recommendations for Research Data Management in Large Interdisciplinary Projects

CRC 985 - Functional Microgels and Microgel Systems

Nicole A. Parks [1]
Konstantin W. Kröckert [2]
Fabian Claßen [3]
Walter Richtering [3]
Mathias Müller [1]
Sonja Herres-Pawlis [2]

1. IT Center, RWTH Aachen University, Aachen, Germany.
2. Institute of Inorganic Chemistry, RWTH Aachen University, Aachen, Germany.
3. Institute of Physical Chemistry, RWTH Aachen University, Aachen, Germany.

Date Submitted:
2024-03-01

Keywords:
Data, chemistry, microgels, research data management, collaborative projects, SFB 985, INF, data-producing methods

Data availability:
Data can be found here: https://dx.doi.org/10.22000/1793

Software availability:
No software was specifically developed for this project. The associated Jupyter Notebook can be found within the above-mentioned dataset.

Corresponding Author:
Sonja Herres-Pawlis
sonja.herres-pawlis@ac.rwth-aachen.de

Abstract. Within CRC 985, groups from a variety of chemical institutes, chemical engineering, physics, and the life sciences collaborate on research on microgel systems. Over a funding period of nearly 12 years, the CRC has produced numerous publications, which are associated with a large amount of underlying research data arising from many different methods. To gain a deeper understanding of this research data, the CRC 985 INF project has carried out a survey and thus gathered information on the data producing methods involved. Based on this information, recommendations for data exchange formats, data publication and archival for the current project are made. Furthermore, we propose solutions, especially for data organization and documentation, for similar, interdisciplinary projects.

1 Introduction

The collaborative research center (CRC)[1] 985 *Functional Microgels and Microgel Systems* has studied microgels, soft colloidal macromolecular compounds that find applications in many different fields for over two funding periods, the current third funding period being its final. The project brings together research groups from numerous chemical institutes, chemical engineering, physics, biotechnology, and the life sciences, with RWTH Aachen University, DWI - Leibniz Institute for Interactive Materials, the RWTH Aachen University Hospital (UKA), and Forschungszentrum Jülich (FZJ) cooperating with each other. In total, approximately 40 groups, currently involving approx. 90 principal investigators (PIs), post doctoral researchers, or doctoral researchers, have or are actively contributing to the project and over 300 scientific publications

---

[1] CRCs are long-term yet temporary research projects funded by the German Research Foundation (DFG). They can run a total of 12 years, with individual funding periods of 4 years.

have been produced so far.

In the first funding period, which began in 2012, the research data management (RDM) structure included a Microsoft SharePoint, while Mattermost was introduced as an instant-message communication system. On this basis, information could be shared and communicated across research areas as well as internally in smaller groups. Furthermore, during the previous funding periods, a sample management system was integrated into SharePoint to simplify workflows by providing sample history and tracking sample location, while implementing a universal naming system throughout the project and assigning persistent identifiers (PIDs) [1]. However, until this stage, data was shared and stored in a manner that did not follow any specific guidelines. The best practice of the scientists was therefore to add metadata in the form of individually written texts and to save raw measurement data in an individual project folder. Storing data across projects with the same structure and making it accessible for future projects is challenging with this approach. One reason for this is that different templates would have to be developed individually for different tasks, or new software would have to be developed for this purpose explicitly for this CRC. Similar statements regarding this problem description for projects of this scale have been published in other CRCs [2], [3].

From today's point of view, proficient RDM requires much more, e.g., the sharing and archiving of data according to the FAIR (findable, accessible, interoperable, reusable) principles that were introduced in 2016 [4]. At their core, these guidelines build upon one another to ultimately ensure a dataset's reusability. For research data, they carry implications for both those producing the data, e.g., researchers, but also for those providing infrastructures such as research data repositories [5]. Implementing practices and tools that enable FAIR throughout each stage of a research project also facilitates FAIR in the long run. Large, interdisciplinary projects can benefit from these practices as participants can more easily find, access, and (re)use data produced by their collaborating partners or predecessors, e.g., from previous funding periods.

Fully functional RDM infrastructures and information standards are still a work in progress. The German national research data infrastructure (NFDI) and its discipline-specific consortia aim to move this progress along [6]. In the area of chemistry, NFDI4Chem strives to not only set up a system of repositories for data sharing and archival, but also to establish minimum information and format standards to ensure data remains reusable and interoperable [7]. These efforts should inform the research communities' RDM practices, while the consortia also require researchers' input to best suit their needs.

As part of the CRC 985 Information and Infrastructure (INF) project, we present an overview of the diversity in a research project of this magnitude in terms of the number of data-producing methods and the variety of associated data. A survey to gather relevant information lays the foundation of this work. Based on this information as well as formal and informal exchange with CRC project members, we discuss how to deal with such a variety of data in future projects in terms of project preparation, recommended RDM practices in terms of storage, publication, archival and the accompanying data formats, and communication and awareness among participating researchers. Furthermore, as a project which includes many chemical and chemistry-related disciplines, the information presented here can inform the efforts and goals within NFDI consortia such as NFDI4Chem.

2 Methodology

Figure 1 shows the general approach taken for this work. Stage 1 focused on gathering information within CRC 985. To this end, the INF project compiled a structured questionnaire to survey the data-producing methods and workflows throughout the CRC. It then acquired contacts for RDM-related topics for the various groups and distributed the first version of the questionnaire. Reviewing the initial answers and exchanging information directly with the participants enabled a revision of the questions and expansion of the questionnaire. On the one hand, the questionnaire then directly included explanations on terminology and concepts such as metadata, and on the other hand the survey became more specific to enable the INF project to gather the desired information in a more targeted manner. These surveys and exchanges took place throughout 2021 through 2023.

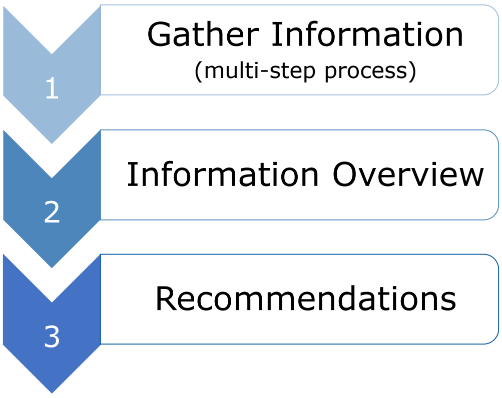

Figure 1: Targeted incremental approach to provide an overview of the project's scope and set the basis for future RDM improvements.

In the second stage, the INF project compiled an overview of the gathered information on data-producing methods. This serves as a resource for CRC 985 and was therefore published on the project's SharePoint for easy reference. Details from this table inform the Results section of this work.

The third stage, recommendations, employs the data collected in the previous stage as well as general information and feedback collected in a rather informal manner in question and answer sessions as part of workshops or presentations. This informed the INF project on the needs of the researchers. By drawing on knowledge provided by Fairsharing.org [8], re3data.org [9], and NFDI4Chem [10] as well as central solutions offered by the RWTH Aachen University, recommendations for current and future projects on infrastructure options, e.g. working data storage, electronic laboratory notebooks (ELNs), and data publishing and archival services, are made. Furthermore, areas that require additional work by infrastructure providers are pinpointed.

3 Results and Discussion

3.1 Stage 1: Gathering Information

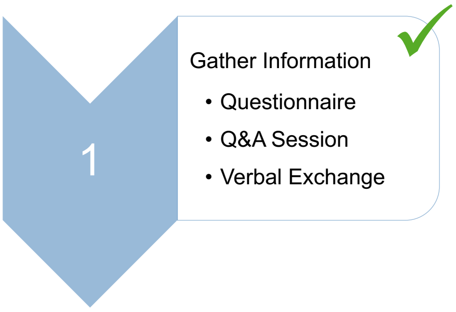

Figure 2: Successful information gathering through a questionnaire that was continuously improved through question and answer sessions and a close exchange with CRC 985 scientists.

The questionnaire created at the beginning of this study was used as a living document. Therefore, updates to the questions occurred throughout the first stage to better explain the questions and thus acquire more detailed information, as outlined in Section 2. The questionnaire successfully gathered information in a structured manner and allowed for a baseline to gain more detailed information. This required close face-to-face exchange between the research project members and members of the INF project. In addition, the INF project held seminars for researchers to raise awareness with respect to RDM. Subsequent question and answer sessions gave a further overview of the methodological diversity as well as other RDM-related concerns, enabling the INF project to provide suggestions to facilitate RDM in the CRC 985. Therefore, by combining a questionnaire as a living document with a close exchange between the data-producing researchers, the first phase was successfully completed (Figure 2). All versions of the questionnaire as well as the completed documents be found within the dataset published on Radar4Chem [11].

3.2 Stage 2: Information Overview

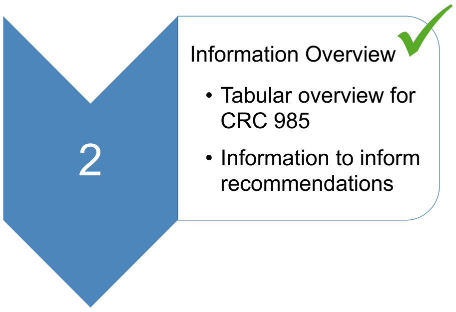

Figure 3: Successful information overview that tabulates all methods and resulting data volumes within CRC 985.

The full content of the information gathered falls outside the scope of the results reported here, with the focus being placed on information regarding data-producing methods, the produced data volume, the generated data types, and data documentation, and working data storage and organization.

The questionnaires resulted in a tabular overview of the data-producing methods employed throughout the CRC 985. Figure 4 provides an overview of these methods by research area, indicated by institute or department names. As shown, the wide variety of methods, from spectroscopy to microscopy to numerical methods, cover a broad context of disciplines. This rather coarse-grained depiction summarizes the methods into wider categories, yet, it should be mentioned that the amount of devices and setups employed throughout the CRC gives rise to a large variety of data, including differences in the data output sizes and file types, even within a specific method. In total, 40 method categories were reported throughout the project. As this reporting was primarily voluntary and researchers may acquire, develop, or even switch methods as a project progress, this number is approximate.

Figure 5 exhibits the resulting multitude of data output sizes. The majority of the methods produce data at or below the 1 GB mark, while five methods, namely high-resolution microscopy methods, such as superresolution fluorescence microscopy or tensiometry, and numerical methods, cross or go far beyond that mark.

The survey results provide an overview of commonly used data formats for raw and exported data. This will be discussed in more detail in Section 3.3, with reported data formats provided in Table 1. During exchange with researchers and due to the responses presented below, it was clear that standard formats were not necessarily well-known, however, and therefore guidance on data formats is required. This information was included on the shared overview table on the SharePoint for project members to reference. An anonymized version of this table is also provided in the published dataset [11].

The questionnaire also addressed data documentation, especially regarding (uniform) metadata. The responses reveal that, for most groups, very little uniform, machine-readable metadata is recorded unless it is contained directly in the output data files. However, this information may not always be contained in the exported version of the data, with which many members reported working. Relevant information is often included directly in the file name, analog or electronic laboratory notebooks, or digitized in plain text, Microsoft Office, or Microsoft Excel files. Only one group mentioned using controlled vocabularies.

It should be noted that, in some cases, project members, especially doctoral students, expressed concerns in terms of data storage best practices, which data should be stored, published, and archived at which stage data should be stored and/or published (raw vs. exported or processed data), data organization, backup systems, and data formats. This was often expressed in informal conversations or workshop or seminar settings.

Thus, the survey provided sufficient results to obtain an overview of the methodological diversity and generated data that led to the successful completion of the second phase (Figure 3). In addition to the data-producing methods, other foundational aspects and concerns regarding RDM were collected and will be addressed in the following.

Method	Advanced Materials for Biomedicine (DWI)	Bioinorganic Chemistry (RWTH-IAC)	Biotechnology (RWTH-Biotec)	Chemosensation (RWTH-Chemosensation)	Fluid Process Engineering (RWTH-FVT)	Functional and Interactive Polymers (DWI)	Mechanoresponsive (bio)materials (DWI)	Molecular Systems Engineering (RWTH-LTT)	Nanostructured Materials (RWTH-IAC)	Photoresponsive Materials (DWI/RWTH-ITMC)	Physical Chemistry (RWTH-IPC)	Theoretical Physics of Living Matter (FZJ-IBI5)
Atomic Force Microscopy (AFM)									■		■	
Cyclic Voltammetry (CV)		■										
Dynamic Light Scattering (DLS)						■	■		■		■	
Electron Paramagnetic Resonance Spectroscopy (EPR)		■										
Electrophysiology (Patch Clamp)				■								
Elemental Analysis (EA)		■										
Energy-dispersive X-ray spectroscopy (EDX)									■			
Fluorescence Spectroscopy		■					■					
Gas Chromatography (GC)		■			■							
High-Performance Liquid Chromatography (HPLC)					■					■		
In-situ Raman Spectroscopy		■										
Infra-Red Spectroscopy (IR)		■				■				■		
Mass Spectrometry (MS)		■								■		
Medium Pressure Liquid Chromatography (MPLC)										■		
Microscopic Probing					■							
Molecular Dynamics								■				
Nanoindentation	■											
Nuclear Magnetic Resonance Spectroscopy (NMR)		■				■	■			■		
Numerical Methods												■
Optical microscopy (confocal, widefield, inverted, fluorescence)	■			■		■	■				■	
Optical microscopy (high-throughput)	■											
Optical microscopy (superresolution)				■							■	
Plate Reader: Colorimetric or Fluorescence-Based Assays			■									
Quantum Mechanics								■				
Raman Spectroscopy		■				■						
Rheometry						■					■	
Scanning Electron Microscopy (SEM)						■			■			
Scanning Transmission Electron Microscopy (STEM)						■						
Size Exclusion Chromatography (SEC)		■				■						
Small Angle X-Ray Scattering (SAXS)											■	
Spectro-Electro		■										
Static Light Scattering (SLS)											■	
Statistical Thermodynamics								■				
Tensiometry					■						■	
Thermogravimetric Analysis (TGA)						■						
Transmission Electron Microscopy (TEM)									■			
UV/Vis Spectroscopy		■					■			■		
UV/Vis/Near Infra-Red Spectroscopy		■			■							
Viscometry											■	
X-Ray Diffraction Analysis		■										

Figure 4: Methods reported according to their area of research in CRC 985. The employed or available methods range from spectroscopy, to microscopy, to numerical, representing the variety of disciplines involved in the project. Nevertheless, many methods are common to chemistry-related research. In total, 40 method categories were reported.

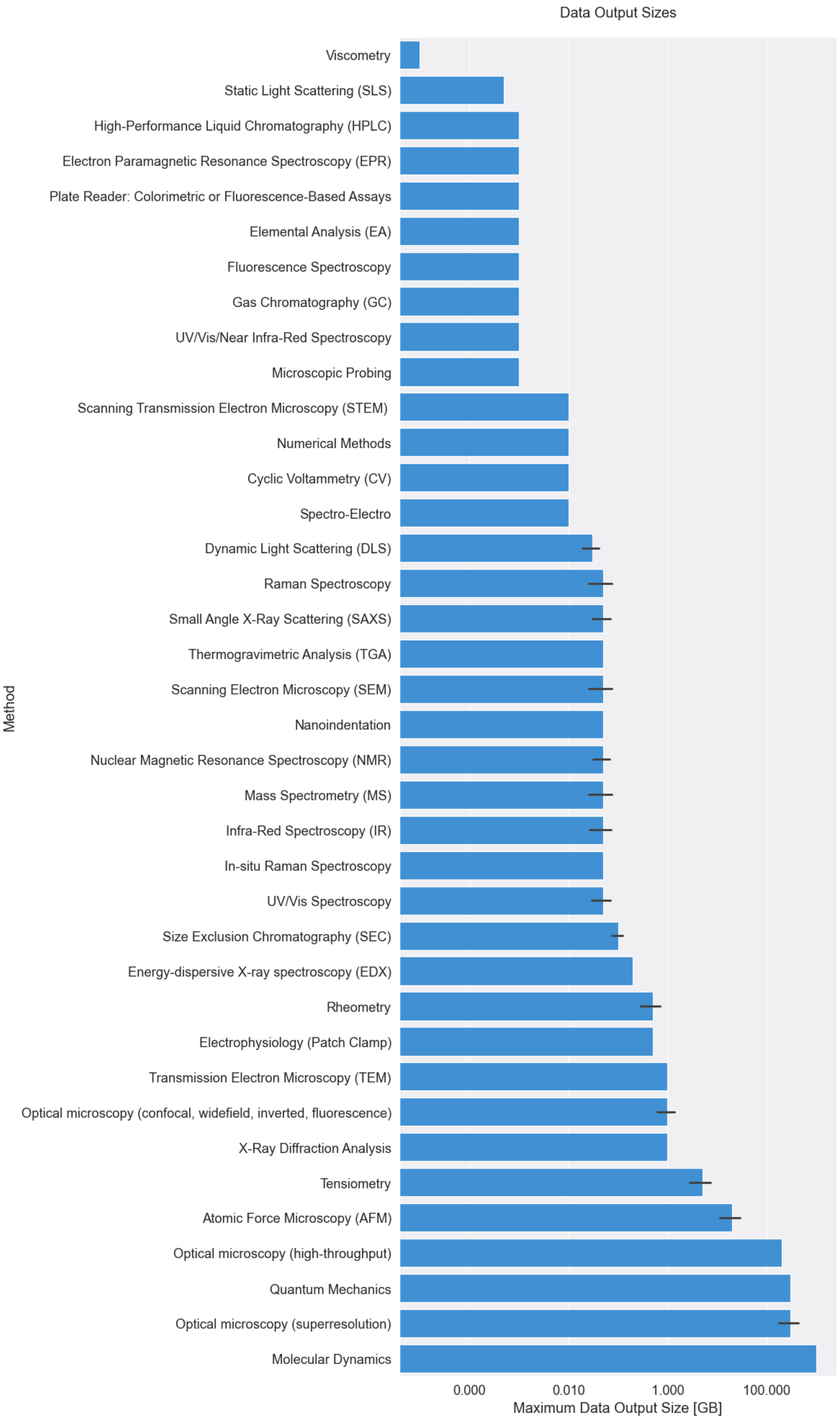

Figure 5: Methods and their output data sizes reported in CRC 985. Most reported output sizes are smaller than 1 GB, with numeric and imaging methods far beyond that point and up to 1 TB. Where applicable, error bars indicate the standard deviation of the data output sizes reported for specific methods.

3.3 Stage 3: Recommendations

Based on the knowledge gained from the presented results, we derived the following recommendations as outlined below. On the one hand, the data-producing method types and file sizes influence aspects such as data publication platforms and recommended file types. On the other hand, the project participants' accounts allow to directly address the concerns and advise on research data management best practices accordingly.

The main concerns reported were:

1. Knowledge and implementation of data organization basics and best practices regarding working data storage and structure
2. Internal data reuse, e.g., the ability to easily build upon a predecessor's work
3. Access to storage space for large amounts of (raw) data
4. Data exchange format standards
5. Knowledge of data documentation best practices and minimum information (metadata) standards
6. Publishing data underlying a journal article publication, e.g. which repository best suits the research data and data access control (open access vs. closed access options)

These concerns were largely reported on a group and not necessarily a project-specific level. Many are also interlinked and can thus be grouped together. Therefore, in the following, we will discuss and make recommendations for data organization within a group, which involves working data storage and organization, data formats, data documentation, including minimum information (metadata) standards as well as archival (covering points 1, 2, 3, 4, 5 above). Together, these points ensure data can be reused by others within the group and also prepare data for publication and reuse by those outside of an organization or project. Then, we recommended repositories based on discipline and/or data acquisition methods employed, and how to reference this data within a journal article (covering point 6 above). Lastly, we will outline how individual groups can tie this all together into collaborative work in a large, interdisciplinary project.

These recommendations fit the data lifecyle, depicted in Figure 6.

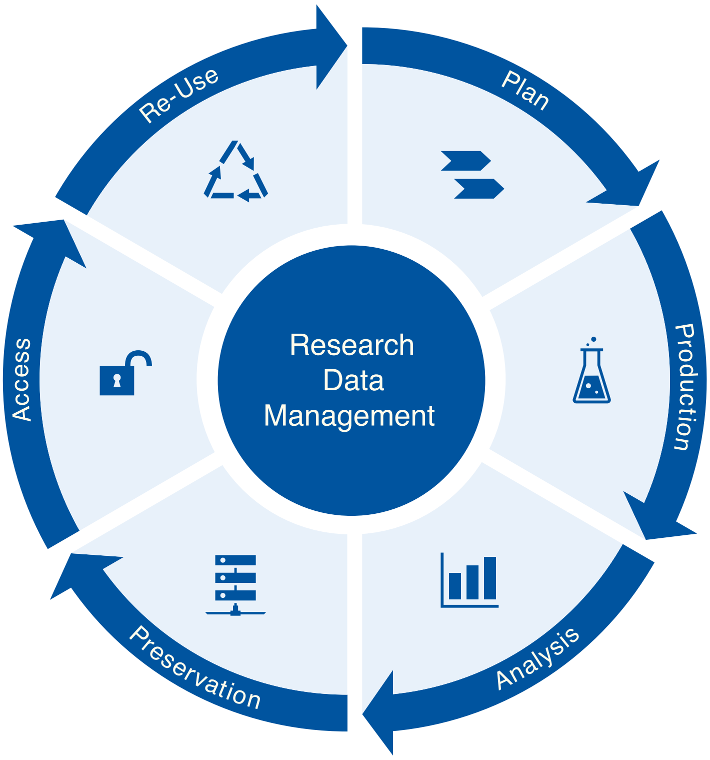

Figure 6: The research data lifecycle depicts the typical stages of research data throughout a project. These includes the planning of the project, which includes detailed planning on which research data will be generated or re-used as well as how it will be stored during and archived after the project. The active research phases include the data production and analysis phases, after which the data is preserved and access rights are determined, such as open-access in a public repository or closed access in an institutional archive. Those who have access to the data can then re-use it in the next project. At this point, the planning stage restarts the cycle [12].

3.3.1 Data Organization and Documentation

Proper data organization, first and foremost, ensures that those currently working with the data can do so efficiently. Furthermore, it enables others to easily understand and therefore reuse or build upon the data, from future doctoral students in the same group to external researchers with whom the data may be shared. A plethora of information exists on the topic; in order to condense this information for group members, the INF project held a workshop on data organization basics. This information was also outlined on guides posted to the CRC's SharePoint. Therefore, we will not go into detail here on organizing directory structures, data backup, and versioning. Rather, we will provide the essential points and then elaborate on supporting infrastructure.

Working project data should be centrally stored, either using institutional server or rented server space from the university's central service providers. Research data storage on individual hard drives should be avoided, while the centralized storage may be synchronized to local devices using various services. Working data should be organized by overarching projects, wherever possible, using templates for project directory structure, while access to directories should be controlled on an administrative level. Furthermore, this organization and structure should be well-documented, e.g., in a group level data management plan as well as plain-text README files contained within the directory structure. This not only aids in onboarding new group members, but also provides a reference for (external data) stewards or data managers, e.g., those involved in INF projects, while providing contextual information for data publication as well as for after the project has ended.

As for research data documentation, without specialized tools, text-based metadata templates in the form of YAML or JSON can also be employed. Based on such a template, details for each dataset are maintained and stored next to the data itself, thereby recording the minimum information for an experiment, measurement, or sample as agreed upon by the group, and by following existing community standards, where available. While raw data files should be maintained in read-only folders within the project structure and may contain all necessary information, text-based metadata files can provide further documentation, especially when the raw data is contained in proprietary data formats. These templates should be established in the planning phase of the research data lifecycle and updated when required throughout the data production and analysis phases (see Figure 6).

This covers very basic data storage and management that does not employ any specialized tools or infrastructure, besides a well-managed institutional server, directory structure, and documentation using agreed-upon templates. It provides group members, especially junior scientists, with the framework to operate in an efficient and organized manner, while producing transparent results that are (re)usable by other group members. However, sophisticated tools and platforms exist, and are being continuously updated and improved, to further assist researchers in effective research data management.

In many natural sciences, the laboratory journal stands as the staple of research documentation. However, analog journals are not machine-readable and do not necessarily follow uniform documentation standards. Digital counterparts, electronic laboratory notebooks (ELNs), offer a powerful solution to managing working data and, especially, documenting research in a digital manner. These platforms exist with a wide variety of styles and target user groups, from the more synthetic chemistry focused Chemotion ELN [13]–[15] to the broadly customizable eLabFTW [16], [17], both of which are used within the CRC 985.

For ELNs, it is important to continue to follow data organization and documentation best practices. While some ELNs, such as the Chemotion ELN, strive to adhere to minimum information standards for supported methods, highly customizable instances or unsupported methods require high-level organization from within the group or institute. As with the templates outlined above, groups or institutes should agree on the information to record for their experiments and create templates for the ELN. eLabFTW, for example, enables custom metadata and allows for the creation of experiment templates. Chemotion has recently also expanded to include LabIMotion [18] which enables custom modules for non-chemistry or not yet included methods. Therefore, an ELN must be centrally managed and documented within the group, analogues to the basic data organization and storage outlined above. This includes providing templates and usage guidelines, as well as training group members.

In cases where ELNs are not quite suited or increased storage is required while metadata management is at the forefront, the RWTH Aachen IT Center has developed Coscine (short for Collaborative Scientific Integration Environment) [19], [20]. This platform primarily aims to organize and manage working research data in ongoing projects. On a group level, Coscine offers various storage types, called resources, with a storage quota of up to 125 TB per project for participating universities or groups involved in NFDI-related projects. Custom metadata application profiles can be generated to fit group needs, which result in a fillable metadata form that includes metadata validation for input values. Data within a project or sub-project is organized into resources, each of which has been assigned a specific application profile and a persistent identifier (PID) in the form of an ePIC [21], which leads to a contact page. Therefore, groups can customize their data documentation and storage structure to fit their needs and incorporate community-specific minimum information standards. Details pertaining to the collaborative aspects of this platform will be discussed in Section 3.3.3.

Working from a basis of well-structured and well-documented data organization, established during the planning phase and implemented during the data production phase of the research data lifecylce (Figure 6), provides the foundation for RDM in collaborative projects. Maintenance of these practices and proper onboarding of group members ensures adherence and avoids uncertainty.

3.3.2 Data Formats

Vendor software typically directs data formats for output data, which may be proprietary. Data export options often exist, as reported throughout the project. Interoperable data requires open and standardized data formats, which do not exist for every method [22]. For many methods, open export formats such as TEXT and comma-separated values (CSV) were reported, however, the associated metadata may be lost or incomplete upon export. A workaround, as outlined above, involves maintaining complete and well-structured metadata in plain text-based file formats, within ELNs, or on platforms such as Coscine. In the case of standard open formats, tools such as ELNs and research data repositories may only accept certain standard data formats or may automatically convert the data accordingly [23]. Furthermore, vendor software may also include converters for such formats. As standard open data exchange formats exist for certain analysis methods within the CRC and since many of them were not mentioned in the survey responses, we gathered recommendations and summarized these in Table 1, sourcing information from FAIRsharing [8] and NFDI4Chem's Knowledge Base [10], as well as the Chemotion Repository documentation [24].

This information has also been shared on the CRC 985 SharePoint along with the method information outlined above. Gathering this information specifically arose from communication over the common misconception that data should always be published as CSV or TEXT files. While this may present a solution for data types for which no standards have been established, standards for certain data types do, indeed, exist and may simply lack awareness.

Table 1: Data exchange formats recommended by FAIRsharing, NFDI4Chem, and the Chemotion Repository for selected methods reported within CRC 985 and common data formats or file extensions reported throughout the project. Formats sourced from FAIRsharing.org are cited accordingly, while those listed on NFDI4Chem's Knowledge Base and the Chemotion Repository Documentation are denoted accordingly. We recommend the adoption of formats printed in bold font.

method	data exchange format or file extension	data exchange formats within CRC 985
Chromatography	**ANDI-MS** [25] CSV[a] TXT[a]	CSV PDF .vdt .gcd
Colorimetric or Fluorescence-based Assays	-	.ruc (raw) ASCII (export including meta-data)
Computational Chemistry	CHARMM Card File Format (CRD) [26]	ASCII .log .cosmo .energy .out .gjf .xyz CSV (processed)
Cyclic Voltammetry (CV)	TXT[a]	.nox
Electrophysiology (patch clamp)	-	.dat
Electron Paramagnetic Resonance Spectroscopy (EPR)	TXT[a]	.spe TXT (export)
Elemental Analysis (EA)	TXT[a]	TXT
Energy-dispersive X-ray spectroscopy (EDX)	-	TXT JPEG (export) PNG (export)
Fluorescence spectroscopy	**JCAMP-DX**[a]	OPJ FDS TXT (export) PDF (export)

Continued on next page

(Continued)

IR Spectroscopy (IR)	**JCAMP-DX** [27][a] AnIML [28][b]	.ispd TXT (export) PDF (export)
Mass Spectrometry (MS)	**JCAMP-DX** [27] AnIML [28][b] **mzML** [29][a]	.d .bad Xcalibur Raw file TXT .jws
Mechanical Surface Analysis (nanoindentation)	- (standard data model: CWA 17552:2020 [30]	TXT
Microscopy	**OME-TIFF** [31]	.nid .spm .jpk-qi-image .jpk-qi-data **TIFF**[e] LIF DM4 (TEM) JPEG (export) PNG (export) AVI (video) CSV TXT
NMR	NMR-STAR [32] CCPN [33] **NMR-ML** [34] **NMReData** [35] (assignments)[a] AniML [28][b] **JCAMP-DX** (raw)[a]	.mrnova FID PDF (export)
Raman Spectroscopy	**JCAMP-DX**[a] AniML [28][b]	.icRaman .sps TXT (export) CSV (export) .spc (export) .xlsx (export)

Continued on next page

(Continued)

Rheometry	-	.rdf .tri .iwp CSV (export)
Dynamic Light Scattering	CSV[b]	**ASC**[d] .dts **.zmes**[d] CSV (export) TXT (export)
Static Light Scattering	-	.d80 .txt (export, not all parameters included)
Small Angle X-Ray Scattering (SAXS)	-	.mpa .info .edf .dat
Spectroelectrochemistry	-	.str8 TXT (export)
Tensiometry	PNG (contact angle measurements)[a]	.krs **.zip (export, contains all .krs and XML)**[d] XLSX (export or analysis results)
Thermal Analysis	-	.stad .spp TXT (export) CSV (export)
UV/Vis Spectroscopy	CSV[a]	
JCAMP-DX[c]	.dsw .bsk .bkn .str .jws .jwb .ksd .sre (ASCII) TXT (export) CSV (export)	

Continued on next page

(Continued)

X-Ray Diffraction Analysis (XRD)	**CIF** [36] (single crystal)[a] **.xyd** (powder)[a]	binary encoded frames (images) .p4p .hkl .res CIF .x

[a] NFDI4Chem Knwoledge Base

[b] under development according to FAIRsharing.org

[c] Chemotion Repository

[d] (meta)data accessible by common tools

[e] preferably method-specific TIFF-formats that include extended metadata

The existing standard data exchange formats listed in Table 1 provide guidelines on formats to chose from, while recommended standards and common formats are highlighted in bold font. The exact format choice for each method will depend on available software and export or conversion tools and also the formats data types specific repositories will accept for publication (see, for example, Chemotion Repository requirements in [24]).

Notably, many methods do lack specific standards, for which the above-mentioned practice of documenting data appropriately and sharing data along with the associate metadata in open, text-based formats is advised. As the varying efforts such as the NFDI consortia continue their work, more standards will become available. Furthermore, minimum information standards will continue to direct how data should be formatted and documented, further guiding format standards.

Table 1 as well as the published overview [11] also serve to inform the standards community on which formats researchers are employing in their day-to-day work and where standards are lacking. It may also serve providers of ELNs and repositories in terms of which formats are in use and which may be supported by their tools.

Where standards are not yet available, providing data in an accessible format is advisable. For example, in dynamic light scattering, devices supplied by Malvern Pananalytical commonly provide data in files with the .zmes extension. While not necessarily a standard format, the data may be accessed and viewed using common SQLite tools. Similarly, dynamic light scattering data provided in ASC files may be accessed by common text editors. Many of these include metadata alongside the data, while exporting to TXT and CSV files may not and additional metadata annotation would be required.

As with data organization and documentation, format standards must be agreed upon as part of the planning stage of the data lifecycle (Figure 6), communicated within the group, and updated as more standards become available.

3.3.3 Collaboration

Up until now, the discussion has focused on the group level. Having a well-documented approach to data organization, documentation, and the tools used helps in identifying how collaborative projects such as CRCs and the contained sub-projects can best manage data.

The CRC 985 INF project addressed sample tracking throughout a collaborative project involving many different groups and institutes in previous funding periods [1], as described in Section 1. This system aimed to solve a specific problem with sample traceability within the project, while enabling project members to directly attach associated data to a sample. However, as shown in Figure 4, some research within the CRC may not involve physical samples. Furthermore, SharePoint relies on database storage that cannot accommodate larger data sets. It is therefore not suitable for all methods and the associated (raw) data, e.g., superresolution fluorescence microscopy (SRFM) and numerical methods. For these cases, other systems can provide the necessary solutions. It should also be noted that the metadata describes the sample rather than any attached data, and therefore would still require external documentation to fully describe the dataset belonging to the sample if not included directly within in the files.

A shared or central ELN instance, that is used by all the members of the project, could provide one solution, yet, this did not prove realistic in this CRC for multiple reason, from varying user and group needs to the lack of a centralized solution offered by the university. As individual groups and institutes have indeed implemented ELNs, exchange formats between these could assist in collaborations in such projects. Solutions here may be provided by exchange formats between ELNs, a goal of initiatives such as the ELN Consortium [37], which currently involved ten ELN providers, including those mentioned above.

Another option is using Coscine. Its intent is to create a collaborative work environment. Roll management occurs on a project level, therefore, members can be given access to their respective sub-project, with all data relevant to the project collected and documented in one place. A REST API (Representational State Transfer Application Programming Interface) also allows for automated data workflows, e.g., between local servers or ELNs and the platform. As such, metadata, data, and identifiers may be mirrored between platforms, giving members a working-group agnostic option. Its large storage capacity assists researchers where local servers or systems that rely on a database structure such as SharePoint and some ELNs reach their limits. The system has been used for groups producing large amounts of data within the CRC, such as tensiometry and other imaging techniques.

In large interdisciplinary projects, one-size-fit-all-systems prove difficult. However, having clear guidelines for individual groups that enable FAIR allows for these systems to be linked to one centralized system that facilitates data sharing. Due to a lack of a REST API, the CRC's SharePoint and therefore the original sample management system prove difficult for automated workflows. Coscine and the major ELNs do offer solutions here, which therefore enable one central platform for data sharing and collaboration within a project, while local systems are maintained.

In the process of this study, besides the questionnaire, the sample management system was also optimized, leading to more information when storing data. This study thus also directly contributed to improved data management, and shows that for specific cases it can be helpful to use a variable platform as is desired for future data storage.

3.3.4 Data Publication and Archival

Figure 7: Several recommendations could be made for active data storage, including data formats, documentation, and archival for a project on the scale of CRC 985.

Aside from facilitating research within groups as well as large projects, the aim to make data reusable according to FAIR also includes making the (meta)data available to others while describing how to access the data (Figure 6: Access and Re-Use). Therefore, a data policy was established during the second funding period [1], which stipulated that all data underlying a published journal article should be published as well.

Various options exist for such publications, with the three common categories being: (1) institutional repositories, (2) general repositories, and (3) community-specific repositories. Where possible, community-specific repositories are preferred, as these are able to provide detailed metadata templates, enabling researchers to fully describe the published data. When using general or institutional repositories, adding as many (optional) metadata fields is best practice, while providing plain-text README files for additional context. As institutional repositories may be used for reporting purposes, importing published datasets is also important, analogous to text publications. Within these categories, we make the following recommendations for data sharing and archival CRC 985, outlined in Table 2, which completes the final objective of this study (Figure 7).

Table 2: Repositories recommended for CRC 985 and projects with similar data types. Institutional repositories correspond to research institutes involved in the current project.

Repository	Type	Description [8]	Date Size Limits
Jülich DATA [38]	institutional	A registry service to index all research data created at or in the context of Forschungszentrum Jülich, which may also be used to publish research data and software.	10 GB per file (depends on Dataverse installation); prefers links to larger datasets [39]

Continued on next page

(Continued)

RWTH Publications Research Data [40]	institutional	As part of the general RWTH Publications repository, data and software can be published by all RWTH Aachen University members and affiliates.	100 GB per file; 1 TB maximum over all files (gigamove) [41]
Chemotion Repository [42]	discipline-specific	The repository supports the storage of data related to chemical samples or reactions, with a focus on data from synthetic and analytical work. While not a requirement, data may be submitted directly via the Chemotion ELN.	None; might limit it to 50 MB in future [43]
Cambridge Structural Database (CSD) [44]	discipline-specific	Established in 1965, the Cambridge Structural Database (CSD) is the a repository for small-molecule organic and metal-organic crystal 3D structures. Database records are automatically checked and manually curated by one of our expert in-house scientific editors. Every structure is enriched with chemical representations, as well as bibliographic, chemical and physical property information, adding further value to the raw structural data.	50 MB per file; 100 MB for the total size of files uploaded; exception for bigger files via email contact [45]

Continued on next page

(Continued)

Inorganic Crystal Structure Database (ICSD) [46]	discipline-specific	The world's largest database for fully determined inorganic crystal structures and contains the crystallographic data of published crystalline inorganic structures. Organometallic and theoretical structures have been added within the past years.	None; contact for file sizes > 10 TB [47]
ioChem-BD [48], [49]	discipline-specific	IoChem-BD is a digital repository of Computational Chemistry and Materials results. A set of modules and tools aimed to manage large volumes of quantum chemistry results from a wide variety of broadly used simulation packages.	default 1 GB per upload; > 100 MB not to be uploaded by web interface [50]
NOMAD Repository & Archive [51]	discipline-specific	The NOMAD Repository and Archive stands for open access of scientific materials data. It enables the confirmatory analysis of materials data, their reuse, and repurposing. All data is available in their raw format as produced by the underlying code (Repository) and in a common, machine-processable, and well-defined data format (Archive).	32 GB per upload (maximum of 10 non-published uploads per user) [52]

Continued on next page

(Continued)

RADAR4Chem [53], [54]	chemistry - general	A low-threshold and easy-to use service for sustainable publication and preservation of research data from all disciplines of chemistry. Currently, exclusive to publicly funded research institutions and universities in Germany.	10 GB per project [53]
Suprabank [55]	discipline-specific	Curated, open resource for intermolecular interaction.	10 GB per user (can be adapted) [56]
zenodo [57]	general	EU discipline-agnostic repository for data and other research results.	50 GB per data set [58]

Certain repositories are also tied to ELNs, therefore providing direct data and metadata workflows. Going a step further, data may also be converted to standard open formats, as is the case with Chemotion ELN and Chemotion Repository, as mentioned above.

The published data should then be explicitly referenced via their DOI within the article using a data availability statement, which journals are increasingly requiring [59]. They may also be cited within the article itself. Especially in cases which involve multiple published datasets, this provides additional context for the reader.

Within the CRC, aside from the repositories of the respective institutions, the Chemotion Repository, for those using the Chemotion ELN, RADR4Chem, as well as the two institutional repositories (see Table 2) have been fairly accepted. These guarantee storage and accessibility for 10 years or more, conforming with German Research Foundation (DFG) requirements; the data herein is therefore successfully be deemed archived, while it can also be accessed and reused in accordance with the research data lifecycle in Figure 6.

As shown in Figure 5, much of the data volume falls into smaller sizes, with imaging and numerical methods requiring larger storage if all data were to be published. Therefore, much of the produced can be published without too much concern for data volume. For some methods, such as Atomic Force Microscopy, not all extracted data must be published, yet the scripts employed to do so could be. Hence, the data may be reproduced in the same manner when needed, while the published data volume is held to a minimum in cases where repositories limit quota. Furthermore, repositories may offer more quota upon request.

Typically, however, projects will amass more data than that, which has been published. This therefore requires additional archive resources. For project members in CRC 985, the above-mentioned Coscine also serves as an archiving space. The entire SharePoint, including the sample management system will be archived under the CRC's Coscine project, while members can gain access to the system to archive their data as needed.

3.4 Possible Requirements for Future CRCs and INF Projects

The overarching role of INF projects within the CRC has largely been left out of the above discussion. These central projects, however, can play a vital part in setting up and implementing the above aspects.

Three aspects were identified within the CRC 985 INF project that should be considered for future projects:

1. Support for project-wide data management plans and guidelines during project planning stage
2. End-of-life plan for implemented infrastructure solutions
3. Sustainability of software solutions

To elaborate on 1., many workflows within research groups evolve naturally to fit the needs of those carrying out much of the practical work, i.e., the individual doctoral researchers. However, these tend to be highly individualistic and can be difficult to alter in order to streamline data workflows. Therefore, providing clear guidelines on data organization and associated tools is vital both within the group, but also across the project and should be established in the planning phase. INF projects need to be involved at this stage and assist with infrastructure planning and selection. Thus, overarching solutions can be available at the beginning of a project to avoid implementing solutions and tools and altering workflows during ongoing work. Thus, individual workflows can then be developed within a given framework that facilitates data storage, documentation, and exchange.

In terms of 2., the selected solutions require a detailed end-of-life management. It will not always be possible to foresee which services and dependencies may become outdated over the lifetime of a project. However, precautions to safeguard any and all data managed by these services in a structured manner must exist.

As for 3., the software solutions developed by the INF project, e.g., data workflow scripts, should be designed to outlive the project. The aspect of maintenance comes into play. Therefore, INF projects should directly include individuals within the groups who are able to maintain these solutions after the INF project is no longer available.

Overall, detailed, high-level planning for data management and the implementation of infrastructure solutions should involve INF projects at a very early stage of the project. Then, throughout the project, members must be onboarded and continuously informed on common practices, guidelines, and policies to ensure adherence.

4 Conclusion

Information on the data-producing methods and the associated data formats and data sizes in CRC 985 were collected in order to gain an overview of the diversity and derive RDM concepts and structures for CRCs. The collected information is based on a structured survey, which collected most of the details on the methods themselves, while formal as well as informal discussions in various settings provided further feedback and deeper insight. Based on the information as a whole, recommendations for this ongoing as well as future projects are made.

The gathered information paints a picture of the varied disciplines and the accordingly varied data types and sizes. This underlines the need for standardized open exchange formats, as many of the open export formats reported do not necessarily contain the required complete information in the form of structured metadata to fully understand the acquired data. In order to assist in this, tools from plain-text metadata templates to structured ELNs and data management platforms provide essential machine-readable solutions for data documentation, assisting in data interoperability and reuse.

The workflows and the RDM practices for each stage of the research data lifecyle (see Figure 6) should be clearly defined and documented on a group level in advance. This information can then feed into large projects such as CRCs, enabling informed decisions regarding RDM and collaboration within the planning phase. In this way, data stewards within the INF project can then establish policies, workflows, and infrastructures for collaboration within these institutional frameworks while working closely with researchers.

For projects of the size of CRC 985, a one-size-fits-all solution, such as a uniform ELN and repository where all (meta)data can be recorded in a well-structured manner, does not exist due to the variety of analytical and experimental methods employed and the associated different data formats and size requirements. Therefore, discipline-specific solutions found on a group level require collaboration platforms that support RDM. Within CRC 985, Microsoft SharePoint serves as such as collaboration platform, however, expectations regarding RDM evolved over the project duration. FAIR data requires more structured and defined metadata on various levels. More appropriate platforms for RDM have become available, including platforms such as the RWTH Aachen University's Coscine. This shows that, in addition to a minimum standard which should be defined a prior to the data production phase of the research data lifeycle (see Figure 6), a certain flexibility should also be implemented to meet evolving requirements in later funding periods.

With the requirement to publish all data underlying a text publication, ELNs and RDM platforms can greatly assist researchers' workflows in FAIR data publication and archival in subject-specific repositories by providing automated workflows. With much of this work still being in-progress by infrastructure providers, future research projects will be able to greatly benefit, while current work provides vital insight for these efforts.

5 Acknowledgements

The authors acknowledge German Research Foundation (DFG) funding under the project numbers 191948804 (CRC 985) and 441958208 (NFDI4Chem) as well as for the funding and support within the framework of the DALIA project with the funding code 16DWWQP07B, funded by the the Federal Ministry of Education and Research (BMBF) and the funding measure from the EU's Capacity Building and Resilience Facility.

6 Roles and contributions

Nicole A. Parks: Conceptualization, Investigation, Writing, Visualization, Data Curation – original draft

Konstantin W. Kröckert: Conceptualization, Investigation, Writing – original draft

Fabian Claßen: Conceptualization, Writing – original draft

Walter Richtering: Project Administration, Writing - review & editing

Mathias Müller: Project Administration, Writing - review & editing

Sonja Herres-Pawlis: Project Administration, Supervision, Writing – review & editing

References

[1] F. Claus, S. Kirchmeyer, M. S. Müller, and W. Richtering, "Das INF-Projekt im SFB 985 Funktionelle Mikrogele und Mikrogelsysteme," *Bausteine Forschungsdatenmanagement*, no. 2, pp. 104–111, Nov. 2019. DOI: 10.17192/bfdm.2019.2.8097. (visited on 04/06/2023).

[2] M. Schröder, H. LeBlanc, S. Spors, and F. Krüger, "Intra-consortia data sharing platforms for interdisciplinary collaborative research projects," *it - Information Technology*, vol. 62, no. 1, pp. 19–28, Feb. 2020, ISSN: 2196-7032, 1611-2776. DOI: 10.1515/itit-2019-0039. (visited on 02/03/2023).

[3] H.-J. Götze, M. Alten, H. Burger, *et al.*, "Data Management of the SFB 267 for the Andes — from Ink and Paper to Digital Databases," in *The Andes*, O. Oncken, G. Chong, G. Franz, *et al.*, Eds., Springer Berlin Heidelberg, 2006, pp. 539–556, ISBN: 978-3-540-24329-8. DOI: 10.1007/978-3-540-48684-8_26. (visited on 03/08/2023).

[4] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data*, vol. 3, no. 1, p. 160 018, Mar. 2016, ISSN: 2052-4463. DOI: 10.1038/sdata.2016.18. (visited on 01/27/2023).

[5] A. Kraft, *The FAIR Data Principles for Research Data*, Sep. 2017. (visited on 03/15/2023).

[6] N. Hartl, E. Wössner, and Y. Sure-Vetter, "Nationale Forschungsdateninfrastruktur (NFDI)," *Informatik Spektrum*, vol. 44, no. 5, pp. 370–373, Oct. 2021, ISSN: 1432-122X. DOI: 10.1007/s00287-021-01392-6. (visited on 09/21/2023).

[7] C. Steinbeck, O. Koepler, F. Bach, *et al.*, "NFDI4Chem - Towards a National Research Data Infrastructure for Chemistry in Germany," *Research Ideas and Outcomes*, vol. 6, e55852, Jun. 2020, ISSN: 2367-7163. DOI: 10.3897/rio.6.e55852. (visited on 04/22/2022).

[8] *FAIRsharing | Home*. [Online]. Available: https://fairsharing.org/ (visited on 04/04/2023).

[9] *Re3data - Registry of Research Data Repositories*. DOI: 10.17616/R3D. [Online]. Available: https://www.re3data.org/ (visited on 12/21/2022).

[10] *NFDI4Chem Knowledge Base | NFDI4Chem Knowledge Base*. [Online]. Available: https://knowledgebase.nfdi4chem.de (visited on 04/04/2023).

[11] N. A. Parks, K. Kröckert, F. Claßen, M. Müller, S. Herres-Pawlis, and W. Richterig, *Dataset belong to the publication "Data-Producing Methods in CRC 985: Recommendations for Research Data Management in Large Interdisciplinary Projects"*, 2023. DOI: 10.22000/1793.

[12] FDM-Team RWTH Aachen Univesity, *Research Data Life Cycle*, 2022. (visited on 04/27/2022).

[13] P. Tremouilhac, A. Nguyen, Y.-C. Huang, *et al.*, "Chemotion ELN: An Open Source electronic lab notebook for chemists in academia," *Journal of Cheminformatics*, vol. 9, no. 1, p. 54, Sep. 2017, ISSN: 1758-2946. DOI: 10.1186/s13321-017-0240-0. (visited on 04/25/2022).

[14] PiTrem, *ComPlat/chemotion_ELN: Chemotion ELN 0.9.1*, Jun. 2021. DOI: 10.5281/zenodo.4899080. [Online]. Available: https://zenodo.org/record/4899080 (visited on 01/26/2023).

[15] *Chemotion*. [Online]. Available: https://eln.chemotion.net/home (visited on 04/25/2022).

[16] N. Carpi, A. Minges, and M. Piel, "eLabFTW: An open source laboratory notebook for research labs," *The Journal of Open Source Software*, vol. 2, no. 12, p. 146, Apr. 2017, ISSN: 2475-9066. DOI: 10.21105/joss.00146. (visited on 04/05/2023).

[17] *eLabFTW - Open Source Laboratory Notebook*. [Online]. Available: https://www.elabftw.net (visited on 04/05/2023).

[18] *The LabIMotion Extension | Chemotion*, Mar. 2023. [Online]. Available: https://chemotion.net/docs/labimotion (visited on 04/17/2023).

[19] M. Politze, F. Claus, B. D. Brenger, M. A. Yazdi, B. P. A. Heinrichs, and A. Schwarz, "How to Manage IT Resources in Research Projects? Towards a Collaborative Scientific Integration Environment," *European journal of higher education IT*, 2020. DOI: 10.18154/RWTH-2020-11948.

[20] *Coscine | The research data management platform*. [Online]. Available: https://coscine.de/ (visited on 03/17/2023).

[21] *Persistent Identifiers for eResearch*. [Online]. Available: https://www.pidconsortium.net/ (visited on 03/16/2023).

[22] D. Rauh, C. Blankenburg, T. G. Fischer, *et al.*, "Data format standards in analytical chemistry," *Pure and Applied Chemistry*, vol. 94, no. 6, pp. 725–736, Jun. 2022, ISSN: 1365-3075. DOI: 10.1515/pac-2021-3101. (visited on 03/16/2023).

[23] P. Tremouilhac, P.-C. Huang, C.-L. Lin, *et al.*, "Chemotion Repository, a Curated Repository for Reaction Information and Analytical Data," *Chemistry–Methods*, vol. 1, no. 1, pp. 8–11, 2021, ISSN: 2628-9725. DOI: 10.1002/cmtd.202000034. (visited on 04/18/2023).

[24] *For Data Files | Chemotion*, https://chemotion.net/docs/repo/details_standards/files, Oct. 2023. (visited on 01/16/2024).

[25] FAIRsharing Team, *FAIRsharing record for: Analytical Data Interchange Protocol for Chromatographic Data*. DOI: 10.25504/FAIRSHARING.D7795C. (visited on 03/29/2023).

[26] FAIRsharing Team, *FAIRsharing record for: CHARMM Card File Format*, 2015. DOI: 10.25504/FAIRSHARING.7HP91K. (visited on 03/29/2023).

[27] FAIRsharing Team, *FAIRsharing record for: Joint Committee on Atomic and Molecular Physical data - working group on Data eXchange*, 2018. DOI: 10.25504/FAIRSHARING.V8NVE2. (visited on 03/29/2023).

[28] FAIRsharing Team, *FAIRsharing record for: Analytical Information Markup Language*, 2015. DOI: 10.25504/FAIRSHARING.6CS4BF. (visited on 03/29/2023).

[29] FAIRsharing Team, *FAIRsharing record for: Mz Markup Language*, 2015. DOI: 10.25504/FAIRSHARING.26DMBA. (visited on 03/29/2023).

[30] FAIRsharing Team, *FAIRsharing record for: CWA 17552:2020 Engineering materials - Electronic data interchange - Instrumented indentation test data*. DOI: 10.25504/FAIRSHARING.5C379F. (visited on 03/28/2023).

[31] FAIRsharing Team, *FAIRsharing record for: Open Microscopy Environment - Tagged Image File Format*, 2015. DOI: 10.25504/FAIRSHARING.CQ8TG2. (visited on 03/28/2023).

[32] FAIRsharing Team, *FAIRsharing record for: NMR Self-defining Text Archive and Retrieval format*, 2015. DOI: 10.25504/FAIRSHARING.2CHXXC. (visited on 03/29/2023).

[33] FAIRsharing Team, *FAIRsharing record for: Collaborative Computing Project for NMR*, 2015. DOI: 10.25504/FAIRSHARING.AVW5Q. (visited on 03/29/2023).

[34] FAIRsharing Team, *FAIRsharing record for: Nuclear Magnetic Resonance Markup Language*, 2015. DOI: 10.25504/FAIRSHARING.ES03FK. (visited on 03/29/2023).

[35] FAIRsharing Team, *FAIRsharing record for: Nuclear Magnetic Resonance Extracted Data Format*. DOI: 10.25504/FAIRSHARING.8AE3D0. (visited on 03/29/2023).

[36] FAIRsharing Team, *FAIRsharing record for: Crystallographic Information Framework*, 2015. DOI: 10.25504/FAIRSHARING.ZR52G5. (visited on 03/29/2023).

[37] *The ELN Consortium*. [Online]. Available: https://github.com/TheELNConsortium (visited on 04/06/2023).

[38] Re3data.Org, “Jülich DATA,” 23 dataverses, 29 datasets, 1170 files, 2021. DOI: 10.17616/R31NJMYC. (visited on 04/06/2023).

[39] *Dataset + File Management — Jülich DATA documentation*. [Online]. Available: https://apps.fz-juelich.de/fdm/staging/mode-of-access/user/dataset-management.html%5C#file-upload (visited on 10/13/2023).

[40] Re3data.Org, “RWTH Publications Research Data,” 319 research datasets, 2018. DOI: 10.17616/R33N6J. (visited on 04/06/2023).

[41] *GigaMove - RWTH AACHEN UNIVERSITY IT Center - English*. [Online]. Available: https://www.itc.rwth-aachen.de/cms/it-center/Services/Kollaboration/~smiti/GigaMove/?lidx=1 (visited on 10/13/2023).

[42] FAIRsharing Team, *FAIRsharing record for: Chemotion repository*, 2018. DOI: 10.25504/FAIRSHARING.IAGXCR. (visited on 04/18/2023).

[43] *Frequently Asked Questions (FAQ) | Chemotion*, Oct. 2023. [Online]. Available: `https://chemotion.net/docs/repo/faq` (visited on 10/13/2023).

[44] FAIRsharing Team, *FAIRsharing record for: The Cambridge Structural Database*, 2015. DOI: `10.25504/FAIRSHARING.VS7865`. (visited on 04/18/2023).

[45] *Deposit - The Cambridge Crystallographic Data Centre (CCDC)*. [Online]. Available: `https://www.ccdc.cam.ac.uk/deposit/upload` (visited on 10/13/2023).

[46] FAIRsharing Team, *FAIRsharing record for: Inorganic Crystal Structure Database*. DOI: `10.25504/FAIRSHARING.A95199`. (visited on 04/18/2023).

[47] D. A. Steudel, D. S. Rühl, D. R. Hinek, and S. Rehme, “Scientific Manual ICSD Database,”

[48] C. Bo, M. Alvarez, N. Lopez, F. Maseras, J. M. Poblet, and C. De Graaf, *ioChem-BD Find central service*, Nov. 2017. DOI: `10.19061/iochem-bd-find`. (visited on 04/18/2023).

[49] FAIRsharing Team, *FAIRsharing record for: ioChem-BD*, 2018. DOI: `10.25504/FAIRSHARING.LWW6A1`. (visited on 09/21/2023).

[50] *Set upload limits — ioChem-BD documentation*. [Online]. Available: `https://docs.iochem-bd.org/en/latest/faqs/admin/setup-upload-limits.html` (visited on 10/13/2023).

[51] FAIRsharing Team, *FAIRsharing record for: NoMaD Repository*, 2018. DOI: `10.25504/FAIRSHARING.AQ20QN`. (visited on 04/18/2023).

[52] *How to upload data — NOMAD Repository and Archive documentation*. [Online]. Available: `https://nomad-lab.eu/prod/rae/docs/upload.html` (visited on 10/13/2023).

[53] *RADAR4Chem | RADAR*. [Online]. Available: `https://radar.products.fiz-karlsruhe.de/de/radarabout/radar4chem` (visited on 04/18/2023).

[54] FAIRsharing Team, *FAIRsharing record for: RADAR*. DOI: `10.25504/FAIRSHARING.601A27`. (visited on 04/18/2023).

[55] FAIRsharing Team, *FAIRsharing record for: SupraBank*, 2018. DOI: `10.25504/FAIRSHARING.VJWUT7`. (visited on 04/18/2023).

[56] *SupraBank*. [Online]. Available: `https://suprabank.org/terms%5C_of%5C_service` (visited on 10/13/2023).

[57] FAIRsharing Team, *FAIRsharing record for: Zenodo*, 2018. DOI: `10.25504/FAIRSHARING.WY4EGF`. (visited on 04/18/2023).

[58] *Zenodo - Research. Shared.* [Online]. Available: `https://help.zenodo.org/faq/` (visited on 10/13/2023).

[59] N. A. Parks, T. G. Fischer, C. Blankenburg, *et al.*, “The current landscape of author guidelines in chemistry through the lens of research data sharing,” *Pure and Applied Chemistry*, Feb. 2023, ISSN: 1365-3075. DOI: `10.1515/pac-2022-1001`. (visited on 04/06/2023).