Handreichung zum Umgang mit Forschungsdaten in DFG-Anträgen der Fachkollegien 4.21 (Verfahrenstechnik, Technische Chemie) und 4.22 (Strömungsmechanik, Technische Thermodynamik und Thermische Energietechnik)

Stand: 28. Oktober 2026

Präambel

Die vorliegende Handreichung soll allgemein gültige Dokumente wie die Leitlinien zum Umgang mit Forschungsdaten oder den Kodex „Leitlinien zur Sicherung guter Wissenschaftlichen Praxis“ fachspezifisch ergänzen und Antragstellenden, Gutachtenden sowie Gremienmitgliedern als Orientierung für die Formulierung, Begutachtung und Bewertung von DFG-Anträgen dienen.

Grundsätzlich gilt, dass entstehende Forschungsdaten sowie Fragen des zugehörigen Datenmanagements so vielfältig und heterogen sind wie die in den Fachkollegien vertretenen Disziplinen selbst. Bestehende Unterschiede zwischen einzelnen Fächern oder Fachcommunitys dürfen daher nicht zum Nachteil einzelner Antragstellender gereichen.

Dennoch ist ein verantwortungsvoller Umgang mit den in einem Forschungsvorhaben entstehenden Forschungsdaten ein elementarer Bestandteil desselben und es besteht in den Fachkollegien ein breiter Konsens darüber, dass in jedem DFG-Antrag, insbesondere unter Zuhilfenahme der DFG-Checkliste, angemessene projektspezifische Überlegungen zum Umgang mit Forschungsdaten anzustellen sind. Als Grundlage empfiehlt sich die Berücksichtigung der FAIR-Prinzipien für Forschungsdaten (findability, accessibility, interoperability, and reusability). Darauf aufbauend, aber keineswegs abschließend, wird auf die Best-Practice-Forschungsdatenkonzepte verwiesen, die sich in koordinierten Forschungsverbünden des Fachgebietes bewährt haben und/oder in den einschlägigen NFDI-Konsortien (z.B. NFDI4ING oder NFDI4cat) stetig weiterentwickelt werden. Sehr positiv wird zudem die Darstellung eigener Praxisbeispiele (z.B. anhand von Links auf Datensätze) in der Einzelförderung gesehen.

Allgemeinlautende Verweise, etwa rein auf Anwendung der FAIR-Prinzipien ohne weitere Erläuterungen oder auf zentrale Maßnahmen an der betreffenden Forschungseinrichtung, werden hingegen als nicht ausreichend angesehen. Bei Fortsetzungsanträgen und in Abschlussberichten ist ein Rückblick auf erfolgte Aktivitäten zum Datenmanagement erwartbar und dieser soll auch in die Begutachtung und Bewertung einfließen. Bei interdisziplinär ausgerichteten Forschungsprojekten ist dafür Sorge zu tragen, dass der Umgang mit Forschungsdaten den Anforderungen aller beteiligten Disziplinen genügt.

Die oben verlinkte, von der DFG zur Verfügung gestellte Checkliste zum Umgang mit Forschungsdaten beinhaltet die folgenden sechs zentralen Punkte, zu denen in jedem DFG-Antrag möglichst konkrete und verbindliche Angaben in Bezug auf das geplante Forschungsprojekt getätigt werden sollen:

1. Datenbeschreibung
2. Dokumentation und Datenqualität
3. Speicherung und technische Sicherung während des Projektverlaufs
4. Rechtliche Verpflichtungen und Rahmenbedingungen
5. Datenaustausch und dauerhafte Zugänglichkeit der Daten
6. Verantwortlichkeit und Ressourcen

Diese werden im ersten Teil dieses Dokumentes fachspezifisch kommentiert.

Neben Forschungsvorhaben in der Einzelförderung bieten gerade auch koordinierte Verfahren eine hervorragende Plattform, um ein gemeinsames Forschungsdatenmanagement auf hohem Niveau

aufzubauen und Standards für Forschungscommunitys zu entwickeln. Insbesondere ortsverteilte Verbünde können hier prägend wirken. Daher werden im zweiten Teil dieses Dokumentes Use Cases aus dem Fachgebiet vorgestellt. Diese Best-Practice-Beispiele ergänzen die allgemeinen Kommentierungen im ersten Teil und werden in diesem Dokument laufend aktualisiert.

Es ist bewusst, dass der Schutz personenbezogener Daten sowie weitere gesetzliche Rahmenbedingungen (z.B. Exportkontrollen oder Dual-Use-Risken) den freien Zugang zu projektspezifischen Daten auch bei Einhaltung der FAIR-Prinzipien einschränken können.

Erster Teil: Fachspezifische Kommentierung der sechs zentralen Punkte der DFG-Checkliste zum Umgang mit Forschungsdaten

Bei jedem der sechs zentralen Punkte werden zunächst die Leitfragen der Checkliste zitiert. Auf diese folgen dann die jeweiligen fachspezifischen Kommentierungen. Diese sind stets als Anregungen zu verstehen. Sämtliche genannten Aspekte werden nicht in jedem Fall gleichermaßen relevant sein und entsprechend adressiert werden müssen.

1. Datenbeschreibung

Leitfragen
Auf welche Weise entstehen in Ihrem Projekt neue Daten? Werden existierende Daten wiederverwendet? Welche Datentypen, im Sinne von Datenformaten (z. B. Bilddaten, Textdaten oder Messdaten) entstehen in Ihrem Projekt und auf welche Weise werden sie weiterverarbeitet? In welchem Umfang fallen diese an bzw. welches Datenvolumen ist zu erwarten?

Kommentierung
Bereits im Zuge der Projektplanung soll abgeschätzt werden, welche Primär-, Sekundär- oder Metadaten wie, in welcher Größenordnung und in welchen Formaten erzeugt, verarbeitet und nach welchen Standards gespeichert werden müssen, und ob sich aus der Datenmenge oder -vielfalt besondere Herausforderungen für das Datenmanagement ergeben. Bei sehr großen Datenmengen kann gegebenenfalls eine Filterung der speicherbaren Datensätze notwendig werden, was in diesem Fall erläutert werden soll. Es sollte vermieden werden, dass die Speicherung von Daten aufwendiger ist als deren erneute Erzeugung.

2. Dokumentation und Datenqualität

Leitfragen
Welche Ansätze werden verfolgt, um die Daten nachvollziehbar zu beschreiben (z. B. Nutzung vorhandener Metadaten- bzw. Dokumentationsstandards oder Ontologien)? Welche Maßnahmen werden getroffen, um eine hohe Qualität der Daten zu gewährleisten? Sind Qualitätskontrollen vorgesehen und wenn ja, auf welche Weise? Welche digitalen Methoden und Werkzeuge (z. B. Software) sind zur Nutzung der Daten erforderlich?

Kommentierung
Für die Dokumentation der Daten sowie ihre Zuordnung und spätere Nutzung ist ihre eindeutige Identifikation unerlässlich. Hierbei ist eine Orientierung an fachüblichen Terminologien, Taxonomien, Semantiken oder Ontologien unter Zuhilfenahme sowohl administrativer als auch fachlicher Metadaten unerlässlich.

Die Datenqualität kann nicht einheitlich für alle Felder definiert werden. Daher soll in einem Forschungsantrag erläutert werden, wie die Datenqualität im Projekt definiert ist und welche konkreten Maßnahmen zur Gewährleistung einer durchgängig hohen Datenqualität – auch mit Blick auf eine spätere, möglichst weitgehende Nachnutzung – ergriffen werden. Dabei soll auch auf die zur (Nach-)Nutzung notwendigen Methoden und Werkzeuge eingegangen werden.

3. Speicherung und technische Sicherung während des Projektverlaufs

Leitfragen
Auf welche Weise werden die Daten während der Projektlaufzeit gespeichert und gesichert? Wie wird die Sicherheit sensibler Daten während der Projektlaufzeit gewährleistet (Zugriffs- und Nutzungsverwaltung)?

Kommentierung
Im Antrag sollte erläutert werden, wie die Speicherung der Forschungsdaten, ggf. in Verbindung mit zugehöriger Software, erfolgt. Eine geordnete Ablage und nachvollziehbare Speicherung sind anzuraten. Bestehende Vorgaben der jeweiligen Forschungseinrichtung sollen erwähnt werden. Generell ist für Forschungsdaten aus allen DFG-geförderten Vorhaben eine Archivierung für zehn Jahre verpflichtend (Kodex Leitlinien zur Sicherung guter Wissenschaftlichen Praxis, Leitlinie 17). Dies schließt sowohl analog wie digital vorliegende Daten ein.

In begründeten Einzelfällen mögen bestimmte Daten nicht aufbewahrt werden können. Diesbezügliche Abweichungen von den Standards der Community sind explizit im Antrag darzulegen.

4. Rechtliche Verpflichtungen und Rahmenbedingungen

Leitfragen
Welche rechtlichen Besonderheiten bestehen im Zusammenhang mit dem Umgang mit Forschungsdaten in Ihrem Projekt? Sind Auswirkungen oder Einschränkungen in Bezug auf die spätere Veröffentlichung bzw. Zugänglichkeit zu erwarten? Auf welche Weise werden nutzungs- und urheberrechtliche Aspekte sowie Eigentumsfragen berücksichtigt? Existieren wichtige wissenschaftliche Kodizes bzw. fachliche Normen, die Berücksichtigung finden sollten?

Kommentierung
Rechtliche Verpflichtungen und Rahmenbedingungen sollen bereits in der Planungsphase des Forschungsvorhabens überprüft und im Antrag benannt werden. Wo es Dritte gibt, die Rechte an den Daten haben, sollen sich daraus evtl. ergebende Einschränkungen dokumentiert werden.

5. Datenaustausch und dauerhafte Zugänglichkeit der Daten

Leitfragen
Welche Daten bieten sich für die Nachnutzung in anderen Kontexten besonders an? Nach welchen Kriterien werden Forschungsdaten ausgewählt, um diese für die Nachnutzung durch andere zur Verfügung zu stellen? Planen Sie die Archivierung Ihrer Daten in einer geeigneten Infrastruktur? Falls ja, wie und wo? Gibt es Sperrfristen? Wann sind die Forschungsdaten für Dritte nutzbar?

Kommentierung
Aus der Vorhabensbeschreibung soll klar hervorgehen, warum welche Daten als extern archivierungswürdig angesehen werden. Eine dauerhafte Zugänglichkeit der Daten bedingt nicht nur ihre Archivierung, sondern nach Möglichkeit auch einen verlässlichen Zugang (s. auch Kodex Leitlinien zur Sicherung guter Wissenschaftlichen Praxis, Leitlinie 13). Alle relevanten Daten sollen neben den Daten, die bereits über Publikationen zugänglich sind, dauerhaft zugänglich archiviert werden. Dazu gehören beispielsweise weitere projektbezogene Messungen oder Daten aus Messreihen, die zu einem Forschungsprojekt mit resultierenden Publikationen gehören, jedoch noch nicht konkret über eine Publikation einsehbar und nachnutzbar sind. Alle Daten oder Softwarepackete können auch separiert von Publikationen mit eigener DOI veröffentlicht werden.

Eine sinnvolle Auswahl dieser Daten soll mit Augenmaß eigenverantwortlich von den Projektleitenden (Antragstellenden) erfolgen und begründet sein. Der überwiegende Teil der Daten, insbesondere diejenigen, die auf etablierten Methoden und der Messung mit kommerziell erhältlichen Geräten basieren, lässt sich standardisiert handhaben. Eine Archivierung und dauerhafte Zugänglichkeit ist für diesen Großteil der Daten technisch gut realisierbar. Für einen geringen Teil von Daten, die nicht einem solchen Standard zugeordnet werden können, mag eine Archivierung und Gewährleistung von dauerhaftem Zugang aufgrund eines großen Aufwands, oder der Datengröße, nicht möglich oder praktikabel sein. In diesen Ausnahmefällen soll eine Erläuterung gegeben werden.

Es ist anzuraten, dass sich Antragstellende bereits vor Projektbeginn mit den projektspezifischen Anforderungen an ein Repositorium befassen, entsprechend geeignete Repositorien identifizieren und ggf. Kontakt zu deren Betreibenden aufnehmen. Im Zuge einer Kontaktaufnahme mit Serviceeinrichtungen sollen auch zeitliche und finanzielle Bedarfe geklärt werden, um diese rechtzeitig im Forschungsvorhaben berücksichtigen zu können.

Für eine spätere Nachnutzbarkeit sollen archivfähige, standardisierte Datenformate verwendet werden. Die technische Kompatibilität verschiedenster Datensätze und Softwarelösungen stellt eine weitere große Herausforderung des Datenmanagements dar. Sollten im Projekt Lösungen für eine praktikable

Umwandlung von Datenformaten entwickelt werden, ist es wünschenswert, diese Lösungen zur Nutzung bereitzustellen, um eine generelle Standardisierung von Datenformaten weiter voranzutreiben.

6. Verantwortlichkeiten und Ressourcen

Leitfragen
Wer ist verantwortlich für den adäquaten Umgang mit den Forschungsdaten (Beschreibung der Rollen und Verantwortlichkeiten innerhalb des Projekts)? Welche Ressourcen (Kosten; Zeit oder anderes) sind erforderlich, um einen adäquaten Umgang mit Forschungsdaten im Projekt umzusetzen? Wer ist nach Ende der Laufzeit des Projekts für das Kuratieren der Daten verantwortlich?

Kommentierung
Im Antrag soll klar dargelegt werden, wer im Projektverlauf für die einzelnen Schritte des Forschungsdatenmanagements zuständig ist. Die letztendliche Verantwortung für ein sorgfältiges Datenmanagement an sich sowie eine abschließende Qualitätskontrolle der im Projekt entstandenen Daten liegt grundsätzlich bei den Projektleitenden (Antragstellenden). Für Projektmitarbeitende, die Forschungsdaten in der Praxis erheben, ist eine umfangreiche Einweisung in das Forschungsdatenmanagement unerlässlich, um die Datenqualität sicherzustellen.
Die Projektbeschreibung soll ebenfalls eine Abschätzung des für das Datenmanagement nötigen Aufwands enthalten. Dem ist die hierfür zur Verfügung stehende institutionelle Grundausstattung und Infrastruktur gegenüberzustellen, gerade im Hinblick auf eine etwaige Beantragung zusätzlicher projektspezifischer Mittel.

Zweiter Teil: Use Cases aus dem Fachgebiet

Zur Veranschaulichung fachüblicher Standards und Entwicklung darauf aufbauender neuer Ideen für ein erfolgreiches Datenmanagement folgt die Beschreibung aktueller Use Cases aus dem Fachgebiet. In diese Aufstellung aufgenommen werden nur Projekte, die ein anhand der Veröffentlichung konkreter Daten praktisch gelebtes und durch mindestens eine Fortsetzungsbegutachtung qualitätsgesichertes Datenmanagement nachweisen können.

- Seite 5: SPP 2170
- Seite 8: SPP 2403
- Seite 11: SFB 1411

DFG SPP 2170 FAIR DATA MANAGEMENT

DFG SPP 2170 “InterZell” Koordination	427990970/ 2019-ongoing I. und II. Finanzierungsphase

DFG SPP 2170 “Novel Production Processes through Cross-Scale Analysis, Modeling and Design of Cell-Cell and Cell-Bioreactor Interactions” (InterZell)

Das Research Daten Management (RDM) des DFG-Schwerpunktprogramm SPP 2170 ist darauf ausgelegt, die enormen Datenmengen (Big Data) aus Experimenten, Simulationen und Modellierungen strukturiert zu verwalten, langfristig gemäß den Guidelines der DFG zu sichern [1] und die FAIR Prinzipien (Findable, Accessible, Interoperable, Reusable) [2] zu erfüllen. Der verbindlicher Datenmanagement Plan (DMP) wurde zu Beginn der I. Finanzierungsphase erstellt und publiziert [3]. Das Koordinationsprojekt bietet allen Projektpartnern den Zugriff auf ein Repositorium [3] und ermöglicht allen Projekten die Nutzung eines Daten Repositorium (DaRUS) für die Speicherung. Dafür wurde mit den institutionellen IT-Departments der Projektpartner
die entsprechenden Protokolle freigegeben und Zugriff auf das Repositorium gewährt. Für die Early Stage Researcher (ESR) wurden Schulungen und Workshops angeboten. Auf dem Kickoff der Finanzierungsphasen erhielten alle Projektpartner Überblick über das Datenmanagement durch Experten des Repositoriums der Universität Stuttgart. Das RDM des SPP2170 basiert auf den FAIR-Prinzipien [2] als Community Approach [4]. Unabhängig davon obliegt die Verantwortung für das Datenmanagement und der Umgang mit sensiblen Daten und Datenschutz den einzelnen durch die DFG geförderten Projektteilnehmern [5].

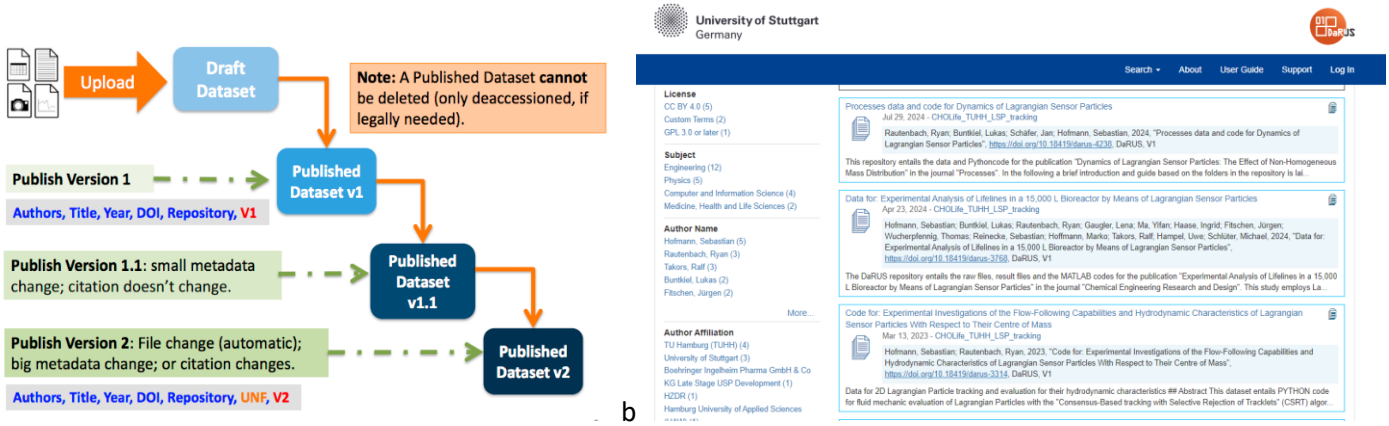


Fig. 1: DaRUS Focus Team etablierte den Daten-Publishing Prozess, entnommen. https://www.izus.uni-stuttgart.de/en/fokus/darus/publication/ (a). Beispiel DaRUS Publikation eines SPP 2170 Projekts unter Leitung von Prof. Ralf Takors, Universität Stuttgart, Prof. Michael Schlüter, TUHH (b)

FAIR Datenmanagement des SPP 2170 https://darus.uni-stuttgart.de/dataverse/InterZell-IBVT
Folgende Kernpunkte des SPP 2170 Research Datenmanagements (RDM) wurden vereinbart:

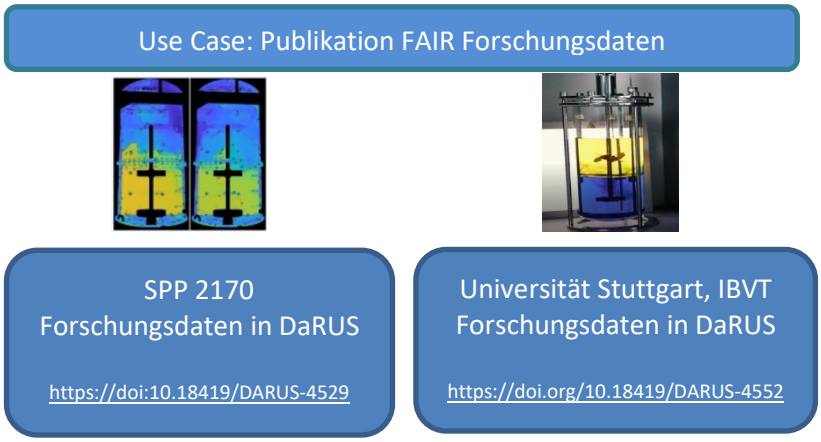
1. **Speicherung und zentrale Plattform (DaRUS):** Die Forschungsdaten können im Datenrepositorium der Universität Stuttgart (DaRUS) gespeichert werden, mit Access und Publikation (Fig. 1). Die Daten-Zwischenspeicherung erfolgt projektspezifisch bei den Projekten.
2. **Strukturierung, Datenbeschreibung, Dokumentation:** Die Beschreibung der Daten erfolgt ab Beginn des Data Life Cycle. Experimente und Modellierungen, sowie unstrukturierte Daten werden durch disziplinspezifische Kriterien und Metatags definiert. Semantik und verfügbare Ontologien sind anzuwenden.
3. **Verantwortlichkeit:** Die Koordination erfolgt durch Prof. Dr. Ralf Takors (IBVT Stuttgart) in Zusammenarbeit mit dem DaRUS-Team. Für das RDM der Projekte des SPP 2170 sind die jeweiligen Principal Investigators (PI) eigenständig verantwortlich.
4. **Zitierfähigkeit:** Veröffentlichte Datensätze im DaRUS-Repository sind für die weltweite Forschungsgemeinschaft zitierbar, sie erhalten eine DOI und sind mit ORCID verknüpft.
5. **Ressourcen:** Weitere Informationen sind im DMP erhältlich [3].
6. **Rechtliche Verpflichtungen und Rahmenbedingungen:** Jedes Forschungsprojekt innerhalb des SPP 2170 ist für die Einhaltung der Rahmenbedingungen und rechtlichen Verpflichtungen eigenständig zuständig.

Weitere inhaltliche Anforderungen

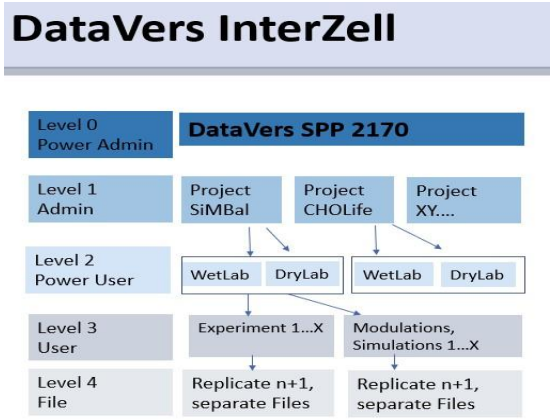
Gemäß den DFG-Leitlinien, berücksichtigen die Projekte des SPP 2170 folgende Aspekte:

i. **Dokumentation:** Diese erfolgt ab der Datenerzeugung in der frühesten Phase des Data Life Cycle.
ii. **Metadaten:** Ausführliche Beschreibung der Datensätze zur Nachnutzung ab Datenerstellung.
iii. **Ontologie und Semantik:** Alle SPP 2170 Projektteilnehmer nutzen das Template für Engineering/Life Science Daten im DaRUS.
iv. **Langzeitarchivierung:** Speicherung für mindestens 10 Jahre [1].
v. **Qualität:** Maßnahmen zur Sicherung der Datenqualität erfolgen bei Datenerstellung, Data Stewards sind für Support zuständig. Die PI´s sind für Qualitätskontrolle, Lizensierung und Handling sensibler Daten verantwortlich.
vi. **Proprietäre Daten** sind zu vermeiden.

Weitere Anforderungen sind im RDM des SPP 2170 aufgeführt [3]. Alle SPP 2170-Projekte beachten die Richtlinien zu interoperablen Daten und verwenden die bereitgestellten Metadaten, die in DaRUS als Template hinterlegt sind. Die Vorlagen für Dokumentation und Metadaten sind obligatorisch. Ontologien und Methoden werden im Template integriert und können übernommen werden.



a



b entnommen von [3]

https://doi.org/10.18419/DARUS-829

Fig. 2: Beispiele Best Practice Datenmanagement in Bioengineering, Repositorium DaRUS (2a) und Struktur des SPP 2170 mit Dataverses und Administration (2b)

References

[1] DFG Guidelines https://www.dfg.de/resource/blob/169298/ua-inf-empfehlungen-200901.pdf
[2] Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18
[3] Rehnert, M., Takors, R. 2022, "Data Management of a Biotechnology Network as a Contribution to FAIR Data Management", https://doi.org/10.18419/DARUS-829, DaRUS, V1
[4] Rehnert, M., Takors, R. 2023: Fair research data management as community approach in bioengineering. Engineering in Life Sciences, Volume 23, Issue 1 e2200005 https://doi.org/10.1002/elsc.202200005
[5] Rehnert, M., Takors, R. 2026: Managing complexity in research data management for heterogeneous biotechnological data: a perspective on interacting as data steward. *Bioprocess Biosyst Eng* 49, 2321–2331. https://doi.org/10.1007/s00449-026-03380-0.

Acknowledgement

The support of this work by the German Research Foundation (DFG) within the priority program “InterZell” (SPP2170) is gratefully acknowledged. We thank our colleagues of the SPP2170 and the DaRUS team who provided training and expertise that greatly assisted the use case.

SPP 2170 Use Case im Überblick

Data Storage	Erfolgt als FAIR RDM sowie nach den Erfordernissen der projekteigenen Institutionen.
Origin of the data:	Labor data, Modelling/Simulation data, omics, Prozessdaten etc.
Daten Format:	Heterogene Daten wie Programm code, pdf, csv, images, etc. sind möglich. Proprietäre Daten sollten vermieden werden.
Repository:	Das Repository DaRUS ist seit 2019 für den SPP 2170 verfügbar.
Datasets	Für Datasets > 100 GB wurde eine neue SOP erstellt, Datasets < 100 werden über die API gespeichert.
Struktur	Metadaten-Templates sind vorhanden. Die Struktur für die Dataverses wurden nach dem Kickoff bereitgestellt.
DaRUS Access	Alle Mitglieder des SPP erhalten Access. Eine Anmeldung ist erforderlich.

Das RDM erfolgt über den gesamten Daten Lebenszyklus. Es wurde ein verbindlicher DMP erstellt [3].

Schulung und Kommunikation: Alle Mitglieder des SPP 2170 erhielten Schulungen zu den Grundsätzen des Datenmanagements: Promovierende wurden im Rahmen der Summer School 2020/2024 geschult. Alle Mitglieder des SPP 2170 wurden beim Kick-off des Programms über die FAIR-Prinzipien und DaRUS und entsprechende Richtlinien informiert. Diese Schulungen wurden vom FOCUS-Team, Universität Stuttgart für das SPP-2170-Programms durchgeführt. Zugang zu den entsprechenden Ressourcen und Lernmaterialien wurde durch das FOCUS Team gewährleistet. Für Support steht der SPP 2170 Koordinator zur Verfügung.

Verantwortlichkeiten und Lokales Zugriffs-Management: Allen PI´s wurde die Plattform DaRUS vorgestellt. Für jede einzelne Institution wurde der Kontakt mit der IT-Abteilung hergestellt und über EPN eingebunden. Dadurch sind die über Deutschland verteilten Institutionen berechtigt, auf das Repositorium der Universität Stuttgart zuzugreifen. DaRUS kann zur Veröffentlichung von Daten genutzt werden, die von den Forschenden innerhalb ihrer Projekte gespeichert wurden.

Access und Qualitätskontrolle erfolgt über die Universität Stuttgart gemäß den FAIR-Prinzipien Die zehn Projekte innerhalb des DFG-Schwerpunktprogramms SPP 2170 sind jeweils für ihre selbst generierten sowie gemeinsam genutzten und verteilten Daten verantwortlich. Dabei orientieren sie sich an den FAIR-Datenprinzipien, den DFG-Richtlinien sowie den geltenden Sicherheitsstandards. Die Vergabe von Persistent Identifikatoren (DOI) erfolgt im DaRUS. ORCID Felder sind im DaRUS verfügbar und deren Nutzung wird empfohlen. Die Struktur und Administrierung wurde durch den Speaker und Koordinator erstellt (Fig. 2b) und konnte auf weitere Anwendungsfälle ausgeweitet werden (2a).

Änderungs- und Risikomanagement: Die Verantwortung für die Daten liegt bei den jeweiligen Projektleitungen (Project Owner), entsprechend der mit dem Fördergeber vereinbarten Regelungen.

FAIR-Datenmanagement und Daten Lebenszyklus: Die Projekte sind über den gesamten Lebenszyklus der Daten hinweg für die Umsetzung des FAIR-Datenmanagements verantwortlich. Support erfolgt durch die Koordination des SPP 2170. Alle Projekte des SPP 2170 stellen ihre Forschungsdaten gemäß den DFG-Richtlinien bereit. Das Datenmanagement ist in verschiedene Hierarchielevel untergliedert. Labordaten (z. B. Experimente, Sequenzierung, Bioreaktor-Kultivierung, Prozessdaten), generelle Daten (z. B. Modellierung, Simulation, Software, Programmcode) sind für die Projektpartner zugänglich (Power-User-Zugriff auf Level 2). Alle SPP 2170-Projekte stellen sicher, dass ihre Daten gemäß den DFG-Richtlinien FAIR sind. Es liegt in der Verantwortung jedes der zehn Projekte, FAIR RDM anzuwenden.

Use Case: Pragmatisches Forschungsdatenmanagement im SPP 2403 „Carnot Batteries: Inverse Design from Markets to Molecules“

Das Forschungsdatenmanagement im SPP 2403 „Carnot Batteries: Inverse Design from Markets to Molecules“[1] ist als pragmatischer, alltagstauglicher Arbeitsablauf angelegt. Das Programm umfasst 17 Teilprojekte mit experimentellen und simulationsbasierten Arbeiten. Entsprechend heterogen sind die entstehenden Daten: Messreihen, Daten zu Fluiden und Fluidgemischen, Simulationsdaten aus CFD- und MD-Rechnungen, Auswerteskripte, Programmcodes, Parameterdateien, Labor- und Versuchsdokumentationen sowie Daten aus Energiesystem- und Marktmodellen. Ziel des Forschungsdatenmanagements ist daher nicht, allen Teilprojekten ein einheitliches System aufzuzwingen, sondern Daten so zu dokumentieren, intern zu teilen und zu veröffentlichen, dass sie auch nach längerer Zeit noch verständlich sind und von anderen sinnvoll genutzt werden können.

Eine wichtige **Prämisse** im SPP ist, dass Forschungsdatenmanagement nur dann funktioniert, wenn es für die Forschenden selbst einen klaren Nutzen hat. In Schulungen und Diskussionen wurde deshalb bewusst nicht nur auf gute wissenschaftliche Praxis und FAIR-Prinzipien verwiesen, sondern auch darauf, dass gute Dokumentation den eigenen Arbeitsalltag erleichtert: Man kann auch nach Monaten nachvollziehen, welches Experiment, welcher Parametersatz, welcher Simulationslauf oder welches Modellszenario zu welchem Zweck durchgeführt wurde und zu welchem Ergebnis geführt hat. Publikationen lassen sich schneller vorbereiten, Rückfragen von Kooperationspartnern oder Gutachtenden sind leichter zu beantworten, und beim Wechsel von Mitarbeitenden geht weniger Wissen verloren. Gerade wegen des Zeitdrucks im Forschungsalltag muss Forschungsdatenmanagement möglichst früh und einfach in die normalen Abläufe eingebaut werden.

Die Koordination des Forschungsdatenmanagements erfolgt im SPP durch einen **Data Steward** mit einer 25%-Stelle. Er hat die Datenpolitik des SPP beim Kick-off-Workshop vorgestellt und anschließend in Online-Tutorials und Workshops vertieft, unter anderem zur Nutzung von Coscine[2]. Seine Rolle ist nicht, die fachliche Verantwortung für die Daten zu übernehmen – diese bleibt bei den erzeugenden Wissenschaftlerinnen und Wissenschaftlern und den jeweiligen PIs. Der Data Steward unterstützt durch Schulung, Beratung, Vorschläge für Metadatenschemata und die konkrete Kuratierung von Datensätzen vor der internen Weitergabe oder Veröffentlichung. Er ist zudem leicht erreichbare Ansprechperson, informiert regelmäßig über neue Entwicklungen und stellt gelungene Beispiele aus dem SPP vor. Diese persönliche Unterstützung hat sich als sehr hilfreich erwiesen, weil viele Probleme erst am tatsächlich vorliegenden Datensatz sichtbar werden und nicht durch allgemeine RDM-Regeln lösbar sind. Hilfreich bei Auswahl, Planung und Einrichtung der FDM-Plattformen war zudem der *Research Data Service* der Universität Duisburg-Essen, der insbesondere in der Anfangsphase beratend zur Seite stand.

Der empfohlene **FDM-Workflow** besteht aus der Nutzung weniger, gut bekannter Werkzeuge. Für die tägliche Dokumentation von Experimenten, Messungen und teilweise auch Simulationsläufen wird eLabFTW[3] empfohlen; dort werden neben der Beschreibung auch Metadaten, Stichworte und Verweise auf Dateien, Programmversionen, Commits oder Auswertungen festgehalten. Für Codeentwicklung, Auswerteskripte und Dokumentation wird GitLab[4] genutzt, gegebenenfalls ergänzt durch ReadTheDocs. Besonders hilfreich sind hier verständliche Commit-Kommentare, README-Dateien und klare Hinweise, welche Codeversion zu welchem Datensatz oder welcher Publikation gehört.

[1] https://www.uni-due.de/spp2403/
[2] https://coscine.rwth-aachen.de/
[3] https://www.elabftw.net/
[4] https://about.gitlab.com/

Für den **internen Austausch** im SPP und die Ablage ausgewählter Forschungsdaten nutzen wir Coscine, das in der von uns genutzten RWTH-Umgebung bisher kein öffentliches Repositorium ist. Für unseren Zweck ist das passend: Coscine dient als geschützter Arbeitsbereich für den Datenaustausch zwischen Gruppen. Gerade in einem ortsverteilten Verbund ist es sonst nicht trivial, Daten strukturiert und mit Zugriffsrechten nur zwischen Teilprojekten auszutauschen. In Coscine können auch noch nicht publikationsreife Datensätze abgelegt werden, die für andere Teilprojekte bereits nützlich sind; Metadaten können früh mitgeführt und später verbessert werden. Falls Coscine künftig auch öffentliche Daten unterstützt, kann dies eine zusätzliche Option werden.
Für **Veröffentlichungen** wurde eine Zenodo-Community des SPP[5] eingerichtet, in der publikationsbezogene Daten und teilweise auch Code mit DOI bereitgestellt werden. Zum Ende der ersten Förderperiode enthält sie acht Einträge. Bei 17 Teilprojekten zeigt dies, dass der Workflow genutzt wird, aber auch, dass die Veröffentlichung kuratierter Datensätze Zeit braucht und oft an den Publikationsfortschritt gekoppelt ist. Einige Gruppen stellen Daten weiterhin als Supplementary Material bereit; wir sehen dies nicht als grundsätzlich schlechtere Lösung. Zenodo hat Vorteile bei DOI, Zitierbarkeit, Versionierung und Metadaten, Supplementary Material ist dafür unmittelbar mit dem Artikel verbunden. Wichtig ist in beiden Fällen, dass die Daten auffindbar, verständlich beschrieben und mit der Publikation verknüpft sind. Ein grundsätzliches Problem bleibt die **Auffindbarkeit von Datensätzen**: Anders als bei Publikationen gibt es keine etablierten, zentralen Suchmaschinen oder Indizes wie Scopus oder Web of Science für Forschungsdaten. Wir versuchen dies durch aussagekräftige Titel, Abstracts und Keywords in Zenodo zumindest teilweise auszugleichen.
Coscine und Zenodo haben sich bisher als zwei komplementäre Bausteine erwiesen, nicht als durchgängige Pipeline: Coscine dient dem geschützten internen Austausch mit Metadatenschema, Zenodo-Datensätze entstehen dagegen meist publikationsnah und werden gezielt für die Veröffentlichung zusammengestellt. Ein direkter Transfer eines Datensatzes von Coscine nach Zenodo hat bisher nicht stattgefunden – die Anforderungen an einen intern nützlichen und einen veröffentlichten Datensatz überschneiden sich, sind aber nicht identisch.
Entsprechend unterscheiden sich auch die **Metadaten**: In Coscine nutzen wir explizite Metadatenschemata, in Zenodo dagegen vor allem die dort vorgesehenen Felder (Titel, Autorinnen und Autoren, Beschreibung, Keywords, Lizenz, DOI, Verknüpfungen) sowie ausführliche, menschenlesbare **README**-Dateien. Separate maschinenlesbare Metadatendateien wie metadata.yaml oder metadata.json wurden bisher nicht systematisch beigelegt. Uns war es bisher wichtiger, verständliche und tatsächlich nutzbare Datensätze zu veröffentlichen, als die Einstiegshürde durch zusätzliche formale Anforderungen zu erhöhen; zusätzliche metadata.yaml-Vorlagen sollen aber als nächstes getestet werden.
Ein zentrales Element ist ein **einfaches, weiterentwickelbares Metadatenverständnis** statt eines für alle Teilprojekte identischen Schemas. Der Data Steward erarbeitet zusammen mit den Gruppen Vorschläge, die zum jeweiligen Datentyp und Forschungsgebiet passen: In experimentellen oder thermodynamischen Arbeiten etwa Angaben zu Fluiden, Temperatur-, Druck- und Leistungsbereichen, Geräten und Messbedingungen; bei CFD- oder MD-Simulationen Geometrien, Randbedingungen, Kraftfelder, Zeitschritte, Konvergenzkriterien und Softwareversionen; in Energiesystem- und Marktmodellen modellierte Regionen, Jahre, Szenarien sowie Annahmen zu Preisen, Lastprofilen und Technologien. Solche fachnahen Angaben sind oft wichtiger als ein formal perfektes, aber inhaltlich wenig hilfreiches Schema.
Meist reichen wenige gut gewählte **Angaben**: Wer hat die Daten erzeugt, zu welchem Teilprojekt und welcher Publikation gehören sie, warum wurde die Untersuchung durchgeführt, was wurde gemessen, simuliert oder modelliert, welche Materialien, Parameter, Softwareversionen oder Randbedingungen waren wichtig, welche Einheiten haben die

[5] https://zenodo.org/communities/spp-2403-carnot-batteries

Variablen, welche Datei enthält Rohdaten, welche ausgewertete Daten oder Skripte, und wer beantwortet Rückfragen? Solche Angaben sind schnell erstellt, machen aber einen großen Unterschied für die spätere Nutzung. Auf ein strikt einheitliches Vokabular wurde bewusst verzichtet, weil die Hürde niedrig bleiben sollte; das wurde in der Praxis nicht bereut.
Bei der Datenweitergabe setzen wir auf **einfache, allgemein nutzbare Formate**: Datensätze werden meist als ZIP-Archive mit README-Datei, Unterordnern, Tabellen und Skripten bereitgestellt, bevorzugt in nicht-proprietären Formaten wie CSV oder TXT. Die Daten sollten möglichst ohne Spezialsoftware sichtbar sein; nach Möglichkeit sollte nicht erst ein Skript ausgeführt werden müssen, um zu erkennen, was enthalten ist. Spezielle Formate werden nur verwendet, wenn sie für die jeweilige Community einen klaren Vorteil bieten. Eine vollständig standardisierte Ordnerstruktur gibt es bisher nicht, dafür sind die Datenarten zu unterschiedlich und die Einstiegshürde soll niedrig bleiben.
Die **Kuratierung** erfolgt vor allem im Dialog: Der Data Steward sieht Datensätze in Coscine oder Zenodo durch, prüft Metadaten, README-Dateien, Dateinamen, Keywords und Verständlichkeit und gibt Rückmeldung an die Datengeber. Besonders hilfreich ist es, wenn zusätzlich eine fachkundige Kollegin oder ein fachkundiger Kollege versucht, einen *exemplarischen Datensatz nachzuvollziehen*: Ist klar, welche Datei wofür gedacht ist, sind Einheiten und Abkürzungen erklärt, lässt sich eine zentrale Abbildung mit den bereitgestellten Daten und Skripten reproduzieren? Solche einfachen Prüfungen zeigen schnell, ob ein Datensatz nur für die eigene Gruppe verständlich ist oder auch für andere.
In der praktischen Kuratierung haben sich **KI-Werkzeuge** als hilfreiche Arbeitserleichterung erwiesen – nicht zur ungeprüften Erstellung von Metadaten, aber um aus Stichpunkten, einer Ordnerstruktur oder kurzen Dateibeschreibungen erste Entwürfe für README-Dateien, Variablenerklärungen, Abstracts oder Keywordlisten zu erstellen, welche anschließend von den erzeugenden Wissenschaftlerinnen und Wissenschaftlern geprüft und korrigiert werden. Dies spart spürbar Zeit und senkt die Hemmschwelle, Datensätze verständlich zu beschreiben.
In der **Kommunikation** mit den Teilprojekten war wichtig zu vermitteln, dass nicht automatisch alle Rohdaten veröffentlicht werden müssen – hier bestanden anfangs Sorgen. Die Entscheidung, welche Daten weitergegeben werden, liegt bei den erzeugenden Wissenschaftlerinnen und Wissenschaftlern unter Verantwortung der jeweiligen PIs. Sie beurteilen am besten, welche Daten für andere nützlich, für die Nachvollziehbarkeit einer Publikation notwendig oder rechtlich beziehungsweise praktisch problematisch sind. Ziel ist nicht die maximale Datenmenge, sondern brauchbare und gut erklärte Datensätze.
Nicht alle Gruppen im SPP nutzen die empfohlenen Werkzeuge im gleichen Umfang; einige folgen **etablierten lokalen Abläufen**. Das wird akzeptiert, solange nachvollziehbare Dokumentation, sichere Speicherung, interne Austauschbarkeit und Veröffentlichung geeigneter Daten gewährleistet sind. Diese Flexibilität ist wichtig, weil ein zu starres System eher dazu führt, dass Forschungsdatenmanagement als zusätzliche Last wahrgenommen wird.

Als bisherige **Erfolgsfaktoren** sehen wir vor allem vier Punkte: Erstens muss der Nutzen für die Forschenden selbst deutlich werden. Zweitens braucht es einfache Werkzeuge, die in die tägliche Arbeit passen. Drittens hilft ein Data Steward erheblich bei Schulung, Metadatenvorschlägen und Kuratierung. Viertens sollten Daten nicht erst am Projektende „aufgeräumt“ werden – je früher Kontextinformationen erfasst werden, desto geringer der spätere Aufwand. Für die nächste Phase planen wir kurze Checklisten für Coscine-Ablagen und Zenodo-Veröffentlichungen sowie einfache Vorlagen für README-Dateien, metadata.yaml-Dateien und typische Ordnerstrukturen. Auch die Planung der Datenübergabe bei Personalwechseln ist ein nächster Schritt. Diese Punkte ergeben sich aus den bisherigen Erfahrungen im SPP. Weiterhin wäre eine bessere übergreifende *Auffindbarkeit* von Forschungsdatensätzen wünschenswert, etwa vergleichbar mit Scopus oder Web of Science für Publikationen.

Forschungsdatenmanagement im SFB 1411

Einordnung und Zielsetzung

Forschungsdatenmanagement (Research Data Management, RDM) ist ein integraler Bestandteil der wissenschaftlichen Wertschöpfungskette und wurde im Sonderforschungsbereich SFB1411 im Rahmen eines eigenen Projektes (Informationsmanagement und Informationsinfrastruktur, INF) umgesetzt. Aus unserer Erfahrung sollte ein tragfähiges Konzept für das Forschungsdatenmanagement dabei folgende Ebenen zusammenführen:

- *Orientierung an den FAIR-Prinzipien (Findability, Accessability, Interoperability, Reusability)*
- *Strukturierte und einheitliche Erfassung von Metadaten*
- *Nutzung eines elektronischen Laborbuchs*
- *Digitale Verknüpfung von Forschungsgruppen und Prozessschritten*
- *Integration in zentrale Hochschulstrukturen*
- *Strategien zur Langzeitspeicherung*
- *Publikation von Roh- und Prozessdaten*

Das übergreifende Ziel ist es, Forschungsdaten über ihren gesamten Lebenszyklus konsistent zu erfassen, zu beschreiben, zu speichern und entlang von Prozessketten zu verknüpfen. Im Zentrum steht die Integration experimenteller und simulativer Daten - im Beispiel von SFB1411 von der Partikelsynthese über die Charakterisierung bis zur Modellierung. Dadurch werden Partikelsysteme und ihre Prozessparameter mit Charakterisierungsergebnissen, daraus entstehenden Materialien und parallelen Simulationsergebnissen eindeutig verknüpft, langfristig nachverfolgbar gemacht und für eine spätere Nutzung zugänglich gehalten. Zu Beginn der ersten Förderperiode des SFB (2020) wurde openBIS als beste Lösung für das interdisziplinäre Konsortium identifiziert: eine flexible Open-Source-Plattform, die Electronic Lab Notebook (ELN)- und Laboratory Information Management System (LIMS)-Funktionalitäten vereint und eine strukturierte Verknüpfung von Daten und Metadaten ermöglicht.

Implementierung von RDM im SFB 1411

Das entwickelte RDM erweist sich als praxistauglich. Als technischer Kern wurde openBIS als ELN- und Datenmanagementplattform etabliert und zunächst vom INF-Team gemeinsam mit einer Pilotgruppe aus Promovierenden, Postdocs und PIs auf die spezifischen Anforderungen des SFB zugeschnitten. Bewährt hat sich insbesondere die Möglichkeit, Daten nicht nur abzulegen, sondern über Metadaten, Objektstrukturen und Parent-Child-Beziehungen entlang von Prozessketten zu verknüpfen. Dadurch werden experimentelle und simulative Arbeitsschritte nachvollziehbar miteinander verbunden und Forschungsergebnisse reproduzierbar dokumentiert.

Use Case: Durchgängige Prozesskette von Experiment zu Simulation

Ein zentraler Anwendungsfall des im SFB 1411 etablierten RDM ist die durchgängige Abbildung von Prozessketten und interagierenden Projekten. Ausgangspunkt bildet häufig die Synthese eines Partikelsystems. Die dabei entstehenden Proben werden anschließend in einem Charakterisierungsprojekt detailliert analysiert, beispielsweise mittels tomographischer Verfahren. Die so erzeugten Bild- und Strukturdaten werden zusammen mit den zugehörigen Metadaten, wie Messparametern und Probenhistorie, in der Plattform openBIS gespeichert. Abbildung 1 zeigt exemplarisch einen Eintrag in der openBIS-Oberfläche und veranschaulicht, wie Daten und zugehörige Metadaten innerhalb der Plattform strukturiert erfasst und dargestellt werden.

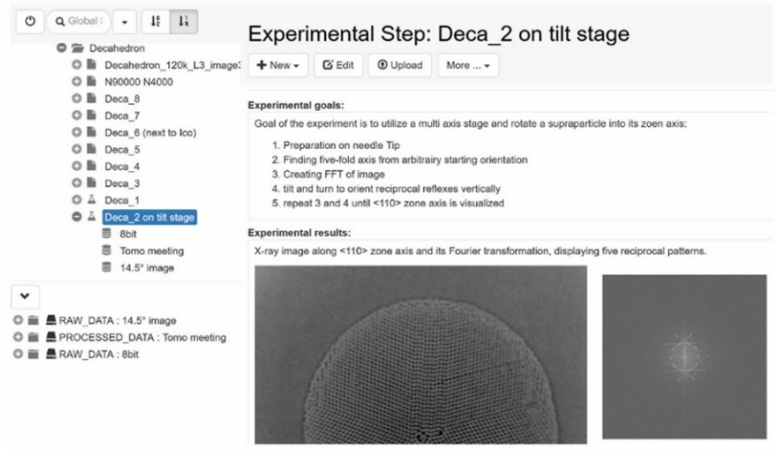


Abbildung 1: Benutzeroberfläche von openBIS aus der Perspektive der Forschenden. Im oberen linken Bereich ist die hierarchische Ordnerstruktur (Projekte, Experimente, Datensätze) dargestellt. Auf der rechten Seite werden strukturierte Metadaten (z.B. Probenparameter, Messbedingungen, Vorgehensweise) erfasst und mit den zugehörigen Daten verknüpft. Im gezeigten Beispiel handelt es sich um Röntgenmikroskopie-Daten, die als TIFF-Dateien vorliegen; sowohl die Rohdaten als auch die prozessierten Datensätze sind im System hinterlegt und werden im unteren linken Bereich visualisiert, sodass eine direkte Gegenüberstellung und Nachvollziehbarkeit der Verarbeitungsschritte möglich ist. Abbildung reproduziert aus [1].

Die gewonnenen Partikel- und Strukturinformationen werden in nachgelagerten Simulationsprojekten weiterverwendet. Dort dienen sie als Eingangsgrößen für numerische Modelle, mit denen beispielsweise physikalische Eigenschaften der Partikel berechnet werden. Die Simulationsergebnisse werden wiederum im System abgelegt und mit den zugrunde liegenden experimentellen Daten verknüpft. Durch die konsequente Nutzung von Metadaten und die Organisation entlang von Prozessketten entsteht eine durchgängige, nachvollziehbare Datenstruktur, in der alle Schritte über sogenannte Parent-Child-Beziehungen miteinander verbunden sind. Abbildung 2 zeigt exemplarisch eine solche Organisation anhand von Herstellung, Charakterisierung und Simulation kolloidaler photonischer Kristalle. Dies ermöglicht nicht nur die vollständige Reproduzierbarkeit einzelner Ergebnisse, sondern auch die gezielte Nachnutzung von Daten in anderen Kontexten.

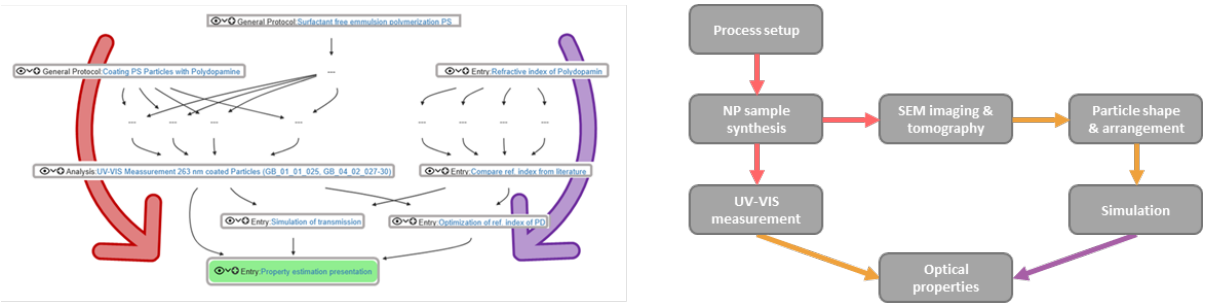


Abbildung 2: Schematische Darstellung der durchgängigen Prozesskette im SFB 1411. Die Abbildung zeigt die Zusammenarbeit zwischen Synthese, Charakterisierung und Simulation innerhalb eines gemeinsamen Projekts. Links ist ein Prozessgraph dargestellt, der die durchgängigen Workflows von Synthese (roter Pfad) und Simulation (violetter Pfad) abbildet; zur besseren Übersicht ist eine vereinfachte Darstellung rechts gegeben, bei der zusätzliche Verknüpfungen mit Charakterisierungsschritten (orangene Pfeile) angegeben sind. Abbildung angepasst aus [1].

Zentrale Aufgabengebiete

Die zentrale Erfahrung der Nutzung der RDM-Plattform durch die Pilotgruppe war die Notwendigkeit der Einbettung in klare organisatorische Abläufe (Workflows, Zuständigkeiten) um eine wirksame Nutzung zu gewährleisten. Daraus ergeben sich folgende zentrale Aufgabengebiete des INF-Teams neben der eigentlichen Etablierung und Pflege der Plattform:

- *Data Stewardship:* Die Schaffung und Umsetzung organisatorischer Abläufe und Aktivierung aller Beteiligter ist notwendig, um die Nutzung der RDM Plattform zu gewährleisten. Dies wurde SFB 1411 wurde durch folgende Maßnahmen realisiert. i) Regelmäßiges Vorstellen der RDM-Aktivitäten und Plattformen in Seminaren und Retreats sowie dezidierte Workshops und Schulungen zur Bedienung der Plattformen. ii) Dauerhafte und niederschwellige Zugänglichkeit von Beratung und Unterstützung durch wissenschaftliche Mitarbeiter, um Hierarchien in Anfragen zu vermeiden, sowie offene Zoom-Sprechstunden ohne Terminkoordination. iii) Verfügbarkeit von Hands-on Tutorials, Videoanleitungen, schriftliche Handreichungen und Checklisten zum Selbstabgleich aller Mitglieder, welche unter „RDM How to: Zenodo data publications within CRC 1411“, „CRC 1411 Zenodo Data Publication Videos [SerienID : 4666]“ und „CRC 1411 OpenBIS Introduction Videos [SerienID : 4658]“ verfügbar sind. Das INF-Team fungiert in der Praxis damit als Data-Stewarding-Struktur mit direkter Ansprechbarkeit für alle Forschenden. Die bereitgestellten Materialien decken zentrale Aspekte ab, darunter Systemeinführung, Arbeiten mit Objekttypen und Metadaten, Datenhierarchien sowie Upload-, Export- und Rollenmodelle. Ergänzend werden typische Anwendungsfälle (z. B. Datenverknüpfung, Zenodo-Publikation) anhand kompakter, praxisnaher Einheiten vermittelt.
- *Metadaten Vereinheitlichung:* Die Entwicklung geeigneter (Meta-) Datenstrukturen ist notwendig für einheitliche Nutzung von Daten. Neben standardisierten Formaten wurden aus der Pilotgruppe Herausforderungen identifiziert und projektspezifische Lösungen entwickelt, um komplexe Prozessketten durch einheitliche Datenformate und Metadatenstrukturen abzubilden. Beispielhaft entwickelte Datenformate – konkret für die Datenformate der Analysetechniken UV-VIS **.uvvis* und Partikelgrößenverteilung **.psd* – können über „RDM How to: Exemplary standardized CRC 1411 data formats“ eingesehen werden. Diese werden kontinuierlich abgestimmt und weiterentwickelt, und fließen perspektivisch in Nationale Forschungsdateninfrastruktur (NFDI) Initiativen ein.
- *Qualitätssicherung:* Das INF Team gewährleistet eine Qualitätssicherung veröffentlichter Forschungsdaten. Dazu wurde ein klar geregelter Publikationsworkflow etabliert. Datenpublikationen werden parallel zur wissenschaftlichen Veröffentlichung vorbereitet, beim INF-Team eingereicht und in einem iterativen Prozess kuratiert, um einheitliche Struktur und Qualitätsstandards zu gewährleisten. Das iterative Vorgehen zeigt dabei Probleme auf, die nachgehend in Workshops und Seminaren thematisiert und verbessert werden können. Nach Freigabe erfolgt die Veröffentlichung (z. B. über Zenodo: CRC 1411 Zenodo Community) mit DOI und Verknüpfung zur Publikation. Aktuell umfasst dies 69 Datensätze (62 Datensätze zu Publikationen,

4 Preprints, 1 Technical Notes, 2 Software) mit insgesamt 9710 Downloads (Stand: 28.05.2026), und belegt so eine aktive Nachnutzung in der wissenschaftlichen Community.

Strukturbildende Wirkung

Das etablierte RDM entfaltet Wirkung über den SFB hinaus und trägt zur Entwicklung institutioneller RDM-Strukturen an der FAU maßgeblich bei. Der im SFB eingerichtete openBIS-Server wurde inzwischen universitätszentral in den Betrieb des *FAU Competence Center for Research Data and Information* (FAU CDI) überführt, welcher heute Betrieb, Wartung und Hosting übernimmt, und damit eine langfristige tragfähige Infrastruktur sichert. Die im INF-Projekt aufgebaute Expertise wurde zudem nachfolgenden Verbundprojekten genutzt, darunter SFB 1452, SFB 1719 und RTG 3103, bei denen die Expertise in CRC1411 durch Beratung, Training und Transfer katalysierend wirkte. Weitere Einheiten (Arbeitsgruppen, Lehrstühle) werden durch operative Unterstützung, Beratung und Schulungen unterstützt, um die strukturelle Verankerung von RDM an der FAU zu gewährleisten. Parallel dazu wird ein aktiver Austausch mit NFDI-Konsortien, wie z. B. FAIRmat gepflegt. Die im SFB gewonnenen Erfahrungen fließen in die Weiterentwicklung und Harmonisierung offener, einheitlicher Daten- und Metadatenformate ein, wie sie in mehreren NFDIs vorangetrieben werden.

Herausforderungen und *Lessons Learned*

- *Plattformauswahl*: Eine erste zentrale Entscheidung betrifft die Wahl eines geeigneten ELN. Die Herausforderung besteht darin, ein System zu finden, das einerseits flexibel genug ist, um unterschiedliche Fachkulturen und Datentypen abzubilden, andererseits aber ausreichend Struktur für eine einheitliche Datendokumentation vorgibt. Zum Zeitpunkt der Implementierung im SFB 1411 (2020) war openBIS für das Konsortium die geeignete Lösung, da es als open-source ELN-LIMS-System an die Anforderungen des Verbundes angepasst und in die vorhandenen Strukturen eingebunden werden konnte. Da sich unterschiedliche ELN-LIMS-Systeme jedoch sehr dynamisch weiterentwickeln, muss die Plattformwahl zum jeweiligen Implementierungszeitpunkt sorgfältig neu geprüft werden. Dabei sind nicht nur technische Funktionen, sondern auch Fragen der langfristigen Wartung, der Anschlussfähigkeit an Hochschulstrukturen und der Einbindung in übergreifende Forschungsdateninfrastrukturen zu berücksichtigen.
- *Interdisziplinarität des Konsortiums*: In einem interdisziplinären Verbund treffen unterschiedliche Fachsprachen, Datenpraktiken und Anforderungen aufeinander. Die Herausforderung besteht darin, gemeinsame Strukturen zu schaffen, ohne die fachliche Spezifik einzelner Teilprojekte zu verlieren. Besonders zu beachten ist daher, dass Datenformate und Metadatenstrukturen von allen beteiligten Gruppen verstanden und genutzt werden können. Damit Daten nicht nur abgelegt, sondern verstanden, verknüpft und nachgenutzt werden können, müssen gemeinsame (Meta-)Datenstrukturen entwickelt werden.
- *Aktivierung der Mitwirkenden*: Eine zentrale Herausforderung besteht darin, Forschende zur konsequenten Nutzung der Infrastruktur zu motivieren und den konkreten Nutzen im Arbeitsalltag sichtbar zu machen. Besonders zu beachten ist, dass zusätzliche Schritte wie Metadatenannotation, Datenstrukturierung und Qualitätssicherung zunächst als Mehraufwand wahrgenommen werden können. Im SFB wurde dem durch persönliche Ansprechbarkeit des INF-Teams, regelmäßige Schulungen, Hands-on-Tutorials, Videoanleitungen und praxisnahe Beispiele begegnet. Pilotanwendungen und konkrete Use Cases halfen dabei, die Vorteile der Plattform für Zusammenarbeit, Nachvollziehbarkeit und Datenpublikation sichtbar zu machen.
- *Qualitätssicherung:* Datenpublikationen erfordern eine verlässliche Prüfung von Struktur, Metadatenqualität und technischer Nachnutzbarkeit, ohne die wissenschaftliche Verantwortung der Autorinnen und Autoren zu ersetzen. Besonders zu beachten ist, dass Datenpublikation frühzeitig mit der Fachpublikation mitgedacht werden muss. Dafür wurde im SFB ein verbindlicher Publikationsworkflow etabliert. Dadurch wird Datenpublikation zu einem planbaren Bestandteil wissenschaftlicher Veröffentlichung und nicht zu einem nachgelagerten Zusatzaufwand.
- *Große Datenmengen:* Große und heterogene Datenmengen stellen weiterhin eine technische und organisatorische Herausforderung dar. Nicht alle Daten lassen sich gleichermaßen effizient im ELN selbst abbilden, visualisieren oder verarbeiten.

[1] F. Plass, S. Englisch, B. Apeleo Zubiri, L. Pflug, E. Spiecker, M. Stingl: *Using OpenBIS as Virual Research Environment: An ELN-LIMS Open-Source Database Tool as a Framework within the CRC 1411 Design of Particulate Products*. Data Sci J. (2023)