



Handlungsempfehlung zum Umgang mit Forschungsdaten

Fachforum Materialwissenschaft und Werkstofftechnik

Ansprechpartnerin in der DFG-Geschäftsstelle:

Cosima Schuster
Ingenieurwissenschaften 1
Materialwissenschaften und Werkstofftechnik
Tel.: 0228/885-2271
E-Mail: cosima.schuster@dfg.de

Stand: 02.02.2023

Vorbemerkung

Die vorliegende Handreichung des Fachforums Materialwissenschaft und Werkstofftechnik ergänzt die fachspezifische Empfehlung aus 2018 durch aktualisierte Empfehlungen zum Forschungsdatenmanagement. Sie soll zur Orientierung für Antragsstellende und Gutachtende dienen. Grundlage sind die Leitlinien der DFG zum Umgang mit Forschungsdaten ([Leitlinien zum Umgang mit Forschungsdaten \(dfg.de\)](#)) und der Kodex „Leitlinien zur Sicherung guter Wissenschaftlichen Praxis“ ([Home | Wissenschaftliche Integrität \(wissenschaftliche-integritaet.de\)](#)).

Forschungsdaten werden im Bereich Materialwissenschaft und Werkstofftechnik hauptsächlich als experimentelle Messdaten, als Daten aus der Bildgebung/Mikroskopie sowie als Daten aus Simulationen und Modellierungen erzeugt. Daten sind konsequenter Weise somit zudem auch in Programmcodes und Realproben enthalten. In der Zusammenführung und Sicherung dieser verschiedenen Datensätze liegt eine Herausforderung für das Fachgebiet. Dies gilt nicht zuletzt, da die Verzahnung von Experiment und Computersimulation, aber auch Hochdurchsatzmethoden eine zunehmend wichtige Rolle spielen. Da die erzeugten Daten sowohl durch datenintensive Charakterisierungsmethoden als auch in den Simulationen in Umfang und Größe stetig zunehmen, ist ein sorgfältig geplantes, aber auch ressourcenschonendes Forschungsdatenmanagement unabdingbar. Mit den entstehenden Forschungsdaten verantwortungsvoll umzugehen, ist einer der elementaren Bestandteile von Forschungsvorhaben.

Datenmanagement bedeutet allerdings nicht, alles zu speichern. Datenspeicherung ist teuer, kostet Energie und erzeugt CO₂. Insbesondere ist zu überlegen und zu begründen, welche Forschungsdaten langfristig archiviert und bereitgestellt werden sollen, um sicherzustellen, dass die Forschungsergebnisse nachvollziehbar sind.

Aus Sicht des Fachforums ist aufgrund zahlreicher Aktivitäten im Bereich des Forschungsdatenmanagements, eine Kommentierung der Vorgaben als Handreichung für die Antragstellenden notwendig. Neben Projekten mit nur einer antragsstellenden Einrichtung können auch Forschungsverbände in Normal- und koordinierten Verfahren durch ein abgestimmtes, gemeinsames Forschungsdatenmanagement qualitativ hochwertige und interoperable Datenräume erzeugen, welche die Vernetzung der Teilprojekte unterstützen. Insbesondere für ortsverteilte Verbände ergibt sich daraus ein hoher Nutzen. Für die nachhaltige Umsetzung des Forschungsdatenmanagements ist es wichtig, Wissenschaftlerinnen und Wissenschaftler in frühen Karrierephasen für ein sorgfältiges Datenmanagement zu sensibilisieren und geeignete Werkzeuge an die Hand zu geben.

Die vorliegende Handreichung verdeutlicht dabei aus Sicht des Fachforums wesentliche Mindestanforderungen an dem Umgang mit Forschungsdaten, um zu sensibilisieren, einen Rahmen für dem Umgang mit Forschungsdaten aufzuzeichnen.

Angaben zum Umgang mit Forschungsdaten in Anträgen an die DFG

Im Vorfeld des Antrags / der Arbeiten sollten Überlegungen zum Datenmanagement angestellt werden, die in einem Datenmanagementplan (DMP) zusammengefasst werden können. Die Überlegungen sollten zumindest schriftlich festgehalten, umgesetzt und überprüft werden. Dabei sollten zu den folgenden Punkten konkrete und verbindliche Angaben gemacht werden, die jeweils die wesentlichen Informationen wiedergeben sollten (weitere Beschreibungen finden Sie im Fragenkatalog ([forschungsdaten_checkliste_de.pdf \(dfg.de\)](https://www.dfg.de/foerderung/foerderungsberechtigten/faq_forschungsdaten_checkliste_de.pdf)):

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

1. Datenbeschreibung

1.1 Grundlegende Erläuterungen und Definitionen:

Der Begriff der Forschungsdaten umfasst alle Daten, die aus einem Forschungsprozess als Teil wissenschaftlicher Arbeit hervorgehen, gesammelt und weiterverarbeitet werden. Als Grundlage eines verantwortlichen Datenmanagements ist es unabdingbar, die im Projekt erzeugten und verwendeten Daten genau zu beschreiben und zu speichern. Zentrale Begriffe, die im Zusammenhang mit der Datenbeschreibung Einsatz finden, sind Primär-, Sekundär- und Metadaten. Primärdaten (Primäre Forschungsdaten) sind Rohdaten oder Urdaten, die unmittelbar aus einem Experiment oder einer Simulation hervorgehen oder in Form von Programm-Codes generiert werden. Sekundärdaten (Sekundäre Forschungsdaten) entstehen durch weitere Verarbeitung aus Primärdaten. Metadaten sind Daten, die es ermöglichen Forschungsdaten zu beschreiben, zu verwalten und zu klassifizieren.

1.2 Kernanforderungen:

Im Rahmen der Projektplanung muss dargelegt werden, welche Datentypen in welcher Menge erzeugt und nach welchen Standards die unterschiedlichen Datentypen gespeichert werden. Dies impliziert als Mindestanforderung, dass zu den im Projekt entstehenden oder verwendeten Daten Art und Format im Antrag beschrieben bzw. der Umfang abgeschätzt werden müssen.

1.3 Zusätzliche Erläuterungen:

Ergänzend können Angaben zur Wiederverwendung existierender Daten, zur Weiterverarbeitung und Aufbereitung (Auswertung) im Projektverlauf generierter Daten die Erläuterungen zur Datenbeschreibung sinnvoll ergänzen.

2. Dokumentation und Datenqualität

2.1 Grundlegende Erläuterungen und Definitionen:

Die Dokumentation von Daten umfasst alle Schritte, die notwendig sind, um Informationen, die aus Experimenten oder Simulationen hervorgehen oder in Programmcodes enthalten sind, zur weiteren Verwendung dauerhaft nutzbar, auffindbar und nachvollziehbar zu machen. Die Verantwortung für die Dokumentation der Daten und die Sicherstellung der Datenqualität liegt sowohl bei den Antragsstellenden (Projektleitenden) als auch bei den beteiligten Mitarbeitenden. Für Projektmitarbeitende, die die Forschungsdaten tagtäglich erheben, ist eine umfangreiche Einweisung in das Forschungsdatenmanagement wesentlich, um die Datenqualität sicherzustellen. Die Datenqualität ist dadurch gekennzeichnet, dass die Daten den an sie gestellten wissenschaftlichen Anforderungen hinsichtlich der Korrektheit der Erfassung und des Inhalts genügen. Welche Form der Dokumentation sinnvoll ist, ist in Abhängigkeit von den Messdaten, der sonstigen Infrastruktur sowie der Einrichtungen der Antragsstellenden festzulegen.

2.2 Kernanforderungen:

Dokumentation

Nach Einschätzung der Antragsstellenden sind wesentliche Schritte der Dokumentation von Daten als Grundlage einer sorgfältigen Projektplanung anzugeben. Dies gilt für Daten in Form von Simulationen oder Programmcode sowie in Form experimenteller Ergebnisse einschließlich der zugehörigen Proben. Für experimentelle Arbeiten impliziert die nachvollziehbare Dokumentation der Daten Erläuterungen zu standardisierten Messprotokollen, die in Abhängigkeit von den jeweiligen Anforderungen der Antragsstellenden entwickelt wurden. Über diese muss der chronologische Messablauf für eine Probe, die Gesamtzahl der Messungen sowie die ggf. vorgenommene Nichtberücksichtigung von Messdaten, der Grund für die Nichtberücksichtigung sowie der Ort der Archivierung der Proben nachvollziehbar sein. Primärdaten müssen auf den Mess-, Prüf- oder Fertigungsrechnern abgespeichert und anschließend mit den zugehörigen Metadaten auf Auswertungsrechner überführt werden. Händisch oder analog erfasste Messdaten sind zu digitalisieren und mit ihren Metadaten zu dokumentieren. Die verwendeten Mess- oder Analysegeräte, das verwendete Probenmaterial und das Messdatum müssen als Mindestanforderungen mit der jeweiligen Probe dokumentiert werden. Letztere muss auffindbar archiviert werden.

Die finale Speicherung der Daten sollte auf einem anderen System mit Sicherungskonzept erfolgen und die Möglichkeit zur Langzeitarchivierung der Daten bieten.

Datenqualität

Die Datenqualität ist über die Kriterien der statistischen Signifikanz und Konsistenz der Daten zu bewerten und anhand dieser vor dem Hintergrund des projektspezifischen Arbeitsprogramms zu diskutieren.

2.3 Zusätzliche Erläuterungen:

Dokumentation

Digitale Laborjournale, insbesondere vorab schematisierte, können ein sinnvolles Instrument zur Datendokumentation sein und Vorteile für eine geordnete Ablage von Daten, ihre Digitali-

sierung, ihre Recherchierbarkeit und die Einbindung unterschiedlicher Messdaten bieten. Es ist sinnvoll, digitale Daten möglichst in strukturierter Form (in Key-Value-Paaren) zu speichern, um eine spätere Überführung in maschinenlesbare Formate (z.B. JSON) zu erleichtern.

Sofern möglich, ist es ratsam die Datensätze in nicht-proprietäre Formate umzuwandeln und zu versionieren. Weitere Kriterien für die Langzeitarchivierung werden beispielsweise durch das deutsche Kompetenznetzwerk für die Langzeitarchivierung und Verfügbarkeit digitaler Ressourcen (nestor, www.langzeitarchivierung.de) benannt und schließen die Verwendung verbreiteter Dateiformate, weniger unterschiedlicher und abwärtskompatibler Formate sowie möglichst unkomprimierter Daten ein.

Für die Dokumentation der Daten sowie ihre Zuordnung und spätere Nutzung ist ihre eindeutige Identifikation sehr hilfreich. Dies ist beispielsweise im Sinne eines unveränderlichen Identifikators (engl. Persistent Identifier, PID) für Proben und Messungen realisierbar. Zu einer transparenten Dokumentation der Daten können eine Konvention zur Datei-Benennung sowie projekt- oder publikationsbezogene Ordnerstrukturen für die Ablage beitragen. Für die Metadaten können Metadatenschema oder einfache Readme-Dateien (vgl. [Guidelines for creating readme style metadata \(open.ac.uk\)](https://open.ac.uk)) verwendet werden, die es ermöglichen Daten zu jeder Zeit zu finden und einzuordnen. Eine systematische Übersicht über das Thema Metadatenschemata gibt unter anderem Data Cite Commons (<https://schema.datacite.org/>). Die Auffindbarkeit und Nachnutzbarkeit der Metadaten sollte analog denen der Primär- und Sekundärdaten sichergestellt werden.

Datenqualität

Zusätzlich zu den Mindestanforderungen können konkrete Angaben zur Qualitätskontrolle der Forschungsdaten sowie etwaige Maßnahmen zur Gewährleistung einer hohen Datenqualität Teil der Ausführungen zum Umgang mit Forschungsdaten sein.

Des Weiteren können Konzepte diskutiert werden, über die vermieden wird, dass die Datenqualität durch Faktoren wie die Datenaufzeichnung oder -eingabe und die Speicherung beeinflusst werden.

3. Speicherung und technische Sicherung während des Projektverlaufs

3.1 Grundlegende Erläuterungen und Definitionen:

Forschungsdaten, für die die langfristige Aufbewahrung durch die Projektleitung als notwendig erachtet wird, sind in eine Speicherumgebung zu überführen. Für die jeweilige Speicherumgebung ist die langfristige Zugänglichkeit dabei wesentlich.

Die Speicherung und Archivierung der Daten umfasst alle Maßnahmen die notwendig sind, um Forschungsdaten für einen angemessenen Zeitraum von in der Regel 10 Jahren aufzubewahren. Details zur Datenarchivierung sind in den Regeln der Guten Wissenschaftlichen Praxis (Leitlinie 17, Archivierung) benannt.

3.2 Kernanforderungen:

Die Antragsstellenden haben grundsätzlich sicherzustellen, dass an den jeweiligen Einrichtungen eine redundante Datenspeicherung umgesetzt wird. Im Antrag muss anhand der wesentlichen Schritte erläutert werden, wie die Speicherung und Archivierung erfolgen. Zentrale Elemente dieser Beschreibung sind die Art der Sicherung, der Schreibschutz, der Schutz der Daten vor unkontrolliertem Zugriff sowie das Ordnungssystem. Zudem ist darzustellen, wo die Daten und ggf. Sicherungskopien (Daten-Back-Ups) gespeichert und archiviert werden. Als Speicherort müssen Fileserver, SharePoint-Lösungen oder vergleichbare Lösungen herangezogen werden. insbesondere bei Gemeinschaftsprojekten oder koordinierten Programmen müssen die Zugriffs- und Nutzungsverwaltung zwischen den Standorten dargestellt werden.

3.3 Zusätzliche Erläuterungen:

Bestehende Vorgaben und Einrichtungen (z.B. Virtual Research Environments) der jeweiligen Forschungseinrichtung können erwähnt werden.

4. Rechtliche Verpflichtungen und Rahmenbedingungen

4.1 Grundlegende Erläuterungen und Definitionen:

In der Nachnutzung von fremden Forschungsdaten oder für den urheberrechtlichem Schutz der eigenen Daten sind die rechtlichen Rahmenbedingungen zu beachten.

4.2 Kernanforderungen:

Rechtliche Verpflichtungen und Rahmenbedingungen (Nutzungs- und Urheberrecht sowie Recht auf fremdes geistiges Eigentum und Patentrecht) müssen bereits im Vorfeld des Vorhabens überprüft und im Antrag kurz benannt werden. Die rechtlichen Verpflichtungen beziehen sich übergreifend auf die Aufnahme, Dokumentation, Speicherung, Archivierung und Nachnutzung der erhobenen oder verwendeten Daten. Dies gilt insbesondere auch in koordinierten Programmen, in denen Daten (Codes, Experimentaldaten o.ä.) weiterverarbeitet werden. Sofern das Projekt auf bereits existierenden Daten beruht, muss bedacht werden, ob die Einwilligung zur Nutzung durch den Urheber der Daten notwendig ist und gegeben wurde.

4.3 Zusätzliche Erläuterungen:

Als Ergänzung zu den Mindestanforderungen können wissenschaftliche Kodizes, fachliche Normen sowie rechtliche Besonderheiten erörtert werden. Zudem können Auswirkungen oder Einschränkungen auf die spätere Veröffentlichung erläutert werden, die aus rechtlichen Verpflichtungen und Rahmenbedingungen resultieren.

5. Datenaustausch und dauerhafte Zugänglichkeit der Daten

5.1 Grundlegende Erläuterungen und Definitionen:

Nicht alle Daten verfügen über das gleiche Potenzial für Nachnutzung. Die Kriterien für die Freigabe zur Nachnutzung sind projektspezifisch. Nur durch eine systematische und kriterienbasierte Auswahl der Forschungsdaten, die zur Verfügung gestellt werden, ist der Nutzen durch die weitere Verwendung der Daten für Dritte gewährleistet. Dies impliziert notwendigerweise auch die dauerhafte Archivierung und Zugänglichkeit der Daten unter Einsatz geeigneter hard- oder softwaretechnischer Infrastruktur sowie die klare Kommunikation der Wege über die der Datenaustausch sichergestellt wird. Aufgrund der schnell fortschreitenden technischen Entwicklungen ergeben sich in immer kürzer werdenden Intervallen Anpassungen, die im Blick zu behalten sind. Die dauerhafte Zugänglichkeit der Daten impliziert dabei, dass die Daten ohne zeitliche Limitierung einem spezifischen Adressatenkreis über gängige Suchmaschinen zugänglich sein müssen. Dieser Adressatenkreis ist dadurch charakterisiert, dass die inbegriffenen Personen und Einrichtungen dem in- oder ausländischen Forschungssektor zugeordnet werden können.

5.2 Kernanforderungen:

Die Systematik und Grundkriterien des Datenaustausches sind darzustellen. In den Ausführungen zum Umgang mit Forschungsdaten ist darzustellen, welche Daten als extern archivierungswürdig und als nachnutzbar angesehen werden. Es ist ebenfalls zu erörtern, wenn keine Daten für die allgemeine Nachnutzung freigegeben werden können oder die Freigabe individuell nach definierten Kriterien erfolgt. Falls Daten einer Karenzzeit oder Sperrfrist unterliegen, muss dies unter Angabe der Dauer der Zeit (Frist) dargestellt werden.

5.3 Zusätzliche Erläuterungen:

Alle relevanten und reproduzierbaren Daten, die Publikationen zugrunde liegen, sollten neben den Daten, die bereits über Publikationen aufbereitet zugänglich sind, öffentlich zugänglich archiviert bzw. publiziert werden. Dies kann ggf. als zusätzliche Information in Form supplementärer Datenpublikationen bei den Artikeln oder über zertifizierte, fachspezifische oder fachübergreifende Repositorien oder Datenzentren (z.B.: zenodo oder figshare) erfolgen. Auch nicht mit Publikationen verknüpfte Daten können dort veröffentlicht werden. Die Auswahl dieser Daten sollte mit Augenmaß eigenverantwortlich und bewusst von den Projektleitenden erfolgen und begründet sein. Kriterien für die Auswahl der Forschungsdaten für die Nachnutzung und die Löschung von Daten sollten in diesem Zusammenhang definiert werden.

Eine Übersicht über verfügbare Repositorien wird unter anderem im Research Infrastructure Portal der DFG (vgl. <https://risources.dfg.de>), in der Registry of Research data Repositories (vgl. <https://www.re3data.org>) oder durch Data Cite Commons (vgl. <https://commons.datacite.org/repositories>) geführt. Durch die Vergabe einer DOI kann die Zitierfähigkeit sichergestellt werden.

Sollten im Projekt Lösungen für eine praktikable Umwandlung von Datenformaten entwickelt werden, ist es wünschenswert, diese Lösungen zur Nutzung der Allgemeinheit bereitzustellen, um eine generelle Standardisierung von Datenformaten weiter voranzutreiben.

6. Verantwortlichkeit und Ressourcen

6.1 Grundlegende Erläuterungen:

Im Zusammenhang mit dem Umgang mit den Forschungsdaten obliegen allen Beteiligten eines Projekts grundsätzlich unterschiedliche Rollen und Verantwortlichkeiten. Die adäquate Umsetzung der notwendigen Schritte innerhalb des Umgangs mit Forschungsdaten erfordert oftmals monetäre, zeitliche oder zusätzliche infrastrukturelle Ressourcen.

6.2 Kernanforderungen:

Verantwortlichkeit

Im Antrag muss klar dargelegt werden, wer im Projektverlauf für welche einzelnen Schritte des Forschungsdatenmanagements die Hauptzuständigkeit besitzt. Die Verantwortung für ein sorgfältiges Datenmanagement sowie eine abschließende Qualitätskontrolle liegt grundsätzlich bei den Projektleitenden.

Ressourcen

Die langfristige Finanzierung der Infrastruktur zum Umgang mit Forschungsdaten ist sicherzustellen und im Antrag zu dem jeweiligen Forschungsvorhaben darzulegen.

6.3 Zusätzliche Erläuterungen:

Der zeitliche und ressourcenspezifische Aufwand kann auch im Hinblick auf die Beantragung projektspezifischer Mittel berücksichtigt werden. Es sollte festgelegt werden, wer nach Ende der Laufzeit des Projekts für das Kuratieren der Daten verantwortlich ist.