

# Handlungsempfehlung zum Umgang mit Forschungsdaten des Fachforum Materialwissenschaft und Werkstofftechnik

Stand: 10.10.2025

# Ansprechpartnerin in der DFG-Geschäftsstelle:

Cosima Schuster
Ingenieurwissenschaften 1
Materialwissenschaften und Werkstofftechnik

Tel.: 0228/885-2271

E-Mail: <a href="mailto:cosima.schuster@dfg.de">cosima.schuster@dfg.de</a>



## A. Einleitung

## a. Vorbemerkung

Die vorliegende Handreichung des Fachforums Materialwissenschaft und Werkstofftechnik soll zur Orientierung für Antragsstellende und Gutachtende dienen. Mit den entstehenden Forschungsdaten verantwortungsvoll umzugehen, ist ein elementarer Bestandteil von Forschungsvorhaben. In der Qualitätssicherung und Archivierung der unterschiedlichen Datensätze und dem Zusammenführen der teils sehr verschiedenartigen Datenarten liegen die besonderen Herausforderungen für das Fachgebiet. Dies gilt nicht zuletzt, da die Verzahnung von Experiment und Computersimulation, aber auch Hochdurchsatzmethoden und Ansätze des inversen Werkstoffdesigns sowie der Einsatz von Methoden der künstlichen Intelligenz wie das Maschinelle Lernen eine zunehmend wichtigere 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.

Grundlage sind die Leitlinien der DFG zum Umgang mit Forschungsdaten (<u>Leitlinien zum Umgang mit Forschungsdaten (dfg.de)</u>) und der Kodex "Leitlinien zur Sicherung guter Wissenschaftlichen Praxis" (<u>Home | Wissenschaftliche Integrität (wissenschaftliche-integritaet.de)</u>).

# b. Abkürzungen

Im Text verwendete Abkürzungen sind die Definition von

- "FAIR"en Daten, also auffindbaren, zugänglichen, interoperablen und wiederverwendbaren Daten aus "findable, accessible, interoperable, reusable" sowie
- **NFDI** für die Nationale Forschungsdateninfrastruktur (<u>www.nfdi.de</u>).

## c. Definition von Forschungsdaten

Forschungsdaten entstehen im Bereich Materialwissenschaft und Werkstofftechnik

- als experimentelle Messdaten,
- als Daten aus der Bildgebung/Mikroskopie
- Daten aus Simulationen und Modellierungen
- Programmcodes und Realproben sind ebenfalls Bestandteil von Forschungsdaten.

## d. Einschränkungen

Es ist dabei klarzustellen, dass zwei häufig genannte Punkte **nicht** Teil eines überzeugenden Datenmanagements sind.

• Zur Auswahl der zu archivierenden Daten: Zum ersten bedeutet Datenmanagement nicht, alle jemals erzeugten Daten und Proben zu speichern und zu archivieren. Datenspeicherung ist teuer, kostet Energie und erzeugt CO<sub>2</sub>. Vielmehr ist mit Blick auf die Kodex-Leitlinien zu überlegen und zu begründen, welche Forschungsdaten zu archivieren sind, um die Nachvollziehbarkeit der Forschungsergebnisse zu sichern. Beispielsweise ist bei der Lagerung physischer Proben zu bedenken, dass sich diese bei Lagerung verändern können und dann diesem Zweck nicht mehr dienen. Daher ist das Protokoll der Probenherstellung das in dieser Hinsicht nützlichere Forschungsdatum, das archiviert werden sollte.



• Zum Unterschied "FAIR" und "OPEN": Ein zweites häufiges Missverständnis ist die Vorstellung, dass "FAIR"e Daten uneingeschränkt öffentlich zugängliche Daten sind. FAIR und OPEN sind keine deckungsgleichen Begriffe. Existenz und Beschreibung der Daten, insbesondere derer, die zur Nachnutzung freigegeben werden sollen, können bereits durch das Veröffentlichen der zugehörigen Metadaten zugänglich gemacht werden. Die dazugehörigen Daten sind damit ebenfalls "findable" und über die Urheberangabe "accessible", wobei die Gewährung auf den Zugang vom Datenurheber bestimmt wird.

# e. Erläuterungen zu Metadaten

Metadaten umfassen dabei:

- Bibliographische Metadaten, beispielsweise die Angaben in Literaturdatenbanken zu
  - Urheber/Autor der Daten
  - Publikationsort, Publikationsdatum
     Hierzu wird häufig JavaScript Object Notation (<a href="https://www.json.org/json-en.html">https://www.json.org/json-en.html</a>)
     verwendet, XML-Standards finden sich bei DataCite <a href="http://schema.datacite.org/">https://schema.datacite.org/</a>
     und DublinCore <a href="https://www.dublincore.org/specifications/dublin-core/dcmi-terms/">https://www.dublincore.org/specifications/dublin-core/dcmi-terms/</a>.
- Administrative Metadaten, die die interne Verwaltung der Daten beschreiben. Als Grundlage eines verantwortlichen wissenschaftlichen Arbeitens ist es unabdingbar, die im Projekt erzeugten und verwendeten Daten nachvollziehbar zu benennen und auffindbar zu speichern. Zu einer transparenten Dokumentation der Daten trägt eine Konvention zur Datei-Benennung sowie projekt- oder publikationsbezogene Ordnerstrukturen bei. In der digitalen Aufbereitung (und damit auch Benennung und Beschreibung) von Datensätzen ist eine Orientierung an fachspezifischen Taxonomien, Semantiken oder Ontologien hilfreich. In den zugehörigen Metadaten wird abgebildet:
  - Dateiname, Dateiformat
  - Speicherort, ULR
  - Zugangsberechtigungen

# Hierzu wird verwiesen auf:

- o PROV-O, https://www.w3.org/TR/prov-o/
- DCAT, <u>Data Catalog Vocabulary (DCAT) Version 2</u>
- Fachspezifische Metadaten, wie
  - Chemische Zusammensetzung
  - Herstellungsverfahren
  - Prozessierungsschritte
  - Charakterisierungsmethode
  - Angaben zum Gerät
  - Angaben zur Software
  - Angaben zur Rechnerarchitektur
  - o ....



Welche dieser Metadaten in welcher Granularität im projektspezifischen Datenmanagement zu berücksichtigen sind, ist anhand der Forschungsfragen und verwendeten Methoden und Werkzeuge festzulegen.



## B. Handreichung zum Umgang mit Forschungsdaten in Anträgen an die DFG

Das Fachforum Materialwissenschaft und Werkstofftechnik sieht es weiterhin als wichtig an, die Vorgaben in den Anträgen als Handreichung für die Antragstellenden zu kommentieren. Die vorliegende Handreichung verdeutlicht dabei aus Sicht des Fachforums wesentliche Mindestanforderungen an den Umgang mit Forschungsdaten und zeigt an Beispielen auf, wie weitergehende Standards zu erreichen sind. Im Vorfeld der Antragsstellung / der Arbeiten Überlegungen zum Datenmanagement angestellt werden, die in einem Datenmanagementplan (DMP) zusammengefasst und schriftlich festgehalten werden können. Insbesondere Forschungsverbünde können durch ein abgestimmtes, gemeinsames Forschungsdatenmanagement interoperable Datenräume erzeugen, welche die enge Vernetzung der Teilprojekte und damit hochwertige wissenschaftliche Ergebnisse ermöglichen. Gerade für ortsverteilte Verbünde oder interdisziplinäre Projekte mit heterogenen Datenarten ergibt sich daraus eine große Herausforderung, aber auch ein hoher Nutzen. Für die nachhaltige Umsetzung des Forschungsdatenmanagements ist es wichtig, Wissenschaftler\*innen in frühen Karrierephasen für ein sorgfältiges Datenmanagement zu sensibilisieren und geeignete Werkzeuge an die Hand zu geben. Dabei sollten zu den folgenden Punkten der Checkliste konkrete und verbindliche Angaben gemacht werden.

- 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

Erläuterungen und weitere Beschreibungen zu diesen Punkten sind auf der DFG-Webseite im Fragenkatalog zu finden unter <u>forschungsdaten\_checkliste\_de.pdf</u> (dfg.de)

Auf der Homepage des NFDI MatWerk Konsortiums, <a href="https://nfdi-matwerk.de/solutions">https://nfdi-matwerk.de/solutions</a> oder <a href="https://nfdi-matwerk.de/solutions/via-data-life-cycle">https://nfdi-matwerk.de/solutions/via-data-life-cycle</a> werden Angebote und Werkzeuge zum Datenmanagement angegeben, die ebenfalls der Strukturierung dieser Checkliste folgen.

# 1. Datenbeschreibung

# 1.1 Grundlegende Erläuterungen und Definitionen:

Der Begriff der Forschungsdaten umfasst alle Daten, die in einem Forschungsprozess entstehen, gesammelt, weiterverarbeitet oder verwendet werden. Daten können analog, digital oder als physische Proben vorliegen. 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 beschreiben und klassifizieren Forschungsdaten.

## 1.2 Kernanforderungen:



Im Rahmen der Projektplanung sollte abgeschätzt werden, welche Datentypen in welcher Größenordnung erzeugt und nach welchen Standards die unterschiedlichen Datentypen gespeichert werden, und ob sich aus der Datenmenge oder -vielfalt besondere Herausforderungen für das Datenmanagement ergeben. Bei sehr großen Datenmengen ist gegebenenfalls eine Filterung der überhaupt abspeicherbaren Datensätze notwendig. Dies sollte erläutert werden. Eine bloße Auflistung unterschiedlicher Dateiformate ist nicht zielführend und daher weder ausreichend noch notwendig.

# 2. Dokumentation und Datenqualität

# 2.1 Grundlegende Erläuterungen:

#### Dokumentation:

Die Dokumentation von Daten umfasst alle Schritte, die notwendig sind, um Informationen, die aus Experimenten oder Simulationen hervorgehen oder in Programmcodes enthalten sind, langfristig auffindbar und nachvollziehbar zu machen.

## Datenqualität:

Es sollte erläutert werden, wie die Datenqualität für das jeweilige Projekt definiert ist. Die Daten müssen den an sie gestellten wissenschaftlichen Anforderungen hinsichtlich der Korrektheit der Erfassung und des Inhalts genügen. Konkrete Angaben zur Qualitätskontrolle der Forschungsdaten sowie die Maßnahmen zur Gewährleistung einer hohen Datenqualität sind Teil der Ausführungen zum Umgang mit Forschungsdaten. 

Diese Angaben sind oftmals bereits in der Versuchsplanung und im konkreten Arbeitsprogramm erläutert, so dass darauf verwiesen werden kann.

# 2.2 Kernanforderungen:

## **Dokumentation**

Für experimentelle Arbeiten impliziert die nachvollziehbare Dokumentation der Daten Erläuterungen zu den für das Projekt entwickelten und dann standardisiert genutzten Messprotokollen. Messprotokolle enthalten den chronologischen Messablauf für eine Probe, die Gesamtzahl der Messungen sowie die ggf. vorgenommene Nichtberücksichtigung von Messdaten, der Grund für die Nichtberücksichtigung sowie den Ort der Archivierung der Proben. Händisch oder analog erfasste Messdaten sind zu digitalisieren, damit Sekundär- und Metadaten maschinenlesbar vorliegen. Für numerische Arbeiten sind die ausgeführten Konvergenz-Studien zu dokumentieren.

Die verwendeten Mess- oder Analysegeräte, das verwendete Probenmaterial und das Messdatum sind mit der jeweiligen Probe als fachspezifische Metadaten zu dokumentieren. Beispiele für entsprechende Metadaten bei numerisch-simulativen Arbeiten sind die Parameter in den Eingabe-Dateien, die genutzte Software (mit Versionsangabe) oder die genutzten Rechner-Plattformen.

## Datengualität



Die Datenqualität ist über die Kriterien der konsistenten Erfassung von Daten und der statistischen Signifikanz zu bewerten und anhand dieser vor dem Hintergrund des projektspezifischen Arbeitsprogramms zu diskutieren. Gerade im Bereich grundlegender oder explorativer Forschung kann es unumgänglich sein, bei stichprobenartigen Experimenten Einschränkungen hinsichtlich der statistischen Absicherung zu akzeptieren. Die Verknüpfung mit Metadaten trägt zur Auffindbarkeit und Überprüfbarkeit von Daten bei und sichert Transparenz und trägt somit ebenfalls zur Datenqualität bei.

#### 2.3 Hinweise:

#### Dokumentation

Digitale Laborjournale und elektronische Laborbücher sind sinnvolle Instrumente zur Datendokumentation und bieten deutliche Vorteile für eine geordnete Ablage, Digitalisierung und Recherchierbarkeit von Daten sowie die Einbindung unterschiedlicher Messdaten. Ihre Verwendung wird daher empfohlen.

Sofern möglich, sind die Datensätze in maschinenlesbare, 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, <a href="https://www.langzeitarchivierung.de">www.langzeitarchivierung.de</a>) 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 hilfreich. Dies ist beispielsweise im Sinne eines unveränderlichen Identifikators (engl. Persistent IDentifier, PID) für Proben und Messdaten realisierbar, der spätestens bei Projektabschluss zugewiesen werden sollte.

#### Datengualität

Unter Qualitätssicherung ist die Gesamtheit der Maßnahmen zu verstehen, die sicherstellen sollen, dass die Daten in Aspekten wie Genauigkeit, Vollständigkeit, Konsistenz, Zuverlässigkeit und Aktualität den Anforderungen des Projekts entsprechen. Insbesondere bei Verwendung von Methoden des maschinellen Lernens ist im Antrag (bereits im Arbeitsprogramm) darzulegen, wie die Datenqualität sichergestellt wird und nach welchen Kriterien das Training als erfolgreich gewertet wird. Die ML-Modelle können sehr komplex sein, so das kleine Ungenauigkeiten in der Eingabe zu drastischen Änderungen der Ausgaben führen können. Ein Testen dieser Systeme in Abhängigkeit der nahezu unbegrenzten Trainingsmöglichkeiten ist daher schwierig und sollte ausführlich beschrieben werden.

# 3. Speicherung und technische Sicherung während des Projektverlaufs

# 3.1 Grundlegende Erläuterungen und Definitionen:

Für die Forschungsdaten, die im Laufe der Projektbearbeitung erzeugt werden, ist eine geeignete Ablage sicherzustellen. Eine Konvention zur Datei-Benennung sowie projekt- oder publikationsbezogene Ordnerstrukturen sind Teil der Datenorganisation.



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 sicherzustellen, dass an den jeweiligen Einrichtungen eine redundante Datenspeicherung umgesetzt wird. Zentrale Elemente sind die Art der Sicherung, der Schreibschutz, der Schutz der Daten vor unkontrolliertem Zugriff sowie das Ordnungssystem. Es ist darzustellen, wie die Daten und ggf. Sicherungskopien (Daten-Back-Ups) gespeichert und archiviert 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 (unter anderem Nutzungs- und Urheberrecht sowie Patentrecht) sollten bereits im Vorfeld des Vorhabens überprüft werden. Die rechtlichen Verpflichtungen beziehen sich übergreifend auf die Aufnahme, Dokumentation, Speicherung, Archivierung und Nachnutzung der erhobenen oder verwendeten Daten. Sofern das Projekt auf bereits existierenden Daten beruht, externen wie auch bei koordinierten Projekten, muss bedacht werden, ob die Einwilligung zur Nutzung durch den Urheber der Daten notwendig ist und gegeben wurde. Im Kontext der Biomaterialien sind die Vorgaben zur Nutzung von Patientendaten zu berücksichtigen.

## 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:

Datenaustausch

Bei Gemeinschaftsprojekten oder koordinierten Programmen müssen die Zugriffs- und Nutzungsverwaltung zwischen den Projektpartnern und zwischen den Standorten definiert



sein. Aber auch innerhalb einer Arbeitsgruppe sind die Zugriffsrechte und -Möglichkeiten zu klären. Dies beinhaltet die klare Kommunikation der Wege über die der Datenaustausch erfolgt. Insbesondere in internationalen Kooperationen aber auch allgemein ist der Zugriff auf und Austausch der Daten gegen unbefugten Zugriff abzusichern. Die entsprechenden Infrastrukturen (z.B. Virtual Research Environments, Definition der administrativen Metadaten) der jeweiligen Forschungseinrichtung sind zu benennen.

# Dauerhafte Zugänglichkeit

Die Daten, für die die langfristige Aufbewahrung durch die Projektleitung als notwendig erachtet wird, sei es für die Archivierung zur Nachvollziehbarkeit des Forschungsprozesses oder für die Bereitstellung zur Nachnutzung, sind in eine geeignete Speicherumgebung zu überführen. Die finale Speicherung der Daten sollte auf einem anderen System erfolgen als dem, das zur Speicherung während des Forschungsprojekts genutzt wird.

Zur Nachvollziehbarkeit der Forschungsergebnisse sind die Daten, die Abschlussarbeiten, Promotionen und Publikationen zugrunde liegen, öffentlich zugänglich zu archivieren bzw. mit den Publikationen zu veröffentlichen.

Da die erzeugten Daten projektspezifisch ein unterschiedliches Potenzial für eine Nachnutzung aufweisen, sind die Kriterien für die Auswahl zur Freigabe der Daten zur Nachnutzung zu überlegen. Kriterien können Datenqualität, allgemeines Interesse, besondere Parametersätze oder Randbedingungen sein. Eine zur Verfügungstellung impliziert die dauerhafte Archivierung und Zugänglichkeit der Daten unter Einsatz geeigneter hard- oder softwaretechnischer Infrastruktur. Die Daten beziehungsweise der Speicherort der Daten müssen dann ohne zeitliche Limitierung über gängige Suchmaschinen auffindbar sein.

# 5.2 Kernanforderungen:

Datenaustausch

Die technische Umsetzung des Datenaustausches zwischen Projektpartnern ist darzustellen.

## Dauerhafte Zugänglichkeit

Alle relevanten und reproduzierbaren Daten, die Publikationen, insbesondere den Artikeln in wissenschaftlichen Zeitschriften, zugrunde liegen, sollten öffentlich zugänglich archiviert bzw. publiziert werden. Dies kann in der Publikation selbst, oder 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.

Zu weiteren Nachnutzungsmöglichkeiten ist darzustellen, ob und welche Daten über die Nachvollziehbarkeit und Qualitätssicherung hinaus als nachnutzbar angesehen werden und wie diese in besonders aufbereiteter Form zur Verfügung gestellt werden. Vorgesehene Karenzzeiten oder Sperrfristen sind zu benennen.

## 5.3 Beispiele:

Auch nicht mit Journal-Artikel verknüpfte Daten können in Repositorien als eigenständige Publikationen veröffentlicht werden. Die Auswahl dieser Daten sollte mit Augenmaß



eigenverantwortlich und bewusst von den Projektleitenden erfolgen. Kriterien für die Auswahl der Forschungsdaten für die Publikation/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. <a href="https://risources.dfg.de">https://risources.dfg.de</a>), in der Registry of Research Data Repositories (vgl. <a href="https://www.re3data.org">https://www.re3data.org</a>) oder durch Data Cite Commons (vgl. <a href="https://commons.datacite.org/repositories">https://commons.datacite.org/repositories</a>) 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 den Beteiligten eines Projekts unterschiedliche Rollen und Verantwortlichkeiten. 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 adäquate Umsetzung der notwendigen Schritte innerhalb des Umgangs mit Forschungsdaten erfordert oftmals monetäre, zeitliche oder zusätzliche infrastrukturelle Ressourcen. Die dazu nötigen Mittel können projektbezogen beantragt werden, sofern sie die Grundanforderungen überschreiten, die sich aus der Befolgung der Leitlinien des Kodex durch die wissenschaftliche Einrichtung ergeben.

## 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. Es sollte festgelegt werden, wer nach Ende der Laufzeit des Projekts für das Kuratieren der Daten verantwortlich ist.

## Ressourcen

Die verfügbare Infrastruktur zum Umgang mit Forschungsdaten ist knapp darzulegen.

