

UFZ-Regelung | IR-5/18 | 17.06.2024

Regeln für den verantwortungsvollen Umgang mit Forschungsdaten am Helmholtz-Zentrum für Umweltforschung GmbH – UFZ

In Kraft gesetzt durch	
Geschäftsführung	
Prof. Dr. Rolf Altenburger	Dr. Sabine König
in GF-Sitzung Nr. 380 am 17.06.2024	

Dokumentenhistorie

Datum	Version	
13.09.2018	1	Verabschiedung „Grundsätze zum verantwortungsvollen Umgang mit Forschungsdaten am Helmholtz-Zentrum für Umweltforschung GmbH – UFZ“
17.06.2024	2	Ergänzung der Grundsätze durch die Regeln für den verantwortungsvollen Umgang mit Forschungsdaten am Helmholtz-Zentrum für Umweltforschung GmbH – UFZ

Inhalt:

1	Grundsätze zum verantwortungsvollen Umgang mit Forschungsdaten am UFZ	1
2	Geltungsbereich der Regeln für den verantwortungsvollen Umgang mit Forschungsdaten am UFZ.....	2
3	Governance des Forschungsdatenmanagements am UFZ.....	2
4	Verantwortlichkeit für den verantwortungsvollen Umgang mit Forschungsdaten	3
5	Qualifizierung und Weiterentwicklung des Personals	3
6	Wissenschaftliche Anerkennung.....	4
7	Relevanz von Daten.....	5
8	Nutzung von Forschungsdateninfrastrukturen	5
9	Planung des Umgangs mit Forschungsdaten	5
10	Datenzugang und offene Daten	6
11	Qualitätssicherung	7
12	Langzeitverfügbarkeit und Nachnutzung.....	7
13	Rechtliche Aspekte.....	8
14	Einbettung in nationale und internationale Initiativen und Konsortien	8
15	Definitionen.....	9
a)	Datenlebenszyklus.....	9
b)	Datenmanagementplan (DMP)	9
c)	Datenqualität	9
d)	Forschungsdaten	9
e)	Forschungsdatenmanagement (FDM)	10
f)	Forschungsdateninfrastruktur	10
g)	Metadaten.....	10
h)	Metadatenstandard.....	10
i)	Open Data	10
j)	Open Science	11
k)	Persistente Identifikatoren/Identifier (PID).....	11
l)	Repositorium	11
m)	Standards	11
16	Inkrafttreten.....	11
	Anhang 1 – FAIR Prinzipien	12

1 Grundsätze zum verantwortungsvollen Umgang mit Forschungsdaten am UFZ

Das UFZ ist eines der weltweit führenden Forschungszentren im Bereich der integrierten Umweltforschung. Es unterstützt Politik, Wirtschaft und Zivilgesellschaft dabei, die Folgen menschlichen Handelns für die Umwelt besser zu verstehen und Optionen für gesellschaftliche Entscheidungsprozesse zu erarbeiten. Für die Behandlung komplexer Umweltprobleme setzt das UFZ seine umfangreichen Erfahrungen in der integrierten Umweltforschung und seine innovativen wissenschaftlichen Infrastrukturen ein. Es unterstützt verschiedene wissenschaftliche Communities auf nationaler und internationaler Ebene und sieht sich dem gesellschaftlichen Ziel verpflichtet, qualitätsgesicherte Forschungsdaten zum Zustand und zur Veränderung unserer Umwelt verlässlich und nach anerkannten Standards zu erheben, zu verarbeiten und langfristig für die nationale und internationale Gemeinschaft zu erhalten und nutzbar zu machen.

Forschungsdaten werden in allen wissenschaftlichen Disziplinen und daher in allen Departments des UFZ erhoben. Sie reichen beispielsweise von klassischen Labor-, Feld- und Fernerkundungsdaten bis hin zu Forschungsdaten aus Kohortenstudien oder der qualitativen und quantitativen Sozialforschung.

Orientiert an den Empfehlungen für Richtlinien der Helmholtz Gemeinschaft zum Umgang mit Forschungsdaten¹, an den Leitlinien zum Datenmanagement nach den FAIR-Prinzipien der Europäischen Kommission² und den Leitlinien zum Umgang mit Forschungsdaten der DFG³ setzen die Regelungen in diesem Dokument die folgenden Grundsätze in die Praxis um.

1. Das UFZ bekennt sich zu den Prinzipien des offenen Zugangs zu Wissen, Forschungsergebnissen und Technologien. Dabei sind gesetzliche Regelungen, insbesondere der Schutz personenbezogener Daten, ebenso wie die wissenschaftlichen Interessen und vertraglichen Vereinbarungen mit Kooperationspartnern sowie gegebenenfalls Verwertungsinteressen verbunden mit einer zumindest zeitweiligen Vertraulichkeit von Forschungsdaten zu berücksichtigen. Im Einzelfall kann das UFZ dafür angemessene Embargofristen zur exklusiven Erstnutzung benennen. Die Forschenden des UFZ wahren die Grundsätze und Regeln guter wissenschaftlicher Praxis.⁴
2. Das UFZ erkennt die Notwendigkeit an, die erforderlichen infrastrukturellen und organisatorischen Rahmenbedingungen für das Forschungsdatenmanagement (Fachpersonal, Qualifizierung und Infrastrukturen) dauerhaft bereitzustellen. Die nähere Ausgestaltung bedarf einer weiteren Konzeptionierung.

¹ Mitgliederversammlung der Helmholtz-Gemeinschaft (Ed.) (2016) *Die Ressource Information besser nutzbar machen! Positionspapier zum Umgang mit Forschungsdaten in der Helmholtz-Gemeinschaft*, Potsdam: Helmholtz Open Science Office. <https://doi.org/10.2312/os.helmholtz.021>

² European Commission: Directorate-General for Research & Innovation (2016) *H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020*, v3.0. Abrufbar unter https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf [geprüft 2023-09-30]

³ Deutsche Forschungsgemeinschaft - DFG (2019) Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex. <https://doi.org/10.5281/zenodo.6472827>

⁴ Helmholtz-Zentrum für Umweltforschung GmbH - UFZ (2023) *Regeln zur Sicherung guter wissenschaftlicher Praxis im Helmholtz-Zentrum für Umweltforschung GmbH - UFZ*. UFZ-Regulation IR-17/12, Abrufbar unter <https://www.ufz.de/index.php?de=36448> [geprüft 2023-09-30]

3. Als übergeordnetes Prinzip für die nachhaltige Bereitstellung der am UFZ erhobenen Forschungsdaten erkennt das UFZ die FAIR-Prinzipien an⁵ (siehe **Anhang 1**).
4. Das UFZ unterstützt nationale und internationale Erklärungen zum Umgang mit Forschungsdaten.

2 Geltungsbereich der Regeln für den verantwortungsvollen Umgang mit Forschungsdaten am UFZ

- (1) Die Regeln für den verantwortungsvollen Umgang mit Forschungsdaten am Helmholtz-Zentrum für Umweltforschung GmbH – UFZ (im Folgenden „Regeln des FDM“) richten sich an alle am UFZ tätigen Personen (inklusive wissenschaftliche Gäste), welche im Rahmen von Forschungstätigkeiten Daten erhoben haben, erheben und/oder mit Forschungsdaten arbeiten. Sie bilden einen verbindlichen Rahmen für das Forschungsdatenmanagement (FDM) am UFZ.
- (2) Im Rahmen von Forschungskollaborationen gelten die Regeln des FDM, soweit die Beteiligten keine gleich- oder höherwertigeren Vorgaben treffen. Die beteiligten Forschenden des UFZ sind für die Einhaltung dieser Regeln verantwortlich. Werden im Rahmen von Forschungsprojekten Forschungsdaten-Richtlinien entwickelt oder von Drittmittelgebern vorgeschrieben, die über gleich- oder höherwertigere Vorgaben verfügen, haben diese Vorrang vor diesen Regeln.
- (3) Die Regeln des FDM beziehen sich auf digitale Forschungsdaten und digitale Repräsentationen von analogen Daten.
- (4) Die Regeln des FDM wurden unter Berücksichtigung der *Regeln zur Sicherung guter wissenschaftlicher Praxis am UFZ (IR-17/12)* erstellt und ersetzen die *Grundsätze zum verantwortungsvollen Umgang mit Forschungsdaten am Helmholtz-Zentrum für Umweltforschung GmbH – UFZ (IR-5/18)* (im Folgenden „Grundsätze des FDM“).

3 Governance des Forschungsdatenmanagements am UFZ

- (1) 2019 wurde die Einheit Research Data Management (RDM) mit dem Ziel etabliert, den verantwortungsvollen Umgang mit Forschungsdaten am UFZ durch ein professionelles Beratungs-, Qualifizierungs- und Unterstützungsangebot sowie die Weiterentwicklung der bestehenden Dateninfrastrukturen im gesamten Datenlebenszyklus am UFZ zu unterstützen. Hierzu wurden der Einheit die strategische und personelle Verantwortung zur Ausgestaltung eines professionellen Datenmanagements basierend auf existierenden und zukünftigen Anforderungen durch die Geschäftsführung und die Themenbereiche übertragen. Damit wurde der Einheit die Verantwortung für die Umsetzung der Grundsätze des FDM übertragen, was die Konkretisierung von Verantwortlichkeiten und Handlungsfeldern sowie die notwendigen Schritte zur Implementierung umfasst. Die Einheit RDM ist für Fragen des

⁵ Wilkinson et al. (2016): *The FAIR Guiding Principles for scientific data management and stewardship*. Scientific Data, 3, 160018, <https://doi.org/10.1038/sdata.2016.18>

FDM im DataHub des Forschungsbereichs Erde und Umwelt der Helmholtz-Gemeinschaft sowie weiterer nationaler und internationaler Initiativen und Konsortien zuständig.

- (2) Zur Sicherstellung der zielgerichteten Entwicklung einer modernen und nachhaltigen Infrastruktur wurde das RDM Board eingerichtet. Es bildet die Schnittstellen zu strategischen Entwicklungen in den Themenbereichen und deren wissenschaftlichen Fach-Communities, nimmt die Priorisierung von Entwicklungen vor und berät die Geschäftsführung und die Themenbereiche in Bezug auf datenbezogene strategische Aspekte. Hierzu entsendet jeder Themenbereich zwei Vertreter:innen in das RDM Board, die mit den strategischen Zielen des Themenbereiches selbst, der Programmforschung (PoF) und den einhergehenden Aspekten des Datenmanagements vertraut sind.
- (3) Jedes Department im UFZ benennt eine Ansprechperson für Forschungsdatenmanagement, welche durch geeignete Angebote aus dem Bereich RDM weitergebildet wird. Diese Person unterstützt im jeweiligen Department in Fragen des Datenmanagements, dient als Ansprechperson für die RDM Board Vertreter:innen des jeweiligen Themenbereiches und erfasst und kommuniziert Bedarfe im Zusammenhang mit Datenmanagement an die Einheit RDM.

4 Verantwortlichkeit für den verantwortungsvollen Umgang mit Forschungsdaten

- (1) Alle Mitarbeitende des UFZ, die Forschungsdaten erheben und/oder weiterverarbeiten, müssen mit diesen verantwortungsvoll umgehen.
- (2) Die Geschäftsführung, Themenbereichs- und Departmentleitenden schaffen Voraussetzungen, um den nachhaltigen Umgang mit den Forschungsdaten im gesamten Datenlebenszyklus zu ermöglichen.
- (3) Departmentleitende sind verantwortlich, das Personal (inkl. Studierende, Stipendiat:innen) über die Regeln des FDM zu informieren und die Einhaltung der Vorgaben zu prüfen.
- (4) Projektverantwortliche und/oder Departmentleitende sind verantwortlich, beim Wechsel ihrer Mitarbeitenden den Verbleib der Originaldaten am Entstehungsort sicherzustellen, Vorkehrungen zur sachgerechten Weitergabe von Primärdaten und zur Klärung der Zugangsrechte zu treffen.

5 Qualifizierung und Weiterentwicklung des Personals

- (1) Das UFZ stellt seinen Mitarbeitenden ein Beratungs-, Qualifizierungs- und Unterstützungsangebot zu FDM zur Verfügung.⁶ Die Angebote sind an den spezifischen Erfordernissen der Fachdisziplinen und Infrastrukturen ausgerichtet. Dies erfolgt in Zusammenarbeit und Abstimmung mit dem Qualifizierungsangebot der Abteilung Personal

⁶ RDM Guidelines UFZ, <https://rdm.pages.ufz.de/guidelines/>

und Unternehmenskultur (PACE) und den zentrenübergreifenden Angeboten der Helmholtz-Gemeinschaft (z. B. durch HIDA⁷ und HIFIS⁸). Angebote erfolgen u.a. in den Bereichen:

- Grundlegende Aspekte und Praktiken des FDM,
 - Digitalisierung in der Forschung und Umgang mit Forschungsdateninfrastrukturen.
- (2) Mitarbeitenden soll ermöglicht werden, Angebote zur Qualifizierung, Weiterbildung und zum Selbstlernen zu FDM zu nutzen, um ihre Arbeitsweisen, Workflows und Technologien den Prozessen der Digitalisierung anzupassen.
- (3) Studierende, Promovierende und Praktikant:innen sind von ihren Betreuenden am UFZ über den adäquaten Umgang mit Forschungsdaten zu informieren. Fachspezifische Kompetenzen und Standards sind zu vermitteln.

6 Wissenschaftliche Anerkennung

- (1) Das UFZ erkennt die Erzeugung und Verarbeitung von Forschungsdaten sowie deren Bereitstellung nach den FAIR-Prinzipien als eine zentrale und unverzichtbare Leistung im Forschungsprozess an. Qualitätsgesicherte Forschungsdaten zählen zum wissenschaftlichen Output des UFZ und ihre Erzeugung und Veröffentlichung wird entsprechend gewürdigt. Dies erfolgt u.a. durch den Einbezug von Datenpublikationen in das wissenschaftliche Reporting und die Jahresgespräche.
- (2) Um den Einbezug von Datenpublikationen in das wissenschaftliche Reporting zu sichern, sind alle Forschenden dazu verpflichtet, ihre Datenpublikationen der Publikationserfassung am UFZ⁹ zu melden.
- (3) Ergänzend zu etablierten Metriken für Textpublikationen werden am UFZ Metriken für Forschungsdaten und Forschungssoftware ermittelt. Die Umsetzung orientiert sich an den Vorgaben und Empfehlungen der Helmholtz-Gemeinschaft.
- (4) Eine zitierbare Datenpublikation fördert die Reproduzierbarkeit von darauf aufbauenden Untersuchungen. Forschungsdatenpublikationen sollen daher durch einen persistenten Identifikator (PID), z. B. DOI oder Handle, zitierbar gemacht werden.
- (5) Die Forschenden machen die Verwendung von Daten durch Zitation kenntlich und halten dabei die anerkannten Prinzipien zur Datenzitation (Data Citation Principles¹⁰) ein.

⁷ Helmholtz Information & Data Science Academy (HIDA), <https://www.helmholtz-hida.de/>

⁸ Helmholtz Federated IT Services (HIFIS), <https://hifis.net/>

⁹ Publizieren am UFZ, <https://www.ufz.de/index.php?de=50166>

¹⁰ Martone M. (ed.) (2014): Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. San Diego CA: FORCE11. <https://doi.org/10.25490/a97f-egyk>.

7 Relevanz von Daten

Forschungsdaten können eine unterschiedliche Relevanz hinsichtlich der strategischen Bedeutung für Themenbereiche und Programmforschung am UFZ aufweisen. Daraus ergeben sich Konsequenzen für das Vorgehen bei der Speicherung der Daten sowie für die Entwicklung und den Betrieb erforderlicher Forschungsdateninfrastrukturen. Strategisch relevante Daten werden dabei bevorzugt behandelt und berücksichtigt.

8 Nutzung von Forschungsdateninfrastrukturen

- (1) Der Begriff Forschungsdateninfrastrukturen umfasst die Kombination aus Hardware, Software und personeller Unterstützung, die es ermöglicht, unter Anwendung verschiedener Methoden Daten zu erzeugen bzw. zu sammeln, zu analysieren, zu simulieren und schließlich zu publizieren und/oder zu archivieren.
- (2) Das UFZ implementiert und unterhält eine Forschungsdateninfrastruktur und stellt damit eine, an den FAIR-Prinzipien orientierte Aufbewahrung, Langzeitverfügbarkeit, Datensicherheit und die technische Verfügbarkeit von digitalen Forschungsdaten sicher. Darüberhinausgehende nutzungs- oder projektspezifische Anforderungen können bei Verfügbarkeit erforderlicher Ressourcen, die bereits bei der Projektantragstellung berücksichtigt werden sollten, über das RDM-Board umgesetzt werden.
- (3) Vom UFZ werden entsprechende leistungsfähige Infrastrukturen durch die Einheit RDM und die IT-Abteilung (WKDV) zur Verfügung gestellt und unter Wahrung der Verhältnismäßigkeit weiterentwickelt.
- (4) Datenerzeugende und -verarbeitende Forschungsinfrastrukturen (z. B. wissenschaftliche Großgeräte, Ausbauinvestitionen) sollen an die durch die Einheit RDM bereitgestellten Dateninfrastrukturen angebunden werden. Die Bereitstellung und der Betrieb von Basisdiensten und Speicherinfrastrukturen erfolgen durch die WKDV. Stehen in den Fachdisziplinen anerkannte nationale und internationale Forschungsdateninfrastrukturen als Alternativen zu den vom UFZ bereitgestellten Basisdiensten zur Verfügung, ist eine Nutzung dieser zu prüfen, zu ermöglichen und gegebenenfalls zu bevorzugen.
- (5) Stehen in den Fachdisziplinen anerkannte nationale und internationale Forschungsdateninfrastrukturen zur Verarbeitung, Speicherung und Archivierung von Daten und Metadaten bereit, die zur Nutzung eine Anbindung an die vom UFZ bereitgestellten Infrastrukturen benötigen, ist eine Anbindung an diese zu prüfen und unter Wahrung der Verhältnismäßigkeit zu realisieren.

9 Planung des Umgangs mit Forschungsdaten

- (1) In Abstimmung der Wissenschaftler:innen mit der Einheit RDM sowie unter Berücksichtigung der einschlägigen Anforderungen der Forschungsfördernden und -partner:innen werden sämtliche Schritte des Forschungsdatenmanagements – von der Datenerhebung bis zur Veröffentlichung und Zugänglichmachung der Forschungsdaten – festgelegt, die Festlegung

in einem Datenmanagementplan dokumentiert und entsprechend umgesetzt. Die inhaltliche Tiefe der Planung soll sich an internationalen Empfehlungen orientieren.¹¹ Dabei sind ethische, rechtliche und Verwertungsaspekte sowie die Einhaltung der Regeln der guten wissenschaftlichen Praxis zu berücksichtigen. Auf Ebene von Departments oder Arbeitsgruppen kann eine Konkretisierung erforderlicher Inhalte (z. B. in Bezug auf in den Communities etablierte disziplinspezifische Metadatenstandards) erfolgen.

- (2) Der Datenmanagementplan soll als lebendiges Dokument verstanden werden, welches im Verlauf eines Projektes ergänzt und angepasst werden kann. Er soll spätestens mit Abschluss des Projektes im UFZ Archiv oder einer anderen geeigneten Stelle zusammen mit den relevanten Datensätzen archiviert werden. Die Einheit RDM sowie die Ansprechpersonen für Forschungsdatenmanagement der Departments beraten zu inhaltlichen und organisatorischen Fragen bei der Erstellung von Datenmanagementplänen.
- (3) Die Einheit RDM berät in Fragen der Berücksichtigung und Einwerbung von Mitteln für das Datenmanagement in Projektanträgen.

10 Datenzugang und offene Daten

- (1) Das UFZ unterstützt und fördert die langfristige Sicherung und den grundsätzlich offenen Zugang zu Forschungsdaten aus öffentlich geförderter Forschung. Forschende, Vorgesetzte und Projektleitende entscheiden gemeinsam und möglichst frühzeitig über die öffentliche Bereitstellung der Forschungsdaten.
- (2) Soweit möglich, sollen die zugehörigen Forschungsdaten einschließlich relevanter Metadaten mit einer wissenschaftlichen Publikation zum frühestmöglichen Zeitpunkt offen zugänglich gemacht werden. Dabei sind Rechte am geistigen Eigentum, gesetzliche Regelungen, insbesondere der Schutz personenbezogener Daten, ethische Aspekte ebenso wie die wissenschaftlichen Interessen und vertraglichen Vereinbarungen mit Kooperationspartner:innen sowie ggf. Verwertungsinteressen zu berücksichtigen.
- (3) Bei der Veröffentlichung und Bereitstellung von Forschungsdaten sind diese mit Lizenzangaben zu versehen. Die Forschenden überprüfen bei der Auswahl der Infrastrukturen zur Bereitstellung und Veröffentlichung von Forschungsdaten die Zugangs- und Lizenzbedingungen. (→ Rechtliche Aspekte)
- (4) Grundlegende Metadaten, die der Auffindbarkeit von Forschungsdaten dienen, sollen ohne Zugangsbeschränkungen (gemeinfrei) verfügbar sein, sofern dem keine Geheimhaltungsbelange entgegenstehen.

¹¹ Science Europe (2021): Practical Guide to the International Alignment of Research Data Management - Extended Edition. <https://doi.org/10.5281/zenodo.4915862>.

- (5) In den Metadaten erfasste Personen, Textpublikationen, Institutionen, Software sowie in Beziehung stehende Daten sollen sofern möglich über geeignete persistente Identifikatoren (PID) referenziert werden (z. B. ORCID¹², ROR¹³, DOI¹⁴).

11 Qualitätssicherung

- (1) Die Qualitätssicherung soll den gesamten Datenlebenszyklus abdecken und somit von der Datenerhebung über Workflows und Methoden der Bearbeitung, deren Aufbewahrung und Sicherung, bis zum kontrollierten Löschen der Daten reichen.
- (2) Ein Kernelement der Qualitätssicherung ist die Erfassung von Metadaten nach fallspezifisch üblichen Standards. Die Metadaten enthalten Angaben zur Datenqualität und die an ihnen durchgeführten qualitätssichernden Maßnahmen (z. B. Methoden, Software). Datenerzeugende und -verantwortliche Personen sowie die Methoden sollen während des gesamten Lebenszyklus nachvollziehbar sein. Die datenerzeugenden und -verantwortlichen Personen sowie die bei der Datenverarbeitung eingesetzten Methoden und Software sind, sofern möglich, mit PIDs zur referenzieren. Wenn möglich sind die zur Erfüllung der FAIR-Anforderungen nötigen Metadaten zu erfassen.
- (3) Die Forschungsdaten sind korrekt, vollständig und unverfälscht zu managen, die Integrität der Forschungsdaten muss gewährleistet sein. Das UFZ stellt entsprechende Infrastrukturen (z. B. redundante Datenspeicher) zur Verfügung (→Nutzung von Forschungsdateninfrastrukturen).

12 Langzeitverfügbarkeit und Nachnutzung

- (1) Forschungsdaten sollen auch nach ihrer Erstverwertung für eine Nachnutzung langfristig in einer geeigneten, vertrauenswürdigen Art und Weise nachhaltig abgelegt und gesichert sowie ggf. veröffentlicht werden. Für die Bereitstellung und sachgerechte Nachnutzung von Forschungsdaten ist es notwendig, dass bei der Datenerhebung neben der Beschreibung der reinen Prozedur der Datenerfassung auch technische und deskriptive Metadaten erfasst werden (→ Qualitätssicherung).
- (2) Die Einheit RDM stellt eine Archivierungslösung für die langfristige Speicherung von Forschungsdaten bereit, um die sich gemäß den *Regeln zur Sicherung guter wissenschaftlicher Praxis am UFZ (IR-17/12, in der jeweils geltenden Fassung)* ergebenden Anforderungen zu erfüllen. Für die langfristige Speicherung können auch externe Fachrepositorien und Datenzentren unter Beachtung der hier geregelten Maßnahmen genutzt werden.
- (3) Die Forschenden berücksichtigen des Weiteren die Hinweise zur Sicherung, Aufbewahrung und Nutzung von Primärdaten und Protokollen aus den *Regeln zur Sicherung guter wissenschaftlicher Praxis am UFZ (IR-17/12, in der jeweils geltenden Fassung)*.

¹² ORCID – Open Researcher and Contributor ID. <https://orcid.org/>

¹³ ROR – Research Organisation Registry ID. <https://ror.org/>

¹⁴ DOI – DataCite Digital Object Identifier. <https://datacite.org/>

- (4) Für das Format der Daten sind bevorzugt freie und in der jeweiligen Disziplin etablierte Standardformate zu wählen, um die Interoperabilität zu erleichtern und langfristige Lesbarkeit zu gewährleisten. Die Sicherstellung der Nachnutzbarkeit macht es insbesondere bei Daten, die in urheberrechtlich geschützten (proprietären) Datenformaten abgelegt werden, notwendig, auch den Zugriff auf Software zum Lesen und Verarbeiten dieser Daten langfristig zu gewährleisten.

13 Rechtliche Aspekte

- (1) Die Inhaberschaft an Daten wird in den Arbeitsverträgen bzw. Gastverträgen zwischen den Forschenden und dem UFZ geregelt.
- (2) Jeder Datensatz ist mit einer Lizenzinformation (Verwertungs- bzw. Nutzungsbedingungen) zu versehen. Hierbei sind vorrangig geeignete offene, standardisierte Lizenzen zu wählen. Empfohlen wird die Anwendung der Creative Commons Lizenzen¹⁵, und speziell die CC0 Lizenz für Metadaten und die CC-BY Lizenz für Daten.
- (3) Hinweise zur Verarbeitung personenbezogener Daten werden in den *Regeln zur Sicherung guter wissenschaftlicher Praxis am UFZ, Teil A, Ziff. 6 „Rechtliche Rahmenbedingungen, Forschungsethik und Nutzungsrechte“ (IR-17/12, in der jeweils geltenden Fassung)* behandelt und sind durch die Forschenden zu berücksichtigen.

14 Einbettung in nationale und internationale Initiativen und Konsortien

- (1) Das UFZ kooperiert mit anderen wissenschaftlichen Institutionen und beteiligt sich an nationalen und internationalen Initiativen und Konsortien zur Erarbeitung gemeinsamer Standards und Strukturen im Bereich FDM und Open Science. Durch Beteiligung unter anderem in Projekten der Programmforschung der Helmholtz-Gemeinschaft (PoF), der Helmholtz-Inkubator Plattformen¹⁶, im DataHub¹⁷ des Forschungsbereichs Erde und Umwelt, in mehreren domänenspezifischen Konsortien der Nationalen Forschungsdateninfrastruktur (NFDI)¹⁸ und der European Open Science Cloud (EOSC)¹⁹ wird internationale Anschlussfähigkeit sichergestellt und das digitale „Datenökosystem“ mitgestaltet.
- (2) Forschende des UFZ werden durch die Einheit RDM unterstützt, ihre Datenprodukte in Infrastrukturen nationaler und internationaler Initiativen zu integrieren sowie dabei, diese Infrastrukturen zu nutzen und in die eigene Arbeit zu integrieren.

¹⁵ Creative Commons Lizenzen. <https://creativecommons.org/licenses/?lang=de>

¹⁶ Helmholtz Inkubator. <https://www.helmholtz.de/forschung/im-fokus/information-data-science/helmholtz-inkubator/>

¹⁷ Helmholtz DataHub. <https://earth-data.de/>

¹⁸ NFDI – Nationale Forschungsdateninfrastruktur. <https://www.nfdi.de/>

¹⁹ EOSC – European Open Science Cloud. <https://eosc-portal.eu/>

15 Definitionen

a) Datenlebenszyklus

Der Datenlebenszyklus ist ein Modell, das den zyklischen Charakter der Arbeit mit Daten aller Art, inkl. Informationen, in ihren verschiedenen Bearbeitungs- und Nutzungsstadien im Prozess der wissenschaftlichen Wertschöpfung beschreibt. Als wesentliche Stadien dieses Zyklus gelten die Datengenerierung (z. B. Messungen), die Datenaufbereitung, die Datenauswertung/-analyse, die Speicherung bis hin zur Langzeitarchivierung sowie die Verfügbarmachung durch Veröffentlichung (z. B. in Datenbanken und Repositorien, als Journal-Publikationen, auf Online-Plattformen) bis hin zur Nachnutzung in weiteren oder neuen Forschungskontexten, die sich auch durch die Lehre ergeben können.²⁰

b) Datenmanagementplan (DMP)

Ein Datenmanagementplan (DMP) ist ein wichtiges Werkzeug, um den Umgang mit den eigenen Forschungsdaten zu strukturieren. DMP können sowohl als Checklisten dienen, als auch zur laufenden Dokumentation der Erhebung bis zur langfristigen Speicherung bzw. Veröffentlichung der Daten.²¹

c) Datenqualität

Der Begriff Datenqualität bezeichnet die Güte und Verlässlichkeit von Datenobjekten selbst. Voraussetzung für deren wissenschaftliche Nutzbarkeit ist, dass die enthaltenen Daten mit dokumentierten Standards (z. B. von Messgeräten, Berechnungsalgorithmen) und anerkannten Methoden gewonnen werden, und dass diese Standards transparent sind und Nachhaltigkeit sichern. Wenn immer möglich sollte die Unsicherheit in einem Datum geeignet quantifiziert werden.

Die Bewertung von Datenqualität richtet sich nach den zu definierenden Ansprüchen, etwa an die Genauigkeit von Messwerten, abhängig von der jeweiligen Forschungsfrage und damit vom Verwendungszweck. Zudem wird Datenqualität dadurch bestimmt, dass Datensätze und -sammlungen ausreichend Informationen (in Form von möglichst standardisierten technischen und fachlichen Metadaten) über ihre Generierung, Weiterverarbeitung und Analyseverfahren beinhalten. Dies ist eine wesentliche Voraussetzung für eine Nachnutzung der Daten und die Reproduzierbarkeit von Forschungsergebnissen.²²

d) Forschungsdaten

Zu Forschungsdaten zählen u. a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Umfragedaten oder Beobachtungsdaten, methodische Testverfahren sowie Fragebögen. Da Forschungsdaten in einigen Fachbereichen auf der Analyse von Objekten basieren (z. B. Gewebe-, Material-, Gesteins-, Wasser- und Bodenproben, Prüfkörper, Installationen, Artefakte und Kunstgegenstände), muss der Umgang mit diesen ebenso nachhaltig sein und eine fachlich

²⁰ Rfll – Rat für Informationsinfrastrukturen (2016): Begriffserklärungen. Rfll Berichte No. 1. <http://nbn-resolving.de/urn:nbn:de:101:1-201607146410>

²¹ Helmholtz Metadata Collaboration (2021): Glossar. <https://helmholtz-metadaten.de/de/glossar>

²² Rfll – Rat für Informationsinfrastrukturen (2016): Begriffserklärungen. Rfll Berichte No. 1. <http://nbn-resolving.de/urn:nbn:de:101:1-201607146410>

adäquate Nachnutzungsmöglichkeit, wann immer sinnvoll und möglich, bedacht/berücksichtigt/ermöglicht werden.²³

e) Forschungsdatenmanagement (FDM)

Unter Forschungsdatenmanagement versteht man den Prozess der Transformation, Selektion und Speicherung von Forschungsdaten mit dem Ziel, diese langfristig und unabhängig vom Datenerzeuger zugänglich, nachnutzbar und nachprüfbar zu halten. Es können dazu an allen Punkten des Datenlebenszyklus strukturierte Maßnahmen ergriffen werden, um die wissenschaftliche Aussagekraft von Forschungsdaten zu erhalten, deren Zugänglichkeit durch Dritte für Auswertung und Analyse zu bewahren und die Nachweiskette zu sichern.²⁴

f) Forschungsdateninfrastruktur

Der Begriff Forschungsdateninfrastrukturen umfasst die Kombination aus Hardware, Software und personeller Unterstützung, die es ermöglicht, unter Anwendung verschiedener Methoden Daten zu erzeugen bzw. zu sammeln, zu analysieren, zu simulieren und schließlich zu publizieren und/oder zu archivieren.

g) Metadaten

Metadaten oder Metainformationen sind strukturierte Daten, die Informationen über Merkmale anderer Daten enthalten.²⁵ Metadaten bezeichnen alle zusätzlichen Informationen, die zur Interpretation der Forschungsdaten notwendig oder sinnvoll sind und die eine (automatische) Verarbeitung der Forschungsdaten durch technische Systeme ermöglichen.²⁶

h) Metadatenstandard

Um die Interoperabilität, d. h. die Verknüpfung und gemeinsame Bearbeitung, von Metadaten zu gewährleisten, wurden für bestimmte Zwecke Metadatenstandards geschaffen. Sie dienen einer inhaltlich als auch strukturell gleichförmigen Beschreibung von ähnlichen Daten. Ein Metadatenstandard kann oft durch ein sogenanntes Mapping in einen anderen Metadatenstandard überführt werden.²⁷

i) Open Data

Als Open Data (aus englisch open data ‚offene Daten‘) werden Daten bezeichnet, die von allen zu jedem Zweck genutzt, weiterverbreitet und weiterverwendet werden dürfen. Einschränkungen der Nutzung sind nur erlaubt, um Ursprung und Offenheit des Wissens zu sichern, beispielsweise durch Nennung des Urhebers oder die Verwendung einer Share-alike-Klausel²⁸.

²³ Deutsche Forschungsgemeinschaft (DFG) (2021): Umgang mit Forschungsdaten. https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf

²⁴ forschungsdaten.info (2022): Forschungsdaten und Forschungsdatenmanagement. Glossar. <https://forschungsdaten.info/praxis-kompakt/glossar/#c269836>

²⁵ Helmholtz Metadata Collaboration (2021): Glossar. <https://helmholtz-metadaten.de/de/glossar>

²⁶ forschungsdaten.org (2015): Metadaten. <https://www.forschungsdaten.org/index.php/Metadaten>

²⁷ forschungsdaten.info (2022): Forschungsdaten und Forschungsdatenmanagement. Glossar. <https://forschungsdaten.info/praxis-kompakt/glossar/#c269914>

²⁸ https://de.wikipedia.org/wiki/Open_Data

j) Open Science

Der Begriff Open Science bündelt Strategien und Verfahren, die allesamt darauf abzielen, die Chancen der Digitalisierung konsequent zu nutzen, um alle Bestandteile des wissenschaftlichen Prozesses über das Internet offen zugänglich und nachnutzbar zu machen. Damit sollen Wissenschaft, Gesellschaft und Wirtschaft neue Möglichkeiten im Umgang mit wissenschaftlichen Erkenntnissen eröffnet werden.²⁹

k) Persistente Identifikatoren/Identifizier (PID)

Persistente Identifizier (PID) sind global eindeutige Zeichenketten zur Identifikation von Objekten. Sie werden in einer verteilten Infrastruktur verwaltet und können durch diese aufgelöst werden.³⁰

l) Repositorium

Ein Repositorium ist eine Datenbank bzw. ein Datenarchiv zur Speicherung und Veröffentlichung von digitalen Forschungsdaten. Es dient vorrangig dazu, diese für einen sehr langen, idealerweise unbegrenzten Zeitraum aufzubewahren sowie verfügbar, zitierbar und nachnutzbar zu halten. Es können zudem über ein entsprechendes Rechte- und Lizenzmanagement die Zugriffs- und Nutzungsbedingungen für Forschungsdaten geregelt und festgelegt werden.

m) Standards

Standards und Regeln dienen bei der Prozessierung von Daten, Informationen und Wissen dem Ziel der Normierung und damit einer Form der Qualitätssicherung (Qualität im Sinne von Wiederverwendbarkeit) in inhaltlicher, formaler, rechtlicher, organisatorischer und technischer Hinsicht. Die Digitalisierung aller Bereiche der Wissenschaft und die hohe Dynamik digitaler Technik erfordern Standardisierungen insbesondere bei Daten und Metadaten, Austauschformaten, Schnittstellen, Datenmodellen, Auszeichnungssprachen und Vokabularen. Standardisierungen bzw. die Einhaltung von Standards beim Datenmanagement sind aufwändige, jedoch wissenschaftlich und wirtschaftlich sinnvolle Investitionen in die Qualitätssicherung und Nachhaltigkeit.³¹

16 Inkrafttreten

Die Regelung tritt zum 17.06.2024 in Kraft und ersetzt die Regelung „Grundsätze zum verantwortungsvollen Umgang mit Forschungsdaten am Helmholtz-Zentrum für Umweltforschung GmbH – UFZ“ in der Fassung vom 13.09.2018.

²⁹ Open Science AG (2014): Mission Statement. V 1.0.

³⁰ Helmholtz Metadata Collaboration (2021): Glossar. <https://helmholtz-metadaten.de/de/glossar>

³¹ RfII – Rat für Informationsinfrastrukturen (2016): Begriffserklärungen. RfII Berichte No. 1. <http://nbn-resolving.de/urn:nbn:de:101:1-201607146410>

Anhang 1 – FAIR Prinzipien

Das FAIR Prinzip:

- steht für „**F**indable, **A**ccessible, **I**nteroperable and **R**e-usable“ – auffindbar, zugänglich, integrierbar und nachhaltig verwertbar;
- dient dazu, nicht nur Menschen, sondern im Wesentlichen auch Maschinen die automatische Suche und Benutzung von Forschungsdaten und eine nachhaltige Nachnutzung dieser zu ermöglichen;
- erleichtert Erkundung, Zugang, Integration, Wiederverwendung, Zitation und Referenzierung von Forschungsdaten;
- sollte nicht nur auf Forschungsdaten Anwendung finden, sondern auch auf die Tools und Workflows, die zur Generierung dieser Daten verwendet wurden.

Die Benutzung von „(meta)data“ bedeutet, dass die Prinzipien nicht nur auf die Daten, sondern auch auf die Metadaten angewendet werden sollen.

To be **F**indable:

F1. (meta)data are assigned a globally unique and persistent identifier

F2. data are described with rich metadata (defined by R1 below)

F3. metadata clearly and explicitly include the identifier of the data it describes

F4. (meta)data are registered or indexed in a searchable resource

To be **A**ccessible:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol

A1.1 the protocol is open, free, and universally implementable

A1.2 the protocol allows for an authentication and authorization procedure, where necessary

A2. metadata are accessible, even when the data are no longer available

To be **I**nteroperable:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles

I3. (meta)data include qualified references to other (meta)data

To be **R**e-usable:

R1. meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1. (meta)data are released with a clear and accessible data usage license

R1.2. (meta)data are associated with detailed provenance

R1.3. (meta)data meet domain-relevant community standards