

Leitlinie für Forschungsdatenmanagement (FDM) am Leibniz-Institut für Deutsche Sprache (IDS)

in der Fassung vom 18.1.2023

PRÄAMBEL

Eines der wichtigsten satzungsmäßigen Ziele des Leibniz-Instituts für Deutsche Sprache (IDS) ist es, die deutsche Sprache in ihrem gegenwärtigen Gebrauch und in ihrer neueren Geschichte zu dokumentieren. Das IDS erkennt die grundlegende Bedeutung von Forschungsdaten und ihrer Dokumentation an, um dieses Ziel zu erreichen, und qualitativ hochwertige Forschung und wissenschaftliche Integrität zu erhalten und ist bestrebt, diesbezüglich den höchsten Standard anzustreben. Das IDS erkennt weiterhin an, dass korrekte und leicht auffindbare Forschungsdaten wesentliche Grundlage eines jeden Forschungsprojektes sind. Sie sind notwendig für die Nachvollziehbarkeit und Reproduzierbarkeit von Forschungsprozessen und deren Ergebnissen. Forschungsdaten haben einen langfristigen Nutzen für Forschung und Wissenschaft und das Potenzial für eine umfassende Nachnutzung und Verbreitung in der Gesellschaft. FDM-Leitlinien unterstützen Forschende und den wissenschaftlichen Nachwuchs im Umgang mit Forschungsdaten und tragen zu einem zukunftsfähigen Forschungsumfeld bei.

1. DEFINITIONEN

CARE-Prinzipien: CARE-Prinzipien für insbesondere linguistische Forschungsdaten stehen für den gemeinsamen Nutzen (*Collective benefit*), Selbstbestimmung der Verwendung (*Authority to Control*), Verantwortung (*Responsibility*) und Ethik (*Ethics*). Sie wurden in der Linguistik entwickelt, um die FAIR-Prinzipien zu ergänzen.

Datengeber: jede (natürliche oder juristische) Person, wie z. B. ein Presseverlag oder eine Forschungseinrichtung, die Daten an das IDS übermittelt.

Datenmanagementplan: Im Zuge des Forschungsdatenmanagements legen Projekte Datenmanagementpläne an, in denen die Prozesse im Zuge des Datenlebenszyklus von Forschungsdaten sowie hierfür benötigte Ressourcen und Services dokumentiert werden.

Datenübernahme: Der Prozess des Transfers von Forschungsdaten von anderen Institutionen oder von natürlichen Personen an das IDS.

Datenschutz: Schutz vor missbräuchlicher Verarbeitung personenbezogener Daten gemäß geltendem Recht und insbesondere der Datenschutz-Grundverordnung (DSGVO).

FAIR-Prinzipien: FAIR-Daten sind Daten, die den Grundsätzen der Auffindbarkeit (*Findability*), Zugänglichkeit (*Accessibility*), Interoperabilität (*Interoperability*) und Wiederverwendbarkeit (*Re-Usability*) entsprechen. Diese Grundsätze wurden von der Forschungsgemeinschaft formuliert und werden u. a. von der Europäischen Kommission unterstützt.

Forschende: Forschende sind alle in der Forschung aktiven Mitglieder des IDS, einschließlich Mitarbeiter/-innen, Promovierende, aber auch Hilfskräfte und Praktikanten. Ebenso eingeschlossen sind Personen, die nicht unmittelbar dem IDS angehören, die Einrichtungen aber für ihr Forschungsvorhaben nutzen (z. B. Gastforschende oder Kooperationspartner).

Forschungsdaten: Unter Forschungsdaten werden alle digitalen Daten zusammengefasst, die Ausgangspunkt, Gegenstand, Arbeitsschritt oder Ergebnis von Forschungsprozessen sind. Forschungsdaten können im Laufe ihres Forschungsdatenlebenszyklus und in der Laufzeit von Forschungsprojekten kontinuierlich verschiedene Formen annehmen (unterschiedliche

Varianten der Primärdaten, aufbereitete Daten, gemeinsam genutzte Daten, zur weiteren Bearbeitung veröffentlichte Daten). Sie können zu verschiedenen Zeitpunkten und Aufbereitungszuständen unterschiedliche Zugangsberechtigungen besitzen, z. B. als offene, zugriffsbeschränkte und nichtöffentliche Daten. Typische Beispiele von Forschungsdaten am IDS sind Texte, Audiodaten, Videodaten, Transkripte, Annotationen, Metadaten, Messdaten, Umfragedaten, Daten methodischer Testverfahren oder Simulationen, lexikalische Daten, Grammatiken, Protokolle. Zu den Forschungsdaten am IDS gehört insbesondere auch Forschungssoftware (z. B. Quellcodes, Tools, Skripte).

Forschungsdatenmanagement: Das Management von Forschungsdaten umfasst deren Planung, Erfassung, Verarbeitung, Kuratation, Dokumentation und Aufbewahrung. Es sichert den Zugang, die Nachnutzung, Reproduzierbarkeit und Qualitätssicherung aller Forschungsdaten, die wissenschaftlichen Ergebnissen zugrunde liegen, nach aktuellen technologischen, rechtlichen und ethischen Standards und Prinzipien (z. B. FAIR und CARE).

Grammatiken: Grammatiken sind systematische Beschreibungen syntaktischer Regelmäßigkeiten und Konventionen in natürlicher Sprache. Sie behandeln also den Aufbau und die Erscheinungsformen sprachlicher Einheiten wie Morphemen, Wörtern oder Phrasen. Als Darstellungsform kommen dabei zunehmend digitale Formate zum Einsatz, z. B. im grammatischen Informationssystem "grammis" oder in Form phänomenspezifischer Datensätze.

Korpora: Korpora sind digitale Sammlungen von Sprachdaten. Schriftliche Korpora werden aus schriftlichen Texten (z. B. Zeitungen, Literatur etc.) zusammengestellt. Bei den Korpora gesprochener Sprache handelt es sich um Sammlungen von Audio- und/oder Videoaufzeichnungen mündlicher Äußerungen. Korpora beinhalten zudem Metadaten. Typische Metadaten schriftlicher Korpora sind z. B. Autorschaft, Entstehungszeitpunkt, Entstehungsort; typische Metadaten mündlicher Korpora sind etwa sprecherbezogene Metadaten sowie Metadaten zum Sprechereignis. Korpora können zudem linguistische Annotationen beinhalten (z. B. Lemmatisierung, Part-of-Speech-Tags, (grammatische) Abhängigkeiten und Transkriptionen).

Lexikalische Daten: Lexikalische Daten beschreiben den Wortschatzaufbau und die Verwendung und Verbreitung von Wörtern in Sätzen, Texten und multimodaler Kommunikation. Die Daten liegen z. B. als Listen, Datenbanken, annotierten Sprachdaten, oder in Form regelmäßig strukturierter Wörterbuchtexte und multimedialen Formaten wie Karten, Abbildungen, Visualisierungen von Daten oder Hörbeispielen vor. Die konkrete Struktur der Daten richtet sich dabei an der zugrundeliegenden Forschungsfrage und der Funktion der Ressource aus und kann sich im Grad ihrer Formalisierung stark darin unterscheiden, ob sie für die Nutzung durch menschliche Benutzende oder eine maschinelle Nutzung gedacht sind. Retrodigitale, wissenschaftliche Wörterbücher enthalten zudem in der Regel umfangreiche Quellenverzeichnisse und Bibliographien der in ihnen zitierten Textbelege. Um- und Hilfetexte von Wörterbüchern dokumentieren wissenschaftliche Entscheidungen zur Datenkonzeption und geben Hinweise zur Benutzung; Glossare erklären die verwendete Fachterminologie der Metasprache.

Metadaten: Metadaten sind strukturierte Daten zur einheitlichen Beschreibung von Daten und Sprachressourcen. Sie sorgen dafür, dass den Beschreibungen von Instanzen ein- und derselben Ressourcenklasse eine einheitliche Struktur zugrunde liegt und erleichtern so das Suchen, Finden und Selektieren relevanter Sprachressourcen aus einer Vielzahl möglicher Sprachressourcen¹.

Nutzende: Nutzende sind alle natürlichen oder juristischen Personen, unabhängig von ihrer Zugehörigkeit vom IDS, denen zu anderen Zwecken als Archivierung Zugang zu Daten gewährt wird. Unterschiedenen werden interne und externe Nutzende. Interne Nutzende sind

¹ vgl. https://wiki.dnb.de/download/attachments/43523047/201209_metadaten.pdf

Angehörige des IDS, d. h. Personen, die sich über die DFN-AAI nach Auswahl der Organisation „Leibniz-Institut für Deutsche Sprache“ anmelden können. Alle anderen Nutzenden zählen zu den externen.

Open Access: Open Access steht in der Wissenschaft für den freien, öffentlichen und kostenlosen Zugang zu wissenschaftlichen Informationen im Internet.

Für weitere Informationen vgl. die Open-Access-Leitlinie des IDS (<https://www.ids-mannheim.de/bibliothek/open-access-am-ids/leitlinie/>).

Sprachressourcen: Sprachressourcen sind eine Klasse elektronisch vorliegender Sprachdaten, die mit linguistischen Informationen (z. B. Annotationen) angereichert sind. Sprachressourcen können beispielsweise Korpora, Lexika, Wortnetze oder Daten gesprochener Sprache sein.

Standard: Ein Standard ist ein Dokument, das Anforderungen, Spezifikationen, Richtlinien oder Merkmale bietet, die konsequent genutzt werden können, um sicherzustellen, dass die Materialien, Produkte, Prozesse und Dienstleistungen für ihren Zweck geeignet sind. Standards sind verpflichtet, ein gemeinsames Verständnis oder eine gemeinsame Interpretation von Forschungsdaten herzustellen und insbesondere ihre Interoperabilität zu gewährleisten.

Wissenschaftliche Publikation: Wissenschaftliche schriftliche Arbeit, Fachpublikation oder auch wissenschaftliche Textpublikation von einem Autor, einer Autorin oder mehreren Autor/-innen, deren Veröffentlichung i.d.R. bei einem wissenschaftlichen Verlag schon erfolgt oder vorgesehen ist.

2. GELTUNGSBEREICH

Diese Leitlinie für das Management von Forschungsdaten richtet sich an alle Forschenden, die im weitesten Sinne mit Forschungsdaten des IDS befasst sind. Sie wurde am 18.1.2023 vom Vorstand des IDS verabschiedet. Im Falle von Drittmittelprojekten sollte, soweit möglich, diese Leitlinie berücksichtigt werden. Spezifische Vereinbarungen mit Drittmittelgebern in Bezug auf das Datenmanagement haben Vorrang vor dieser Leitlinie.

3. RECHTE AM GEISTIGEN EIGENTUM

Das Urheberrecht garantiert jedem den Schutz seiner geistigen Schöpfungen. Ob Forschungsdaten dem Schutz des Urheberrechtsgesetzes unterliegen, ist davon abhängig, ob entweder die Anforderungen an die geistige Schöpfungshöhe oder die Voraussetzungen des Datenbankurheberrechts erfüllt werden.

Aufgrund ihrer Spezifität erfüllen Sprachdaten in der Regel die Schöpfungshöhe und sind daher urheberrechtlich geschützt. Darüber hinaus können Forschungsdatensätze (z. B. Korpora) durch das Datenbankherstellerrecht geschützt werden.

Die Verwertungsrechte an Forschungsdaten stehen dem Urheber kraft Gesetz zu. Diese Rechte sind nicht übertragbar. Das IDS erlangt die Nutzungsrechte an den Forschungsdaten über vertragliche Vereinbarungen, wie z. B. Arbeitsverträge, Werkverträge oder Datenübertragungsverträge. Nutzungs- und Verwertungsrechte an Forschungsdaten werden in einem Arbeitsvertrag oder einer sonstigen vertraglichen Vereinbarung zwischen den Forschenden und dem IDS definiert.

Das Datenbankherstellerrecht an am IDS zusammengestellten Forschungsdatensätzen steht grundsätzlich per Gesetz (§ 87a UrhG) dem IDS zu.

Auch Software kann urheberrechtlich geschützt sein. An Software, die von IDS-Mitarbeiter/-innen im Rahmen ihrer Tätigkeit und vom IDS beauftragten Dienstleistern entwickelt wurde, ist ausschließlich das IDS zur Ausübung aller vermögensrechtlichen Befugnisse per Gesetz (§ 69b UrhG) berechtigt.

Nutzungs- und Verwertungsrechte, sowie Datenbankherstellerrecht können weiterhin durch zusätzliche Übereinkünfte definiert werden (z. B. in einer Zuwendungsvereinbarung oder einem Konsortialvertrag).

4. UMGANG MIT FORSCHUNGSDATEN

Forschungsdaten müssen identifizierbar, zugänglich, zurückverfolgbar, interoperabel und, wenn möglich, für die spätere Nutzung und Überprüfung bzw. Reproduktion verfügbar sein.

Vorbereitung und Planungen:

Vor der Erhebung oder einer geplanten Übernahme von Forschungsdaten sind aktuelle technologische Standards, Metadatenstandards und Qualitätsstandards der Daten, Kosten und Dauer der Datenerhaltung, nach Möglichkeit zudem auch datenschutzrechtliche, urheberrechtliche, lizenzrechtliche und ethische Aspekte zu klären. Im IDS gelten insgesamt diesbezüglich institutsweit einheitliche Prozesse, aber auch spezifische Regelungen.

Datenmanagementpläne definieren die Verantwortlichkeiten, Zuständigkeiten, Prozesse und Bedarfe für die Erhebung, Erschließung und Speicherung der Forschungsdaten über alle Phasen des Datenlebenszyklus hinweg.

Dokumentation:

Aus Gründen der Nachvollziehbarkeit, Anschlussfähigkeit und Nachnutzbarkeit sind Forschungsdaten zu dokumentieren. Die Dokumentation erfüllt den jeweiligen Standard für den Umgang mit Forschungsdaten. Die Dokumentation erfolgt dabei für alle Phasen eines Datenlebenszyklus. In Dokumentationen werden z. B. Ablageorte, Informationen über Probanden, Erhebungsorte, Annotationsrichtlinien, Codebooks, Versionen von Daten und Berechtigungen möglichst detailliert für die Nutzung im Projektkontext und für die Nachnutzung beschrieben.

Speicherung:

Es ist von besonderer Bedeutung, die Integrität von Forschungsdaten zu bewahren. Forschungsdaten müssen auf eine korrekte, vollständige, unverfälschte und verlässliche Art und Weise gemäß den aktuellen technischen und organisatorischen Standards gespeichert werden.

Die den Forschungsdaten zugrundeliegenden Daten (z. B. Rohdaten, Analyseprogramme, Annotationsprogramme) werden in Abhängigkeiten von gesetzlichen Vorgaben oder vertraglichen Vereinbarungen und IDS-spezifischen Policies ebenfalls gespeichert.

Für Forschungsdaten, die im Laufe eines Forschungsprozesses entstehen, sollen im Projekt Prozesse zur Bewertung der Relevanz entstehender Daten, zu ihrer Speicherung sowie zu Löschrufen nicht langfristig zu speichernder Daten definiert und umgesetzt werden.

Forschungsdaten und -unterlagen sind mindestens so lange aufzubewahren und zugänglich zu halten, wie es gemäß den Rechten am geistigen Eigentum oder den Auflagen der Forschungsförderer im Rahmen der anwendbaren gesetzlichen und vertraglichen Bestimmungen erforderlich ist.

Wenn Forschungsdaten und zugehörige Unterlagen nach Ablauf der Speicherfrist gelöscht werden sollen oder aus rechtlichen bzw. ethischen Gründen gelöscht oder vernichtet werden müssen, so darf dies nur unter Berücksichtigung jeglicher rechtlicher oder ethischer Gesichtspunkte geschehen. Die Löschung muss nachvollziehbar sein und dokumentiert werden. Bei der Entscheidung über Erhalt oder Löschung der Daten müssen die Interessen und vertraglich festgelegten Bestimmungen von Drittmittelgebern, und sonstigen Beteiligten, insbesondere von Datengeber/-innen, Mitwirkenden und Kollaborationspartner/-innen, berücksichtigt werden. Dabei müssen Aspekte der Sicherheit und Vertraulichkeit bedacht werden.

Verfügbarmachung und Verteilung:

Zur Publikation vorgesehene Forschungsdaten sollen in einem geeigneten Repositorium oder Archivierungssystem, wie beispielsweise dem Forschungsdaten-Repositorium/Digitalen Langzeitarchiv des Leibniz-Instituts für Deutsche Sprache oder in anderen anerkannten Fachrepositorien, abgelegt werden. Eine Dokumentation der Forschungsdaten bzw. Metadaten muss

hierfür vorliegen. Die Daten sollten mit persistenten Identifikatoren (PID, z. B. DOI oder Handle) versehen werden.

In Übereinstimmung mit den Rechten am geistigen Eigentum und unter der Voraussetzung, dass keine Rechte Dritter, gesetzliche Bestimmungen oder andere Schutzrechte (etwa Persönlichkeitsrechte oder Urheberrechte) dies verbieten, sollten nach Möglichkeit Forschungsdaten mit einer freien Lizenz versehen und offen verfügbar gemacht werden. Forschungsdaten, für die besondere Schutzrechte greifen, werden mit Zugangsbeschränkungen versehen, die die jeweiligen Berechtigungen berücksichtigen.

Forschungsdaten, die zur Nachnutzung vorgesehen sind, sollen in zitierbarer Form zur Verfügung gestellt werden. Bei Bedarf werden Empfehlungen zur Zitation für die entsprechenden Datensätze gegeben.

5. VERANTWORTLICHKEITEN, RECHTE, PFLICHTEN

Die Verantwortung für das Forschungsdatenmanagement während und nach der Projektlaufzeit liegt beim IDS und seinen Forschenden. Dabei müssen die „Regeln zur Sicherung guter wissenschaftlicher Praxis am Leibniz-Institut für Deutsche Sprache“ beachtet werden.

5a. Verantwortlichkeiten der Forschenden

Forschende

i. gehen mit Forschungsdaten so um, dass die Grundsätze und Anforderungen dieser Leitlinie erfüllt werden.

ii. sammeln, dokumentieren, speichern und archivieren Forschungsdaten und die damit verbundene Dokumentation, so dass ein Zugang bzw. eine ordnungsgemäße Löschung möglich ist. Dies beinhaltet auch die Vereinbarung von Abläufen und Verantwortlichkeiten in gemeinsamen Forschungsprojekten. Derartige Informationen sollen Bestandteil eines Datenmanagementplans (DMP) sein, der die Sammlung, Verwaltung, Aufbewahrung, Nutzung und Veröffentlichung der verwendeten Daten dokumentiert und die Voraussetzungen für Integrität und Vertraulichkeit der Daten beschreibt.

iii. planen, soweit möglich, die weitere Nutzung der Daten insbesondere nach Projektabschluss. Dies umfasst sowohl die Festlegung von Nutzungs- und Verwertungsrechten nach Projektende, einschließlich der Zuweisung entsprechender Lizenzen, als auch die Regelung von Datenspeicherung und -archivierung im Fall eines Ausscheidens aus dem IDS.

iv. erfüllen alle relevanten organisatorischen, regulatorischen, institutionellen und sonstigen vertraglichen und gesetzlichen Bestimmungen sowohl in Bezug auf Forschungsdaten als auch auf die Verwaltung zugehöriger Forschungsunterlagen (zum Beispiel bei Kontext- oder Herkunftangaben) und treten dafür ein. Sie greifen dazu auf die Beratungs- und Fortbildungsangebote nach 5b.ii zurück.

5b. Verantwortlichkeiten des IDS

Das IDS

i. unterstützt ihre Organisationseinheiten und stellt angemessene Mittel und Ressourcen für Forschungsförderung, Dienstleistungen, den Betrieb von Organisationseinheiten, Infrastrukturen und Mitarbeiterqualifizierung bereit.

ii. fördert die Einhaltung der „Regeln zur Sicherung guter wissenschaftlicher Praxis am Leibniz-Institut für Deutsche Sprache“.² Dazu stellt es Vorlagen für DMPs bereit, betreibt Monitoring und bietet Qualifizierungsmaßnahmen sowie Unterstützung und Beratung an. Dies geschieht

² Vgl. https://www.ids-mannheim.de/fileadmin/org/gremien/ombud/Regeln_GWP_am_IDS_-_2022-01-18.pdf

in Übereinstimmung mit aktuellen Richtlinien, Verträgen mit Drittmittelgebern, internen Satzungen, Verhaltenskodizes und weiteren relevanten Leitfäden.

iii. entwickelt Mechanismen und stellt Dienste bereit, um Forschungsdaten zu speichern, sicher aufzubewahren und abzulegen, damit der Zugang zu den Forschungsdaten während und nach Abschluss von Forschungsprojekten gewährleistet werden kann.

iv. stellt den Zugang zu den oben beschriebenen Diensten und Infrastrukturen bereit, so dass die Forschenden die Auflagen von Drittmittelgebern und weiteren Rechtsträgern einhalten können und ihre in dieser Leitlinie beschriebenen Verantwortlichkeiten wahrnehmen können.

6. GÜLTIGKEIT

Wie vom Vorstand des IDS beschlossen, tritt diese Leitlinie am 19.1.2023 in Kraft.