

Leitfaden für die Publikation und Archivierung von Forschungsdaten an der TU Braunschweig

Die TU Braunschweig bietet ihren Forschern die Möglichkeit, ihre Forschungsdaten auf [LeoPARD](#) zu archivieren und, falls gewünscht, zu publizieren. Im Folgenden wird auf die wichtigsten Punkte eingegangen.

1) Forschungsdaten vorbereiten

Checkliste

- Ich habe meine Daten in einer nachvollziehbaren Ordnerstruktur gespeichert.
- Meine Ordner- und Dateibenennungen sind auch für Dritte nachvollziehbar, d.h. sie sind systematisch, inhaltsbezogen und eindeutig.
- Die Daten sind so gestaltet, dass sie selbständige Nachvollziehbarkeit und Nachnutzbarkeit ermöglichen.
- Ich habe archivierungswürdige Daten identifiziert, bzw. zu veröffentlichte Daten ausgewählt, die die wissenschaftliche Relevanz meiner Arbeit abbilden.
- Meine Daten liegen im besten Falle in standardisierten, verbreiteten Dateiformaten für kostenfreie Darstellungsprogramme vor (z.B. XML, CSV, TXT, TIFF).
- Ich habe darauf geachtet, dass die Daten nicht schreibgeschützt sind.

Ordnerstrukturierung

Um eine möglichst nachvollziehbare, klare Strukturierung der Forschungsdaten zu erlangen, eignet sich eine hierarchische Struktur. Wie diese genau aussieht, hängt vom Datensatz ab und muss jedes Mal individuell festgelegt werden. Bei heterogener Struktur der Daten (viele verschiedene Dateierarten) kann man Ordnerarten anlegen, z.B.:

- Materialien → Daten, Dokumentationen, Publikationen
- Daten → Texte, Fotos, Modelle, Tabellen

Sind die Daten homogener, können sie auch nach Zeit, Ort oder Methode unterteilt werden, z.B.:

- Projekt_XY → Untersuchungsgebiet01 → Messungen

Wichtig ist dabei, dass die gewünschten Dateien schnell und intuitiv gefunden werden können.

Dateibenennung

Bei der Benennung von Dateien sollte von Anfang an eine Namenskonvention festgelegt werden, die während des gesamten Projekts konsistent angewendet, insbesondere wenn mehrere Parteien an einem Projekt beteiligt sind. Die Benennung sollte Aufschluss über den Inhalt der Daten geben und helfen, sie eindeutig zu identifizieren. Dafür bieten sich besonders das Erstellungsdatum, eine kurze Beschreibung, der Ort, Projektname, Probennummer, Analyseverfahren oder Versionsnummer an.

Zudem sollten folgende Regeln beachtet werden:

1. Dateinamen sollten kurz aber erklärend sein.
2. Es sollten keine Sonderzeichen, Leerzeichen, Satzzeichen, oder Umlaute verwendet werden.

3. Es sollten Großbuchstaben und Unterstriche stattdessen verwendet werden.
4. Großbuchstaben beeinflussen die Sortierung, weswegen auf eine konsistente Benennung geachtet werden sollte.
5. Als Datumsformat eignet sich am besten: YYYYMMDD.
6. Wenn Nummern angegeben werden, sollten diese immer zweistellig oder gar dreistellig sein, sodass die 0 als Platzhalter dient (z.B. Probe01 und nicht Probe1).
7. Die einzelnen Elemente des Dateinamens sollten so angeordnet sein, dass die Datei schnellstmöglich gefunden wird, Namensanfänge wie „Test“ oder „Entwurf“ sollten vermieden werden.
8. Wenn verschiedene Versionen einer Datei vorhanden sind, sollte ein V mit Nummerierung verwendet werden, um Ordnung zwischen den Versionen zu halten (V01, V02, etc.).
9. Wiederholungen von Informationen aus Ordnernamen in Dateinamen sollten vermieden werden.

Sinnvolle Elemente für die Dateibenennung

Sinnvolle Elemente	Hinweise und Beispiel
Name des Verfassers oder Initialen	LG
Inhaltliche/ Fachliche Beschreibung	Schlagwörter
Datum der Erstellung v.a. für Versionskontrolle	YYYYMMDD
Bearbeitungsstatus	Original, Entwurf, Teil
Version	V02

Beispiele für Dateibenennungen

Gut	Schlecht
20200813Agenda	13Aug2020Agenda
2020UBWorkshopFDM	Der_2020_Workshop_zum_Thema_Forschungsdatenmanagement_der_UB!
20200804_ProjektA_Protokoll_LG	Protokoll Laura

Vertiefende Informationen: http://www.data.cam.ac.uk/files/gdl_tilsdocnaming_v1_20090612.pdf

Die Namenskonvention der Dateien und Ordnerstruktur sollten in der beiliegenden README-Datei (s. Abschnitt 2) des Datensatzes erläutert werden, falls Kürzel verwendet werden.

Datenselektion

Um zu entscheiden, welche Daten geeignet für eine Archivierung/ Publikation sind, können sie in folgende Kategorien eingeteilt werden.

Kategorie 1: Sehr wertvolle Forschungsdaten

Forschungsdaten, die

- einzigartig oder sehr aufwändig wiederherzustellen sind und
- für eine große Community und/oder für ein breites Spektrum an Nachnutzungsmöglichkeiten interessant sind.

→ Unbedingt archivieren / publizieren!

Kategorie 2: Begrenzt wertvolle Forschungsdaten

Forschungsdaten, die entweder

- mit vertretbarem Aufwand wiederherzustellen sind oder
- nur für kurze Zeit, eine kleine Zielgruppe oder nur zum Nachweis von Forschungsergebnissen interessant sind.

→ Archivieren/ publizieren, wenn Aufwand und Kosten nicht dagegen sprechen.

Kategorie 3: Wenig wertvolle Forschungsdaten

Forschungsdaten, die

- nur zu Testzwecken entstanden,
- nicht lesbar oder nicht ausreichend beschrieben sind.

→ Daten sind nicht geeignet für Archivierung/ Publikation.

Dateiformate

Nicht jedes Dateiformat ist für eine Langzeitarchivierung geeignet. Damit Forschungsdaten auch nach einigen Jahren lesbar und zu verwenden sind, empfehlen sich offene, nicht proprietäre Dateiformate.

	Empfehlenswert	Weniger empfehlenswert
Text	PDF/A, TXT, HTML	Word (.docx/.doc)
Tabellen	CSV, SPSS portable	Excel (.xlsx/.xls)
Präsentation	PDF/A	PowerPoint (.pptx/.ppt)
Multimedia	MP4, WAV, AVI	WMV
Bilder	TIFF, JPEG2000, PNG	GIF, JPG
Datenaustausch	XML, RDF, JSON	SDXF

Weiterführende Informationen:

<https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/formate-erhalten/>

2) Dokumentation vorbereiten

Checkliste

- Ich habe eine umfassende Dokumentation betrieben, die eine Reproduzierbarkeit und Nachnutzung des Datensatzes ermöglicht. Dazu gehören
 - Erhebung der Forschungsdaten
 - Verarbeitung der Forschungsdaten
 - Struktur des Datensatzes
- Ich habe eine separate README-Datei (TXT oder PDF/A) mit beschreibenden Metadaten angelegt.

Metadaten

Metadaten beschreiben Forschungsdaten. Sie dienen der besseren Dokumentation und Auffindbarkeit von Daten. Typische Metadaten sind Titel, Verlag, Autor:in oder Datum. Darüber hinaus enthalten Metadaten beschreibende Informationen über den Kontext der Daten, wie Messgerät, Messort oder verwendete Software. Diese helfen, den Datensatz besser zu verstehen. Metadaten erleichtern das langfristige Verständnis der erhobenen Forschungsdaten und begünstigen ihre Nachnutzung, weshalb sie essentiell für eine Archivierung, bzw. Publikation sind.

Die Verknüpfung von Daten und Metadaten kann z. B. über einen persistenten Identifier oder durch eine README-Datei geschehen.

README-Datei

Eine README-Datei ist eine Textdatei, die begleitend zu einem Datensatz zur Verfügung gestellt wird und dabei hilft, ihn zu beschreiben und zu verstehen. Sie enthält vorwiegend kontextuelle oder strukturelle Metadaten, um die Interpretierbarkeit und Nachnutzung zu erleichtern. Solche Informationen können z. B. Einheiten von Messparametern, Informationen über Verarbeitungsschritte oder auch eine Übersicht über Einzeldateien eines Datenpaketes sein. Insgesamt gilt: Halten Sie die README-Datei kurz und einfach, mit den zum Verständnis des Datensatzes notwendigen Informationen.

Bei sehr heterogenen Datensätzen (z.B. bestehend aus Geodaten, Matlab-Skripten und Tabellen mit Messdaten) können a) alle Informationen in einer README-Datei beschrieben werden, wobei Bezug auf einzelne Datenspezifika genommen wird, oder es kann b) ein "Haupt"-README erstellt werden, das den Zusammenhang des Gesamtdatensatzes beschreibt und ergänzend spezifische READMEs für Unterdatensätze oder auch einzelne Dateien, welche im jeweiligen Unterverzeichnis abgelegt werden. Bei mehreren README-Dateien sollte darauf geachtet werden, dass die Benennung der READMEs Aufschluss darauf gibt, welchen Datensatz/ Datei sie beschreiben. Es sollte auf eine einheitliche Formatierung der README-Dateien geachtet werden.

Wir haben eine [README-Vorlage](#) für Sie erstellt. Nutzen Sie diese und passen Sie sie für Ihren Datensatz an, indem Sie Punkte streichen oder ergänzen.

3) Exkurs Forschungssoftware

Checkliste

- Meine Software ist versioniert.
- Meine Software habe ich mit zugehörigem Quell- oder Sourcecode versehen.
- Ich habe den Zweck der Software in der beiliegenden README-Datei formuliert.
- Der Code ist modularisiert und nachvollziehbar gestaltet, z.B. mittels selbsterklärender Namen oder Kommentare.

Veröffentlichung von Forschungssoftware

Software nimmt einen immer größeren Stellenwert in der Forschung ein. Sie hilft nicht nur bei der Analyse von Daten, sondern ist oftmals auch für deren Herstellung entscheidend. Oftmals braucht man die zugrundeliegende Software, um Forschungsdaten richtig verstehen und reproduzieren zu können. Zudem kann selbst entwickelte Forschungssoftware ein eigenständiges Forschungsergebnis darstellen. Deshalb sollte Forschungssoftware wie Forschungsdaten zur Nachnutzbarkeit publiziert werden.

Metadaten für Forschungssoftware

Forschungssoftware benötigt eigene Metadaten, um sie gut zu beschreiben und nachvollziehbar zu machen. Ein Beispiel für einen Metadatenstandard ist [CodeMeta](#).

Archivierung von Forschungssoftware

Die Langzeitarchivierung von Software gestaltet sich schwierig. Zum einen muss sie verständlich beschrieben sein und zum anderen muss dafür gesorgt werden, dass sie ausführbar gehalten wird. Darin liegt ein Problem, denn oft ist Software stark abhängig von der Rechenumgebung in der oder für die sie entwickelt wurde, sprich von Betriebssystem, Hardware oder Software-Bibliotheken. Diese ändern sich mit fortschreitender Technik sehr schnell, sodass entwickelte Software nicht mehr gelesen oder ausgeführt werden kann. Um dieses Problem zu lösen, gibt es noch keine endgültige Lösung. Ein Ansatz ist die Nutzung von Emulation, bei der zusätzlich zur eigentlichen Software auch noch der ausführende Rechner samt Betriebssystem nachgebildet wird.

Weiterführende Informationen

Projekt Suresoft: <https://suresoft.dev/>

4) Rechtliche Aspekte abklären und Lizenz wählen

Checkliste

- Ich habe mich vergewissert, dass ich mit der Veröffentlichung keine Rechte Dritter (z. B. Datenschutz, Urheberrecht, Geheimhaltungsabreden) verletze.
- Ich habe eine geeignete Lizenz (z.B. Creative Commons) für die Nachnutzung ausgewählt.

Beim Umgang mit Forschungsdaten gibt es viele rechtliche Fragestellungen zu beachten. Es ist ratsam bereits am Anfang eines Projekts rechtliche Belange und Verantwortlichkeiten abzuklären. Die

folgende Abbildung bietet einen ersten Überblick, welche Fragen man vor der Publikation von Forschungsdaten bedacht haben muss.



entnommen aus: Elisabeth Böker, & Peter Brettschneider. (2020, Mai). Forschungsdaten veröffentlichen - Organisatorische und rechtliche Fragen. Zenodo. <http://doi.org/10.5281/zenodo.3864901>, CC BY 4.0

Urheberrecht

Mit dem Urheberrecht werden geistige Schöpfungen und Leistungen geschützt. Liegt ein Urheberrecht vor, ist die Nutzung, Weitergabe und Verarbeitung des Werks nur mit Einwilligung des Urhebers möglich. Forschungsdaten unterliegen nur selten dem Urheberrecht, es kann jedoch Ausnahmen geben. Oftmals werden publizierte Forschungsdaten allerdings mit Lizenzen versehen, die eine mögliche Nachnutzung klären. Datenbanken bilden im Übrigen eine Ausnahme und sie genießen laut deutschem Recht einen spezifischen Schutz, wodurch die Ersteller das alleinige Recht zur Verbreitung und Vervielfältigung haben (§ 87b UrhG). Allgemein endet der urheberrechtliche Schutz 70 Jahre nach dem Tod des Urhebers. Danach gelten die Werke als gemeinfrei.

Beschreibende Metadaten von Forschungsdaten sind i.d.R. als nicht urheberrechtlich geschützt zu betrachten, da es sich meist um relativ kurze, rein beschreibende Darstellungen handelt.

Neben einem Urheberrecht können Forschungsdaten in bestimmten Fällen auch einem Patentrecht unterliegen, wenn die Forschungsdaten technische Lehre beschreiben. In solchen Fällen darf nur der Patentinhaber die patentierten Daten nutzen.

Arbeits- und dienstvertragliche Vorgaben

Open-Access-Leitlinie der TU Braunschweig: <https://doi.org/10.24355/dbbs.084-202008280945-0>

Leitlinie zum Umgang mit Forschungsdaten der TU Braunschweig:
<https://doi.org/10.24355/dbbs.084-201911211552-0>

Es ist im Allgemeinen darauf zu achten, was im Arbeitsvertrag bezüglich der Nutzungsrechte (gehören i.d.R. dem Arbeitgeber) an erzeugten Forschungsdaten festgehalten ist und ob z.B. ein Betriebsgeheimnis vorliegt.

Auch Förderer von Forschungsprojekten können Ansprüche haben, z.B. wenn es um die Rechte an erstellter Software im Zuge des Forschungsprojektes geht. Im Zweifelsfall sollte im Vorfeld einer Veröffentlichung nachgefragt werden, um ggf. eine Zustimmung für die Verbreitung einzuholen.

Datenschutz

Beim Umgang mit personenbezogenen Daten müssen verschiedene datenschutzrechtliche Aspekte beachtet werden. Durch eine Anonymisierung können datenschutzrechtliche Restriktionen vermieden werden. „Anonymisieren ist das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbar natürlichen Person zugeordnet werden können“ (§ 3 Abs. 6, BDSG). Des Weiteren gibt es die Möglichkeit zur Pseudonymisierung. Diese beschreibt „die Verarbeitung personenbezogener Daten in einer Weise, in der die Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzliche Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die Daten keiner betroffenen Person zugewiesen werden können“ (§ 46 Abs. 5, BDSG neu).

Liegt ein Personenbezug vor, dürfen Forschungsdaten nur mit der Unterzeichnung einer Einwilligungserklärung der Probanden erhoben und weiterverarbeitet werden.

Muster für Einwilligungserklärungen:

<https://www.econstor.eu/dspace/bitstream/10419/97181/1/785263330.pdf>

Der Bereich Datenschutz ist ein weites Feld. Die TU Braunschweig hat eigens einen [Datenschutzbeauftragten](#), der für jegliche Fragen rund ums Thema Datenschutz zur Verfügung steht.

Publikationsvereinbarungen mit dem Repositorium

Repositorien stellen Nutzungsverträge /-vereinbarungen für die Publikation/ Archivierung von Forschungsdaten auf. Diese gilt es im Vorfeld gut durchzulesen und es sollte sichergestellt sein, dass alle angegebenen Vorgaben erfüllt sind.

Vertragsbedingungen von LeoPARD:

<https://leopard.tu-braunschweig.de/content/publish/contract.xml>

Datennachnutzungsvereinbarungen

Zugriffsberechtigung

Es ist empfehlenswert, Forschungsdaten so offen wie möglich zur Verfügung zu stellen. Jedoch können die Daten auch mit einem Embargo versehen (Zugriffverzögerung für eine bestimmte Zeit) oder eingeschränkt werden, wer Zugriff auf die Daten erhält. Sensible Forschungsdaten, die z.B. personenbezogene Informationen enthalten, die sich nicht vollständig anonymisieren lassen, wären ein Fall für einen solchen eingeschränkten Zugriff.

Folgende Varianten bietet LeoPARD für Forschungsdaten an:

1. Publikation (uneingeschränkter Zugriff)

Bei dieser Variante gibt es keine Beschränkungen für den Zugriff auf die Daten. Jeder kann die Daten einsehen und herunterladen. Durch die Wahl einer passenden Nutzungslizenz kann bestimmt werden, wie die Daten nachgenutzt werden dürfen.

2. Embargo (verzögerter Zugriff)

Ein Embargo kann für die Veröffentlichung des Datensatzes festgelegt werden. Bis zum Ende des Embargos ist der Datensatz gesperrt und für andere nicht zugänglich. Erst nach Beendigung der selbst gewählten Embargofrist wird der Datensatz veröffentlicht.

Die Verwendung eines Embargos kann sinnvoll sein, wenn die Veröffentlichung von Forschungsdaten verzögert werden soll, z. B. bis zur Veröffentlichung eines zugehörigen Zeitschriftenartikels oder aus kommerziellen Gründen bei der Anmeldung eines Patents.

3. Publikation (eingeschränkter Zugriff)




Verschiedene Arten von Zugriffsbedingungen können gewählt werden. LeoPARD ermöglicht es z.B., dass Forschungsdaten nur von Mitgliedern der TU Braunschweig eingesehen werden können. Wenn das nicht genügt, kann auch ein eingeschränkter Zugriff ausgewählt werden, bei dem nur die Metadaten des Datensatzes einsehbar sind, der Datensatz selbst aber nicht. Bei Interesse eines Forschers kann sich dieser beim Autoren melden und dieser entscheidet individuell, ob der Forscher Zugriff zu dem Datensatz erhalten soll. Dies kann hilfreich sein, wenn rechtliche oder ethische Gründe es nicht erlauben, den Datensatz frei zur Verfügung zu stellen. Zudem ist es so auch möglich mit dem interessierten Forscher abzuklären, wofür er die Daten nutzen möchte.





4. Archivierung

Natürlich ist es auch möglich, dass die Forschungsdaten gemäß der guten wissenschaftlichen Praxis für mindestens 10 Jahre archiviert werden, ohne dass sie öffentlich gemacht werden. In diesem Fall können nur Mitarbeiter der UB den Datensatz einsehen, um ihn zu verwalten.

Lizenzen für Daten

Wenn Sie Ihre Forschungsdaten veröffentlichen oder fremde Forschungsdaten nachnutzen möchten, müssen die Nachnutzungsrechte geklärt sein. Oftmals werden publizierte Forschungsdaten mit Lizenzen versehen, die festlegen, unter welchen Bedingungen Forschungsdaten genutzt werden können, ohne den Urheber der Daten vorher um Erlaubnis fragen zu müssen. Zugleich erläutern sie, wie die Nachnutzung deklariert werden muss (Zitierhinweis). Etabliert haben sich dafür die [Creative Commons Lizenzen](#), die ab Version 4.0 auch für Forschungsdaten gelten.

Lizenz	Abkürzung	Beschreibung	Open-Access-konform
	CC0	Keine Rechte vorbehalten, gemeinfrei	✓
	CC BY	Namensnennung: uneingeschränkte Weiterverarbeitung und kommerzielle Nutzung erlaubt	✓
	CC BY-SA	Namensnennung + gleiche Lizenzvergabe (SA = share alike): uneingeschränkte	✓

		Weiterverarbeitung und kommerzielle Nutzung unter gleicher Lizenzvergabe	
	CC BY-NC	Namensnennung + nicht kommerziell: uneingeschränkte Weiternutzung aber nicht zu kommerziellen Zwecken ist erlaubt	x
	CC BY-NC-SA	Namensnennung + nicht kommerziell + gleiche Lizenzvergabe: uneingeschränkte Weitervergabe für nichtkommerzielle Zwecke ist unter der gleichen Lizenzvergabe gestattet	x
	CC BY-ND	Namensnennung + keine Bearbeitung: eine kommerzielle Nutzung ist gestattet, aber keine Bearbeitung	x
	CC BY-NC-ND	Namensnennung + nicht kommerziell + keine Bearbeitung: Weiterverarbeitung und kommerzielle Nutzung sind nicht gestattet	x

ACHTUNG: Eine einmal vergebene Lizenz kann nicht wirksam zurückgezogen werden!

Lizenzen für Software

Für erstellte Software gibt es eigene Open-Source-Software-Lizenzen. Gebräuchliche Lizenzen sind:

GPL: GNU General Public License

LGPL: GNU Lesser General Public License

MIT: MIT-Lizenz

Apache Lizenz

EUPL: European Union Public License

Ausführliche Auflistung von Open-Source-Lizenzen:

<https://ifross.github.io/ifrOSS/Lizenzcenter>

Hilfe bei der Wahl der passenden Lizenz:

<https://creativecommons.org/choose/?lang=de>

<https://choosealicense.com/>

5) Publizieren

Sind alle genannten Punkte bedacht und umgesetzt? Dann steht einer Archivierung bzw. Publikation Ihrer Forschungsdaten nichts mehr im Weg. Eingereicht werden die Daten entweder selbstständig per Shibboleth und der TU Braunschweig Kennung auf [LeoPARD](#) ([Anleitung](#) zur Selbsteinreichung) oder ein Mitarbeiter des Forschungsdaten-Teams (forschungsdaten@tu-braunschweig.de) unterstützt Sie gern dabei (z.B. bei größeren Datenmengen). In beiden Fällen unterzeichnen Sie zunächst das Antragsformular, welches Sie digital an das Forschungsdaten-Team (forschungsdaten@tu-braunschweig.de) schicken können.

Nach der Einreichung der Daten werden diese durch die Mitarbeiter der UB geprüft und wenn alles in Ordnung ist, werden die Forschungsdaten archiviert oder mit einer DOI versehen und veröffentlicht.