

Handreichungen für das Forschungsdatenmanagement: Abschluss und Publikation von Forschungsvorhaben

Abschlussphase

Aufbewahrung und Veröffentlichung

Die Technische Hochschule Wildau bekennt sich als Institution zu den „Leitlinien zur Sicherung der guten wissenschaftlichen Praxis“ (GWP) der DFG und nimmt in ihrer Leitlinie zum Umgang mit Forschungsdaten an der Technischen Hochschule Wildau¹ Bezug auf die GWP. In der Leitlinie 13 (der GWP) über die *Herstellung von öffentlichem Zugang zu Forschungsergebnissen* und der Leitlinie 17 über die *Archivierung* nimmt die GWP Bezug zum Umgang bei der Publikation von Forschungsdaten. Zu den wichtigsten Punkten gehören:

- Forschungsergebnisse und -daten, sowie dazugehörige zentrale Materialien, gegebenenfalls eingesetzte Forschungssoftware sollen in adäquater Weise aufbewahrt werden
- Orientierung soll entlang den Standards eines Fachgebiets erfolgen
- Daten sollen mindestens 10 Jahre vorgehalten werden
- Publikation sollte so offen wie möglich und so geschlossen wie nötig sein
- Können bestimmte Daten nicht vorgehalten werden, muss dies begründet sein (bspw. aus Datenschutzgründen)
- Bei der Publikation sind die FAIR-Prinzipien zu beachten

Datenselektion

Grundsätzlich gilt es, Daten dahingehend auszuwählen, ob sie für eine Veröffentlichung geeignet sind, die Daten also potenziell Relevanz für Dritte aufweisen oder ob die Daten archiviert werden. Hinweise zum Vorgehen bei der Datenselektion geben diese Handreichungen:

- <https://www.tu-braunschweig.de/forschung/forschungsdaten-transparenz/forschungsdaten/grundkurs->

¹ TH Wildau. Leitlinie zum Umgang mit Forschungsdaten an der Technischen Hochschule Wildau. Amtliche Mitteilungen 08/2025 vom 27.02.2025. https://www.th-wildau.de/files/2_Dokumente/Amtliche_Mitteilungen/08_2025_FDM_Leitlinie_THWildau.pdf

[forschungsdatenmanagement/datenorganisation-dokumentation/datenselektion](#)

- <https://www.dcc.ac.uk/sites/default/files/documents/publications/Five%20Steps%20to%20decide%20what%20data%20to%20keep.pdf>

PID

Ein Persistenter Identifikator (oder engl. Persistent Identifier, PID) ist eine eindeutige und dauerhafte (= persistente) digitale Referenz, die aus einer definierten Kombination von Ziffern und/oder alphanumerischen Zeichen besteht. Sie kann auf ein digitales Objekt (z. B. einen Datensatz), aber auch auf Personen und Organisationen verweisen. Zudem befinden sich PID für Software, Forschungsinstrumente, Datenmanagementpläne, Repositorien oder wissenschaftliche Konferenzen in Entwicklung. Im Gegensatz zu einer URL verweist ein PID nicht auf einen Standort, sondern auf das Objekt selbst, sodass es selbst dann noch auffindbar bleibt, wenn sich der Standort ändert.

Ein verbreiteter Standard für digitale Objekte ist [DOI](#) (Digital Object Identifier), ein weiterer ist [URN](#) (Uniform Resource Name). Für Personen hat sich [ORCID](#) (Open Researcher Contributor Identifier) etabliert, für Organisationen [ROR](#) (Research Organization Registry). Ein weiterer Standard ist die [GND](#) (Gemeinsame Normdatei), mit der Entitäten, also Personen, Körperschaften, Konferenzen, Geografika, Sachbegriffe und Werke repräsentiert werden können.

Weitere Informationen:

- <https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/persistente-identifikatoren/>
- <https://projects.tib.eu/pid-service/persistent-identifiers/persistent-identifiers-pids/>

Datenbanken

Definition: In einer Datenbank können Daten gesammelt und verwaltet werden. Sie eignen sich auch dafür, große Datenmengen strukturiert und langfristig zu speichern. Sogenannte Datenbankverwaltungssysteme (DMBS) ermöglichen dabei die Organisation von Daten und das Herstellen von Zusammenhängen, auch komplexe Abfragen sind möglich. Datenbanken sind dafür konzipiert, dass mehrere Benutzer gleichzeitig auf die Daten zugreifen können.

(Forschungs-)Datenbanken können urheberrechtlichen Schutz genießen: "Werden Forschungsdaten in einer Datenbank gesammelt, kann dies einen eigenen Urnehmerschutz begründen und sollte daher von Projektseite mitbedacht werden (vgl. §4 UrhG). Datenbanken unterliegen nach deutschem Recht einem spezifischen Schutz, der den Erstellerinnen oder Erstellern der Datenbank das alleinige Recht zu ihrer Verbreitung und Vervielfältigung gewährt (vgl. § 87b UrhG)." (Quelle: forschungsdaten.info, <https://forschungsdaten.info/themen/rechte-und-pflichten/urheberrecht/>)

Repositorien

Definition: Repositorien sind verwaltete Speicherorte zur Aufbewahrung digitaler Objekte, die meist öffentlich zugänglich sind. Es ist zwischen institutionellen, disziplinspezifischen und disziplinübergreifenden Repositorien zu unterscheiden.

Im Rahmen des vom BMBF und MWFK geförderten Projekts „IN-FDM-BB“ wird RADAR-BB als brandenburgweites, disziplinübergreifendes Repositorium für Forschungsdaten entwickelt, für die geeignete fachspezifische Repositorien nicht existieren. Somit wird RADAR-BB auch für die Technische Hochschule Wildau zur Verfügung stehen.

Weitere disziplinübergreifende Repositorien, die sich in der FDM-Community bewährt haben:

- [Zenodo](#)
- [Figshare](#)
- [Dryad](#)

Übersicht: Eine internationale Übersicht an Repositorien bietet der Dienst „Registry of Research Data Repositories“: <https://www.re3data.org/>

In Deutschland angesiedelte Repositorien sind beim Portal für Forschungsinfrastrukturen: „RIsources“ zu finden: <https://risources.dfg.de/>

Repositorium finden – Kriterien zur Auswahl

Bei der Auswahl eines geeigneten Repositoriums helfen die im Folgenden aufgeführten Qualitätsmerkmale, die sicherstellen sollen, dass die Forschungsdaten u. a. auffindbar, nachnutzbar, zitierbar und langfristig verfügbar sind:

- Die Datensätze erhalten einen persistenten Identifikator wie z. B. DOI (Digital Object Identifier) oder URN (Uniform Resource Name).

- Es ist möglich, den Datensatz mit Metadaten ausführlich zu beschreiben. Angaben zu Entstehung / Kontext des Datensatzes können in einem Textfeld oder in einer ReadMe-Datei gemacht werden.
- Die Datensätze werden durch Indexierung für Internet-Suchmaschinen öffentlich gemacht.
- Der Download des Datensatzes ist möglich.
- Das Repositorium erlaubt es, verschiedene Zugriffsoptionen auf den Datensatz einzustellen.
- Datensätze können mit einer Lizenz (z. B. Creative Commons) versehen werden.
- Idealerweise ist ein Zertifikat für vertrauenswürdige Archive vorhanden, z. B. CoreTrustSeal oder nestor-Siegel für vertrauenswürdige digitale Langzeitarchive.

Datenzentren

Definition: “Unter einem Datenzentrum ist ein zentrales Repositorium zu verstehen, das für eine bestimmte Wissens- und/oder Gedächtnisinstitution die Speicherung, Verwaltung und Verbreitung von Forschungsdaten und zugehörigen Informationen übernimmt.” (Quelle: [forschungsdaten.info](https://forschungsdaten.info/praxis-kompakt/glossar/#c269828), (<https://forschungsdaten.info/praxis-kompakt/glossar/#c269828>))

Übersicht: Eine umfangreiche Übersicht über vorhandene Datenzentren bietet RatSWD (<https://www.konsortswd.de/angebote/forschende/alle-datenzentren/>)

Datenjournale

Definition: Datenjournale sind Publikationen mit dem Hauptzweck, Datensätze zu veröffentlichen. Neben den Forschungsdaten wird eine ausführliche Beschreibung veröffentlicht, um deren Nachnutzung zu ermöglichen.

Übersicht fachübergreifende Datenjournale:

- Data Science Journal: <https://datascience.codata.org/>
- Data in Brief: <https://www.sciencedirect.com/journal/data-in-brief>
- (MPDI) Data: <https://www.mdpi.com/journal/data>
- F1000Research: <https://f1000research.com/>
- Scientific data: <https://www.nature.com/sdata/>
- Patterns: <https://datascience.codata.org/>

Weiterführende Informationen:

- https://www.forschungsdaten.org/index.php/Data_Journals
- <https://www.uni-wuerzburg.de/rdm/informationen/datenpublikation/>

Selbstentwickelte Software

Für die Veröffentlichung von Software sind neben den FAIR Prinzipien auch die FAIR Principles for Research Software zu beachten. Weiterhin sollte der Umgang mit eigens entwickelter Software entlang des Forschungsdatenlebenszyklus mit einem Softwaredatenmanagementplan (SMP) geplant werden.

FAIR Principles for Research Software: <https://zenodo.org/records/6623556>

Softwaredatenmanagementplan (SMP): <https://zenodo.org/records/1460504>

Bei der Publikation von Software können zwei verschiedene Wege gewählt werden, die mit Vor- und Nachteilen einhergehen. Weit in der Community verbreitet ist das zur Verfügung stellen von Software über Github. Dies hat den Vorteil, dass aufgesetzte Repositories immer wieder erweitert werden können. Jedoch entspricht dieser Weg in vielerlei Hinsicht nicht den FAIR-Prinzipien. Daher bietet es sich an, Software auch in einem Repository zu veröffentlichen. Damit geht der Vorteil einher, dass mit der Vergabe eines persistenten Identifikators (bspw. DOI) die Software zitierfähig und publiziert ist. Jedoch bildet die Publikation nur einen Zeitpunkt im Prozess der Softwareentwicklung ab und die Software sollte bereits einen hohen Reifegrad besitzen bzw. fertiggestellt sein, bevor man sie zur Publikation bringt.

Ein publizierter Datensatz sollte gut dokumentiert sein und folgende Elemente enthalten:

- Licence-txt
- ReadMe zum Copyright
- Software als Ordner mit allen notwendigen Dateien
- Manual oder auch technischer Report zur Anleitung der Software

Bei der Publikation von Software ist insbesondere auf die Vergabe einer entsprechenden Lizenz zu achten. Kriterien für die Auswahl richten sich nach den anvisierten Nutzungs- bzw. Nachnutzungsszenarien für die Software. Eine Übersicht zu Lizenzen bieten die folgenden Webseiten:

- <https://ifross.github.io/ifrOSS/Lizenzcenter#open-source-lizenzen>
- <https://www.tldrlegal.com/>

Literaturempfehlung:

- Biernacka, K. & Schulz, S. (2022). Forschungsdatenmanagement in der Informatik. Logos Verlag Berlin. <https://doi.org/10.30819/5490>

Datenschutzanforderungen und Datenzugang

Bei der Veröffentlichung der Daten sollte ein möglichst vollständiger Zugang zu den Daten angestrebt werden. Es gilt: so offen wie möglich, so geschlossen wie nötig. Dieser Grundsatz gilt gerade in Bezug auf Daten, bei denen die DSGVO Anwendung findet. Hier sind auch die Zugangs- und Nutzungsbedingungen der Daten zu klären. All diese Aspekte können bereits in Form eines DMPs vor einer Veröffentlichung geklärt werden.

Für die Veröffentlichung von Forschungsdaten, die im Zusammenhang mit personenbezogenen Daten stehen, sind einige Überlegungen zu treffen. Die folgenden Punkte sollten beachtet werden:

- Vorgaben der Drittmittelförderer und der eigenen Einrichtung hinsichtlich der Datenarchivierung sind zu prüfen und zu beachten.
- Prüfung, ob Daten gelöscht werden sollten.
- Daten sollten, wenn möglich, verfügbar gemacht werden.
- Es sollte geklärt werden, ob nur eine beschränkte Nachnutzung möglich ist, die sich aus den Eigenschaften der Daten (personenbezogene Daten) ergibt.
- Es sollte überlegt werden, ob Daten als Public Use File (PUF) ohne Restriktionen, oder als Scientific Use File (SUF) unter Restriktionen veröffentlicht werden sollten.
- Ort der Veröffentlichung ist zu klären (Repositorium oder Datenzentrum).
- Zeitpunkt der Veröffentlichung festlegen.

Weiterführende Informationen:

- RatSWD [Rat für Sozial- und Wirtschaftsdaten] (2020): Handreichung Datenschutz, 2. vollständig überarbeitete Auflage, RatSWD Output 8 (6), <https://doi.org/10.17620/02671.50>
- Kienbaum, Janna; Fischer, Patryk & Paßmann, Sven (2023): Forschungsdatenmanagement bei personenbezogenen Daten - eine Handreichung, <https://doi.org/10.5281/zenodo.7428524>

Lizenzen

Eine Lizenz ist ein vertraglich vereinbartes Nutzungsrecht. Mit ihr legt die rechteinhabende Person fest, ob, wie und unter welchen Bedingungen ein Werk weiter- oder nachgenutzt werden kann. Es gibt kommerzielle Lizenzen, die für die Nutzung

eine Gebühr verlangen und freie Lizenzen, die eine unentgeltliche Nutzung ermöglichen.

Freie Lizenzen erlauben die Nutzung eines urheberrechtlich geschützten Inhalts für jedermann. Allerdings kann es Einschränkungen hinsichtlich der Verwendung oder Bearbeitung geben. Es haben sich standardisierte Lizenzen herausgebildet, die die häufig vorkommenden Bedingungen abbilden. Üblich sind:

- [Creative Commons \(CC\)](#), die für Texte, Abbildungen und Daten geeignet sind
- [GNU General Public License \(GPL\)](#), die für Software konzipiert wurde
- [Open Data Commons \(ODC\)](#), die für Datenbanken eingerichtet wurden sowie
- [Community Data License Agreement \(CDLA\)](#), das für Daten konzipiert wurde.

Die Verwendung einer freien Lizenz ist unbedingt zu empfehlen, weil sie die Nutzung von Daten erheblich vereinfacht, denn damit fällt der Aufwand weg, in jedem Einzelfall eine Gestattung der rechteinhabenden Person einzuholen. Eine freie Lizenz lässt sich wie eine generalisierte Erlaubnis verstehen, mit der der*die Urheber*in den Nutzenden mitteilt, dass das Werk unter diesen Bedingungen in dieser Art verwendet werden darf.

Lizenzen werden in der Regel über ein Repositorium vergeben. Es gilt zu bedenken, dass eine einmal vergebene Lizenz nicht zurückgezogen werden kann. Außerdem sind die Vorgaben der eigenen Institution und der Förderer hinsichtlich der FAIR-Prinzipien und Open-Access-Richtlinien zu beachten.

Bei der Nachnutzung von Forschungsdaten ist auf die jeweiligen Lizenzvorgaben dieser Daten zu achten. Diese sind bindend, weshalb eine potentielle Nutzung vorab mit dem*der Urheber*in geklärt werden muss.

Weitere Informationen:

- <https://forschungsdaten.info/themen/rechte-und-pflichten/forschungsdaten-veroeffentlichen/>
- https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/Infomaterialien/IN_Freie_Lizenz_en_-_einfach_erklaert_Broschuere_2017_07_20.pdf

Ansprechpersonen

Beratung zum Forschungsdatenmanagement und Datenmanagementplänen

Blanka Goßner

Forschungsdatenmanagement | Zentrum für Forschung und Transfer

Technische Hochschule Wildau | Hochschulring 1 | 15745 Wildau

Haus 13 | Raum 0.41

Telefon: +49 3375 508 322

Mail: blanka.gossner@th-wildau.de

Beratung bei der Beantragung drittmittelgeförderter Forschungsprojekte

Zentrum für Forschung und Transfer

Mail: forschung@th-wildau.de