



## Leitfaden zur Publikation von Forschungsdaten

Eine Empfehlung

Forschungsdatenmanagement

Dr. Max Hassenstein  
forschungsdaten@tiho-hannover.de

### 1. Vorbemerkung

Mit Forschungsdaten sind konkret alle Daten und Informationen gemeint, welche im Rahmen des Forschungsprozesses zur Beantwortung der Forschungsfrage angefallen. Typische Forschungsdaten sind u. a. Laborwerte, Video- und Tonaufzeichnungen, Mess- oder Beobachtungsdaten, Befragungsergebnisse oder Mitschriften und Texte. Verworfenen Daten, die rein aus Testzwecken generiert wurden, sind nicht als Forschungsdaten anzusehen.

Die Leitlinien zur Sicherung guter wissenschaftlicher Praxis sehen vor, dass Forschende ihre Forschungsdaten öffentlich zugänglich machen<sup>1</sup>. Dies fördert zum einen Transparenz und Reproduzierbarkeit und somit auch den wissenschaftlichen Diskurs, wie auch die Möglichkeit, unabhängige Studiendaten für Metaanalysen zu kombinieren. Dementsprechend verpflichtet sich die Stiftung Tierärztliche Hochschule Hannover (TiHo) gemäß ihrer Forschungsdaten-Policy<sup>2</sup> diesen Prinzipien und erwartet in diesem Rahmen von ihren Forschenden und Studierenden eine Publikation ihrer Forschungsdaten.

Neben den Aspekten der Transparenz und Reproduzierbarkeit stellt eine Publikation der Forschungsdaten die langfristige Sicherung und Verfügbarkeit sicher. Dritte können einfacher an Forschungsergebnisse anknüpfen und bestehende Daten ressourcenschonend nachnutzen, was insbesondere in mit öffentlichen Mitteln geförderte Forschungsvorhaben geboten ist. Ein Vorteil der Veröffentlichung von Forschungsdaten für die Datenschaffenden ist die Erhöhung der Sichtbarkeit durch zusätzliche Zitationen<sup>3</sup>.

Das vorliegende Dokument dient als Orientierung für wissenschaftlich tätige Personen an der TiHo in Bezug auf die Publikation von Forschungsdaten. Dabei werden die typischen Schritte der Datenpublikation aufgegriffen und Empfehlungen dargelegt.

### 2. Schritte der Datenpublikation

Die Publikation von Daten kann in mehrere Schritte unterteilt werden (Abb. 1).

---

<sup>1</sup> Deutsche Forschungsgemeinschaft (2022) Guidelines for Safeguarding Good Research Practice. Code of Conduct. <https://doi.org/10.5281/zenodo.6472827>

<sup>2</sup> <https://www.tiho-hannover.de/forschung/gute-wissenschaftliche-praxis/forschungsdatenmanagement/forschungsdatenpolicy>

<sup>3</sup> Colavizza G et al. (2020) The citation advantage of linking publications to research data. PLoS One;15(4):e0230416. <https://doi.org/10.1371/journal.pone.0230416>.



Abbildung 1: Schema zum Prozess der Datenpublikation.

### Schritt 1: Überblick verschaffen

Allen voran ist im Kontext der Datenpublikation wichtig, einen Überblick über sämtliche Forschungsdaten eines Projekts zu erlangen. Dabei sollten alle Daten des Forschungsprojekts herangezogen werden, die im Rahmen des Forschungsprozesses entweder entstanden, modifiziert oder nachgenutzt wurden.

Im Allgemeinen ist es ratsam, von Projektbeginn an in einem zentral abgelegten Dokument festzuhalten, welche Forschungsdaten entstanden sind und wo diese gespeichert sind – insbesondere, wenn die Daten aus projektbezogenen Gründen nicht zentral gespeichert werden können, z. B. falls diese bei Kooperationspartnern verbleiben müssen oder in Laborinformationssystemen zwischengespeichert sind. Sollte ein solches Dokument nicht vorliegen, sollte es nach Möglichkeit nachträglich angelegt werden.

Selbst Forschungsdaten, die sich der ersten Einschätzung nach nicht zur Publikation eignen, sollten dennoch aufgelistet und erst im Rahmen der späteren Datenselektion in gemeinsamer Entscheidung mit dem Projektteam ggf. aussortiert werden. Im Fall der Aussortierung der Forschungsdaten sollte die Begründung unmittelbar und nachvollziehbar dokumentiert werden. Dieser Schritt ist wichtig, da bestimmte Förderer eine Begründung für zurückgehaltene Daten fordern können.

### Schritt 2: Vorgaben prüfen

Anhand der zuvor erarbeiteten Übersicht aller im Projekt angefallenen Forschungsdaten sollten geltende Vorgaben in Bezug auf die Datenpublikation geprüft und berücksichtigt werden.

Zunächst sollte, sofern vorliegend, der Datenmanagementplan (DMP) des Forschungsprojekts beachtet werden. Dieser enthält üblicherweise eigens festgelegte Leitplanken, Informationen und das zuvor angestrebte Vorgehen hinsichtlich der Datenpublikation.

Konkrete Forderungen Dritter hinsichtlich der Forschungsdatenpublikation bzw. spezifischen Anforderung an diese können von Seiten der Förderer festgeschrieben worden sein, z. B. im Rahmen vertraglicher Regelungen. Die Deutsche Forschungsgemeinschaft (DFG) sieht bspw. vor, dass im Rahmen ihrer geförderten Forschungsvorhaben entstandene Forschungsdaten öffentlich zugänglich gemacht werden. Zu beachten sind auch mögliche Journal- und Verlagspolicies, die im Rahmen der Open Science eine Datenpublikation erwarten und oftmals eine Begründung erwartet wird, falls Daten nicht geteilt werden. Des Weiteren können fachspezifische Empfehlungen der Fachgesellschaften bzw. Fachcommunities zum Tragen kommen. Auch Ethikkommissionen können Empfehlungen aussprechen, z. B. wenn Forschungsdaten sensible Informationen wie Standorte schützenswerter Arten enthalten. Schließlich können auch Datenrepositorien Vorgaben hinsichtlich der hochzuladenden Daten machen, daher wird empfohlen, sich grundsätzlich mit den typischen Anforderungen, die sich üblicherweise auf Dokumentation bzw. Metadaten und Qualitätsaspekte bezieht, vertraut zu machen. Im Zusammenhang mit der

Dokumentation von Forschungsdaten ist die Bezeichnung Metadaten („Daten über Daten“) ge-  
läufig. Diese Bezeichnung steht oftmals im engeren Sinn synonym für computerleserlich struk-  
turierte Informationen (z. B. XML, JSON) über die Daten wie Messzeitpunkt, Messgerät, Mess-  
ort, Autor, Beschreibungen zur Organisation der Daten.

Forschungsdaten sollen u. a. nur dann veröffentlicht werden, wenn die Rechte Dritter nicht ver-  
letzt werden. Besonders nennenswerte rechtliche Aspekte im Forschungsdatenmanagement  
sind das Urheberrecht bei Daten mit schöpferischer Höhe und der Datenschutz im Falle von  
personenbezogenen Daten. Falls Dritte ein Urheberrecht an den (nach-) genutzten Daten ha-  
ben, oder falls Geheimhaltungsabreden, z. B. festgelegt in Verträgen, bestehen, muss eine Er-  
laubnis eingeholt werden. Für die Veröffentlichung von Daten, die dem Datenschutz unterliegen,  
muss die eindeutige Einwilligung der Betroffenen vorliegen. Ein nützliches Flussdiagramm mit  
den wichtigsten rechtlichen Aspekte wird auf dem Informationsportal [forschungsdaten.info](https://forschungsdaten.info) be-  
reitetgestellt<sup>4</sup>.

### Schritt 3: Datenselektion

Die Entscheidung darüber, ob und welche Forschungsdaten publiziert werden, sollte nicht von  
subjektiven Entscheidungen abhängen. Im Vordergrund sollten stets Überlegungen zur Eignung  
der Daten zum Nachweis der Forschungsergebnisse sowie dessen Nachnutzungspotenzial für  
die wissenschaftliche Gemeinschaft stehen. Des Weiteren müssen geltende Vorgaben an eine  
Datenpublikation (s. Schritt 2) beachtet werden.

Sollten bspw. im Rahmen eines Forschungsprojekts umfassende Forschungsdaten generiert  
worden sein und alle angestrebten wissenschaftlichen Fachartikel gemäß Projektantrag publi-  
ziert worden sein, sollten keine Daten aus strategischen Gesichtspunkten wie dem Offenhalten  
von Zusatzpublikationen zurückgehalten werden. Gerade wenn die Forschungsdaten als wertvoll  
erachtet werden, sollte umso mehr eine Datenpublikation angestrebt werden, zumal entspre-  
chende Würdigungen der Datenschaffenden durch Lizenzen ermöglicht werden (s. Schritt 5).

Neben den bestehenden Vorgaben zur Datenpublikation (s. Schritt 2) können Forschende er-  
gänzende Kriterien zur Entscheidungsfindung hinsichtlich der Forschungsdatenpublikation her-  
anziehen (Tabelle 1).

Tabelle 1: *Kriterien für und gegen die Publikation von Forschungsdaten*

Kriterien, die für eine Publikation sprechen	Kriterien, die gegen eine Publikation sprechen
+ Einzigartigkeit der Daten	– Testdaten
+ Aufwendigkeit der Datenerzeugung	– Verworfenen Daten und Fehlmessungen
+ Relevanz der Daten für wissenschaftliche Ge- meinschaft oder die Allgemeinheit	– Mangelnde mittel- und langfristige Relevanz
+ Nutzen dem Nachweis der Forschungsergebnisse	– Nicht dokumentierte/unzureichend beschriebene Daten
+ Qualität der Daten	– Ethische Aspekte wie Datenschutz oder beson- ders schützenswerte Daten wie z. B. Orte gefähr- deter Arten
+ Ausreichend dokumentierte Daten und Metada- ten	– Rechtliche Bedenken wie ungeklärtes Urheber- recht
	– Nachgenutzte, bereits publizierte Daten

<sup>4</sup> Forschungsdaten publizieren? Die wichtigsten rechtlichen Aspekte. <https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/daten-publizieren/>

Wichtige Kriterien von publikationswürdigen Forschungsdaten sind, dass diese z. B. einzigartig und von Interesse für die wissenschaftliche Gemeinschaft bzw. der Öffentlichkeit sind. Des Weiteren verfügen relevante Forschungsdaten zwingend über eine ausreichende Datenqualität sowie Dokumentation inkl. Metadaten. Nicht publiziert werden sollten u. a. Testdaten/Testmessungen und verworfene Daten sowie Daten von mangelnder Qualität und/oder ohne Dokumentation.

Sollten Abwägungen ergeben, dass sämtliche bzw. ein Großteil der Forschungsdaten den Anforderungen einer Forschungsdatenpublikation nicht entsprechen und somit nicht publiziert werden können, dann sollte eine Begründung verfasst und leicht auffindbar abgespeichert werden, um die Entscheidung langfristig nachvollziehen zu können. Sofern die Forschungsdaten nicht veröffentlicht werden können, aber eine Datenpublikation vertraglich oder durch Förderer festgeschrieben ist, sollte umgehend Kontakt mit entsprechender Kontaktstelle aufgenommen werden, um aktiv den Konflikt zu lösen.

Die Datenqualität ist von hoher Bedeutung für unverfälschte wissenschaftliche Ergebnisse. Auch bei der Forschungsdatenpublikation ist auf die Datenqualität zu achten. Anhand mehrerer Dimensionen lässt sich die Datenqualität charakterisieren und bewerten<sup>5</sup>. Wichtige Datenqualitätsdimensionen sind u. a. die Genauigkeit (Accuracy) und Vollständigkeit (Completeness). Es ist empfehlenswert, vor der Datenpublikation die wichtigsten bzw. relevanten Datenqualitätsdimensionen zu überprüfen und die Erfüllung dieser sicherzustellen.

#### Schritt 4: Daten und Dokumentation vorbereiten

Nach der Datenselektion muss eruiert werden, welche der Daten tatsächlich publiziert werden können. Die allgemeine Maßgabe entsprechend der Leitlinien ist, so offen und transparent wie möglich zu gestalten, um die Nachvollziehbarkeit der Forschung zu gewährleisten. Allerdings ist eine Dokumentation der Daten zur Nachvollziehbarkeit, auch unabhängig einer Publikation, zwingend erforderlich. Daher sollten originale Forschungsdaten auch nicht ohne Dokumentation veröffentlicht werden.

Die FAIR-Prinzipien<sup>6</sup> gelten als allgemeine Leitprinzipien des Forschungsdatenmanagements und sollten sich insbesondere im Zuge der Datenpublikation widerspiegeln. Das FAIR-Akronym steht für **F**indable/auffindbar, **A**ccessible/zugänglich, **I**nteroperable/interoperabel, **R**esuable/wiederverwendbar. Grob lassen sich diese Prinzipien wie folgt zusammenfassen:

F) Die (Meta-)Daten sind mit einer einzigartigen und dauerhaften Kennung (z. B. einer DOI<sup>7</sup>) versehen, werden von umfassenden Metadaten begleitet und die (Meta-)Daten sind indexiert bzw. registriert, um eine Auffindbarkeit zu gewährleisten.

A) Die (Meta-)Daten sind zugänglich, wenn sie anhand einer dauerhaften Kennung unter der Verwendung eines quelloffenen, standardisierten Kommunikationsprotokolls (z. B.

---

<sup>5</sup> Hassenstein, M., Vanella, P. (2022) Data Quality—Concepts and Problems. Encyclopedia 2, no. 1: 498-510. <https://doi.org/10.3390/encyclopedia2010032>

<sup>6</sup> Wilkinson MD et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016 Mar 15;3:160018. doi: 10.1038/sdata.2016.18.

<sup>7</sup> DOI = digital object identifier. Eine DOI kann üblicherweise durch ein Repositorium erzeugt werden, welche dann auch in die Metadaten eingepflegt werden sollte.

HTTP) wiederauffindbar sind. Dies ist grundsätzlich der Fall, sofern das vorgesehene Repositorium über den Browser erreicht werden kann.

I) Zur Gewährleistung der Interoperabilität müssen die (Meta-)Daten in einem Standard(-format) bereitgestellt werden, bspw. in XML oder CSV.

R) Zwecks Wiederverwendbarkeit müssen die (Meta-)Daten genau beschrieben werden, einschließlich ihrer Herkunft bzw. Erzeugung, und durch die Angabe von Lizenzen und zugehörigen Veröffentlichungen begleitet werden. Die gängigen Standards der Fachbereiche sollten hier berücksichtigt werden.

Forschungsdaten lassen sich auf vielfältige Weise beschreiben und dokumentieren. Nennenswerte, typische Dokumentationslösungen sind das (elektronische) Laborbuch, das Data Dictionary bzw. Codebook, oder die readme-Datei oder ähnliche textbasierte Dokumente zur Datenbeschreibung. Mit wachsendem Umfang der Daten steigt auch die Dringlichkeit einer readme-Datei. Hier kann ein readme-File im txt-Format sinnvoll sein, bei sehr vielen Daten kann ein längeres Dokument oder gar eine Datensatzbeschreibung als alleinstehende Publikation als „Data Paper“ in einem Datenjournal sinnvoll sein. Eine typische readme-Datei enthält u. a. Informationen wie Titel, Kurzbeschreibung, Datenkuratoren, Methodik, Datumsangaben, benötigte Ressourcen.

Im Rahmen der Datenpublikation ist darauf zu achten, dass gängige und möglichst frei verwendbare bzw. nicht-proprietäre Datenformate verwendet werden. So soll z. B. ein Dokument im PDF-Format anstelle des MS Word-Formats (.docx) bereitgestellt werden, oder eine Datentabelle als CSV anstelle einer MS Excel-Tabelle (.xlsx) (Tabelle 2).

Tabelle 2: *Empfohlene Dateiformate zur Publikation von Forschungsdaten*

Dateityp	Empfohlenes Format
<b>Datentabellen</b>	CSV, TXT
<b>Sequenzdaten</b>	FASTA, FASTQ
<b>Textdokumente</b>	PDF, XML
<b>Bilddateien</b>	PNG, JPEG, TIFF, SVG (Vektorgraphiken)
<b>Videodateien</b>	MP4, AVI, MOV
<b>Audiodateien</b>	WAV, FLAC, MP3 (verlustbehaftetes, aber speicherschonendes Format)

Alle anfallenden und zugehörigen Daten sollten im Rahmen der Vorbereitung zur Publikation sinnvoll benannt und strukturiert werden. Bei zunehmender Menge an Einzeldateien wird es ggf. erforderlich, Forschungsdaten in eine Ordnerstruktur -ähnlich wie im lokalen Arbeitsverzeichnis- zu überführen. Die TiHo stellt Empfehlungen zur Dateibenennung und Ordnerstruktur bereit, welche sich auf reine Forschungsdatenverzeichnisse übertragen lassen und somit für eine nachvollziehbare Ordnerstruktur gesorgt werden kann.

### Schritt 5: Publikation

Bei der Wahl einer geeigneten Plattform zur Forschungsdatenpublikation sollten neben möglichen Vorgaben (s. Schritt 2) Überlegungen zur Zielgruppe herangezogen werden: Während

sich fachspezifische Repositorien (z. B. gfbio<sup>8</sup> u. a. für Biodiversitätsdaten) zur Erreichung einer bestimmten Forschungsgemeinschaft eignen, kann die strategische Platzierung von Forschungsdaten in allgemein ausgerichteten Repositorien (z. B. Zenodo<sup>9</sup>) sinnvoll sein, wenn bspw. die Forschungsdaten über den eigenen Wissenschaftsbereich hinaus von Relevanz sind.

Um geeignete, fachspezifische Repositories ausfindig zu machen, kann das Register re3data.org<sup>10</sup> verwendet werden, welches vielzählige Filterungs- und Sortierungsoptionen bietet- somit kann gezielt nach bestimmten Kriterien und Anforderungen gesucht werden (z. B. Zertifizierungen, Qualitätskontrollen, Metadatenstandards).

Bei der Wahl des Repositoriums sollte auf die Dauer der Datenbereitstellung geachtet werden, um möglichen Anforderungen wie der guten wissenschaftlichen Praxis (s. Schritt 2) gerecht zu werden. Auch das Setzen eines Embargodatums wird von vielen Repositorien unterstützt. Dies kann sehr nützlich sein, wenn die Forschungsdaten im Vorfeld zur allgemeinen Veröffentlichung einem kleineren Personenkreis zu Verfügung gestellt werden sollen, wie es bspw. ein Peer-Review-Verfahren eines Fachartikels erfordern könnte.

Schließlich ist im letzten Schritt die Festlegung einer Lizenz wichtig, um die Möglichkeiten der Nachnutzung publizierter Forschungsdaten zu steuern und an Dritte zu kommunizieren. Die Lizenzwahl hängt vorwiegend von den geltenden Vorgaben ab. Anhand von Lizenzen u. a. vorgegeben werden, dass die geteilten Daten nur unter der Namensnennung der Datenschaffenden genutzt werden dürfen und/oder dies ausschließlich zu nichtkommerziellen Zwecken geschieht. Grundsätzlich wird, möglichst freizügige Lizenzen zu verwenden (Abb. 2). Im Rahmen der Forschungsdatenpublikation sind die Creative Commons<sup>11</sup>-Lizenzen gängig und den gängigen Datenpublikationsplattformen auswählbar.

Es ist wichtig zu beachten, dass einmal vergebene Lizenzen nicht zurückgezogen werden können, d. h. die Vergabe einer restriktiveren Lizenz ist üblicherweise im Nachgang nicht mehr möglich!

---

<sup>8</sup> <https://www.gfbio.org/>

<sup>9</sup> <https://zenodo.org/>

<sup>10</sup> <https://www.re3data.org/>

<sup>11</sup> <https://creativecommons.org/>

