

Definition

Repositorien für Forschungsdaten dienen dazu, digitale Daten aus der eigenen Forschung zu veröffentlichen oder zu archivieren und sie für Interessierte zugänglich zu machen. Die Motivation hierfür ist, dass Forschungsergebnisse aus Publikationen¹ mit Hilfe der veröffentlichten Forschungsdaten nachvollziehbar und reproduzierbar werden. Es ist auch möglich einen Datensatz ohne dazugehörige Publikation zu veröffentlichen, damit möglichst viele andere Forscher diesen für ihre eigenen Forschungsfragen verwenden können.

Die Veröffentlichung der Forschungsdaten eines Projekts wird bereits von vielen Drittmittelgebern verlangt bzw. empfohlen und zählt zur guten wissenschaftlichen Praxis². Bei der Auswahl eines Datenrepositoriums sollte auf verschiedene Qualitätsmerkmale geachtet werden, die zum Verständnis und der Sichtbarkeit des Datensatzes positiv beitragen können.

Qualitätsmerkmale für Repositorien



Persistente Identifikatoren (PID) für Datensätze: Um einen Datensatz möglichst eindeutig referenzieren zu können, benötigt er einen Identifikator, der auf die entsprechende Ressource im Internet verweist. Hierzu wird von dem Repository (selten auch von der eigenen Einrichtung oder Verlagen) für den Datensatz in der Regel ein DOI (Digital Object Identifier) vergeben. Andere PID-Systeme, die ebenfalls fachbezogen weite Verbreitung finden, sind URN (Uniform Resource Name) oder hdl (Handle-System).



Persistente Identifikatoren für Autoren und Autorinnen: Bei der Angabe der Autoren und Autorinnen ist es ebenfalls wichtig diese mit einem eindeutigen Identifikator versehen zu können. Dieser bleiben selbst dann bestehen, wenn die Forschenden ihre Namen ändern oder ihren Arbeitsort wechseln und ist deshalb wichtig, um den Datensatz dauerhaft mit den entsprechenden Personen zu verknüpfen. Gängige Identifikatoren für wissenschaftliche Autoren und Autorinnen sind beispielsweise ORCID und ResearcherID.



Metadaten: Um Metadaten dienen dazu den Datensatz möglichst ausführlich zu beschreiben. Dazu gehören unter anderem Angaben wie z.B.: Autoren, Erhebungs- und Veröffentlichungsdatum, Organisation/Institut, Dateiformat, Schlüsselwörter oder Fachbereiche. Je nach Fachrichtung können auch spezifischere Angaben üblich sein, wie zum Beispiel Geolokalisation.



Download- und Exportmöglichkeiten: Das Repository sollte die Möglichkeit bieten, den Datensatz oder auch nur Teile davon herunterzuladen zu können. Weiterhin sollten auch verschiedene Exportmöglichkeiten zum Zitieren existieren, wie z.B. bib für Latex oder JSON-LD für Webseiten, um den Datensatz leicht in verschiedenen Systemen referenzieren zu können. Auch die Anzeige des Zitats des Datensatzes in verschiedenen Stilen als Klartext kann hilfreich sein.



Beschreibung bzw. Dokumentation: Es sollte die Möglichkeit bestehen, in einem Textfeld die Entstehung und den Kontext des Datensatzes zu beschreiben. Dies kann aber auch mit einer sogenannten Readme-Datei im Datensatz selber umgesetzt werden.



Zugriffsmöglichkeiten: Diverse Zugriffsoptionen geben die Möglichkeit, den Zugriff auf den Datensatz zu beschränken. Der Datensatz kann entweder frei zugänglich, eingeschränkt nur für eine bestimmte Benutzergruppe, auf Anfrage oder erst nach einer bestimmten Embargofrist frei erhältlich sein.

¹ Mit Publikationen sind wissenschaftliche Werke in Schriftform gemeint, wie Artikel oder Bücher.

² Kodex der DFG: https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf

Repositorien für Forschungsdaten



Lizenzen: Die Lizenz eines Datensatzes definiert, wie dieser weiterverwendet werden darf. Durch die Auswahl einer passenden Lizenz können Einschränkungen bezüglich Namensnennung, Monetarisierung, Veränderung oder auch Wiederveröffentlichung festgelegt werden. Standardlizenzen werden zum Beispiel von CC (Creative Commons), ODC (Open Data Commons) und GPL (GNU General Public License) angeboten.



Überblick/Vorschau des Datensatzes: Da viele Benutzer sofort Informationen oder eine Übersicht in Form von Visualisierungen zu einem Datensatz wünschen, ist es sinnvoll, wenn das Repository bereits Ordnerstrukturen, Bild- oder PDF-Viewer anbietet, ohne den Datensatz zuerst herunterladen zu müssen.



Versionierung: Es kann immer wieder vorkommen, dass ein Datensatz sich verändert oder aktualisiert werden muss. In diesem Fall ist es sinnvoll, wenn man den Datensatz als neue Version anlegen kann, anstatt ihn als neuen Datensatz eintragen zu müssen.



Anmeldung und Bearbeitung: Die Registrierung und das Anmelden sollten einem so leicht gemacht werden wie möglich, zum Beispiel über die ORCID des Autors oder der Autorin. Nach der Veröffentlichung sollte der Autor oder die Autorin weiterhin die Möglichkeit haben den Datensatz, insbesondere die Metadaten, jederzeit bearbeiten zu können, ohne das Support-Desk kontaktieren zu müssen.



Auffindung durch Suchmaschinen: Nicht immer werden die Datensätze eines Repositoriums durch Indexierung für Internet-Suchmaschinen öffentlich gemacht. Indem man einen Beispieldatensatz per Titel, Beschreibung oder Schlüsselwörtern in gängigen Suchmaschinen sucht, kann man herausfinden, ob dieser dort indiziert und damit auffindbar ist.

Die aufgeführten Qualitätsmerkmale sind nicht immer auf den ersten Blick ersichtlich. Oft hilft es aber, sich ein paar Beispieldatensätze auf dem Repository anzuschauen, um einen ersten Eindruck zu bekommen.

Die **FAIR-Prinzipien** geben Grundsätze vor, um Forschungsdaten nachhaltig nutzbar zu machen. Forschungsdatenrepositorien helfen dabei einen Teil dieser Prinzipien zu erfüllen. So werden zum Beispiel hochgeladene Datensätze bereits automatisch mit einem persistenten Identifikator versehen und wichtige Metadaten müssen vor der Veröffentlichung ausgefüllt werden. Dennoch müssen die Dokumentation und das Einhalten von fachspezifischen Standards vom Forschenden selbst übernommen werden.



Angebote und Services

Im Folgenden möchten wir einige empfehlenswerte Angebote und Services präsentieren. Als erste Anlaufstelle für die Suche nach einem geeigneten fachbezogenen Repository kann re3data dienen.

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Re3data.org listet über 2000 Repositorien aus verschiedenen Fachgebieten. Diese werden dort zusammen mit Informationen zur Einhaltung wichtiger Qualitätsmerkmale präsentiert. Mehrere deutsche Institute und Förderer unterstützen diese Plattform bereits, damit Forschende sie als unterstützendes Angebot bei der Suche nach geeigneten Repositorien verwenden können.



Fachübergreifende Repositorien für Forschungsdaten

Konnte man über das re3data Verzeichnis kein geeignetes fachbezogenes Repository finden, so stehen einem generelle Forschungsdatenrepositorien als Alternative zur Verfügung.

Alle der folgenden Repositorien bieten ihren Service grundlegend kostenfrei an, d.h. Datensätze können kostenlos hochgeladen und öffentlich gemacht werden. Bei einigen der Angebote existieren jedoch Einschränkungen bezüglich der möglichen Anzahl der Dateien oder der Größe eines Datensatzes.



Zenodo ist ein disziplinübergreifendes Repository für jede Art von wissenschaftlichen Datensätzen, wie z.B. auch für Publikationen, Quellcode oder Präsentationen. Es besitzt unter anderem eine Integration von *GitHub* und Social Media Statistiken. Die Plattform wird von der *Europäischen Kommission* finanziert, ist aus dem Projekt *OpenAIRE* entstanden und wird am *CERN* gehostet.

Link: zenodo.org



Das *EUDAT* Projekt bietet mit *B2Share* eine Plattform, um Daten interdisziplinär zugänglich zu machen und zu erhalten. Mit *B2NOTE*, *B2SAFE*, *B2HANDLE*, *B2DROP*, *B2FIND*, *B2STAGE*, *B2ACCESS* und *easyDMP* werden weitere Services geboten, um den Umgang mit Forschungsdaten zu vereinfachen. Die Plattform ist aus dem *Horizon 2020* Förderungsprogramm der Europäischen Union entstanden.

Link: b2share.eudat.eu



Das Repository *OSF Home* des *Open Science Frameworks* ist Open Source und soll den Wissenschaftlern und Wissenschaftlerinnen helfen, ihre Forschungsprozesse möglichst offen und transparent zu gestalten. Mit *OSF Preprints*, *Meetings*, *Institutions*, *Registries* und *Collections* werden weitere Services geboten, um dieses Vorgehen zu unterstützen. Das Angebot wurde vom *Center for Open Science* entwickelt und wird in den USA gehostet.

Link: osf.io



Dryad ist eine us-amerikanische, disziplinübergreifende Lösung für die Veröffentlichung aller Arten von Forschungsdaten. Die Open Source Plattform wird von verschiedenen amerikanischen Instituten und Förderern unterstützt. Seit 2019 wird *Dryad* von der *California Digital Library* gehostet. In Kooperation mit *Zenodo* sollen neue Services entwickelt werden.

Link: datadryad.org



Figshare ist ein Repository für Forschungsdaten, welches besonders für visuelle Grafikdatensätze wie Poster, Diagramme oder Videos ausgelegt ist. Die Plattform wird von der *Nature Publishing Group* angeboten, welche Teil der Verlagsgruppe *Springer Nature* und in Großbritannien ansässig ist.

Link: figshare.com



Die *Digitale Bibliothek Thüringen* ist eine lokale Lösung für die Veröffentlichung von Forschungsdaten. Sie bietet Forschenden Thüringer Hochschulen neben der Open-Access-Veröffentlichung von Publikationen auch die Publikation von Datensätzen an.

Link: db-thueringen.de

Wenn Sie bei der Suche nach einem geeigneten Repository Hilfe benötigen oder offene Fragen bezüglich der Organisation und Aufbereitung Ihrer Daten haben, hilft Ihnen gerne das Thüringer Kompetenznetzwerk Forschungsdatenmanagement! Kontaktieren Sie uns unter: <https://forschungsdaten-thueringen.de/kontakt.html>