



UNIVERSITÄT
LEIPZIG

CRASHKURS FORSCHUNGSDATENMANAGEMENT

AG Forschungsdatenmanagement
Onlineversion, April 2020

WAS SIE IN DIESEM KURS ERWARTET

Am Ende dieses Kurses wissen Sie,

... was Forschungsdaten sind.

... was Forschungsdatenmanagement ausmacht.

... wie Sie Ihre Daten organisieren.

... wie Sie Forschungsdaten suchen, nachnutzen und selbst veröffentlichen können.

... welche rechtlichen Vorgaben und Vorgaben der Forschungsförderer Sie beachten müssen.

Wir wünschen Ihnen viel Erfolg mit unseren Materialien und sind dankbar für Feedback an forschungsdaten@uni-leipzig.de!

Ihr Team der AG Forschungsdaten

KURSIINHALT

1. Grundlagen – Begriffe und Konzepte
2. Datendokumentation und Metadaten
3. Datenspeicherung und -organisation
4. Langzeitarchivierung
5. Datenveröffentlichung
6. Rechtliche und ethische Aspekte
7. Datennachnutzung
8. Datenmanagementpläne

1. GRUNDLAGEN – BEGRIFFE UND KONZEPTE

FORSCHUNGSDATEN (FD)

... sind alle Daten, die während eines wissenschaftlichen Forschungsprozesses entstehen, verarbeitet werden oder dessen Ergebnis sind.

Aufgrund der Fächervielfalt sind FD sehr heterogen.

Das können z. B. sein: Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten, Objekte aus Sammlungen oder Proben, methodische Testverfahren (Fragebögen, Interviews), Transkripte, Beobachtungen, Software, Code, Simulationen.

Quellen:

[Kindling, Maxi; Schirnbacher, Peter \(2013\). „Die digitale Forschungswelt“ als Gegenstand der Forschung, in: Information – Wissenschaft & Praxis Band 64, Heft 2–3: S. 27–136. DOI: 10.1515/iwp-2013-0017.](#)

[Universität Leipzig \(2019\). Grundsätze für das Management von Forschungsdaten an der Universität Leipzig, Zugriff: 20.04.2020.](#)

[Deutsche Forschungsgemeinschaft \(2015\). Leitlinien zum Umgang mit Forschungsdaten, Zugriff: 20.04.2020.](#)

MIT WELCHEN FORSCHUNGSDATEN ARBEITEN SIE?



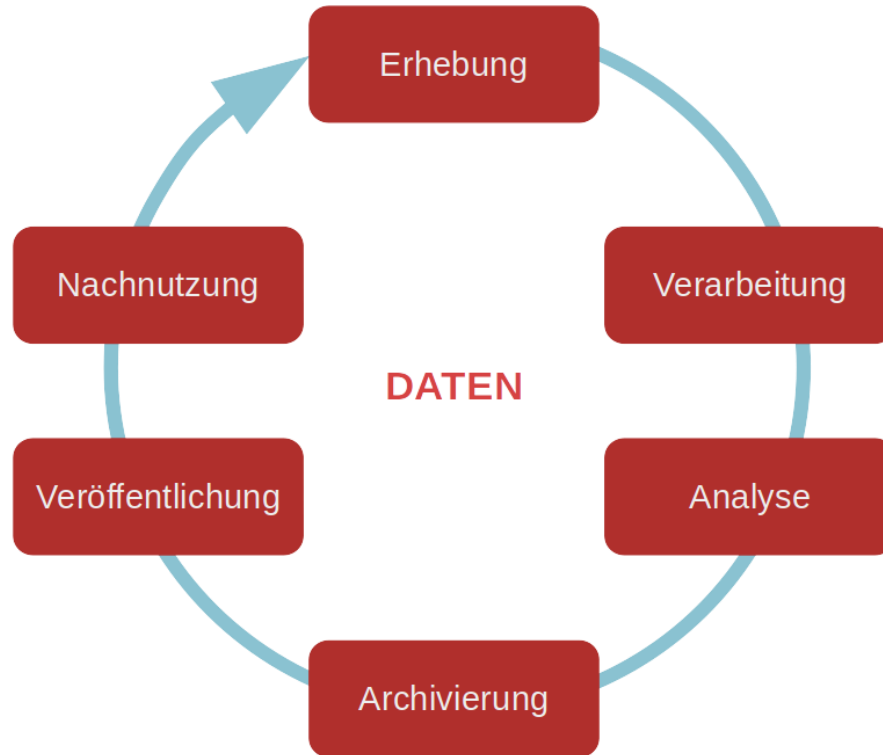
Quelle: pixabay.com.
Dieses Werk ist lizenziert unter einer
[CC0 1.0 Universal \(CC0 1.0\) Public Domain Dedication](https://creativecommons.org/publicdomain/zero/1.0/).
<https://creativecommons.org/publicdomain/zero/1.0/deed.de>

FORSCHUNGSDATENLEBENSZYKLUS

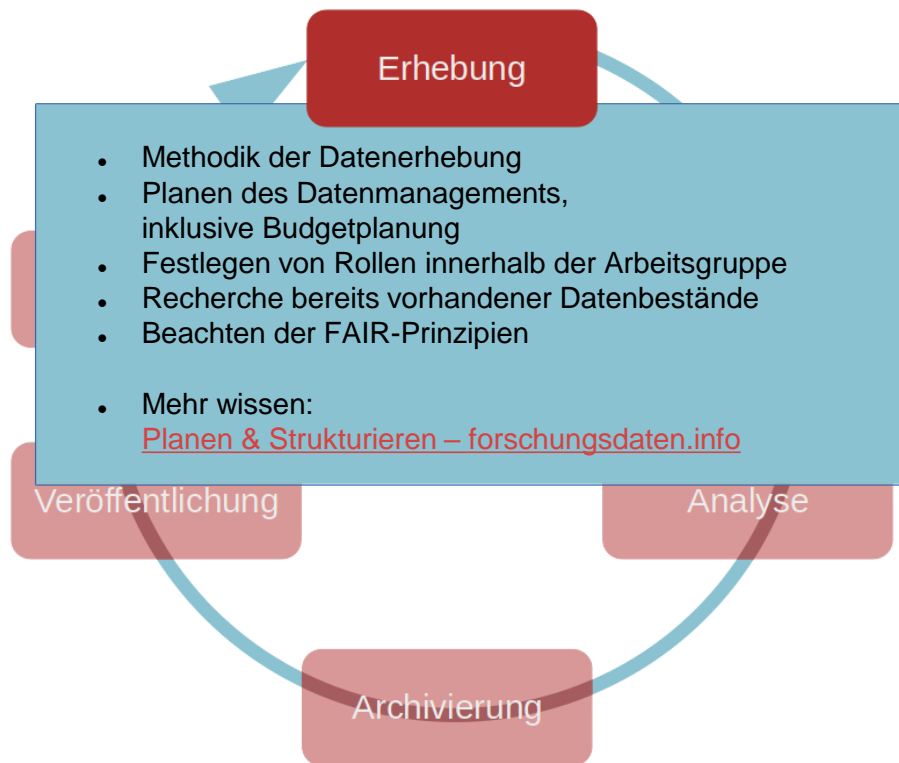
ist eine schematische Darstellung der Phasen, die Daten in einem Forschungsprozess idealerweise durchlaufen. Hieraus ergeben sich die Aufgaben im Forschungsdatenmanagement (FDM) für Forschende und Einrichtungen der wissenschaftlichen Infrastruktur.

Die Phasen des Forschungsdatenlebenszyklus bilden zudem die Grundlage für das Aufstellen eines Datenmanagementplans sowie für die Kostenkalkulation im Gesamtprojekt.

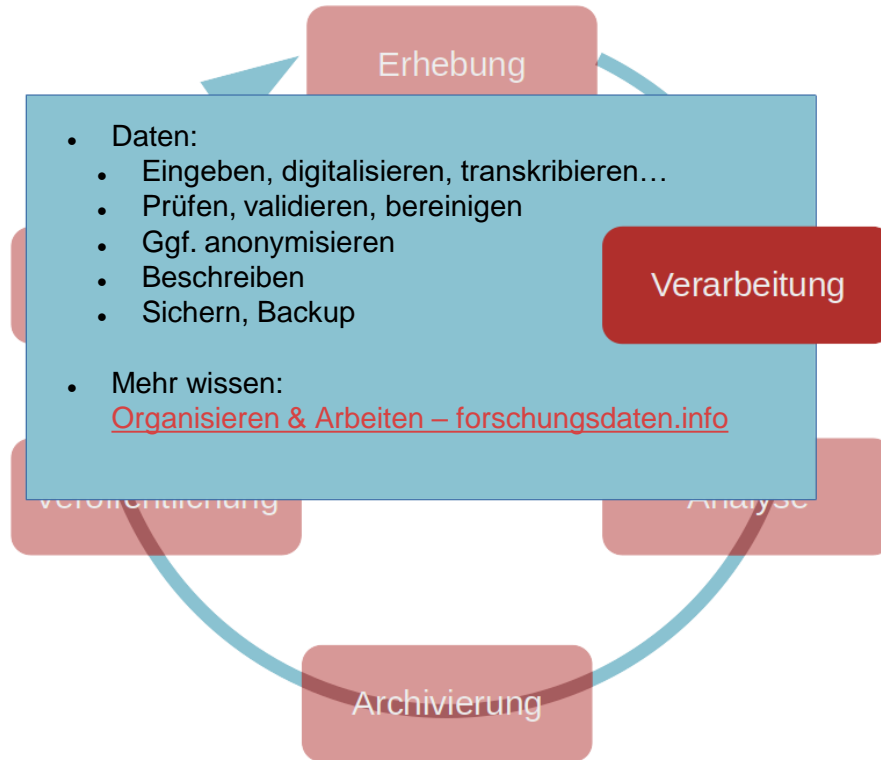
FORSCHUNGSDATENLEBENSZYKLUS



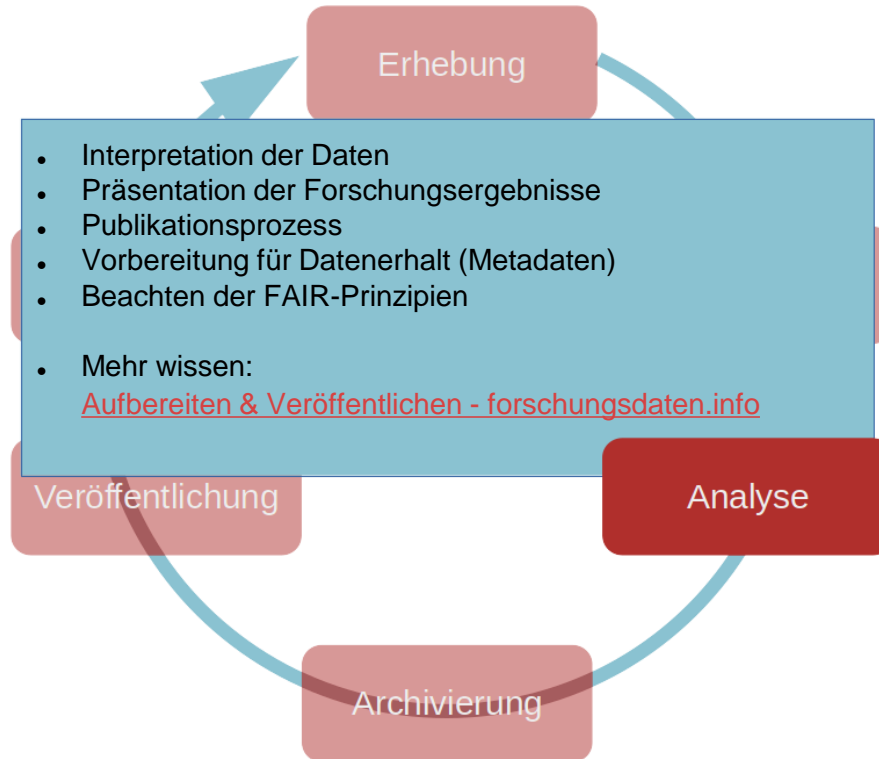
FORSCHUNGSDATENLEBENSZYKLUS



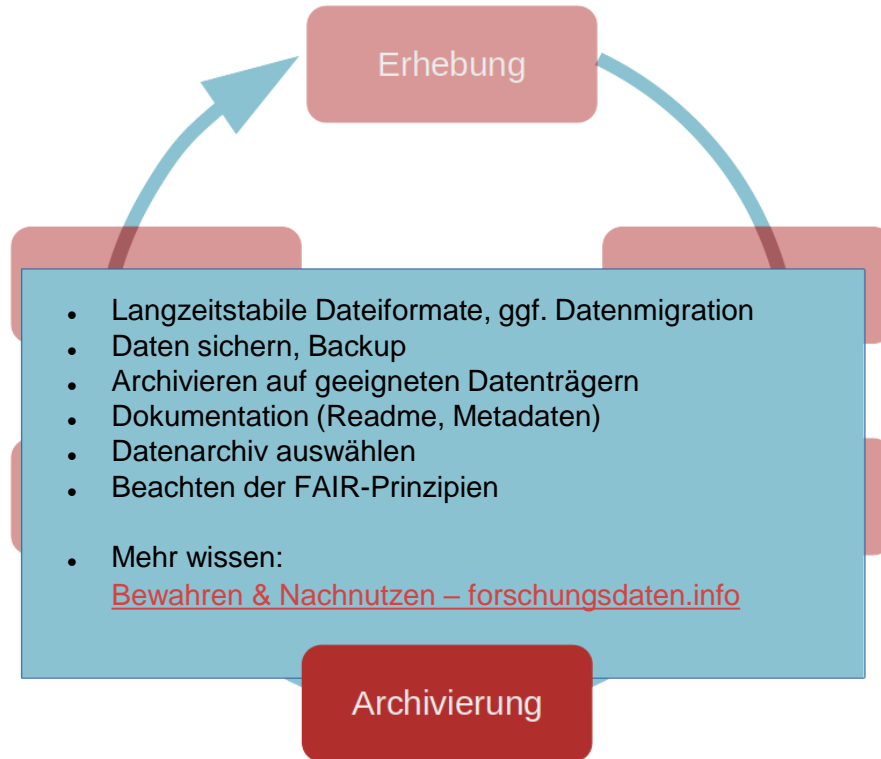
FORSCHUNGSDATENLEBENSZYKLUS



FORSCHUNGSDATENLEBENSZYKLUS



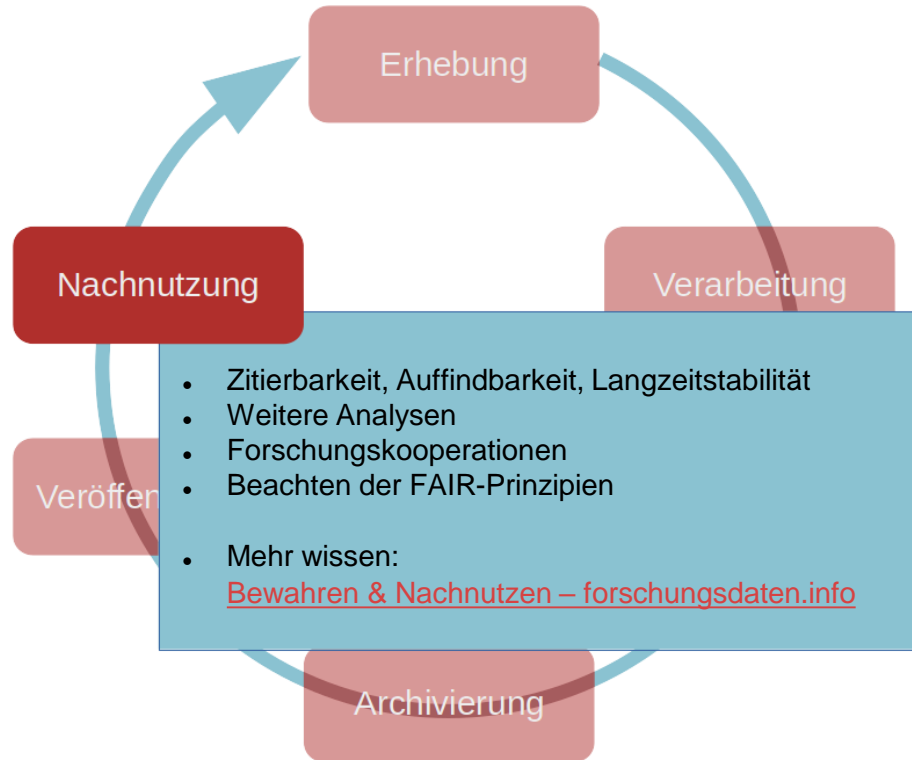
FORSCHUNGSDATENLEBENSZYKLUS



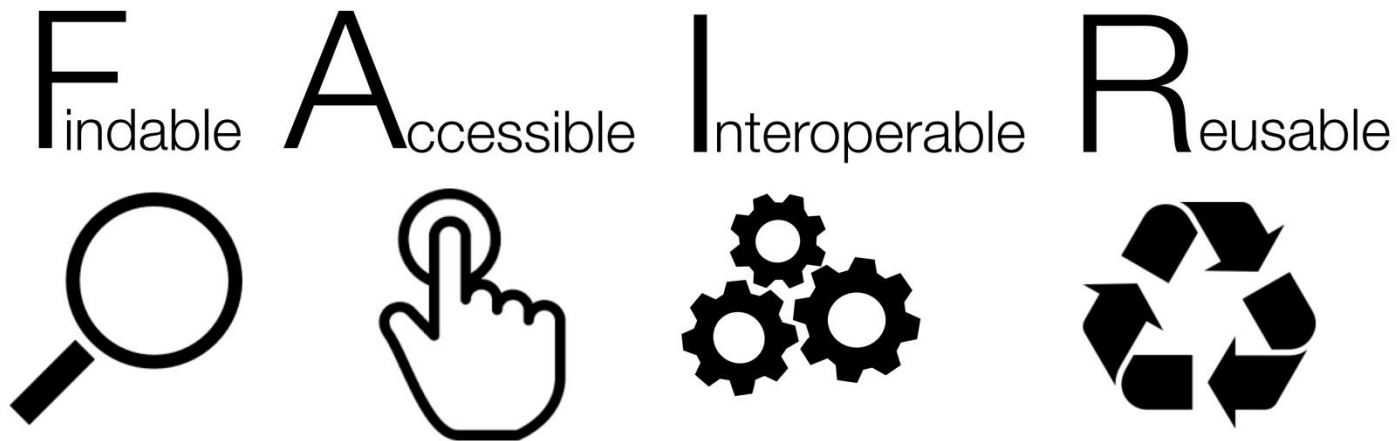
FORSCHUNGSDATENLEBENSZYKLUS



FORSCHUNGSDATENLEBENSZYKLUS



FAIR-PRINZIPIEN



Quelle: Pundir, Sangya. https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg, CC-BY-SA-4.0 <https://creativecommons.org/licenses/by-sa/4.0/deed.en>

FAIR-PRINZIPIEN

... sind eine international anerkannte, unverbindliche Richtlinie der Datendokumentation und -veröffentlichung.

Sie helfen, Forschungsdaten entlang ihres Lebenszyklus so zu organisieren, zu beschreiben, zu speichern, zu archivieren und zu veröffentlichen, dass diese sowohl von Menschen als auch von Maschinen gelesen und interpretiert werden können.

Die FAIR Prinzipien dienen damit in erster Linie dem nachhaltigen Datenerhalt und der -nachnutzung.

FAIR-PRINZIPIEN

Findable:

Auffinden von Daten und deren Metadaten, menschen- und maschinenlesbar

Accessible:

Daten und Metadaten mittels Standard-Kommunikationsprotokollen abrufbar, herunterladbar oder lokal nutzbar

Interoperable:

(teil)automatisierter Austausch, Interpretation und Kombination von Daten

Reusable:

Nachnutzung von Daten unter verständlichen Bedingungen

- Grundsatz des Datenzugangs: “As open as possible, as closed as necessary.”

FORSCHUNGSDATENMANAGEMENT

... umfasst alle Aktivitäten, die mit

- Aufbereitung,
- Speicherung,
- Archivierung und
- Veröffentlichung von FD verbunden sind.

FDM begleitet den Forschungsprozess von den ersten Planungen bis zur Archivierung, Nachnutzung oder Löschung der Daten. Ziel ist es, diese langfristig zugänglich, nachnutzbar und nachvollziehbar zu halten.

WOZU FORSCHUNGSDATENMANAGEMENT?

- Grundvoraussetzung der guten wissenschaftlichen Praxis
- Glaubwürdigkeit, Nachvollziehbarkeit der eigenen Forschung durch Dritte
- Reproduzierbarkeit von Forschungsergebnissen
- Erfüllen der Vorgaben von Forschungsförderern und Verlagen
- Minimieren des Datenverlustrisikos
- Vereinfachung zukünftiger Nachnutzung der eigenen Daten und Interpretierbarkeit der Daten langfristig sicherstellen
- Unterstützung von Open Science, Ermöglichen neuer Erkenntnisse, Metaanalysen und Kooperationen

ASPEKTE DES FORSCHUNGSDATENMANAGEMENTS

Folgende Aspekte sind im FDM zu beachten und zugleich Teil eines Datenmanagementplans (DMP):

- Grundlegende Beachtung der FAIR-Prinzipien
- Ordnung und Strukturierung der erhobenen oder generierten Daten
- Dokumentation und Metadaten
- Regelmäßige Speicherung und Backup
- Langzeitarchivierung der FD
- Datensicherheit
- Evtl. Publikation von FD
- Auffinden und Nachnutzbarkeit
- Rechtliche und ethische Aspekte

FORSCHUNGSDATEN-POLICIES

Policies definieren Richtlinien und Verantwortlichkeiten zum Umgang mit FD. Sie werden von verschiedenen Instanzen verabschiedet und sind unbedingt zu beachten:

- **Institutionell:**
[Grundsätze für das FDM an Uni Leipzig](#)
- **Forschungsförderer:**
[Leitlinien zum Umgang mit Forschungsdaten \(DFG\)](#)
[Horizon 2020 Online Manual – Data Management](#)
- **Fachgemeinschaften:**
[Umgang mit Forschungsdaten im Fach Psychologie \(DGP\)](#)
[CESSDA - Sozialwissenschaften](#)
- **Verlage und Repositorien:**
[Springer Nature – Research Data Policy](#)
[Zenodo – General Policies](#)

„GRUNDSÄTZE FÜR DAS FDM AN DER UL“ (VERÖFFENTLICHT AM 17.04.2019)

- **Präambel:**
Ziel, Bedeutung, Hinweis auf gute wissenschaftlichen Praxis
- **Geltungsbereich:**
für Mitglieder und Angehörige der UL
- **Prinzipien:**
Bedeutung der FAIR-Prinzipien, Speicherung, Publikation
- **Verantwortlichkeiten der Forschenden:**
 - Verantwortung für das Management Ihrer FD
 - Anhalten zum Aufstellen eines DMPs
 - Gesetze einhalten, z. B. zu Personenschutz, Eigentum, Verträgen
 - Verantwortung für Mitarbeiter_innen
- **Verantwortlichkeiten der Hochschulleitung:**
 - Voraussetzungen schaffen: Infrastruktur, Beratungs- und Weiterbildungsangebot

VERLAGS-POLICIES – BEISPIEL SPRINGER NATURE

Auch Verlage verabschieden Richtlinien zum Umgang mit FD. Diese sollten für eine Veröffentlichung beachtet werden:

Typ 1 – Es wird dazu ermutigt, Daten zu teilen und zu zitieren (*[Photosynthesis Research](#)*)

Typ 2 – Es wird dazu ermutigt, Daten zu teilen und Datenzugänglichmachung nachzuweisen (*[Plant and Soil](#)*)

Typ 3 – Es wird dazu ermutigt, Daten zu teilen und Aussagen zur Datenverfügbarkeit sind verpflichtend (*[Palgrave Communications](#)*)

Typ 4 – Es ist eine Voraussetzung, Daten zu teilen, dies nachzuweisen und ein Peer Review der Daten zu ermöglichen (*[Scientific Data](#)*)

2. DATENDOKUMENTATION UND METADATEN

DATENDOKUMENTATION – INHALTE

- Beschreibung des Forschungsvorhabens
- Projektziele, Hypothesen
- Informationen zur Erhebung der Daten (Methoden, Einheiten, Zeiträume, Orte, verwendete Technik)
- Maßnahmen zur Datenbereinigung
- Struktur der Daten und deren Beziehungen zueinander
- Erläuterung von Variablen, Labels und Codes
- Unterschiede zwischen verschiedenen Datensatz-Versionen
- Informationen zum Zugang und Nutzungsbedingungen (Lizenzen)
- Verantwortlichkeiten und ggf. Urheberrecht an Daten

DATENDOKUMENTATION – VORGEHEN

- Generierung und Bearbeitung der Daten im Forschungsprozess möglichst ausführlich beschreiben
- Dokumentation mit Projektidee beginnen
- Ziel: Nachvollziehbarkeit, Validierung und Reproduzieren der Forschungsergebnisse
- Ausgestaltung richtet sich nach Disziplin und Gepflogenheiten des Fachbereiches
- FAIR-Prinzipien als Richtlinie in der Datendokumentation

- Menschenlesbare Dokumentation: Readme-Datei(en)
- Maschinenlesbare Dokumentation: strukturierte Metadaten

DATENDOKUMENTATION – README-DATEI

- ist eine menschenlesbare Textdatei, die einen Datensatz/einen Programmcode beschreibt, dient dessen Interpretation
- selbsterklärende Dateinamen wählen, sollte mit dem beschriebenen Datensatz assoziiert werden können
- möglichst in einem offenen Textformat (.txt) verfassen, proprietäre Formate (wie MS Word) vermeiden
- lieber mehr Informationen geben, als zu wenige
- Hilfestellung: [Readme-Template](#)

- Beispiele:
[readme.txt \(EDMOND\)](#), [Readme.txt \(GFZ-Dataservice\)](#)

DATENDOKUMENTATION – METADATEN

- sind strukturierte Angaben über Daten, maschinenlesbar
- ermöglichen das Interpretieren, Finden, Durchsuchen, Nutzen von Forschungsdaten
- möglichst umfassende, standardisierte Metadaten vergeben
- standardisiertes Vokabular zur Vergabe von Schlagworten zur Beschreibung der Datensätze verwenden, um Suche und Wiederauffindbarkeit zu vereinfachen (Klassifikationen, Thesauri: [Basel Register of Thesauri, Ontologies & Classifications](#))
- an fachspezifischen Standards orientieren: [Verzeichnis Metadaten-Standards](#), [RDA – Metadata Directory](#)
- einfacher, weit verbreiteter Metadaten-Standard im XML-Format: [Dublin Core](#)

DATENDOKUMENTATION – METADATEN

Metadaten können verschiedene Angaben enthalten. Denkbar sind:

- **Inhaltliche Metadaten**

- Titel des Datensatzes
- Datenverantwortliche_r/
Primärforscher_in (ggf. ORCID)
- Mitwirkende (ggf. ORCID)
- Rechte (Lizenzen)
- Sprache
- Methodik
- Abstract
- ...

- **Administrative/Technische Metadaten**

- Datum/Uhrzeit
- Standort (ggf. Koordinaten)
- Datenart, Dateiformat, -größe
- Probe
- Gerätetyp
- ...

DATENDOKUMENTATION – METADATEN

- Fachspezifisches Metadatenset – z. B. [Metadatenset Forschungsdaten Bildung](#)
- Beispiel: [Darwin Core](#) im XML-Format (Biodiversity Information Standard):

```
<?xml version="1.0"?>
<dwr:SimpleDarwinRecordSet
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://rs.tdwg.org/dwc/xsd/simpledarwincore/ http://rs.tdwg.org/dwc/xsd/tdwg_dwc_s
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:dwc="http://rs.tdwg.org/dwc/terms/"
  xmlns:dwr="http://rs.tdwg.org/dwc/xsd/simpledarwincore/">
  <dwr:SimpleDarwinRecord>
    <dcterms:type>PhysicalObject</dcterms:type>
    <dcterms:modified>2009-02-12T12:43:31</dcterms:modified>
    <dcterms:rightsHolder>Museum of Vertebrate Zoology</dcterms:rightsHolder>
    <dcterms:rights>Creative Commons License</dcterms:rights>
    <dwc:institutionCode>MVZ</dwc:institutionCode>
    <dwc:collectionCode>Mammals</dwc:collectionCode>
    <dwc:occurrenceID>urn:catalog:MVZ:Mammals:14523</dwc:occurrenceID>
    <dwc:basisOfRecord>PreservedSpecimen</dwc:basisOfRecord>
    <dwc:country>Argentina</dwc:country>
    <dwc:countryCode>AR</dwc:countryCode>
    <dwc:stateProvince>Neuquén</dwc:stateProvince>
    <dwc:locality>25 km al NNE de Bariloche por Ruta 40 (=237)</dwc:locality>
  </dwr:SimpleDarwinRecord>
</dwr:SimpleDarwinRecordSet>
```

PERSISTENTE IDENTIFIER (PID)

- verknüpfen digitale Informationen mit einer festen Adresse (Zeichenkette) und machen diese dadurch dauerhaft auffindbar und zitierbar
- verhindern Informationsverlust durch defekte Links auf digitale Informationen
- im Gegensatz zu URLs sind PIDs unveränderlich
- Informationen können bspw. Publikationen, FD, Software-Code sein

- Beispiele:
 - Digital Object Identifier (DOI) zur eindeutigen Zuordnung von Publikationen und FD-Sätzen
 - ORCID zur eindeutigen Zuordnung wissenschaftlicher Leistungen zu einer Person

DIGITAL OBJECT IDENTIFIER (DOI)

- sichert dauerhafte Auffindbarkeit und Zitierbarkeit einmal veröffentlichter FD
- Metadaten müssen nur einmal eingegeben werden und sind mit publiziertem Datensatz verknüpft

Einzigartige Folge von alphanumerischen Zeichen:

- Präfix (Organisationskennung)
- Suffix (Objektkennung)

Beispiel:

- 10.1594/PANGAEA.909429
- [https://https://doi.org/10.1594/PANGAEA.909429](https://doi.org/10.1594/PANGAEA.909429)



OPEN RESEARCHER AND CONTRIBUTOR ID (ORCID)

... ist ein Profildienst und Identifier für Wissenschaftler_innen.

- eindeutige Zuordnung von Personen zu Werken
- von Forschenden selbst angelegt und gepflegt
- verbunden mit zahlreichen Datenquellen
- löst Problem von Namensänderungen und Namensvettern
- implementiert auch in (Daten-)Repositorien

- Beispiel:

<https://orcid.org/0000-0002-8321-6629>



3. DATENSPEICHERUNG UND -ORGANISATION

SPEICHERUNG

- Schaffen einer nachvollziehbaren Speicherstruktur ist essentiell für die Arbeit mit Daten, besonders in der Gruppe
- Ordnerstrukturen:
 - inhaltlich zusammengehörige Daten in eine Datei/einen Ordner innerhalb eines Unterordners, maximal 2–3 Unterordner
- Funktionale Benennungskonvention für Dateien etablieren:
 - Versionskontrolle
- Sichere und nachhaltige Speichermedien wählen, bspw. institutioneller Speicher, passwortgeschützte Speichermedien, sicherer Cloudservice

DATEIBENENNUNG – TIPPS

- Aussagekräftige Namen
- Einheitliches Schema, logische Struktur
- Datumsangabe zur chronologischen Sortierung: JJJJMMTT
- Vermeidung von Leer- und Sonderzeichen, Umlauten
- Dokumentierte Namenskonventionen oder genutzte Abkürzungen, z. B.
 - [Sediment]_[Probe]_[Instrument]_[JJJJMMTT].csv
 - [Projekt]_[Interview]_[Ort]_[Personen-ID]_[JJJJMMTT].mp4
- Ggf. Werkzeuge für gleichzeitige Umbenennung mehrerer Dateien nutzen

DATEIBENENNUNG – TOOLS ZUR UMBENENNUNG

Windows:

[Ant Renamer](#)

[Rename It](#)

[Bulk Rename Utility](#)

Mac:

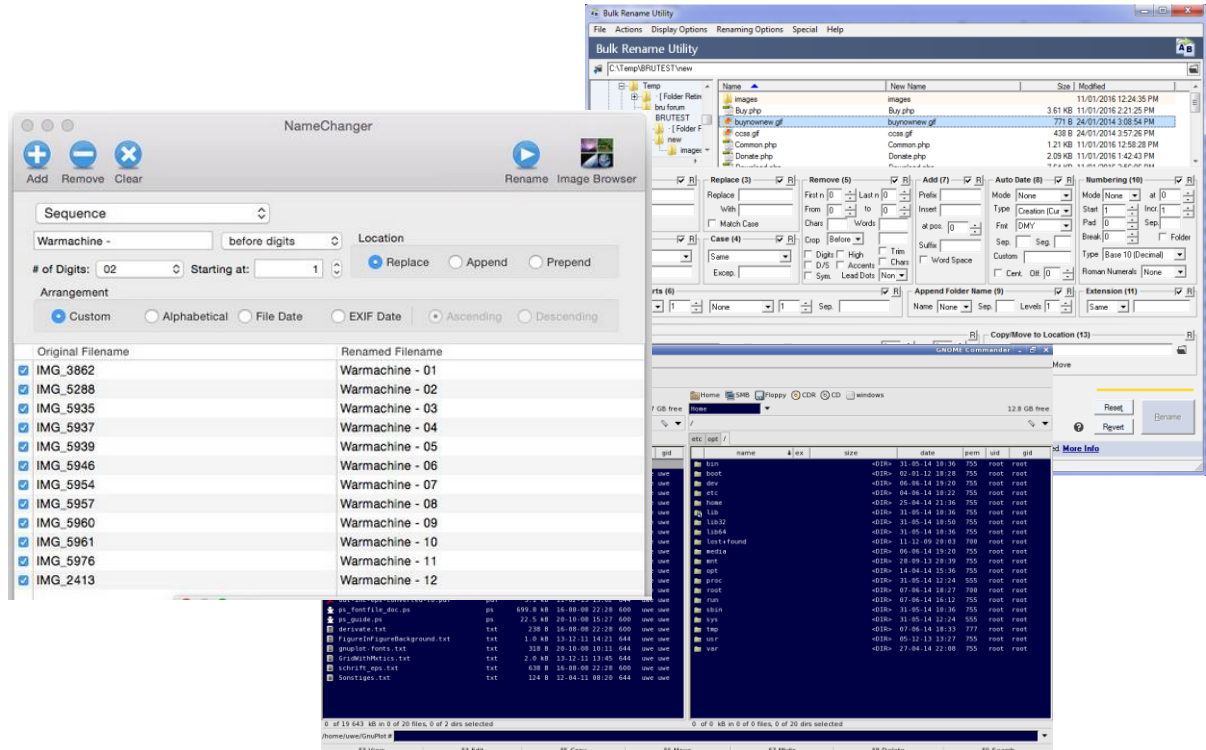
[Renamer 6](#)

[Name Changer](#)

Linux:

[GNOME Commander](#)

[GPRename](#)



SPEICHERMEDIEN

Worauf man bei der Wahl eines Speichermediums achten sollte:

- Metadaten, Skalierbarkeit, Möglichkeit zur Verschlüsselung
- Größe, Kosten
- Zugriffsgeschwindigkeit, Lese- und Schreibgeschwindigkeit, Zuverlässigkeit
- Gemeinsamer Zugriff, Zugangskontrolle, Rechteverwaltung
- Wartbarkeit
- Schnittstellen
- Backup

SPEICHERMEDIEN – EIGENER PC

Vorteile

- selbst verantwortlich für Sicherheit und Backup
- maximale Kontrolle

Nachteile

- was mit dem PC geschieht, geschieht mit dem Backup
- eventuell fehlende Ressourcen und Know-how
- Einzellösungen aufwendig

SPEICHERMEDIEN – MOBIL (STICK, EXTERNE FESTPLATTE)

Vorteile

- einfach zu transportieren
- kann im verschließbaren Schrank oder Safe aufbewahrt werden

Nachteile

- Verlust, Diebstahl, ...
→ besonders unsicher
- Inhalte ungeschützt, falls nicht verschlüsselt
- externe Festplatte: stoß- und verschleißanfällig

SPEICHERMEDIEN – INSTITUTIONELL (CLOUD, VIRTUELLE LAUFWERKE)

Vorteile

- Backup der Daten ist sichergestellt
- professionelle Durchführung und Wartung
- Speicherung gemäß Datenschutzrichtlinien der Institution

Nachteile

- Geschwindigkeit eventuell vom Netzwerk abhängig
- Zugriff auf Backups eventuell verzögert durch Dienstweg
- eventuell unklar welche Sicherheitskriterien und -strategien eingesetzt werden

SPEICHERMEDIEN – EXTERN (CLOUD EINES KOSTENLOSEN ODER KOSTENPFLICHTIGEN ANBIETERS)

Vorteile

- einfach zu nutzen und zu verwalten
- Backup der Daten ist sichergestellt
- für mobiles Arbeiten nutzbar
- professionelle Durchführung und Wartung

Nachteile

- je nach Anbieter kann die Verbindung auch unsicher sein
- abhängig vom Zugang zum Internet (Up- & Download evtl. langsam)
- Zugriff auf Backups eventuell verzögert
- Datenschutz?

LEITFRAGEN BEI DER WAHL DES SPEICHERMEDIUMS

- Wie viel Speicherplatz benötige ich?
- Welche Datentypen habe ich und wie häufig werde ich diese ersetzen?
- Wer benötigt Zugang?
- Ist es notwendig einen Remote-Zugang zu den Daten zu haben?
- Wie wichtig ist schneller Zugriff?

LEITFRAGEN BEI DER WAHL DES SPEICHERMEDIUMS

- Wird simultaner und synchroner Zugriff benötigt?
- Welche Schritte kann ich mit diesem Medium vornehmen, um meine Daten vor Verlust zu schützen? (Passwort, Verschlüsselung, physischer Schutz u. a.)
- Welche Speicherlösungen sind für personenbezogene Daten geeignet?
- Wie häufig werde ich ein Backup machen und wo wird dieser gespeichert?
- Wie viel finanzielle Mittel stehen mir zur Verfügung?

SCHUTZ VOR DATENVERLUST

Risiken

- unbeabsichtigte Änderungen oder Löschen
- technische Defekte
- Katastrophen (Unwetter, Brand)
- Diebstahl etc.



Strategien

- Sicherung wichtiger Dateien auf räumlich getrennten Datenträgern
- Speicherung auf Universitätsservern mit automatischem, regelmäßigem Backup



BACKUP

- Backup-Varianten
 - **Voll-Backup:**
Sicherung des gesamten Datenbestands, speicherintensiv
 - **Inkrementelles Backup:**
Sicherung der Änderungen gegenüber dem vorherigen inkrementellen oder Voll-Backup, alle Backup-Schritte sind zur Wiederherstellung der Daten nötig, wenig speicherintensiv
 - **Differenzielles Backup:**
Sicherung der Änderungen gegenüber dem letzten Vollbackup, Vollbackup und 1 Backup-Schritt sind zur Wiederherstellung nötig
- Sicheres Backup: Die 3-2-1-Regel
 - mindestens 3 Kopien einer Datei, auf mindestens 2 unterschiedlichen Medien, wovon mindestens eins dezentral ist
- Testen Sie die Datenwiederherstellung zu Beginn sowie in regelmäßigen Abständen!

SCHUTZ SENSIBLER DATEN

Schützen Sie Ihre (sensiblen) Daten:

- Hardware (bspw. separater abschließbarer Raum, Verschlüsselung)
- Datenverschlüsselung
 - Software: [GnuPG](#), [OpenPGP](#)
- Sicherheit der Passwörter
 - Tool: [Keepass](#), Webseite: [Prism Break](#)
- Mindestens zwei Personen sollten Zugang zu Ihren Daten haben
 - beide sollten das Passwort und die Speicherorte kennen

4. LANGZEITARCHIVIERUNG

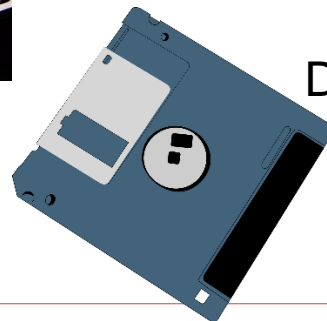
WIE LANGE IST LANGZEIT?



CD: 5–10 Jahre



Festplatte: 3–10 Jahre



Diskette: 10–20 Jahre

Quelle: pixabay.com.

Dieses Werk ist lizenziert unter einer
[CC0 1.0 Universal \(CC0 1.0\) Public
Domain Dedication.](https://creativecommons.org/publicdomain/zero/1.0/)

<https://creativecommons.org/publicdomain/zero/1.0/deed.de>

ARCHIVIERUNG – ABGRENZUNG ZU BACKUP

Backup

- automatische Sicherung aller Daten, um Datenverlust vorzubeugen
- technische Ursachen, z. B. defekt
- menschliches Versagen, z. B. versehentlich gelöscht
- alle Versionen

Archivierung

- Sicherung ausgewählter Daten, um diese langfristig aufzubewahren
- nur endgültige Versionen
- Integritätssicherung
- Langzeitspeicherung
- Durchsuchbarkeit

LANGZEITARCHIVIERUNG

- Daten sind langfristig recherchierbar, auffindbar, zugänglich, lesbar
- Gute wissenschaftliche Praxis: 10 Jahre Verfügbarkeit von Forschungsdaten
- Langzeitarchivierung umfasst somit Maßnahmen, die geplant, kontrolliert und vorgenommen werden müssen
- Geld und Zeit einplanen für die Nutzung einer geeigneten Infrastruktur, Daten zu dokumentieren und mit Metadaten zu versehen
- frühzeitige Planung empfohlen

LANGZEITARCHIVIERUNG

- Zu archivierende Daten sollten
 - unverschlüsselt,
 - nicht komprimiert,
 - patentfrei,
 - in offenen, langzeitstabilen Dateiformaten gespeichert (Übersicht: [Katalog Archivischer Dateiformate](#)),
 - in offenen, dokumentierten Standards erstellt sein.
- ggf. Migration: Testen!
- Grundprinzip: seltenere Migration, längere Lebensdauer

LANGZEITSTABILE DATEIFORMATE

- Wahl des Dateiformats hängt mit der beabsichtigten Nutzung der Daten zusammen: Sollen diese unkompliziert nachgenutzt oder möglichst lang aufbewahrt werden können?
- Tipps zur Wahl des richtigen Formats:
 - in der eigenen Community etablierte Formate verwenden
 - offene Formate verwenden, um die Daten unabhängig von kommerzieller Software nutzen zu können (bspw. erfordert das Statistikprogramm SPSS ein spezielles Format .sav, das von anderen Programmen nicht gelesen werden kann)
 - offene Formate erhöhen die Wahrscheinlichkeit, dass Daten auch zukünftig noch gelesen werden können
 - Formate wählen, die ggf. verlustfrei in andere Formate konvertiert werden können

LANGZEITSTABILE DATEIFORMATE

Dateiformat	Empfehlung	Vermeiden
Tabellen	CSV, TSV, ODS, XLSX, HTML/XML	XLS, SPSS
Text	TXT, PDF/A, ODT, DOCX	DOC, PPT, PDF
Multimedia	Container: MKV Codec: FLAC, FFV1	MOV, Ogg, Flash, WMA, WMV
Bilder/Grafiken	TIFF, DNG, PNG, SVG, JPEG2000	GIF, JPEG

ANFORDERUNGEN AN LANGZEITARCHIVE

- **Technische Anforderungen**

Welche physischen Speicher werden genutzt? Welche Backup-Strategie wird angewendet? Werden Daten migriert oder komprimiert etc.

- **Kosten**

Welche Kosten fallen für die Nutzenden an? In welcher Höhe und auf welcher Grundlage (Speichervolumen, Serviceleistungen)?

- **Zugänglichmachung der Daten**

Unter welchen Bedingungen werden Metadaten und Daten bereitgestellt? Wer darf auf die Daten in welcher Form zugreifen? Welche Beschränkungen gibt es?

- **Siegel für vertrauenswürdige Langzeitarchive**

Beispiele: [Core Trust Seal](#), [nestor-Siegel](#), [ISO 16363](#), DIN 31644

- **Langlebigkeit des Dienstleisters**

- **Beachtung der FAIR-Prinzipien**

5. DATENVERÖFFENTLICHUNG

WARUM FORSCHUNGSDATEN VERÖFFENTLICHEN?

- Eigenen Forschungsoutput erhöhen:
veröffentlichte Forschungsdaten als eigenständige, zitierbare, nachnutzbare wissenschaftliche Leistung
- Steigerung der Zitationsrate von Publikationen, deren Daten öffentlich frei verfügbar sind
- Datenverlust minimieren:
in geeigneten Repositorien veröffentlichte Daten werden langzeitverfügbar gehalten, eigene Arbeit wird gesichert

WARUM FORSCHUNGSDATEN VERÖFFENTLICHEN?

- Vorgaben von Forschungsförderern erfüllen
- Anderen Wissenschaftler*innen die Arbeit mit qualitativ hochwertigen Daten ermöglichen:
Veröffentlichen von besonders aufwändig erhobenen, einzigartigen Forschungsdaten
- Open Science:
neue Erkenntnisse und wissenschaftliche Kooperationen fördern

PUBLIKATIONSWEGE

- Anhang zu einem begutachteten Artikel („Supplementary Material“, [Beispiel](#))
- Eigenständige Publikation in einem Forschungsdaten-Repository
 - fachspezifisch, Beispiele: [Datorium](#), [Pangaea](#)
 - institutionell, Beispiele: [OpARA](#), [Edmond](#), [DB Thüringen](#)
 - disziplinübergreifend, z. B. [Zenodo](#)
- Data Journals
 - Beispiele: [Earth System Science Data](#), [Data in Brief](#)
 - [Liste Data Journals](#)

REPOSITORIEN FINDEN

Verzeichnis von Forschungsdatenrepositorien: re3data.org



QUALITÄTSMERKMALE VON REPOSITORIEN

- Vergabe von persistenten Identifikatoren für Datensätze/Software (Beispiel: DOI, URN, PURL, Handle-System)
- Anwendung von persistenten Identifikatoren für Autor*innen (ORCID)
- Vergabe von Metadaten
- Download- und Exportmöglichkeiten (Zitierweise in verschiedenen Formaten, bspw. Bibtex)
- Möglichkeit zur Beschreibung des Datensatzes (Klartext auf der Webseite, Readme-Datei)

QUALITÄTSMERKMALE VON REPOSITORIEN


- Verschiedene Zugriffsmöglichkeiten zur Beschränkung des Datenzugangs (Rechtemanagement, auf Anfrage, nur Metadaten frei verfügbar ...)
- Transparente Lizenzvergabe (Auswahlmöglichkeiten verschiedener Lizenzmodelle)
- Versionierungen des Datensatzes (Vorhalten und Zugang verschiedener Versionen)
- Auffinden des Datensatzes durch Suchmaschinen (Schnittstellen)
- Siegel (Beispiel: Core Trust Seal, nestor-Siegel)

LIZENZEN VERGEBEN




Voraussetzung: Bin ich befugt, Daten zu veröffentlichen und Lizenzen zu vergeben?

- Verantwortlichkeiten im Datenmanagementplan prüfen, rechtliche Aspekte beachten
- Personenbezogene und sensible Daten nicht veröffentlichen,
Hilfestellung: [Forschungsdaten veröffentlichen?](#)
- Leitfrage: Was sollen andere mit meinen Daten tun dürfen?
- Öffentliche Lizenzen als Werkzeug zur rechtssicheren (Nach-)Nutzung von Forschungsdaten, Vertragsstatus zwischen Autor*in und Nutzenden
- Meistverbreitet: [Creative Commons](#) – ab 4.0 auch für Forschungsdaten

LIZENZEN VERGEBEN – BEISPIEL: CREATIVE COMMONS

  CC0 (Public Domain Dedication)

  CC BY 4.0 (Namensnennung)

   CC BY-SA 4.0 (Namensnennung – Weitergabe unter gleichen Bedingungen)

   CC BY-ND 4.0 (Namensnennung – Keine Bearbeitung)

   CC BY-NC 4.0 (Namensnennung – Nicht kommerziell)

...

 Public Domain Mark

Vergeben Sie für Ihre veröffentlichten FD stets eine Lizenz, um eine rechtssichere Nachnutzung zu ermöglichen!

6. RECHTLICHE UND ETHISCHE ASPEKTE

RECHTLICHE UND ETHISCHE ASPEKTE

Vorabinformation:

Die AG Forschungsdaten der Universität Leipzig **ist nicht befugt**, rechtsverbindliche Aussagen für das Management Ihrer Daten zu treffen. Diese können lediglich empfehlenden Charakter haben.

- Bei Fragen zum Datenschutz und zur Verarbeitung personenbezogener Daten kontaktieren Sie bitte den Datenschutzbeauftragten der Universität:
dsb@uni-leipzig.de
- Ethische Fragen beantwortet Ihnen die Geschäftsstelle des Ethikbeirates:
ethikbeirat@uni-leipzig.de
- Fragen zu Patent- und Lizenzvorgaben richten Sie bitte an den Transferservice der Universität:
dezernat1@zv.uni-leipzig.de

RECHTLICHE HERAUSFORDERUNGEN

Urheberrecht	Wissenschaftsrecht	Arbeits-/Dienstrecht	Förderbedingungen	Policies
Unterliegen FD dem Urheberrecht (Schöpfungshöhe)?	Bestehen Vorgaben im hochschuleigenen Recht?	Wem gehören die an der UL erhobenen FD?	Welche Bedingungen geben Förderer/Industrie vor?	Welche Bedingungen für FDM geben Policies vor?
Datenschutz	Verträge	Grundrechte	Internationales Recht	EU-Recht
Welche Daten sind schützenswert (DSGVO, BDSG)?	Bestehen Absprachen zum geistigen Eigentum an FD (Auftragsforschung)?	Welche verfassungsrechtlichen Grenzen sind zu beachten?	Welche Bestimmungen außerhalb Deutschlands bestehen?	Was bringt bspw. die European Open Science Cloud für FD?
Zoll- und Exportbestimmungen	Wettbewerbsrecht	Patentrecht	Verlagsrecht	Quelle
Wie wirken sich Security Trade Control- und Zollbestimmungen auf Produktentwicklungen aus?	Werden Daten im unternehmerischen Geschäftsverkehr genutzt?	Was ist zu beachten, wenn Forschungsdaten Patentreife erreichen? Zu welchem Zeitpunkt dürfen Daten, die im Zusammenhang mit dem Patentrecht stehen, veröffentlicht werden?	Welche Embargofristen gelten für die Veröffentlichung von Forschungsdaten im Open Access?	Hartmann, Thomas. (2019). Rechtsfragen: Institutioneller Rahmen und Handlungsoptionen für universitäres FDM. Zenodo. https://doi.org/10.5281/zenodo.2654306

FORSCHUNGSDATEN UND URHEBERRECHT

- Für Forschungsdaten allein kann kein Urheberrechtsschutz angenommen werden, denn diese stellen in der Regel keine eigenständige geistige Schöpfung dar und erfüllen damit nicht die erforderliche Schöpfungshöhe.
- Daher müssen rechtliche Fragen in Bezug auf die Namensnennung, Veröffentlichung und (Nach-)Nutzung von FD im Voraus eines Forschungsprojekts geklärt werden.
- Vereinbarungen sollten schriftlich, am besten im DMP, festgehalten werden.
- Empfehlung:
FD zunächst so behandeln, als fielen sie unter das Urheberrecht, um begründete Entscheidungen für die gesamten Schritte im Datenlebenszyklus treffen und Verantwortlichkeiten festlegen zu können.
- Mehr wissen: [Urheberrecht – forschungsdaten.info](https://www.forschungsdaten.info/urheberrecht)

SCHUTZ PERSONENBEZOGENER DATEN

- Erfüllung von rechtlichen Datenschutzvorgaben (DSGVO, BDSG)
- Schutz der Identität, vor allem bei sensiblen/vertraulichen Themen
→ ggf. Pseudonymisierung/ Anonymisierung (Tool: [Amnesia](#))
- Schutz der Teilnehmenden vor kommerzieller Nutzung der Angaben
- Personenbezogene Angaben sind grundsätzlich von anderen erhobenen Daten getrennt aufzubewahren und dürfen nur zu wissenschaftlichen Zwecken zusammengeführt werden
- Eine Archivierung personenbezogener Daten ist nur mit Einwilligung der Teilnehmenden möglich

INFORMIERTE EINWILLIGUNG

- Teilnehmende müssen darüber informiert werden, was mit ihren Daten passiert.
- Teilnehmende müssen einer Teilnahme auf Basis dieser Informationen zustimmen.
- Datenanalyse und -archivierung benötigen separate Einwilligungen.
- Vorlagen und Checklisten bietet unter anderem <https://www.forschungsdaten-bildung.de/einwilligung>
- Häufiges Problem bei der Archivierung: keine Einwilligung vorhanden
- Strikte Formulierungen in Einwilligungserklärungen

ANONYMISIERUNG UND PSEUDONYMISIERUNG

- **Anonymisierung**

- Entfernen aller identifizierenden Informationen/Details
- Rückschlüsse auf Person unmöglich machen unter Einsatz aller bekannten Mittel
- Ersetzen sensibler Informationen mit einer Beschreibung, die Bezug zu originärem Kontext hat
- Anonymität vor Vollständigkeit/Information

- **Pseudonymisierung**

- Ersetzen sensibler Informationen durch ein Pseudonym (Code)

ETHIK IM FORSCHUNGSDATENMANAGEMENT

- Erhebung von sensiblen Daten nur mit positivem Votum der Ethikkommission
- Ethisch relevante Punkte in Forschungsvorhaben:
 - Versuche an und mit Menschen (Befragungen, Studien etc.)
 - Versuche, die personenbezogene Daten sammeln und archivieren
 - Forschung, bei der Proben wie etwa Gewebe von Menschen entnommen werden
 - Tierversuche
 - Umweltrelevante Forschung
 - Austausch von Wissen und Technologie mit Drittländern
 - Sicherheitsrelevante Forschung (Dual Use), bei der Potential zum Missbrauch der Forschungsergebnisse besteht
- Mehr wissen: [Ethik & FDM – forschungsdaten.info](https://www.forschungsdaten.info)

RECHT UND ETHIK IM FORSCHUNGSDATENMANAGEMENT

- Ethik und Recht sind komplexe Themen im FDM und müssen in den meisten Fällen aus mehreren Blickwinkeln betrachtet werden. Die Beschäftigung mit diesen Aspekten lässt zudem noch einen großen Interpretationsraum zu, da es noch keine hinreichend gesicherten Erkenntnisse und nur wenige Grundsatzurteile zu Referenzzwecken gibt.
- Daher empfehlen wir:
 - Nehmen Sie frühzeitig kompetente Beratung in Anspruch.
 - Klären Sie die Rechte an ihren Forschungsdaten von Beginn des Forschungsprozesses an in schriftlicher Form (vgl.: DMP), auch in Hinblick auf die Zeit nach Projektende und einem eventuell anstehenden Arbeitgeberwechsel.
 - Beziehen Sie möglichst alle Beteiligten an der Datenerhebung, -analyse und -veröffentlichung ein. Ziel sollte eine konsensbasierte Vereinbarung sein.

LITERATUR ZU RECHTLICHEN RAHMENBEDINGUNGEN

Aufgrund der Komplexität der rechtlichen Einordnung von Forschungsdaten möchten wir Ihnen an dieser Stelle weitere Informationsquellen zum Thema bereitstellen:

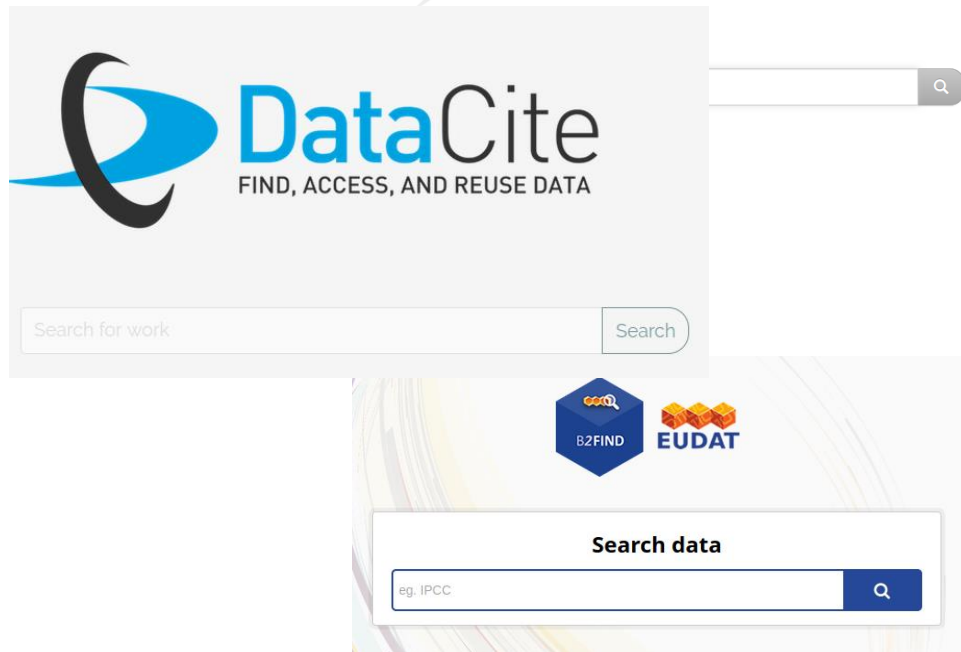
- Kreuzter, Till; Lahmann, Henning (2019). Rechtsfragen bei Open Science: Ein Leitfaden. Hamburg University Press, <https://doi.org/10.15460/HUP.195>
- Ostendorff, Phillip; Linke, David (2019). Best-Practices im Umgang mit rechtlichen Fragestellungen zum Forschungsdatenmanagement (FDM), Bibliotheksdienst Band 53: Heft 10–11, S. 710–723, <https://doi.org/10.1515/bd-2019-0098>
- Hartmann, Thomas (2019). Rechtsfragen: Institutioneller Rahmen und Handlungsoptionen für universitäres FDM. Zenodo, <https://doi.org/10.5281/zenodo.2654306>
- [Projekt DataJus – Kurzgutachten zu rechtlichen Rahmenbedingungen des FDM](#)

7. DATENNACHNUTZUNG

FORSCHUNGSDATEN FINDEN

Verschiedene Möglichkeiten:

- direkt in Repositorien
- mittels Metasuchmaschinen, z. B.
 - [BASE](#)
 - [B2FIND](#) (EUDAT)
 - [geisDataSearch](#)
- [DataCite Metadata Search](#)
- [Google Dataset Search](#)



FORSCHUNGSDATEN ZITIEREN

Vorschlag DataCite:

- Urheber (Veröffentlichungsdatum): Titel, Version. Publikationsagent [d.h. Repositorium/ Datenarchiv/ Journal]. Genereller Ressourcentyp. Identifikator [d. h. DOI, PURL, ...].

Beispiel:

- Katarzyna Biernacka, Dominika Dolzycka, Kerstin Helbig, Petra Buchholz (2019): Train-the-Trainer Konzept zum Thema Forschungsdatenmanagement, Version 2.0. Zenodo. Datensatz. <https://doi.org/10.5281/zenodo.2581292>.

FORSCHUNGSDATEN RECHTSSICHER WEITERVERBREITEN

Rechte und Pflichten aus CC-Lizenzen ([Beispiel CC BY 4.0 International](#)):

Sie dürfen:

- Teilen: in jedwedem Format/Medium vervielfältigen und weiterverbreiten
- Bearbeiten: remixen, verändern und darauf aufbauen, für beliebige Zwecke, sogar kommerziell

Unter folgenden Bedingungen:

- Namensnennung: Zitiervorschläge nutzen, angemessene Urheber- und Rechteangaben machen, einen Link zur Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden
- Keine weiteren Einschränkungen, die anderen rechtlich irgendetwas untersagen, was die Lizenz erlaubt.

8. DATENMANAGEMENTPLÄNE

WAS SIND DATENMANAGEMENTPLÄNE?

Der Datenmanagementplan (DMP) beschreibt:

- die Entstehung/Erhebung der Daten
 - die Datendokumentation
 - den Datenzugang
 - die Verantwortlichkeiten im FDM
 - die Datenarchivierung
 - die Modalitäten der Datenveröffentlichung
- ... über den gesamten Lebenszyklus hinweg.
- DMPs sind dynamische, projektbegleitende Dokumente, regelmäßige Aktualisierungen sind erforderlich → **Living Documents**

BESTANDTEILE EINES DATENMANAGEMENTPLANS

- Administrative Informationen (Projektname, Datenurheber, weitere Mitwirkende, Kontakt, Förderprogramm usw.)
- Projekt- und Datensatzbeschreibung
- Datentypen, -formate, -umfang, Datenerhebung/Methodik
- Angaben Datenbenennung, zu Metadaten und Standards
- Datenaustausch, Zugriffsregelungen, Veröffentlichung und Nachnutzung
- Archivierung und Sicherung der Daten
- Verantwortlichkeiten
- Kosten

Der Umfang kann zwischen wenigen Absätzen und mehreren Seiten variieren.

WARUM EINEN DMP SCHREIBEN?

- verbindliche Grundlage für Umgang mit FD im Forschungsprozess
- erleichtert Verständnis und Bereitstellung der (eigenen) Daten
- Abstimmung und Kooperation mit Projektpartnern
- legt Verantwortlichkeiten fest
- regelt Zugriffsrechte
- gewährleistet sichere Datenspeicherung und Backups
- hilft bei Kosteneinschätzung
(Kostenkalkulation im FDM: Handreichung Kostenbeantragung)
- **kann/muss Bestandteil eines Förderantrags sein**

ANFORDERUNGEN DER FORSCHUNGSFÖRDERER

Förderer	Forderung	Abgabe bei Antrag	Inhalt	Updates
Horizon 2020	DMP	Nein. Erster Plan innerhalb der ersten 6 Projektmonate	Inhalte des Horizon 2020 Template	Ja, falls signifikante Änderungen auftreten und zum Projektende
Deutsche Forschungsgemeinschaft	Angaben zum Umgang mit Forschungsdaten	Ja	Inhalte der Leitlinie zum Umgang mit Forschungsdaten	Nein
BMBF	DMP erforderlich in Abhängigkeit vom Programm	Ja	Inhalt hängt vom jeweiligen Programm ab	Nein
BMBF Bildungsforschung	DMP	Ja	Inhalte laut Checkliste	Ja

DATENMANAGEMENTPLÄNE –TOOLS

- [DMP online](#)
- [RDMO](#) – Research Data Management Organizer
- [DMP Tool](#)

- [Muster DMPs](#) für Horizon 2020, DFG und BMBF auf der Webseite der HU Berlin
- [Research Data Managment Hub](#) (Schweiz) – Checkliste, Richtlinien, Beispiele (SNSF)
- [ETH Zürich](#)

MEHR WISSEN?

Besuchen Sie unseren Vertiefungskurs in der Villa Tillmanns:

„Datenmanagementpläne erstellen“

- Anmeldung über die Kompetenzschule der Research Academy Leipzig
- Onlinematerial zum Kurs und weiteren relevanten Themen wird auf unserer Webseite bereitgestellt: [Forschungsdatenmanagement an der Uni Leipzig](#)

FRAGEN? WIR BERATEN SIE GERN!

Zentraler Kontakt:

forschungsdaten@uni-leipzig.de

Zuständigkeiten:

- Dezernat 1: Forschung und Transfer:
 - Erstberatung, Fördermittel beantragen, DMP-Erstellung, Schulungen
- Universitätsbibliothek:
 - Veröffentlichung von Forschungsdaten, PIDs, Lizenzen, Nachnutzung
- Universitätsrechenzentrum:
 - Datensicherheit, Storage, Backup, Langzeitarchivierung, HPC, Big Data



UNIVERSITÄT LEIPZIG

Dezernat 1: Forschung und Transfer

- Pia Voigt (pia.voigt@zv.uni-leipzig.de, Telefon: 97-35075)
- Dr. Barbara Weiner (barbara.weiner@zv.uni-leipzig.de, Telefon: 97-35071)

Universitätsbibliothek

- Dr. Stephan Wünsche (wuensche@ub.uni-leipzig.de, Telefon: 97-30564 oder 97-35074)

Universitätsrechenzentrum

- Sebastian Frericks (sebastian.frericks@uni-leipzig.de)
- Dr. Stefan Kühne (stefan.kuehne@uni-leipzig.de, Telefon: 97-33303)
- Dr. Romy Elze (romy.elze@uni-leipzig.de, Telefon: 97-32304)
- Clemens Hoffmann (hoffmann@informatik.uni-leipzig.de, Telefon: 97-32347)



UNIVERSITÄT
LEIPZIG

Diese Präsentation wurde erstellt unter Nutzung von:

Dolzycka, Dominika; Biernacka, Katarzyna; Helbig, Kerstin; Buchholz, Petra (2019): Train-the-Trainer Konzept zum Thema Forschungsdatenmanagement, Version 2.0, Berlin, <https://doi.org/10.5281/zenodo.2581292> .

– CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>). Es wurden Änderungen vorgenommen.



Der Text dieser Präsentation steht unter der Lizenz
Creative Commons Attribution 4.0 International (CC BY 4.0).