



Forschungsdatenmanagement in der Psychologie

Fachspezifisches Train-the-Trainer-Konzept

Fachspezifische Informationen

Version 2.0


Sven Paßmann


Sibylle Söring

Freie Universität Berlin

Berlin 2023

DIE AUTORINNEN UND AUTOREN

Sven Paßmann  0000-0001-9251-8269

Sibylle Söring  0000-0002-1698-3289

HERAUSGEBER

Freie Universität Berlin mit Projektschwerpunkt „Fachspezifisches Train-the-Trainer-Modul“ (AP 5).

Diese Publikation wurde im Rahmen des Verbundprojekts „FDNext“ durch die Deutsche Forschungsgemeinschaft gefördert – Projektnummer: 429 828 830

Gefördert durch



IMPRESSUM

„Forschungsdatenmanagement in der Psychologie: Fachspezifisches Train-the-Trainer-Konzept. Fachspezifische Informationen. Version 2“ von Sven Paßmann und Sibylle Söring ist lizenziert unter der Creative Commons Attribution 4.0 International (CC BY 4.0).



ZITATIONSVORSCHLAG

Paßmann, Sven; Söring, Sibylle. (2023). Forschungsdatenmanagement in der Psychologie: Fachspezifisches Train-the-Trainer-Konzept (Version 2). DOI: 10.5281/zenodo.8113417.

DOI

<https://doi.org/10.5281/zenodo.8113417>



Inhaltsverzeichnis

Einleitung.....	6
Danksagung	8
Forschungsdaten (Einheit 3).....	9
1. Forschungsdaten: Definition	9
2. Spezifika von Forschungsdaten in der Psychologie.....	10
3. Forschungsdatenlebenszyklus	10
Forschungsdatenmanagement (Einheit 4).....	12
1. Motivation.....	12
2. Forschungsdatenmanagement.....	13
a) Was ist Forschungsdatenmanagement?	13
b) Aufgaben des Forschungsdatenmanagements.....	14
c) Die FAIR-Prinzipien	15
3. (Psychologie-spezifische) Forschungsdaten-Policies, Leitlinien und Empfehlungen.....	16
a) Was ist eine Forschungsdaten-Policy?	16
b) Psychologie-spezifische Empfehlungen	16
c) Leitlinien von Forschungsförderern	17
d) Institutionelle Policies und Leitlinien in der Psychologie	18
e) Projektspezifische Policies	19
Rechtliche Aspekte (Einheit 5).....	21
1. Ebenen der Regelung.....	21
2. Datenschutz und Persönlichkeitsrecht	22
3. Anonymisierung der in der Psychologie relevanten Daten.....	23
a) Anonymisierung qualitativer und quantitativer Daten	23
b) Anonymisierung von Daten aus bildgebenden Verfahren.....	24
c) Anonymisierung anderer Forschungsdaten und weitere Schutzmaßnahmen	24
4. Urheberrecht und verwandte Schutzrechte	25
5. Vertragsrecht	26
Datenmanagementplan (Einheit 6)	27
1. Motivation.....	27
2. Was ist ein Datenmanagementplan?.....	27
3. Bestandteile eines Datenmanagementplans	28
4. Anforderungen der Forschungsförderer	28
5. Muster und Werkzeuge für die Psychologie	29
Organisation von Daten (Einheit 7).....	32
1. Motivation.....	32
2. Verzeichnisstruktur	32
3. Dateibenennung	33
4. Umbenennen mehrerer Dateien gleichzeitig	34
5. Versionskontrolle	34
Dokumentation und Metadaten (Einheit 8)	36

1.	Motivation.....	36
2.	Datendokumentation.....	36
3.	Codebooks/Data Dictionaries für die Psychologie.....	37
4.	Was sind Metadaten?	38
5.	Metadatenstandards	38
6.	Psychologie-spezifische Metadatenstandards	39
7.	Thesaurus, Normdaten und Klassifikationen in der Psychologie	39
8.	Elektronische Laborbücher (ELB) in der Psychologie	41
Zugriffssicherheit (Einheit 9)		43
1.	Motivation.....	43
2.	Verschlüsselung	43
3.	Überprüfen der Nutzbarkeit	43
4.	Passwortschutz und Zugriffsrechte	44
Speicherung und Backup (Einheit 12)		45
1.	Motivation.....	45
2.	Datenspeicherung.....	45
3.	Backup-Varianten	46
4.	3-2-1-Regel: Hinweise für ein sicheres Backup.....	47
Langzeitarchivierung (Einheit 13)		48
1.	Motivation.....	48
2.	Begriff „Langzeitarchivierung“	48
3.	Nachhaltige Dateiformate in der Psychologie.....	48
4.	Langzeitarchive und deren Anforderungen	49
Infrastrukturen für das Forschungsdatenmanagement (Einheit 14)		51
1.	Nationale und regionale Angebote für das Forschungsdatenmanagement	51
2.	Institutionelles Forschungsdatenmanagement	52
3.	Forschungsdatenmanagement-Infrastrukturen	52
Publikation von Forschungsdaten (Einheit 15)		53
1.	Ziele und Motivation.....	53
2.	Leitfragen für die Datenauswahl	53
3.	Publikationswege.....	54
4.	Repositorien.....	55
a)	Fachspezifische Repositorien.....	55
b)	Generische Repositorien	55
c)	Institutionelle Repositorien	56
5.	Auswahl eines Repositoriums.....	56
6.	Lizenzen.....	57
7.	Persistente Identifier	59
a)	Digital Object Identifier (DOI).....	59
b)	Open Researcher and Contributor ID (ORCID).....	60
c)	Research Activity ID (RAiD)	60
8.	Zugriffsklassen/Agreements	60
Literaturverzeichnis		62

Weiterführende Ressourcen	65
Einheit 3	65
Einheit 4	65
Einheit 5	65
Einheit 6	66
Einheit 7	66
Einheit 8	67
Einheit 9	67
Einheit 12	67
Einheit 13	67
Einheit 15	67

Abkürzungsverzeichnis

BMBF	Bundesministerium für Bildung und Forschung
CC	Creative Commons
DFG	Deutsche Forschungsgemeinschaft
DINI	Deutsche Initiative für Netzwerkinformation
DMP	Datenmanagementplan
EC	European Commission (Europäische Kommission)
FD	Forschungsdaten
FDM	Forschungsdatenmanagement
PPTX	PowerPoint-Präsentation
TN	Teilnehmende
UAG	Unter-Arbeitsgruppe
WL	Workshopleitende

Einleitung

Das vorliegende Konzept eines Train-the-Trainer Workshops zum Thema Forschungsdatenmanagement für Psycholog*innen (Paßmann & Söring, 2023) entstand im Rahmen des von der Deutschen Forschungsgemeinschaft (DFG) im Zeitraum 2020-2023 geförderten Projektes FDNNext. Es richtet sich zuvorderst an Multiplikator*innen aus dem Fachbereich Psychologie, denen es ein fachbezogenes Instrument zur Kompetenzentwicklung im Forschungsdatenmanagement (FDM) an die Hand gibt. Zu diesem Zweck wurden zahlreiche, für die Psychologie relevanten Aspekte des FDM zusammengetragen und in das generische FDM-Train-the-Trainer-Konzept (Birnacka et al., 2021) des durch das BMBF geförderte Vorgänger-Projekt FDMentor¹ integriert.

Um den unterschiedlichen Erfahrungsstufen der Trainer*innen in FDM und/oder Didaktik Rechnung zu tragen, wurden zwei Hauptdokumente (FDNNext_TtTKonzept_01_Fachinformationen_V2, FDNNext_TtTKonzept_02_DidaktikLehrdrehbücher_V2) sowie eine Methodensammlung (FDNNext_TtTKonzept_03_MethodenMaterialien_V2) erstellt, die unabhängig voneinander, aber auch ergänzend genutzt werden können (siehe Tabelle 1). Es empfiehlt sich jedoch, wegen der starken Verschränkung (die beiden Hauptdokumente verweisen stets aufeinander, siehe unten), immer alle Dokumente für ein besseres Verständnis der Fach- als auch Didaktikebene vorliegen zu haben.

Tabelle 1: Übersicht der Dokumente.

Dokument	Inhalt	Zielgruppe
„FDNNext_TtTKonzept...“		
_01_Fachinformationen_V2	Grundlegende Aspekte des FDM mit fachspezifischen Informationen für Psycholog*innen	unerfahren in FDM
_02_DidaktikLehrdrehbücher_V2	Fachliche Informationen zu Didaktik; Ziele der Lerneinheiten; Übersicht der jeweils verwendeten Methoden und Übungen; Lehrdrehbücher	unerfahren in Didaktik
_03_MethodenMaterialien_V2	Übersicht und Beschreibung der im Workshop verwendeten Methoden; Materialsammlung	
_Arbeitsmaterialien_FDNNext_TtTKonzept.zip	Sammlung verschiedenster Materialien zur Ausgestaltung eines Workshops (Präsentationsfolien, Lehrdrehbücher, Sammlung fachspezif. Informationen)	

Das vorliegende Dokument „FDNNext_TtTKonzept_01_Fachinformationen_V2“ bezieht sich dabei ausschließlich auf die fachspezifischen Aspekte des FDM (**Lerneinheiten 3-9 und 12-16**) und kann von Trainer*innen mit Erfahrungen in Didaktik als inhaltliche Grundlage für die Workshop-Planung genutzt werden. Jede Lerneinheit referiert dabei auf einen spezifischen Aspekt des FDM und enthält

¹ FDMentor Projektseite, Zugriff am 23.11.2023, <https://www.forschungsdaten.org/index.php/FDMentor>; Veröffentlichte Materialien im Rahmen des Projektes auf Zenodo, Zugriff am 23.11.2023, <https://zenodo.org/communities/fdmentor/?page=1&size=20>.

Hintergrundwissen sowie weitergehende Informationen zu den einzelnen Themengebieten. Die hinter jedem Titel einer Lerneinheit in Klammern vermerkte Nummer verweist auf das zweite Dokument „FDNext_TtTKonzept_02_DidaktikLehrdrehbücher_V2“ und den darin enthaltenen Lehrdrehbüchern für die jeweilige Lerneinheit, welche auch mögliche, im Workshop zu erreichende Ziele definiert.

Thematische Einheiten, die sich ausschliesslich auf didaktische Komponenten beziehen (**Lerneinheiten 1-2, 10-11 und 17-21**), werden in dem Dokument „FDNext_TtTKonzept_02_DidaktikLehrdrehbücher_V2“ behandelt. Besonderer Fokus sind hier die sozialen, didaktischen und organisatorischen Komponenten eines Workshops inkl. der zu deren Verständnis notwendigen Hintergrundinformationen. Für die konkrete Ausgestaltung einzelner Themengebiete enthält das Dokument exemplarische Lehrdrehbücher für einzelne Lerneinheiten zur Orientierung und definiert auch hier mögliche, im Workshop zu erreichende Ziele.



Abbildung 1: Komponenten des Train-the-Trainer-Konzepts. Die angegebenen Nummern entsprechen den Einheitsnummern im jeweiligen Dokument (FI, Fachinformation; DI, didaktische Informationen).

Die Nachnutzung der Konzept-Dokumente soll die Planung und Durchführung eines insgesamt zweitägigen (Online-)Workshops ermöglichen, dessen Inhalt auf sieben Zeitstunden pro Tag verteilt ist. Jeder Tag enthält zudem jeweils eine Mittagspause (eine Stunde) und zwei kürzere Pausen (von je 15 min). Dabei sind die einzelnen Lerneinheiten so konzipiert, dass diese sowohl eigenständig als auch in thematischen Gruppen genutzt werden können, z. B. „Forschungsdaten“ zusammen mit „Forschungsdatenmanagement“. Aber auch „Rechtliche Aspekte“ und „Zugriffssicherheit“, als auch „Speicherung und Backup“ und „Langzeitarchivierung“ bieten sich als thematische Einheiten an. Am

Ende des Dokuments finden sich darüber hinaus weiterführende Lektürehinweise und das Literaturverzeichnis.

Die Version 1 des fachspezifischen Train-the-Trainer Konzeptes wurde für die vorliegende Version um folgende Informationen erweitert:

- Definitionen der DGPs hinsichtlich Roh-, Primär- und Sekundärdaten
- Befunde aus Befragungen unter Psycholog*innen, die die häufigsten Fehler im Umgang mit Forschungsdaten aufzeigen
- projektspezifische Policies
- methodische Ansätze der Anonymisierung von Forschungsdaten
- Psychologie-spezifische Metadaten-Standards
- Psychologie-spezifische ELNs
- nachhaltige Datenformate
- institutionelle Repositorien
- Definition von Langzeitarchiven
- Literaturverzeichnis

Danksagung

Das hier vorliegende Konzept basiert auf dem „Train-the-Trainer Konzept zum Thema Forschungsdatenmanagement (Version 4)“ (Biernacka et al., 2021) des BMBF-geförderten Vorgänger-Projektes „FDmentor“² (Laufzeit des Projektes: 2017-2019; Autor*innen: Katarzyna Biernacka, Petra Buchholz, Sarah Ann Danker, Dominika Dolzycka, Claudia Engelhardt, Kerstin Helbig, Juliane Jacob, Janna Neumann, Carolin Odebrecht, Britta Petersen, Benjamin Slowig, Ute Trautwein-Bruns, Cord Wiljes, Ulrike Wuttke); zahlreiche generische Inhalte konnten übernommen werden. Den Autor:innen gilt unser Dank für die umfangreichen Vorarbeiten, auf denen aufgebaut werden konnte.

² FDmentor Projektwebseite, Zugriff am 23.11.2023, <https://www.forschungsdaten.org/index.php/FDmentor>; Veröffentlichte Materialien im Rahmen des Projektes auf Zenodo, Zugriff am 23.11.2023, <https://zenodo.org/communities/fdmentor/?page=1&size=20>.

Forschungsdaten (Einheit 3)³

1. Forschungsdaten: Definition

In der Vergangenheit haben sich – je nach Fachdisziplin und Projektkontext - unterschiedliche Definitionen von Forschungsdaten etabliert. Kindling/Schirmbacher (2013) betrachten digitale Forschungsdaten als „[...] alle digital vorliegenden Daten [...], die während des Forschungsprozesses entstehen oder ihr Ergebnis sind“ (Kindling & Schirmbacher 2013). Laut der Deutschen Forschungsgemeinschaft (DFG) zählen zu Forschungsdaten „[...] u. a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Umfragedaten, Objekte aus Sammlungen oder Proben, die in der wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden. Methodische Testverfahren, wie Fragebögen, Software und Simulationen können ebenfalls zentrale Ergebnisse wissenschaftlicher Forschung darstellen und sollten daher ebenfalls unter den Begriff Forschungsdaten gefasst werden“⁴.

Die Technische Universität Berlin fasst diesen Begriff noch weiter⁵. Forschungsdaten sind hier „alle Informationen (unabhängig von Form oder Darstellung), die während eines Forschungsprozesses entstehen oder sein Ergebnis sind, einschließlich der Informationen, die zur Nachvollziehbarkeit und Reproduktion der Ergebnisse notwendig sind.“ In der Aufzählung werden u. a. „Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Objekte aus Sammlungen oder Proben, Umfragen und Interviews, aber auch Skripte, Zeitschriebe, Berechnungen, Software und Code“ als spezifische Forschungsdaten erwähnt. Die Freie Universität Berlin bestimmt in ähnlicher Weise „Informationen als Forschungsdaten, die die Dokumentation, Nachvollziehbarkeit und – abhängig vom Forschungsgebiet – Reproduzierbarkeit der Ergebnisse gewährleisten (Metadaten)“⁶.

In ihrer fachspezifischen Empfehlung von 2016 bezeichnet die Deutsche Gesellschaft für Psychologie (DGPs) Forschungsdaten zunächst noch als Primärdaten, welche „vollkommen unbearbeitete (d. h. untransformierte, nicht-aggregierte etc.) quantitative und qualitative Daten, zum Beispiel bei Experimenten alle manipulierten und gemessenen Variablen für jeden Experimentaldurchgang jeder Person“ sind. In den nachfolgenden Überarbeitungen wird diese Definition erweitert und umfasst neben den Primär- nunmehr auch Roh- und Sekundärdaten (Gollwitzer et al., 2020):

Als **Rohdaten** werden die Ursprungsaufzeichnungen bezeichnet, d.h. die erste „nicht-flüchtige“⁷ Form der Daten (z. B. Kreuze auf einem Papierfragebogen, Zeichnungen, Audio- oder Videoaufnahmen, Blickbewegungsmessungen, neuro- oder peripher-physiologische Aufzeichnungen wie z. B. EEG, Herzrate). Sie liegen bereits oft, aber nicht notwendigerweise, in digitalisierter Form vor.

Als **Primärdaten** werden die Daten bezeichnet, die eine erste Übertragung der Rohdaten in ein digitales Format darstellen (z. B. den Code „1“ für eine Ja-Antwort in einem Fragebogen; d.h. untransformierte, nicht-aggregierte etc. in digitaler Form vorliegende quantitative oder qualitative Daten). Wenn die Antworten der Probandinnen und Probanden mittels einer Software oder im Rahmen

³ Die Vergabe der Nummern für jede Einheit beziehen sich auf die Lerneinheiten im Teildokument „FDNext_TtTKonzept_02_DidaktikLehrdrehbücher_V2“ des fachspezifischen TtT-Konzeptes.

⁴ Deutsche Forschungsgemeinschaft. „Leitlinien zum Umgang mit Forschungsdaten.“ Zugriff am 23.11.2023, http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf.

⁵ Forschungsdatenpolicy der Technischen Universität Berlin. Zugriff am 23.11.2023, <https://www.tu.berlin/arbeiten/wichtige-dokumente/richtlinien-leitlinien/forschungsdaten-policy>.

⁶ Definition „Forschungsdaten“ der Freien Universität Berlin. Zugriff 23.11.2023, <https://www.fu-berlin.de/sites/forschungsdatenmanagement/ueber-forschungsdaten/forschungsdaten/index.html>.

⁷ (Nicht-)Flüchtiges Datum ist ein Begriff aus der IT-Forensik. Als nicht-flüchtig werden alle schriftlich oder digital festgehaltenen Informationen bezeichnet, welche in einem Speicher ohne weitere Energiezufuhr (z.B., wenn Energie zur Aufrechterhaltung der Speicherung benötigt wird) dauerhaft Bestand haben. Siehe auch „Leitfaden IT-Forensik“ des Bundesamtes für Sicherheit in der Informationstechnik. Zugriff am 23.11.2023, https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Cyber-Sicherheit/Themen/Leitfaden_IT-Forensik.pdf?__blob=publicationFile&v=2.

eines Online-Survey erhoben und sofort digital gespeichert wurden, sind Rohdaten und Primärdaten auch äquivalent.

Sekundärdaten stellen Transformationen von Variablen, Aggregationen einzelner Beobachtungen, Berechnungen von Skalenwerten oder von ereigniskorrelierten Potentialen durch Mittelung bzw. der Ausschluss von Werten dar, von denen angenommen wird, dass sie die Robustheit der Analyseergebnisse beeinträchtigen (z.B. Extrem- und Ausreißerwerte, ungültige Fälle).

2. Spezifika von Forschungsdaten in der Psychologie

Aufgrund des Forschungsobjekts „Mensch“ und seines durch kognitive Prozesse geleiteten Verhaltens sind die in der Psychologie erhobenen Daten sehr häufig durch einen starken Personenbezug gekennzeichnet. Dies führt dazu, dass die Erhebung der Daten von Personen immer auch mit dem Recht dieser Personen auf den Schutz ihrer Daten kollidiert, die in der Charta der Europäischen Union verankert ist⁸. Potentielle Teilnehmende einer Studie müssen der Erhebung und Verarbeitung ihrer Daten deshalb durch eine informierte Einwilligung uneingeschränkt zustimmen (siehe **Einheit 5 "Rechtliche Aspekte"**, S. 21).

Bei einigen Datenformaten (bspw. Video- und Audioaufnahmen, Genetik, M/EEG oder MRT-Daten) lässt sich zudem der Personenbezug auch bei anonymisierter Erhebung nicht gänzlich von dem jeweiligen Individuum trennen. Das Gehirn, und ebenso die Gehirnaktivität, könnten mit entsprechend weit entwickelten Programmen aufgrund der individuellen Lebensgeschichte, welche sich in spezifischen Verknüpfungen zwischen den Gehirnarealen widerspiegelt, eindeutig einer Person zuordnen. MRT-Aufnahmen enthalten zudem neben den Daten von Interesse auch Informationen über Knochenstruktur und Bindegewebe, welche eine eindeutige Rekonstruktion des Gesichtes einer Person ermöglichen (Schimke et al., 2011). Darüber hinaus besitzen menschliche Stimmen personenspezifische Merkmale⁹ wie Stimmhöhe, Stimmfrequenz usw., bestimmt durch die individuelle Größe und Form des Kehlkopfes oder ethnische Zugehörigkeit. Videoaufnahmen zeichnen neben dem Gesicht oft auch spezifische Körpermerkmale, wie z. B. Tattoos, Frisur, Körperbewegungen (vor allem im klinischen Kontext) usw. auf.

Auch wenn Methoden existieren, die eine nachfolgende Zuordnung dieser Daten zu konkreten Personen erschweren (siehe **Einheit 5 "Rechtliche Aspekte"**, S. 21), kann eine komplette Anonymisierung nicht immer gewährleistet werden, was bei einer geplanten Publikation der Daten beachtet werden muss (siehe **Einheit 15 "Publikation von Daten"**, Punkt 2, S. 53). Diese, die Psychologie betreffenden, Besonderheiten der Daten müssen somit im Sinne des Forschungsdatenmanagements reflektiert und fachspezifisch umgesetzt werden.

3. Forschungsdatenlebenszyklus

Der Lebenszyklus von Forschungsdaten stellt die Schritte und Aspekte eines Forschungsvorhabens in Bezug auf das Forschungsdatenmanagement dar. Der Lebenszyklus nach [forschungsdaten.info](https://www.forschungsdaten.info)¹⁰ basiert auf der vom UK Data Service erstellten Version¹¹ und beinhaltet die Planung, Erhebung,

⁸ EU-Charta, Recht auf Privatsphäre (Art. 7) und den Schutz der Daten (Art. 8). Zugriff am 23.11.2023, <http://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:12012P/TXT&amp;from=DE>.

⁹ Zhang, Z. (2016). Mechanics of human voice production and control. *J. Acoust. Soc. Am.*, 140(4), 2614-2635.

¹⁰ Der Datenlebenszyklus, Zugriff am 23.11.2023,

<https://www.forschungsdaten.info/themen/informieren-und-planen/datenlebenszyklus/>.

¹¹ Data Lifecycle vom UK Data Service, Zugriff am 23.11.2023, <https://ukdataservice.ac.uk/learning-hub/research-data-management/>.

Aufbereitung und Analyse, Teilen und Publizieren, Archivierung sowie die Nachnutzung von Daten (siehe Abbildung 2).

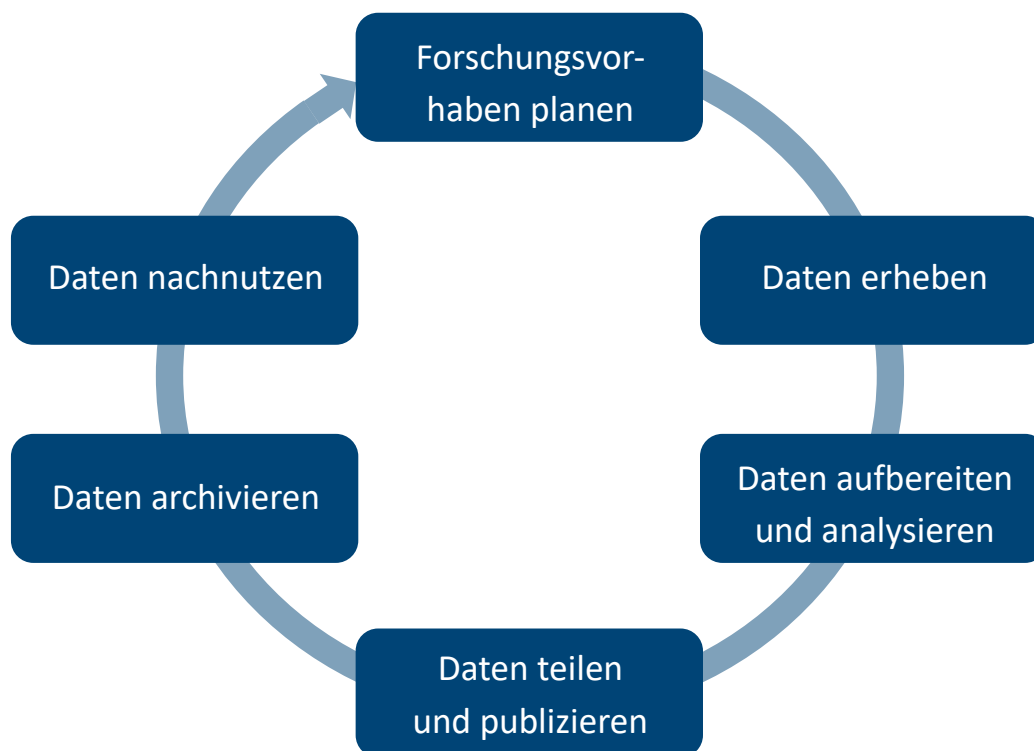


Abbildung 2: Forschungsdatenlebenszyklus nach [forschungsdaten.info](https://www.forschungsdaten.info/).

Forschungsdatenmanagement (Einheit 4)

1. Motivation

Der Kodex „Leitlinien zur Sicherung guter wissenschaftliche Praxis“ der DFG¹² formuliert verschiedene Aspekte der wissenschaftlichen Integrität, die durch ein nachhaltiges, den sogenannten FAIR-Kriterien (siehe unten **Punkt 2c**; Wilkinson et al., 2016) entsprechendes Forschungsdatenmanagement (FDM) adressiert und umgesetzt werden können. Ziel ist es, Forschungsdaten langfristig und personenunabhängig zugänglich, nachnutzbar und nachprüfbar zu halten¹³, in dem es Aspekte der Qualitätssicherung (in Bezug auf Erhebung, Verarbeitung, Auswertung und Archivierung bis hin zur Veröffentlichung von Daten) einbezieht. FDM adressiert zudem die rechtlichen und ethischen Rahmenbedingungen eines Forschungsprozesses (Recht auf Forschungsfreiheit, Nutzungsrecht von Daten, Ethikvoten, Datenschutz). Instrumente wie der Datenmanagementplan regeln darüber hinaus die datenbezogenen Verantwortlichkeiten innerhalb eines Projekts und dokumentieren verwendete Datentypen, Software, Speicherung, Archivierung und Zugang zu den im Vorhaben generierten bzw. beforchten Daten.

Nicht zuletzt durch die Replikationskrise erlangte FDM in der Psychologie¹⁴ besondere Relevanz als notwendiger Bestandteil einer validen Forschung durch das Nachvollziehen der Ergebnisse und damit deren Reproduzierbarkeit. Kovacs et al. (2021) konnten anhand von Publikationen aus der Psychologie im Verlag Elsevier¹⁵ die uneindeutige Datenbenennung (N = 86 von N = 506), Fehler in der Versionskontrolle (N = 62 von N = 506), Fehler im Datentransfer (N = 35 von N = 506) sowie Verlust von Daten und Dokumentationen (N = 30 von N = 506) als häufigste Mängel im Forschungsprozess nachweisen. Als häufigster Grund für diese Fehler wurde dabei vor allem schlechte Projektvorbereitung oder -management (wie z. B. schlechte oder fehlende Planung, Standards, Kommunikation zwischen Forschenden, Zeitmanagement, usw.) bzw. persönliche Schwierigkeiten (z. B. Sorglosigkeit, Unaufmerksamkeit, Kontrollmangel, Übermut, usw.) identifiziert.

Diese Umstände erhöhten das Bewusstsein für die Relevanz eines entsprechend systematisierten FDM, das mittel- und langfristig die Transparenz, Nachhaltigkeit, Interoperabilität und Zugänglichkeit datenbasierter Forschung zum Ziel hat. In einer Studie, die unter Psycholog*innen deren Motivation zur professionalisierten Umsetzung von FDM erfragte, wurde das Vermeiden des Datenverlustes am höchsten bewertet, gefolgt von dem Anspruch, die Replikation von Forschungsergebnissen zu ermöglichen und Transparenz in der Forschung herzustellen (Borghi & Gulick, 2021).

Aus diesem Grund:

- geben sich z. B. größere Forschungsprojekte, die als Verbund mehrere wissenschaftliche Institute auf nationaler und internationaler Ebene umfassen, eine für alle gleichermaßen gültige Leitlinie im Umgang mit Forschungsdaten;
- verabschieden wissenschaftliche Institute oder Fachbereiche entsprechende Policies im Umgang mit Forschungsdaten an deren Einrichtungen;
- unterstützen Forschungsförderer oder spezifische Förderprogramme diese Bestrebungen, indem sie ein nachhaltiges FDM während der Projektlaufzeit als auch darüber hinaus fordern.

¹² Deutsche Forschungsgemeinschaft. Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex. Zenodo, 2019, <http://doi.org/10.5281/zenodo.3923602>.

¹³ Forschungsdatenmanagement, Zugriff am 23.11.2023, <https://forschungsdaten.info/themen/informieren-und-planen/was-ist-forschungsdatenmanagement/>.

¹⁴ Positionspapier der Psychologie-Fachschaften Konferenz zur Replikationskrise in der Psychologie. Zugriff am 23.11.2023, <https://psyfako.org/wp-content/uploads/30-PsyFaKo-STN-Open-Science-Kulturwandel-zu-einer-offenen-Wissenschaft.pdf>.

¹⁵ Verlag Elsevier, Zugriff am 23.11.2023, <https://www.elsevier.com/de-de>.

2. Forschungsdatenmanagement

a) Was ist Forschungsdatenmanagement?

Forschungsdaten zählen zu den zentralen Ressourcen der Wissenschaft, was einem systematischen und verantwortungsvollen Umgang mit ihnen eine grundlegende Bedeutung beimisst. Im Rahmen des FDM werden „die eigenen Arbeitsprozesse, die die Erzeugung von und den Umgang mit Forschungsdaten betreffen, möglichst effizient und zielorientiert“ (Meyermann, 2012) über den gesamten Zeitraum des Projektes gesteuert. Damit begleitet FDM den gesamten Forschungsprozess von der ersten Planung bis hin zur Archivierung, Nachnutzung oder Löschung der Daten.

Eine durch ein standardisiertes FDM umgesetzte Strategie im Umgang mit Daten hilft im Forschungsprozess nicht nur Daten zu verwalten, sondern steuert und vereinheitlicht den Umgang mit ihnen auch. Ein standardisiertes FDM verringert sowohl den zukünftigen Aufwand bei eigener und fremder Nachnutzung der Daten als auch das Risiko eines Datenverlustes. Zudem stehen bestimmte Förderquellen nur Projekten mit einem dezidierten FDM-Plan zur Verfügung.

Durch das Verschriftlichen einer FDM-Strategie entsteht ein erster, sogenannter Datenmanagementplan. Er beinhaltet sowohl technische als auch organisatorische, strukturelle, rechtliche und ethische Aspekte des Umgangs mit Daten für die Dauer eines Projektes. Aber auch weiterreichende Aspekte, wie beispielsweise die Archivierung und Zugänglichmachung nach Projektende oder die eigene oder fremde Nachnutzung der Daten sollten hier berücksichtigt werden.

Ausgehend vom Forschungsdatenlebenszyklus beschreibt ein sogenannter Forschungsdaten-Workflow – mehr noch als ein Datenmanagementplan – die einzelnen Verarbeitungsschritte der Forschungsdaten in Abhängigkeit von der gewählten Software und den benötigten Infrastrukturen und Services. Dabei wird eine prozessorientierte Perspektive eingenommen, die es Datenerstellenden erlaubt, auch Datenübergaben und -konvertierungen, die bspw. zwischen Datenverarbeitung und -analyse notwendig sind, konkret abzubilden. In einem Workflow werden auch Verantwortlichkeiten (Rollen und Akteure) festgehalten.

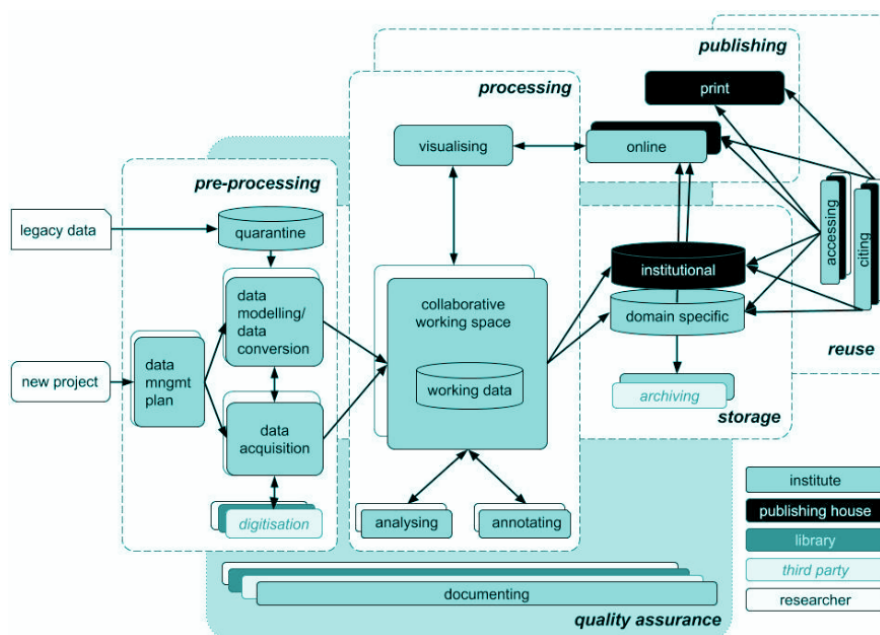


Abbildung 3: Beispielhafter Datenmanagementworkflow nach Wissik & Āurĉo 2015 (Wissik & Āurĉo, 2016)

Im Projekt COGITATE¹⁶ wird beispielhaft ein solcher, auf den FAIR-Prinzipien basierender Workflow¹⁷ in das Projekt implementiert, und zur Nachnutzung¹⁸ veröffentlicht.

Die genannten Arbeitsschritte und Instrumente im Rahmen des Forschungsdatenmanagements haben dabei folgende Ziele:

- Standardisierung der Datenpflege:
 - Nachvollziehbarkeit der Datenerhebung und –verarbeitung (z. B. keine verteilte Datenablage in unterschiedlichen Versionen auf verschiedenen Rechnern) sowie der daraus resultierenden Ergebnisse
 - Erleichterung der Zusammenarbeit zwischen Mitarbeitenden innerhalb eines Projektes bzw. Projektpartnern
 - ein gemäß den FAIR-Prinzipien (Wilkinson et al., 2016; siehe unten, **Punkt 2c**) schnelleres Auffinden von Daten, bspw. durch aussagekräftige Benennung
 - Sicherstellung der plattformunabhängigen Datenverarbeitung durch Nutzung freier Datenformate, den Transfer der Daten in zukünftige eigene oder fremde Projekte oder die (halb-)automatische Verarbeitung durch Metadaten
 - Vorbeugung von Datenverlust, u. a. aufgrund defekter Hard- oder Software oder von Ursprungsversionen von Dateien
- Nachnutzbarkeit:
 - Weitergabe und Nachnutzung von Daten unter Verwendung von Einwilligungserklärungen
 - Ermöglicht optimierten Mitteleinsatz, bspw. Kostenersparnis durch Nachnutzung statt neuer Erhebung (Erhalt von Primär- und Sekundärdaten)
 - Wissenserhalt, da Daten unabhängig von Personen, Projekten oder Institutionen zugänglich sind
- Erfüllung von Auflagen der Drittmittelgeber
- Referenzierbarkeit (Piwowar et al., 2007)

b) Aufgaben des Forschungsdatenmanagements

Forschungsdatenmanagement umfasst sämtliche Schritte des Forschungsprozesses. Zentrale Aufgaben sind (Corti et al., 2019):

- Planung des Umgangs mit Forschungsdaten zu Beginn eines Forschungsprojektes sowie ggf. Darstellung der geplanten Maßnahmen in Förderanträgen
- Festlegen einer Ordnerstruktur und Dateinamenskonventionen
- Dokumentation von Forschungsdaten und Auszeichnung mit Metadaten
- Backup und Langzeitarchivierung von Forschungsdaten
- IT-Sicherheit und Zugriffsrechte für Forschungsdaten
- Langzeitarchivierung von Forschungsdaten
- Publikation von Forschungsdaten
- Auffinden und Nachnutzen bestehender Forschungsdaten
- Berücksichtigung von Datenschutz- und Urheberrecht im Umgang mit Forschungsdaten

¹⁶ Projekt COGITATE. Zugriff am 23.11.2023, <https://www.arc-cogitate.com/project>.

¹⁷ Implementierung eines FAIR-Workflows in einem Neuroscience Projekt. Zugriff am 23.11.2023, <https://datacite.org/implementing-fair-workflows-project/>.

¹⁸ Chen, X., & Jagerhorn, M. (2022). Implementing FAIR Workflows along the research lifecycle. *Procedia Computer Science*, 211, 83-92.

c) Die FAIR-Prinzipien

FORCE11, eine Gruppe Forschender und Mitarbeitenden aus Bibliotheken, Archiven, Verlagen und Forschungsförderern, hat 2016 Prinzipien zur Erschließung von Forschungsdaten verfasst (Wilkinson et al., 2016).

Diese sogenannten FAIR-Prinzipien umfassen vier Ziele¹⁹:

- **Findable** (Auffindbarkeit)
- **Accessible** (Zugänglichkeit)
- **Interoperable** (Interoperabilität, d. h. mit anderen Daten kompatibel und von möglichst vielen Analyse- und Visualisierungsapplikationen nutzbar)
- **Reusable** (Wiederverwendbarkeit)

Durch die Anwendung bzw. Umsetzung der FAIR-Prinzipien werden Daten und Metadaten menschen- und maschinenlesbar, wodurch sie effektiver auffindbar und nachnutzbar sind. Darüber hinaus sollten Daten und deren Metadaten so archiviert werden, dass sie mit Hilfe von Standard-Kommunikationsprotokollen leicht und langfristig von Mensch und Maschine abgerufen, heruntergeladen oder lokal genutzt werden können. Die Daten sollten in einer Form vorliegen, in der sie in einer (teil-)automatisierten Weise mit anderen Datensätzen ausgetauscht, interpretiert und kombiniert werden können. Eine gute Beschreibung der Daten und deren Metadaten – idealerweise in standardisierter Form (siehe **Einheit 8 „Dokumentation und Metadaten“, Punkt 5, S. 38**) – gewährleistet dabei die Wiederverwendung der Daten für zukünftige Forschung und den Vergleich mit anderen kompatiblen Datenquellen. Dazu zählen außerdem ordnungsgemäßes Zitieren der Daten und eine für Mensch und Maschine eindeutige Darstellung der Nachnutzungsbedingungen.

Das Hauptziel der FAIR-Prinzipien ist die Nachnutzbarkeit von Forschungsdaten (Boeckhout et al., 2018). Das bedeutet nicht, dass alle Daten uneingeschränkt offen und zugänglich sein müssen. Eine eingeschränkte Zugänglichkeit, beispielsweise aus datenschutzrechtlichen Gründen, widerspricht nicht den FAIR-Prinzipien, sofern dafür die Metadaten verfügbar und zugänglich gemacht worden sind. Ein sorgfältig, bereits von Beginn an geplantes Forschungsdatenmanagement ist eine der Grundlagen sowohl für FAIR wie auch für Open Data (eines der sechs Prinzipien von Open Science (Gallagher et al., 2019)), denn viele der für eine möglichst weitreichende Umsetzung dieser beiden Prinzipien notwendigen Schritte werden bereits sehr früh im Forschungsprozess angegangen. Darüber hinaus unterstützt das Forschungsdatenmanagement auch Aspekte, die durch FAIRness und Offenheit nicht notwendigerweise abgedeckt werden - etwa Langzeitarchivierung und gute Datenqualität (Higman et al., 2019). Verschiedene Tools²⁰ bieten online Fragenkataloge zur selbstständigen Überprüfung der FAIRness der eigenen Forschungsdaten an. Je höher die FAIRness und damit, je offener die Daten sind, desto wahrscheinlicher ist die Nachnutzung.

¹⁹ GO FAIR International Support and Coordination Office. „FAIR Principles.“ Zugriff am 23.11.2023, <https://www.go-fair.org/fair-principles/>.

²⁰ Bspw. FAIRsFAIR, FAIR Aware. Zugriff am 23.11.2023, <https://www.fairsfair.eu/fair-aware> oder Australian Research Data Commons, FAIR self-assessment tool. Zugriff am 23.11.2023, <https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/>. Kollektion verschiedener Tools. Zugriff am 23.11.2023, <https://fairassist.org#!/>.

3. (Psychologie-spezifische) Forschungsdaten-Policies, Leitlinien und Empfehlungen

a) Was ist eine Forschungsdaten-Policy?

Policies, Leitlinien und Empfehlungen beschreiben Vorgaben zum Umgang mit Forschungsdaten. Hierzu zählen

- fachspezifische Empfehlungen (siehe Psychologie-spezifische Empfehlung)
- Leitlinien von Forschungsförderern
- institutionelle Policies und Leitlinien
- projektspezifische Policies
- Journal- und Verlags-Policies und Leitlinien

In den nachfolgenden Kapiteln werden insbesondere die ersten vier Formen näher erläutert und mit Beispielen ergänzt.

b) Psychologie-spezifische Empfehlungen

Für einige Fachgebiete gibt es bereits spezifische Empfehlungen für den Umgang mit Forschungsdaten (u. a. Psychologie, Genetik, Linguistik, Bildungsforschung, Sozialwissenschaften²¹). Solche fachspezifischen Anforderungen adressieren die jeweils unterschiedlichen Datentypen (bspw. bildgebende Formate, Interviews, Umfragen, Farbspektren usw.), Datenquellen (z. B. Personen, chemische/physikalische Experimente usw.) und Praktiken im Umgang mit Daten. Gleichzeitig dient die Formulierung dieser Standards innerhalb eines Fachbereichs dazu, Vergleichbarkeit und Interoperabilität der Daten zu spezifischen Fragestellungen herzustellen (z. B. Erhebungen zum Thema Stress).

So gab die DGPs 2016²² eine in der Folge mehrfach überarbeitete Empfehlung zum allgemeinen Umgang mit Forschungsdaten²³ für die Psychologie in Deutschland heraus. In der Handreichung wird zwischen **Rohdaten**, **Primärdaten** und **Sekundärdaten** unterschieden (siehe **Einheit 3 "Forschungsdaten"**, Punkt 1, S. 9).

Hinsichtlich **personenbezogener Daten** adressiert die Empfehlung die informierte Einwilligung sowie die Veröffentlichung der Daten zur Nachvollziehbarkeit der Ergebnisse und Nachnutzung durch andere Forscher*innen. In diesem Zusammenhang wird zwischen sogenannten **Typ 1-Daten** (Veröffentlichung als Bestandteil einer Publikation) und **Typ 2-Daten** (Veröffentlichung nach Projektabschluss) unterschieden, sowie Hinweise zur Wahl eines geeigneten Repositoriums für die Publikation und/oder die Archivierung dieser Daten gegeben.

Die Revision von 2020 nimmt zudem Aspekte der Zugriffs- und Nutzungsrechte Dritter (Stichwort **Zugriffsklassen**), strukturelle Anreize der Bereitstellung entsprechender Dienste an den Instituten zur Umsetzung der Leitlinie sowie den Umgang mit Konfliktfällen (z. B. bei Verstößen gegen gesetzliche Vorgaben) in einem Psychologie-spezifischen Kontext in den Blick.

²¹ Übersicht über fachspezifische Leitlinien zum Umgang mit Forschungsdaten, Zugriff am 23.11.2023, https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/forschungsdaten/empfehlungen/index.html.

²² DGPs „Empfehlungen zu einem offenen und nachhaltigem Datenmanagement“. Zugriff am 23.11.2023, <https://www.dgps.de/schwerpunkte/transparente-wissenschaft/umgang-mit-forschungsdaten/>.

²³ Management und Bereitstellung von Forschungsdaten in der Psychologie: Überarbeitung der DGPs-Empfehlungen. Zugriff am 23.11.2023, <https://psyarxiv.com/hcxtm/>.

Für die Lebenswissenschaften liefern zudem die „Gute Klinische Praxis (GCP)“ und die „Grundsätze der Guten Laborpraxis (GLP)“ Regularien für das Forschungsdatenmanagement. Diese beiden Grundsätze sind jedoch in Deutschland gesetzlich bindend²⁴.

c) Leitlinien von Forschungsförderern

Für die Psychologie können Projektmittel von verschiedenen Forschungsförderern eingeworben werden, welche ebenfalls Leitlinien für das Forschungsdatenmanagement erstellt haben: die Deutsche Forschungsgesellschaft (DFG²⁵) und das vom European Research Council (ERC) geführte Programm Horizon 2020²⁶.

So sind i.d.R. ausführliche Angaben zum Umgang mit Forschungsdaten im Rahmen der Drittmittelbeantragung verpflichtend. Dies beinhaltet u. a.:

- die Methodik der Datenerhebung bzw. eine Begründung, warum existierende Daten nicht genutzt werden
- die Ansätze zur Beschreibung der Daten (z. B. Metadaten, Dokumentationsstandards)
- die Art der Methoden und Tools zur Verarbeitung der Daten
- Ort der Langzeitarchivierung der Daten sowie geplante Nutzungs- und Zugriffsrechte
- die für das Projekt relevanten rechtlichen Verpflichtungen und Bedingungen:
 - Auswirkungen oder Einschränkungen in Bezug auf eine spätere Veröffentlichung oder Zugänglichkeit
 - Aspekte der Nutzung und des Urheberrechts
 - Eigentumsfragen
- die Verantwortlichkeiten und Ressourcen
 - Beschreibung der Rollen und Verantwortlichkeiten innerhalb des Projekts
 - benötigte Ressourcen (Kosten, Zeit oder andere), um angemessenen Umgang mit Forschungsdaten zu realisieren
 - Verantwortlichkeiten nach Beendigung des Projektes

Einreichungen beim ERC sehen die Vorlage eines Datenmanagementplans innerhalb von 6 Monaten nach Projektstart vor. Dabei kommen ähnliche Anforderungen wie seitens der DFG²⁷ zum Tragen, wie etwa die Beschreibung der zu erhebenden Daten, Protokolle zur Strukturierung der Daten, die genutzten Metadatenstandards, die Benennung der persistenten Identifier, die Datenpflege (z. B. Backups, Bereinigung), Aspekte der Langzeitarchivierung sowie geplante Nutzungs- und Zugriffsrechte).

Für Anträge, die beim BMBF im Rahmen spezifischer Richtlinien bzw. Förderinitiativen eingereicht werden, können Informationen in Bezug auf den Umgang mit Forschungsdaten erforderlich sein, abhängig vom jeweiligen Förderprogramm. So wird beispielhaft im Programm zur Förderung von

²⁴ Bundesministerium für Justiz und Verbraucherschutz „Verordnung über die Anwendung der Guten Klinischen Praxis bei der Durchführung von klinischen Prüfungen mit Arzneimitteln zur Anwendung an Menschen.“ Zugriff am 23.11.2023, https://www.bfarm.de/SharedDocs/Downloads/DE/Arzneimittel/KlinischePruefung/GCP-Verordnung.pdf?__blob=publicationFile,

²⁵ Deutsche Forschungsgemeinschaft. „Leitlinien zum Umgang mit Forschungsdaten.“ Zugriff am 23.11.2023, <https://zenodo.org/record/6472827#.Yt40rcHP2Zw>

²⁶ Europäische Kommission. „Guidelines on FAIR Data Management in Horizon 2020. Version 3.0.“ Zugriff am 23.11.2023, https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

²⁷ DFG-Checkliste zur Beschreibung des Umgangs mit Forschungsdaten. Zugriff am 27.03.2023, https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf.

Forschungsprojekten zum Thema „Nicht-staatliche Hochschulen“ (Bundesanzeiger vom 27.04.2022²⁸) eine Verpflichtung verlangt, Daten mit „etwaiger Relevanz zur Nutzung durch Dritte nach Abschluss des Projekts in weitergabefähiger Form auf der Basis gängiger Standards einer geeigneten Einrichtung/einem Forschungsdatenzentrum zur Verfügung zu stellen mit dem Ziel, langfristige Datensicherung, Sekundärauswertungen oder eine Nachnutzung zu ermöglichen“. Diese Vorgabe soll mittels eines eigenen Forschungsdatenmanagementplans unterstützt werden, der bei Einreichung vorgelegt werden muss. Da Förderprogramm-unabhängige Richtlinien seitens des BMBF fehlen, sollte sich im Falle des Fehlens konkreter Vorgaben bei Förderrichtlinien an den gängigen Empfehlungen der DGPs orientiert werden.

d) Institutionelle Policies und Leitlinien in der Psychologie

Zahlreiche der Hochschulen und außeruniversitären Forschungseinrichtungen haben bereits eine institutionelle Forschungsdaten-Policy eingeführt, um den Umgang mit Forschungsdaten zu regeln²⁹.

Fachspezifische Forschungsinstitute im Bereich der Psychologie, wie bspw. das Leibniz-Institut für Psychologie, das ZI Mannheim³⁰ oder das IfaDo³¹, veröffentlichen entsprechend eigene Policies. Die Humboldt Universität Berlin verweist zudem in Bezug auf die Verhaltensforschung neben den DGPs-Richtlinien auch auf die FDM-Empfehlung des Rates für Sozial- und Wirtschaftsdaten³².

In institutionellen Policies kann neben der Empfehlung, Forschungsdaten und Forschungsergebnisse im Open Access-Format zu veröffentlichen, auch die Kosten- und Ressourcensteuerung (Hiemenz & Kuberek, 2018a), d.h. die personellen, organisatorischen und technischen Ressourcen und Services der Forschungseinrichtung für das FDM angegeben werden. Abbildung 4 zeigt die Elemente einer institutionellen Forschungsdaten-Policy (Hiemenz & Kuberek, 2018b) und listet inhaltliche Punkte, die in diesen Kategorien geregelt werden.

²⁸ Richtlinie zur Förderung von Forschungsprojekten zum Thema „Nicht-staatliche Hochschulen“, Bundesanzeiger vom 27.04.2022. Zugriff am 23.11.2023, <https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2022/04/2022-04-27-Bekanntmachung-Nicht-staatl-Hochschulen.html>.

²⁹ Liste der Forschungsdaten-Policies an deutschen Hochschulen: forschungsdaten.org: „Data Policies“. Zugriff am 23.11.2023, https://www.forschungsdaten.org/index.php/Data_Policies#Institutionelle_Policies.

³⁰ ZI Mannheim, „Handling your data“. Zugriff am 23.11.2023, <https://www.zi-mannheim.de/en/service/rechtsverstoesse-melden/handling-your-data.html>, „Privacy policy“, Zugriff am 23.11.2023, <https://www.zi-mannheim.de/en/service/privacy-policy.html>.

³¹ Forschungsdatenpolicy des IfaDo. Zugriff am 23.11.2023, <https://www.ifado.de/datenmanagement/>.

³² Humboldt Universität Berlin. Zugriff am 23.11.2023, https://www.konsortswd.de/wp-content/uploads/RatSWD_Output3_Forschungsdatenmanagement.pdf.

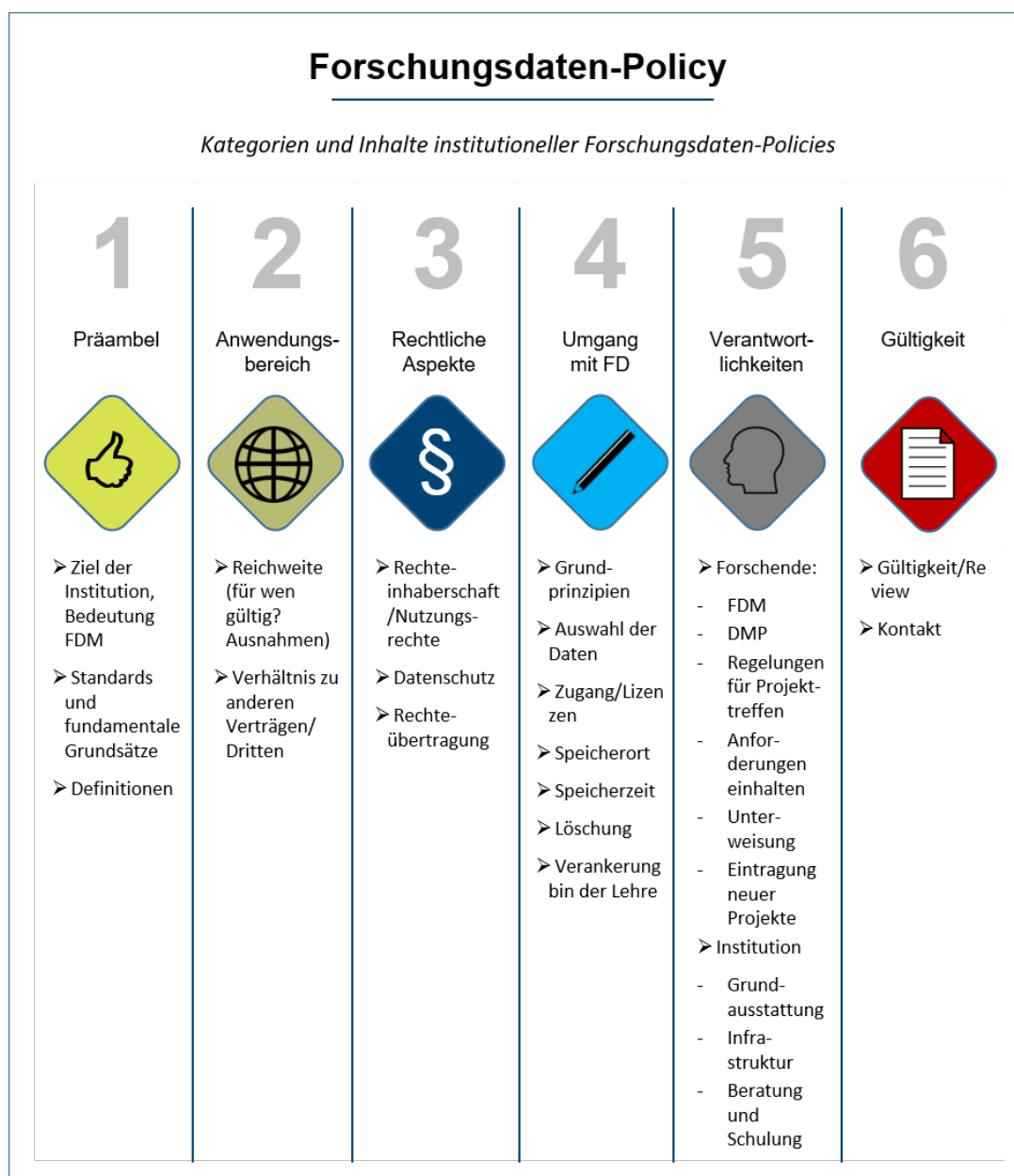


Abbildung 4: Visuelle Aufbereitung einer Forschungsdaten-Policy. Entnommen aus dem strategischen Leitfaden zur Erstellung einer institutionellen Forschungsdatenpolicy (Hiemenz & Kuberek, 2018b).

e) Projektspezifische Policies

Die von der DFG finanzierten Förderprogramme „Sonderforschungsbereich“, „Transregio“ und „Forschungsprojekt“, sowie einige Ausschreibungen des BMBF adressieren dezentral verteilte Forschungsgruppen, die sich einer gemeinsamen Forschungsfrage widmen. Da sich die Leitlinien und Policies sowohl auf institutioneller als auch internationaler Ebene unterscheiden können, ist die Erstellung einer projektspezifischen Policy für die jeweilige Forschungsgruppe zum Umgang mit den hier erhobenen Daten sinnvoll. Ebenso wie bei einer institutionellen Policy wird dabei die Art und Weise der Erhebung, Speicherung, Analyse und Weiterverwertung der Daten für alle Standorte verbindlich festgelegt und damit synchronisiert.

Im Rahmen des Projektes FDNext wurde zu diesem Zweck ein strukturierter Leitfaden zur Erstellung einer projektspezifischen Policy (Schmiederer & Kuberek, 2022) konzipiert und veröffentlicht. Dieser

informiert über Inhalt und Struktur einer Policy inkl. der dafür notwendigen Vorüberlegungen und stellt Checklisten zum systematischen Vorgehen bei der Erstellung zur Verfügung. Ein Online-Generator zur Unterstützung der Erstellung einer Forschungsdatenpolicy ist ebenfalls in Planung.

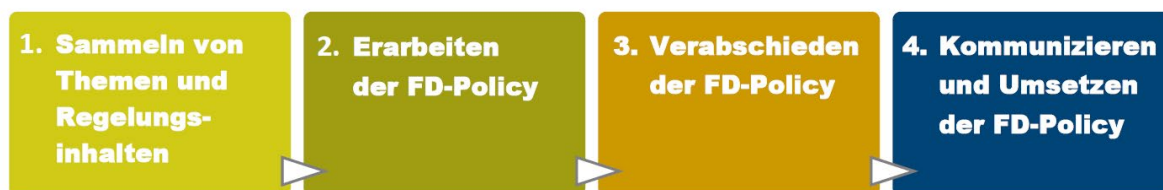


Abbildung 5: Die vier Phasen zur Erstellung einer Forschungsdatenpolicy für Forschungsprojekte.

Rechtliche Aspekte (Einheit 5)

1. Ebenen der Regelung

Je nach der Beschaffenheit der Daten müssen rechtliche und ethische Rahmenbedingungen, die auf sehr verschiedenen Ebenen verortet sind, beim Forschungsdatenmanagement berücksichtigt werden (siehe Tabelle 2). So gelten, wenn es um personenbezogene Daten geht (siehe **Einheit 5 "Rechtliche Aspekte", Punkt 2**, S. 22), die EU-Datenschutz-Grundverordnung (DSGVO), das Bundesdatenschutzgesetz, die Landesdatenschutzgesetze und, falls vorhanden, die Landeshochschulgesetze sowie je nach Fall weitere datenschutzrechtliche Bestimmungen, z. B. im Sozialgesetzbuch, dem Genetik-, Arzneimittel- oder Schulgesetz. Eine im Rahmen des Projektes FDNext veröffentlichte Handreichung³³ bietet für Wissenschaftler*innen der Psychologie, der Bildungs- und Erziehungswissenschaften hilfreiche Informationen im Umgang mit personenbezogenen Daten.

Für Aspekte jenseits der personenbezogenen Daten kann es auch auf institutioneller Ebene Richtlinien, Policies oder auch eigene Ethikrichtlinien geben. Grundsätzlich gelten für alle Forschende in Deutschland die „Leitlinien zur Sicherung guter wissenschaftlicher Praxis“³⁴ der Deutschen Forschungsgemeinschaft. Eine Übersicht über die in Deutschland gültigen Leitlinien bietet das Informationsportal des KonsortSWD³⁵.

Tabelle 2: Beispiele rechtlicher Anwendungsfelder im Kontext der Datenpublikation³⁶.

Patentrecht	Urheberrecht	Wettbewerbsrecht	Datenschutz
Was ist zu beachten, wenn FD Patentreife erlangen (können)?	Unterliegen FD überhaupt dem Urheberrechtsgesetz?	Werden Daten im unternehmerischen Geschäftsverkehr (nicht) einvernehmlich genutzt?	Welche FD sind "schützenswert"?
Wissenschaftsrecht	Grundrechte	Internationales Recht	EU-Recht
Können Lizenz- und Veröffentlichungsvorgaben für FD per Mandatierung erfolgen?	Welche verfassungsrechtlichen Grenzen sind zu beachten?	Welche Rechtsbestimmungen bestehen außerhalb Deutschlands?	Was bringt bspw. die "European Data Economy" für FD?
Verträge	Arbeits-/Dienstrecht	Förderbedingungen	Policies
Bestehen Absprachen zum "geistigen Eigentum" an FD?	Wem "gehören" die an Hochschulen erhobenen FD?	Welche Bedingungen geben Förderer (DFG; Industrie) vor?	Welche rechtliche Verbindlichkeit können Policies entfalten?

³³ Kienbaum, J., Fischer, P., & Paßmann, S. (2023). Forschungsdatenmanagement bei personenbezogenen Daten - eine Handreichung. Zenodo. <https://doi.org/10.5281/zenodo.7428524>.

³⁴ Deutsche Forschungsgemeinschaft. *Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex*. Zenodo, 2022. Zugriff am 23.11.2023, https://zenodo.org/record/6472827#.Yt_WXdrwBaQ

³⁵ Übersicht der in Deutschland gültigen Leitlinien auf KonsortSWD. Zugriff am 23.11.2023, <https://www.konsortswd.de/themen/best-practice-forschungsethik/ethikleitlinien-kodices/>.

³⁶ nach Hartmann, Thomas: „terra incognita – digitale Forschungsdaten auf der Suche nach einer rechtlichen Heimat“. Beitrag bei der Workshoptagung „Rechtliche Aspekte bei digitalen Forschungsdaten“ am 31.01.2018, Europa-Universität Viadrina, Frankfurt (Oder), Zugriff am 23.11.2023, https://www.forschungsdaten.org/images/0/08/Hartmann_TerraIncognita-Forschungsdaten-RechtlicheHeimat.pdf.

In manchen Fällen ist es notwendig, den Zugang zu den Daten einzuschränken, vor allem wenn ethische oder rechtliche Aspekte eine Rolle spielen, oder die Forschung Vertraulichkeitsabreden unterliegt, bspw. bei Auftragsforschung (siehe **Einheit 15 „Publikation von Forschungsdaten“**, **Punkt 6**, S. 59 und **Punkt 8**, S. 60).

2. Datenschutz und Persönlichkeitsrecht

In der europäischen Union regelt die **DSGVO** die Belange des Datenschutzes und des Persönlichkeitsrechts ihrer Bürger*innen, welches grundsätzlich den Umgang mit personenbezogenen Daten unter Erlaubnisvorbehalt untersagt. Sie regelt auf europäischer Ebene das Recht auf den Schutz personenbezogener Daten und stellt den freien Verkehr personenbezogener Daten sowohl innerhalb der EU als auch gegenüber Drittstaaten sicher.

Durch eine sogenannte „**informierte Einwilligung**“ (informed consent) des Teilnehmenden ist es jedoch Forschenden gestattet, personenbezogene Daten zu erheben und zu verarbeiten. Hierbei handelt es sich um eine unmissverständlich abgegebene Einwilligung seitens des Teilnehmenden in Form einer informierten Erklärung oder einer sonstigen eindeutigen zustimmenden Handlung, die erkennen lässt, dass der*diejenige über den genauen Zweck der Datenerhebung, -verarbeitung und -nutzung aufgeklärt ist. Diese Zustimmung kann auch mündlich erfolgen (z. B. bei Analphabeten). Bei Studien, die eine **Täuschung** beinhalten (zur Vermeidung des Bestätigungsfehlers, engl. *confirmation bias* (Mercier, 2016)), ist nach der Aufklärung über das wahre Ziel³⁷ die Einholung einer zweiten Einwilligung notwendig, welche sich nur noch auf die Verwendung der Daten selbst beziehen muss³⁸.

Die Artikeln 4 und 9 der DSGVO gehen ausführlicher auf die Definition von **personenbezogenen Daten** ein. Im ersten Artikel werden personenbezogene Daten als „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person [...] beziehen“ definiert. Als identifizierbar werden die Betroffenen angesehen, wenn sie direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen identifiziert werden können, die Ausdruck der physischen, psychologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Personen sind.

Als eine besondere Form der personenbezogenen Daten gelten diejenigen, die eines erhöhten Schutzes bedürfen (Art. 9 DSGVO). Dazu gehören Daten, „aus denen die rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen oder Gewerkschaftszugehörigkeit hervorgehen“, sowie genetische und biometrische Daten zur eindeutigen Identifizierung einer natürlichen Person, Gesundheitsdaten oder Daten zum Sexualleben oder der sexuellen Orientierung einer natürlichen Person.

Ist die Publikation von personenbezogenen Forschungsdaten geplant, muss diese Information in die Aufklärung aufgenommen werden. Wichtig ist dabei vor allem, alle geplanten Nutzungszwecke in der Einwilligung zu erfassen (so u. a. die Langzeitarchivierung, Datenpublikation, weitere Auswertungen zu anderen Fragestellungen usw.), da es im Nachhinein häufig nicht mehr gelingt, die Erweiterung der Einwilligung einzuholen.

³⁷ Deutsche Gesellschaft für Psychologie (DGPs). (2018). Ethisches Handeln in der psychologischen Forschung. S. 41.

³⁸ Deutsche Gesellschaft für Psychologie (DGPs). (2018). Ethisches Handeln in der psychologischen Forschung. S. 45; Rat für Sozial- und Wirtschaftsdaten (RatSWD) (Ed.) (2017): Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften, RatSWD Output, No. 9 (5), Rat für Sozial- und Wirtschaftsdaten (RatSWD), Berlin, <https://doi.org/10.17620/02671.1>.

Die folgenden Seiten bieten Informationen, Hilfestellungen und Templates zur Erstellung von entsprechenden Einwilligungserklärungen:

- Hilfestellungen und eine Sammlung von Templates auf TransMIT-Zentrum für wissenschaftlich-psychologische Dienstleistungen (DGPs)³⁹
- Sammlung von Hilfestellungen für Einwilligungserklärungen auf den Seiten von KonsortSWD⁴⁰
- Sammlung auf Forschungsdaten-Bildung⁴¹

3. Anonymisierung der in der Psychologie relevanten Daten

Insbesondere in der Psychologie ist die Nutzung personenbezogener Daten weit verbreitet. Hier hat die DGPs für die Psychologie eine datenschutzrechtliche Empfehlung zur Erstellung einer Einwilligungserklärung im Rahmen von Forschungsvorhaben erstellt⁴². Neben den bereits genannten Aspekten adressiert diese Empfehlung u. a. auch die Studienbeschreibung, Datenauswertung und -nachnutzung. Enthalten sind zudem ein Template für eine Einwilligungserklärung, dass die Vorgaben der DSGVO umsetzt, sowie Kontaktangaben zu den Datenschutzbeauftragten des Bundes und der Länder.

a) Anonymisierung qualitativer und quantitativer Daten

Um Daten einfacher und rechtssicher verarbeiten zu können, existieren verschiedene Möglichkeiten, den Personenbezug zu entfernen bzw. eine Anonymisierung der Daten zu erreichen (Meyermann & Porzelt, 2014). So können **qualitative Daten** beispielsweise durch die Aggregation (Mittelung der Datenwerte) mittels herkömmlicher Analysesoftware wie R oder SPSS anonymisiert werden. Eine weitere Option der Anonymisierung ist das komplette Entfernen personenbezogener Daten oder das Ersetzen dieser durch Informationen, die nicht auf einzelne Personen rückführbar sind (bspw. PLZ statt Straße). Letzteres kann für **qualitative Daten** zum Beispiel mittels der Software QualiAnon⁴³ bewerkstelligt werden, während für quantitative Daten mit der Unterstützung des EC Horizon Programms das Tool Amnesia⁴⁴ entwickelt wurde, welches eine Anonymisierung der Daten auf einem sehr hohen k-Anonymitätslevel⁴⁵ ermöglicht. Ein weiteres Tool für quantitative Daten ist ARX⁴⁶.

³⁹ TransMIT-Zentrum, Zugriff am 23.11.2023, <https://zwpd.transmit.de/zwpd-dienstleistungen/zwpd-ethikkommission/vorlagen-antragstellung>.

⁴⁰ Hilfestellungen für Einwilligungserklärungen liefert KonsortSWD, Zugriff am 23.11.2023, <https://www.konsortswd.de/ratswd/best-practice-forschungsethik/einwilligungsunterlagen/>.

⁴¹ Hilfestellungen und Templates in Bezug auf Einwilligungserklärungen auf Forschungsdaten Bildung, Zugriff am 23.11.2023, <https://www.forschungsdaten-bildung.de/einwilligung>.

⁴² Datenschutzrechtliche Empfehlung der DGPs. Zugriff am 23.11.2023, https://zwpd.transmit.de/images/zwpd/dienstleistungen/ethikkommission/0.1a_datenschutzrechtliche_empfehlung_en_einwilligungforschungsvorhaben.pdf.

⁴³ Anonymisierungs-Applikation QualiAnon. Zugriff am 23.11.2023, <https://www.qualiservice.org/de/helpdesk/webinar/tools.html>.

⁴⁴ Amnesia, Zugriff am 23.11.2023, <https://amnesia.openaire.eu/index.html>.

⁴⁵ Sweeney, L. (2002): K-Anonymität beschreibt den Grad der Nicht-Unterscheidbarkeit der Informationen einer spezifischen Person von k-1 anderen Individuen. Je höher der Wert k, desto höher die Nicht-Unterscheidbarkeit, desto anonymier die Daten:

Beispiel: Wenn durch eine erfolgte Anonymisierung ein Datensatz in Gruppen von Individuen unterteilt werden kann (Anonymisierung erfolgte durch die Bildung von Clustern (z. B., Altersgruppe 20-30 Jahre)) – dann bestimmt die Mindestanzahl der zugeordneten Individuen zu einer dieser Gruppen den k-Wert.

Bitte beachten Sie: Da diese Methode bekannte Mängel aufweist, sollte der k-Wert immer im Kontext des Datensatz (Einzigartigkeit der Daten, Anzahl der individuellen Daten mit identischen sensiblen Attributen, Zusatzwissen, welches nicht Bestandteil des Datensatzes ist, usw.) bewertet werden.

⁴⁶ ARX, Zugriff am 23.11.2023, <https://arx.deidentifier.org/>.

Für Daten im klinischen Kontext ist die Safe Harbour Methode⁴⁷ hilfreich (in den USA gesetzlich verankert), in der 18 Identifier aufgeführt sind, welche anonymisiert werden müssen (z. B. Namen, geographische Kategorien kleiner als Bundesstaat/Bundesland, Geburtsdatum, Telefonnummern usw.).

b) Anonymisierung von Daten aus bildgebenden Verfahren

In Bezug auf MRT-Daten wird empfohlen, neben der Vergabe von Dateinamen ohne Personenbezug auch die Informationen in den Aufnahmen zu entfernen, die für die Forschungsfrage u. U. keine Relevanz haben (Knochenstruktur, extra-kraniales Gewebe)⁴⁸. Dafür gibt es zwei weit verbreitete Ansätze, das sogenannte „skull stripping“ (Abisolieren des Schädels) und das „defacing“ (Entfernen von Gesichtsstrukturen). Zur Umsetzung sind verschiedene Möglichkeiten entwickelt worden, zum einen spezifische Programmroutinen (z. B. ft_defacevolumes für Fieldtrip⁴⁹) oder speziell dafür entwickelte, open-source Software (z. B. MRI_deface⁵⁰, Pydeface⁵¹). Um die (Meta-)Daten in diesen Aufnahmen zu entfernen, die zur Identifizierung spezifischer Probanden (Geschlecht, Alter, Name usw.) benutzt werden können, gibt es ebenfalls frei verfügbare Software (u. a. die De-Identification Toolbox⁵² oder GDCM⁵³).

Da für M/EEG Daten gerade die individuellen Muster der Gehirnaktivität Grundlage der Forschung sind, sind Methoden zur De-Identifizierung wenig sinnvoll. Nichtsdestotrotz sollte auf größtmögliche Anonymität bei den Dateinamen und Metadaten geachtet werden.

c) Anonymisierung anderer Forschungsdaten und weitere Schutzmaßnahmen

Die rasante Entwicklung von digitalen Medien führt heutzutage mehr als zuvor zur unbedachten Veröffentlichung von personenbezogenen Daten – zum Beispiel in Form von Bildern oder Videos. Das Recht am eigenen Bild umfasst jedoch das Recht eines jeden Einzelnen zu entscheiden, für welchen Zweck die Bilder oder andere Formen von Abbildungen der eigenen Person zur Verfügung stehen. Nur mit Einwilligung können diese verbreitet oder veröffentlicht werden.

Als eine Einwilligung gilt sowohl eine Entlohnung als auch die Beantwortung von Fragen vor laufender Kamera. Ausnahmen hiervon bilden:

- zeitgeschichtliche Abbildungen
- Bilder, auf denen die Person nicht als das Hauptmotiv dargestellt wird (z. B. bei einer Aufnahme einer Demonstration oder einer belebten Einkaufsstraße)
- Bilder von Versammlungen o. ä.

⁴⁷ Safe Harbour Methode zur Anonymisierung von (Gesundheits-)Daten Zugriff am 23.11.2023, <https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#safeharborguidance>

⁴⁸ De-Identifizierungsmethoden: Schimke, N., Kuehler, M., & Hale, J. (2011). Preserving privacy in structural neuroimages. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6818 LNCS, 301–308. https://doi.org/10.1007/978-3-642-22348-8_26.

⁴⁹ De-Identifizierung in Fieldtrip. Zugriff am 23.11.2023, https://www.fieldtriptoolbox.org/faq/how_can_i_anonymize_an_anatomical_mri/.

⁵⁰ MRI_deface. Zugriff am 23.11.2023, <https://surfer.nmr.mgh.harvard.edu/fswiki/AutomatedDefacingTools>.

⁵¹ Pydeface. Zugriff am 23.11.2023, <https://github.com/poldracklab/pydeface>.

⁵² De-Identification Toolbox. Zugriff am 23.11.2023, <https://www.nitrc.org/projects/de-identification/>.

⁵³ GDCM zur Entfernung von Metadaten. Zugriff am 23.11.2023, <https://sourceforge.net/projects/gdcm/>.

Video- und auch Tonaufnahmen können jedoch editiert und damit anonymisiert bzw. de-personalisiert werden. Beispiele hierfür sind das Verfremden (Verpixeln, schwarzer Balken) der Personen oder Informationen sowie die Überarbeitung von Tonaufnahmen, um Stimmen zu verfremden mittels verschiedener Applikationen (bspw. Freeware wie VideoPad Video Editor⁵⁴ oder Avidemux⁵⁵ bzw. kostenpflichtige Editierprogramme wie z. B. Vegas⁵⁶).

Pseudonyme, d. h. die Ersetzung einer Information durch eine ähnliche, aber nicht mehr personenbeziehbare Information, sind ebenfalls eine Möglichkeit, um qualitative Daten ohne informierte Einwilligung weiterverarbeiten zu können. Eine Veröffentlichung ist aber ohne entsprechende Maßnahmen zur Sicherung gesetzlicher Vorgaben nicht ohne weiteres möglich (z. B. Einschränkung durch Zugriffsklassen oder Lizenzen). Weitere Schutzmaßnahmen bieten Zugriffsbeschränkungen mittels Passwortschutz und die Vergabe von Zugriffsrechten. So kann Zugang zu den Daten z. B. nur an bestimmte IPs oder nur für definierte Personenkreise gewährt werden (siehe **Einheit 9 "Zugriffssicherheit"**, S. 43).

4. Urheberrecht und verwandte Schutzrechte

Das **Urheberrecht** dient dem Schutz der schöpferischen Leistung von Kreativschaffenden und der Kreativwirtschaft. Es erteilt den Urheber*innen das alleinige Recht an der Veröffentlichung, Verarbeitung, Vervielfältigung, Aufführung und dem Verleih ihres Werkes. Damit ein Werk als "urheberrechtlich geschützt" gilt, muss es im Zuge einer persönlichen, geistig-schöpferischen Leistung entstanden sein. Nach §2 Abs. 1 UrhG gehören dazu "Werke der Literatur, Wissenschaft und Kunst". Ideen, Fakten oder Informationen werden demnach nicht vom Urheberrecht geschützt, da sie lediglich die Basis eines Werkes bilden, jedoch nicht das Werk selbst sind (Kreutzer & Lahmann, 2021).

Somit sind maschinell erzeugte und unbearbeitete Rohdaten urheberrechtlich nicht schutzfähig. Dahingehend können jedoch die für eine Veröffentlichung angefertigten Texte, Zeichnungen, individuell gestalteten Grafiken, Fotografien, Videos etc. urheberrechtlich geschützt sein – jedoch nur, wenn sie den obengenannten Kriterien entsprechen. Dies muss im Einzelfall geprüft werden (Leibniz Universität Hannover & Technische Informationsbibliothek, 2018).

Die sogenannten **Leistungsschutzrechte** - auch "verwandte Schutzrechte" genannt - sind Schutzrechte für z. B. Fotograf*innen, Sänger*innen, Interpret*innen oder Vermittler*innen von Inhalten. Der Leistungsschutz beinhaltet unter anderem aber auch den Schutz des Datenbankstellers, Schutz der Lichtbilder oder den Schutz des Herstellers von Tonträgern (weitere Schutzrechte können im UrhG Teil 2 "Verwandte Schutzrechte" nachgelesen werden). Die Leistungsschutzrechte stehen demjenigen zu, der die jeweils geschützte Leistung erbracht hat.

Daneben können aber auch **gewerbliche Schutzrechte** zu berücksichtigen sein, bspw. eine vertragliche Vereinbarung (durch z. B. Arbeitsverträge), die eine Übertragung des Nutzungsrechts an den Daten auf die Institution vorsieht. Des Weiteren können vertragliche Regelungen zu Nutzungsrechten auch in Förderrichtlinien von Geldgebern oder Richtlinien der beteiligten Gruppen, Abteilungen oder Institutionen zu finden sein. Auch Open-Content-Lizenzen (siehe **Einheit 15 "Publikation"**, **Punkt 6**, S. 57) sind Verträge über die Einräumung von Nutzungsrechten. Die Verträge müssen dabei nicht schriftlich, sondern können auch implizit oder mündlich geschlossen werden (Leibniz Universität Hannover & Technische Informationsbibliothek, 2018). Obwohl die Nutzungsrechte übertragen werden, bleiben jedoch die Urheberschaft und damit die Verwertungsrechte bei den Ersteller*innen der Daten.

⁵⁴ VideoPad Video Editor. Zugriff am 23.11.2023, https://www.nchsoftware.com/videopad/de/index.html?ns=true&kw=videopad%20video%20editor&gclid=CjwKCAjw6raYBhB7EiwABge5KI_FtxfW39dUG-AzhdxtysgkPTTgjIsveEHIGNvqRADqtX5PI2k7TxcC7h4QAvD_BwE.

⁵⁵ Avidemux. Zugriff am 23.11.2023, <http://avidemux.sourceforge.net/>.

⁵⁶ Vegas Software. Zugriff am 23.11.2023, <https://www.vegascreativesoftware.com/de/>.

Wenn anzunehmen ist, dass die Forschung oder ihre Ergebnisse wirtschaftlich nutzbar sind, ist es empfehlenswert, sich an die Wissenstransferstellen der eigenen Einrichtung zu wenden. Hier kann geklärt werden, wie die Lizenzierung und Nachnutzung der Daten zu regeln ist. Auch über das Patentrecht kann man sich dort ausführlich informieren.

5. Vertragsrecht

Weitere Rechte und Pflichten für den Umgang mit Forschungsdaten können sich auch aus sonstigen vertraglichen Vereinbarungen ergeben. Dies können sowohl Verträge mit Fördermittelgebern als auch mit wissenschaftlichen oder wirtschaftlichen Kooperationspartner sein.

Sollen Daten außerhalb von Kooperationsverträgen zur gemeinsamen Bearbeitung an Kolleg*innen weitergegeben werden, bevor sie publiziert wurden, wird dies durch Datenbereitstellungsverträge (data sharing agreements) geregelt. Hier wird festgelegt, unter welchen Bedingungen Daten durch Dritte verarbeitet werden dürfen.

Dabei verfolgt ein solches Agreement zwei Ziele:

- den Schutz der Daten und der*des Datengebenden vor Datenmissbrauch durch Dritte
- die Eindeutigkeit und Dokumentation aller Vereinbarungen zur Datennutzung auf beiden Vertragsseiten

Datenmanagementplan (Einheit 6)

1. Motivation

Ein Datenmanagementplan (DMP) bildet ein strukturiertes Instrument zur Dokumentation des Umgangs mit Forschungsdaten im Forschungsprozess bzw. Projektverlauf. Nicht nur im Kontext der Drittmittelbeantragung, wo er heute im Rahmen verschiedener Förderlinien verpflichtend ist, unterstützt er Forschende bei der Umsetzung nachhaltigen FDMs. Er

- schafft eine verbindliche Grundlage für einen einheitlichen Umgang mit Daten im Forschungsprozess
- erleichtert das Verständnis der eigenen Daten
- erleichtert die Abstimmung zwischen Projektpartnern hinsichtlich der Umsetzung verschiedener Aspekte des FDM (u. a. Datenerhebung und -analyse, Speicherung, Langzeitarchivierung, Publikation und Nachnutzung, usw.)
- hilft mögliche Probleme frühzeitig zu erkennen und dafür Lösungen zu skizzieren
- legt Verantwortlichkeiten fest
- regelt Zugriffsrechte
- hilft Datenduplikate, Datenverlust und Sicherheitslücken zu vermeiden.

Ein Datenmanagementplan kann, als Instrument der Qualitätssicherung, des Monitorings und der Dokumentation, den gesamten Forschungsprozess bzw. Projektverlauf begleiten (living document). Als Active Data Management Plan⁵⁷ trägt er zur Nachnutzbarkeit der Daten bei (Reusability).

Eine hilfreiche Vorüberlegung für einen Datenmanagementplan ist es, den Prozess ausgehend von der Datenarchivierung her zu beschreiben. Aus diesen Überlegungen ergibt sich die Notwendigkeit, bereits sehr früh die bspw. zu verwendenden Formate, Standards, Metadaten, Lizenzen etc. festzulegen.

2. Was ist ein Datenmanagementplan?

Ein Datenmanagementplan (DMP) ist ein Dokument, in dem der beabsichtigte Umgang mit Forschungsdaten beschrieben wird; er schließt Maßnahmen sowohl während des Forschungsprozesses als auch nach dessen Abschluss mit ein. Das Dokument enthält dabei alle Informationen, die die Sammlung, Aufbereitung, Speicherung, Archivierung und Veröffentlichung von Forschungsdaten beschreiben und dokumentieren, und kann – je nach Komplexität des Projektes – zwischen wenigen Absätzen und mehreren Seiten variieren. Die Diversität der Forschungsdaten sowie des Umgangs mit ihnen bestimmt die Länge eines Datenmanagementplans. Er sollte – wenn möglich – auch mit allen Projektbeteiligten abgestimmt sein.

⁵⁷ RDA Interest Group Active Data Management Plans. Zugriff am 23.11.2023, <https://www.rd-alliance.org/node/9136/outputs>.

3. Bestandteile eines Datenmanagementplans

Ein DMP adressiert i.d.R. folgende Punkte:

- Projekttitle, Laufzeit und Forschungsfrage(n)
- Verantwortliche*r für das Datenmanagement
- Nachnutzung existierender Daten
- zu erhebende Daten:
 - Beschreibung der zu erfassenden Daten, Datentypen und -formate
 - erwarteter Speicherbedarf
 - Methoden der Datenerhebung, verwendete Hard- und Software
- Datenorganisation:
 - Datenspeicherung
 - Backup
 - Ordnerstruktur
 - Dateinamenskonventionen
 - Dokumentation und Metadaten
- Rechtliche Aspekte, z. B.:
 - Datenschutz
 - Urheberrecht
- Datenaustausch und -zugang:
 - im Projekt
 - mit externen Partnern und Dienstleistern
- Langzeitsicherung und Archivierung
- Datenpublikation
- Kosten des Datenmanagements
- Qualitätssicherung
- Zugriff und Nachnutzung

4. Anforderungen der Forschungsförderer

In Deutschland werden Datenmanagementpläne bereits von einer Reihe von Forschungsförderern bei der Antragstellung gefordert. Forschungsförderer wie die Europäische Kommission (European Commission, EC), die Deutsche Forschungsgemeinschaft (DFG), die Volkswagen Stiftung (VW Stiftung) und das Bildungsministerium für Bildung und Forschung (BMBF) erwarten zunehmend die Bereitstellung eines Datenmanagementplans zu Förderbeginn (EC⁵⁸), Angaben zum Umgang mit zu erhebenden Forschungsdaten (DFG)⁵⁹ sowie – je nach Förderrichtlinie – einen Verwertungsplan für Projektergebnisse oder einen detaillierten Datenmanagementplan bei Antragstellung (BMBF⁶⁰, VW Stiftung⁶¹) (siehe Tabelle 3).

⁵⁸ European Commission, H2020 Programme. „Guidelines on FAIR Data Management in Horizon 2020.“ Version 3.0. Zugriff am 23.11.2023,

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

⁵⁹ Deutsche Forschungsgemeinschaft. „Leitlinien zum Umgang mit Forschungsdaten.“ Zugriff am 23.11.2023, http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf.

⁶⁰ Verbund Forschungsdaten Bildung. „Checkliste zur Erstellung eines Datenmanagementplans in der empirischen Bildungsforschung.“ Version 1.1. *fdbinfo* Nr. 2 (2015). Zugriff am 23.11.2023, https://www.forschungsdaten-bildung.de/files/fdbinfo_2.pdf.

⁶¹ VolkswagenStiftung. „OPEN SCIENCE. OPEN ACCESS – OPEN DATA – OPEN SOURCE“ Zugriff am 23.11.2023,

https://www.volkswagenstiftung.de/sites/default/files/documents/Open_Science_Policy_und_Umsetzung_VolkswagenStiftung.pdf.

Tabelle 3: Vergleich der Anforderungen der Förderer bezüglich des Datenmanagementplans (Stand: 26.09.2022).

Förderer	Plan gefordert?	Abgabe bei Antrag?	Inhalt	Updates
EC Horizon Europe	Ja	Nein, erster Plan innerhalb der ersten 6 Projektmonate	Inhalte des Horizon Europe Template	Updates bei signifikanten Änderungen und zum Projektende
DFG	Angaben zum Umgang mit Forschungsdaten	Ja	Inhalte der Leitlinie zum Umgang mit Forschungsdaten	Nein
BMBF	abhängig vom Programm	Falls notwendig, ja	Inhalt abhängig vom Programm; Bildungsforschung-Checkliste	abhängig vom Programm
Volkswagenstiftung	Ja	Ja	Inhalte des Science Europe Templates	Nein

5. Muster und Werkzeuge für die Psychologie

Zahlreiche Instrumente können Forschende bei der Erstellung eines Datenmanagementplans unterstützen. Eine Auswahl ist in Tabelle 4 aufgelistet.

Bei DataWiz⁶² handelt es sich um ein vom Leibniz-Institut für Psychologie entwickeltes, leistungsstarkes Tool, mit dem ein an die Anforderungen der großen Forschungsförderer (DFG, EU Horizon, BMBF) angepasster Datenmanagementplan als Output generiert werden kann. Der ebenfalls in Deutschland weit verbreitete Research Data Management Organiser (RDMO⁶³) hingegen ist eher generischer Art, bietet aber umfassende Hilfen auch für das projektbegleitende Management und die Administration von Forschungsdaten und kann mithilfe fachspezifischer Templates an die Bedürfnisse der eigenen Disziplin angepasst werden. Weitere, ebenfalls in der Psychologie verbreitete Tools sind das vor allem in Großbritannien übliche DMPonline⁶⁴, und der für die Bildungswissenschaften entwickelte, aber auch in der Psychologie nutzbare STAMP⁶⁵.

⁶² Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID). „DataWiz.“ Zugriff am 23.11.2023, <https://datawiz.leibniz-psychology.org/DataWiz/>.

⁶³ Research Data Management Organiser. Zugriff am 23.11.2023, <https://rdmorganiser.github.io/>.

⁶⁴ DMPonline. „Plan to make data work for you.“ Zugriff zur Zeit der Erstellung der Version 2 des fachspezifischen TtT nicht möglich, <https://dmponline.dcc.ac.uk/>.

⁶⁵ Standardisierter Datenmanagementplan für die Bildungsforschung (STAMP), Zugriff am 23.11.2023, <https://www.forschungsdaten-bildung.de/stamp-nutzen>.

Tabelle 4: Muster und Werkzeuge zur Erstellung eines Datenmanagementplans

Checklisten	Mustervorlagen
Checkliste Verbund Forschungsdaten Bildung ⁶⁶	Musterplan EU Horizon 2020 (Version 3.0) ⁶⁷
Checkliste OSF.io ⁶⁸	Musterplan DFG ⁶⁹
Checkliste Digital Curation Centre ⁷⁰	Musterplan BMBF ⁷¹
Checkliste Guidance University of Bath ⁷²	Musterplan ICPSR ⁷³
Leitfaden, entwickelt im WissGrid-Projekt (Ludwig & Ehnke, 2013)	Musterplan VW-Stiftung ⁷⁴
	Mustersammlung Digital Curation Centre ⁷⁵

DMP-Tools, allgemein		
	Ziel	Zusätzl. Informationen
DMPTool ⁷⁶	Erstellung DMP, insb. für amerikanische Forschungsförderer	
RDMO	Erstellung DMP, insb. für deutsche Forschungsförderer	bietet fachspezif. Templates; auf forschungsdaten.info ohne Installation via Shibboleth nutzbar ⁷⁷
DMPonline ⁷⁸	Erstellung DMP, insb. für britische Forschungsförderer; DMP für EC-Förderung	DMPTool/DMPonline basieren auf dem gleichen Softwarecode (DMPRoadmap)

⁶⁶ Verbund Forschungsdaten Bildung (2015). „Checkliste zur Erstellung eines Datenmanagementplans in der empirischen Bildungsforschung“. Version 1.1. fdbinfo Nr. 2. Zugriff am 23.11.2023, https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdbinfo_2.pdf.

⁶⁷ DMP-Musterplan EU Horizon 2020 (Version 3.0). Zugriff am 23.11.2023, <https://www.cms.hu-berlin.de/de/dl/dataman/muster-dmp-h2020-v3>.

⁶⁸ Checkliste DMP auf OSF.io, Zugriff am 23.11.2023, <https://help.osf.io/article/144-creating-a-data-management-plan-dmp-document>.

⁶⁹ Muster-Datenmanagementplan DFG. Zugriff am 23.11.2023, <https://www.cms.hu-berlin.de/de/dl/dataman/muster-dmp-dfg>.

⁷⁰ DCC. (2013). „Checklist for a Data Management Plan. v.4.0“. Edinburgh: Digital Curation Centre. Zugriff am 23.11.2023, <http://www.dcc.ac.uk/resources/data-management-plans>.

⁷¹ Muster-Datenmanagementplan BMBF. Zugriff am 23.11.2023, <https://www.cms.hu-berlin.de/de/dl/dataman/muster-dmp-bmbf>.

⁷² Cope, J (2013). „Data Management Planning Guidance for Postgraduate Researchers“. University of Bath, Bath. Zugriff am 23.11.2023, <https://researchportal.bath.ac.uk/en/publications/data-management-planning-guidance-for-postgraduate-researchers>.

⁷³ ICPSR. „Framework for Creating a Data Management Plan“. Zugriff am 23.11.2023, <https://www.icpsr.umich.edu/web/pages/datamanagement/dmp/framework.html>.

⁷⁴ Muster-Datenmanagementplan VW-Stiftung. Zugriff am 23.11.2023, <https://www.cms.hu-berlin.de/de/dl/dataman/muster-dmp-vwstiftung-pdf>.

⁷⁵ DMP-Mustersammlung des DCC. Zugriff am 23.11.2023, <https://www.dcc.ac.uk/resources/data-management-plans/guidance-examples>.

⁷⁶ DMPTool, „Welcome to the DMPTool“. Zugriff am 23.11.2023, <https://dmptool.org/>.

⁷⁷ Research Data Management Organiser. Zugriff am 23.11.2023, <https://rdmorganiser.github.io/>.

⁷⁸ DMPonline, „Plan to make data work for you“. . Zugriff zur Zeit der Erstellung der Version 2 des fachspezifischen TtT nicht möglich, <https://dmponline.dcc.ac.uk/>.

Data Stewardship Wizard ⁷⁹	Erstellung DMP	Zielgruppe: Data Stewards, die Forscher*innen bei der DMP-Erstellung unterstützen
OpenDMP/Argos ⁸⁰	Erstellung DMP	Tool unterstützt das RDA DMP Template ⁸¹

DMP-Tools, Psychologie-spezifisch		
	Ziel	Zusätzliche Informationen
DataWiz	Erstellung DMP, insb. für Psycholog*innen	
BION ⁸²	Erstellung DMP, insb. für Neurowissenschaftler*innen der Universität Gießen	
STAMP	Erstellung DMP	Entwickelt für die Bildungsforschung, ebenso nutzbar für Psycholog*innen

⁷⁹ Data Stewardship Wizard, „Data Stewardship Wizard“. Zugriff am 23.11.2023, <https://ds-wizard.org/>.

⁸⁰ ARGOS, „Plan and follow your data“. Zugriff am 23.11.2023, <https://argos.openaire.eu>.

⁸¹ Allgemeiner Standard für maschinenlesbare Datenmanagementpläne der Research Data Alliance. Zugriff am 23.11.2023, <https://www.rd-alliance.org/group/dmp-common-standards-wg/outcomes/rda-dmp-common-standard-machine-actionable-data-management>.

⁸² BION. Zugriff am 23.11.2023, <https://www.uni-giessen.de/fbz/fb06/psychologie/weitere-inst/bion/bioncu-datenmanagement>.

Organisation von Daten (Einheit 7)

1. Motivation

Eine standardisierte Organisation von Daten gewährleistet,

- dass auch nach Jahren die Art und Weise der Datenerhebung, -analyse und -veröffentlichung nachvollziehbar bleibt
- dass Benennungskonventionen einheitlich und transparent sind
- dass Daten nachgenutzt werden können
- Daten einfacher suchen und schneller finden zu können
- Datenverlust durch Überschreibung oder versehentliches Löschen vorzubeugen
- Maschinenlesbarkeit zu gewährleisten

und führt insgesamt zu einem effizienteren Arbeiten, bietet Transparenz über das eigene Forschungsprojekt und führt durch das Folgen der FAIR-Prinzipien zu mehr Nachhaltigkeit durch die Nachnutzung von Daten.

2. Verzeichnisstruktur

Eine Verzeichnisstruktur (auch Verzeichnisbaum genannt) bezeichnet die hierarchische Anordnung der Ordner, in welchem die zu einem Projekt gehörigen Daten abgelegt werden (siehe Abbildung 5). Die Verzeichnisstruktur sollte klar ersichtlich und damit auch für andere Forschende verständlich sein, um ein leichteres Auffinden der Daten zu gewährleisten. Idealerweise folgen Verzeichnisstrukturen dem Workflow in dem jeweiligen Vorhaben und unterstützen so das meist schrittweise Erstellen, Analysieren und Publizieren der Daten. Für eine gute Übersicht sollten Verzeichnisstrukturen auf Servern mit jenen auf dem lokalen Rechner und/oder einer Cloud übereinstimmen.

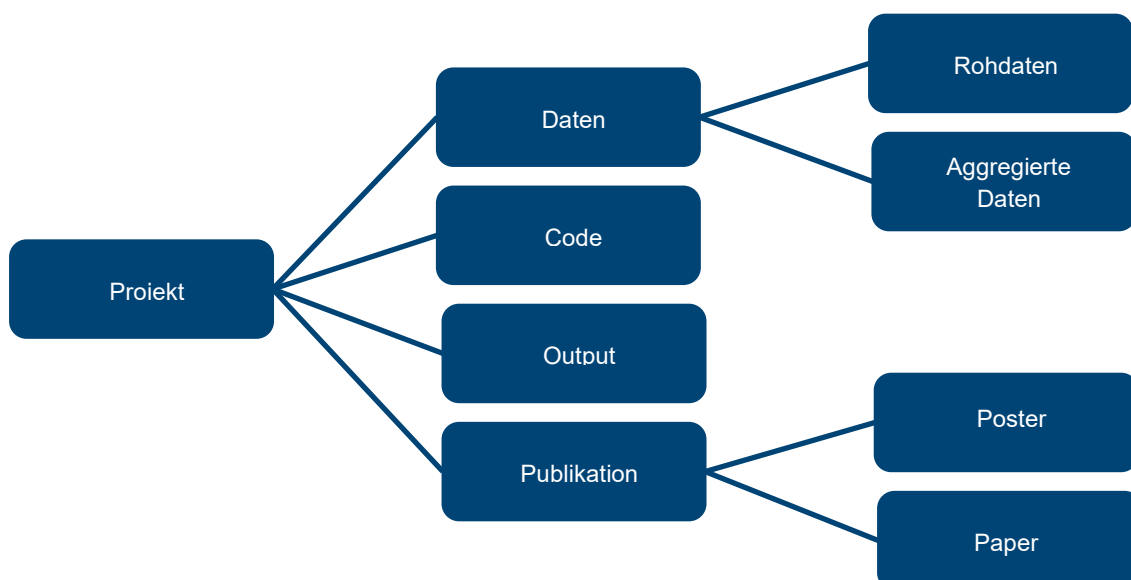


Abbildung 6: Beispiel für eine Verzeichnisstruktur mit Unterordnern.

3. Dateibenennung

Der Dateiname sollte objektiv und intuitiv sowie personenunabhängig nachvollziehbar sein. Die Benennung kann dabei nach den folgenden drei Kriterien erfolgen:

- **System** – wichtig für den späteren Zugriff und Abruf der Daten ist die Berücksichtigung des Systems, unter dem die Datei gespeichert wird.
- **Kontext** – der Dateiname beinhaltet inhaltsspezifische oder deskriptive Informationen, damit unabhängig vom Speicherort klar bleibt zu welchem Kontext die Datei gehört, z. B. „Zeitplan.pdf“ oder „ZeitplanFDNext.pdf“.
- **Konsistenz** – die Namenskonvention sollte vorab gewählt werden, um sicherzustellen, dass sie systematisch befolgt werden kann und die gleichen Informationen (wie z. B. Datum und Zeit) in derselben Reihenfolge beinhaltet (z. B. JJJJMMTT).

Dateinamen sollten so lang wie nötig, aber auch so kurz wie möglich sein, um übersichtlich zu bleiben und innerhalb jedes Betriebssystems lesbar zu sein. Namensbestandteile, die bereits in den Ordnernamen abgelegt sind, müssen nicht nochmal in den Dateinamen wiederholt werden. Um eine einheitliche Namensgebung zu gewährleisten, sind folgende Namensbestandteile hilfreich:

- Inhalt
- Ersteller*in
- Erstellungsdatum
- Bearbeitungsdatum
- Bezeichnung der Arbeitsgruppe
- Publikationsdatum
- Projektnummer
- Versionsnummer

Leerzeichen, Punkte und Sonderzeichen sollten vermieden werden, da sie unter verschiedenen Systemen unterschiedlich interpretiert werden und dies zu Fehlern führen kann. Insbesondere Leerzeichen sollten durch Unterstriche ersetzt werden. Oder man schreibt jeweils den ersten Buchstaben von Wörtern groß.

Um eine chronologische Sortierung zu ermöglichen, empfiehlt es sich, den Namen mit Datumsangabe zu beginnen (JJJJMMTT_Name oder JJJJMMTTName). Beispiele für einheitliche Namensgebung:

- 20200923_VP12_Session1_VALlist1.txt
- 20190520_VP09_Session1_EEGpre_EO.edf
- 20210131_VP23_Interview1.txt

Automatisch generierte Namen (bspw. von der Digitalkamera) sollten der besseren Verständlichkeit wegen entsprechend den obengenannten Konventionen umgeändert werden. Des Weiteren ist es sowohl bei größeren Projekten, als auch bei kleinen Forschungsvorhaben lohnenswert, die gewählten Namenskonventionen schriftlich festzuhalten. Gewählte Abkürzungen sollten dabei insbesondere in einem Datenmanagementplan (siehe **Einheit 6 "Datenmanagementplan"**, S. 27) oder einer README-Datei erläutert werden. Ohne eine solche Datei ist die Rekonstruktion dieser Konventionen nach Jahren oft nur noch schwer möglich.

4. Umbenennen mehrerer Dateien gleichzeitig

Das gleichzeitige Umbenennen mehrerer Dateien kann nützlich sein, u. a.:

- um die automatisch generierten Namen von der Digitalkamera oder anderer Software in einem Arbeitsschritt zu ändern;
- um Leerzeichen oder andere Sonderzeichen aus mehreren Dateinamen in einem Arbeitsschritt zu entfernen bzw. sie zu ersetzen.

Entsprechende Software existiert für die meisten Betriebssysteme, bspw.:

Windows:

- Ant Renamer⁸³
- Rename-IT⁸⁴
- Bulk Rename Utility⁸⁵

Mac:

- Renamer 6⁸⁶ (for Mac)
- Name Changer⁸⁷

Linux:

- GNOME Commander⁸⁸
- GPRename⁸⁹

Unix:

- Unter Unix kann das Kommando „rename“ oder „mv“ hilfreich sein, um mit regulären Ausdrücken Dateien aufzufinden und umzubenennen.

5. Versionskontrolle

Versionen und deren Historie sind nützlich, um einen Überblick über die durchgeführten Schritte zu behalten und diese nachvollziehbar zu machen, und wenn nötig, auch wieder einen Arbeitsschritt zurückzugehen. Bei für die Öffentlichkeit zugänglich gemachten Versionen können diese darüber hinaus auch die Fehlerbehebung unterstützen.

Die verbreitetste Form, Versionen zu kennzeichnen, besteht in der Vergabe von ganzen Zahlen für größere Versionsänderungen und mit einem Unterstrich verbundenen Zahlen für kleinere Veränderungen (z. B. v1, v2, v1_01, v2_03 etc.). Es wird davon abgeraten, Bezeichnungen wie final, final2, revision, definitiv_final zu benutzen.

Bei größeren Projekten (z. B. Verbundprojekte, SFBs) ist die Nutzung einer Versionskontroll-Software (z. B. Git oder Subversion) bei der Verwaltung von Versionen sehr hilfreich. Eine Liste von Open-Source Software zur Versionskontrolle ist auf der Online-Plattform Wikipedia zu finden⁹⁰. Bei

⁸³ Ant Renamer. Zugriff am 23.11.2023, <http://www.antp.be/software/renamer>.

⁸⁴ Rename-IT. Zugriff am 23.11.2023, <http://sourceforge.net/projects/renameit>.

⁸⁵ Bulk Rename Utility. Zugriff am 23.11.2023, <http://www.bulkrenameutility.co.uk/>.

⁸⁶ Renamer 6. Zugriff am 23.11.2023, <http://renamer.com/>.

⁸⁷ Name Change. Zugriff am 23.11.2023, <http://mrrsoftware.com/namechanger/>.

⁸⁸ GNOME Commande. Zugriff am 23.11.2023, <https://gcmd.github.io/>.

⁸⁹ GPRename. Zugriff am 23.11.2023, <http://gprename.sourceforge.net/>.

⁹⁰ Übersicht über Open-Source Software zur Versionskontrolle, Zugriff am 23.11.2023, https://de.wikipedia.org/wiki/Kategorie:Freie_Versionsverwaltungssoftware.

kollaborativen Dokumenten und Speicherorten wie Wikis, Google Docs oder Clouds steht ebenfalls die Funktion der Versionierung und Änderungsverfolgung zu Verfügung.

Beispiele für die Dateibeschriftung mit Versionierung:

- Doe_interview_July2010_V1
- Lipid_analysis_rate_V2
- 2017_01_28_MR_CS3_V6_03

Dokumentation und Metadaten (Einheit 8)

1. Motivation

Damit Forschungsdaten besser auffindbar und nachvollziehbar sind, ist eine Dokumentation der Daten als Basis guter wissenschaftlicher Praxis⁹¹ unverzichtbar. Hilfreich ist hier bspw. eine Checkliste zum Umgang mit Forschungsdaten der DFG⁹² oder rudimentärer die Beantwortung der 5W1H-Fragen (siehe Dokument „**FDNext_TtTKonzept_03_MethodenMaterialien_V2**“, S. 11). Eine Datendokumentation erleichtert zudem erheblich die Nachnutzung der Daten durch eigene oder externe Anschlussforschung. Sie ermöglicht auch die Reproduzierbarkeit von Befunden, die vor allem vor dem Hintergrund der Replikationskrise in der Psychologie⁹³ die Verifizierung von bisherigen Forschungsergebnissen nötig macht.

2. Datendokumentation

Zu einer Datendokumentation gehört unter anderen die gezielte Benennung der Dateien. Darüber hinaus umfasst sie die Ergänzung der Daten, um sogenannte Metadaten (Informationen, welche die Studie und den Erhebungsprozess beschreiben, **siehe Einheit 8 "Dokumentation und Metadaten", Punkt 4**, S. 38), sowie weitere Dokumente⁹⁴, sofern sie zum Verständnis und Interpretation der Daten beitragen (z. B. Protokolle, eMails). Da mit der Zeit Details in Bezug auf Erhebung, Probleme bei den Messungen oder mit Programmen/Gerätschaften usw. in Vergessenheit geraten können, ist es empfehlenswert, die Dokumentation von Daten arbeitsbegleitend zu betreiben. Eine Datendokumentation dient somit der vereinfachten Handhabung und (Nach-)Nutzung der Daten, da sie Entscheidungen bei der Erhebung und Analyse rekonstruierbar und nachvollziehbar macht. Dokumentationsformen sind bspw. ReadMe-Dateien, Tagging der Dateien, Versionierungen, Codebooks/Data Dictionary, Feld- oder Laborbuch, Artikel in einem Data- oder Methoden-Journal oder verwendete Tools.

Zu den grundlegenden Inhalten einer Dokumentation zählen:

- eine Beschreibung des Forschungsvorhabens
- die Projektziele
- Hypothesen
- Informationen zur Erhebung der Daten (Methoden, Einheiten, Zeiträume, Orte, verwendete Instrumente und Software)
- Maßnahmen zur Datenbereinigung
- Angaben zur Struktur der Daten und deren Beziehungen zueinander
- die Erläuterung von Variablen, Labels und Codes
- Angaben zur Versionshistorie

⁹¹ Leitlinien der DFG zur Sicherung guter wissenschaftlicher Praxis. Zugriff am 23.11.2023, https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kode_x_gwp.pdf.

⁹² DFG (2021). „Checkliste zum Umgang mit Forschungsdaten“. Zugriff am 23.11.2023, https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf.

⁹³ Positionspapier der Psychologie-Fachschaften-Konferenz zur Replikationskrise in der Psychologie. Zugriff am 23.11.2023, <https://psyfako.org/wp-content/uploads/30-PsyFaKo-STN-Open-Science-Kulturwandel-zu-einer-offenen-Wissenschaft.pdf>.

⁹⁴ Definition Datendokumentation. Zugriff am 23.11.2023, <https://www.forschungsdaten-bildung.de/daten-dokumentieren>.

- Informationen zum Zugang und Nutzungsbedingungen

Ähnlich wie eine Veröffentlichung in einem Data Journal helfen kann, die Dokumentation der Datensammlung, -aufbereitung und -analyse für das Fachpublikum transparent zu machen (siehe **Einheit 15 „Publikation von Forschungsdaten“**, Punkt 3, S. 54), kann eine Veröffentlichung in einem Methoden-Journal dies in Bezug auf die Art und Weise der Erhebung der Daten inkl. der verwendeten Software und Instrumente erfüllen. Es bleibt anzumerken, dass - im Gegensatz zu Data Journals -, Methoden-Journals ihren Fokus zumeist auf die Veröffentlichung neu entwickelter methodischer Ansätze legen. Sollte jedoch innerhalb einer Studie die Entwicklung eines neuen methodischen Ansatzes ein essentieller Bestandteil sein, kann dies in Fachzeitschriften wie das Journal of visual experiments (JoVE⁹⁵) oder das Psychological Methods⁹⁶ veröffentlicht werden.

3. Codebooks/Data Dictionaries für die Psychologie

Codebooks bzw. Data Dictionaries sind in der Psychologie ein essentieller Bestandteil der Projekt- und Datendokumentation, um die Datenanalyse zu erleichtern und eine spätere Nachvollziehbarkeit der Daten zu ermöglichen. Zahlreiche Anleitungen und Muster für fach- und anwendungsspezifische Codebooks, aber auch Tutorien sind heute u. a. auf diversen Repositorien, wie z. B. Zenodo oder dem Open Science Framework (OSF), zu finden.

Anleitung:

- Shribe-Blog⁹⁷
- Cookbook for Codebook⁹⁸

Fachspezifische Muster:

- Ludwig-Maximilian-Universität München, Codebook⁹⁹ der PsyCourse-Studie (Budde et al., 2019, klinische Längsschnitt- und Beobachtungsstudie zur Klärung der Schizophrenie/Bipolare Störung-Dichotomie)
- Codebook auf SPSS-Basis¹⁰⁰ der ZPID-Nutzerstudie¹⁰¹ (Bedarfserhebung zu Open-Science-Infrastrukturbedarf innerhalb der Psychologie, Abdeckung durch Angebote des Leibniz-Instituts für Psychologie)
- Codebook¹⁰² der PsyCorona-Studie¹⁰³ (Ziel: Identifikation psychologischer und kultureller Faktoren, die die Verbreitung des SARS-CoV-2 beeinflussen)

⁹⁵ Journal of visual experiments (JoVE), Zugriff am 23.11.2023, <https://www.jove.com/de/>.

⁹⁶ Psychological Methods, Zugriff am 23.11.2023, <https://www.apa.org/pubs/journals/met/>.

⁹⁷ Shribe. Zugriff am 23.11.2023, <https://shribe.de/codebuch-erstellen/>.

⁹⁸ Cookbook for Codebook. Zugriff am 23.11.2023, <https://osf.io/e4tqy>.

⁹⁹ Codebook der PsyCourse-Studie an der LMU München. Zugriff am 23.11.2023, <https://data.ub.uni-muenchen.de/251/>.

¹⁰⁰ Codebook auf SPSS-Basis der ZPID-Nutzerstudie, Leibniz-Institut für Psychologie. Zugriff am 23.11.2023, <https://www.psycharchives.org/en/item/9898b1cf-25f0-4119-938b-a349aaa31b9e>.

¹⁰¹ Ergebnisse der ZPID-Studie. Zugriff am 23.11.2023, <https://www.psycharchives.org/en/item/89a4e823-0d35-4d55-ac19-b7fda04f3935>.

¹⁰² Codebook der PsyCorona-Studie. Zugriff am 23.11.2023, <https://osf.io/7kfj5>.

¹⁰³ PsyCorona-Studie der University Groningen. Zugriff am 23.11.2023, <https://www.rug.nl/sustainable-society/research/previous-themes/psycorona/about-the-psycorona-study?lang=en>.

Weitere Muster:

- Codeplan zur Analyse von Veranstaltungsankündigungen in der Weiterbildung, Deutsches Institut für Erwachsenenbildung¹⁰⁴

Die DFG empfiehlt darüber hinaus, in den Studien genutzte Software (Art und Versionen) sowie physische Geräte für die Datenerhebung ebenfalls in die Dokumentation aufzunehmen¹⁰⁵. Dies hilft, Unterschiede in der Datenerhebung und/oder Auswertung durch technisch bedingte Parameter (bspw. physische Unterschiede in der Bauweise, Unterschiede in der Programmierung) im Vergleich mit anderen Studien besser nachvollziehen zu können.

4. Was sind Metadaten?

Metadaten sind strukturierte Daten, die Informationen über andere Daten beinhalten – gleichsam „Daten über Daten“. Sie werden entweder unabhängig oder gemeinsam mit den Daten, die sie beschreiben, gespeichert. Sie können ebenfalls als Primärdaten zur Beantwortung von Forschungsfragen verwendet werden. Hierbei unterscheidet man zwischen **administrativen** (geben Auskunft darüber, welche Art von Anweisungen, Richtlinien und Einschränkungen für eine Datei gelten – z. B. über Zugriffsklassen), **strukturellen** (wie verwandte Objekte miteinander verknüpft sind, bspw. die richtige Reihenfolge der Datenpunkte in einer Interventionsstudie, oder der Interviews in einer Längsschnittstudie) und **deskriptiven** Metadaten (Elemente, die Daten auffindbar und identifizierbar machen, bspw. Titel, Abstract, Author, Keywords, usw.). Sie bilden eine spezifische Untermenge der Dokumentationsangaben und dienen u. a. der Auffindbarkeit der Daten durch Suchmaschinen und in bibliothekarischen Nachweissystemen. Damit sie auch maschinenlesbar sind, zum Beispiel im Semantic Web¹⁰⁶, werden sie häufig im XML-Format¹⁰⁷ gespeichert.

5. Metadatenstandards

Um die Interoperabilität, d.h. die Verknüpfung und das Mapping verschiedener Metadaten zu gewährleisten, stehen sogenannte standardisierte Metadatenvokabulare zur Verfügung. Diese Standards ermöglichen eine inhaltlich und strukturell gleichförmige Beschreibung von ähnlichen Datensätzen. Zu diesem Zweck beinhalten sie eine festgelegte Auswahl an Informationen, die notwendig ist, um diese Daten auffinden und identifizieren zu können. Zu den bekanntesten bibliographischen fachübergreifenden Metadatenstandards zählen Dublin Core¹⁰⁸, DataCite Metadata Schema¹⁰⁹ und MARC21¹¹⁰.

¹⁰⁴ DIE Bonn. Zugriff am 23.11.2023, <https://www.die-bonn.de/docs/codeplan-analyse-veranstaltungsankuendigungen-schrader.pdf>.

¹⁰⁵ Empfehlung der DFG zur Aufnahme von Software und physischen Geräten zur Datenerhebung. Zugriff am 23.11.2023, https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/empfehlungen_forschungsdaten_psychologie_en.pdf.

¹⁰⁶ Das Semantic Web ist eine Erweiterung des Webs um eine maschinenlesbare Schicht und vereinfacht den Datenaustausch zwischen Rechnern.

¹⁰⁷ XML („Extensible Markup Language“) ist eine Auszeichnungssprache zur Darstellung hierarchisch strukturierter Daten im Format einer Textdatei.

¹⁰⁸ DublinCore. Zugriff am 23.11.2023, <https://www.dublincore.org/>

¹⁰⁹ DataCite Metadata Schema. Zugriff am 23.11.2023, <https://schema.datacite.org/>

¹¹⁰ MARC21. Zugriff am 23.11.2023, https://www.dnb.de/DE/Professionell/Metadatendienste/Exportformate/MARC21/marc21_node.html;jsessionid=838CCCAE6368D653EE43AE5BA237C815.intranet231#doc158972bodyText1.

6. Psychologie-spezifische Metadatenstandards

Auch wenn sich viele Fachbereiche spezifische Anforderungen zur Beschreibung und Erschließung ihrer Daten mit Metadaten erarbeitet haben, hat sich für die Psychologie noch kein entsprechender Standard entwickelt. Als Alternative setzte sich jedoch ein für die Sozial- und Wirtschaftswissenschaften existierender Standard (Data Documentation Initiative¹¹¹, DDI) durch. Für einige Teilbereiche innerhalb der Psychologie gibt es jedoch Ansätze dies zu ändern¹¹², in anderen Teilbereichen hingegen gibt es schon Standards: für bildgebende Formate den sogenannten BIDS-Standard¹¹³ und den für das Projekt EBRAINS entwickelten Standard openMINDS^{114,115} (siehe auch **Einheit 13 "Langzeitarchivierung"**, **Punkt 4**, S. 49). Für weitere konkrete Anforderungen innerhalb der Psychologie sollten ebenfalls spezifisch entwickelte Formate in Betracht gezogen werden, wie z. B. für die Statistik¹¹⁶.

Eine im Jahr 2020 veröffentlichte Handreichung beschreibt das Minimum an Metadaten, die in der Psychologie im Rahmen einer Veröffentlichung hilfreich wären (Blask et al., 2020). In einer weiteren Veröffentlichung aus dem 2021 wurde darauf aufbauend ein Nutzenden-orientierter Standard vorgestellt, der spezifisch für die behaviorale Psychologieforschung entwickelt wurde (Blask et al., 2021). Darüber hinaus bietet die DataWiz-Plattform¹¹⁷ für die Psychologie relevante Hinweise, in welcher Art und Weise Metadaten den eigenen Daten hinzugefügt werden können. Ebenfalls hilfreich sind in diesem Zusammenhang verschiedene, von Projekten entwickelte Open-Source-Tools zur Generierung von Metadaten (DataCite Metadata Generator¹¹⁸, CFFNIT¹¹⁹, RADAR¹²⁰).

7. Thesaurus, Normdaten und Klassifikationen in der Psychologie

Um Personen, Institutionen, Projekt-Eckdaten, Orte und vieles mehr ebenfalls eindeutig zu referenzieren, können sie mithilfe von sogenannten Normdateien, Thesauri und Klassifikationen erschlossen werden, welche als kontrollierte Vokabulare bezeichnet werden¹²¹.

Normdateien erleichtern als Identifikatoren beispielsweise die Suche nach Personen bei Namensgleichheit und können von Suchmaschinen eindeutig interpretiert werden. So ist die Psychologie in der sogenannten Gemeinsamen Normdatei (GND, siehe unten) z. B. unter der eindeutigen Nummer „4047704-6“ geführt¹²². Weitere Fachbereiche innerhalb der Psychologie und

¹¹¹ Data Documentation Initiative. Zugriff am 23.11.2023, <https://ddialliance.org/>.

¹¹² GENEMEDE, ein in der Entwicklung befindlicher Metadatenstandard für (Cognitive) Neuroscience. Zugriff am 23.11.2023, <https://genemedede.github.io/>.

¹¹³ BIDS, Metadatenstandard für bildgebende Formate, Zugriff am 23.11.2023, <https://bids.neuroimaging.io/>.

¹¹⁴ openMINDS, Metadatenstandard für das Projekt EBRAINS, Zugriff am 23.11.2023,

<https://github.com/HumanBrainProject/openMINDS>.

¹¹⁵ Präsentation zu openMINDS, Zugriff am 23.11.2023,

<https://flagship.kip.uni-heidelberg.de/jss/HBPM?m=displayPresentation&ml=226&mEID=8809>.

¹¹⁶ DDI-Standards für Statistik. Zugriff am 23.11.2023, <https://rdamsc.bath.ac.uk/subject/Statistics>.

¹¹⁷ Metadata und Dokumentation. Zugriff am 23.11.2023, <https://datawizkb.leibniz-psychology.org/index.php/during-data-collection/what-should-i-know-about-metadata/>

¹¹⁸ DataCite Metadata Generator. Zugriff am 23.11.2023, <https://dhvlab.gwi.uni-muenchen.de/datacite-generator/>.

¹¹⁹ CFFNIT Metadaten. Zugriff am 23.11.2023, <https://citation-file-format.github.io/cff-initializer-javascript/#/start>.

¹²⁰ RADAR als Plugin für RDMO. Zugriff am 23.11.2023, <https://radar.products.fiz-karlsruhe.de/de/radarfeatures/radar-metadaten-schema>; Plugins für RDMO, Zugriff am 23.11.2023, <https://github.com/rdmorganisier/rdmo-plugins>.

¹²¹ Bundesinstitut für Arzneimittel und Medizinprodukte, Zugriff am 23.11.2023,

<https://www.dimdi.de/dynamic/de/klassifikationen/>.

¹²² Die in der GND-Datenbank hinterlegte Nummer für den Begriff „Psychologie“. Zugriff am 23.11.2023, <https://lobid.org/gnd/4047704-6>.

angrenzender Fachgebiete (u. a. Neurologie, Neurophysiologie, Medizin) sind ebenfalls durch Identifikatoren erschlossen.

Wichtige Normdateien sind:

- **GND**¹²³ – Die Gemeinsame Normdatei der Deutschen Nationalbibliothek dient neben der Katalogisierung von Literatur in Bibliotheken als etablierter Standard zur Erschließung und Referenzierung von Information.
- **ISNI**¹²⁴ – Der International Standard Name Identifier dient der eindeutigen Identifikation der öffentlichen Identität von Personen, die an einer Publikation beteiligt sind. Der ISNI ist ein Standard der Internationalen Organisation für Normung (ISO) und ist mit der ORCID vergleichbar.
- **VIAF**¹²⁵ – Der Virtual International Authority File ist eine internationale Normdatei für Personendaten und wird vom Online Computer Library Center (OCLC) verwaltet. Die Normdaten der GND und ISNI bilden einen Bestandteil von VIAF.
- **Open Funder Registry**¹²⁶ – Die Open Funder Registry dient der Identifikation von Forschungsförderern.

Des Weiteren dienen **Klassifikationen** in der Datendokumentation u. a. der Zuordnung von Objekten in (meist hierarchisch strukturierte) Klassen, welche durch bestimmte Merkmale charakterisiert sind. Für die Psychologie ergibt sich z. B. folgende Systematik des ZBW¹²⁷: N.00 Nachbarwissenschaften (bspw. Geschichte, Politik und Politikwissenschaften, Kultur- und Geisteswissenschaften) > N.07 Psychologie > N.07.01 Arbeits- und Organisationspsychologie / N.07.02 Konsumentenpsychologie. Weitere solcher Klassifikationen finden auch für Fachliteratur, Tests, Methoden, Altersgruppen und Evidenzphasen Anwendung¹²⁸.

Bei einem **Thesaurus** handelt es sich um eine natürlich-sprachliche, geordnete Sammlung von Begriffen und deren Beziehungen zueinander. Thesauri und kontrolliertes Vokabular werten Metadaten wesentlich auf und erhöhen die Referenzierbarkeit, Interoperabilität und Auffindbarkeit der Daten. Psychologie-spezifische Thesauri sind hier zu finden:

- Medizin und Biowissenschaften: Thesaurus Medical Subject Headings (MeSH)¹²⁹
- Psychologie:
 - deutschsprachiger Thesaurus PSYINDEX¹³⁰
 - Thesauri der American Psychological Association¹³¹
- Psychologie-spezifische Thesauri mit wirtschaftswissenschaftlicher Ausrichtung:

¹²³ Deutsche Nationalbibliothek. „Gemeinsame Normdatei (GND).“ Zugriff am 23.11.2023, <https://www.dnb.de/>.

¹²⁴ International Standard Name Identifier. „About ISNI.“ Zugriff am 23.11.2023, <https://isni.org/>.

¹²⁵ OCLC, „VIAF: Virtual International Authority File.“ Zugriff am 23.11.2023, <http://viaf.org/>.

¹²⁶ Crossref. „Funder Registry.“ Zugriff am 23.11.2023, <https://www.crossref.org/education/funder-registry/>.

¹²⁷ ZBW Klassifikation Psychologie, Zugriff am 23.11.2023, <https://zbw.eu/stw/version/latest/thsys/70017/about.de.html>.

¹²⁸ PSYINDEX Klassifikation Psychologie. Zugriff am 23.11.2023, <https://www.psyindex.de/ueber/inhalte-aufbau/schlagwoerter-klassifikationen/#update-psyindex-terms-2019>.

¹²⁹ Thesaurus Medical Subject Headings. Zugriff am 23.11.2023, <https://www.nlm.nih.gov/mesh/>.

¹³⁰ PSYINDEX Thesaurus Psychologie. Zugriff am 23.11.2023, <https://www.psyindex.de/ueber/inhalte-aufbau/schlagwoerter-klassifikationen/#update-psyindex-terms-2019>.

¹³¹ APA-Thesaurus für Psychologie. Zugriff am 23.11.2023, <https://www.apa.org/pubs/databases/training/thesaurus>.

- Leibniz-Informationszentrum Wirtschaft (ZBW), Thesaurus für Neurowissenschaften¹³²
- ZBW Psychologie¹³³
- Disziplinspezifische Klassifikationen und Thesauri können über das Basel Register of Thesauri, Ontologies & Classifications (BARTOC)¹³⁴ recherchiert werden.

8. Elektronische Laborbücher (ELB) in der Psychologie

Elektronische Laborbücher (ELB, oder auch ELN für Electronic Lab Notebook) dienen der Dokumentation der Konzeptionierung, Durchführung und Auswertung von wissenschaftlichen Experimenten, Beobachtungen oder Versuchen und den in diesem Zusammenhang erstellten Forschungsdaten. Sie sind die digitalen Versionen von Laborbüchern in Papierformat, die bisher überwiegend in natur- und lebenswissenschaftlichen Disziplinen einen wesentlichen Teil des wissenschaftlichen Arbeitsprozesses darstellten. Mit zunehmender Digitalisierung, v. a. bei der Erhebung von Daten, erfahren die ELB auch eine steigende Wahrnehmung und Nutzung in der Psychologie.

So zeigen sich für die Psychologie Anwendungsmöglichkeiten, wenn es z. B. um die digitale Archivierung ausgefüllter Fragebögen, Daten bildgebender Verfahren (z. B. EEG, MRT) oder die Protokolle einzelner experimenteller Messungen geht. Die Vorteile von ELBs zeigen sich vor allem in der automatisierten und digitalisierten Erhebung von Daten, deren Protokollierung und somit der Nachvollziehbarkeit hinsichtlich verschiedener Verarbeitung- und Analyseprozesse. Fragebögen, Terminvergabe, eMail-Versand usw. können durch automatisierte Prozesse gesteuert werden, was die lückenlose Erhebung von Daten erleichtert. Durch das Vorliegen der Daten in digitaler Form können diese Prozesse durch Wenn-Dann-Verknüpfungen entsprechend den Studienbedingungen gesteuert werden, wenn z. B. Cis*-Frauen oder Trans*-Männer durch die Angabe „Zyklusbeginn“ Terminvorschläge bekommen, um dahingehend eine Interferenz mit den Studienzielen zu vermeiden. Eine weitere Stärke zeigt sich in der automatischen Protokollierung jeder neuen Eingabe und Veränderung hinsichtlich des Autors und der Zeit (siehe **Einheit 7 "Ordnung und Struktur", Punkt 5, S. 34**), so dass jederzeit eine transparente Nachvollziehbarkeit gewährleistet ist.

Es existieren zahlreiche ELB-Softwareangebote - von kommerziellen Anbietern bis hin zu Open-Source-Lösungen, darunter

- REDcap¹³⁵ (Open Source)
- eLabFTW¹³⁶ (Open Source)
- Labfolder¹³⁷ (kommerziell)
- openBIS¹³⁸ (Open Source)
- OSF.io template¹³⁹
- Rspace¹⁴⁰

¹³² ZBW Leibniz-Informationszentrum Wirtschaft, Thesaurus für Neurowissenschaften. Zugriff am 23.11.2023, <https://zbw.eu/stw/version/latest/descriptor/28954-1/about.de.html>.

¹³³ ZBW Leibniz-Informationszentrum Wirtschaft Thesaurus für Psychologie. Zugriff am 23.11.2023, <https://zbw.eu/stw/version/latest/descriptor/15790-5/about>.

¹³⁴ BARTOC.org. Zugriff am 23.11.2023, <http://www.bartoc.org>.

¹³⁵ REDCap, Zugriff am 23.11.2023, <https://projectredcap.org/>.

¹³⁶ eLabFTW ELN. Zugriff am 23.11.2023, <https://www.elabftw.net>.

¹³⁷ LabFolder ELN. Zugriff am 23.11.2023, <https://www.labfolder.com>.

¹³⁸ openBIS Laboratory Notebook. Zugriff am 23.11.2023, <https://openbis.ch>.

¹³⁹ OSF Template für ELN, Zugriff am 23.11.2023, <https://osf.io/6fbrk/>.

¹⁴⁰ Rspace, in R implementierbares ELN, Zugriff am 23.11.2023, <https://www.researchspace.com/>.

In der Regel haben die verschiedenen Disziplinen auch differierende Anforderungen an Funktionalitäten, die eine ELB-Software mit sich bringen sollte, sodass es keine “one-fits-all-Lösung” geben kann. Die ZBMed (Adam 2020) liefert praktische Hinweise, die bei der Entscheidung und Einführung eines ELB in die jeweilige Forschungsgruppe oder das Projekt helfen können. Bei der Einführung von ELBs wird in der Regel eine technische Installation der Software in den vorhandenen IT-Infrastrukturen benötigt, für deren Betreuung auch personelle Ressourcen eingeplant werden müssen. Sollten diese Ressourcen fehlen, zeigt die TU Dresden Möglichkeiten auf, wie ELBs auch mittels herkömmlicher Software erzeugt und in den Arbeitsablauf integriert werden können¹⁴¹.

¹⁴¹ Informationssammlung der TU Dresden. Zugriff am 23.11.2023, <https://tu-dresden.de/forschung-transfer/services-fuer-forschende/kontaktstelle-forschungsdaten/news/elektronisches-laborbuch>.

Zugriffssicherheit (Einheit 9)

1. Motivation

Um Forschungsdaten vor Datenverlust schützen und deren Missbrauch zu verhindern, sollten im Datenmanagement die Aspekte Sicherheit und Zugriffsrechte auch durch entsprechende Maßnahmen reflektiert werden. Insbesondere sensible Daten - personenbezogene Daten oder Daten, deren Schutz vertraglich zugesichert wurde (z. B. Firmengeheimnisse, Auftragsforschung) - erfordern spezifische Schutzmodelle. Aber auch bislang nicht publizierte, wissenschaftliche Ergebnisse bedürfen des Schutzes.

Ein solcher Schutz kann z. B. mittels Verschlüsselung, Backups, Speicherung auf vertrauenswürdigen Speichermedien und einer spezifischen Regelung von Zugriffsrechten sichergestellt werden. Um zu späteren Zeitpunkten entsprechende Maßnahmen nachvollziehen zu können, sollten dazu Protokolle erstellt und sicher aufbewahrt werden. Außerdem sollten nur die Personen Zugriff auf die Daten erhalten, welche diese auch zum Arbeiten benötigen. Gerade bei einer institutionenübergreifenden Zusammenarbeit ist die Regelung dieser Aspekte von besonderer Wichtigkeit.

Es ist zudem sinnvoll, die für die Datensicherheit geeigneten technischen und organisatorischen Maßnahmen einzurichten, zu dokumentieren und immer wieder zu überprüfen, ob diese noch dem aktuellen Bedarf entsprechen.

2. Verschlüsselung

Ein physikalischer Zugriff auf einen Rechner kann einen unbefugten Zugriff auf Daten zur Folge haben. Somit erschließt sich die Notwendigkeit einer Verschlüsselung der Daten, welche jedoch nur sinnvoll ist, wenn auch sämtliche betroffenen Daten verschlüsselt werden und auch allen Beteiligten die Speicherorte bekannt sind. Dies gilt auch für deren Kopien und Backups.

Es können sowohl ausgewählte Speicherorte als auch ganze Datenträger verschlüsselt werden, z. B. durch automatische Verschlüsselungslösungen mit Programmen wie z. B. FileVault¹⁴², BitLocker¹⁴³ oder dm-crypt¹⁴⁴. Dabei müssen die Parameter für alle vorgesehenen Datenspeicherorte gleich eingestellt werden. Es bietet sich ebenfalls eine Dateiverschlüsselung als zusätzliche Schutzmaßnahme bei einem Upload von Daten in eine Cloud-Umgebung an, sowie bei der Weitergabe von Daten per E-Mail. Bei letzterem verhindert sie unberechtigtes Lesen und Verändern von Nachrichtentexten und Anhängen auf dem Weg zum Empfänger (Ende-zu-Ende-Verschlüsselung).

3. Überprüfen der Nutzbarkeit

Durch Festplattenfehler oder Programmabstürze können Daten nicht mehr lesbar werden, weswegen eine Datensicherung nur dann hilfreich ist, wenn die Wiederherstellung der Daten gewährleistet ist. Gelegentlich verursacht auch das Kopieren von Dateien selbst Fehler. Es ist daher grundsätzlich empfehlenswert, die Datenwiederherstellung zu Beginn der Sicherung sowie in regelmäßigen Abständen auf etwaige Fehler in den Daten zu testen, um so einen Datenverlust zu verhindern.

¹⁴² FileVault: Verschlüsselungsfunktion zum Verschlüsseln von persönlichen Daten unter Apple-Betriebssystemen

¹⁴³ BitLocker: Festplattenverschlüsselung von Microsoft

¹⁴⁴ dm-crypt: Verschlüsselungsfunktion unter Linux

Neben der Kontrolle der Lesbarkeit sollte auch eine Virenprüfung durchgeführt werden, um die Überschreibung unbeschädigter Backup-Dateien durch fehlerhafte oder virenbelastete Dateien auszuschließen.

4. Passwortschutz und Zugriffsrechte

Bei der Arbeit mit schutzwürdigen Daten sollten sichere Passwörter vergeben und der Zugriff auf den Kreis der unmittelbar an den Daten arbeitenden Personen beschränkt werden¹⁴⁵.

Ein sicheres Passwort hat folgende Kennzeichen¹⁴⁶:

- es sollte aus mindestens acht Zeichen bestehen (weniger Zeichen erhöhen das Entschlüsselungs-Risiko)
- es enthält Klein- und Großbuchstaben sowie Sonderzeichen und Zahlen
- verwendete Zeichen sollten auf der Tastatur nicht nebeneinanderliegen
- das Passwort sollte möglichst nicht in Wörterbüchern vorkommen

Zu vermeiden sind einfache Passwörter wie 123456, password, 111111, qwertz, abc123 oder admin. Namen, Geburtstage, Begriffe aus Wörterbüchern, Filmfiguren oder Autokennzeichen sollten ebenfalls nicht als Passwort genutzt werden – auch dann nicht, wenn sie rückwärts geschrieben werden. Ebenso ergibt das Anhängen von Zahlen oder Sonderzeichen an ein Wort kein sicheres Passwort.

Über die Vergabe von Berechtigungen wird festgelegt, welche Personen bzw. Personenkreise mit welchen Rechten auf bestimmte Verzeichnisse und Dateien zugreifen dürfen. Dabei ist die Vergabe von abgestuften Lese- und Schreibrechten sowie Ausführungsrechten möglich. Es ist dabei wichtig, die Vergabe wohlgedacht vorzunehmen, um den Arbeitsfluss nicht zu behindern: sind die Zugriffsrechte zu locker vergeben, können Personen unerwünscht auf sensible Daten zugreifen; sind sie hingegen zu restriktiv vergeben, wird eine Nachnutzung erschwert oder gar unterbunden.

¹⁴⁵ Leitlinien der GESIS zum Umgang mit Zugriffsrechten und Sicherheit der Daten. Zugriff am 23.11.2023, https://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-07.pdf

¹⁴⁶ Bundesamt für Sicherheit in der Informationstechnik. „Sichere Passwörter erstellen“. Zugriff am 23.11.2023, https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-Verbraucher/Informationen-und-Empfehlungen/Cyber-Sicherheitsempfehlungen/Accountschutz/Sichere-Passwoerter-erstellen/sichere-passwoerter-erstellen_node.html.

Speicherung und Backup (Einheit 12)

1. Motivation

Um dem Verlust von wertvollen Forschungsdaten entgegenzuwirken, sollten diese regelmäßig gespeichert und zusätzlich durch Backups abgesichert werden. Im Rahmen eines Forschungsdatenmanagementplans kann so für alle Mitglieder einer Forschungsgruppe verbindlich festgelegt werden, wann auf welchen Datenträgern in welchem Format die Daten abgelegt werden sollen, und in welchen Abständen Backups erfolgen sollen. Dies ist vor allem bei institutsübergreifenden Forschungsprojekten sinnvoll, um ein einheitliches Vorgehen zu gewährleisten.

2. Datenspeicherung

Forschungsdaten können auf unterschiedlichen Medien gespeichert werden, die hinsichtlich des Schutzes vor Datenverlust und vor unbefugtem Zugriff ihre jeweiligen Vorteile und Risiken aufweisen (siehe Tabelle 5).

Tabelle 5: Überblick über die Eigenschaften, Vorteile und Risiken der häufigsten Speichermedien und -orte.

	Vorteile	Risiken
Eigener PC	<ul style="list-style-type: none"> • selbst verantwortlich für Sicherheit und Backup • eigene Kontrolle 	<ul style="list-style-type: none"> • evtl. fehlende Ressourcen und Knowhow zum Konfigurieren und Überprüfen der Qualität der Sicherungskopien; • Verlust bzw. Hard- und Softwareprobleme des PC betreffen auch das Backup • Individuallösungen im Rahmen einer Arbeitsgruppe sind aufwendig, kostspielig und ineffizient
Mobile Speichermedien	<ul style="list-style-type: none"> • einfach zu transportieren • können im verschließbaren Schrank oder Safe aufbewahrt werden 	<ul style="list-style-type: none"> • Verlust • Diebstahl • Inhalte sind ungeschützt, wenn sie nicht zuvor verschlüsselt wurden (siehe Einheit 9, Punkt 2, S. 43) • stoß- und verschleißanfällig

<p>Institutionelle Speicherorte</p>	<ul style="list-style-type: none"> • Backup der Daten ist sichergestellt • professionelle Durchführung und Wartung • Speicherung entsprechend den Datenschutzrichtlinien der Institution • Datenschutz über Zugriffsrechte geregelt • für mobiles Arbeiten nutzbar 	<ul style="list-style-type: none"> • Geschwindigkeit vom Netzwerk abhängig • Zugriff auf Backups evtl. verzögert durch Dienstweg • evtl. unklar, welche Sicherheitskriterien und -strategien eingesetzt werden • ggf. mit höheren Kosten verbunden
<p>Externe Speicherorte</p>	<ul style="list-style-type: none"> • einfach zu nutzen und zu verwalten • werden professionell gewartet • für mobiles Arbeiten nutzbar 	<ul style="list-style-type: none"> • je nach Anbieter unsichere Verbindung • Abhängigkeit vom Internetzugang • Upload und Download kann lange dauern • Zugriff auf Backups evtl. verzögert • Datenschutz unklar (ob Sicherheitskriterien und -strategien den Vorgaben für sensible Daten oder der eigenen Institution entsprechen, siehe Freie Universität Berlin¹⁴⁷)

Kostenlose Cloud-Speicherdienste sind keine geeigneten Speicherorte für schützenswerte Daten, da nicht immer sichergestellt wird, wie sicher die Daten hinterlegt sind und was mit den Daten geschieht. Auch wenn es die Aufgabe der Anbieter von Cloud-Diensten ist, missbräuchliche Nutzung von Daten auszuschließen, ist es unklar, ob dies allumfassend gelingt.

Forschende können aber auch auf der materiellen Ebene zur Sicherung ihrer (sensiblen) Daten beitragen (siehe **Einheit 9 "Zugriffssicherheit"**, S. 43). Passwortgeschützte externe Festplatten können z.B. in einem separaten, abschließbaren Raum oder Schrank aufbewahrt werden, und Notebooks vor Diebstahl durch ein Schloss gesichert werden. Dabei sollten mindestens zwei Personen Zugang zu den Daten haben, um auch im Krankheitsfall oder bei Abwesenheit die Verfügbarkeit der Daten zu gewährleisten.

3. Backup-Varianten

Backup bezeichnet die Erstellung einer Sicherungskopie der Daten auf einem anderen Speichermedium. Es sollte bereits frühzeitig eine Strategie entwickelt werden, die regelmäßige Backups und Neukäufe von Ersatzfestplatten als Teil der Arbeitsroutinen miteinschließt, um im Bedarfsfall eine möglichst einfache Datenrekonstruktion sicherzustellen. Viele Betriebssysteme sind zu diesem Zweck bereits mit eigenen Backup-Programmen ausgestattet und erfordern keinen zusätzlichen Installationsaufwand (z. B. die Time Machine-Funktion des Mac OSX Betriebssystems).

¹⁴⁷ Freie Universität Berlin. „Richtlinie zur Auslagerung von Daten in die Cloud“. Zugriff am 23.11.2023, https://www.fu-berlin.de/sites/it-sicherheit/downloads/Richtlinie_Cloud-Datenablage_-_1_0.pdf.

Die Speicherung kann inkrementell oder differenziell erfolgen. Bei beiden Varianten wird zunächst ein Voll-Backup erstellt; anschließend werden bei der

- **inkrementellen Datensicherung** nur die Dateien oder Teile von Dateien gespeichert, die sich seit der letzten inkrementellen Sicherung geändert haben oder neu hinzugekommen sind.
- **differenziellen Datensicherung** alle Daten gespeichert, die sich seit dem letzten Voll-Backup geändert haben oder neu hinzugekommen sind.

4. 3-2-1-Regel: Hinweise für ein sicheres Backup

Es wird empfohlen, die Daten mindestens einmal täglich zu sichern sowie ein wöchentliches Gesamtbackup durchzuführen. Als hilfreich hat sich die "**3-2-1-Regel**" herausgestellt:

- mindestens drei Kopien der Daten,
- auf mindestens zwei verschiedenen Speichermedien
- von denen eine an einem anderen Ort (dezentral) zu hinterlegen ist.

Dezentrale Datenträger sollten von der üblicherweise genutzten Infrastruktur (z. B. Server, PC) abgekoppelt sein und die Datenwiederherstellung zu Beginn sowie in regelmäßigen Abständen auf Fehler überprüft werden.

Die meisten Institutionen bieten zudem automatisierte Lösungen an, bei der alle Daten ausschließlich auf durch das Rechenzentrum bzw. die IT-Abteilung gesicherten Laufwerken abgelegt und durch Backups gesichert werden. Dies verhindert, dass Sicherungen vergessen werden und dass die Konfiguration des Backup-Systems für jeden einzelnen Datenträger einzeln zu erfolgen hat.

Langzeitarchivierung (Einheit 13)

1. Motivation

Um Daten auch in Zukunft recherchier- und ggf. nachnutzbar zu halten, werden sie archiviert. Der Kodex zur Guten Wissenschaftlichen Praxis der DFG fordert hierfür eine i.d.R. 10 Jahre lange Verfügbarkeit der relevanten Forschungsdaten ab dem Zeitpunkt der Veröffentlichung der Forschungsergebnisse (Deutsche Forschungsgemeinschaft, 2013). Darauf basierend haben viele wissenschaftliche Institutionen diese Forderung auch in ihren Forschungsdaten-Policies und Leitlinien aufgenommen.

2. Begriff „Langzeitarchivierung“

„Langzeit“ ist ein Hilfswort für die Beschreibung eines nicht näher definierten Zeitraumes, in dem technologische und soziokulturelle Veränderungen eintreten können, die den Erhalt, den Zugriff auf, die Recherche in und die Nachnutzung von digitalen Forschungsdaten beeinflussen können. Somit muss die digitale Langzeitarchivierung Maßnahmen beinhalten, die geplant, kontrolliert und umgesetzt werden müssen, um diesen Aspekten Rechnung zu tragen.

3. Nachhaltige Dateiformate in der Psychologie

Für eine Langzeitarchivierung ist nicht jedes Dateiformat gleich gut geeignet. Um auch eine zukünftige Arbeit mit Daten zu gewährleisten, sollten Dateiformate gewählt werden, die nicht an kostenpflichtige Software gebunden sind, um z. B. eine höhere Interoperabilität zu erreichen (siehe Tabelle 6). In diesem Zusammenhang unterscheidet man vor allem zwischen sogenannten proprietären und offenen Formaten. Proprietäre Formate sind solche, deren Nutzung eine kostenpflichtige Software erfordert (z. B. Microsoft Office, AutoCAD, SPSS, MaxQDA). Die zu archivierenden Dateien sollten dabei unverschlüsselt, nicht komprimiert, patentfrei und im offenen, dokumentierten Standard erstellt sein.

Mitunter sind proprietäre Dateiformate für die eigene Arbeit unerlässlich. Für die Langzeitarchivierung hingegen sollten etwaige Daten in andere, offene Formate konvertiert werden. Dabei ist zu prüfen, ob die Konvertierung erfolgreich war und das Format valide ist, da Software auch Fehler produzieren kann. Sowohl die Originaldatei als auch die Datei im konvertierten Format sollte gespeichert werden. Eine Übersicht über empfohlene nachhaltige Dateienformate bietet der „Council of Australasian Archives and Records Authorities“¹⁴⁸.

Für die Psychologie existieren neben den fachunabhängigen Empfehlungen^{149,150} für nichtproprietäre Versionen gängiger Text-, Tabellen- und Bildformate auch Empfehlungen für die bildgebenden Verfahren (siehe Tabelle 6). Bei der Aufzeichnung struktureller und funktioneller Daten durch die Magnetresonanztomographie wurden hier schon sehr früh offene Datenformate wie z. B. DICOM¹⁵¹

¹⁴⁸ Übersicht über empfohlene nachhaltige Dateiformate, Zugriff am 23.11.2023, <https://www.caara.org.au/wp-content/uploads/2020/07/Sustainable-Digital-File-Formats-for-Creating-and-Using-Records-V1.0-April-2020.pdf>.

¹⁴⁹ IANUS - Forschungsdatenzentrum für Archäologie & Altertumswissenschaften. „IT-Empfehlungen für den nachhaltigen Umgang mit digitalen Daten in den Altertumswissenschaften. Dateiformate“. Version: 1.0.1.0. Berlin: IANUS - Forschungsdatenzentrum für Archäologie & Altertumswissenschaften, 2017. <http://doi.org/10.13149/000.111000-a>.

¹⁵⁰ UK Data Service. „Recommended formats.“ Zugriff am 23.11.2023, <https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/recommended-formats/>.

¹⁵¹ DICOM Datenformatstandard. Zugriff am 23.11.2023, <https://www.dicomstandard.org/>.

oder NifTI¹⁵² genutzt. In Bezug auf die Aufzeichnung der Gehirnströme mittels Magnet-/Elektroenzephalographie sind die Aufnahmen jedoch häufig noch durch spezifische Datenformate an die Benutzung firmeneigener (proprietärer) Software für die Analyse gebunden (z. B. bei BioSemi, BrainVision, NeuroScan). Durch die Entwicklung spezieller Softwarepakete (Matlab¹⁵³-basierte open source-Anwendungen EEGlab¹⁵⁴ und Fieldtrip¹⁵⁵) ist man jedoch in der Lage, die Verarbeitung und Analysen dieser Daten unabhängig von proprietärer Software zu gestalten^{156,157}.

Dabei empfiehlt es sich, die aufgezeichneten Daten zusätzlich noch in offene Datenformate umzuformatieren, um eine zukünftige Verarbeitung der Daten unabhängiger von den Entwicklungen in der Wirtschaft zu machen. Mit der Entwicklung des Datenformats BIDS¹⁵⁸ wurde Psycholog*innen ein weiterer freier Datenstandard zur Verfügung gestellt. Dieser zeichnet sich vor allem durch die hohe Interoperabilität in Bezug auf Analyseprogramme (siehe die obengenannten Software-Pakete) aus, als auch durch die automatische Implementierung von Metadaten. Zunächst entwickelt für Magnetresonanzaufnahmen, wurde dieser mittlerweile für die Magnet/Elektroenzephalografie erweitert (Niso et al., 2018; Pernet et al., 2019).

Tabelle 6: Empfehlungen nachhaltiger Formate für Langzeitarchivierungen.

Dateiformat	Empfehlung	Vermeiden
Tabellen	CSV, TSV, SPSS portable, XLSX	XLS, SPSS
Text	TXT, HTML, RTF, PDF(/A), DOCX	DOC
Multimedia	Container: MPEG4, MKV Codec: Theora, Dirac, FLAC	QuickTime, Flash
Bilder	TIFF, JPEG2000, PNG	GIF, JPG
Bildgebende Verfahren	European Data Format (EDF) ¹⁵⁹ , General Data Format für biomedizinische Daten (GDF) ¹⁶⁰ , DICOM, BIDS (Standardformat für MRI, MEG und EEG data)	EEG (Firma BrainVision), BDF (Firma BioSemi), CNT (Firma Neuroscan), MEG Formate spezif. Firmen (CTF, FIF, 4d, KIT, KDF, ITAB)

4. Langzeitarchive und deren Anforderungen

Als Langzeitarchive werden im allgemeinen digitale Speicherorte genannt, die nicht bzw. nicht ausschließlich bzw. vorrangig die Veröffentlichung bzw. Zugänglichkeit, sondern die dauerhafte Aufbewahrung von Forschungsdaten ermöglichen. Oft bieten entsprechende Angebote aber auch Möglichkeiten, mittels Zugriffsklassen (siehe **Einheit 15 „Publikation von Forschungsdaten“**,

¹⁵² NifTI Datenformatstandard. Zugriff am 23.11.2023, <https://brainder.org/2012/09/23/the-nifti-file-format/>.

¹⁵³ Matlab. Zugriff am 23.11.2023, <https://www.mathworks.com/products/matlab.html>.

¹⁵⁴ Fieldtrip. Zugriff am 23.11.2023, <https://www.fieldtriptoolbox.org/>.

¹⁵⁵ EEGlab. Zugriff am 23.11.2023, <https://eeglab.org/>.

¹⁵⁶ Von Fieldtrip unterstützte Datenformate. Zugriff am 23.11.2023, <https://www.fieldtriptoolbox.org/faq/dataformat/>.

¹⁵⁷ Von EEGlab unterstützte Datenformate. Zugriff am 23.11.2023, https://sccn.ucsd.edu/eeglab/plugin_uploader/plugin_list_all.php.

¹⁵⁸ BIDS Spezifikationen. Zugriff am 23.11.2023, <https://zenodo.org/record/6094534#.Yt-3cd9CRaQ>.

¹⁵⁹ European Data Format (EDF), Zugriff am 23.11.2023, <https://www.edfplus.info/>.

¹⁶⁰ General Data Format (GDF) for biomedical signals, Zugriff am 23.11.2023, <https://arxiv.org/abs/cs/0608052>.

Punkt 8, S. 60) und Lizenzen (siehe **Einheit 15 „Publikation von Forschungsdaten“**, **Punkt 6**, S. 57) die Nachnutzung der dort archivierten Daten zu regeln (siehe **Einheit 15 „Publikation von Forschungsdaten“**, **Punkt 4**, S. 55). Aus diesem Grund werden - im Gegensatz zur Speicherung (bei der alle Daten eines Projektes in zuvor festgelegten Intervallen gespeichert werden (siehe **Einheit 12 „Speicherung und Backup“**, S. 45),- nur ausgewählte Daten (z. B. die für eine Publikation verwendeten Daten oder Meilensteinversionen) in ein Langzeitarchiv hochgeladen. Als Leitfragen für die Auswahl von Daten für ein Langzeitarchiv können die Empfehlungen für Repositorien herangezogen werden (siehe **Einheit 15 „Publikation von Forschungsdaten, Punkt 2**, S. 53).

Bei der Wahl eines geeigneten Speicherortes für die Langzeitarchivierung sollten folgende Punkte beachtet werden:

- **Technische Anforderungen:** Ein Dienstleister (institutionalisiert oder extern) sollte eine Strategie zur Datenkonvertierung und Migration verfolgen. Darüber hinaus sollte in regelmäßigen Abständen eine Kontrolle der Lesbarkeit der Dateien sowie eine Virenprüfung durchgeführt werden. Alle Schritte sollten dokumentiert werden.
- **Siegel für vertrauenswürdige Langzeitarchive:** Hierzu dienen Siegel mit unterschiedlichen Prüfschwerpunkten (z. B. das nestor-Siegel¹⁶¹ oder das CoreTrustSeal¹⁶²).
- **Kosten der Services:** Es ist immer zu prüfen, ob Dienstleister Kosten für eine Datenspeicherung erheben. Diese können beispielsweise von der Menge der Daten, der Umsetzung von technischen Standards oder der Affiliation der Datenprovider abhängen. Institutionalisierte Repositorien sind hingegen oftmals kostenlos – sollte aber im Vorfeld geprüft werden.
- **Zugänglichmachung der Daten:** Bei der Wahl des Speicherortes ist u.a. das Kriterium der Zugänglichkeit massgeblich.
- **Langlebigkeit des Dienstleisters:** Wirtschaftliche und politische Faktoren haben Einfluss auf die Langlebigkeit der Dienstleister. Im Kontext von Forschungseinrichtungen spielt dieser Aspekt kaum eine Rolle.

¹⁶¹ Nestor-Siegel. Zugriff am 23.11.2023,

https://www.langzeitarchivierung.de/Webs/nestor/DE/Zertifizierung/nestor_Siegel/siegel.html

¹⁶² CoreTrustSeal. Zugriff am 23.11.2023, <https://www.coretrustseal.org/>

Infrastrukturen für das Forschungsdatenmanagement

(Einheit 14)

Nationale, regionale und institutionelle Forschungsdaten-Infrastrukturen unterstützen Forschende bei der Umsetzung von Vorgaben und Empfehlungen wie dem Kodex zur „Guten wissenschaftlichen Praxis“ oder der FAIR-Prinzipien. Entsprechende Service-Portfolios umfassen neben technischer Infrastruktur auch Helpdesks, Beratungsangebote, generische und fachspezifische Schulungen sowie Handreichungen, Leitlinien und weitere Informationsangeboten.

1. Nationale und regionale Angebote für das Forschungsdatenmanagement

Basierend auf einer Empfehlung des Rates für Informationsinfrastrukturen (RFII)¹⁶³ wurde mit der Gründung 2020 der Aufbau einer **Nationalen Forschungsdaten-Infrastruktur**¹⁶⁴ (NFDI) begonnen. Die NFDI soll eine zentrale Rolle dabei einnehmen, passgenaue Lösungen für das fachspezifische FDM zu bündeln und nachhaltig verfügbar zu machen. Während die 27 in der NFDI geförderten Konsortien sich dem fachspezifischen Forschungsdatenmanagement widmen, werden in fünf Querschnitts-Sektionen und einem Basisangebot fachübergreifende Aspekte adressiert. Für die Psychologie sind dabei insbesondere die Konsortien KonsortSWD¹⁶⁵ für die Sozial- und Verhaltenswissenschaften und NFDI4Health¹⁶⁶ (vorrangig hinsichtlich klinischer Psychologie) relevant, sowie das nicht von der NFDI geförderte, aber mit ihr kooperierende NFDI Neuroscience¹⁶⁷.

Darüber hinaus stellt das **Leibniz-Institut für Psychologie** (ZPID¹⁶⁸) eine zentrale, überregionale Infrastruktureinrichtung für die Psychologie in den deutschsprachigen Ländern dar. Durch die Bereitstellung von verschiedenen Services unterstützt es Psycholog*innen bei der Umsetzung eines auf den FAIR-Prinzipien beruhenden FDM. Die für das FDM relevanten Angeboten sind:

- das Repositorium „**PsychArchives**“¹⁶⁹ (siehe Einheit 15 „Publikation von Forschungsdaten“, Punkt 5, S. XX)
- das Assistenzsystem „**DataWiz**“¹⁷⁰ zur Dokumentation des eigenen Forschungsprojektes (siehe Einheit 6 „Datenmanagementplan“, Punkt 5, S. XX)

Neben diesen nationalen Einrichtungen gibt es in Deutschland für die meisten Bundesländer auch verschiedene, projektfinanzierte Landesinitiativen, die auf regionaler Ebene Ressourcen bündeln und auch durch Konzept- und Toolentwicklungen mithelfen, Lücken in verschiedenen FDM-Bereichen zu schliessen (z. B. im Recht, und Policies, usw.). Eine Übersicht über diese, wie auch überregionale und die nachfolgenden institutionellen Infrastrukturen sind zusammen mit den auf Zenodo erhältlichen Konzeptunterlagen für diesen Workshop zu finden (Begleitmaterialien, siehe Einleitung, S. XX).

¹⁶³ Empfehlung des Rates für Informationsstrukturen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, Zugriff am 23.11.2023, <https://rfii.de/download/rfii-empfehlungen-2016/>.

¹⁶⁴ Nationale Forschungsdateninfrastruktur (NFDI), Zugriff am 23.11.2023, <https://www.nfdi.de/>.

¹⁶⁵ KonsortSWD, Zugriff am 23.11.2023, <https://www.konsortswd.de/>.

¹⁶⁶ NFDI4Health, Zugriff am 23.11.2023, <https://www.nfdi4health.de/>.

¹⁶⁷ NFDINeuroscience, Zugriff am 23.11.2023, <https://nfdi-neuro.de/>.

¹⁶⁸ Leibniz-Institut für Psychologie (ZPID), Zugriff am 23.11.2023, <https://leibniz-psychology.org/>.

¹⁶⁹ Psycharchive, vom ZPID bereitgestelltes Repositorium für psychologische Daten, Zugriff am 23.11.2023, <https://www.psycharchives.org/>.

¹⁷⁰ Datawiz, vom ZPID bereitgestelltes Assistenzsystem zur Datendokumentation, Zugriff am 23.11.2023, <https://datawiz.leibniz-psychology.org/DataWiz/>.

2. Institutionelles Forschungsdatenmanagement

Hochschulen und außer-universitäre Forschungseinrichtungen weisen unterschiedlich ressourcenstarke und umfängliche FDM-Infrastrukturen und -Serviceangebote vor. Diese reichen von der einzelnen, halben Stelle bis hin zu gut ausgestatteten, z.T. auch verstetigten Kompetenznetzen. Dabei sind i.d.R. folgende Abteilungen bzw. Einrichtungen beteiligt:

- FDM-Kontakt- bzw. -Servicestelle (wenn vorhanden)
- Bibliothek
- IT/Rechenzentrum
- Forschungsförderung
- rechtliche Anlaufstellen (z. B. Datenschutzbeauftragte, Justizariat)

3. Forschungsdatenmanagement-Infrastrukturen

Hochschulen haben die Aufgabe, ihre Forschenden mit den geeigneten Werkzeugen und Dienstleistungen bestmöglich beim Forschungsdatenmanagement zu unterstützen - sowohl **hinsichtlich der Kompetenzentwicklung und Beratung als auch im Hinblick auf technische Infrastrukturen. Dazu zählen z. B.** folgende Services:

Kompetenzentwicklung, Awareness, Beratung

- Beratung zum Umgang mit Forschungsdaten im Kontext von Projektanträgen bzw. Drittmittelbeantragungen
- Unterstützung bei der Erstellung eines Datenmanagementplans
- Schulungen und/oder Workshops zum Thema Forschungsdatenmanagement
- Informations- und Vernetzungsveranstaltungen
- Handreichungen und Leitlinien
- Rechtliche Beratung zum Thema Forschungsdaten (auch in Kooperation mit Experten)

Technische Infrastruktur

- Archivierung und Speicherung (Langzeitarchivierung)
- Datenpublikation (Repositorium)
- Datenbank-Service
- Datensicherheit: Backup-Service
- Data Sharing: Cloud-Services
- Sync-and-Share-Lösungen
- Versionierungs-Software
- Vergabe von persistenten Identifiern
- DMP-Tool
- Datendokumentation: Elektronisches Laborbuch

Publikation von Forschungsdaten (Einheit 15)

1. Ziele und Motivation

Die Publikation von Forschungsdaten nach den FAIR-Prinzipien (siehe **Einheit 4 „Forschungsdatenmanagement“**, **Punkt 2c**, S. 15) ermöglicht vor allem Anschluss- und Folgeforschung durch andere Forscher und dient der Transparenz der Forschung und der Reproduzierbarkeit der Daten, was die wissenschaftliche Integrität der datenerstellenden Forscher*innen erhöht (eine Replikation beweist die Sorgfalt bei der Datenerhebung und -analyse der ursprünglichen Daten, wenn die Ergebnisse deckungsgleich sind). Zudem gibt es Hinweise, dass Artikel, deren Daten publiziert wurden, häufiger zitiert werden (Pienta et al., 2010; Piwowar & Vision, 2013; Henneken & Accomazzi, 2011; Piwowar et al., 2007), was die Reputation der entsprechenden Forschenden erhöht. Durch die Kontaktaufnahme von an publizierten Daten interessierten Forschenden kann sich zudem auch der Kreis möglicher zukünftiger Kooperationspartner erweitern – wenn sich über die Datennachnutzung hinaus das gemeinsame Interesse an spezifischen Fragestellungen zeigt.

Die Publikation von Daten kann zudem zu einem effizienteren Forschungsprozess beitragen – sie verhindert die doppelte Datenerhebung zu gleichen oder ähnlichen Fragestellungen. Zudem ist nicht jede Forschungseinrichtung in der Lage, ausreichend finanzielle und/oder technische Ressourcen für spezifische Forschung zur Verfügung zu stellen (z. B. übersteigt ein MRT-Gerät oft den finanziellen Rahmen einer Forschergruppe). So können bereitgestellte Daten auch ressourcenschwächeren Forschungskontexten zur Verfügung gestellt werden. Auch im Sinne der Studierendenausbildung in Bezug auf Methodik und Statistik ist die Publikation von Forschungsdaten sinnvoll, da eine erfolgreiche Replikation der Daten den Lernerfolg der Studierenden widerspiegelt.

Nicht zuletzt wird die Zugänglichkeit von Daten zunehmend von Verlagen, Institutionen und Fördermittelgebern gefordert.

2. Leitfragen für die Datenauswahl

Für jede Publikation ist zu entscheiden, unter welchen Bedingungen sie veröffentlicht wird, zum Beispiel, ob sie frei zugänglich gemacht wird (Open Access) oder zugriffsgeschützt archiviert werden soll. Der Wettbewerb innerhalb der Wissenschaft kann eine eingeschränkte oder zeitverzögerte Publikation jedoch sinnvoll machen: Sollen aus den erhobenen Daten noch weitere eigene Publikationen entstehen, so spielen z. B. der Zeitpunkt der Publikation der Daten und die Wahl des Publikationsmodells eine entscheidende Rolle.

Bei der Auswahl können folgende Leitfragen¹⁷¹ helfen:

- Handelt es sich um eine abgeschlossene Datenerhebung oder um einen kumulativen Datensatz (das Projekt beinhaltet mehrere aufeinander folgende Erhebungsphasen)?
- Zu welchem Zeitpunkt im Forschungsprozess werden die Daten publiziert (z. B. sind es noch unbearbeitete Rohdaten, oder bereits Primär- oder Sekundärdaten)?
- Aus welcher Motivation heraus werden sie publiziert?
- Werden Rohdaten oder bearbeitete Daten veröffentlicht?
- Sollen die Daten einem Peer-Review-Prozess unterzogen werden?

¹⁷¹ angelehnt an Martin, Elaine R. (Ed.). *New England Collaborative Data Management Curriculum. Module 6. Data Sharing & Reuse Politics*. Zugriff am 23.11.2023, https://library.umassmed.edu/docs/necdmc_module6.docx.

Nicht immer können Nachnutzungsszenarien vollständig antizipiert werden. Nachfolgende Studien können Datensätze unter gänzlich anderen Forschungsfragen untersuchen (Steiner et al., 2008), als dies in der Ursprungsstudie intendiert war, oder deren Metadaten auswerten (McKiernan et al., 2016). Grundsätzlich empfiehlt sich daher neben der Veröffentlichung auch eine umfassende Dokumentation dieser Forschungsdaten (siehe Einheit 8 „Dokumentation und Metadaten“, Punkt 2, S. 34) auch wenn eine konkrete Anschlussforschung noch unklar ist.

3. Publikationswege

Für die Datenpublikation bieten sich verschiedene Möglichkeiten. Etabliert ist heute **(I)** die Veröffentlichung von aggregierten Daten als Supplement zu einem wissenschaftlichen Artikel (Journal-Publikation). Hinzukommen **(II)** die Datenpublikation in einem generischen, institutionsspezifischen oder fachspezifischen Repositorium als eigenständige Informationsobjekte¹⁷² sowie **(III)** die Veröffentlichung der Datenbeschreibung in sog. Data Journals. Dabei handelt es sich um Fachzeitschriften, die sich auf die Veröffentlichung von Forschungsdaten spezialisiert haben.

(I) Supplement: Daten können als Supplemente zu Veröffentlichungen von wissenschaftlichen Forschungsergebnissen in Fachartikeln (Journals etc.) publiziert werden. Diese Daten ergänzen und belegen die im Artikel präsentierten Forschungsergebnisse. Dabei handelt es sich meist um aggregierte Daten in Form kleiner Tabellen oder Bilder. Die in dieser Form veröffentlichten Daten sind in der Regel nur in Kombination mit dem zugehörigen Artikel auffindbar und zugänglich und entsprechen in den meisten Fällen nicht den FAIR-Prinzipien.

(II) Repositorium: Daten können als ein eigenständiges Informationsobjekt in einem Repositorium publiziert werden. Sofern diese Daten Grundlage eines wissenschaftlichen Artikels sind, können sie in der Regel über die Metadaten mit der Publikation verknüpft werden (siehe Repositorien). In disziplinspezifischen Repositorien können diese Daten zudem besser kontextualisiert bzw. mit anderen Datensammlungen verknüpft werden, indem sie zusätzliche Recherche-, Analyse- und Visualisierungswerkzeuge bereitstellen. Disziplinübergreifende und institutionelle Repositorien sind zwar nicht fachbereichsspezifisch ausgelegt, die Daten selbst sind aber aufgrund ihrer Metadaten entsprechend thematisch such- und auffindbar. Zudem können institutionelle Repositorien eine gute Übersicht der an einer Institution gesammelten Daten bieten, sofern diese auf dem Repositorium abgelegt werden.

(III) Data Journals: Data Journals widmen sich der Veröffentlichung von Daten und entsprechenden Informationen zu diesen publizierten Daten, die in frei zugänglichen oder zugangsbeschränkten Repositorien publiziert werden. Bei den entsprechenden Informationen handelt es sich um ausführliche Dokumentationen von veröffentlichten Daten, deren Eigenschaften sowie Angaben zur potentiellen Nachnutzung. Die Daten im Repositorium und ihre Dokumentation in der Zeitschrift werden mittels eines persistenten Identifiers (siehe Identifier) miteinander verknüpft und sind somit eindeutig auffindbar. Zudem bieten einige dieser Zeitschriften ein Peer-Review-Verfahren an, bei dem der Datensatz sowie seine Dokumentation durch externe Gutachter geprüft werden, z. B. ob die jeweilige Dokumentation die damit verknüpften Daten abbildet, ob die Dokumentation die Daten ausreichend erklärt, und ob die Dateiformate standardisiert sind. Beispiele solcher Datenzeitschriften¹⁷³

¹⁷² Informationsobjekte, auch Entitäten genannt, können aus Datenbanken, Dokumenten, Nachrichten oder jeder weiteren Art von Information bestehen, Zugriff am 23.11.2023, <https://project-base.org/projektmanagement-glossar/informationsobjekt-eeepk/>.

¹⁷³ Übersicht über Data Journals der HU Berlin, Zugriff am 23.11.2023, <https://www.cms.hu-berlin.de/de/dl/dataman/teilen/dokumentation/datajournal>.

stellen die Open-Access-Zeitschrift „Journal of Open Psychology Data“¹⁷⁴ für die Psychologie oder disziplinübergreifend „Data in Brief“¹⁷⁵ dar.

4. Repositorien

Repositorien sind Speicherorte, in denen Objekte archiviert, dokumentiert und publiziert werden können. Dabei handelt es sich um Speicherorte für Daten, die eine vom wissenschaftlichen Artikel zeitlich und räumlich unabhängige Veröffentlichung und Archivierung von Daten ermöglichen, i.d.R. für den Regeln der Guten wissenschaftlichen Praxis folgenden Zeitraum von mindestens 10 Jahren.

Je nach Repository können Daten, Datensätze, Experiment- und Auswertungsbeschreibungen, audiovisuelle Objekte wie Bild- und Videodateien, Modelle von Simulationen oder auch Software veröffentlicht werden. In manchen Repositorien können ganze Forschungsdatensätze in ihrer komplexen Form als Einheit abgelegt werden.

a) Fachspezifische Repositorien

Fachrepositorien bieten den Vorteil der höheren Sichtbarkeit in der jeweiligen Forschungscommunity sowie mehr Möglichkeiten, Daten entsprechend den Anforderungen und Standards des jeweiligen Fachgebiets zu dokumentieren bzw. mit Metadaten zu beschreiben, um die Such- und Auffindbarkeit in den entsprechenden Nachweissystemen zu gewährleisten.

Beispiele für Fachrepositorien in der Psychologie:

- PsychArchives¹⁷⁶
- Forschungsdatenzentrum am ZPID¹⁷⁷
- GenBank¹⁷⁸ für Genforschungsdaten
- openNeuro¹⁷⁹ (**Bitte beachten Sie: die Server von openNeuro liegen in den USA und sind hinsichtlich personenbezogener FD nicht an die DSGVO gebunden**)
- brainlife¹⁸⁰

b) Generische Repositorien

Bei interdisziplinärer Forschung kann die Zuordnung der entstandenen Daten zu einem Fachgebiet schwierig sein. Hier können disziplinübergreifende Repositorien zur Publikation bzw. Speicherung dienen. Sie akzeptieren in der Regel eine Vielzahl von Datentypen und bieten umfangreiche allgemeine Suchfunktionen. Eine Kuratierung der Daten oder andere Formen der inhaltlichen Qualitätskontrolle wird dabei jedoch nur selten geleistet. Beispiele für fachübergreifende Repositorien:

- ZENODO¹⁸¹ Digitale Forschungsdaten und Publikationen aus allen Forschungsgebieten, Produkt des EU OpenAIRE¹⁸²-Projekts (open source)

¹⁷⁴ Ubiquity Press, „Journal of Open Psychology Data.“ Zugriff am 23.11.2023, <https://openpsychologydata.metajnl.com/>.

¹⁷⁵ Elsevier. „Data in Brief.“ Zugriff am 23.11.2023, <https://www.journals.elsevier.com/data-in-brief>.

¹⁷⁶ PsychArchives. Zugriff am 23.11.2023, <https://www.psycharchives.org/>.

¹⁷⁷ Forschungsdatenzentrum am ZPID, Zugriff am 23.11.2023, <https://rdc-psychology.org/de/homepage-deutsch>.

¹⁷⁸ GenBank. Zugriff am 23.11.2023, <http://www.ncbi.nlm.nih.gov/genbank>.

¹⁷⁹ OpenNeuro. Zugriff am 23.11.2023, <https://openneuro.org/>.

¹⁸⁰ Brainlife. Zugriff am 23.11.2023, <https://brainlife.io/about/>.

¹⁸¹ ZENODO. Zugriff am 23.11.2023, <https://zenodo.org>.

¹⁸² OpenAIRE, ein dem Open Science-Paradigma verpflichtetes paneuropäisches Forschungsinformationssystem, hat die Verbesserung der Auffindbarkeit und Wiederverwendbarkeit von Forschungspublikationen und -daten zum Ziel.

- DRYAD¹⁸³ (open source)
- Figshare¹⁸⁴ Digitale Bilddaten aus allen Forschungsgebieten (kommerziell)

c) Institutionelle Repositorien

Institutionelle Repositorien bieten eine Alternative, wenn kein geeignetes Fachrepositorium vorhanden ist. So können für Forschende unklare rechtliche Rahmenbedingungen in Bezug auf den Umgang mit Daten am Standort des externen Repositoriums für eine Publikation im Repositorium der eigenen Institution sprechen. Institutionelle Repositorien, die i.d.R. auch von entsprechenden Beratungs- und Unterstützungsangeboten begleitet werden, sind in der Regel für alle an der Forschungseinrichtung vertretenen Fachgebiete offen und im Rahmen bestimmter Datenvolumen kostenfrei nutzbar. Darüber hinaus bieten sie aber auch die Möglichkeit, lediglich die Metadaten von bereits anderweitig publizierten oder nicht publizierten Daten nachzuweisen.

Beispiele für institutionelle Repositorien an Hochschulen:

- DepositOnce¹⁸⁵ der Technischen Universität Berlin
- edoc-Server¹⁸⁶ der Humboldt-Universität zu Berlin
- OpARA: an der Universität Leipzig¹⁸⁷, der Technischen Universität Dresden¹⁸⁸, der HTW Dresden und der TU Bergakademie Freiberg
- Refubium¹⁸⁹ der Freien Universität Berlin
- PUB¹⁹⁰ der Universität Bielefeld
- heiDATA¹⁹¹ der Universität Heidelberg
- Edmond¹⁹² der Max-Planck-Gesellschaft
- hilData¹⁹³ der Universität Hildesheim

5. Auswahl eines Repositoriums

Für die Suche nach einem geeigneten Repositorium dient das disziplinübergreifende Verzeichnis re3data¹⁹⁴, welches die Suche nach verschiedensten Aspekten wie Fachdisziplin, Art der Daten oder Standort des Servers ermöglicht, ob der Daten-Upload kostenpflichtig ist oder in welchem Repositorium die Datennutzung eingeschränkt ist. Weitere Suchmaschinen sind OpenDOAR¹⁹⁵ oder ROAR¹⁹⁶.

¹⁸³ DRYAD. Zugriff am 23.11.2023, <http://datadryad.org>.

¹⁸⁴ Figshare. Zugriff am 23.11.2023, <https://figshare.com>.

¹⁸⁵ DepositOnce. Zugriff am 23.11.2023, <https://depositonce.tu-berlin.de>.

¹⁸⁶ edoc-Server HU Berlin. Zugriff am 23.11.2023, <https://edoc.hu-berlin.de>.

¹⁸⁷ OpARA an der Universität Leipzig, Zugriff am 23.11.2023, <https://www.urz.uni-leipzig.de/ueber-uns/rechtliche-grundlagen/regeln-opara>.

¹⁸⁸ OpARA an der Technischen Universität Dresden, Zugriff am 23.11.2023, <https://opara.zih.tu-dresden.de/xmlui/>

¹⁸⁹ Refubium. Zugriff am 23.11.2023, <https://refubium.fu-berlin.de>.

¹⁹⁰ PUB. Zugriff am 23.11.2023, <https://pub.uni-bielefeld.de>.

¹⁹¹ HeiDATA. Zugriff am 23.11.2023, <https://heidata.uni-heidelberg.de>.

¹⁹² EDMOND der Max-Planck-Gesellschaft. Zugriff am 23.11.2023, <https://edmond.mpdl.mpg.de/>

¹⁹³ hilData, Zugriff am 23.11.2023, <https://hildata.uni-hildesheim.de/>.

¹⁹⁴ re3data.org - Registry of Research Data Repositories. Zugriff am 23.11.2023, <https://doi.org/10.17616/R3D>.

¹⁹⁵ OpenDOAR Suchmaschine. Zugriff am 23.11.2023, <https://v2.sherpa.ac.uk/opensoar/>.

¹⁹⁶ ROAR Suchmaschine. Zugriff am 23.11.2023, <http://roar.eprints.org/>.

Für die Wahl eines Repositoriums empfiehlt OpenAIRE die folgende Priorisierung¹⁹⁷:

1. Externes, etabliertes fachspezifisches Repositorium
2. Institutionelles Repositorium
3. Kostenfreies generisches Repositorium
4. Suche via re3data.org

Kriterien für die Auswahl eines geeigneten Repositoriums¹⁹⁸:

1. Zertifizierung, z. B. CoreTrustSeal (siehe **Einheit 13 "Langzeitarchivierung", Punkt 4**, S. 49)
2. (Automatisierte) Vergabe von persistenten Identifiern, z. B. Digital Object Identifier (siehe **Einheit 15 „Publikation von Forschungsdaten“, Punkt 7a**, S. 59)
3. Zugang zu Daten: offen, eingeschränkt oder unzugänglich
4. Nutzungsbedingungen für Datenautor*innen sowie Nutzer*innen, z. B. Entgelte, Embargofristen
5. Möglichkeit der Lizenzauswahl

6. Lizenzen

Die Nutzung publizierter Daten durch Dritte kann durch verschiedene Lizenzierungsmodelle geregelt werden. Etabliert haben sich z. B. die kostenfreien Lizenzen von **Creative Commons** (CC) für die offene Nutzung. Dabei stehen folgende Kennungen zur Verfügung:

BY = Namensnennung; Namensnennung des Urhebers erforderlich.

NC = non-commercial (nicht kommerziell); Nutzung unter kommerziellen Zwecken nicht erlaubt.

ND = no derivatives (keine Bearbeitung); Veränderung des Werkes nicht erlaubt

SA = share alike (Weitergabe unter gleichen Bedingungen); Nach der Bearbeitung muss das Werk unter der gleichen Lizenz weitergegeben werden.

Bei der Vergabe der Lizenz bestehen folgende Möglichkeiten:

- CC0 (Gemeinfreiheit/Public Domain)¹⁹⁹
- CC BY (Namensnennung)²⁰⁰
- CC BY-ND (Namensnennung – Keine Bearbeitung)²⁰¹
- CC BY-NC (Namensnennung – Nicht-kommerziell)²⁰²
- CC BY-SA (Namensnennung – Weitergabe unter gleichen Bedingungen)²⁰³
- CC BY-NC-SA (Namensnennung – Nicht-kommerziell – Weitergabe unter gleichen Bedingungen)²⁰⁴

¹⁹⁷ OpenAIRE. „How to select a repository?“ Zugriff am 23.11.2023, <https://www.openaire.eu/find-trustworthy-data-repository>.

¹⁹⁸ re3data.org. „FAQ.“, Zugriff am 23.11.2023, <https://www.re3data.org/faq>.

¹⁹⁹ Creative Commons Public Domain Dedication (CC0). Zugriff am 23.11.2023, <https://creativecommons.org/publicdomain/zero/1.0/>.

²⁰⁰ Creative Commons Namensnennung (CC BY). Zugriff am 23.11.2023, <https://creativecommons.org/licenses/by/4.0/>.

²⁰¹ Creative Commons Namensnennung - keine Bearbeitung (CC BY-ND). Zugriff am 23.11.2023, <https://creativecommons.org/licenses/by-nd/4.0/>.

²⁰² Creative Commons Namensnennung – Nicht-kommerziell (CC BY-NC). Zugriff am 23.11.2023, <https://creativecommons.org/licenses/by-nc/4.0/>.

²⁰³ Creative Commons Namensnennung – Weitergabe unter gleichen Bedingungen (CC BY-SA). Zugriff am 23.11.2023, <https://creativecommons.org/licenses/by-sa/4.0/>.

²⁰⁴ Creative Commons Namensnennung – Nicht-kommerziell – Weitergabe unter gleichen Bedingungen (CC BY-NC-SA). Zugriff am 23.11.2023, <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

- CC BY-NC-ND (Namensnennung – Nicht-kommerziell – Keine Bearbeitung)²⁰⁵

Die Vergabe einer Creative Commons-Lizenz selbst hat **keinen Einfluss auf das Urheberrecht** (siehe **Einheit 5 "Rechtliche Aspekte", Punkt 4**, S. 25), bietet jedoch eine einfache und standardisierte Möglichkeit die Verwendung der Forschungsdaten zu regeln. Es wird dabei empfohlen, die Nutzung der Forschungsdaten so weit wie möglich zu gewähren. Dabei ist zu beachten, dass bei der Vergabe einer CC0-Lizenz der/die Urheber*in auf alle urheberrechtlichen und verwandten Schutzrechte verzichtet, wohingegen alle weiteren CC-Lizenzen die Nutzungsrechte unterschiedlich stark einschränken.

Ein weiteres Lizenzierungsmodell bieten die Open Data Commons (ODC)²⁰⁶:

- Open Data Commons Open Database License (ODbL)²⁰⁷ (entspricht der CC BY-SA Lizenz)
- Open Data Commons Attribution License (ODC-By)²⁰⁸ (entspricht der CC BY Lizenz)
- Open Data Commons Public Domain Dedication and License (PDDL)²⁰⁹ (entspricht der CC0 Lizenz)

Die **ODC-Lizenzen** eignen sich vor allem für Datenbankrechte, Patentrechte und Warenzeichen. Bei den Datenbank-Lizenzen können für die in entsprechend lizenzierten Datenbanken enthaltenen Inhalte jeweils unterschiedliche Bedingungen festgelegt werden.

Wurde im Rahmen eines Forschungsprojektes eine Software zur Analyse psychologischer Daten entwickelt, und möchte man diese Software zur freien Verfügung stellen (z. B. Fieldtrip²¹⁰), benötigt man hingegen eine gesonderte Lizenz, um den besonderen Anforderungen dieses Formats gerecht zu werden (Einverständnis zur Installation, Veränderung, Ausführung, Zweck oder Ort der Nutzung, Anzahl der Benutzer etc.). Die Creative Commons-Lizenzen können dies nicht abdecken. Hierfür empfiehlt sich eine der gängigen Software-Lizenzen wie z. B. die MIT-Lizenz²¹¹, GNU General Public License (GPL)²¹², GNU Lesser General Public License (LGPL)²¹³ oder die Apache-Lizenz²¹⁴.

²⁰⁵ Creative Commons Namensnennung – Nicht-kommerziell – keine Bearbeitung (CC BY-NC-ND). Zugriff am 23.11.2023, <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

²⁰⁶ Open Knowledge Foundation. „Open Definition. Defining Open in Open Data, Open Content and Open Knowledge. Home.“. Zugriff am 23.11.2023, <https://opendatacommons.org/>.

²⁰⁷ Open Knowledge Foundation. „Open Definition. Defining Open in Open Data, Open Content and Open Knowledge. Open Data Commons Open Database License (ODbL).“. Zugriff am 23.11.2023, <https://opendatacommons.org/licenses/odbl/>.

²⁰⁸ Open Knowledge Foundation. „Open Definition. Defining Open in Open Data, Open Content and Open Knowledge. Open Data Commons Attribution License (ODC-By).“. Zugriff am 23.11.2023, <https://opendatacommons.org/licenses/by/>.

²⁰⁹ Open Knowledge Foundation. „Open Definition. Defining Open in Open Data, Open Content and Open Knowledge. Open Data Commons Public Domain Dedication and License (PDDL).“ Zugriff am 23.11.2023, <https://opendatacommons.org/licenses/pddl/>.

²¹⁰ Die für die Toolbox Fieldtrip gültige Lizenz, Zugriff am 23.11.2023, <https://www.gnu.org/licenses/gpl-3.0.html>.

²¹¹ Open-Source Initiative. „The MIT License.“. Zugriff am 23.11.2023, <https://opensource.org/licenses/mit-license.php>.

²¹² GNU General Public License Version 3.0. Zugriff am 23.11.2023, <https://www.gnu.org/licenses/gpl-3.0.en.html>.

²¹³ GNU Lesser General Public License Version 3.0. Zugriff am 23.11.2023, <https://www.gnu.org/licenses/lgpl-3.0.html.en>.

²¹⁴ The Apache Software Foundation. „Apache License 2.0“. Zugriff am 23.11.2023, <https://www.apache.org/licenses/>.

Zu den Lizenzen, die die Idee von Open Data, und somit die Idee von Open Science (siehe **Einheit 4 „Forschungsdatenmanagement“, Punkt 2c**, S. 15) fördern, zählen CC0, CC BY, CC BY-SA, PDDL, ODC-By und ODbL²¹⁵.

Wird jedoch keine Lizenz vergeben, ist eine Nachnutzung ohne das Einverständnis der Urheber*innen nicht möglich, d. h. eine fehlende Lizenz bedeutet nicht automatisch lizenzfrei im Sinne von z. B. CC0 (**keine Lizenz ≠ nicht lizenzfrei**).

7. Persistente Identifier

Um digitale Publikationen langfristig auffindbar, eindeutig referenzier- und zitierbar zu machen, sowie zur Verbesserung der Mensch-/Maschinenlesbarkeit, werden persistente Identifier (PIDs) eingesetzt. Jedes mit diesem Identifier verlinkte Objekt (z. B. eine Publikation oder eine Person) wird hierbei eineindeutig gekennzeichnet, was den Identifier zu einem Teil der Metadaten macht.

Etablierte ID-Systeme sind der „Digital Object Identifier“ (**DOI**) für die Kennzeichnung von Publikationen und Daten (DOI-Dienstleister: CrossRef²¹⁶ für Zeitschriftenartikel, DataCite²¹⁷ für Datensätze) und die „Open Researcher and Contributor ID“ (**ORCID**) für eine eindeutige Identifikation wissenschaftlicher Autor*innen. Die Vergabe von persistenten Identifiern kann u. U. mit Kosten verbunden sein, die in dem Datenmanagement mitbedacht werden sollten: bei der Publikation von Ergebnissen (und damit verbunden die Vergabe eines Identifiers) werden seitens der Journals in der Regel Gebühren erhoben.

a) Digital Object Identifier (DOI)

International etabliert ist der zitierfähige Digital Object Identifier (DOI). Die International DOI Foundation (IDF) garantiert dabei einheitliche Standards und Workflows bei der Vergabe von DOIs (Paskin, 2010), welche durch einen seit Mai 2012 gültigen ISO-Standard²¹⁸ reguliert ist.

Ein DOI

- ist eine einzigartige Folge von alphanumerischen Zeichen
Zugelassene Zeichen sind: a – z, A – Z, 0 – 9, . (Punkt), - (Bindestrich), _ (Unterstrich), : (Doppelpunkt) sowie / (Schrägstrich).
- besteht aus zwei Teilen: einem Präfix, der die vergebende Organisation kennzeichnet, und einem Suffix, der das Objekt identifiziert

Beispiele für die Gestaltung des Suffixes:

- Ursprungs-DOI: 10.1234/abc123
- DOI einer neuen Version: 10.1234/abc123.1
- DOI eines Teils: 10.1234/abc123/2

Ein DOI ermöglicht es u. a. die Daten mit zugehörigen Publikationen eindeutig zu verknüpfen. Dadurch bleiben die Daten dauerhaft zitierbar. DOIs werden häufig durch Repositorien bzw. Institutionen wie Bibliotheken vergeben. DataCite verteilt z. B. sogenannte Nummernkreise an Institutionen (Registrar), die diese dann individuell vergeben (Registrant).

²¹⁵ Open Knowledge Foundation. „Open Definition. Defining Open in Open Data, Open Content and Open Knowledge. Conformant Licenses.“ Zugriff am 23.11.2023, <https://opendefinition.org/licenses/>.

²¹⁶ CrossRef, Zugriff am 23.11.2023, <https://www.crossref.org/>.

²¹⁷ DataCite. Zugriff am 23.11.2023, <https://datacite.org>.

²¹⁸ International Organization for Standardization. ISO 26324:2012 Information and documentation — Digital object identifier system. 2012. Zugriff am 23.11.2023, <https://www.iso.org/standard/43506.html>.

Im Gegensatz zu DOIs sind Handles²¹⁹ als globales Referenzsystem für große Datenmengen weniger persistent und nicht durch einen Standard abgesichert. Sie eignen sich daher vor allem für die Referenzierung von Daten vor dem Publizieren.

b) Open Researcher and Contributor ID (ORCID)

Um zu gewährleisten, dass wissenschaftliche Arbeiten einem*r Autor*in unabhängig von Institutions- und Namenswechseln eindeutig zugeordnet werden, kann die sogenannte ORCID²²⁰ verwendet werden. ORCID vereinfacht zudem die Referenzierung von Publikationslisten und Angaben zur Autor*innenschaft, z. B. durch die Vermeidung einer falschen Referenzierung bei Namensgleichheit in derselben Fachdisziplin.

Die ORCID-ID

- ist ein (alpha-)numerischer 16-stelliger Code.
- Verschafft eine eindeutige wissenschaftliche Identität (auch über Namenswechsel, Tippfehler oder Namensgleichheit hinweg).
- wird von Zeitschriften, Forschungsförderern und Institutionen genutzt.
- wird von Forschenden selbst gepflegt.
- wird von einer gemeinnützigen Initiative betrieben.
- hat Verbindung zu Web of Science, Scopus, Zenodo, DataCite, u. a.

c) Research Activity ID (RAiD)

Der RAiD²²¹ ist ein recht neuer, vom ARDC²²² entwickelter persistenter Identifier für Forschungsprojekte, der angedacht ist, sämtliche Forschungsaktivitäten (u. a., ORCID und DOIs) unter einem Identifier zusammenzufassen. Ähnlich wie ORCID muss der RAiD jedoch selbst gepflegt werden, bietet dafür aber einen tracking record über die Dauer des Projekts.

8. Zugriffsklassen/Agreements

Um den Zugriff auf Forschungsdaten für eine Ansicht oder Nachnutzung zu regeln, können auch sogenannte Zugriffsklassen bestimmt werden, die vor allem in Repositorien Anwendung finden. In der aktuellen Fassung ihrer Forschungsdaten-Policy²²³ definiert die DGPs 4 Klassen (Klasse 0-3), um diesen Zugriff zu regeln. Auf der niedrigsten Stufe (**Zugangsklasse 0**) besitzen veröffentlichte Daten keinerlei Zugangsbeschränkungen und werden unter CCO- oder ODC-Lizenzen Nutzenden zur Verfügung gestellt. Die Abfrage von Informationen über die Identität des Nutzenden, deren*dessen Zugehörigkeit zu einer Forschungseinrichtung und/oder die angedachte Nutzung der Daten ist möglich.

Die Daten der nächsthöheren **Zugangsklasse 1** werden ebenfalls grundsätzlich frei zur Verfügung gestellt; ihre Nutzung ist jedoch nur unter der expliziten Zustimmung der Nutzungsbedingungen des*r Urhebers*in (data sharer) möglich. Jegliche Zuwiderhandlung kann hier somit rechtlich verfolgt werden. Zusätzlich erfragen manche Repositorien hier ebenfalls Informationen über die Identität, deren*dessen Zugehörigkeit und/oder die angedachte Nutzung.

²¹⁹ Handle.net Registry, Zugriff am 23.11.2023, <https://www.handle.net/>.

²²⁰ ORCID, Zugriff am 23.11.2023, <https://orcid.org/>.

²²¹ RAiD, Research Activity Identifier, Zugriff am 28.04.2023, <https://ardc.edu.au/services/ardc-identifier-services/raid-research-activity-identifier-service/>.

²²² ARDC, Australian research data common, Zugriff am 23.11.2023, <https://ardc.edu.au/about-us/>.

²²³ Management und Bereitstellung von Forschungsdaten in der Psychologie: Überarbeitung der DGPs-Empfehlungen, Zugriff am 23.11.2023, <https://psyarxiv.com/hcxtm/>.

Daten mit einer grundsätzlichen Zugangsbeschränkung haben mindestens die **Zugangsklasse 2**. Die Nutzungsbestimmungen werden oftmals mittels individualisierter Nutzungsverträge zwischen Urheber*in und Nutzenden geregelt. Auch hier, wie auch in der nächst höheren Stufe, gilt: Zuwiderhandlungen der explizit zugestimmten Nutzungsbedingungen können rechtliche Folgen nach sich ziehen.

Die **Zugangsklasse 3** bildet die restriktivste Klasse; sie erlaubt den Datenzugriff nur unter der expliziten Zustimmung zu den Bedingungen des Datenschutzgesetzes (z. B. online mittels einer geeigneten gesicherten Verbindung zwischen Repositorium und Datennachnutzenden oder lokal). Hierbei gilt: Art und Umfang der Nutzung unterliegen dem Ermessen des Teilenden und werden durch die datenhaltende Einrichtung ermöglicht.

Mitunter definieren Repositorien Zugriffsklassen, die u. a. den von der DGPs empfohlenen Richtlinien entsprechen, z. B.:

- PsyArchives²²⁴: 4 Klassen (Klassen: 0/0+, 1, **Klassen 2 und 3 sind in Vorbereitung**)
- PsychData²²⁵: 3 Klassen (Klassen: 0,1,2)

Andere Repositorien definieren u. U. gar keine Klassen, regeln aber den Datenzugang z. B. per Vertrag oder Affiliation.

²²⁴ Zugriffsklassen PsychArchives. Zugriff am 23.11.2023, https://www.psycharchives.org/en/about#sharing_levels/.

²²⁵ Zugriffsklassen PsychData. Zugriff am 23.11.2023, <https://rdc-psychology.org/de/datenzugang/>.

Literaturverzeichnis

- Adam, B (2020). Elektronische Laborbücher im Kontext von Forschungsdatenmanagement und guter wissenschaftlicher Praxis - ein Wegweiser für die Lebenswissenschaften: ELN-Wegweiser. Köln. <http://dx.doi.org/10.4126/FRL01-006422868>.
- Biernacka K., Buchholz P., Danker, S.A., Dolzycka, D., Engelhardt, C., Helbig, K., Jacob, J., Neumann, J., Odebrecht, C., Petersen, B., Slowig, B., Trautwein-Bruns, U., Wiljes, C. & Wuttke, U. (2021). Train-the-Trainer-Konzept zum Thema Forschungsdatenmanagement (Version 4). Zenodo. <https://doi.org/10.5281/zenodo.5773203>.
- Blask, K., Jalynskij, M., & Gerhards, L. (2020). Metadata in Psychology 1.0: What researchers really need-Study description of the data referring to the expert interviews conducted in the BMBF-funded project PsyCuraDat. <https://doi.org/10.23668/PSYCHARCHIVES.2750>.
- Blask, K., Gerhards, L., & Jalynskij, M. (2021). PsyCuraDat: Designing a user-oriented curation standard for behavioral psychological research data. *Frontiers in Psychology*, 11, 579397.
- Boeckhout, M., Zielhuis, G. A., & Bredenoord, A. L. (2018). The FAIR guiding principles for data stewardship: fair enough?. *Eur J Hum Gen*, 26(7), 931-936.
- Borghini, J. A. & van Gulick, A. E. (2021). Data management and sharing: Practices and perceptions of psychology researchers. *PLoS ONE* 16(5): e0252047. <https://doi.org/10.1371/journal.pone.0252047>
- Budde, M., Anderson-Schmidt, H., Gade, K., Reich-Erkelenz, D., Adorjan, K., Kalman, J. L., ... & Heilbronner, U. (2019). A longitudinal approach to biological psychiatric research: The PsyCourse study. *Am. J. Med. Genet. Part B: Neuropsychiatric Genetics*, 180(2), 89-102. <https://doi.org/10.1002/ajmg.b.32639>.
- Corti, L., Van den Eynden, V., Bishop, L., & Woollard, M. (2019). *Managing and sharing research data: A guide to good practice*. Sage.
- Deutsche Forschungsgemeinschaft (2013). Sicherung guter wissenschaftlicher Praxis. Denkschrift. Weinheim: Wiley-VCH, S. 21–22. <https://doi.org/10.1002/9783527679188.oth1>.
- Gallagher, R. V., Falster, D. S., Maitner, B., Salguero-Gomez, R., Vandvik, V., Pearse, W., & Enquist, B. J. (2019). The open traits network: Using open science principles to accelerate trait-based science across the tree of life. *Nature Ecology & Evolution*.
- Gollwitzer, M., Abele-Brehm, A., Fiebach, C., Ramthun, R., Scheel, A. M., Schönbrodt, F., & Steinberg, U. (2020). Data management and data sharing in psychological science: Revision of the DGPs recommendations. <https://doi:10.31234/osf.io/24ncs>.
- Henneken, E. A., & Accomazzi, A. (2011). Linking to data-effect on citation rates in astronomy. *arXiv preprint arXiv:1111.3618*.
- Hiemenz, B.M. & Kuberek, M. (2018a). Leitlinie? Grundsätze? Policy? Richtlinie? – Forschungsdaten-Policies an deutschen Universitäten. *o-bib*, 5, S. 1–13. <https://doi.org/10.5282/o-bib/2018H2S1-13>.
- Hiemenz, B., & Kuberek, M. (2018b). Empfehlungen zur Erstellung institutioneller Forschungsdaten-Policies. Das Forschungsdaten-Policy-Kit als generischer Baukasten mit Leitfragen und Textbausteinen für Hochschulen in Deutschland. <https://doi.org/10.14279/depositonce-7521>.
- Higman, R., Bangert, D., & Jones, S. (2019). Three camps, one destination: the intersections of research data management, FAIR and Open. *Insights*, 32(1).

- Kindling, M., & Schirmbacher, P. (2013). „Die digitale Forschungswelt“ als Gegenstand der Forschung/Research on Digital Research/Recherche dans la domaine de la recherche numérique. *Information-Wissenschaft & Praxis*, 64(2-3), 127-136.
- Kovacs, M., Hoekstra, R., & Aczel, B. (2021). The role of human fallibility in psychological research: a survey of mistakes in data management. *Advances in methods and practices in psychological science*, 4(4), 25152459211045930.
- Kreutzer, T., & Lahmann, H. (2021). *Rechtsfragen bei Open Science: Ein Leitfaden* (p. 244). Hamburg University Press.
- Leibniz Universität Hannover & Technische Informationsbibliothek (2018). FAQs zu rechtlichen Aspekten im Umgang mit Forschungsdaten (Version 180215). *Zenodo*. <https://doi.org/10.5281/zenodo.1173546>.
- Ludwig, J., & Enke, H. (2013). *Leitfaden zum Forschungsdaten-Management. Ergebnisse aus dem WissGrid-Projekt*, 15.
- McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., ... & Yarkoni, T. (2016). How open science helps researchers succeed. *elife*, 5, e16800.
- Mercier, H. (2016). Confirmation bias–myside bias. In *Cognitive illusions* (pp. 109-124). Psychology Press.
- Meyermann, A. (2012). Datenmanagement. Eine zentrale Voraussetzung für den Erfolg der Data Sharing-Idee in den Sozialwissenschaften. *DSZ-BO Working Paper*, 2.
- Meyermann, A. & Porzelt, M. (2014). Hinweise zur Anonymisierung qualitativer Daten. Frankfurt am Main. *Leibniz-Institut für Bildungsforschung und Bildungsinformation*, 17 S. - (forschungsdaten bildung informiert; 1). <https://doi.org/10.25656/01:21968>.
- Niso, G., Gorgolewski, K. J., Bock, E., Brooks, T. L., Flandin, G., Gramfort, A., ... & Baillet, S. (2018). MEG-BIDS, the brain imaging data structure extended to magnetoencephalography. *Sci. data*, 5(1), 1-5. <https://doi.org/10.1038/sdata.2018.110>.
- Paskin, N. (2010). Digital object identifier (DOI®) system. *Encyclopedia of library and information sciences*, 3, 1586-1592. <https://doi.org/10.1081/E-ELIS3-120044418>.
- Paßmann, S. & Söring, S. (2023). Forschungsdatenmanagement in der Psychologie: Fachspezifisches Train-the-Trainer-Konzept. *Zenodo*. <https://doi.org/10.5281/zenodo.7560900>.
- Pernet, C. R., Appelhoff, S., Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., & Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Sci Data*, 6(1), 103. <https://doi.org/10.1038/s41597-019-0104-8>.
- Pienta, A. M., Alter, G. C., & Lyle, J. A. (2010). The enduring value of social science research: The use and reuse of primary research data. <http://hdl.handle.net/2027.42/78307>.
- Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PloS one*, 2(3), e308. <https://doi.org/10.1371/journal.pone.0000308>.
- Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175.
- Schimke, N., Kuehler, M., & Hale, J. (2011). Preserving privacy in structural neuroimages. In *Data and Applications Security and Privacy XXV: 25th Annual IFIP WG 11.3 Conference, DBSec 2011, Richmond, VA, USA, July 11-13, 2011. Proceedings 25* (pp. 301-308). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-22348-8_26.
- Schmiederer, S., & Kuberek, M. (2022). *Forschungsdaten-Policies für Forschungsprojekte: ein strukturierter Leitfaden*. Technische Universität Berlin. <https://doi.org/10.14279/depositonce-16196>.

- Steiner, D., Zumbühl, H. J. and Bauder, A. (2008). Two alpine glaciers over the past two centuries: A scientific view based on pictorial sources. In *Orlove, B., Wiegandt, E. and Luckmen, B. H. (Eds.). Darkening Peaks: Glacier Retreat, Science, and Society*, 83-99. Berkeley, University.
- Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International journal of uncertainty, fuzziness and knowledge-based systems*, 10(05), 557-570.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. data*, 3(1), 1-9.
- Wissik, T., & Ďurčo, M. (2016). Research data workflows: from research data lifecycle models to institutional solutions. In *Selected papers from the CLARIN annual conference 2015, October 14–16, 2015, Wrocław, Poland* (No. 123, pp. 94-107). Linköping University Electronic Press.
- Zhang, Z. (2016). Mechanics of human voice production and control. *J. Acoust. Soc. Am.*, 140(4), 2614-2635.

Weiterführende Ressourcen

Einheit 3

- Aust, P., Helbig, K., Schenk, U., Zielke, D., Rosenbaum, A. & Schulze, J. (2016). *Was sind Forschungsdaten?*. Berlin: Humboldt-Universität zu Berlin - Medien-Repository. <https://doi.org/10.18450/dataman/90>.
- Dumontier, M. *Are we FAIR yet? Webinar*. RDA, 2018. Zugriff am 27.04.2023, <https://rd-alliance.org/webinar-are-we-fair-yet>.
- PARTHENOS. "Manage, improve and open up your research and data." Zugriff am 27.04.2023, <https://training.parthenos-project.eu/sample-page/manage-improve-and-open-up-your-research-and-data/>.

Einheit 4

- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... & Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422-1425. <https://doi.org/10.1126/science.aab2374>.
- Fachspezifische Policy der American Psychological Association: Berenson, K. R. (2018). *Managing your research data and documentation*. American Psychological Association.
- Fachspezifische Policy des Schweizer Nationalfond, Zugriff am 27.04.2023, <https://www.scienceurope.org/our-resources/practical-guide-to-the-international-alignment-of-research-data-management>.

Einheit 5

- Brettschneider, P., Biernacka, K., Böker, E., Danker, S. A., Jacob, J., Perry, A., Wiljes, C. & Wuttke, U. (2021). Urheberrecht und Lizenzierung bei Forschungsdaten. *Zenodo*. <https://doi.org/10.5281/zenodo.5243232>.
- Forschungsdaten.info. Rechte und Pflichten 2021. Zugriff am 27.04.2023, <https://www.forschungsdaten.info/themen/rechte-und-pflichten/>.
- Forschungsdaten.info. Rechte und Forschungsdaten - ein Überblick 2021. Zugriff am 27.04.2023, <https://www.forschungsdaten.info/themen/rechte-und-pflichten/recht-und-forschungsdaten-ein-ueberblick/>.
- Guibault, L. & Wiebe, A. (Ed.). *Safe to be open: Study on the protection of research data and recommendations for access and usage*. Göttingen: Universitätsverlag Göttingen, 2013. <https://doi.org/10.17875/gup2013-160>.
- Hartmann, T. (2013). *Zur urheberrechtlichen Schutzfähigkeit von Forschungsdaten*. In „Law as a Service (LaaS) - Recht im Internet- und Cloud-Zeitalter“ herausgegeben von Jürgen Taeger. Band 1, S. 505–515. *Edeweicht: OIWIR, Oldenburger Verlag für Wirtschaft, Informatik und Recht*. <http://hdl.handle.net/11858/00-001M-0000-0014-1208-E>.
- Helbig, K. (2019). Quiz FDM-Rechtsexpertise. *Zenodo*. <https://doi.org/10.5281/zenodo.3479148>.
- Lauber-Rönsberg, A. (2021) 1.4 Rechtliche Aspekte des Forschungsdatenmanagements. In *Praxishandbuch Forschungsdatenmanagement*. Markus Putnings, Heike Neuroth und Janna Neumann (Hrsg.), 89-114. Berlin, Boston: De Gruyter Saur. <https://doi.org/10.1515/9783110657807-005>.
- Lauber-Rönsberg, A., Krahn, P. & Baumann, P. (2018). *Kurzfassung: Gutachten zu den rechtlichen Rahmenbedingungen des Forschungsdatenmanagements im Rahmen des DataJus-Projekts*. Dresden: Technische Universität. Zugriff am 27.04.2023,

https://tu-dresden.de/gsw/phil/irget/jfbimd13/ressourcen/dateien/dateien/DataJus/DataJus_Zusammenfassung_Gutachten_12-07-18.pdf.

- Lauber-Rönsberg, A., Krahn, P. & Baumann, P. (2021). *Forschungsdatenmanagement und Recht. Datenschutz-, Urheber- und Vertragsrecht*. Arbeitshefte der Arbeitsgemeinschaft für juristisches Bibliotheks- und Dokumentationswesen 28. Düns/Feldkirch: W. Neugebauer. ISBN 978-3- 85376-328-5.
- Meyermann, A. & Porzelt, M. (2019). *Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung - eine Handreichung. Version 2.0*. Forschungsdaten Bildung informiert Nr. 6. Zugriff am 27.04.2023, https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert-nr-6_v2.pdf.
- RatSWD - Rat für Sozial- und Wirtschaftsdaten. *Handreichung Datenschutz*. 2. Vollständig überarbeitete Auflage. RatSWD Output 8 (6). Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD), 2020. <https://doi.org/10.17620/02671.50>.
- Scheller, J. (2017). *Rechtliche Rahmenbedingungen der Verwendung von Videos in der Schul- und Unterrichtsforschung. Diskrepanzen zwischen Datenschutzrecht, Förder- und Genehmigungsaufgaben. Version 1.0*. Forschungsdaten Bildung informiert Nr. 5. Zugriff am 27.04.2023, https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert-nr-5.pdf.
- Technische Universität Dresden. „Forschungsprojekt DataJus.“ Zugriff am 27.04.2023, <https://tu-dresden.de/gsw/jura/igetem/jfbimd13/forschung/forschungsprojekt-datajus>.
- Verbund Forschungsdaten Bildung. „Informierte Einwilligung.“ Zugriff am 27.04.2023, <https://www.forschungsdaten-bildung.de/einwilligung>.

Einheit 6

- CESSDA ERIC: „Adapt your Data Management Plan.“ Zugriff am 27.04.2023, https://www.cessda.eu/content/download/4302/48656/file/TTT_DO_DMPExpertGuide_v1.3.pdf.
- Helbig, K. (2015). Hinweise und Checkliste zur Erstellung eines Datenmanagementplans. *Zenodo*, Zugriff am 27.04.2023, <https://doi.org/10.5281/zenodo.33485>
- Helbig, K., Paul-Stüve, T. & Rex, J. (2021). DMP-Toolguide 1.0. *Zenodo*. Zugriff am 27.04.2023, <https://zenodo.org/record/4632308#.YuEKMt9CRaR>
- Helbig, K. & Aust, P. (2015). Datenmanagementpläne für EU, DFG und BMBF. Präsentation. *Zenodo*. Zugriff am 27.04.2023, <https://doi.org/10.5281/zenodo.33482>.
- Neuroth, H., Engelhardt, C., Klar, J., Ludwig, J. & Enke, H. (2018). Aktives Forschungsdatenmanagement. *ABI Technik* 38, S. 55–64. Zugriff am 27.04.2023, <https://doi.org/10.1515/abitech-2018-0008>.

Einheit 7

- CESSDA Training Working Group: „CESSDA Data Management Expert Guide. Bergen: CESSDA ERIC, 2017-2018. Zugriff am 27.04.2023, <https://www.cessda.eu/DMGuide>.
- Corti, L., van den Eynden, V., Bishop, L. & Woollard, M. (2014). *Managing and Sharing Research Data: A Guide to Good Practice*. Los Angeles, CA: SAGE.
- Haenel, V. & Plenz, J. (2014). *Git: verteilte Versionsverwaltung für Code und Dokumente*. 2. Aufl. München: Open Source Press.
- Pilato, C. M., Collins-Sussman, B. & Fitzpatrick, B. W. (2009). *Versionskontrolle mit Subversion*. Köln: O'Reilly.
- Software Carpentry. *Data - Episode 1 - Data Management*. Video. Youtube, 2011. Zugriff am 27.04.2023, <https://www.youtube.com/watch?v=3MEJ38BO6Mo>.

- Software Carpentry. "Data Management: Data Management." Zugriff am 27.04.2023, <https://v4.software-carpentry.org/data/mgmt.html>.
- Technology, Information and Learning Support (TILS). *TILS Document Naming Convention*. Queensland University of Technology, 2009. Zugriff am 27.04.2023, https://www.data.cam.ac.uk/files/gdl_tilsdocnaming_v1_20090612.pdf.

Einheit 8

- CESSDA Training Working Group. „CESSDA Data Management Expert Guide.“ Bergen: CESSDA ERIC, 2017-2018. Zugriff am 27.04.2023, <https://www.cessda.eu/DMGuide>.
- forschungsdaten.info. „Datendokumentation.“ Zugriff am 27.04.2023, <https://www.forschungsdaten.info/themen/beschreiben-und-dokumentieren/datendokumentation/>.
- Social Science Research Council. „*Principles of Documenting Data*.“ In: Managing Qualitative Social Science Data. An interactive online course. Zugriff am 27.04.2023, <https://managing-qualitative-data.org/modules/2/a/>.
- University of Helsinki. „Data Support.“ Zugriff am 27.04.2023, <https://www.helsinki.fi/en/research/services-researchers/data-support>.
- ZB MED (Hrsg.). Elektronische Laborbücher im Kontext von Forschungsdatenmanagement und guter wissenschaftlicher Praxis – ein Wegweiser für die Lebenswissenschaften. Köln, 2019. <https://doi.org/10.4126/FRL01-006415715>.

Einheit 9

- Bundesamt für Sicherheit in der Informationstechnik. „Sicherer Umgang mit Passwörtern Schritt-für-Schritt erklärt.“ Zugriff am 27.04.2023, https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-Verbraucher/Informationen-und-Empfehlungen/Cyber-Sicherheitsempfehlungen/Accountschutz/Sichere-Passwoerter-erstellen/Umgang-mit-Passwoertern/umgang-mit-passwoertern_node.html.
- Leibniz Universität Hannover. „Datenverschlüsselung.“ Zugriff am 27.04.2023, <https://www.luis.uni-hannover.de/de/services/it-sicherheit/praevention/datenverschluesselung/>

Einheit 12

- Hanson, K., Surkis, A. & Yacobucci, K. (2012). Data Sharing and Management Snafu in 3 Short Acts. Video. Youtube. Zugriff am 27.04.2023, https://youtu.be/66oNv_DJuPc.

Einheit 13

- Digital Curation Centre. *Five steps to decide what to keep: a checklist for appraising research data v.1*. Edinburgh: Digital Curation Centre, 2014. Zugriff am 27.04.2023, <https://www.dcc.ac.uk/guidance/how-guides/five-steps-decide-what-data-keep>.
- Ein zukünftiges plattform- und disziplinunabhängiges Datenformat, welches für wide- oder longformat-Daten entwickelt wird, Schnittstellen zu anderen Datenformaten (Excel, Texteditor usw.) und Software (Python, R, Matlab usw.) bereithält, sowie die Vergabe von Metadaten und persistentem Identifier ermöglicht. Zugriff am 27.04.2023, <https://www.konsortswd.de/konsortswd/das-konsortium/services/open-data-format/>

Einheit 15

- Data Archiving and Networked Services. „The PID Forum.“ Zugriff am 27.04.2023, <https://pidforum.org/>.
- Ebel, T. & Meyermann, A. (2015). Hinweise zur Anonymisierung von quantitativen Daten. Version 1.0. Forschungsdaten Bildung informiert Nr. 3. Frankfurt a.M.:

Forschungsdatenzentrum Bildung am DIPF, 2015. Zugriff am 27.04.2023, https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert-nr-3.pdf.

- Forschungsdaten.info. "Persistente Identifikatoren." Zugriff am 27.04.2023, <https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/persistente-identifikatoren/>.
- International DOI Foundation. „DOI.“ Zugriff am 27.04.2023, <https://www.doi.org>.
- Kümmer, S., Lücke, S., Schulz, J., Spenger, M. & Weber, T. (2019). DataCite Best Practice Guide (Version Version 1.0). *Zenodo*. <http://doi.org/10.5281/zenodo.3559800>.
- Henneken, E. A. & Accomazzi, A. (2018). *Linking to data-effect on citation rates in astronomy*. *arXiv*, <https://arxiv.org/abs/1111.3618>.
- Poisot, T. E., Mounce, R. & Gravel, D. (2013). Moving toward a sustainable ecological science: don't let data go to waste!, *Ideas in Ecology and Evolution* 6. <https://doi.org/10.4033/iee.2013.6b.14.f>.
- Project FREYA. Zugriff am 27.04.2023, <https://www.project-freya.eu/en/about/mission>.
- Verbund FDB. „Informierte Einwilligung.“ Zugriff am 27.04.2023, <https://www.forschungsdaten-bildung.de/einwilligung>.
- Vierkant, P., Beucke, D., Deinzer, G., Hartmann, S., Herwig, S., Höhner, K., ... & Summann, F. (2018). *Autorentifizierung anhand der Open Researcher and Contributor ID (ORCID)-Positionspapier*. Humboldt-Universität zu Berlin. <https://doi.org/10.18452/19528>.