

# Forschungsdatenmanagement in kleinen Forschungsprojekten

Eine Handreichung für die Praxis





Rat für Sozial- und Wirtschaftsdaten (RatSWD)

# Forschungsdatenmanagement in kleinen Forschungsprojekten

Eine Handreichung für die Praxis

## Abkürzungsverzeichnis

ALLBUS	.....	Allgemeine Bevölkerungsumfrage der Sozialwissenschaften
CC	.....	Creative Commons
CC0	.....	Public Domain Dedication
CESSDA	.....	Consortium of European Social Science Data Archives
CF	.....	Campus Files
CIS	.....	Centro de Investigaciones Sociológicas
DDI	.....	Data Documentation Initiative
DFG	.....	Deutsche Forschungsgemeinschaft
DMP	.....	Datenmanagementplan
DOI	.....	Digital Object Identifier
DSGVO	.....	Europäische Datenschutzgrundverordnung
ELSST	.....	European Language Social Science Thesaurus
ERIC	.....	European Research Infrastructure Consortium
FAIR	.....	Findable, Accessable, Interoperable, Reusable
FDM	.....	Forschungsdatenmanagement
FDZ	.....	Forschungsdatenzentrum
FDZ-BO	.....	FDZ Betriebs- und Organisationsdaten
FORS	.....	Schweizer Kompetenzzentrum Sozialwissenschaften
GESIS	.....	.GESIS – Leibniz-Institut für Sozialwissenschaften
GND	.....	Gemeinsame Normdatei
ICPSR	.....	Inter-university Consortium for Political and Social Research
ISSDA	.....	Irish Social Science Data Archive
KonsortSWD	.....	Konsortium für Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften
LDSG	.....	Landesdatenschutzgesetze
NEPS	.....	Nationales Bildungspanel
NFDI	.....	Nationale Forschungsdateninfrastruktur
ODC	.....	Open Data Commons
OSF	.....	Open-Science-Bewegung
PUF	.....	Public Use Files
SOCIOSITE	.....	Social Science Information System
SOEP	.....	Sozio-oekonomisches Panel
SUF	.....	Scientific Use Files
TOM	.....	Technische und organisatorische Maßnahmen

## Inhaltsverzeichnis

1	Ziele und Adressat:innen der Handreichung	7
2	Grundlagen	7
3	Forschungsdatenmanagement: Step-by-step	9
3.1	<b>Planungsphase</b>	9
3.1.1	Primär- vs. Sekundärdaten	9
3.1.2	Vielfältigkeit von Daten	10
3.1.3	Fragen der Archivierung und Nachnutzung der generierten Daten	10
3.1.4	Datenmanagementplan	11
3.1.5	FDM und die Kontingenz von Forschungsprozessen	12
3.1.6	Forschungsethische Erwägungen bei der Planung eines Projekts	13
3.2	<b>Datengenerierungsphase</b>	14
3.2.1	Datenschutz einhalten	14
3.2.2	Speicherung, Backup und Sicherung von Forschungsdaten	20
3.2.3	Daten organisieren	23
3.2.4	Daten fortlaufend dokumentieren bzw. kontextualisieren	25
3.3	<b>Aufbereitungs- und Analysephase</b>	28
3.4	<b>Veröffentlichungs-/Archivierungsphase</b>	30
3.4.1	Vorteile und Anforderungen beim Datenteilen	30
3.4.2	Eine geeignete Dateninfrastruktur finden	31
3.4.3	Data Sharing als kooperative Aufgabe	31
3.5	<b>Nachnutzung von Forschungsdaten</b>	33
4	Glossar	36
5	Literaturverzeichnis	38
	Mitwirkende bei der Erstellung	42

## Abbildungsverzeichnis

Abbildung 1: Der Forschungsdaten-Lebenszyklus . . . . .	8
Abbildung 2: Vor- und Nachteile häufig genutzter Speichermöglichkeiten. . . . .	21
Abbildung 3: Benutzungsordnung des Datenarchivs von GESIS . . . . .	34

# 1 Ziele und Adressat:innen der Handreichung

Die Handreichung verfolgt das Ziel, den Forschenden in sozial-, verhaltens- und wirtschaftswissenschaftlichen Projekten den Einstieg in das Forschungsdatenmanagement (FDM) zu erleichtern. Im Fokus stehen dabei einzelne Forschungsprojekte oder kleinere temporäre Projektverbünde, die einen Datenbestand aufbauen und langfristig nutzen und/oder zur Verfügung stellen wollen, ohne bisher auf institutionelle Infrastrukturen zurückzugreifen. Die Broschüre besitzt Überblickscharakter, beschränkt sich auf sozial-, verhaltens- und wirtschaftswissenschaftliche Daten, ist anwendungsorientiert und erhebt keinen Anspruch auf Vollständigkeit.

## 2 Grundlagen

*Forschungsdaten* können je nach Disziplin und Erhebungsmethode sehr unterschiedlich ausfallen. Nach Angabe der Deutschen Forschungsgemeinschaft (DFG) zählen darunter „u. a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten oder Beobachtungsdaten, methodische Testverfahren sowie Fragebögen. Korpora, Software und Simulationen können ebenfalls zentrale Ergebnisse wissenschaftlicher Forschung darstellen und werden daher ebenfalls unter den Begriff Forschungsdaten gefasst“ (DFG, 2021, S. 1).

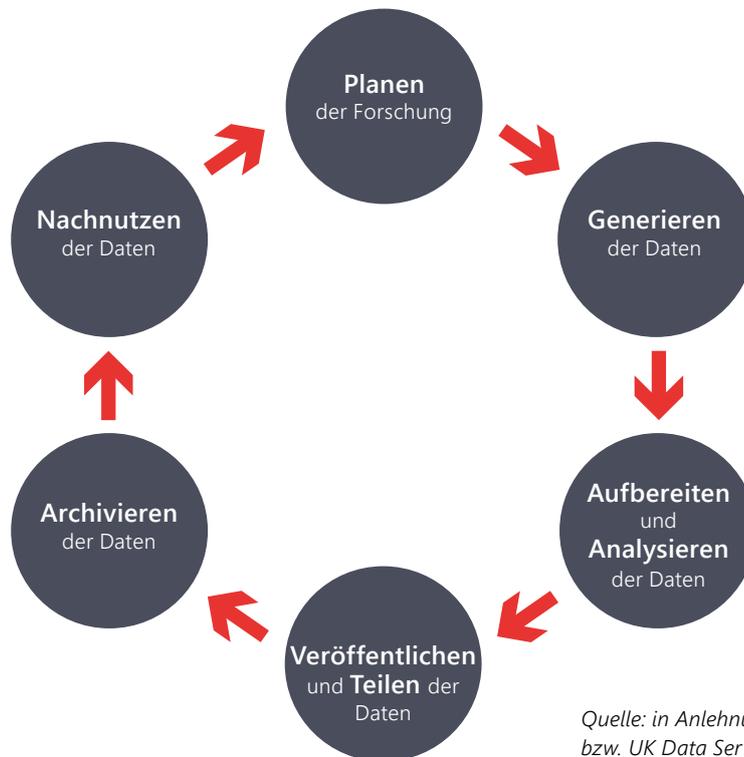
Forschungsdaten sollten nach den „FAIR-Prinzipien“ behandelt werden, d. h. sie sollten nach Möglichkeit auffindbar (*findable*), zugänglich (*accessible*), interoperabel (*interoperable*) und wiederverwendbar (*reusable*) sein (Wilkinson et al., 2016). Bei den vier FAIR-Dimensionen handelt es sich mehr um ein Kontinuum als um ein Entweder-oder (siehe auch Netscher, 2020). Im Gegensatz zu „Open Data“ (alle Daten sind für alle zugänglich) kann die Umsetzung der FAIR-Prinzipien auch bedeuten, dass die Existenz der Daten über Metadaten auffindbar, der Zugang zu den Daten jedoch nur für eine genau zu bestimmende Gruppe an potenziellen Nutzenden unter Auflagen gestattet ist. Ebenso können rechtliche und ethische Gründe gegen eine offene Verfügbarkeit der Daten sprechen. Generell sollten FAIR-Daten „as open as possible and as closed as necessary“ (European Commission, o.J.) sein.

Nicht nur zur Umsetzung der FAIR-Prinzipien ist ein gut dokumentiertes und nachvollziehbares FDM nötig. *FDM* umfasst den gesamten Prozess, den Forschungsdaten in einem Forschungsprojekt durchlaufen – angefangen mit der Planung der Studie, über die Datengewinnung, die Datenaufbereitung und -analyse, die Speicherung der Daten und Metadaten, deren Archivierung und eventuellen Bereitstellung – bis hin zur Wiederverwendung und ggf. Löschung der Daten.

Dieser Prozess kann anhand des *Forschungsdaten-Lebenszyklus* illustriert werden (siehe ► Abbildung 1), der den idealtypischen Prozess darstellt und sich weitestgehend am Management quantitativer Daten orientiert. Abweichungen von diesem idealtypischen Verlauf sind denkbar und unter bestimmten Bedingungen auch sinnvoll. So kann etwa im Rahmen der Planungsphase eruiert werden, ob die passenden Daten zur Beantwortung der Fragestellung bereits erhoben und zugänglich sind, so dass einige Prozessstufen übersprungen werden können. Ebenfalls können beispielsweise Schleifen bei der Gewinnung und Analyse der Daten sinnvoll sein – wie sie etwa die „Grounded Theory“ vorsieht (Strübing, 2018) – oder bei Änderungen in der Erhebung aufgrund von Pretests. Ebenso kann die Veröffentlichung der Daten eingeschränkt werden, wenn eine (faktische) →Anonymisierung<sup>1</sup> der qualitativen Daten deren Interpretierbarkeit so reduzieren würde, dass Nachnutzungen nicht mehr sinnvoll möglich wären. Auch wenn es sich um eine idealtypische Struktur handelt, werden wir uns im Rahmen dieses Beitrags daran orientieren.

<sup>1</sup> Begriffe, die mit einem Pfeil markiert sind, werden im Glossar (Anhang) erläutert.

Abbildung 1: Der Forschungsdaten-Lebenszyklus



Die Durchführung eines aktiven FDM inklusive der Umsetzung der FAIR-Prinzipien ist mit vielfältigen Vorteilen verbunden. Diese Vorteile umfassen sowohl utilitaristische Argumente aus Perspektive der Scientific Community als auch rationale Argumente aus Perspektive der einzelnen Forschenden.

Aus Perspektive der Scientific Community funktioniert wissenschaftlicher Erkenntnisfortschritt generell umso effizienter, je mehr auf bestehendes Wissen und damit auch auf bestehende Daten aufgebaut werden kann. Neben diesem Effizienzgedanken steht jedoch die Qualitätssicherung der Forschungsprozesse und -ergebnisse sowie damit auch der Forschungsdaten im Vordergrund. Die Reproduzier- bzw. Nachvollziehbarkeit von Forschungsergebnissen setzt ein aktives FDM zwingend voraus. Durch das Teilen von Forschungsdaten kann die Anzahl an Datengewinnungen reduziert werden und somit auch das Kollektivgut Teilnahmebereitschaft von Proband:innen geschont werden.

Die einzelnen Forschenden können mittels FDM und Umsetzung der FAIR-Prinzipien die Sichtbarkeit der eigenen Forschung bzw. Person erhöhen. Zudem können so auch ansonsten brachliegende Arbeiten durch die Publikation und potenzielle Zitierbarkeit der Daten sowie der dazugehörigen Dokumentationen zur Steigerung der eigenen Reputation beitragen. Kooperationen können leichter initiiert und die Zusammenarbeit vereinfacht werden (etwa durch verständliche Dokumentation und Nachvollziehbarkeit der erhobenen Daten). Die Gefahr von Datenverlusten oder unerlaubten Zugriffen auf Daten kann zudem minimiert werden. Erfahrungen mit „Open Science“ werden in manchen Disziplinen, wie etwa der Psychologie und den Wirtschaftswissenschaften, auch als Einstellungskriterium berücksichtigt. Nachnutzung verfügbarer Daten kann zum Teil in der universitären Lehre eingesetzt sowie für Abschlussarbeiten genutzt werden. Nicht zuletzt setzt ein Großteil der Drittmittelgeber die Erstellung eines Datenmanagementplans (DMP) bei der Antragstellung voraus und geht davon aus, dass mit öffentlichen Mitteln erhobene Daten im Regelfall auch öffentlich bzw. für die Scientific Community zur Verfügung gestellt werden.

Ein aktives FDM und die Umsetzung der FAIR-Prinzipien umfasst jedoch auch immer die Berücksichtigung ethischer Prinzipien und rechtlicher Regelungen und darf nicht mit einem Teilen von Forschungsdaten um jeden Preis verwechselt werden.

## 3 Forschungsdatenmanagement: Step-by-step

### 3.1 Planungsphase

Vor und mit Projektbeginn sind auf Basis der anvisierten Forschungsfrage eine Reihe konzeptioneller, forschungsstrategischer und organisatorischer Fragen zu klären, die alle auf die eine oder andere Weise Einfluss auf das FDM haben. Davon soll in diesem Abschnitt die Rede sein.



Planen  
der Forschung

#### 3.1.1 Primär- vs. Sekundärdaten<sup>2</sup>

Die erste Frage betrifft die Art empirischer Daten, die zur Bearbeitung der Forschungsfrage erforderlich sind: Sollen neue Daten generiert werden oder soll auf bereits vorhandene Datensätze zurückgegriffen werden? Während es eine gewisse Üblichkeit gibt, die nationalen Surveys oder Panels (etwa SOEP, ALLBUS, NEPS) in Teilauswertungen auf eigene Fragestellungen hin zugeschnitten zu nutzen, ist es vor allem in der →qualitativen Forschung eher üblich, Datenmaterial konkret auf die aktuelle Fragestellung hin orientiert neu zu generieren und dabei von der Spezifik des untersuchten Felds auszugehen.

Die Klärung der Frage, ob es erforderlich ist, eigenes Datenmaterial, also Primärdaten, für ein jeweiliges Projekt zu generieren, erfordert einen Überblick über verfügbares empirisches Material Dritter im angezielten Themenfeld. Informationen darüber lassen sich bei den diversen →Forschungsdatenzentren (FDZ)<sup>3</sup> gewinnen, die eine Vielzahl an Studien teilweise bezogen auf bestimmte Themenfelder bzw. auf bestimmte Datentypen →hosten und →kuratieren. Zurzeit sind im Rahmen der Nationalen Forschungsdateninfrastruktur (NFDI)<sup>4</sup> zentrale Portale im Aufbau, die die Recherche nach passenden Sekundärdatensätzen und den Zugang zu entsprechendem Material erleichtern sollen.

Im Fall der ausschließlichen Nutzung von →Sekundärdaten aus den großen nationalen oder internationalen Forschungsdatenrepositorien erübrigen oder vereinfachen sich eine Reihe von Aufgabenstellungen im Rahmen des projektbezogenen FDM. Die mit der Datengewinnung verbundenen Aufgaben der Etablierung einer →„informierten Einwilligung“ („Informed Consent“) der Teilnehmenden sowie der Aufbereitung und Anonymisierung bzw. →Pseudonymisierung des Datenmaterials sind dann bereits geleistet. Auch die Entscheidung über die Bereitstellung eigener Materialien für Nachnutzungen durch Dritte stellt sich dann nicht mehr in vollem Umfang, da die Daten selbst bereits veröffentlicht sind und lediglich über die Verfügbarmachung von Syntax-Files oder von Codelisten o. ä. der eigenen Auswertung zu entscheiden ist.

Anders stellt sich die Situation bei im eigenen Projekt generiertem Datenmaterial dar: Hier gilt es frühzeitig zu klären, welche Datentypen entstehen (z. B. →Transkripte von Leitfadeninterviews, Beobachtungsprotokolle, Dokumente aus dem Feld, Umfragedaten, →Vignettendaten, Videomaterial, Tondokumente, Social-Media-Feeds, physiologische Messungen, psychometrische Testverfahren), wie das Material gewonnen und weiterverarbeitet werden soll und welche datenschutzrechtlichen bzw. forschungsethischen Erfordernisse sich daraus ergeben.

<sup>2</sup> Die Begriffe Primär- und Sekundärdaten werden in den sozial-, verhaltens- und wirtschaftswissenschaftlichen Fächern teilweise unterschiedlich verwendet. Im Kontext dieses Berichts werden Primärdaten als Daten verstanden, die in ihrem Entstehungskontext genutzt werden und Sekundärdaten als Daten, die außerhalb ihres Entstehungskontextes wiederverwendet werden.

<sup>3</sup> Eine Übersicht der beim RatSWD akkreditierten FDZ ist auf der Seite von KonsortSWD abrufbar: <https://www.konsortswd.de/datenzentren/>

<sup>4</sup> <https://www.nfdi.de/>

### 3.1.2 Vielfältigkeit von Daten

Die Vielfalt an Forschungsdesigns und dabei generierten bzw. genutzten Daten bringt gerade bei der Planung einer Untersuchung die Notwendigkeit einer Vielzahl von Entscheidungen mit sich, die erheblichen Einfluss auf den Fortgang der jeweiligen Studie, aber auch auf das Nachnutzungspotential der generierten Daten haben. Auch wenn nicht alle diese Fragen zu Beginn einer Studie schon endgültig zu beantworten sind, ist es doch erforderlich und hilfreich, sich mit ihnen von Beginn an auseinanderzusetzen. Der Sinn von Plänen liegt ohnehin meist nicht im Plan, sondern in der Planung, also in einer Handlungsdimension der Prozessreflexion. Auf Basis einer reflektierten Planung lassen sich dann auch später im Prozess sichtbar werdende neue Anforderungen, z.B. Variationen im Sampling, der Einbezug weiterer Falldomänen, die Fokussierung auf Teilaspekte der ursprünglich ins Auge gefassten Untersuchung oder die Hinzunahmen weiterer Datentypen, begründen und systematisch berücksichtigen.

Für das FDM ergeben sich also schon bei der Planung der Untersuchung unterschiedliche Fragen und konkrete Aufgaben.

### 3.1.3 Fragen der Archivierung und Nachnutzung der generierten Daten

Bei jeder Studie ist die Frage zu klären, wie die dabei gewonnenen empirischen Daten langfristig gesichert, also archiviert werden können, sowie ob und ggf. wie diese Daten ganz oder teilweise nicht nur zu Überprüfungszwecken im Sinne der Regeln guter wissenschaftlicher Praxis, sondern auch zu weiteren Nachnutzungen durch Dritte (z.B. zur Anschlussforschung oder als Lehrmaterial) verfügbar gemacht werden können und sollen. Forschungsförderorganisationen wie die DFG fordern inzwischen von Antragstellenden die Prüfung dieser Frage und im Fall der Ablehnung einer Nachnutzungen ermöglichenden Archivierung eine Begründung (vgl. Open Science Future, o.J.). Forschende sind also nicht verpflichtet, ihre Daten für Nachnutzungen bereitzustellen. Es müssen aber sachliche Gründe angegeben werden, warum Nachnutzungen entweder möglich und vorgesehen sind oder aber ausgeschlossen bzw. signifikant eingeschränkt werden.<sup>5</sup> Dabei sichern die forschungsfördernden Institutionen zu, dass begründete Ausschlüsse oder Einwilligungen von Datennachnutzungsoptionen für die Frage der Projektbewilligung keine Rolle spielen.



#### Zur Archivierung und Nachnutzung sind u. a. folgende Fragen zu klären

- Ist die informierte Einwilligung so formuliert, dass sie die Möglichkeit der Weitergabe (anonymisierter) Forschungsdaten über das jeweilige Projekt hinaus an andere Instanzen (FDZ sowie konkrete oder noch unbestimmte Nachnutzende) eindeutig mit abdeckt?
- Steht zu befürchten, dass die über die informierte Einwilligung bekannt gegebene Absicht der Datenweitergabe den Rapport zum Feld und den Feldzugang entscheidend erschwert bzw. im Extremfall die Forschung unmöglich macht? (Hier gilt ein Primat der Primärforschung.)
- Welche Maßnahmen sind für eine derartige Weitergabe erforderlich? Also: Muss das Material ganz oder teilweise über die üblichen Projekterfordernisse hinaus anonymisiert oder pseudonymisiert werden? Sind die zeitlichen und personellen Ressourcen für die notwendigen Maßnahmen im Rahmen des Projekts mit eingeplant und erscheint der Aufwand vertretbar?
- Steht der Aufwand für eine Aufbereitung und Weitergabe der Daten in einem vernünftigen Verhältnis zum erwartbaren Mehrwert für Forschung und Lehre?
- Besteht die Aussicht, dass das für eine ethisch und rechtlich vertretbare Weitergabe aufbereitete Material für Anschlussforschungen noch nutzbar ist (Stichwort: →Verpixelung von Videodaten und Fotomaterial)?
- Welches →Repositoryum oder FDZ kann ggf. den Datensatz nach Projektende hosten?
- Welche Art von Unterstützung für die Materialaufbereitung und Datenbereitstellung kann das jeweilige FDZ zu welchen Konditionen leisten?

<sup>5</sup> Dazu auch die Stellungnahme des DFG-Fachkollegiums 111 „Sozialwissenschaften“ (DFG - Fachkollegium 111 „Sozialwissenschaften“, 2020).

### 3.1.4 Datenmanagementplan (DMP)

Der DMP ist das Dokument, in dem all diese Überlegungen systematisch dokumentiert werden. Das Instrument des DMP ist vergleichsweise neu und in vielen Feldern der empirischen Sozial- und Wirtschaftsforschung sowie insbesondere in kleineren Projekten noch nicht sehr etabliert. Inzwischen verlangen allerdings Forschungsförderinstitutionen schon bei der Beantragung von Projekten zunehmend die Einreichung eines DMP (so etwa Horizon Europe). Die DFG fordert in ihren Förderinformationen zwar keinen DMP, wohl aber Informationen zum Umgang mit den Daten: „Grundsätzlich müssen in allen Förderprogrammen, in denen ein konkretes Arbeitsprogramm für ein Forschungsvorhaben beschrieben wird, Ausführungen zum Umgang mit Forschungsdaten beschrieben werden“ (DFG, 2022b).

Auch über diese Vorschriften hinaus ist die Arbeit mit einem DMP ausgesprochen hilfreich für eine kompetente und effiziente Projektbearbeitung. Dabei hat der DMP mehrere Funktionen:



#### Der DMP

- dient der systematischen Reflexion über die Erzeugung, Nutzung und Einbindung empirischer Forschungsdaten in den Forschungsprozess.
- unterstützt damit Transparenz über die Datenerzeugung und -verwendung.
- stiftet Kommunikation zwischen unterschiedlichen institutionellen Akteur:innen im jeweiligen Forschungskontext (insbesondere: forschungsfördernde Institutionen, Projektnehmende, Forschende und FDZ).
- produziert einen Anlass, sich frühzeitig im Projektverlauf auch mit Fragen der dauerhaften und sicheren Archivierung sowie eventuell anzustrebenden Formen der Nachnutzung der im Laufe des Forschungsprozesses generierten bzw. genutzten Daten zu beschäftigen.
- soll als →„Living Document“ die Datennutzungsprozesse flexibel begleiten, Änderungen fortlaufend dokumentieren, nicht aber hinderliche Festlegungen produzieren, die die praktische Forschung beeinträchtigen könnten.

### Was sind die Inhalte eines DMP?

Die genaue Ausgestaltung von DMP ist nicht normiert, sondern variiert auch je nach Förderinstitution.<sup>6</sup> Ein DMP soll jedoch vor allem folgende Informationen enthalten:



1. Eine *Spezifizierung des Datenmaterials*, das im Projekt genutzt, gesammelt und/oder generiert wird. Zu klären ist also:
  - a. Welcher Art sind die zu verwendenden Daten und mit welchen Methoden werden sie gewonnen?
  - b. Ist die Verwendung bereits existierender Daten beabsichtigt und welche Nutzungsbedingungen gelten dafür?
  - c. Auf welche Weise werden bereits existierende Daten zugänglich gemacht?
  - d. Werden unterschiedliche Datenbestände aus verschiedenen Quellen miteinander verknüpft?
  - e. Welche Maßnahmen der Qualitätssicherung werden für die Generierung der Daten eingesetzt?

<sup>6</sup> Eine Übersicht zu den unterschiedlichen Anforderungen und Handreichungen der forschungsfördernden Institutionen liefert das Open-Science-Magazin der ZBW – Leibniz-Informationszentrum Wirtschaft (Open Science Future, o.J.).

2. *Informationen über den Umgang mit Forschungsdaten während des Projekts:*
  - a. Wie sollen die Daten nachvollziehbar beschrieben werden (Metadaten, Dokumentationsstandards)?
  - b. Gibt es in Bezug auf die Daten besondere rechtliche Verpflichtungen und weitere zu beachtende Rahmenbedingungen, die die Zugänglichkeit zu den Daten einschränken oder Einfluss auf die späteren Möglichkeiten der Veröffentlichung haben können?
  - c. Wie werden die Daten, insbesondere sensible Daten, während des Projektverlaufs gespeichert und gesichert? Wie ist der Zugriff organisiert?
  - d. Wer ist während des Projekts und wer nach Laufzeitende für den adäquaten Umgang mit den Daten verantwortlich?
  - e. Welche Ressourcen sind für Aufarbeitung und Archivierung erforderlich?
3. *Informationen über den Umgang mit Forschungsdaten nach Projektende:*
  - a. Sind für das generierte Datenmaterial Nachnutzungsmöglichkeiten geplant?
  - b. Wie soll die ggf. geplante Verfügbarmachung der Daten realisiert werden?
    - i. In einer minimalen Variante im Sinne der Ablage des minimal dokumentierten und kontextualisierten Datensatzes in einem (z. B. lokalen) Repository oder einer anderen Einrichtung der Forschungsdateninfrastruktur.
    - ii. Als systematisch erschlossener und archivierter Datensatz inklusive einer für die praktische Weiterverwendung, etwa in Replikationen, nutzbaren, umfassenden Dokumentation; zugänglich entweder bei den Datenproduzierenden vor Ort oder über ein FDZ.
    - iii. Als gezielt für potenzielle Sekundäranalysen aufbereiteter und mit Metadaten erschlossener Datenbestand, der einem FDZ übergeben wird.

Quelle: vgl. RatSWD (2016) sowie Science Europe (2021)

Die Aufbereitung von Daten für die Archivierung und insbesondere für eventuelle Nachnutzungen bedeutet einen je nach Art der Daten und der Umstände mitunter erheblichen Bearbeitungsaufwand, der bei der Beantragung von Ressourcen bereits mit bedacht werden muss. Zumindest die DFG erlaubt es, die anfallenden zusätzlichen Personal- und Sachkosten gesondert zu beantragen. Ein Problem stellt es aber gerade in kleineren und im FDM nicht sehr versierten Projekten dar, den erforderlichen Aufwand angemessen zu beziffern. Hier kann es hilfreich sein, sich bereits in der Planungsphase mit einem FDZ, dem die Daten später ggf. auch übergeben werden sollen, in Verbindung zu setzen und sich bei der Kalkulation des Aufwands unterstützen zu lassen. Denn dort sitzen Expert:innen für das Kuratieren und Bereitstellen von Forschungsdaten, die die erforderlichen Tätigkeiten über den gesamten Lebenszyklus der archivierten Daten überblicken und auch vergleichend mit anderen Projekten den Aufwand abschätzen können.

### 3.1.5 FDM und die Kontingenz von Forschungsprozessen

Insbesondere in einer Reihe von Designs der qualitativen Sozialforschung (z. B. Grounded Theory, dokumentarische Analyse, ethnographische Feldforschung oder wissenssoziologische Diskursanalyse) erfolgt die genauere Bestimmung und Arrondierung des Datenkorpus häufig erst sukzessive im Projektverlauf und abhängig von ersten Ergebnissen (z. B. im Wege des →„Theoretical Sampling“). Dadurch ist es erforderlich, a) den bei Projektbeantragung vorgelegten DMP als vorläufig zu kennzeichnen und b) diesen dann projektbegleitend fortzuschreiben – auch das entspricht der Idee eines DMP als Living Document.



### Nützliche Links und Hilfestellungen für die **Erstellung eines Datenmanagementplans**

- Adapt your Data Management Plan. A list of Data Management Questions based on the Expert Tour Guide on Data Management (CESSDA, 2019).
- Der Datenmanagementplan. Eine Wegbeschreibung für Daten (forschungsdaten.info, 2022b).
- Management und Bereitstellung von Forschungsdaten in der Psychologie: Überarbeitung der DGPs-Empfehlungen (Gollwitzer et al., 2020).
- Elements of a Data Management Plan (ICPSR, o.J.-a).
- Practical Guide to the International Alignment of Research Data Management – Extended Edition (Science Europe, 2021).
- Checkliste zur Erstellung eines Datenmanagementplans in der empirischen Bildungsforschung (VerbundFDB, 2015).



### Nützliche Tools zur **Erstellung eines Datenmanagementplans**

Für Projektanträge in *Deutschland*:

- Standardisierter Datenmanagementplan für die Bildungsforschung (Stamp) (VerbundFDB, 2022b).
- Research Data Management Organiser (RDMO Arbeitsgemeinschaft, o.J.).

Für Projektanträge für *Horizon Europe*:

- Argos – Plan and follow your data (Argos, o.J.).
- DMPonline – Plan to make data work for you (DMPonline, o.J.).
- Data Stewardship Wizard (Pergl et al., 2019).

Für Kooperationen mit Forschenden aus den *USA*:

- DMPTool – Build your Data Management Plan (DMPTool, o.J.).

## 3.1.6 **Forschungsethische Erwägungen bei der Planung eines Projekts**

Bereits zu Beginn eines Forschungsprojekts ist zu klären, ob für die Beantragung, Durchführung oder Ergebnispublikation ein →Ethikvotum erforderlich ist.



*Erläuterung:* Universitäten, forschungsfördernde Institutionen, aber zunehmend auch Fachzeitschriften erwarten bei Forschung in sensiblen Themenfeldern und/oder mit vulnerablen Gruppen ein entlastendes Votum einer Ethikkommission. In diesem Votum soll kompetent abgeschätzt werden, welche möglicherweise unerwünschten Nebenfolgen mit der geplanten Studie für Beteiligte oder die Gruppen/Milieus, die sie repräsentieren, einhergehen könnten, ob diese Risiken im Lichte des erwartbaren Nutzens für die Forschung akzeptabel erscheinen und welche Maßnahmen vorgesehen sind, um negative Auswirkungen insbesondere für vulnerable Gruppen zu vermeiden bzw. abzumildern. Da die Infrastruktur an fachkompetenten Ethikkommissionen bislang in den Sozialwissenschaften noch eher rudimentär ist, ist es für Projekte wichtig zu klären, ob ein Votum erforderlich ist und welche Ethikkommission dafür in Betracht kommt.

Während die Bedeutung forschungsethischer Erwägungen und deren konsequente Umsetzung im Forschungshandeln unstrittig sind, wird die Frage der Sinnhaftigkeit von Ethikvoten und des flächendeckenden Aufbaus von Ethikkommissionen auch kontrovers diskutiert.<sup>7</sup> Nichtsdestotrotz macht es der institutionelle Druck von Forschungsfördereinrichtungen und Verlagen immer häufiger erforderlich, entsprechend Voten von Ethikkommissionen einzuholen.

Nicht generell zu beantworten ist die Frage, an welche Ethikkommission Forschende sich wenden können, wenn sie entweder eigene Zweifel an der ethischen Angemessenheit ihres Vorhabens entwickeln oder besagte institutionelle Erwartungen bestehen. Denn die Infrastruktur an Ethikkommissionen ist je nach Fach, Forschungseinrichtung und Förderzusammenhang (z. B. bei interdisziplinären Projekten) sehr unterschiedlich ausgebaut. Für eine fachlich angemessene Beurteilung möglicher ethischer Probleme müsste die jeweilige Kommission über methodische und inhaltliche Fachkompetenz verfügen und zugleich über hinreichend Wissen zu den ethischen und rechtlichen Rahmenbedingungen. Das ist nicht immer gegeben, vor allem wenn es um Entscheidungen zu Projekten mit hochspezialisierten Forschungsmethoden geht. Es ist ratsam, zunächst nach einer fachlich halbwegs einschlägigen Ethikkommission an der eigenen Forschungseinrichtung zu suchen und, falls eine solche nicht existiert, die jeweilige Fachgesellschaft zu kontaktieren. Weniger ratsam ist es, auf fachlich fern liegende Kommissionen, etwa die an jeder medizinischen Fakultät existierenden Gremien, auszuweichen, weil hier die Gefahr von sachunangemessenen, weil methodisch nicht hinreichend sensibilisierten Urteilen besteht.



#### Nützliche Links und Veröffentlichungen zum Thema **Forschungsethik**

- FAQ zu Ethikvoten: Informationen aus den Geistes- und Sozialwissenschaften (DFG, 2022a).
- Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften (RatSWD, 2017).
- Best-Practice-Webseite zu Forschungsethik in den Sozial- und Wirtschaftswissenschaften (RatSWD, o.J.–a).



## 3.2 Datengenerierungsphase

Bevor die Daten generiert werden, ist in Forschungsprojekten in der Regel eine Vorbereitungsphase eingeplant. Diese sollte von Forschenden genutzt werden, um ein effizientes und nachhaltiges Datenmanagement zu implementieren. Wichtige Bereiche, in denen vorab organisatorische Maßnahmen und forschungsstrategische Entscheidungen zu treffen sind, sollen in diesem Abschnitt etwas genauer dargestellt werden.

### 3.2.1 Datenschutz einhalten

Da in den Sozial-, Verhaltens- und Wirtschaftswissenschaften meist Menschen den Untersuchungsgegenstand bilden, werden in vielen Projekten personenbezogene Daten verarbeitet. Der Umgang mit personenbezogenen Daten fällt unter die Europäische Datenschutzgrundverordnung (DSGVO), die durch ergänzende nationale Regelungen des Bundesdatenschutzgesetzes (BDSG) und der Landesdatenschutzgesetze (LDSG) sowie durch Spezialgesetze (z. B. Schul- oder Statistikgesetze) weiter ausgestaltet wird. Diese rechtlichen Grundlagen sind von Forschenden zu berücksichtigen, die mit personenbezogenen Daten arbeiten (Art. 2 Abs. 1 DSGVO).

#### Was sind personenbezogene Daten?

Personenbezogene Daten sind „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen“ (Art. 4 Nr. 1 DSGVO). Damit eine Person als *identifizierbar* gilt, ist entscheidend, ob zwischen den verarbeiteten Daten und der Person eine Verbindung hergestellt werden kann. Die Zuordnung zu Kennnummern wie einer Sozialversicherungsnummer, zu Standortdaten wie einer Adresse

<sup>7</sup> Zur Bedeutung forschungsethischer Reflexion z. B. Friedrichs (2014) und Hopf (2016) sowie kritisch z. B. Hammersley (2009).

oder einer Online-Kennung wie einer IP-Adresse sind beispielsweise ausreichend, um von personenbezogenen Daten zu sprechen. Ein anderes Beispiel wäre, wenn eine Person durch die Kombination mehrerer besonderer Merkmale bestimmbar wird, die Ausdruck ihrer „physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität“ sind (Art. 4 Nr. 1 DSGVO).



Grundsätzlich gilt, wann immer Dritte mit einem gewissen (verhältnismäßigen) Aufwand herausfinden könnten, welcher Person die Daten zuzuordnen sind, fallen diese unter das Datenschutzgesetz. Das Datenschutzrecht gilt hingegen nicht bei anonymen bzw. vollständig anonymisierten Daten, bei denen sich auch mit hohem Aufwand kein Personenbezug (mehr) herstellen lässt.

### Was sind besondere Kategorien personenbezogener Daten?

Es gibt zusätzlich besondere Kategorien von personenbezogenen Daten, die auch als sensible Daten bezeichnet werden. Da sie als besonders schutzwürdig eingestuft werden, erfordert der Umgang mit diesen Daten höhere technische, organisatorische und rechtliche Anforderungen.



Zu diesen besonderen Kategorien zählen Daten, aus denen Informationen zu folgenden Themen hervorgehen (Art. 9 Abs. 1 und Art. 10 DSGVO):

- Ethnische Herkunft
- Politische Meinungen
- Religiöse und weltanschauliche Überzeugungen
- Gewerkschaftszugehörigkeiten
- Genetische und biometrische Daten
- Gesundheitsdaten
- Daten zum Sexualleben oder der sexuellen Orientierung
- Daten über Straftaten, strafrechtliche Verurteilungen und Sicherungsmaßnahmen

### Welche Maßnahmen sind bei der Arbeit mit personenbezogenen Daten notwendig?

Je nach individuellem Forschungsgegenstand und Forschungsdesign können unterschiedlich weitreichende Maßnahmen notwendig werden. An dieser Stelle werden *einige grundlegende Maßnahmen* dargestellt, die in der Regel von allen Forschenden in der Arbeit mit personenbezogenen Daten ergriffen werden sollten. Diese Ausführungen sind jedoch ausdrücklich nicht als rechtlich bindend zu verstehen, sondern ausschließlich als Empfehlungen zum Umgang mit personenbezogenen Forschungsdaten.

#### (1) Verantwortliche Person/en festlegen

Spätestens zum Start eines Forschungsprojekts sollte festgelegt werden, welche Person die Verantwortung für die Verarbeitung personenbezogener Daten trägt. Nach der DSGVO ist verantwortlich, wer „über die Zwecke und Mittel der Datenverarbeitung entscheidet“ (Art. 4 Abs. 7 DSGVO). In der Regel ist dies in einem Einzelprojekt die Projektleitung. In einem Kooperationsprojekt besteht fast immer eine gemeinsame Verantwortung der Projektleitung. In diesem Fall ist eine „Vereinbarung zur gemeinsamen Verantwortlichkeit“ notwendig, in der transparent die Aufgabenteilung bei der Erfüllung der Datenschutzpflichten dargelegt wird (Art. 26 Abs. 1 DSGVO). Die zentralen Inhalte der Vereinbarung werden den Studienteilnehmenden zur Verfügung gestellt (Art. 26 Abs. 2 DSGVO). In der Vertretung nach außen gegenüber den Datenschutzbehörden trägt allerdings fast immer die wissenschaftliche Einrichtung die Verantwortung. Die Projektleitung ist somit als eine verantwortliche (Unter-)Stelle innerhalb der wissenschaftlichen Einrichtungen zu betrachten, die die vorgeschriebenen datenschutzrechtlichen Maßnahmen umzusetzen hat (Meyermann & Porzelt, 2019, S. 32). Deshalb sollten alle datenschutzrelevanten Dokumente immer mit der oder dem Datenschutzbeauftragten der Institution abgestimmt und zur Kenntnis weitergeleitet werden.

### (2) Verzeichnis für Verarbeitungstätigkeiten anlegen

Immer, wenn in Forschungsprojekten personenbezogene Daten verarbeitet werden, ist zu Dokumentationszwecken zum Projektstart zusätzlich ein „Verzeichnis für Verarbeitungstätigkeiten“ anzulegen (Art. 30 Abs. 1 DSGVO). Dieses informiert darüber, welche Art von personenbezogenen Daten von welchen Personen(gruppen) zu welchem Zweck verarbeitet werden. Es wird darin auch offengelegt, wie lange die Speicherdauer ist, wer Zugang zu den Daten erhält und welche Maßnahmen zum Schutz der Daten ergriffen werden. Durch diesen schriftlichen Nachweis wird die Rechenschaftspflicht der DSGVO zur Einhaltung der Schutzvorgaben (Art. 5 Abs. 2 DSGVO) erfüllt. Falls eine staatliche Aufsichtsbehörde in begründeten Fällen nachfragt, können Forschende dieses Verzeichnis vorlegen.

### (3) Einwilligungen einholen

Die Verarbeitung von personenbezogenen Daten ist nur dann rechtmäßig, wenn eine *Rechtsgrundlage* vorliegt (Art. 6 DSGVO). Die am häufigsten herangezogene Rechtsgrundlage bei wissenschaftlichen Projekten ist die informierte Einwilligung (Art. 6 Abs. 1 lit. a DSGVO). Dies bedeutet, dass Studienteilnehmende ausreichend über die Datenverarbeitung informiert werden und auf dieser Basis ihr Einverständnis zur Gewinnung, Speicherung und Auswertung sowie, falls geplant, zur (langfristigen) Archivierung und Bereitstellung für (wissenschaftliche) Nachnutzungen ihrer Daten geben. Die Einwilligung muss *freiwillig* für einen *bestimmten Fall* und in *unmissverständlicher Form* erfolgen (Art. 4 Nr. 11 DSGVO). Sie kann anschließend jederzeit ohne Angabe von Gründen widerrufen werden (Art. 7 Abs. 3 DSGVO). In diesem Fall sind die Daten unverzüglich zu löschen. Die Datenverarbeitung bis zum Zeitpunkt des Widerrufs bleibt davon unberührt.

Eine Willensentscheidung setzt eine *Einwilligungsfähigkeit* voraus (indirekt aus Art. 8 DSGVO abzuleiten). Dies bedeutet, die Person muss die dargebotenen Informationen verstehen und die Konsequenzen der Einwilligung abschätzen können. Bei volljährigen Personen mit Geschäftsfähigkeit wird diese Fähigkeit unterstellt, so dass ausschließlich in berechtigten Fällen eine Einzelprüfung stattzufinden hat. Bei Kindern und Jugendlichen unter 16 Jahren ist jedoch in der Regel die Einwilligung der Erziehungsberechtigten oder des gesetzlichen Vormunds *zusätzlich* oder *anstelle* derjenigen des Kindes einzuholen (Art. 8 Abs. 2 DSGVO; Meyermann & Porzelt, 2019, S. 15).

Damit die Einwilligung wirksam wird, muss sie zusätzlich in *Form und Inhalt* bestimmten Anforderungen entsprechen. Obwohl eine Einwilligung auch mündlich oder in anderer Form erteilt werden kann, wird empfohlen, die *Schriftform* zu wählen, solange dadurch nicht der Forschungszweck (erheblich) beeinträchtigt wird. Die DSGVO legt die Beweislast nämlich in die Hände der Forschenden. Im Zweifelsfall besteht für die Forschenden die Pflicht, einen eindeutigen Nachweis zu erbringen, dass die Studienteilnehmenden informiert wurden und in die Datenverarbeitung eingewilligt haben (Art. 7 Abs. 1 DSGVO). Eine Unterschrift unter eine Einwilligungserklärung ist hierfür die sicherste Methode. Falls der Forschungszweck die Schriftform nicht zulassen sollte, weil dadurch der Interaktionsprozess im Feld negativ beeinflusst werden könnte, ist über andere Wege nachzudenken, wie die Einwilligung in Form einer eindeutig bestätigenden Handlung dokumentiert werden kann (z. B. über detaillierte Feldprotokolle, Audio- oder Videoaufzeichnungen; RatSWD, 2017, S. 23).



*Formale Bedingungen*, die eine (schriftliche) Einwilligung und ein dazugehöriger Informativonsteil erfüllen sollten:

- Eine klare, einfache und zielgruppengerechte Sprache (Art. 7 Abs. 2 DSGVO)
- Eine verständliche und leicht zugängliche Darstellungsform (Art. 7 Abs. 2 DSGVO)
- Vollständige Begleitinformationen, die klar von der Einwilligungserklärung getrennt sind oder graphisch abgesetzt werden (Art. 7 Abs. 2 DSGVO)



*Inhalte*, über die Studienteilnehmende in der *Einwilligungserklärung* nach Art. 13 DSGVO mindestens informiert werden sollten:

- Name und Kontaktdaten der projektverantwortlichen Person, ggf. der vertretenden Person und der bzw. des Datenschutzbeauftragten
- Zweck(e) der Datenverarbeitung
- Rechtsgrundlage für die Datenverarbeitung (Einwilligung nach Art. 6 Abs. 1 lit. a DSGVO, bei sensiblen Daten ist zusätzlich Art. 9 Abs. 2 lit. a DSGVO relevant)
- Informationen über (oder Kriterien für) die Speicherdauer (Art. 13 Abs. 2 lit. a DSGVO)
- Informationen zu den wichtigsten Rechten der Studienteilnehmenden in Bezug auf ihre Daten: Auskunftsrecht (Art. 15 DSGVO), Recht auf Berichtigung (Art. 16 DSGVO), Recht auf Löschung (Art. 17 DSGVO), Recht auf Einschränkung der Verarbeitung (Art. 18 DSGVO), Widerspruchsrecht (Art. 21 DSGVO)
- Freiwilligkeit der Studienteilnahme (Art. 4 Nr. 11 DSGVO), soweit es keine vertraglichen oder gesetzlichen Vorschriften verlangt (Art. 13 Abs. 2 lit. e DSGVO)
- Recht, die Einwilligung zu verweigern und auch jederzeit ohne Angabe von Gründen zu widerrufen (Art. 7 Abs. 3 DSGVO), ohne dass daraus ein Nachteil entsteht; die Datenverarbeitung bis zum Zeitpunkt des Widerrufs bleibt rechtmäßig
- Hinweis auf ein Beschwerderecht bei einer Aufsichtsbehörde (Art. 77 DSGVO)
- Ggf. die geplante Datenweitergabe in Länder außerhalb der EU
- Ggf. die Empfänger:innen oder Kategorien von Empfänger:innen der personenbezogenen Daten



*Weitere Inhalte*, die in der *Einwilligungserklärung* aufzuführen sind, falls sie für das Projekt zutreffen:

- Falls eine Archivierung der Daten außerhalb des unmittelbaren Projektkontexts oder die Bereitstellung für die (wissenschaftliche) Nachnutzung geplant ist, ist darüber zu informieren.
- Falls die personenbezogenen Daten nicht direkt unter Mitwirkung der jeweiligen Person erhoben werden, sondern aus anderen Quellen stammen, muss nach Art. 14 DSGVO in der Einwilligungserklärung zusätzlich über die Datenherkunft und die Kategorien personenbezogener Daten, die verarbeitet werden, informiert werden.
- Falls mit besonderen Kategorien personenbezogener Daten nach Art. 9 Abs. 1 DSGVO gearbeitet wird, sind diese ebenfalls ausdrücklich in der Einwilligungserklärung aufzuführen (nach Art. 9 Abs. 2 lit. a DSGVO).

Nicht alle Inhalte müssen immer genau beschrieben werden. Falls beispielsweise durch eine wahrheitsgetreue Angabe zum Studienzweck die Datengewinnung negativ beeinflusst werden würde (z. B. durch einen Antwortbias der Teilnehmenden bei einer Befragung), sind auch allgemeinere Formulierungen zulässig, die dem Erkenntnisinteresse der Studie nicht entgegenstehen. Je nach Disziplin haben sich außerdem teilweise unterschiedliche ethische Standards entwickelt, inwieweit Täuschungen bei der Nennung des Studienzwecks zugunsten eines bedeutsamen Erkenntnisgewinns zulässig sind (RatSWD, 2017, S. 21–22). Teilnehmende müssen bei Täuschungen im Anschluss an die Studie aufgeklärt werden und ihnen muss gestattet sein, die erlangten Daten wieder zurückzuziehen.

Wenn sich der Zweck der Datenverarbeitung im Laufe des Projekts ändert oder erweitert und die Studienteilnehmenden nicht bereits darüber informiert wurden und ihre Zustimmung entsprechend gegeben haben, sind erneut informierte Einwilligungen einzuholen (Art. 13 Abs. 3 DSGVO bzw. Art. 14 Abs. 4 DSGVO).

Die Einwilligungserklärungen können eigenständig formuliert werden. Es ist jedoch auch möglich, bestehende Vorlagen zu nutzen und für das eigene Projekt anzupassen. Nachfolgend sind einige Beispiele aufgeführt.



#### Beispiele für Vorlagen und Hilfestellungen zur Erstellung von informierten Einwilligungen

- Consent Form Wizard (CFW) - Ein Werkzeug zur Erstellung von Einwilligungserklärungen (DARIAH-EU ELDAH, 2021).
- Ethisches Handeln in der psychologischen Forschung. Empfehlungen der Deutschen Gesellschaft für Psychologie für Forschende und Ethikkommissionen (DGPs, 2018).
- Webseite mit Musterformularen für Einwilligungserklärungen (FDZ Qualiservice, o.J.).
- Formulierungsbeispiele für „informierte Einwilligungen“ (VerbundFDB, 2018a).
- Formulierungsbeispiele für „informierte Einwilligungen“ in leichter Sprache (VerbundFDB, 2018b).
- Checkliste zur Erstellung rechtskonformer Einwilligungserklärungen mit besonderer Berücksichtigung von Erhebungen an Schulen (VerbundFDB, 2019).

#### (4) Technische und organisatorische Maßnahmen

Personenbezogene Daten müssen angemessen durch *technische und organisatorische Maßnahmen (TOM)* gesichert werden. Welche Maßnahmen als angemessen angesehen werden, bemisst sich am Stand der „Technik, der Implementierungskosten und der Art, des Umfangs, der Umstände und der Zwecke der Verarbeitung sowie der unterschiedlichen Eintrittswahrscheinlichkeit und Schwere des Risikos“ (Art. 32 Abs. 1 DSGVO und Art. 24 Abs. 1 DSGVO). Neben der nachfolgend erläuterten Pseudonymisierung und Anonymisierung können beispielsweise folgende TOM eingesetzt werden:

- Verschlüsselung von Dateien und Speichersystemen
- Verwendung sicherer Passwörter oder Passphrasen
- Festlegung von Speicherfristen und Vernichtung nicht mehr benötigter Daten
- Personelle Zugriffs-, Zugangsbeschränkungen und -kontrollen
- Speicherung personenbezogener Daten innerhalb der EU
- Geheimhaltungsvereinbarungen und Vereinbarungen zur Einhaltung des Datenschutzes

In Abschnitt 3.2.2 unter Datensicherheit werden noch weitere mögliche TOM aufgeführt. Die Maßnahmen sollten regelmäßig dahingehend überprüft, bewertet und evaluiert werden, ob sie eine sichere Datenverarbeitung gewährleisten (Art. 32 Abs. 1 lit. d DSGVO).

#### (5) Forschungsdaten pseudonymisieren und anonymisieren

Nach dem Grundsatz der Datenminimierung (Art. 5 Abs. 1 lit. c DSGVO) sollten zu jedem Zeitpunkt eines Forschungsprojekts nur so viele personenbezogene Daten verarbeitet werden wie nötig sind, um den Forschungszweck (noch) realisieren zu können. Zwei wichtige Mittel, um eine möglichst hohe personenbezogene Datensparsamkeit zu erreichen, sind die Pseudonymisierung und die Anonymisierung. Dementsprechend verlangt auch das Datenschutzrecht, dass Forschungsdaten anonymisiert bzw. pseudonymisiert werden, wenn und sobald es der Forschungszweck ermöglicht (Art. 89 Abs. 1 S. 3 und S. 4 DSGVO, § 27 Abs. 3. S. 1 und S. 2 BDSG).

Bei einer *Pseudonymisierung* (nach Art. 4 Nr. 5 DSGVO) werden personenbezogene Merkmale eins zu eins ersetzt, z. B. durch einen Alias, einen Code oder durch die Angabe vergleichbarer Eigenschaften (das sogenannte Pseudonym). Es existiert jedoch eine Schlüsseliste, Zuordnungstabelle oder ein Protokoll, mit der sich die Pseudonyme wieder den Originaldaten zuordnen lassen. Auch wenn das Zuordnungsdokument gut gesichert und getrennt von den Daten (z. B. bei einem vertrauenswürdigen → Datentreuhänder) aufbewahrt wird, könnten die Daten dadurch theoretisch re-identifiziert werden. Deshalb sind pseudonyme Daten nach wie vor personenbezogene Daten, die unter das Datenschutzrecht fallen.

Allerdings ist der Personenbezug in diesen Dateien deutlich reduziert, was weniger restriktive Zugangs- und Zugriffskontrollen im Arbeitsprozess erlaubt. Eine Pseudonymisierung kann in der Regel spätestens nach Abschluss der Datengewinnung durchgeführt werden. In einigen Fällen kann bereits in der Feldphase mit Pseudonymen gearbeitet werden (z. B. Personenidentifikatoren bei Panelstudien).

Die *Anonymisierung* wird in der DSGVO nicht definiert. Aus Erwägungsgrund 26 DSGVO lässt sich jedoch ableiten, dass eine *faktische Anonymisierung* zugrunde gelegt wird. Daten gelten demnach als anonymisiert, wenn nur noch mit unverhältnismäßig hohem Aufwand wieder ein Personenbezug hergestellt werden könnte (RatSWD, 2020, S. 18). Nach diesem Konzept ist eine Re-Identifizierung von Personen nicht völlig unmöglich, aber faktisch so aufwändig, dass sie als äußerst unwahrscheinlich eingeschätzt wird. Deshalb fallen die Daten nicht mehr unter die Datenschutzgesetze. Um den Personenbezug aus den Daten zu entfernen, werden vor allem Techniken genutzt, die den Informationsgehalt reduzieren oder verändern.



Wichtige Anonymisierungstechniken sind:

- Die Löschung der Identifikationsmerkmale (inklusive der Schlüsselliste, Zuordnungstabelle oder Anonymisierungsprotokolle)
- Die Aggregation von Informationen (u. a. Gruppierung von Werten eines Merkmals in allgemeinere Kategorien; Verringerung von detailreichen personenbezogenen Informationen durch Ersetzung mit Angaben auf einer höheren inhaltlichen Abstraktionsebene; Rundung; generalisierende Werte)
- Die Maskierung (d. h. Originalwerte werden durch zufällige, aber technisch valide, neue Werte ersetzt; wichtig ist, dass die neuen Werte eine „vergleichbare Bedeutung“ haben, damit z. B. der Sinngehalt eines Texts in der qualitativen Forschung oder der Informationsgehalt einer Verteilung in der quantitativen Forschung erhalten bleibt.)

In der Regel wird meist eine Kombination aus Verfahren genutzt, um die Daten zu anonymisieren. Während bei quantitativen Daten zum Teil generalisierte, automatisierte Regelungen auf den gesamten Datensatz angewendet werden können, ist bei qualitativen Daten der manuelle Aufwand deutlich größer. Hier müssen in der Regel alle Daten durchgesehen, auf Risiken einer möglichen Re-Identifizierbarkeit eingeschätzt und anonymisiert werden. Bei dichten, detailreichen qualitativen Daten kann dies eine große Herausforderung darstellen. Derzeit gibt es wichtige erste Schritte zur Entwicklung stärker (semi-) automatisierter Verfahren zur Anonymisierung von qualitativen Daten (z. B. das Anonymisierungstool QualiAnon, vgl. Nicolai et al., 2021). Außerdem stellt sich bei qualitativen Daten auch die Frage, ob durch den Anonymisierungsprozess nicht ein zu großer Informationsverlust entsteht, der die anschließenden Analysemöglichkeiten durch Dritte zu stark reduziert. Diese Frage muss jede und jeder Forschende für die eigene Studie individuell beantworten.



#### Nützliche Hinweise und Empfehlungen zur **Anonymisierung von Forschungsdaten**

- Empfehlungen zur Anonymisierung quantitativer Daten (Ebel, 2015).
- Hinweise zur Anonymisierung von quantitativen Daten (Ebel & Meyermann, 2015).
- Hinweise zur Anonymisierung qualitativer Daten (Meyermann & Porzelt, 2019).



#### Nützliches Tool für **Pseudonymisierung bzw. Anonymisierung von Forschungsdaten**

- QualiAnon - Qualiservice tool for anonymizing text data (Nicolai et al., 2021). Für Informationen zur Registrierung/Freischaltung siehe: <https://www.qualiservice.org/de/helpdesk/webinar/tools.html>

### 3.2.2 Speicherung, Backup und Sicherung von Forschungsdaten

Die erhobenen, aufbereiteten und analysierten Daten und die dazugehörige Dokumentation sind das zentrale Element jeder Forschung. Deshalb ist es unerlässlich, sich spätestens am Projektbeginn ein Konzept zu überlegen, wo digitale Daten und Dokumentation in der Projektphase sicher gespeichert werden können und wie sich Datenverluste und unautorisierte Zugriffe verhindern lassen.

#### **Speicherorte auswählen**

Die Wahl eines geeigneten Speicherorts und ggf. eines Rechtemanagements hängt stark mit den individuellen Bedürfnissen des jeweiligen Projekts zusammen. Nachfolgend werden einige Aspekte aufgelistet, die bei der Auswahl der Speicherlösung eine Orientierung bieten können.



#### **Wichtige Aspekte für die Auswahl einer geeigneten Speicherlösung**

*Speicherplatz:* Wie groß sind die Datenmengen, die im Projekt insgesamt anfallen werden? Welche Speichersysteme können den benötigten Speicherplatz bereitstellen?

*Datenzugang:* Wer braucht Zugang zu den Daten und wie wird der Zugang bereitgestellt? Falls Personen aus anderen wissenschaftlichen Einrichtungen mitarbeiten, kann ein Speichersystem notwendig werden, das einen externen Zugang ermöglicht.

*Teamarbeit:* Wird es notwendig, simultan an den Dateien zu arbeiten? Dann könnte eine Sync&Share-Funktion des Speichersystems sinnvoll sein.

*Speicherort:* Wo liegen die Daten? Befinden sich die Daten auf Servern der eigenen Forschungseinrichtung, auf Servern in Europa oder anderen Staaten (z. B. den USA)? Dies ist bei personenbezogenen Daten wichtig, da jeweils das Datenschutzrecht des Landes gilt, in dem sich der Datenserver befindet. Die DSGVO erlaubt die Datenspeicherung außerhalb der EU nur, wenn die betroffenen Personen ihre Einwilligung gegeben haben oder ein gleichwertiges Datenschutzniveau in dem Land vorliegt. Es ist zu empfehlen, ein lokales Speicherangebot der eigenen Einrichtung zu nutzen, welches einer internen Prüfung unterzogen wurde und in aller Regel den europäischen Datenschutzrichtlinien entspricht.

*Zugriffsbeschränkungen:* Falls mit personenbezogenen Daten gearbeitet wird, sollte das Speichersystem Möglichkeiten bereitstellen, den Zugriff auf einzelne Ordner zu beschränken. Eventuell wird es in diesen Fällen auch notwendig, die Daten an einem besonders geschützten Speicherort aufzubewahren, der nur sehr wenigen Personen zugänglich ist und erst im weiteren Forschungsprozess, wenn die Anonymisierung stattgefunden hat, diese auf ein gemeinsames Speichersystem zu übertragen.

*Backup:* Bietet das Speichersystem ein regelmäßiges, automatisiertes Backup, durch das sich Datenverluste leichter vermeiden lassen? In welchen Abständen wird das Backup durchgeführt und wie lange ist die Vorhaltezeit des jeweiligen Backups?

*Speicherdauer:* Wie lange wird der Speicher benötigt? Viele Speichermedien haben eine sehr begrenzte Lebensdauer. Einige von ihnen (z. B. externe Festplatten, gebrannte DVDs oder CDs), können nicht einmal die Anforderung der guten wissenschaftlichen Praxis nach einer sicheren Aufbewahrung für mindestens zehn Jahre garantieren.

*Budget:* Wieviel Geld steht für Speicherung und Backup zur Verfügung und wie teuer sind die jeweiligen Speicherlösungen inklusive Backup?

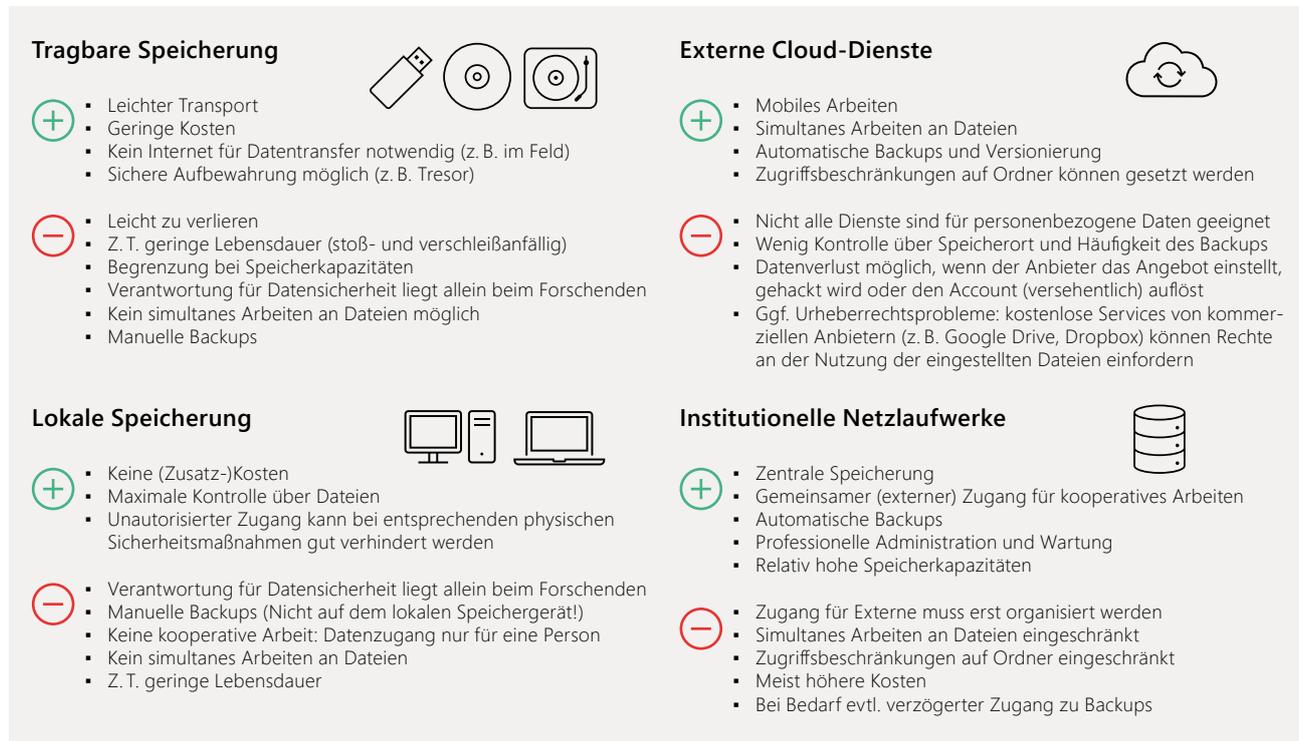
#### **Welche Speicherlösungen gibt es und was sind ihre jeweiligen Vor- und Nachteile?**

Es ist eine Vielzahl von Speicherlösungen möglich. Häufig werden eine oder mehrere der folgenden Speicher gewählt:

- Tragbare Speicher (z. B. USB-Sticks, externe Festplatte, CD/DVD)
- Lokale Speicher im eigenen Arbeitsraum (v. a. Desktop-PC, Notebook)
- Cloud-Dienste (u. a. kommerzielle oder nicht-kommerzielle Anbieter sowie extern oder lokal an der eigenen Einrichtung gehostete Clouds)
- Institutionelle Netzlaufwerke

Die nachfolgende ► Abbildung 2 soll in einer kurzen Übersicht die wichtigsten Vor- und Nachteile dieser vier Speichermöglichkeiten aufzeigen.

Abbildung 2: Vor- und Nachteile häufig genutzter Speichermöglichkeiten



Quelle: Eigene Darstellung orientiert an CESSDA Training Team (2020) und Eisentraut & Hutflasz (2021)

### Backup-Strategie festlegen

Backups der gesammelten Dateien zu erstellen, die im Forschungsprozess entstehen, ist unverzichtbar. Denn regelmäßige Backups schützen vor Datenverlusten, die z. B. durch Hardware- oder Software-Fehler, durch Virus- oder Hackerangriffe sowie durch individuelle Anwendungsfehler (z. B. Überschreiben, versehentliches Löschen von Daten) oder Naturkatastrophen (z. B. Brände, Überschwemmungen, Erdbeben) auftreten können. Falls ein solches Ereignis eintritt, sparen Backups sehr viel Zeit und Geld.

### Was ist ein gutes Backup?

Im Optimalfall wird ein Backup automatisch und regelmäßig durchgeführt, so dass immer eine aktuelle Sicherungskopie vorliegt. Ein gutes Backup-System sichert nicht nur die neueste Version der Daten, sondern hält auch ältere Versionen zurück, auf die bei Bedarf zugegriffen werden kann. Beispielsweise wenn ein Bearbeitungsfehler in den Daten erst zu einem späteren Zeitpunkt auffällt oder wenn bei einer Fehlersuche in den Daten einzelne Arbeitsschritte rekonstruiert werden müssen. Das Backup sollte nicht nur Daten, sondern in regelmäßigen Abständen, wann immer möglich, auch notwendige (nicht-proprietäre) Softwareanwendungen einbeziehen, die für die Reproduzierbarkeit der Daten und Forschungsergebnisse notwendig sind.



### Tipps für die Entwicklung einer guten Backup-Strategie

*Verantwortlichkeiten definieren:* Es sollte eine hauptverantwortliche Ansprechperson für das Thema benannt und festgeschrieben werden sowie welche Rolle den Projektmitgliedern im Backup-Prozess zukommt (z. B. Durchführen manueller Backups, Kontrolle der Backups).

*3-2-1-Regel anwenden:* Nach dieser Regel sollten mindestens drei Datenkopien auf mindestens zwei unterschiedlichen Typen von Speichersystemen angelegt werden. Mindestens eines der Speichersysteme sollte sich an einem externen Standort befinden. Falls die Speichermengen sehr hoch oder die Daten sehr sensibel sind, kann es im Einzelfall auch notwendig sein, mit weniger als drei Kopien zu arbeiten, diese aber besonders gut gegen Datenverluste abzusichern.

*Bevorzugt regelmäßige automatische Backup-Systeme wählen:* Im Optimalfall wird ein Backup automatisch in regelmäßigen, relativ kurzen Abständen durchgeführt, so dass immer eine aktuelle Sicherungskopie vorliegt und ältere Versionen vorgehalten werden, die für Rekonstruktionsprozesse bei spät entdeckten Fehlern verwendet werden können. Dies erleichtert Forschenden die Arbeit und verhindert Datenverluste. Falls aus berechtigten Gründen ein Speichersystem gewählt wird, bei dem ein manuelles Backup notwendig wird, sind Routinen festzulegen, in welchen Zeitabständen und in welcher Form (d. h. Backup-Varianten) Sicherungskopien erstellt werden.

*Dienste der eigenen Einrichtung nutzen:* Wissenschaftliche Einrichtungen bieten in der Regel Speicherdienste, in denen ein automatisiertes, regelmäßiges Backup integriert ist. Bei diesen Diensten wurde in der Regel auch von der jeweiligen Forschungseinrichtung geprüft, dass die Datenschutzrichtlinien eingehalten werden.

*Personenbezogene Daten schützen:* Backup-Dateien, die personenbezogene Daten enthalten, sind genauso wie die Originaldateien zu schützen. Deshalb sollten die gleichen Schutzmaßnahmen ergriffen werden (z. B. Datenverschlüsselung oder Beschränkung des Datenzugriffs).

*Backups kontrollieren:* Falls es das jeweils verwendete Backup erlaubt, sollte zur Sicherstellung der Datenintegrität in regelmäßigen Abständen getestet werden, ob die Datenwiederherstellung funktioniert.

*Aufbewahrungsdauer der Backups definieren:* Backups sind Teil des Löschkonzepts für Forschungsdaten. Es sollte bereits zu Beginn definiert werden, wie lange Backups aufbewahrt werden. Insbesondere personenbezogene Daten sollten spätestens am Ende des Forschungsprojekts zuverlässig und endgültig aus allen Backup-Speichern entfernt werden (z. B. durch Löschsoftware wie dban, Eraser oder BCWipe) (vgl. Art. 17 Abs. 1 lit. a DSGVO).

### **Datensicherheit erhöhen**

Alle Forschenden sollten Maßnahmen ergreifen, um ihre digitalen Daten zu schützen. Ansonsten besteht die Möglichkeit, dass unautorisierte Personen auf die Daten zugreifen, sie (absichtlich oder unabsichtlich) manipulieren, löschen, unrechtmäßig selbst verwenden oder an Dritte weitergeben. Forschungsprojekte, die mit personenbezogenen Daten arbeiten, haben in dieser Hinsicht eine noch größere Verantwortung, vor allem wenn es sich um sensible Informationen zu Personen handelt. Das Datenschutzrecht hat für diesen Fall vorgesehen, dass Forschende die personenbezogenen Daten mit spezifischen *technischen und organisatorischen Maßnahmen (TOM)* schützen müssen.



### Tipps zur Erhöhung der Datensicherheit im Projekt

#### Technische Maßnahmen

- Sichere Passwörter für Computer- und Speichersysteme verwenden
- Verschlüsselung von Daten und Festplatten
- Virussoftware nutzen und Firewalls aktivieren
- Regelmäßige (Sicherheits-)Updates von Software und Betriebssystemen
- Keine offenen unverschlüsselten WLAN-Verbindungen nutzen
- Versenden von Daten per E-Mail vermeiden
- Löschen nicht mehr benötigter Daten über sichere Wege (durch Löschsoftware wie dban, Eraser oder BCWipe, durch physische Vernichtung tragbarer Speichersysteme oder durch Schreddern von Papierdokumenten)

#### Organisatorische Maßnahmen

##### Physische Sicherheitsmaßnahmen

- Tragbare Speichermedien an sicheren, abschließbaren Orten lagern (z. B. Tresore, Dokumentenboxen)
- Automatische Desktopsperrern einrichten
- Türen und Fenster bei Abwesenheit abschließen

##### Zugriffsbeschränkungen/-kontrollen

- Einführen eines Rechtemanagements
- Personenanzahl mit Zugriffsrechten minimieren

##### Sensibilisierung, Regeln und Kommunikation im Projektteam

- Checklisten zu den technischen Maßnahmen entwickeln
- Sensibilisierung und aktive Kommunikation im Projektteam zu den Themen
- Beschäftigte, die mit personenbezogenen oder vertrauenswürdigen Forschungsdaten arbeiten, zur Geheimhaltung und auf spezielle Verhaltensregeln im Umgang mit den Daten verpflichten. Unter Umständen ist dies schon im Arbeitsvertrag geregelt.

### 3.2.3 Daten organisieren

Zu Beginn jedes Forschungsprojekts sollte überlegt werden, wie die Daten am effizientesten für die tägliche Projektarbeit organisiert werden können. Gleichzeitig ist im Blick zu behalten, dass die Datenorganisation möglichst nachhaltig erfolgt, so dass im Falle einer geplanten Archivierung und (eingeschränkten) Zugänglichmachung der Forschungsdaten in einem Repository oder einem FDZ wenig zusätzlicher Aufwand für die Forschenden am Projektende notwendig wird.

#### Ordnerstruktur aufbauen

Eine gute Ordnerstruktur liegt vor, wenn ein Ordnerverzeichnis inklusive Dateien an andere Forschende weitergegeben werden kann und sich diese selbstständig erschließen können, was in der Studie gemacht wurde und was die einzelnen Forschungsschritte waren. Sofern diese Bedingung vorliegt, sollte auch gewährleistet sein, dass die primärforschenden Personen im täglichen Arbeitsprozess einen guten Überblick über ihre Dateien behalten und diese leicht auffinden.



### Tipps für das Anlegen von Ordnerstrukturen

Falls es in der eigenen Fachdisziplin bereits Best-Practice-Regeln gibt, sollten diese genutzt werden.

Grundsätzlich sollten die Ordnerstrukturen

1. *hierarchisch* angelegt sein,
2. *maximal 3 Unter-Ordner-Ebenen* umfassen,
3. *klar und konsistent* aufgebaut sein.

Falls quantitativ geforscht wird, empfiehlt es sich, die Dateien so abzuliegen, dass eine vollständige *Reproduzierbarkeit* der Ergebnisse gewährleistet ist.

Es sollte eine „Readme-Datei“ angelegt werden, in der alle Regeln zu den Ordnerstrukturen aufgeführt sind.

Bei Projektteams empfiehlt es sich, vorab *Verantwortlichkeiten* festzulegen, welche Personen bei Bedarf neue Ordnerstrukturen in gemeinsamen Verzeichnissen erstellen und den Ordnerzugriff regeln.

### Benennungskonventionen für Dateien und Ordner festlegen

Es gibt in den Sozial-, Verhaltens- und Wirtschaftswissenschaften bislang keine eindeutig etablierten Benennungskonventionen. Einige allgemeine Hinweise existieren jedoch, die für die Entwicklung eigener Regeln bei der Benennung von Ordnern und Dateien hilfreich sein können.



### Tipps für die Benennung von Dateien und Ordnern

Die Namensgebung sollte *konsistent* erfolgen. Einheitliche Benennungsregeln für Ordner und Dateien im gesamten Forschungsprozess erleichtern es, Dinge aufzufinden und den Überblick zu behalten. Konsistent bedeutet, dass immer dieselben festgelegten Elemente in der gleichen Reihenfolge in Dateinamen genutzt werden (z. B. Datum, Schlagworte zum Inhalt, Namenskürzel zur bearbeitenden Person, Version: YYYYMMDD\_stichprobeX\_ag\_v001).

Die Namen müssen *maschinenlesbar* sein. Deshalb sind Leerzeichen und Sonderzeichen (z. B. \$, @, %, #, &, \*, \*, ., (), !, /, ?) zu vermeiden, da sie nicht von allen Computersystemen korrekt gelesen werden können. Ordner- und Dateinamen sollten stattdessen aus den Buchstaben *a-z*, den Zahlen *0-9*, und Unterstrichen (*\_*) bestehen. Großbuchstaben und Bindestriche (*-*) können verwendet werden, solange die verwendeten Softwareprogramme zwischen Groß- und Kleinschreibung bzw. zwischen Binde- und Gedankenstrichen (*-*) unterscheiden können. In Anlehnung an Programmiersprachen werden für Namensgebungen von Dateien häufig *vier Varianten* empfohlen: *→snake\_case*, *→kebab-case*, *→camelCase* und *→PascalCase*.

Die Namen sollten den *Inhalt* beschreiben. Je besser der gewählte Name die Datei- oder Ordnerinhalte abbildet, umso leichter wird er von „menschlichen“ Nutzenden verstanden. Dies erleichtert die Auffindbarkeit innerhalb der Dateistruktur und erhöht die Arbeitseffizienz. Alle Metadaten, die bereits im Dateinamen enthalten sind, müssen zudem nicht mehr an anderer Stelle erklärt werden.

Es sollten Namen gewählt werden, die *so kurz wie möglich, aber so lange wie nötig* sind, um den Inhalt der Dateien oder des Ordners zu verstehen. Ein grober Richtwert für die maximale Länge eines Dateinamens sind 32 Zeichen. Ordnernamen sollten deutlich kürzer sein. Eine Orientierung bietet die Tatsache, dass auf den meisten Betriebssystemen ein Dateiname inklusive Ordnerpfad nicht länger als 255 Zeichen sein darf.

Die Dateinamen sollten so gewählt werden, dass sie *automatisch eine sinnvolle Sortierung der Dateien* in Ordnern erzeugen. Da die Reihenfolge der Elemente im Dateinamen festlegt, wie die Dateien anschließend sortiert werden, sollten immer die Merkmale im Dateinamen vorne platziert werden, nach denen sortiert werden soll. Aus diesem Grund werden häufig Zahlen an den Anfang von Dateinamen gestellt, um die chronologische Reihenfolge abzubilden und die neueste Datei leicht zu finden. Meistens wird hierfür das Erstellungsdatum verwendet: YYYYMMDD (z. B. 20220420).

Die Namenskonventionen sollten dem *eigenen Arbeitsstil entsprechen*. Forschende arbeiten täglich mit den Daten und kennen die Besonderheiten des Projekts und der Daten am besten. Deshalb sollten sie entscheiden, mit welchen Bezeichnungen sie am besten klarkommen.

Die *Benennungsregeln* sollten in einer Textdatei (sogenannter Readme-Datei) *dokumentiert* werden. Alle Teammitglieder sollten Zugang zu den Benennungsregeln haben und sie konsistent anwenden.

### 3.2.4 Daten fortlaufend dokumentieren bzw. kontextualisieren

Eine wesentliche Voraussetzung für die (Nach)-Nutzung von Forschungsdaten ist eine gute *Dokumentation* bzw. *Kontextualisierung*. Gemeint ist damit, dass Forschende die Generierung und die weitere Bearbeitung der Daten so darstellen und erläutern, dass auch andere Forschende das Vorgehen nachvollziehen und ggf. die Daten auswerten können. Die Dokumentation soll also Sekundärnutzende dabei unterstützen, die Daten „bestmöglich zu verstehen und zu verwenden, quasi ‚in die Schuhe der Primärforschenden zu schlüpfen‘, obwohl ihnen die unmittelbare Erfahrung der Primärerhebung fehlt“ (Steinhardt et al., 2020, S. 18). In der qualitativen Forschung wird dabei auch von „Kontextualisierung“ gesprochen, weil die interpretative Herangehensweise in der qualitativen Forschung stets auf ein Sinn-Verstehen abzielt, und Sinn sich aus dem Bezug auf Kontext(e) ergibt; somit müssen die Daten mit den entsprechenden Kontextinformationen angereichert werden. Für qualitative wie für quantitative Daten gilt, dass eine gute Forschungsdokumentation in aller Regel ein zentrales Qualitätskriterium ist, von dem die Nachvollziehbarkeit der Forschungsergebnisse und die Nachnutzbarkeit der Daten abhängt. Deshalb sollten Forschende diesem Punkt besondere Aufmerksamkeit widmen.

Zwar dürften die *Vorteile* einer sorgfältigen und detaillierten Forschungsdokumentation evident sein, aber in der Forschungspraxis wird die Dokumentation oft zugunsten vermeintlich wichtigerer Aufgaben vernachlässigt. Sie ist allerdings nicht nur Bestandteil guter wissenschaftlicher Praxis, indem sie Transparenz und Nachvollziehbarkeit gegenüber Dritten schafft. Eine gute Forschungsdokumentation trägt auch zur Sichtbarkeit – und damit zur Nutzung und Zitation – der Forschungsdaten bei. Zudem bietet sie Möglichkeiten dafür, den hinsichtlich akademischer Reputation oft vernachlässigten Bereich der Datengewinnung aufzuwerten und die dort erbrachten Leistungen zu honorieren. Nicht zuletzt ist eine gute Forschungsdokumentation auch für das datengenerierende Forschungsprojekt selbst hilfreich, indem Informationen zum Forschungsvorgehen festgehalten werden und damit sowohl im Projekt weitergegeben als auch systematisch für Auswertungen genutzt werden können.

Das *Vorgehen* bei der Forschungsdokumentation wird in der sozialwissenschaftlichen Methodenlehre kaum behandelt. Es kann je nach Disziplin und Daten ganz unterschiedlich aussehen und auf verschiedenen Ebenen, mit unterschiedlichen Verfahren und in unterschiedlicher Detailtiefe erfolgen bzw. erforderlich sein. Konzeptuell unterschieden wird oft zwischen einer Dokumentation auf der Ebene des gesamten Forschungsprojekts (*study-level*) und auf der Ebene der einzelnen Datengewinnung, z. B. eines Interviews (*data-level*). Zusätzlich kann eine dem Forschungsprojekt übergeordnete Ebene mitgedacht werden, die den zeitgenössischen, kulturellen, regionalen etc. Kontext der Forschung umfasst, denn dieser kann vor allem mit zunehmender zeitlicher und örtlicher Distanz zum Projekt weniger bekannt sein.

Das *Hauptziel* der Dokumentation bzw. Kontextualisierung ist stets, die Entstehung der Daten (langfristig) nachvollziehbar zu machen. Dafür bieten sich verschiedene Darstellungsoptionen an, von denen die drei verbreitesten in den folgenden Absätzen erläutert werden: per Metadaten, weiteren Forschungs- bzw. Kontextmaterialien und per Bericht. Der Umfang der Dokumentation kann je nach Daten unterschiedlich sein: „So kann z. B. schon der Berechnungscode (z. B. do-file, sps-file) ausreichen, um einem Nachnutzer zu zeigen, wie ein Datensatz auf Basis frei verfügbarer und gut dokumentierter Primärdaten erstellt wurde“ (ZBW et al., 2014, S. 21).

Wichtig ist in jedem Fall ein guter Zugang zu zentralen Bestandteilen der Dokumentation (z. B. Metadaten, Bericht), denn dies hilft potenziellen Nachnutzenden bei der Entscheidung, ob die Daten für sie relevant sein könnten.

Die wichtigste *Empfehlung* für die Dokumentation ist, diese prozessbegleitend durchzuführen, d. h. soweit wie möglich in den Forschungsprozess zu integrieren. Im Vergleich zu einer Erstellung erst kurz vor Abschluss des Forschungsprojekts wird auf diese Weise nicht nur der Gesamtaufwand für die Dokumentation verringert, sondern oft auch die Qualität verbessert, da der Forschungsprozess nicht aufwändig aus der Erinnerung rekonstruiert werden muss und auf diese Weise vermieden wird, dass die Darstellung von den Forschungsergebnissen her verfasst wird. Zudem sei auf Hilfen und Empfehlungen zur Dokumentation verwiesen, die an die Bedarfe des eigenen Forschungsprojekts angepasst werden können (z. B. Heuer et al., 2020; ZBW et al., 2014; VerbundFDB, 2019).

### **Dokumentation mittels Metadaten**

Sowohl für quantitative als auch für qualitative Forschungsdaten ist eine Darstellung über Metadaten unverzichtbar. Als „Daten über Daten“ geben Metadaten Informationen sowohl zu inhaltlich-fachlichen als auch zu formell-technischen Aspekten der Daten. Im Unterschied zu anderen Formen der Dokumentation dienen Metadaten vorrangig der Information von externen Dritten und werden in der Regel auch erst mit der Datenarchivierung erfasst. Ihre Rolle für die Such- und Auffindbarkeit – und damit auch für die (potenzielle) Nachnutzung – der Forschungsdaten kann jedoch nicht hoch genug eingeschätzt werden. Bei quantitativen Daten ist es auch denkbar, dass bereits die Metadaten erschöpfend Auskunft über die Daten und deren Eignung für Nachnutzungen geben – beispielsweise, wenn diese nicht nur auf der Ebene des Datensatzes, sondern auch für die Einzeldaten (z. B. Variablen) detailliert vorliegen.

Eine wichtige Empfehlung ist, Forschungsdaten möglichst mittels weit verbreiteter →*Metadatenstandards* zu beschreiben. Auf diese Weise wird die Erfüllung der FAIR-Kriterien Interoperabilität und Auffindbarkeit deutlich verbessert. Bei Metadaten gibt es allgemeine Standards wie „Dublin Core“, der die Beschreibung von Dokumenten und Objekten mithilfe von 15 Kernelementen ermöglicht. Darüber hinaus gibt es zahlreiche disziplinspezifische Standards, die detaillierte Darstellungen erlauben. Weit verbreitet in den Sozial-, Verhaltens- und Wirtschaftswissenschaften sind beispielsweise der Metadatenstandard der „Data Documentation Initiative“ (DDI) sowie das *data* Metadatenchema, das den Metadatenstandard von „DataCite“ um spezifisch sozial- und wirtschaftswissenschaftliche Metadaten ergänzt.

Für die in den Metadaten enthaltenen Informationen sollten ebenfalls möglichst einheitliche Begriffe verwendet werden. Dies verbessert nicht nur die Such- und Auffindbarkeit der Daten, sondern ermöglicht auch den Vergleich und die Verknüpfung von unterschiedlichen Datensätzen. Es wird deshalb empfohlen, beim Ausfüllen der Metadatenfelder nach Möglichkeit ein sogenanntes →„kontrolliertes Vokabular“ zu verwenden. Dabei handelt es sich um Vorgaben bzw. Definitionen von Begriffen und Regeln, die in Wortlisten oder strukturierten Thesauri zusammengefasst sind. So ist für die Sozialwissenschaften beispielsweise der vom Consortium of European Social Science Data Archives (CESSDA) erstellte „European Language Social Science Thesaurus“ (ELSST; CESSDA and Service Providers, 2022) relevant. Zu den kontrollierten Vokabularen zählen auch →Normdaten – wie die von Wissenschaftseinrichtungen im deutschsprachigen Raum oft genutzte „Gemeinsame Normdatei“ (GND; Deutsche Nationalbibliothek, 2022) – sowie nationale und internationale Klassifikationen, z. B. von Wirtschaftszweigen, Berufen oder Schultypen bzw. Bildungsniveaus.

Bedauerlicherweise sind sowohl Metadatenstandards als auch kontrollierte Vokabulare oft auf die Anforderungen quantitativer Forschung bzw. Daten zugeschnitten. Eine adäquate Darstellung qualitativer Forschung ist ein Desiderat, für das bislang nur vereinzelt Lösungen vorliegen. Generell ist bei Fragen zur Verwendung von Metadaten eine Kontaktaufnahme mit FDM-Personal bzw. Forschungsdateninfrastrukturen zu empfehlen, die über entsprechende Angebote und Anforderungen informieren können.

### **Dokumentation mittels (Kontext-)Materialien**

Während Metadaten einen guten (ersten) Eindruck von Forschungsdaten vermitteln können, bieten „Kontextmaterialien“ detaillierte Einblicke und Informationen über die Daten. Dabei handelt es sich um Dokumente bzw. *Materialien*, die – *unabhängig von einer Archivierungsabsicht – in den unterschiedlichen Phasen eines Forschungsprojekts entstehen*: Vom Projektantrag über einen Zwischenbericht bis zum Abschlussbericht; vom Votum der Ethikkommission über das Muster für die informierte Einwilligung bis zum Anonymisierungskonzept; von Vorüberlegungen zur Fallauswahl und Methodik über Erhebungsinstrumente (z. B. Fragebogen, Interview-Leitfaden, Feld- und Methodenbericht) und Aufbereitungsverfahren (z. B. Transkriptionsregeln, Anonymisierungsmaßnahmen) bis hin zu Auswertungshilfen (z. B. Memos und Fallbeschreibungen, Kategoriensysteme, Softwarebefehle und Codebooks).

Gemeinsam ist diesen Informationen und Materialien, dass sie aufgrund ihrer Herkunft aus dem Forschungsprozess für das Verständnis der Forschungsdaten und für deren Nachnutzung von Bedeutung sein können. Forschende sollten somit *überlegen, bei welchen Materialien eine Bereitstellung möglich und sinnvoll ist*. Dabei sind stets, – wie bei den Forschungsdaten selbst – ethische und rechtliche Aspekte zu berücksichtigen, z. B. sensible und personenbezogene Daten ggf. zu anonymisieren. Hinsichtlich des Zugangs zu diesen Materialien können FDZ sicherstellen, dass Kontextmaterialien den gleichen Schutz erfahren wie die Forschungsdaten selbst – zumal die Grenze zwischen diesen Kategorien oft fließend ist. Im Interesse der Nachnutzbarkeit ist darauf zu achten, dass „offenen“ gegenüber geschlossenen und

proprietären Formaten und Standards der Vorzug gegeben wird. Die zentrale Empfehlung lautet somit, zu überlegen, welche Materialien für das Verständnis der Forschungsdaten wichtig sein könnten und diese soweit möglich zur Verfügung zu stellen.

### **Dokumentation mittels (Daten- bzw. Studien-)Bericht**

Eine dritte – und für die Nachvollziehbarkeit und Nachnutzbarkeit von Forschungsdaten zentrale – Form der Forschungsdokumentation ist *das Verfassen eines eigenständigen Texts zur Beschreibung und Erläuterung der Daten*. Ein solcher Text zur Datenentstehung kann je nach Daten und Disziplin unterschiedliche Formen annehmen: Z. B. als bloßes Dokument, als Arbeitspapier oder als Fachaufsatz. Insbesondere FDZ verlangen oft einen entsprechenden Bericht – der z. B. als „Studienreport“, „Daten- und Methodenbericht“ oder „Kontextbogen“ bezeichnet wird –, weil auf diese Weise Sekundärnutzenden die für die Nachnutzung erforderlichen Kontextinformationen gebündelt zur Verfügung gestellt werden können. In der Regel enthält ein solcher Bericht Angaben zum Forschungsprojekt (z. B. Institutionen und Personen; Forschungsfrage, Vorarbeiten und Konzepte), zur Vorbereitung und Durchführung der Datengenerierung (Auswahl von Methoden, Teilnehmende etc.) sowie zu allen weiteren Schritten bei der Aufbereitung und Auswertung, die zu den finalen Daten geführt haben (z. B. Transkription, Anonymisierung, Auswertungsverfahren). Darüber hinaus können auch Angaben zu den Forschungsergebnissen, zu im Projektkontext entstandenen Beiträgen, zu weiteren Kontextmaterialien, und zu Nachnutzungspotentialen der Daten enthalten sein.

Während die zuvor genannten Kontextmaterialien quasi urwüchsig im Forschungsprozess entstehen, erfordert ein eigenständiger Bericht zusätzliche Arbeit von den Forschenden, die bei der Projektplanung berücksichtigt werden sollte. Allerdings reduziert sich auch hier erfahrungsgemäß der Aufwand, wenn das Verfassen des Berichts projektbegleitend durchgeführt wird. Zudem kann beim Verfassen des Berichts oftmals auf Vorarbeiten (DMP, andere Kontextmaterialien) zurückgegriffen werden. Der Bericht bietet darüber hinaus aber auch die Möglichkeit, implizite und informelle Kontexte darzustellen, die aus den anderen (Kontext-)Materialien nicht ersichtlich werden.

Nicht zuletzt ist ein solcher Bericht zur Forschungsdokumentation eine *wichtige Methodenpublikation* des Forschungsprojekts, die die Möglichkeit bietet, den Forschungsprozess ausführlich darzustellen. FDZ verlangen häufig nicht nur einen solchen Bericht, sondern bieten oft die Möglichkeit, diesen Bericht – mit einem persistenten Identifikator versehen – auch als Publikation zu veröffentlichen.



### **Nützliche Tools und Links zum Thema **Datenschutz****

- Interaktive virtuelle Assistenten (iVA) zu Regelungen der Datenschutzgesetze (DSGVO, BDSG, LDSG; BERD@BW, Projektlaufzeit 2019-2022).
- Open Source PIA Software zur Erstellung einer Datenschutzfolgeabschätzung (CNIL, 2021).
- Die Datenschutz-Folgenabschätzung nach Art. 35 DSGVO: Ein Handbuch für die Praxis (Martin et al., 2020).
- Handreichung Datenschutz: 2. vollständig überarbeitete Auflage (RatSWD, 2020).
- Webseite zum Thema Datenschutz in den Sozial- und Wirtschaftswissenschaften (RatSWD, o.J.–b).
- Webseite zu datenschutzrechtlichen Aspekten und Handreichungen (VerbundFDB, 2022a).



### Nützliche Veröffentlichungen & Links zu **Speicherung, Backup, Sicherung und Organisation von Forschungsdaten**

- Open Science Training Handbook (Bezjak et al., 2018).
- CESSDA Data Management Expert Guide (CESSDA Training Team, 2020)
- During Data Collection – Overview (DataWiz Knowledge Base, o.J.–a)
- E-Learning Plattform zu Open Science (FOSTER, o.J.).
- Beitrag zur Dateioorganisation in empirischen Forschungsprojekten (Recker & Brislinger, 2019)
- Learning Hub zu Research Data Management (UK Data Service, o.J.-a).
- Managing and Sharing Data. Best Practice for Researchers (van den Eynden et al., 2011)
- Webseite zum Forschungsdatenmanagement (VerbundFDB, o.J.).



### Nützliche Veröffentlichungen zur **Kontextualisierung und Dokumentation von Forschungsdaten**

- Kontextualisierung qualitativer Forschungsdaten für die Nachnutzung: eine Handreichung für Forschende zur Erstellung eines Studienreports (Heuer et al., 2020)
- Leitfaden zur Kontextualisierung von qualitativen Befragungsdaten (VerbundFDB, 2021).
- Auffinden, zitieren, dokumentieren. Forschungsdaten in den Sozial- und Wirtschaftswissenschaften (ZBW et al., 2014)



## 3.3 Aufbereitungs- und Analysephase

Es ist wichtig, alle Schritte der Datenaufbereitung und Datenanalyse möglichst vollständig und nachvollziehbar zu dokumentieren. Dies erlaubt es anderen Forschenden, die Analyseergebnisse zu reproduzieren und erhöht die Transparenz der eigenen Forschung. Eine Reihe von Studien konnte zeigen, dass es selbst bei Vorliegen der Analyse-Files schwierig ist, die publizierten Ergebnisse zu reproduzieren (z. B. Hardwicke et al., 2021; Stodden et al., 2018). Im Folgenden sollen Prinzipien der Datenaufbereitung und Datenanalyse besprochen werden. Einschränkend muss erwähnt werden, dass die Ausführungen sich vor allem auf Studien mit quantitativer Orientierung beziehen. Bei anderer methodischer Ausrichtung (z. B. qualitative Forschung) können die zentralen Schritte der Datenaufbereitung und Datenanalyse differieren.

Es kann zwischen den Rohdaten und den aufbereiteten Daten unterschieden werden. In den Rohdaten-Files liegt die erste Übertragung der Rohdaten (z. B. Fragebögen, Audioaufzeichnungen von Interviews) in ein digitales Format vor. Die Rohdaten-Files werden im Rahmen der Datenaufbereitung in Analyse-Files transformiert. Typische Verarbeitungsschritte bei der Datenaufbereitung sind z. B. die Überprüfung und Transformation von Variablen, die Bildung von Skalenwerten, die Behandlung von Ausreißern und fehlenden Daten sowie die Kodierung von offenen Antworten. Für die Umsetzung dieser Schritte empfiehlt sich ein Workflow der Datenaufbereitung, der sich nach Brislinger & Moschner (2019, S. 108) aus drei Bausteinen zusammensetzt:



### Drei Bausteine zur Umsetzung des *Workflows der Datenaufbereitung* nach Brislinger und Moschner (2019, S. 108):

1. Eine hierarchische Ordnerstruktur, die eine strukturierte Ablage der Daten und Arbeitsschritte ermöglicht. Es empfiehlt sich eine Readme-Datei anzulegen, in der die Ordnerstruktur beschrieben wird (siehe für das Anlegen einer Ordnerstruktur Abschnitt 3.2.3).
2. Rohdaten-Files, aufbereitete Daten-Files und Analyse-Files. Diese liegen in verschiedenen Versionen vor, die die Schritte der Erfassung und Bearbeitung der Daten sichtbar machen.
3. Programmcode (Syntax-Files in R, Stata, SPSS etc.), der die Schritte der Variablenkodierung, Datentransformation und Fehlersuche ausführt.

Die wichtigste Regel für die Datenaufbereitung ist, niemals Veränderungen an den Rohdaten-Files der originalen Daten vorzunehmen, sondern jeweils neue Daten-Files mit den transformierten Variablen abzuspeichern. Es wird deshalb empfohlen, die Rohdaten-Files gleich zu Beginn in einem separaten Ordner schreibgeschützt abzulegen und an diesem Ordner keine Veränderungen vorzunehmen. Außerdem sollte jeder Verarbeitungsschritt durch Programmcode (Syntax-File) und nicht „per Hand“ durchgeführt werden (also durch Bedienung von Dropdown-Menüs oder Point-and-Click in der grafischen Benutzeroberfläche). Dies hat zum einen den Vorteil, dass alle Transformationen der Rohdaten dokumentiert werden. Zum anderen können so Fehler in der Datenaufbereitung (z. B. falsche Kodierung einer Variablen), die vielleicht erst in einem späteren Stadium (z. B. bei der Datenanalyse oder Publikation der Ergebnisse) entdeckt werden, auch nachträglich noch leicht korrigiert werden. Da die Rohdaten in völlig un bearbeiteter Form vorliegen, können mit Hilfe der angepassten Syntax-Files alle Verarbeitungsschritte reproduziert werden und so ein neuer Datensatz erstellt werden, der die korrigierten Werte umfasst.

Um die Arbeitsschritte bei der Datenaufbereitung und -analyse zu dokumentieren, sollten Änderungen in den Daten-Files und Syntax-Files immer auch zu neuen Dateinamen führen. Mit dem Dateinamen können z. B. der Inhalt und das Erstellungsdatum gekennzeichnet werden (siehe Abschnitt 3.2.3 zu Benennungskonventionen). Dies kann im Verlauf eines Projekts zu sehr vielen Dateien führen, insbesondere wenn mehrere Personen gemeinsam an der Datenaufbereitung arbeiten und auch festgehalten werden soll, wer welche Änderungen an den Daten oder Syntax-Files über die Zeit vorgenommen hat. Es ist deshalb häufig zu empfehlen, mit einem Versionierungssystem zu arbeiten, in dem die Änderungen an den Dateien über die Zeit und Personen dokumentiert werden. Dazu bietet sich Software zur Versionskontrolle (z. B. git/GitHub, Bazaar) an, mit der die verschiedenen Versionen der Dateien verwaltet werden können. Eine Versionskontrolle ermöglicht nicht nur die effiziente Dokumentation und Reproduktion eines jeden Arbeitsschritts, sondern erleichtert auch das kollaborative Arbeiten, da jede Änderung protokolliert wird und somit einer Person eindeutig zugeschrieben werden kann.



### Weitere Tipps für einen *effizienten und reproduzierbaren Workflow* (Christensen et al., 2019; Long, 2009):

1. Die Syntax-Files für die Datenaufbereitung und Datenanalyse sollten ausführliche Kommentare enthalten, die die einzelnen Arbeitsschritte erklären.
2. Die Datenaufbereitung und Datenanalyse sollten voneinander getrennt und mit separaten Syntax-Files vorgenommen werden.
3. Die Benennung von Variablen sollte möglichst informativ sein (z. B. „männlich“ anstatt „geschlecht“, wenn 1 = männlich und 0 = weiblich).
4. Es sollte immer dokumentiert werden, welche Version einer Statistik-Software für die Analysen verwendet wurde (inkl. eventuell notwendiger Zusatzpakete).

In Abhängigkeit von der Fragestellung und dem Untersuchungsdesign kann es ratsam sein, sowohl die Hypothesen als auch die zu ihrer Prüfung geplanten z. B. statistischen Analysen bereits vorher zu dokumentieren (Präregistrierung der geplanten Analysen; Christensen et al., 2019). Damit soll verhindert werden, dass die Hypothesen im Nachhinein an die Analyseergebnisse angepasst werden oder die statistischen Analysen so lange variiert werden (z. B. durch jeweils unterschiedliche Spezifikation der Modelle), bis das gewünschte Ergebnis gefunden wird. Des Weiteren erlaubt die Präregistrierung eine klare Trennung von konfirmatorischen (Prüfung von Hypothesen) und exploratorischen Untersuchungen. Weniger sinnvoll ist eine Präregistrierung in Designs der qualitativen Sozialforschung, in denen explizit vorgesehen ist, dass die Gewinnung der Daten sukzessive in Abhängigkeit von ersten Ergebnissen angepasst wird (z. B. Theoretical Sampling).

In der quantitativen Sozialforschung wird es inzwischen als gute wissenschaftliche Praxis angesehen, dass für eine einzelne Veröffentlichung (z. B. ein Artikel in einer Zeitschrift) die Daten und der Programmcode zur Reproduktion der Ergebnisse zur Verfügung gestellt werden. Eine Reihe von Journals (z. B. das American Journal of Political Science) setzt voraus, dass gemeinsam mit dem Manuskript ein „Replication Package“ eingereicht wird, mit dem die berichteten Befunde verifiziert werden können (Christensen & Miguel, 2019). Ein Replication Package umfasst üblicherweise (1) eine Readme-Datei, die den Aufbau des Replication Package beschreibt, (2) die Daten-Files für die statistischen Analysen inklusive einer Beschreibung der Variablen (Codebook), (3) die Syntax-Files für die statistischen Analysen und (4) Informationen, wie die Inhalte des Manuskripts (z. B. Tabellen, Abbildungen) aus den Daten-Files reproduziert werden können.

Individualdaten in den Sozialwissenschaften umfassen häufig private bzw. sensible Informationen über die teilnehmenden Personen (siehe Abschnitt 3.2.1). Es wird deshalb aus datenschutzrechtlichen Gründen nicht immer möglich sein, die originalen Daten-Files im Rahmen eines Replikation Packages öffentlich zur Verfügung zu stellen. In manchen Fällen kann zusätzlicher Schutz der Teilnehmenden durch Zugangsbeschränkungen zu den Daten gewährleistet werden (siehe Abschnitt 3.4). Ist auch dies nicht möglich, sollten zumindest die Programmcodes für die statistischen Analysen zur Verfügung gestellt werden.



#### Nützliche Tools und Links zur **Aufbereitung und Analyse von Forschungsdaten**

- BITSS Resource Library mit Materialien für Forschungstransparenz und Reproduzierbarkeit (BITSS, o.J.).
- Webseite mit Informationen über Versionierung (DataWiz Knowledge Base, o.J.–b)
- Toolbox zur Unterstützung von Open-Science-Praktiken (OSC - LMU, o.J.)
- Webseite zur Integration von Transparenz und Replizierbarkeit in die Forschungsausbildung (Project TIER, o.J.)
- A handbook for reproducible, ethical and collaborative research (The Turing Way Community, 2021).
- Webseite zum Forschungsdatenmanagement (VerbundFDB, o.J.).

Veröffentlichen  
und Teilen  
der  
Daten

Archivieren

### 3.4 Veröffentlichungs-/Archivierungsphase

Auch kleine Forschungsprojekte sollten frühzeitig darüber nachdenken, ob und ggf. wo und wie sie die im Projekt generierten Forschungsdaten nach Projektabschluss archivieren und für weitere wissenschaftliche Nutzungen zur Verfügung stellen können. Diese Überlegungen sollten bereits *als Teil der Projektplanung vor einer Antragstellung* beginnen, denn nur so kann sichergestellt werden, dass die erforderlichen Tätigkeiten und Ressourcen in die Projektplanung einfließen (z. B. als Bestandteil eines DMP) und ggf. auch als Kosten in einem Projektantrag berücksichtigt werden können.

Der verbreitete englische Begriff des →„Data Sharings“ wird mit Öffnen, Teilen, Bereitstellen, Publizieren oder Veröffentlichen von Forschungsdaten übersetzt. Hinter allen Begriffen steht aber eine ähnliche Praxis: Forschungsprojekte wenden sich als Datengebende an eine Infrastruktureinrichtung für Forschungsdaten –

z. B. ein FDZ oder an ein institutionelles oder allgemeines Repositorium – und stellen darüber die Forschungsdaten (und ergänzende Materialien) für weitere Nutzungen zur Verfügung. Wichtig ist, dass aus rechtlichen und ethischen Gründen sensible und personenbezogene Daten oft nicht frei zugänglich gemacht werden können, sondern bestimmte Schutzmaßnahmen erfordern. Das Datenteilen über eines der im RatSWD akkreditierten FDZ bietet den Vorteil, dass diese Anbieter eine Erfüllung der Qualitätskriterien des RatSWD sowie eine Orientierung an den FAIR-Prinzipien gewährleisten und zugleich den erforderlichen Schutz der Daten – z. B. durch Vorgaben zur Nutzung, abgestufte Zugangswege oder Embargos – ermöglichen können.

### 3.4.1. Vorteile und Anforderungen beim Datenteilen

Forschungsfördernde Institutionen, Fachvereinigungen und Wissenschaftsverlage bzw. Zeitschriften erwarten zunehmend von Forschenden, dass diese ihre Forschungsdaten archivieren und zugänglich machen, oder darlegen, warum das Datenteilen nicht möglich oder sinnvoll ist. Die *Vorteile* des Data Sharings für die Scientific Community liegen in einer größeren Transparenz und Nachvollziehbarkeit von Forschungsergebnissen, aber auch in einer besseren (Nach)-Nutzung der oft aufwändig gewonnenen Forschungsdaten und einem daraus resultierenden größeren Datenangebot. Auf weitere Vorteile für Datengebende wie auch -nutzende wurde bereits in Abschnitt 3.2 verwiesen.

Auf der anderen Seite ist es denkbar, dass Forschungsdaten nicht oder nur eingeschränkt zur Verfügung gestellt werden können, weil ethische, rechtliche oder fachliche Erwägungen dem Datenteilen entgegenstehen. Insbesondere in der qualitativen Forschung stellen die Offenheit des Forschungsprozesses und die Heterogenität und Komplexität der Datenprodukte besondere Anforderungen an die Aufbereitung und Dokumentation der Daten. Allerdings gibt es auch für diese – oftmals besonders informationsreichen und damit für vielfältige Nachnutzungen interessanten – nicht-standardisierten Forschungsdaten in zunehmenden Maße FDM-Angebote in Form von Informationsmaterialien und Tools sowie Beratungen und Workshops, die insbesondere kleine Forschungsprojekte nutzen sollten. Auf diese Weise können Forschende einen Überblick über die Möglichkeiten und Grenzen des Datenteilens und die damit verbundenen Anforderungen gewinnen und vor diesem Hintergrund zu einer wohlüberlegten Entscheidung kommen.

### 3.4.2 Eine geeignete Dateninfrastruktur finden

Die Frage nach einer geeigneten Forschungsdateninfrastruktur zum Teilen der Daten ist oft eng verbunden mit dem „Ob“ und „Wie“ des Datenteilens und sollte deshalb ebenfalls so früh wie möglich – d. h. bereits in der Planungsphase vor einer Antragstellung – erwogen werden.<sup>8</sup> Mittlerweile gibt es eine fast unüberschaubare Anzahl von Infrastrukturen zum Archivieren und Teilen von Forschungsdaten. Einen umfangreichen Überblick bietet die Webseite des „Registry of Research Data Repositories“ (re3data). Diese ermöglicht auch ein Filtern des Angebots nach zahlreichen Kriterien. Um auf diese Weise eine geeignete Einrichtung für die Forschungsdaten zu finden, sollten Forschende sich mit den Standards im Fach sowie den Bedarfen des eigenen Projekts auseinandersetzen. Dazu zählen zum einen Fragen des Schutzniveaus der Daten und der angebotenen Zugangswege, der Such- und Auffindbarkeit, und der Eignung der Infrastruktur für bestimmte Datentypen. Zum anderen können auch Überlegungen zum Bedarf an Beratung und Unterstützung eine Rolle spielen oder das Leistungsangebot hinsichtlich der Datenkuratierung (z. B. technische und andere Qualitätsprüfungen, rechtliche Prüfungen und ggf. ergänzende Aufbereitungen).

Als Faustregel gilt, dass allgemeine bzw. offene Repositorien (z. B. Zenodo) in der Regel (nur) für Daten geeignet sind, bei denen keine Zugangsrestriktionen (z. B. aus ethischen oder rechtlichen Gründen) erforderlich sind, während ansonsten spezialisierte Anbieter zu wählen sind. Bisweilen wird auch explizit die Nutzung einer auf die Datenart bzw. Disziplin spezialisierten Forschungsdateninfrastruktur empfohlen, während institutionseigene oder allgemeine Repositorien nur als zweite und dritte Wahl genannt werden (OpenAIRE, 2022). Diese Empfehlungen beruhen auf der Überlegung, dass Forschungsdaten oft bestimmter technischer und organisatorischer Schutzmaßnahmen bedürfen, die von frei zugänglichen

<sup>8</sup> Nicht immer kann diese Entscheidung allerdings so früh im Projektverlauf getroffen werden, weil nicht immer absehbar ist, ob die Gewinnung von Primärdaten durch die Absicht des Datenteilens beeinträchtigt wird. Ggf. muss daher eine einmal getroffene Entscheidung revidiert oder über die (Teil)-Bereitstellung von Datenmaterial erst zu einem späteren Zeitpunkt entschieden werden.

Repositorien nicht angeboten werden (können). Darüber hinaus ermöglichen spezialisierte Forschungsdateninfrastrukturen zumeist eine bessere Auffindbarkeit der Daten durch die Verwendung von Metadatenstandards und können durch Beratung und Datenkuratierung zu einer höheren Datenqualität beitragen. Letzteres ist auch ein Vorteil von spezialisierten Forschungsdateninfrastrukturen gegenüber institutionseigenen Repositorien – bei denen zudem ggf. geklärt werden sollte, ob eine Langzeitarchivierung/-verfügbarkeit überhaupt möglich bzw. intendiert ist.

In den Sozial-, Verhaltens- und Wirtschaftswissenschaften hat sich das Modell einer dezentralen Forschungsdateninfrastruktur mit auf bestimmte Themen, Disziplinen oder Datenarten spezialisierten FDZ bewährt. Bei den derzeit 42 (Stand Januar 2023) vom RatSWD akkreditierten FDZ ist zudem die Erfüllung bestimmter Qualitätskriterien und die Orientierung an den FAIR-Prinzipien sichergestellt. Es ist allerdings zu beachten, dass nicht alle dort aufgeführten FDZ in der Lage bzw. dafür ausgerichtet sind, externe Forschungsdaten aufzunehmen. Auch diesbezüglich helfen aber die Webseiten von KonsortSWD und einzelner FDZ mit weiteren Informationen sowie eine direkte Kontaktaufnahme mit dem FDZ weiter.

### 3.4.3 Data Sharing als kooperative Aufgabe

Das Archivieren und Bereitstellen von Forschungsdaten ist idealerweise als Kooperation zwischen Forschenden und FDZ angelegt. Deshalb ist eine zentrale Empfehlung an Forschende, frühzeitig direkten Kontakt mit einem FDZ zu suchen. Auf diese Weise können einerseits mögliche Anforderungen an die Daten und die Dokumentation vonseiten des FDZ kommuniziert werden; andererseits bieten viele FDZ auch Beratung und Unterstützung für die Planung der erforderlichen Aufgaben an. Ein frühzeitiger Kontakt mit dem FDZ ist auch deshalb geboten, weil so mögliche Gebühren des FDZ für die Archivierung, Kuratierung und Bereitstellung der Daten in den Projektantrag aufgenommen werden können. Zudem kann es je nach Umfang der Datenvorbereitung für die Archivierung sinnvoll sein, dass Forschende die entsprechenden Tätigkeiten als ein eigenes kleines Arbeitspaket in einen Projektantrag aufnehmen und für dieses die notwendigen Ressourcen beantragen. Ein für die Sozialwissenschaften zentrales Angebot ist die kostenlose Basis-Archivierung von GESIS.<sup>9</sup>

Die Vorbereitung und die Dokumentation der Daten durch die Forschenden sollte möglichst projektbegleitend erfolgen, d. h. in den Verlauf des Forschungsprozesses integriert werden. Erfahrungsgemäß müssen die spätere Archivierung und Bereitstellung der Daten insbesondere beim Einholen einer Informierten Einwilligung, bei der Anonymisierung bzw. Pseudonymisierung der Daten und bei der Forschungsdokumentation mitgedacht werden. Bei Fragen sollte stets der Kontakt zu FDM-Personal der eigenen Institution oder zum jeweiligen FDZ gesucht werden, damit sichergestellt ist, dass die Daten ohne große Reibungsverluste archiviert und bereitgestellt werden können.

Vor der Übermittlung der Daten an ein FDZ wird zumeist ein *Datenübergabe- bzw. Datenüberlassungsvertrag* unterzeichnet, der die Rechte und Pflichten von Forschenden und FDZ regelt und in dem Forschende oft auch Bedingungen bzw. Anforderungen für die Bereitstellung/Nachnutzung festlegen können (z. B. Nutzungsarten, Zugangswege, Embargos). Je nach Vereinbarung und ggf. Leistungsangebot des FDZ werden die Daten nach der Übergabe auf bestimmte – technische, rechtliche oder inhaltliche – Aspekte geprüft und die (Langzeit)-Archivierung vorbereitet. Dabei kann sich die Notwendigkeit von inhaltlichen Klärungen und ggf. Nachbearbeitungen ergeben. Abgeschlossen wird die Archivierung in der Regel mit der Veröffentlichung der Metadaten und ggf. weiterer frei zugänglicher Materialien, anhand derer Nachnutzungsinteressierte Aufschluss über die Daten erlangen können. Die Datensätze selbst sollten einen persistenten Identifikator (z. B. DOI) erhalten, damit sie sichtbar und zitierbar sind (vgl. ZBW et al., 2014, S. 17).

Je nach erforderlichem Schutzniveau der Daten bieten FDZ oft unterschiedliche Zugangswege an. Dabei wird unterschieden zwischen allgemein zugänglichen – aber dafür ggf. sehr stark anonymisierten – Public Use Files (PUF) oder auch Campus Files (CF) für die akademische Lehre und weniger zugänglichen – aber dafür ggf. weniger stark anonymisierten – Scientific Use Files (SUF) für die wissenschaftliche Forschung. Es ist auch möglich, dass Forschungsdaten nur vor Ort an Gastarbeitsplätzen in einem besonders gesicherten „Safe Room“ nachnutzbar sind; dies ist z. B. oft bei Daten der Fall, bei denen eine Anonymisierung technisch oder inhaltlich nicht oder nur eingeschränkt möglich bzw. sinnvoll ist.

<sup>9</sup> <https://data.gesis.org/sharing/>

Die Bedingungen, unter denen Forschungsdaten geteilt und nachgenutzt werden können, werden von den Datengebenden festgelegt. Viele FDZ bieten Datengebenden die Möglichkeit, diese Bedingungen detailliert zu spezifizieren, u. a. um der Vielfalt und besonderen Sensibilität von (personenbezogenen) Forschungsdaten gerecht zu werden. Diese Festlegungen geschehen oft in Form eines Datenüberlassungsvertrags zwischen FDZ und Datengebenden, in der z. B. Embargos, Zugangswege und Nutzungsoptionen festgehalten werden können. Demensprechend müssen Nachnutzende einen Nutzungsvertrag unterzeichnen, in dem sie sich auf diese Bedingungen für die Datennutzung verpflichten (vgl. auch den von KonsortSWD entwickelten Mustervertrag für Datennutzungen: Schallaböck et al., 2022).

Eine Alternative dazu sind →Lizenzen. Diese legen Nutzungsbedingungen in standardisierter Form fest. Die Standardisierung hat den Vorteil, dass Nachnutzende einfacher Forschungsdaten aus verschiedenen Quellen zusammenführen können, aber sie bieten für Datengebende weniger Gestaltungsspielraum. Wenngleich es eigens für Forschungsdaten entwickelte Lizenzmodelle gibt (z. B. „Open Data Commons“, ODC), so werden die „Creative Commons“ (CC)-Lizenzen am häufigsten genutzt. Bei Letzterem erlaubt es ein modulares System, aus vier Lizenzbausteinen sechs Lizenzen zusammensetzen, die sich vor allem dahingehend unterscheiden, ob Bearbeitungen der Daten geteilt werden dürfen und ob eine kommerzielle Nutzung erlaubt ist. Wenn keine Einschränkungen der Datennutzung erforderlich sind, dann wird vor allem die offene Lizenz CC BY, die lediglich eine Namensnennung/Attribution vorsieht, oder der vollständige Verzicht auf Nutzungsbedingungen mittels der „Public Domain Dedication“ (CC0) empfohlen (für Details vgl. Brettschneider et al., 2021). Generell ist zu beachten, dass nicht nur Nutzungsverträge, sondern auch Lizenzen rechtlich bindend sind und somit festlegen, was mit den Daten gemacht werden darf und was nicht.

Abschließend ist wichtig zu wissen, dass sich die von einzelnen FDZ angebotenen Dienstleistungen durchaus erheblich voneinander unterscheiden können, u. a. in Bezug auf die Beratung, die Datenkuratierung und die Formen der Zugänglichmachung (und auch in Bezug auf die damit einhergehenden Gebühren). Somit ist und bleibt die frühzeitige Kontaktaufnahme mit dem FDZ unerlässlich für Forschende.

### 3.5 Nachnutzung von Forschungsdaten

Der Aufbau von Datenpools zur Nachnutzung von Forschungsdaten im Sinne der FAIR-Prinzipien kann als *Gemeinschaftsgut* verstanden werden, das die Verbreitung eines treugeberischen Nutzungsverhaltens und die Institutionalisierung der Funktion eines Treuhänders als Funktionsvoraussetzungen hat (Ostrom, 1990). Damit Forschungsdaten aus verschiedenen Einrichtungen wie Universitäten, außeruniversitären Instituten oder kommerziellen Organisationen Datentreuhändern übermittelt werden, ist Anerkennung durch Zitation geboten und es bedarf des Vertrauens, dass diese Daten im Anschluss der Forschungsöffentlichkeit zur Verfügung stehen, genutzt werden und Missbrauch ausgeschlossen ist. Darüber hinaus verlangt die Archivierung von Forschungsdaten andere Organisationsprinzipien als die Katalogisierung nach Autor:in und Titel in Bibliothekskatalogen. Metadaten zum Datensatz und seiner Genese sind ebenso erforderlich wie Indikatoren zu den oft vielfältigen Themen, die in Datensätzen der Sozial-, Verhaltens- und Wirtschaftsforschung behandelt werden, um archivierte Daten auffindbar zu machen. Diesen Leitlinien folgte bereits 1960 die Gründung des „Zentralarchivs für empirische Sozialforschung“ – heute Teil der GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS) – als erstem Archiv für sozial- und wirtschaftswissenschaftliche (Umfrage)-Daten in Europa (Mayer & Dalhoff D., 1998). Das Konzept einer forschungsnahen Institution, die als Treuhänder Daten archiviert, hat seither horizontal und vertikal erhebliche Verbreitung gefunden.

Für *Datennutzende* ist nicht zuletzt angesichts schwindender Rücklaufquoten bei großen, standardisierten Befragungen der stetig anwachsende Datensatz für Sekundäranalysen von großem Interesse (Birke & Mayer-Ahuja, 2017; Medjedović, 2014; Mochmann, 2014). Die zeitlichen Ressourcen der beforschten Subjekte sind begrenzt und vermehren sich nicht mit wachsender Nachfrage seitens der Sozial-, Verhaltens- und Wirtschaftswissenschaften. Schon aus diesem Grund sind Sekundäranalysen bei inhaltlichen, insbesondere aber auch bei methodischen Fragestellungen in Erwägung zu ziehen. Großforschungsinstrumente wie beispielsweise ALLBUS, NEPS oder SOEP ermöglichen Fragestellungen über einen engen Untersuchungszweck hinaus. Ferner lassen sich langfristige Entwicklungen ex post, wenn schon



Nachnutzen  
der Daten

nicht im Längsschnitt, so doch in mehrfachen Querschnitten auf der Basis vorhandener Daten analysieren. Sekundäranalysen verlangen allerdings sorgfältiger Recherche über die Inhalte und Methoden potenzieller Datensätze zur Abschätzung des „Data Fit“ zur Fragestellung und des damit verbundenen Nachnutzens.

Im *Datenarchiv der GESIS* stehen heute mehr als 6.500 Studien für Sekundäranalysen zur Verfügung. Die Datensätze können online<sup>10</sup> gesucht und Fragebögen, Variablenlisten, Codebücher etc. können eingesehen werden. Je nach Klassifikation (siehe ► Abbildung 3) seitens der Produzent:innen werden die Daten durch den Datenservice zur Verfügung gestellt. Sensible Daten, die besonderen Schutzes bedürfen, werden nur unter gesonderten vertraglichen Verpflichtungen weitergegeben (Off-Site-Zugang) oder stehen ausschließlich an einem Gastarbeitsplatz im sogenannten „Safe Room“ zur Verfügung (On-Site-Zugang). Verschiedene Zugangswege sind bei vielen FDZ vorzufinden. Ähnlich ist bei vielen FDZ auch die Unterstützung für Datengebende bei der Aufbereitung der Forschungsdaten zum Zweck der Nachnutzung verfügbar. Diese FDZ helfen bei Anonymisierung, Dokumentation und Klassifikation und sichern die langfristige Verfügbarkeit der Daten durch Archivierung.

### Abbildung 3: Benutzungsordnung des Datenarchivs von GESIS

Kategorie 0	Daten und Dokumente sind für alle freigegeben.
Kategorie A	Daten und Dokumente sind für die akademische Forschung und Lehre freigegeben.
Kategorie B	Daten und Dokumente sind für die akademische Forschung und Lehre freigegeben, wenn die Ergebnisse nicht veröffentlicht werden. Sollte eine Veröffentlichung oder eine weitergehende Verarbeitung der Ergebnisse geplant sein, ist eine Genehmigung über das Datenarchiv einzuholen.
Kategorie C	Daten und Dokumente sind für die akademische Forschung und Lehre nur nach schriftlicher Genehmigung der Datengebenden zugänglich. Das Datenarchiv holt dazu schriftlich die Genehmigung unter Angabe der Nutzenden und des Auswertungszweckes ein.

Quelle: *GESIS Datenarchiv (2007)*

Zentralisierung erwies sich angesichts der wachsenden Vielfalt spezifischer Forschungsfelder und Methoden dauerhaft als wenig effizientes Organisationsprinzip, zumal seit der Verfügbarkeit von Technologien zur Suche bei dezentraler Speicherung. Deren Nutzung empfiehlt insbesondere die „Open-Science-Bewegung“ (OSF) und dies nicht nur zur Bereitstellung von Texten und Befunden, sondern insbesondere auch von Erhebungsmaterialien, Laboraufzeichnungen etc. sowie Datensätzen (vgl. u. a. MPG, 2003). Die Kombination von Repositorien, Lizenzen (siehe Kapitel 3.4.3) und Suchmaschinen (siehe Kapitel 3.4.3) ermöglicht prinzipiell eine dezentrale Bereitstellung und Recherche von Datensätzen im Internet (BMBF, 2021). Ferner können Repositorien hinsichtlich der langfristigen Nutzbarkeit und Zitierbarkeit zertifiziert werden (vgl. DIN 31644:2012-04 und zugehörige Siegel wie „Nestor“ oder „CoreTrustSeal“). Im Unterschied zu Publikationen oder natur- und technikwissenschaftlichen Daten, deren Bereitstellung idealtypisch lediglich eine Beziehung zwischen Urheber:innen und Nutzenden impliziert, sind im Falle von sozial-, verhaltens- und wirtschaftswissenschaftlicher Forschungsdaten aber in der Regel auch die Informant:innen involviert und damit datenschutzrechtliche Aspekte zu beachten. Die damit verbundenen Rechte und Pflichten – von der Information bis zum Widerruf – verlangen Beratung und Betreuung und damit Aufgaben, die einer stark vorangetriebenen Rationalisierung widerstreben. Das heißt institutionelle Publikationsserver sind für die Verbreitung sozial-, verhaltens- und wirtschaftswissenschaftlicher Forschungsdaten nur eingeschränkt nutzbar.

Diese Disziplinen haben in den vergangenen Dekaden mit der *Ausdifferenzierung der FDZ* auf die gewachsene thematische und methodische Vielfalt reagiert. Die heutige Landschaft der FDZ umfasst spezialisierte Einrichtungen für bestimmte Datensätze, Disziplinen, Forschungsgegenstände oder

<sup>10</sup> unter <https://www.gesis.org/angebot/daten-finden-und-abrufen>

Forschungsmethoden. Aus Perspektive der Datennutzung bleibt die dezentrale Struktur der Archivierung allerdings (zunehmend) verborgen. Die Suche im erwähnten Portal der GESIS oder auch von KonsortSWD umfasst alle angeschlossenen FDZ und liefert auf Basis des vergebenen persistenten Identifikators eine zitierfähige Quelle, den Link bzw. Zugangsweg zum Datensatz sowie dessen Beschreibung.

Für den *europäischen Raum* bietet CESSDA European Research Infrastructure Consortium (ERIC) ein Metaportal zu den Datenarchiven der europäischen Mitgliedsländer. Die Recherchen bleiben im Ergebnis allerdings hinter den Funden in den jeweiligen nationalen Archiven bzw. Portalen zurück. Bei CESSDA sind jedoch sowohl die Links zu den Datenarchiven der Mitglieder, beispielsweise zu CESSDA-France/Réseau Quetelet (PROGEDO), UK Data Service oder zum Schweizer Kompetenzzentrum Sozialwissenschaften (FORS) hinterlegt als auch diejenigen zu den Nicht-Mitgliedern, beispielsweise zu UniData in Italien, Irish Social Science Data Archive (ISSDA) oder zur Datenbank des Centro de Investigaciones Sociológicas (CIS) in Spanien.

Für die *USA* soll auf das Inter-university Consortium for Political and Social Research (ICPSR) hingewiesen werden, dass nicht nur die universitätsübergreifende Datenbank sozial- und politikwissenschaftlicher Studien beheimatet, sondern darüber hinaus Informationen zu anderen Datenarchiven und deren Nutzungsmöglichkeiten enthält. Eine Übersicht – nicht nur von Datenarchiven, sondern auch anderer Statistiken und Datenquellen *weltweit* – wird von dem Social Science Information System (SOCIOSITE) an der Universität Amsterdam gepflegt.

Für die *Datenproduktion* stellt sich die Situation anders dar. Denn die Spezialisierungen der FDZ bedingen, dass eine passgenaue Unterstützung nur bei entsprechenden Datensätzen gewährleistet werden kann. Während ein allgemeines Survey-Projekt z.B. bei GESIS als generalistischer Einrichtung gut beraten ist, findet ein Mixed-Method-Vorhaben der Organisationsforschung im FDZ Betriebs- und Organisationsdaten (FDZ-BO) zweckmäßige Unterstützung. Werden vorwiegend Methoden der qualitativen Sozialforschung eingesetzt, bietet sich das FDZ Qualiservice an.

Die Debatte um „Open Access“ in der Wissenschaft fordert und die technologische Entwicklung ermöglicht die laufende Errichtung weiterer, auch den Forschungsprozess begleitender Speichermöglichkeiten. An größeren Universitäten sind in den vergangenen Jahren vermehrt Research Data Services eingerichtet worden, die beim FDM bis hin zur Datenpublikation und Nachnutzung unterstützen. Hochschulrepositorien sind daher, sofern vorhanden, als Alternative zu den außeruniversitären FDZ zu erwägen. Dies erleichtert nicht nur die vorgelagerten Schritte des FDM. Für kleinere Hochschulen, aber auch für etablierte Communities bleiben die zugehörigen FDZ und ihre disziplinäre Spezialisierung in Forschungsgegenständen und -methoden zentrale Ansprechpartner für die Archivierung.



### Nützliche Links zur **Nachnutzung von Forschungsdaten**

#### *National:*

- DSS (Data Services for the Social Sciences) (GESIS, o.J.).

#### *International:*

- DataCite (International Data Citation Initiative) (DataCite, o.J.).
- ICPSR (Inter-university Consortium for Political and Social Research) (ICPSR, o.J.-b).
- e3data.org (Registry of Research Data Repositories) (re3data.org, o.J.).
- UK Data Service (UK Data Service, o.J.-b).

## 4 Glossar

→ <b>Anonymisierung</b>	Das Verändern von personenbezogenen Daten, sodass sie entweder nicht mehr Personen zugeordnet werden können ( <i>vollständige</i> Anonymisierung, äußerst schwierig und selten) oder nur mit extrem hohem Aufwand zuordenbar sind ( <i>faktische</i> Anonymisierung, Normalfall bei Surveydaten).
→ <b>camelCase</b>	Die Worte werden aneinandergereiht. Dabei wird das erste Wort klein- und jedes darauffolgende Wort großgeschrieben.
→ <b>Data Sharing</b>	Hier das Teilen von Forschungsdaten mit Dritten, die an der Primärforschung nicht beteiligt sind/waren.
→ <b>Ethikvotum</b>	Die Beurteilung eines Forschungsantrags durch eine Ethikkommission mündet in einer schriftlichen Stellungnahme, die als Ethikvotum bezeichnet wird.
→ <b>Forschung, qualitativ</b>	Damit sind hier alle Forschungsansätze gemeint, die interpretativ und/oder rekonstruktiv mit qualitativen Daten arbeiten, gleichviel, welcher konkrete Forschungsansatz gewählt wird.
→ <b>Forschungsdatenzentrum (FDZ)</b>	Einrichtungen, die eigene und fremde Forschungsdaten archivieren, aufbereiten und fallweise für Sekundärnutzungen verfügbar machen. FDZ für die Sozial- und Wirtschaftswissenschaften werden zur Qualitätssicherung vom Rat für Sozial- und Wirtschaftsdaten akkreditiert.
→ <b>Hosten von Daten/Studien</b>	Das Aufbewahren u. ggf. Verfügbarmachen von Forschungsdaten oder ganzen Studien; Kernaufgabe von Forschungsdatenzentren.
→ <b>Informierte Einwilligung (Informed Consent)</b>	Als informierte Einwilligung (Informed Consent) bezeichnet die Einverständniserklärung, die zwischen Forschenden und den Teilnehmenden einer empirischen Studie abgeschlossen wird und in der genau bezeichnet ist, wozu sich die Teilnehmenden bereit erklären, wie die von ihnen gewonnenen Daten aufbereitet werden und was mit diesen Daten geschieht. Forschende sichern in diesem vertragsförmigen Dokument die Einhaltung relevanter, z. B. datenschutzrechtlich oder forschungsethisch gebotener Weisen des Umgang mit dem im Forschungsprozess gewonnenen Wissen und den entsprechenden Daten zu (z. B. strikte Anonymisierung, Aufbewahrungsfristen, ggf. konditionierte Weitergabe an Dritte).
→ <b>Kebab-case</b>	Die Wörter werden durch einen Bindestrich voneinander getrennt.
→ <b>Kontrolliertes Vokabular</b>	Wortschatz, der eindeutig zuordenbar ist.
→ <b>Kuratieren von Daten/ Studien</b>	Man versteht darunter das Organisieren bzw. Betreuen von Daten/Studien.
→ <b>Living Document</b>	Dokumente, die permanent fortgeschrieben und damit aktuell gehalten werden.
→ <b>Lizenz</b>	Lizenzen sind Standardverträge, die allen Nachnutzenden von Forschungsdaten bestimmte Handlungen erlauben. Über Lizenzen können Datengebende die Bedingungen der Nachnutzung festlegen und diese leicht verständlich darstellen.
→ <b>Metadatenstandards</b>	Metadaten sind Informationen über Daten. Im Rahmen von Metadatenstandards wird definiert, welche strukturierten Informationen über Daten vorzuliegen haben.

→ <b>Normdaten</b>	Bei Normdaten handelt es sich um „normierte Datensätze“, die eindeutig Entitäten zuordenbar sind. Die Zuordnung kann dann unabhängig von der Schreibweise erfolgen.
→ <b>PascalCase</b>	Alle Worte werden großgeschrieben und aneinandergereiht.
→ <b>Pseudonymisierung</b>	Pseudonymisierung nach Art. 4 Nr. 5 DSGVO ist „die Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden“.
→ <b>Repositorium/ Repositorien</b>	Verwaltete Orte zur Verwahrung von Dokumenten; hier: verwaltete Speicher von Forschungsdaten.
→ <b>Sekundärdaten</b>	Daten, die in anderen empirischen Studien gewonnen, aufbereitet und zur Nachnutzung verfügbar gemacht werden. Sekundärdaten unterscheiden sich von Primärdaten auch dadurch, dass sie nicht mit Blick auf die je konkrete Fragestellung eines aktuellen Projekts hin gewonnen wurden.
→ <b>Snake_case</b>	Die Wörter werden durch einen Unterstrich voneinander getrennt.
→ <b>Theoretical Sampling</b>	Theoretical Sampling ist eine im Rahmen der Grounded Theory vorgeschlagene Strategie der sukzessiven Entwicklung der Fallauswahl in qualitativ-empirischen Untersuchungen. Dabei richten sich die Sampling-Schritte und -kriterien nach Erfordernissen der im Forschungsverlauf schrittweise entwickelten gegenstandsbezogenen Theorie.
→ <b>Transkripte</b>	Wortgetreue Abschriften von Audio- oder Video-Dokumenten wie z.B. Interviewmitschnitte, Videos von Gruppendiskussionen oder von Interaktionen an öffentlichen Plätzen.
→ <b>Verpixelung</b>	Die teilweise Maskierung von digitalen Fotos bzw. Video-Frames, in der Regel zur Verhinderung der Identifizierung von Personen und Orten im Sinne der Anonymisierung.
→ <b>Verzeichnis für Verarbeitungstätigkeiten</b>	Ein Verzeichnis für Verarbeitungstätigkeiten wird nach Art. 30 Abs. 1 DSGVO immer notwendig, wenn personenbezogene Daten verarbeitet werden. Es informiert insbesondere darüber welche Kategorien von personenbezogenen Daten von welchen Personen/Einrichtungen zu welchem Zweck verarbeitet werden und welche Schutzmaßnahmen ergriffen werden. In der Regel gibt es an den wissenschaftlichen Einrichtungen oder bei den Landesdatenschutzbeauftragten Dateivorlagen, die zur Dokumentation der Datenschutzmaßnahmen verwendet werden können.
→ <b>Vignettendaten</b>	Daten, die im Rahmen von Vignettenanalysen (Factorial Surveys) gewonnen werden. Bei Vignettenanalysen werden Proband:innen in der Regel mehrere kurze fiktive Fallbeschreibungen vorgelegt, in denen einzelne Textbausteine (Dimensionen) experimentell variiert werden und zu denen eine Einschätzung der Proband:innen erhoben wird.

## 5 Literaturverzeichnis

- Argos. (o.J.). *Plan and follow your data*. <https://argos.openaire.eu/splash/> (Zugriff am 03.02.2023).
- BDSG (Bundesdatenschutzgesetz vom 30. Juni 2017 (BGBl. I S. 2097), das zuletzt durch Artikel 10 des Gesetzes vom 23. Juni 2021 (BGBl. I S. 1858; 2022 I 1045) geändert worden ist). (30.06.2017). [http://www.gesetze-im-internet.de/bdsg\\_2018/](http://www.gesetze-im-internet.de/bdsg_2018/)
- BERD@BW (Business, Economic and Related Data Baden-Württemberg). (Projektlaufzeit 2019-2022). *Interactive Virtual Assistant 1: Anwendungsbereich DSGVO*. <https://www.berd-nfdi.de/iva1/> (Zugriff am 03.02.2023).
- Bezjak, S., Clyburne-Sherin, A., Conzett, P., Fernandes, P., Görögh, E., Helbig, K., Kramer, B., Labastida, I., Niemeyer, K., Psomopoulos, F., Ross-Hellauer, T., Schneider, R., Tennant, J., Verbakel, E., Brinken, H. & Heller, L. (2018). *Open science training handbook*. <https://doi.org/10.5281/zenodo.1212496>
- Birke, P. & Mayer-Ahuja, N. (2017). Sekundäranalyse qualitativer Organisationsdaten. In S. Liebig, W. Matiaske & S. Rosenbohm (Hrsg.), *Springer Reference Wirtschaft. Handbuch Empirische Organisationsforschung* (S. 105–126). Springer Gabler. [https://doi.org/10.1007/978-3-658-08493-6\\_7](https://doi.org/10.1007/978-3-658-08493-6_7)
- BITSS (Berkeley Initiative for Transparency in the Social Sciences). (o.J.). *Resource library*. <https://www.bitss.org/resource-library/> (Zugriff am 03.02.2023).
- BMBF (Bundesministerium für Bildung und Forschung - Digitale Zukunft). (2021). *Lizenzvertrag und Offene Lizenzen*. <https://www.bildung-forschung.digital/digitalezukunft/de/wissen/urheberrecht/urheberrecht-in-der-wissenschaft/lizenzvertrag-und-offene-lizenzen/lizenzvertrag-und-offene-lizenzen.html> (Zugriff am 03.02.2023).
- Brettschneider, P., Axtmann, A., Böker, E. & Suchodoletz, D. von (2021). Offene Lizenzen für Forschungsdaten: Rechtliche Bewertung und Praxistauglichkeit verbreiteter Lizenzmodelle. *O-Bib. Das offene Bibliotheksjournal*, 8(3), 1–22. <https://doi.org/10.5282/O-BIB/5749>
- Brislinger, E. & Moschner, M. (2019). Datenaufbereitung und Dokumentation. In U. Jensen, S. Netscher & K. Weller (Hrsg.), *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten: Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten* (S. 97–114). Verlag Barbara Budrich. <https://doi.org/10.3224/84742233.07>
- CESSDA (Consortium of European Social Science Data Archives). (2019). *Adapt your data management plan: A list of data management questions based on the expert tour guide on data management*. [https://static-archive.CESSDA.eu/content/download/4302/48656/file/TTT\\_DO\\_DMPExpertGuide\\_v1.3.pdf](https://static-archive.CESSDA.eu/content/download/4302/48656/file/TTT_DO_DMPExpertGuide_v1.3.pdf)
- CESSDA and Service Providers. (2022). *The European Language Social Science Thesaurus (ELSST)* [Data set]. <https://doi.org/10.5281/Zenodo.7097110>
- CESSDA Training Team. (2020). *CESSDA data management expert guide*. Bergen. CESSDA ERIC. <https://doi.org/10.5281/zenodo.3820473>
- Christensen, G., Freese, J. & Miguel, E. (2019). *Transparent and reproducible social science research: How to do open science*. University of California Press.
- CNIL (National Commission on Informatics and Liberty). (2021). *PIA* [Computer software]. <https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assessment> (Zugriff am 03.02.2023).
- DARIAH-EU ELDAH (Digital Research Infrastructure for the Arts and Humanities, Ethics and Legality in Digital Arts and Humanities). (2021). *Consent Form Wizard (CFW)*. <https://consent.dariah.eu/node/2> (Zugriff am 03.02.2023).
- DataCite (International Data Citation Initiative). (o.J.). *Welcome to DataCite: Connecting research, advancing knowledge* [Homepage]. <https://datacite.org/index.html> (Zugriff am 03.02.2023).
- DataWiz Knowledge Base. (o.J.-a). *During data collection: Overview*. <https://datawizkb.leibniz-psychology.org/index.php/during-data-collection/> (Zugriff am 03.02.2023).
- DataWiz Knowledge Base. (o.J.-b). *Versioning*. <https://datawizkb.leibniz-psychology.org/index.php/during-data-collection/what-should-i-know-about-versioning/> (Zugriff am 03.02.2023).
- Deutsche Nationalbibliothek. (2022). *Gemeinsame Normdatei (GND)*. <https://www.dnb.de/gnd> (Zugriff am 03.02.2023).
- DFG - Fachkollegium 111 „Sozialwissenschaften“. (2020). *Stellungnahme des Fachkollegiums 111 Sozialwissenschaften zum Forschungsdatenmanagement in der Soziologie, der Politikwissenschaft und der Kommunikationswissenschaft*. [https://www.dfg.de/download/pdf/foerderung/grundlagen\\_dfg\\_foerderung/forschungsdaten/fachkollegium111\\_forschungsdatenmanagement.pdf](https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/fachkollegium111_forschungsdatenmanagement.pdf)

- DFG (Deutsche Forschungsgemeinschaft). (2021). *Umgang mit Forschungsdaten: Checkliste für Antragstellende zur Planung und zur Beschreibung des Umgangs mit Forschungsdaten in Forschungsvorhaben*. [https://www.dfg.de/download/pdf/foerderung/grundlagen\\_dfg\\_foerderung/forschungsdaten/forschungsdaten\\_checkliste\\_de.pdf](https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf)
- DFG (Deutsche Forschungsgemeinschaft). (2022a). *FAQ: Informationen aus den Geistes- und Sozialwissenschaften*. [https://www.dfg.de/foerderung/faq/geistes\\_sozialwissenschaften/](https://www.dfg.de/foerderung/faq/geistes_sozialwissenschaften/) (Zugriff am 03.02.2023).
- DFG (Deutsche Forschungsgemeinschaft). (2022b). *Forschungsdaten in den Förderprogrammen*. [https://www.dfg.de/foerderung/grundlagen\\_rahmenbedingungen/forschungsdaten/forschungsfoerderung/index.html](https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/forschungsdaten/forschungsfoerderung/index.html) (Zugriff am 03.02.2023).
- DGPs (Deutsche Gesellschaft für Psychologie). (2018). *Ethisches Handeln in der psychologischen Forschung: Empfehlungen der Deutschen Gesellschaft für Psychologie für Forschende und Ethikkommissionen* (1. Aufl.). Hogrefe.
- DIN 31644:2012-04. *Information und Dokumentation: Kriterien für vertrauenswürdige digitale Langzeitarchive*. Berlin. Beuth Verlag. <https://doi.org/10.31030/1854419>
- DMPonline. (o.J.). *Plan to make data work for you*. <https://dmponline.dcc.ac.uk/> (Zugriff am 03.02.2023).
- DMPTool. (o.J.). *Build your data management plan*. University of California. <https://dmptool.org/> (Zugriff am 03.02.2023).
- DSGVO (Europäische Datenschutz-Grundverordnung). (04.05.2016). *Amtsblatt der Europäischen Union (OJ L 119)*, 1–88. [http://eur-lex.europa.eu/legal-content/DE/TXT/?uri=OJ:JOL\\_2016\\_119\\_R\\_0001](http://eur-lex.europa.eu/legal-content/DE/TXT/?uri=OJ:JOL_2016_119_R_0001)
- Ebel, T. (2015). *Empfehlungen zur Anonymisierung quantitativer Daten*. GESIS - Leibniz-Institut für Sozialwissenschaften. [https://www.gesis.org/fileadmin/upload/institut/wiss\\_arbeitsbereiche/datenarchiv\\_analyse/Anonymisierung\\_quantitativer\\_Daten-0150512.pdf](https://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Anonymisierung_quantitativer_Daten-0150512.pdf)
- Ebel, T. & Meyermann, A. (2015). *Hinweise zur Anonymisierung von quantitativen Daten*. Version 1.2 (forschungsdaten bildung informiert Nr. 3). Leibniz-Institut für Bildungsforschung und Bildungsinformation (DIPF). <https://doi.org/10.25656/01:21970>
- Eisentraut, M. & Hutlesz, T. (2021). *Organisation und Speicherung von Forschungsdaten: Forschungsdatenmanagement in der empirischen Bildungsforschung* [Online-Seminar]. Verbund Forschungsdaten Bildung (VerbundFDB).
- European Commission. (o.J.). *Data management*. [https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management\\_en.htm](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm) (Zugriff am 03.02.2023).
- FDZ Qualiservice (Forschungsdatenzentrum Qualiservice). (o.J.). *Helpdesk: Downloads*. <https://www.qualiservice.org/de/helpdesk.html#downloads> (Zugriff am 03.02.2023).
- forschungsdaten.info. (2022a). *Der Datenlebenszyklus: Stationen des Forschungsdatenmanagements*. <https://www.forschungsdaten.info/themen/informieren-und-planen/datenlebenszyklus/> (Zugriff am 03.02.2023).
- forschungsdaten.info. (2022b). *Der Datenmanagementplan: Eine Wegbeschreibung für Daten*. <https://www.forschungsdaten.info/themen/informieren-und-planen/datenmanagementplan/> (Zugriff am 03.02.2023).
- FOSTER. (o.J.). *Resources: Open science*. <https://www.fosteropenscience.eu/resources> (Zugriff am 03.02.2023).
- Friedrichs, J. (2014). *Forschungsethik*. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (1. Aufl., S. 81–91). Springer VS. [https://doi.org/10.1007/978-3-531-18939-0\\_3](https://doi.org/10.1007/978-3-531-18939-0_3)
- GESIS (GESIS – Leibniz-Institut für Sozialwissenschaften). (o.J.). *Data Services for the Social Sciences (DSS)*. <https://www.gesis.org/institut/abteilungen/data-services-for-the-social-sciences> (Zugriff am 03.02.2023).
- GESIS Datenarchiv. (2007). *Benutzungsordnung: GESIS Abt. Datenarchiv und Datenanalyse*. [https://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/\\_bgordnung\\_bestellen/Benutzungsordnung.pdf](https://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/_bgordnung_bestellen/Benutzungsordnung.pdf)
- Gollwitzer, M., Adele-Brehm, A., Fiebach, C., Ramthun, R., Scheel, A., Schönbrodt, F. & Steinberg, U. (2020). *Management und Bereitstellung von Forschungsdaten in der Psychologie: Überarbeitung der DGPs-Empfehlungen*. <https://doi.org/10.31234/osf.io/hcxtm>

- Hammersley, M. (2009). Against the ethicists: On the evils of ethical regulation. *International Journal of Social Research Methodology*, 12(3), 211–225. <https://doi.org/10.1080/13645570802170288>
- Hardwicke, T., Bohn, M., MacDonald, K., Hembacher, E., Nuijten, M., Peloquin, B., deMayo, B., Long, B., Yoon, E. & Frank, M. (2021). Analytic reproducibility in articles receiving open data badges at the Journal Psychological Science: An observational study. *Royal Society Open Science*, 8(1). <https://doi.org/10.1098/rsos.201494>
- Heuer, J.-O., Kretzer, S., Mozygemba, K., Huber, E. & Hollstein, B. (2020). *Kontextualisierung qualitativer Forschungsdaten für die Nachnutzung: Eine Handreichung für Forschende zur Erstellung eines Studienreports* (Qualiservice Working Papers 1-2020). Bremen. <https://doi.org/10.26092/ELIB/166>
- Hopf, C. (2016). Forschungsethik und qualitative Forschung. In C. Hopf (Hrsg.), *Schriften zu Methodologie und Methoden qualitativer Sozialforschung* (S. 195–205). Springer VS. [https://doi.org/10.1007/978-3-658-11482-4\\_9](https://doi.org/10.1007/978-3-658-11482-4_9)
- ICPSR (Inter-university Consortium for Political and Social Research). (o.J.-a). *Elements of a data management plan*. <https://www.icpsr.umich.edu/web/pages/datamanagement/dmp/elements.html> (Zugriff am 03.02.2023).
- ICPSR (Inter-university Consortium for Political and Social Research). (o.J.-b). *This is ICPSR* [Homepage]. <https://www.icpsr.umich.edu/web/pages/> (Zugriff am 03.02.2023).
- Long, J. (2009). *The workflow of data analysis using Stata* (1. Aufl.). Stata Press.
- Martin, N., Schiering, I., Mester, B., Hallinan, D. & Jensen, M. (2020). *Die Datenschutz-Folgenabschätzung nach Art. 35 DSGVO: Ein Handbuch für die Praxis*. Fraunhofer Verlag. <https://doi.org/10.24406/publica-fhg-300193>
- Mayer, E. & Dalhoff D. (1998). Ein Gespräch mit Erwin K. Scheuch. *ZA-Information*(42), 13–26. [https://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/za\\_information/ZA-Info-42.pdf](https://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/za_information/ZA-Info-42.pdf)
- Medjedović, I. (2014). *Qualitative Sekundäranalyse: Zum Potenzial einer neuen Forschungsstrategie in der empirischen Sozialforschung* (1. Aufl.). Springer VS. <https://doi.org/10.1007/978-3-658-05488-5>
- Meyermann, A. & Porzelt, M. (2019). *Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung: Eine Handreichung*. Version 2.1 (forschungsdaten bildung informiert Nr. 6). Leibniz-Institut für Bildungsforschung und Bildungsinformation (DIPF). <https://doi.org/10.25656/01:21990>
- Mochmann, E. (2014). Quantitative Daten für die Sekundäranalyse. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (1. Aufl., S. 233–244). Springer VS. [https://doi.org/10.1007/978-3-531-18939-0\\_14](https://doi.org/10.1007/978-3-531-18939-0_14)
- MPG (Max-Planck-Gesellschaft). (2003). *Berlin declaration on open access to knowledge in the sciences and humanities*. <https://openaccess.mpg.de/Berlin-Declaration> (Zugriff am 03.02.2023).
- Netscher, S. (2020). *It's not such 'A Fair Way Off' to process open data: Facing requirements on open access and the FAIR data principles* (GESIS Blog: Growing knowledge in the social sciences). <https://doi.org/10.34879/GESISBLOG.2020.3>
- Nicolai, T., Mozygemba, K. & Hollstein, B. (2021). *QualiAnon: Qualiservice tool for anonymizing text data* [Computer software]. FDZ Qualiservice. Universität Bremen [für Informationen zur Registrierung/Freischaltung siehe: <https://www.qualiservice.org/de/helpdesk/webinar/tools.html>] (Zugriff am 03.02.2023).
- Open Science Future. (o.J.). *Was Forschungsförderer in Sachen Forschungsdatenmanagement fordern: Und was bei Drittmittelanträgen zu beachten ist*. Leibniz-Informationszentrum Wirtschaft (ZBW). <https://open-science-future.zbw.eu/was-forschungsfoerderer-in-forschungsdatenmanagement-fordern/> (Zugriff am 03.02.2023).
- OpenAIRE. (2022). *Guides for researchers: How to find a trustworthy repository for your data*. <https://www.openaire.eu/find-trustworthy-data-repository> (Zugriff am 03.02.2023).
- OSC - LMU (Open Science Center - Ludwig-Maximilians-Universität München). (o.J.). *Toolbox*. <https://www.osc.uni-muenchen.de/toolbox/index.html> (Zugriff am 03.02.2023).
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.
- Pergl, R., Hooft, R., Suchánek, M., Knaisl, V. & Slifka, J. (2019). "Data Stewardship Wizard": A tool bringing together researchers, data stewards, and data experts around data management planning. *Data Science Journal*, 18(1), Artikel 59, 1–8. <https://doi.org/10.5334/dsj-2019-059>
- Project TIER. (o.J.). *Teaching Integrity in Empirical Research (TIER)* [Homepage]. <https://www.projecttier.org> (Zugriff am 03.02.2023).

- RatSWD (Rat für Sozial- und Wirtschaftsdaten). (o.J.-a). *Best-Practice Forschungsethik: Best-Practice-Sammlung zur Forschungsethik im Forschungsprozess*. <https://www.konsortswd.de/ratswd/best-practice-forschungsethik/> (Zugriff am 03.02.2023).
- RatSWD (Rat für Sozial- und Wirtschaftsdaten). (o.J.-b). *Datenschutz: Hinweise zum Datenschutz in der Forschung*. <https://www.konsortswd.de/ratswd/best-practice-forschungsethik/datenschutz-forschungsethik/> (Zugriff am 03.02.2023).
- RatSWD (Rat für Sozial- und Wirtschaftsdaten). (2016). *Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften: Orientierungshilfen für die Beantragung und Begutachtung datengenerierender und datennutzender Forschungsprojekte* (RatSWD Output Series, 5. Berufungsperiode Nr. 3). Berlin. <https://doi.org/10.17620/02671.7>
- RatSWD (Rat für Sozial- und Wirtschaftsdaten). (2017). *Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften* (RatSWD Output Series, 5. Berufungsperiode Nr. 9). Berlin. <https://doi.org/10.17620/02671.1>
- RatSWD (Rat für Sozial- und Wirtschaftsdaten). (2020). *Handreichung Datenschutz: 2. vollständig überarbeitete Auflage*. Überarbeitet von Prof. Dr. Matthias Bäcker und Dr. Sebastian Golla in Zusammenarbeit mit dem RatSWD (RatSWD Output Series, 6. Berufungsperiode Nr. 8). Berlin. <https://doi.org/10.17620/02671.50>
- RDMO Arbeitsgemeinschaft. (o.J.). *Research Data Management Organiser (RDMO)* [Computer software]. <https://rdmorganiser.github.io/>
- re3data.org (Registry of Research Data Repositories). (o.J.). *re3data.org: Registry of Research Data Repositories* [Homepage]. <https://www.re3data.org/> (Zugriff am 03.02.2023).
- Recker, J. & Brislinger, E. (2019). Dateiorganisation in empirischen Forschungsprojekten. In U. Jensen, S. Netscher & K. Weller (Hrsg.), *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten: Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten* (S. 81–95). Verlag Barbara Budrich. <https://doi.org/10.3224/84742233.06>
- Schallaböck, J., Hoffstätter, U., Buck, D. & Linne, M. (2022). *Mustervertrag Datennutzung KonsortSWD (1.0.0)*. <https://doi.org/10.5281/zenodo.5828114>
- Science Europe. (2021). *Practical guide to the international alignment of research data management: Extended edition*. <https://doi.org/10.5281/Zenodo.4915862>
- Steinhardt, I., Fischer, C., Heimstädt, M., Hirsbrunner, S., Ikiz-Akinci, D., Kressin, L., Kretzer, S., Möllenkamp, A., Porzelt, M., Rahal, R.-M., Schimmler, S., Wilke, R. & Wünsche, H. (2020). *Das Öffnen und Teilen von Daten qualitativer Forschung: Eine Handreichung* (Weizenbaum Series Nr. 6). Berlin. Weizenbaum Institute for the Networked Society. <https://doi.org/10.34669/WI.WS/6>
- Stodden, V., Seiler, J. & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11), 2584–2589. <https://doi.org/10.1073/pnas.1708290115>
- Strübing, J. (2018). *Qualitative Sozialforschung: Eine komprimierte Einführung* (2. Aufl.). De Gruyter Oldenbourg. <https://doi.org/10.1515/9783110529920>
- The Turing Way Community. (2021). *The Turing Way: A handbook for reproducible, ethical and collaborative research (1.0.1)*. <https://doi.org/10.5281/zenodo.5671094>
- UK Data Service. (o.J.-a). *Learning hub: Research data management*. <https://ukdataservice.ac.uk/learning-hub/research-data-management/> (Zugriff am 03.02.2023).
- UK Data Service. (o.J.-b). *Welcome to the UK Data Service* [Homepage]. <https://ukdataservice.ac.uk/> (Zugriff am 03.02.2023).
- van den Eynden, V., Corti, L., Woollard, M., Bishop, L. & Horton, L. (2011). *Managing and sharing data: Best practice for researchers*. Colchester. UK Data Archive, University of Essex. <https://dam.ukdataservice.ac.uk/media/622417/managingsharing.pdf>
- VerbundFDB (Verbund Forschungsdaten Bildung). (o.J.). *Daten managen*. Leibniz-Institut für Bildungsforschung und Bildungsinformation (DIPF). <https://www.forschungsdaten-bildung.de/daten-managen> (Zugriff am 03.02.2023).
- VerbundFDB (Verbund Forschungsdaten Bildung). (2015). *Checkliste zur Erstellung eines Datenmanagementplans in der empirischen Bildungsforschung*. Version 1.1 (forschungsdaten bildung informiert Nr. 2). Leibniz-Institut für Bildungsforschung und Bildungsinformation (DIPF). <https://doi.org/10.25656/01:22298>

- VerbundFDB (Verbund Forschungsdaten Bildung). (2018a). *Formulierungsbeispiele für „informierte Einwilligungen“*. Version 2.1 (forschungsdaten bildung informiert Nr. 4). Leibniz-Institut für Bildungsforschung und Bildungsinformation (DIPF). <https://doi.org/10.25656/01:22301>
- VerbundFDB (Verbund Forschungsdaten Bildung). (2018b). *Formulierungsbeispiele für „informierte Einwilligungen“ in leichter Sprache*. Version 1.1 (forschungsdaten bildung informiert Nr. 5). Leibniz-Institut für Bildungsforschung und Bildungsinformation (DIPF). <https://doi.org/10.25656/01:22302>
- VerbundFDB (Verbund Forschungsdaten Bildung). (2019). *Checkliste zur Erstellung rechtskonformer Einwilligungserklärungen mit besonderer Berücksichtigung von Erhebungen an Schulen*. Version 2.0 (forschungsdaten bildung informiert Nr. 1). Leibniz-Institut für Bildungsforschung und Bildungsinformation (DIPF). <https://doi.org/10.25656/01:22297>
- VerbundFDB (Verbund Forschungsdaten Bildung). (2021). *Leitfaden zur Kontextualisierung von qualitativen Befragungsdaten*. Version 1.2 (forschungsdaten bildung informiert Nr. 9). Leibniz-Institut für Bildungsforschung und Bildungsinformation (DIPF). <https://doi.org/10.25656/01:22386>
- VerbundFDB (Verbund Forschungsdaten Bildung). (2022a). *Datenschutzrechtliche Aspekte*. Leibniz-Institut für Bildungsforschung und Bildungsinformation (DIPF). <https://www.forschungsdaten-bildung.de/datenschutzrechtliche-aspekte> (Zugriff am 03.02.2023).
- VerbundFDB (Verbund Forschungsdaten Bildung). (2022b). *Stamp nutzen: Standardisierter Datenmanagementplan für die Bildungsforschung*. Leibniz-Institut für Bildungsforschung und Bildungsinformation (DIPF). <https://www.forschungsdaten-bildung.de/stamp-nutzen> (Zugriff am 03.02.2023).
- Wilkinson, M. D., Dumontier, M. & Aalbersberg, I. et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(160018). <https://doi.org/10.1038/sdata.2016.18>
- ZBW (Leibniz-Informationszentrum Wirtschaft), GESIS (GESIS - Leibniz-Institut für Sozialwissenschaften) & RatSWD (Rat für Sozial- und Wirtschaftsdaten). (2014). *Auffinden, zitieren, dokumentieren: Forschungsdaten in den Sozial- und Wirtschaftswissenschaften*. Hamburg. <https://doi.org/10.4232/10.fisuzida2015.2>

## Mitwirkende bei der Erstellung

### Mitglieder der AG FDM Services für kleinere Forschungsprojekte: Wegweiser für den Kulturwandel

**Prof. Dr. Hansjörg Dilger**

Freie Universität Berlin

**Prof. Dr. Michael Eid**

Freie Universität Berlin, RatSWD

**Dr. Anne Gresser**

Julius-Maximilians-Universität Würzburg

**Prof. Dr. Christiane Gross**, *Co-Vorsitz*

Julius-Maximilians-Universität Würzburg, RatSWD

**Dr. Jan-Ocko Heuer**

Universität Bremen

**Prof. Dr. Oliver Lüdtke**

Leibniz-Institut für Pädagogik der Naturwissenschaften und Mathematik (IPN),  
Christian-Albrechts-Universität zu Kiel, RatSWD

**Prof. Dr. Wenzel Matiaske**

Helmut-Schmidt-Universität/Universität der Bundeswehr Hamburg

**Prof. Dr. Laura Seelkopf**

Ludwig-Maximilians-Universität München, RatSWD

**Prof. Dr. Jörg Strübing**, *Co-Vorsitz*

Eberhard Karls Universität Tübingen, RatSWD

### Geschäftsstelle des RatSWD

Marie Eilers







# Impressum

**Herausgeber:**

Rat für Sozial- und Wirtschaftsdaten (RatSWD)  
Geschäftsstelle  
Am Friedrichshain 22  
10407 Berlin  
office@ratswd.de  
<https://www.ratswd.de>

**Redaktion:**

Marie Eilers

**Gestaltung/Satz:**

Claudia Kreuz

Berlin, März 2023

**RatSWD Output:**

Die RatSWD Output Series dokumentiert die Arbeit des RatSWD in seiner 7. Berufungsperiode (2020–2023). In ihr werden seine Stellungnahmen und Empfehlungen veröffentlicht und auf diesem Weg einer breiten Leserschaft zugänglich gemacht.

Die Geschäftsstelle des RatSWD wird als Teil von KonsortSWD im Rahmen der NFDI durch die Deutsche Forschungsgemeinschaft (DFG) gefördert - Projektnummer: 442494171.



Diese Veröffentlichung ist unter der Creative-Commons-Lizenz (CC BY 4.0) lizenziert:

<https://creativecommons.org/licenses/by/4.0/>

Ausgenommen von der oben genannten Lizenz sind Teile, Abbildungen und sonstiges Drittmaterial, wenn anders gekennzeichnet, sowie die Bilder auf Seite 1 und 48 (Deemerwha Studio – stock.adobe.com).

**DOI:** 10.17620/02671.72

**Zitationsvorschlag:**

RatSWD (Rat für Sozial- und Wirtschaftsdaten) (2023). *Forschungsdatenmanagement in kleinen Forschungsprojekten - Eine Handreichung für die Praxis*. (RatSWD Output Series, 7. Berufungsperiode Nr. 3). <https://doi.org/10.17620/02671.72>

■ **Der Rat für Sozial- und Wirtschaftsdaten (RatSWD)** berät seit 2004 die Bundesregierung und die Regierungen der Länder in Fragen der Forschungsdateninfrastruktur für die empirischen Sozial-, Verhaltens- und Wirtschaftswissenschaften. Im RatSWD arbeiten zehn durch Wahl legitimierte Vertreterinnen und Vertreter der sozial-, verhaltens- und wirtschaftswissenschaftlichen Fachdisziplinen mit zehn Vertreterinnen und Vertretern der wichtigsten Datenproduzenten zusammen.

Der RatSWD ist Teil des Konsortiums für die Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften (KonsortSWD) in der Nationalen Forschungsdateninfrastruktur (NFDI). Er versteht sich als institutionalisiertes Forum des Dialoges zwischen Wissenschaft und Datenproduzenten und erarbeitet Empfehlungen und Stellungnahmen. Dabei engagiert er sich für eine Infrastruktur, die der Wissenschaft einen breiten, flexiblen und sicheren Datenzugang ermöglicht. Diese Daten werden von staatlichen, wissenschaftsgetragenen und privatwirtschaftlichen Akteuren bereitgestellt. Derzeit hat der RatSWD 42 Forschungsdatenzentren akkreditiert und fördert deren Kooperation.

