

Handreichung: Forschungsdatenzentren gründen

Buck, Daniel; Croisier, Johannes; Eder, Christina; Hoffstätter, Ute; Jansen, Malte; Meyermann, Alexia; Siegers, Pascal

Veröffentlichungsversion / Published Version

Arbeitspapier / working paper

Empfohlene Zitierung / Suggested Citation:

Buck, D., Croisier, J., Eder, C., Hoffstätter, U., Jansen, M., Meyermann, A., Siegers, P. (2022). *Handreichung: Forschungsdatenzentren gründen*. (RatSWD Working Paper Series, 280). Berlin: Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.70>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

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

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:

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

RatSWD Working Paper Series

RatSWD ■
Rat für Sozial- und
WirtschaftsDaten

280

Handreichung:
Forschungsdatenzentren
gründen.

Daniel Buck, Johannes Croisier,
Christina Eder, Ute Hoffstätter,
Malte Jansen, Alexia Meyermann,
Pascal Siegers

Juli 2022

www.ratswd.de

Working Paper Series

des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

Die RatSWD Working Papers-Reihe startete Ende 2007. In dieser Online-Publikationsreihe werden konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

RatSWD Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen.

Die Inhalte der RatSWD Working Papers stellen ausdrücklich die Meinung der jeweiligen Autorinnen bzw. Autoren dar und nicht die des RatSWD. Die Zuwendungsgeber des RatSWD haben die Publikationen nicht beeinflusst.

Herausgeberin oder Herausgeber der RatSWD Working Papers-Reihe ist die/der Vorsitzende des RatSWD:

seit 2020 Monika Jungbauer-Gans

2014–2020 Regina T. Riphahn

2009–2014 Gert G. Wagner

2007–2008 Heike Solga

Handreichung: Forschungsdatenzentren gründen.

Autoren (alphabetisch): Daniel Buck, Johannes Croisier,
Christina Eder, Ute Hoffstätter, Malte Jansen,
Alexia Meyermann, Pascal Siegers

Verbessert durch Kommentare von: Benjamin Fuchs, Tina Hinz,
Dana Müller, Tobias Roth, Stefan Schweers.

Juli 2022

<https://doi.org/10.17620/02671.70>

Inhalt

1	Sicher und FAIR: Forschungsdatenzentren für die Wissenschaft.....	5
2	Datenrepositorium oder FDZ?.....	6
2.1	Datenrepositorium.....	7
2.2	Forschungsdatenzentren.....	8
3	Grundfunktionen von Forschungsdatenzentren: Aufbereitung & Dokumentation, Langzeitsicherung und Datenzugang.....	10
3.1	Aufbereitung und Dokumentation.....	11
3.2	Langfristsicherung.....	12
3.3	Datenbereitstellung und Datenzugänge.....	13
3.4	Aufnahme externer Daten.....	14
3.5	Weitere Aufgaben.....	14
4	Personalbedarf und Personalstrukturen in Forschungsdatenzentren.....	15
5	Benötigte IT-Infrastrukturen für den Betrieb eines FDZ.....	17
5.1	Hardware.....	18
5.2	Software.....	18
5.3	IT-Sicherheit.....	19
6	Rechtliche Grundlagen und Verträge.....	19
6.1	Rechtliche Rahmenbedingungen.....	19
6.2	Vertragswerke und Nutzungsbedingungen.....	21
7	Erfahrungen aus der Praxis von FDZ-Neugründungen.....	22
7.1	Aufbauphase und Anschubfinanzierung.....	22
7.2	Laufender Betrieb.....	24
7.3	Entwicklung: Innovation und Angebotserweiterung.....	24
8	Akkreditierung durch den RatSWD.....	25
	Literaturverzeichnis.....	27

Zusammenfassung:

Die Verbesserung des Zugangs zur Forschungsdaten ist derzeit ein zentrales wissenschaftspolitisches Thema. Forschungsdatenzentren (FDZ) haben sich in den vergangenen 20 Jahren als Modell für einen sicheren Datenzugang etabliert. In ihrer Datenstrategie sieht die Bundesregierung deshalb die Stärkung der bestehenden Forschungsdatenzentren und die Gründung neuer Forschungsdatenzentren vor.

Diese Handreichung soll einen ersten Einblick in Aufgaben und Strukturen von Forschungsdatenzentren geben. Zwar unterscheiden sich Forschungsdatenzentren in Abhängigkeit von Umfang und Art des Datenangebotes und der Nutzungsfrequenz stark voneinander, bestimmte Aufgaben muss jedoch jedes Forschungsdatenzentrum erfüllen, um einen nutzungsfreundlichen Datenzugang anbieten zu können. Diese Handreichung fasst daher die Erfahrungen von 41 Forschungsdatenzentren zusammen, die sich seit 2001 durch den Rat für Sozial- und Wirtschaftsdaten haben akkreditieren lassen.

Forschungsdatenzentren sind vor allem für die Einrichtung eines Zugangs zu schützenswerten Daten notwendig, andere Daten können auch über eines der zahlreichen Repositorien für Forschungsdaten veröffentlicht werden (vgl. Abschnitt 2). Allerdings ist die Bereitstellung von Daten immer mit der Notwendigkeit der Aufbereitung, Dokumentation und langfristigen Sicherung der Daten verbunden (vgl. Abschnitt 3). Dafür wird ausreichend und qualifiziertes Personal benötigt, das an der Schnittstelle zwischen Fachwissenschaft und Informationsinfrastrukturen arbeitet (vgl. Abschnitt 4). Die Leistungsfähigkeit des Forschungsdatenzentrums hängt entscheidend von der Qualität der IT-Infrastruktur ab (vgl. Abschnitt 5). Bei der Bereitstellung von schützenswerten Daten und im operativen Betrieb des Forschungsdatenzentrums müssen schließlich zahlreiche rechtliche Fragen berücksichtigt werden (vgl. Abschnitt 6).

Die Handreichung schließt mit einem kurzen Einblick in den Prozess der Gründung eines Datenzentrums (vgl. Abschnitt 7) und die Akkreditierung durch den Rat für Sozial- und Wirtschaftsdaten (vgl. Abschnitt 8).

1 Sicher und FAIR: Forschungsdatenzentren für die Wissenschaft¹

“Daten bilden die Grundlage der digitalen Gesellschaft. Mehr Daten innovativ, verantwortungsvoll und gemeinwohlorientiert zu nutzen, kann das Zusammenleben in Deutschland, in Europa und in der Welt bedeutsam verbessern und natürliche Ressourcen schützen“ (Bundeskanzleramt 2021, S. 5).

Daten dienen als Entscheidungsgrundlage in Politik, Verwaltung und Wirtschaft. Sie sind Währung und Geschäftsmodell in der digitalen Welt und die Grundlage für wissenschaftliche Erkenntnis und Innovation. Die Forderung nach einem offenen Zugang (Open Data) zu Daten aus Wissenschaft, Verwaltung und Wirtschaft spiegelt diese Bedeutung wider. Open Data bezeichnet die Idee, die Weiter- und Wiederverwendung von administrativen, wirtschaftlichen und wissenschaftlichen Daten zu wissenschaftlichen und wirtschaftlichen Zwecken durch einen möglichst uneingeschränkten Zugang zu ermöglichen. Dafür müssen Daten frei, also ohne Registrierung und Bezahlung, verfügbar sein. Sie müssen für Menschen und Maschinen lesbar und mit entsprechenden Zusatzinformationen (Metadaten - Daten über Daten) und einer transparenten Lizenz versehen sein, die deutlich macht, was mit diesen Daten getan werden darf und was nicht (Lauber-Rönsberg 2021; Mauer & Recker 2019: 120ff.).

Hier überschneiden sich die Forderungen der Open-Data-Idee mit den Zielen der FAIR-Initiative², die bestrebt ist, Daten auffindbar (findable), leicht zugänglich (accessible), automatisiert zusammenführbar (interoperable) und durch ausreichende Aufbereitung und Dokumentation nachnutzbar (reusable) zu machen (Wilkinson et al. 2016). Die Einhaltung der FAIR-Prinzipien ermöglicht eine Wiederverwertung in neuen Nutzungskontexten. Redundante Datenerhebungen werden so vermieden und der wissenschaftliche und wirtschaftliche Wert existierender Daten besser ausgeschöpft.

Allerdings sind nicht alle Daten dafür geeignet, offen zur Verfügung gestellt zu werden. Dies betrifft alle Daten, für die das Datenschutzrecht angewendet werden muss, die der Vertraulichkeit unterliegen oder deren Nutzung gesetzlich auf bestimmte Nutzungszwecke eingeschränkt ist. Dennoch können die FAIR-Prinzipien auch für solche schutzbedürftigen Daten umgesetzt werden, indem die Daten qualitativ hochwertig dokumentiert und transparente Regelungen für den Zugang zu Daten und Metadaten etabliert werden.

Mit der aktuellen Datenstrategie geht die Bundesregierung den Weg hin zu möglichst offenen, in jedem Fall aber FAIRen Daten einen Schritt weiter und fordert alle Ministerien dazu auf, in ihrem Geschäftsbereich Forschungsdatenzentren (FDZ) zu gründen, um den Zugang zu

¹ Die Erstellung dieser Handreichung wurde gefördert durch die Deutsche Forschungsgemeinschaft als Teil der NFDI mit der Projektnummer 442494171.

² <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/1.-Plan/FAIR-data>.

amtlichen Daten und Forschungsdaten aus den Ministerien und Ressortforschungseinrichtungen für die wissenschaftliche Nachnutzung zu verbessern. Auch die FDZ in der Wissenschaft sollen von Bund und Ländern gemeinsam gestärkt werden (Bundeskanzleramt 2021, S. 36).

Viele Akteure in Verwaltung und Wissenschaft werden mit der Frage konfrontiert, wie die in der eigenen Organisation erzeugten Daten im Einklang mit den gesetzlichen Bestimmungen für die wissenschaftliche Forschung und wirtschaftliche Verwertung bereitgestellt werden können. Im Bereich der Wirtschafts- und Sozialdaten hat sich das Modell der Forschungsdatenzentren etabliert, um den Datenzugang für die Wissenschaft zu organisieren. Der Rat für Sozial- und Wirtschaftsdaten (RatSWD) hat Kriterien für den Betrieb eines FDZ entwickelt und bietet eine Akkreditierung an. Seit der Gründung des RatSWD wurden 41 FDZ (Stand: Juni 2022) akkreditiert. Sie ermöglichen den Zugang zu Daten aus öffentlichen Verwaltungen, Sozialversicherungen und akademischen Forschungsinstituten. Der Datenbestand der FDZ deckt alle Bereiche der Sozial-, Wirtschafts- und Verhaltenswissenschaften sowie angrenzende Fachdisziplinen ab (z. B. die oder die Sprachwissenschaften).³ Die bestehenden FDZ können als Blaupause für die Gründung neuer FDZ dienen. Zwar sind die einzelnen FDZ aufgrund der Vielfalt der Datentypen und der rechtlichen Grundlagen der Datenbereitstellung unterschiedlich organisiert, bestimmte Aufgaben und Strukturen sind jedoch bei allen FDZ zu finden.

Dieses Papier konzentriert sich auf die Darstellung dieser Grundbausteine, die für den Aufbau und Betrieb eines FDZ benötigt werden. Es soll eine erste Orientierung bieten und eine Einführung in die Aufgaben und die Funktionsweise von FDZ – auch im Unterschied zu Repositorien für Forschungsdaten – geben.

2 Datenrepositorium oder FDZ?

Wie voranstehend beschrieben, fordert die Bundesregierung alle Ministerien dazu auf, in ihrem Geschäftsbereich FDZ zu gründen, um den Zugang zu amtlichen Daten und Forschungsdaten aus den Ministerien und Ressortforschungseinrichtungen für die wissenschaftliche Nachnutzung zu verbessern. Neben der Gründung eines eigenen FDZ besteht jedoch auch die Möglichkeit, Daten der eigenen Organisation über bestehende Repositorien oder FDZ zugänglich zu machen. Um entscheiden zu können, welches der geeignete Weg ist, gilt es zunächst, den eigenen Bestand an Daten zu sichten: Um welche Daten handelt es sich? Und in welcher Form können oder sollen diese zugänglich gemacht werden?

³ Eine ständig aktualisierte Übersicht findet sich auf den Webseiten von KonsortSWD:
<https://www.konsortswd.de/datenzentren/>

2.1 Datenrepositorium

Liegen keine rechtlichen oder ethischen Gründe für die Einschränkung des Datenzugangs vor, sollten Daten unter einer freien Nutzungslizenz über ein existierendes Datenrepositorium angeboten werden. Ein Repositorium funktioniert wie ein Bibliothekskatalog, in dem die wichtigsten Merkmale der Daten (Metadaten) beschrieben werden. Darüber hinaus ist das Herunterladen von Daten und Dokumenten möglich, wobei unter Umständen eine Registrierung erforderlich ist.

Mehrere bekannte Repositorien für Forschungsdaten kommen für die Publikation von Daten in Frage:

- (1) Zenodo (<https://zenodo.org/>): disziplinübergreifendes Repositorium für jede Art von wissenschaftlichen Datensätzen, wie z. B. auch für Publikationen, Quellcode oder Präsentationen. Die Nutzung ist kostenfrei.
- (2) Harvard Dataverse (<https://dataverse.harvard.edu/>): disziplinübergreifendes Repositorium für wissenschaftliche Forschungsdaten und Auswertungscode, das vor allem zu Replikationszwecken von internationalen Zeitschriften genutzt wird. Die Nutzung ist kostenfrei.
- (3) Radar (<https://www.radar-service.eu/de/home>): disziplinübergreifendes Repositorium für wissenschaftliche Forschungsdaten des FIZ Karlsruhe.
- (4) GESIS Datenservices (<https://www.gesis.org/datenservices/daten-teilen>): Repositorium für Daten aus den Sozialwissenschaften. Die Nutzung der Basisdienste ist kostenfrei.

Eine Suche nach einschlägigen Forschungsdatenrepositorien und FDZ aus allen Fachdisziplinen ist bei <https://www.re3data.org/search> möglich. Über Datentyp, Fachdomäne oder Land können geeignete Repositorien gefunden werden. Science Europe hat in der jüngeren Vergangenheit Minimal Kriterien für die Auswahl von Repositorien für Forschungsdaten entwickelt, die hilfreiche Anhaltspunkte für die Auswahl auflisten (Science Europe 2021).⁴ Wird hingegen ein Weg gesucht, um regelmäßig Daten für die Nachnutzung bereitzustellen, stellen die FAIR-Prinzipien Leitlinien für die Gestaltung von Forschungsdateninfrastrukturen dar (Betancort et al. 2020):

- (1) fachspezifische Repositorien mit angepassten Metadatenstandards sind generischen Repositorien vorzuziehen, weil diese in der Regel eine bessere Metadaten- und Dokumentationsqualität anbieten. Die Datensuche in generischen Repositorien ist immer noch schwierig, weil die Inhalte der Daten oft nicht ausreichend beschrieben werden.

⁴ https://www.scienceeurope.org/media/4brkxxe5/se_rdm_practical_guide_extended_final.pdf

- (2) Repositorien sollten für die digitale Langzeitarchivierung zertifiziert sein. Die Zertifizierung garantiert die Einhaltung grundlegender Qualitätsstandards für die digitale Langzeitsicherung von Daten. Beispiele für verbreitete Zertifikate sind das CoreTrustSeal⁵ und das nestor Siegel⁶.
- (3) Repositorien sollten die Daten sichtbar machen, indem (i) die Metadaten unter einer freien Nutzungslizenz für die Nutzung durch Suchmaschinen wie z.B. die GoogleDatasetSearch angeboten werden, (ii) die Metadaten nach Standards des W3C-Consortiums für allgemeine Suchmaschinen aufbereitet werden und (iii) eine Schnittstelle für die automatische Übernahme der Metadaten existiert (bevorzugt eine OAI-PMH Schnittstelle).
- (4) Die Daten müssen mit Persistenten Identifikatoren (z. B. Digital Object Identifiers, DOI) versehen werden, was die Datenzitation erleichtert und die Sichtbarkeit der Daten erhöht.
- (5) Die Daten sollten unter einer klaren "Lizenz" nutzbar sein, die transparent regelt, wer was mit den Daten tun darf und welche Nutzungszwecke oder Operationen untersagt sind (Mauer & Recker 2019).

Wenn größere Mengen Forschungsdaten innerhalb einer Organisation anfallen, kann auch der Betrieb eines eigenen Repositoriums für Forschungsdaten sinnvoll sein, sofern dies die oben genannten Kriterien erfüllt. So bietet Radar (siehe oben) Repositoriumslösungen *as a service* für Institutionen an und *Dataverse* ist eine Open-Source Software, die von jeder Organisation genutzt werden kann, wenn die für den Betrieb notwendige IT-Kompetenz im Haus vorhanden ist. Für sozialwissenschaftliche Forschungsdaten bietet GESIS den Dienst *SowiDataNet* an, um das Datenteilen zu vereinfachen, weil kein eigenes lokales Repositorium betrieben werden muss.

Auch Daten, die in Repositorien publiziert werden, müssen aufbereitet und dokumentiert werden, um nachnutzbar zu sein. Ressourcen für das Forschungsdatenmanagement müssen unabhängig von den Zugangswegen bei jeder Form der Datengenerierung eingeplant werden, damit qualitativ hochwertige Daten und Dokumente für die Nutzung durch andere erstellt werden können.

2.2 Forschungsdatenzentren

„Ein Forschungsdatenzentrum ermöglicht der Wissenschaft oder weiteren berechtigten Nutzenden einen flexiblen, sicheren, rechtskonformen und zum Teil umfangreichen Zugang zu Daten“ (Datenstrategie der Bundesregierung 2021, S. 112). FDZ werden zumeist in Organisationen gegründet, die selbst Daten für wissenschaftliche oder statistische Zwecke

⁵ <https://www.coretrustseal.org/>

⁶ https://www.langzeitarchivierung.de/Webs/nestor/DE/Zertifizierung/nestor_Siegel/siegel.html

produzieren und diese für Nachnutzung mit externen Forschenden teilen möchten oder müssen.⁷

Den Datenzugang über ein FDZ (oder eine vergleichbare Struktur) zu organisieren, ist dann notwendig, wenn Daten sonst nicht für die wissenschaftliche Forschung bereitgestellt werden können. Mindestens drei Gründe können dafür entscheidend sein:

- (1) Datenschutz und Vertraulichkeit: Daten, die von natürlichen oder juristischen Personen unter Zusicherung von Anonymität oder Vertraulichkeit erhoben werden, dürfen nur anonymisiert und mit besonderen Schutzmaßnahmen geteilt werden. Anonymisiert bedeutet, dass eine Identifikation der Personen oder Organisationen im Datensatz nicht möglich ist, auch wenn Zusatzinformationen herangezogen werden. Mit der Anonymisierung wird allerdings auch das analytische Potential der Daten verringert, weil nicht nur direkte Identifikatoren aus den Daten entfernt werden, sondern auch indirekte Identifikatoren und andere sensible Merkmale. Wenn für bestimmte Forschungszwecke der Zugriff auf die sensiblen Merkmale erhalten bleiben soll oder aber eine Anonymisierung nicht vollständig möglich ist (bspw. bei Videodaten), können Daten mit besonderem Zugriffsschutz verfügbar gemacht werden. Hierzu zählt beispielsweise die Bereitstellung über einen Gastwissenschaftsarbeitsplatz.⁸ Zu beachten ist, dass neben der Datenschutz-Grundverordnung (DSGVO) und dem Bundesdatenschutzgesetz (BDSG) im Einzelfall auch weitere Gesetze wie bspw. das Sozialgesetzbuch (SGB), das Ausländerzentralregistergesetz (AZRG), das Bundesstatistikgesetz (BStatG) oder Landesschulgesetze Anwendung finden können.
- (2) Urheberrecht und Lizenzierung: Daten, die aus Angeboten kommerzieller Informationsanbieter generiert werden (z. B. Textkorpora aus Zeitungen/Webseiten, Material aus den sozialen Medien oder bestimmte psychologische Testverfahren) unterliegen dem Urheberrecht und den Allgemeinen Geschäftsbedingungen der Dienstbietenden. Diese Regelungen untersagen häufig die Weitergabe von Originalmaterialien oder schränken diese erheblich ein. Die Auswertung durch Dritte kann in vielen Fällen dennoch über andere Zugangswege umgesetzt werden.
- (3) Forschungsethische Bedenken: Die Publikation von Daten kann für Probanden mit erheblichen Risiken verbunden sein. Bei einer Datenerhebung in autoritär regierten Staaten können zum Beispiel Repressionen für Studienteilnehmende drohen. Daher können auch forschungsethische Erwägungen für Alternativen zu einem offenen Datenzugang sprechen.⁹

⁷ Teilweise liegt der Schwerpunkt der FDZ auch auf der Archivierung externer, also nicht selbst produzierter Datenbestände. In jedem Fall liegt aber eine enge Verzahnung von Forschungs- und Datenkompetenz in den FDZ vor.

⁸ vgl. RatSWD Tätigkeitsbericht 2019, S. 32f. <https://doi.org/10.17620/02671.56>.

⁹ vgl. <https://www.konsortswd.de/ratswd/themen/forschungsethik/> für aktuelle Diskussionen und Publikationen im Rahmen von KonsortSWD.

Liegen solche Gründe vor, können Daten oder zumindest einzelne Merkmale aus den Daten die Organisation, die sie erhoben hat, nicht verlassen. Dann ist die Gründung eines Forschungsdatenzentrums mittlerweile ein etablierter Weg, angemessene Zugangswege zu ermöglichen und langfristig zu entwickeln.

Den FDZ kommt damit eine doppelte Schnittstellenfunktion zu: Zum einen stellen sie den Interessenausgleich den Forschungsinteressen der datennutzenden Wissenschaftlerinnen und Wissenschaftlern und den Schutzinteressen der Merkmalsträger:innen (zum Beispiel Teilnehmer:innen an Umfragen) her, deren Informationen die Datengrundlage bilden. Das betrifft in erster Linie die Einhaltung von Vorgaben aus den rechtlichen Regelungen zum Datenschutz. Zum anderen organisieren FDZ das Forschungsdatenmanagement in den Organisationen und vermitteln Dokumentationsstandards oder Standards der Datenbereitstellung in die Forschungsinstitute und Behörden.

3 Grundfunktionen von Forschungsdatenzentren: Aufbereitung & Dokumentation, Langzeitsicherung und Datenzugang

Ein FDZ erfüllt drei Grundfunktionen:

- a) die Aufbereitung und Dokumentation von Forschungsdaten zum Zweck der Sichtbarkeit, Auffindbarkeit und inhaltlichen Nachvollziehbarkeit der Daten,
- b) die sichere Aufbewahrung (Archivierung) dieser Daten zum Zweck des langfristigen (Wert-)Erhalts und ihrer physischen Unversehrtheit sowie
- c) die systematische, rechtskonforme und transparente Bereitstellung der Daten zum Zweck der Nachnutzung entsprechend des Schutzbedarfs (auch im Sinne einer Datentreuhänderfunktion).

Es können sowohl Daten der eigenen Institution aufbereitet und archiviert werden als auch Daten anderer Institutionen bzw. von externen Forschenden aufgenommen werden. Darüber hinaus erfüllen FDZ noch weitere Aufgaben, wie die umfassende Beratung von Nutzenden zum Umgang mit den Datenbeständen sowie Öffentlichkeitsarbeit zur Bekanntmachung des Datenangebots, von Fortbildungsangeboten für die Nutzung der Daten und eigener Forschungstätigkeit, die wichtig ist für den Erwerb von Wissen über das spezifische Forschungsfeld ohne das eine sachgerechte Aufbereitung und Anonymisierung der Daten sowie Beratung von Forschenden kaum möglich ist.

3.1 Aufbereitung und Dokumentation

Vor der Aufnahme eines Datenbestands in das FDZ findet eine Eingangsprüfung statt, die technische, dokumentarische und rechtliche Prüfungen beinhaltet. Die rechtliche Prüfung umfasst eine Auseinandersetzung mit der DSGVO und anderer rechtlicher Vorschriften bezüglich der Verarbeitung und Weitergabe der Daten an Dritte¹⁰, die beispielsweise im Rahmen der Einwilligungserklärung eingeholt worden sein kann, sowie urheberrechtliche Fragen bezüglich der Forschungsdaten und Instrumente (s. *Rechtliche Grundlagen und Verträge*). Erst nach erfolgreicher Klärung dieser Fragen sollte mit der Aufbereitung begonnen werden, da im Einzelfall eine Weitergabe der Daten an Dritte eventuell nicht möglich ist (siehe auch Watteler & Ebel 2019).

Die Aufbereitung und Dokumentation eines Datensatzes kann sehr unterschiedliche Tätigkeiten umfassen. Am Beispiel eines rechteckigen Datensatzes¹¹ können diese Aufbereitungsaufgaben in folgenden Arbeitsschritten bestehen¹²: Prüfung der Konsistenz von Daten und Dokumentationen, Prüfung auf Vollständigkeit der Daten und Korrektheit der Wertebereiche, Prüfung des Datensatzes auf Vollständigkeit, Prüfung der Variablenlabels und Wertelabels auf Rechtschreibfehler und Vollständigkeit, Prüfung der Filterführung, Plausibilitätschecks, Erstellung von abgeleiteten Variablen, Fehlerkorrekturen bzw. Markierung von Fehlern, Erstellung und Anwendung eines Codierkonzepts für fehlende Werte, Längs- und/oder Querschnittsharmonisierung der Daten über verschiedene Zeiträume und/oder geographische Einheiten, Anonymisierung der Daten, Verknüpfung der Daten mit Kontextdaten (z. B. Geodaten) und Integration mehrerer Datensätze, beispielsweise zur Erstellung einer Zeitreihe oder bei international vergleichenden Projekten. Der genaue Auftrag der FDZ bei der Bearbeitung der Daten hängt von der Einbettung in die Arbeitsprozesse einer Organisation ab. In manchen Fällen ist das FDZ-Personal für die Prüfung, Aufbereitung und Anonymisierung zuständig. In anderen Fällen übergeben die Fachabteilungen (fast) fertig aufbereitete Daten, die im FDZ noch geprüft und dokumentiert werden.

Bei qualitativen Daten werden bspw. Informationen zu den Transkripten (z. B. transkribierende Person, Datum, usw.) auf Einheitlichkeit und Vollständigkeit geprüft und bei Videos technische und beschreibende Metadaten erzeugt. Die Prüfung der Anonymität und Sensibilität des Materials stellt bei nicht-standardisierten, unstrukturierten Daten einen besonders hohen Aufwand dar.

Die Dokumentation der Daten besteht aus mehreren Teilen. Zum einen müssen Dokumente angefertigt werden, die es den Datennutzenden ermöglichen, die Entstehung der Daten im

¹⁰ Hier ist insbesondere auf "besondere Kategorien personenbezogener Daten" zu achten, welche beispielsweise bei Umfragedaten eine gesonderte explizite Einwilligung voraussetzen <https://dsgvo-gesetz.de/art-9-dsgvo/>

¹¹ Rechteckige Datensätze, ähnlich einem Tabellenblatt in einem Datenverarbeitungsprogramm. Entsprechend der gängigen Konvention repräsentiert jede Spalte dabei ein Merkmal (eine Variable), jede Zeile einen Fall und in einer Zelle findet sich der Wert eines Falles für eine Variable.

¹² Die Handreichung "Aufbereitung quantitativer Daten" des Verbund Forschungsdaten Bildung beschreibt die Aufbereitung quantitativer Daten: <https://www.forschungsdaten-bildung.de/aufbereitung?la=de>

Detail nachzuvollziehen. So sollten jedem Datensatz Dokumente zur Seite gestellt werden, die es den Nutzenden erlauben, schnell einen Überblick über Charakteristika, Güte und Art der Daten zu erlangen. Solche sogenannten Provenienz-Informationen sind unerlässlich für eine optimale Nachnutzbarkeit von Forschungsdaten im Sinne der FAIR-Prinzipien. Für einen Datensatz aus einer standardisierten Befragung muss also ein Dokument vorliegen, das den Fragebogen enthält, sowie Angaben zur Codierung der Antwortmöglichkeiten. Um die Qualität der Daten zu beurteilen, benötigen Forschende darüber hinaus Informationen zur Grundgesamtheit, der Auswahl der Befragten, zum Befragungszeitraum, der Erhebungsmethode, dem Erhebungsinstitut, möglichen Fehlerkorrekturen und so weiter.

Dokumentationen von Daten sollten allgemeine Metadatenstandards wie Dublin Core¹³ oder fachspezifische Metadatenstandards, wie zum Beispiel DDI¹⁴ für sozial- und wirtschaftswissenschaftlichen Daten nutzen (Betancort et al. 2020). Zudem sollten sie kompatibel mit dem Metadatenschema für die Registrierung von persistenten Identifikatoren sein (bspw. da|ra¹⁵).

3.2 Langfristsicherung

Der Langfristsicherung von Daten und zugehörigen Dokumentationen kommt eine große Bedeutung zu, sichert sie doch die Replizierbarkeit von Forschungsergebnissen und damit einen elementaren Bestandteil der guten wissenschaftlichen Praxis. Unter Langfristsicherung ist dabei nicht das bloße Speichern von Daten und Dokumenten über Zeit gemeint (Mauer & Recker 2019), vielmehr geht es um die systematische Archivierung der Informationen und die langfristige Sicherstellung der Lesbarkeit, Interpretierbarkeit und Nutzbarkeit auch bei Technologiewechseln (z. B. Auf- und Abwärtskompatibilität verschiedener Softwareversionen, Aufkommen neuer (proprietärer) Datenformate.¹⁶ Durch redundante Speicherung auf verschiedenen voneinander unabhängigen Systemen (u. a. verschiedene Server, wenn möglich an verschiedenen Standorten) sind die Daten zudem gegen physischen Verlust zu sichern. Die Speicherorte müssen überdies gegen ungewollten Zugriff und Manipulation von innen und außen gesichert sein.

Verschiedene Drittmittelgeber verlangen mittlerweile eine Archivierung und Bereitstellung von Forschungsdaten für mindestens 10 Jahre, 'echte' Langfristsicherung kennt jedoch kein Verfallsdatum. Diese Langfristsicherung verlangt folglich eine Formatkonvertierung für dauerhaft lesbare Daten.

Um Datengebenden die Orientierung in der vielfältigen Landschaft der Archive, FDZ und Repositorien zu erleichtern, werden verschiedene Gütesiegel vergeben. Eines dieser Zertifikate

¹³ <https://dublincore.org/>

¹⁴ <https://ddialliance.org/>

¹⁵ <https://www.da-ra.de/>

¹⁶ Als "historisches" Beispiel sei der Wechsel von Lochkarten zu digitalen Formaten genannt.

ist das "CoreTrustSeal"¹⁷, das Kriterien und Standards für die transparente und qualitativ hochwertige Archivierung und Kuratierung von Forschungsdaten vorgibt.

3.3 Datenbereitstellung und Datenzugänge

Die dritte zentrale Aufgabe eines FDZ liegt in der Datenbereitstellung mit Nutzendenservice. Das Veröffentlichen von Daten für die (wissenschaftliche) Nachnutzung ist zentral für die Reproduzierbarkeit und Überprüfbarkeit von Forschungsergebnissen und damit für Legitimität und Glaubwürdigkeit von Forschung insgesamt. Zugleich ist die Bereitstellung effizient, denn sie erhöht den Nutzen der Investition in die ursprüngliche Erhebung, Aufbereitung und Dokumentation der Daten.

Der Datenzugang über ein FDZ sollte möglichst transparent, niedrighschwellig und zeitnah gestaltet sein. Alle (berechtigten) Nutzenden sollten gleichzeitig Zugang zu den Datenbeständen erhalten, Gebühren und Sperrfristen¹⁸ sollten auf ein Minimum reduziert werden oder ganz entfallen und die Daten sollten möglichst zeitnah nach der Erhebung, Aufbereitung und Dokumentation verfügbar gemacht werden. Gleichzeitig muss den Charakteristika der Daten und beispielsweise den Anforderungen an den Datenschutz Rechnung getragen werden. Eine eingehende Prüfung, welche der möglichen Datenzugangswege für einen Datensatz oder eine Datenkollektion gewählt werden können, ist somit zwingend notwendig.

Die meisten FDZ bieten je nach Anonymisierungsgrad der Daten differenzierte Zugänge. Im einfachsten Fall können Daten komplett frei bzw. nach einer Registrierung oder dem Abschluss eines Nutzungsvertrages zum Download angeboten werden. Für Daten, die stärker geschützt werden müssen, kann eine kontrollierte Datenfernverarbeitung (Schiller et al., 2017 S. 7, Schiller & Welpton, 2014) und/oder Gastwissenschaftsarbeitplätze (GWAP) bzw. Secure Data Center (SDC) in den Räumen des FDZ oder der Einrichtung angeboten werden. Dokumente und Metadaten sollten, soweit möglich, ohne Einschränkung bereitgestellt werden.

Auswahl und Gestaltung der Zugangswese sind sehr komplex, so dass eine detaillierte Beschreibung den Rahmen dieses Papier sprengen würde. Eine ausführliche Darstellung unterschiedlicher Zugangswege wird in der KonsortSWD Handreichung „Datenzugang. Einführung in das Thema Zugang zu Daten der Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften bei Forschungsdatenzentren“ publiziert (Hoffstätter et al. 2022).

Die verschiedenen Datenzugänge bringen Vor- und Nachteile mit sich und erfordern unterschiedliche Personalressourcen (Systemwartung und -administration, Lizenzmanagement, Outputkontrolle etc.) und IT-Infrastrukturen. Auch die Sicht der Nutzenden sollte

¹⁷ <https://www.coretrustseal.org/>

¹⁸ Erstverwertungsrechte der Primärforschenden, laufende Projekte, laufende peer-review Verfahren u. ä. können zeitlich begrenzte Embargos begründen.

hierbei nicht vernachlässigt werden, da für sie bei den verschiedenen Datenzugängen ebenfalls Vor- und Nachteile entstehen.

3.4 Aufnahme externer Daten

Meist stellen FDZ Daten aus der eigenen Organisation für die Sekundärnutzung zur Verfügung. Darüber hinaus nimmt ein Teil der beim RatSWD akkreditierten FDZ auch Forschungsdaten von Externen auf. Dies trifft oft auf FDZ zu, die Forschungsdaten zu einem bestimmten Forschungsfeld (z. B. Arbeitsmarktforschung oder Bildungsforschung) oder themenübergreifend zu spezifischen Datentypen anbieten (z. B. die qualitative Sozialforschung).

Sollen auch externe Datenbestände aufgenommen werden, sollte initial ein Konzept für den Ablauf, Ansprechpersonen und FDZ-interne Verantwortlichkeiten erstellt und transparent gemacht werden. Zudem muss ein Datenüberlassungsvertrag oder Archivierungsvertrag erstellt werden (vgl. Rechtliche Grundlagen und Verträge). Es ist mit sehr heterogenem Vorwissen der externen Datengebenden und daher mit variierendem Beratungsbedarf und Aufbereitungs- und Prüfaufwänden zu rechnen.

Der Verbund Forschungsdaten Bildung¹⁹ etwa koordiniert unter anderem eine verteilte Archivierung von externen Forschungsdaten aus dem Bereich der Bildungsforschung über kooperierende FDZ.

3.5 Weitere Aufgaben

Neben den obigen Aspekten sind einige weitere Aufgaben und Themenfelder zu berücksichtigen, die den Betrieb eines FDZ prägen. Dies sind etwa Aufgaben der Öffentlichkeitsarbeit und Vernetzung des FDZ mit den Zielgruppen. Fortbildungen und Workshops sowie individuelle Beratung sind für Datennutzende sehr hilfreich, um sie bei der Nutzung besonders komplexer Daten zu unterstützen und die Datenbestände zu bewerben. Daneben ist es auch wichtig, Datengebende zu adressieren: Workshops zu Datenaufnahmen können Datengebende bereits frühzeitig in ihren Forschungsprozessen darin unterstützen, Wissen zur Erhebung, Aufbereitung und Dokumentation ihrer Daten im Kontext von Sekundärnutzungen zu erwerben. Dies hilft, Fehler bei der Datenerhebung zu vermeiden, die am Ende eine Archivierung und Sekundärnutzung der Daten im ungünstigsten Fall unmöglich machen können (zum Beispiel fehlerhaft formulierte Einwilligungserklärungen). Es sind ebenfalls Ressourcen für den Aufbau und die Pflege von Netzwerken mit anderen FDZ einzuplanen, um voneinander zu lernen und gemeinsam an der Weiterentwicklung der nationalen Forschungsdateninfrastruktur (NFDI) mitzuwirken. Das FDZ wird in der Regel mit unterschiedlichen Stakeholdern kommunizieren. Dazu zählt auch das regelmäßige Erstellen von Berichten zu verschiedenen Themen und insbesondere der Datennutzung für verschiedene

¹⁹ <https://www.forschungsdaten-bildung.de>

Adressaten (etwa Institutsleitung, Beirat, RatSWD). Je nach Ausrichtung des FDZ sind auch Ressourcen für wissenschaftliche Forschung durch das FDZ-Personal zu berücksichtigen.

Darüber hinaus wird initial und auch weiterführend eine intensive Zusammenarbeit mit zentralen Stellen innerhalb der Institution (oder mit externen Dienstleistenden) anfallen, insbesondere mit der/dem Datenschutzbeauftragten (Anonymisierungskonzepte für die Datenbestände, Prüfung von Einverständniserklärungen, Datenschutzkonzept für die Daten der Nutzenden), IT-Sicherheit (Konzeption und Einrichtung der Datenzugangswege), IT-Administration (Softwarelizenzen), und Justizariat (Verträge erstellen und anpassen, ggf. urheberrechtliche Fragen klären).

4 Personalbedarf und Personalstrukturen in Forschungsdatenzentren

Die Frage nach dem (quantitativen) Personalbedarf eines FDZ kann pauschal nicht beantwortet werden. Sie hängt von Art und Umfang des zu betreuenden Datenbestandes, geltenden rechtlichen Regelungen, Einbindung in vorhandene Strukturen und Kooperationen mit externen Dienstleistenden und nicht zuletzt vom Nutzungsaufkommen ab. Ein Blick in den Tätigkeitsbericht 2020 der beim RatSWD akkreditierten FDZ²⁰ bestätigt dies, zeigt er doch, dass die Personalstruktur der FDZ sehr unterschiedlich ist. Laut diesem Bericht verfügt ein FDZ im Durchschnitt über 7,7 Vollzeitäquivalente (VZÄ), die Spannweite reicht von <1 VZÄ bis zu 34 VZÄ. Knapp drei Viertel der beim RatSWD akkreditierten FDZ haben maximal 10 VZÄ, andererseits verfügen rund 16 % der FDZ über mehr als 15 VZÄ. Über die Jahre ist tendenziell ein Anstieg in der Anzahl der Mitarbeitenden zu verzeichnen, gleichzeitig steigen aber auch die Nutzungszahlen und die Zahl der bereitgestellten Datensätze. Zudem verlangen die verschiedenen Datenzugangswege nach unterschiedlich hohem Personaleinsatz. Für die Nutzung sensibler Daten müssen Ressourcen für die Nutzendenverwaltung und -betreuung sowie die Prüfung und Freigabe der Auswertungsergebnisse (sogenannte Outputkontrollen) vorgesehen werden.

Die Arbeit der FDZ findet an der Schnittstelle zwischen Fachwissenschaften und Informationswissenschaften/Data Science statt. Das Aufgabenprofil ist vielfältig. Aufgrund des starken Wettbewerbs mit der Privatwirtschaft bei der Rekrutierung von Datenspezialisten gestaltet sich die Personalgewinnung teilweise schwierig.

Aus dem Tätigkeitsbericht 2019 wird deutlich, dass mindestens Teile des Personals in fast allen FDZ auch wissenschaftlich mit den Daten arbeiten und dass diese enge Verzahnung als Grundstein für die qualitativ hochwertige Beratung von Nutzenden angesehen wird. Geforscht wird in den FDZ sowohl zu inhaltlichen als auch zu methodischen Fragestellungen, aber auch

²⁰ <https://www.konsortswd.de/aktuelles/publikationen/taetigkeitsberichte/>

die technische Weiterentwicklung der im FDZ verwendeten oder bereitgestellten Tools und Produkte, sowie ethische und rechtliche Themen spielen eine Rolle.

Als Anhaltspunkt lässt sich eine Minimalausstattung eines FDZ anhand der verschiedenen zu leistenden Arbeiten definieren. Je nach Datenbestand und weiteren Anforderungen werden in den verschiedenen Personalkategorien unterschiedlich viele Personen benötigt:

- a) wissenschaftliche Leitung (EG14/A14 bis EG15/A15)
Aufgaben: Personalführung, Entscheidung über Tools und Standards, Verantwortung für Datenzugang, Vertragswesen, Projektmanagement, Drittmittelanträge/-projekte, eigene wissenschaftliche Forschung;
- b) wissenschaftliche Datenkurator*innen (EG 13/A13)
Aufgaben: Konzeption von Datenprodukten, Entwicklung der Workflows für Datenprüfung, Datenaufbereitung, Datendokumentation und Datenpublikation, Anonymisierung von Daten, Qualitätssicherung der Daten, Tools und Metadatenstandards, Daten und Metadatenmanagement (bspw. Datenbanken), Outputkontrolle für GWAP/SDC, eigene wissenschaftliche Forschung;
- c) Softwareentwickler*in (je nach Qualifikation und Aufgabenspezifikation EG10/A10 bis EG13/A13)
Aufgaben: Konzeption, Entwicklung und Pflege der Software/Webanwendungen (Website, Datenkatalog, ggf. weiteres);
- d) Systemadministrator*in (EG9/A9 bzw. EG10/A10)
Aufgaben: Administrieren der Server, ggf. Remote Access und GWAP/SDC, Lizenzmanagement Software;
- e) wissenschaftliche und/oder juristische Sachbearbeiter*in (EG11/A11 bzw. EG 12/A12)
Aufgaben: Operative Datenbearbeitung, Dokumentation der Daten in hoher Metadatenqualität, operative Datenbereitstellung, (datenschutz-)rechtliche Prüfung von Nutzungsanfragen;
- f) Assistenz (etwa EG8/A8)
Aufgaben: Unterstützung bei Planungsprozessen, Vertragsverwaltung, Nutzendenbetreuung am GWAP/SDC.

Der Personalaufwand kann durch die Vergabe von Tätigkeiten an externe Dienstleistenden reduziert werden. Der Betrieb von Gastwissenschaftsarbeitsplätzen (GWAP) für die on-site Nutzung beispielsweise ist sehr ressourcenintensiv, verlangt er doch nicht nur nach einer entsprechenden IT-Infrastruktur (vgl. Abschnitt 5), sondern auch nach geschultem Personal zur Betreuung der Nutzenden vor Ort, zur Vor- und Nachbereitung der Forschungsaufenthalte und zur Outputkontrolle. Auch Remote Access Verfahren, die Nutzenden den großen Vorteil ermöglichen, standortunabhängig auf sensible Daten zuzugreifen, erfordert technische und

fachliche Personalressourcen.²¹ Gleiches gilt für die Aufbereitung und Dokumentation (komplexer) Datensätze auf Variablenebene nach internationalen Standards oder den dauerhaften Betrieb einer eigenen Bereitstellungsinfrastruktur. Hier kann es sinnvoll sein, in der Anfangsphase oder dauerhaft externe Dienstleistenden einzubeziehen.

5 Benötigte IT-Infrastrukturen für den Betrieb eines FDZ

Wie der Personalbedarf hängt auch die benötigte IT-Infrastruktur direkt vom Datenangebot und den geltenden rechtlichen Regelungen ab. Datenschutzrechtlich sensible Daten verlangen nach einem besonders geschützten Zugang, datenschutzrechtlich unbedenkliche Daten können hingegen unter Umständen zum freien Download zur Verfügung gestellt werden. Für umfangreiche Datenbestände kann ein Datenkatalog hilfreich sein, für größere kumulierte Datensätze bietet sich die Möglichkeit eines individuell angepassten Datensatzes²² an.

Für die langfristige datenschutzkonforme Archivierung und Bereitstellung von qualitativ hochwertigen Daten sind mit Blick auf die (IT-)Infrastruktur verschiedene Voraussetzungen zu erfüllen. Unter anderem müssen Konzepte und Workflows erstellt und implementiert werden für

- die Website mit allgemeinen Informationen zum Forschungsdatenzentrum.
- die Software und Tools, mit denen die Daten und Dokumente aufbereitet und bereitgestellt werden etc.,
- einen Datenkatalog, der – ähnlich einem Bibliothekskatalog – einen Überblick über die im FDZ verfügbaren Daten enthält. Idealerweise findet sich zu jedem vorhandenen Datensatz zentrale Informationen über diesen Datensatz. Idealerweise besteht hier sogar die Möglichkeit eines Dokumenten- oder Datendownloads oder eine Warenkorbfunktion zur Bestellung,
- das Zugriffsmanagement auf die Daten und Dokumente von innen und außen,
- die Verwaltung von Nutzungsanträgen, Verträgen und Kundendaten (Customer Relationship Management), falls die Zahl an vertriebenen Datenprodukten und Nutzenden groß genug ist und
- die Langfristsicherung von Daten und Dokumenten, idealerweise in einem redundanten und verteilten Speichersystem,

²¹ Ein Papier des RatSWD zu Remote Access Verfahren gibt neben einer technischen Beschreibung Anhaltspunkte über Personalressourcen: https://www.konsortswd.de/wp-content/uploads/RatSWD_WP_261.pdf

²² Ein Beispiel ist der Download Wizard des European Social Survey, erreichbar unter <https://www.europeansocialsurvey.org/downloadwizard/> oder das Informationssystem des SOEP am DIW auf <https://paneldata.org/>

Je nach Datenart, geltenden rechtlichen Bestimmungen sowie Anforderungen der Datengebenden und der Zielgruppe können unterschiedliche Softwarelösungen eingesetzt werden. Diese Lösungen können in der (teilweisen) Vergabe von Aufträgen an externe Dienstleistenden (u. a. *da|ra*, SDN, *DataServices*) liegen. Eine Referenzimplementation für einzelne oder alle Elemente der IT-Infrastruktur eines FDZ wurde bislang noch nicht publiziert.

5.1 Hardware

Zur Bearbeitung umfangreicher Datensätze sind leistungsfähige Rechner oder Server notwendig. Zudem verlangt die (langfristige) Archivierung von Datensätzen und Dokumenten idealerweise eine verteilte Serverlandschaft, um gegen Datenverlust geschützt zu sein. Bei schützenswerten Daten muss sichergestellt werden, dass auch mögliche externe Dienstleister, die für Speicher und Back-up Lösungen beauftragt werden, angemessene Schutzmaßnahmen umsetzen (zum Beispiel Auftragsdatenverarbeitungsverträge, Verschlüsselungsverfahren).

Die Größe der Server ist dabei direkt abhängig vom Umfang des Datenbestandes. Sofern eine Infrastruktur für einen Datenfernzugriff aufgebaut werden soll, sind hierfür leistungsfähige Server mit Wartungsverträgen für schnelle Reparaturen bei Systemausfällen vorzusehen, um die Verfügbarkeit des Dienstes für die Nutzenden sicherzustellen. Der Leistungsbedarf bei Servern für den kontrollierten Datenfernzugriff ist eine Funktion des Datenvolumens und der Anzahl paralleler Zugriffe auf die Daten.

5.2 Software

Zur Aufbereitung²³ und Dokumentation eines Datensatzes sind verschiedene Software-Pakete notwendig. Die (quantitativen) Daten werden im Allgemeinen mit einem in den Sozialwissenschaften gängigen Statistikprogramm (z. B. *Stata*, *R*, *SAS*, *SPSS*) aufbereitet und dann zur Nachnutzung in einem oder mehreren dieser Datenformate bereitgestellt. Es ist zudem möglich, ein offenes Datenformat wie ASCII zu verwenden. Die Datenformate (und ggf. bereitgestellte Software) sollten an den Bedarfen der Zielgruppe orientiert sein. Weit verbreitet sind in den Communities inzwischen die Programmiersprachen *R* und *Python*, die ohne weitere Kosten verwendet werden können. Eine Befragung unter Nutzenden kann Aufschluss über die Softwarebedarfe geben. Zum Schutz sensibler Daten sollte eine Software zur Verschlüsselung von Daten und sensiblen Zusatzinformationen verfügbar sein.

Für die Erstellung der Studienbeschreibungen (Metadaten) nach etablierten Standards sind wiederum spezielle Programme (Metadateneditoren) notwendig. Zudem müssen die Daten eine Versionierung und eine persistente Identifikation, wie eine DOI oder URN, bekommen (bspw. über *da|ra* Registrierungsagentur), um die FAIR-Prinzipien einzuhalten (Betancort et al. 2020).

²³ Für die Anonymisierung/Pseudonymisierung von Textdaten bietet Qualiservice das Tool *QualiAnon* an: <https://www.qualiservice.org/de/news/qualiservice-anonymisierungstool-qualianon-nutzbar.html>

Es wird eine Website/Webanwendung benötigt, um etwa die Leistungsangebote des FDZ, die Rahmenbedingungen der Datennutzung und Kontaktmöglichkeiten zu kommunizieren. Die Konzeption, Erstellung und Pflege ist zentraler Bestandteil der Öffentlichkeitsarbeit des FDZ.

Zudem wird eine Webanwendung im Sinn eines Datenkatalogs²⁴ benötigt, um die einzelnen durch das FDZ angebotenen Datenbestände zusammen mit den beschreibenden Metadaten zu präsentieren und diese auch über die Webanwendung (insb. REST-API) anfordern zu können.

Je nachdem, welche Datenzugangswege (vgl. Abschnitt Datenbereitstellung und Datenzugänge) eingesetzt werden, können für entsprechende Zugangswege (insbesondere für den Datenfernzugriff und Gastwissenschaftsarbetsplätze) weitere Lizenzen für Analyse-Software und Software für die Infrastruktur anfallen (bspw. Betriebssysteme und Virtualisierungssoftware für Remote Server System).

Wenn mit einem hohen Aufkommen an Datennutzung gerechnet wird, sollte eine softwaregestützte Verwaltung von Datennutzenden (Stammdaten), Datennutzungsanträgen und -verträgen in Betracht gezogen werden.

5.3 IT-Sicherheit

Eine datenhaltende und vertreibende Institution benötigt zum einen ein Sicherheitskonzept, das die archivierten Daten und Dokumente vor unerlaubten Zugriffen von innen und außen schützt. Dazu zählen Firewalls, VPN-Tunnel und andere geeignete Maßnahmen, um Zugriffe von außen zu kontrollieren und gegebenenfalls zu verhindern. Es muss jedoch auch sichergestellt werden, dass datenschutzrechtlich sensible Daten, beispielsweise Adressdaten, zum einen von den Befragungsdaten getrennt verwahrt werden und zum anderen, dass nur Personen darauf Zugriff haben, die diesen für die Erfüllung ihrer Aufgaben benötigen. Die Wahl der IT-Sicherheitsmaßnahmen hängt dabei insbesondere vom Schutzbedarf der im FDZ vorhandenen Daten ab.

6 Rechtliche Grundlagen und Verträge

6.1 Rechtliche Rahmenbedingungen

Ein FDZ wird im Regelfall als Teil einer bereits bestehenden Organisation gegründet, so dass zunächst die rechtlichen Rahmenbedingungen dieser Institution auch für das FDZ gelten. Im FDZ muss insbesondere die Datenschutz-Grundverordnung (DSGVO) im Rahmen der

²⁴ Eine Standard-Software für diese Anforderungen gibt es nicht. Viele FDZ des RatSWD nutzen Eigenentwicklungen. International wird häufig Dataverse genutzt, eine Open-Source Software, die auf die eigenen FDZ-Prozesse angepasst werden kann.

Datenübernahme in ein FDZ, der langfristigen Speicherung sowie der Nutzung, Datenaufbereitung und Bereitstellung der Daten an Dritte eingehalten werden (Watteler & Ebel 2019). Sofern für die Daten der Institution spezifische Gesetze gelten, die die Datenbereitstellung für Dritte betreffen, sind auch diese anzuwenden.²⁵ Eine Organisation, die ein FDZ plant, sollte die rechtlichen Rahmenbedingungen insofern im Vorfeld mit der internen Rechtsabteilung oder einer rechtlichen Beratung klären.

Im Aufbau eines neuen FDZ in einer bestehenden Institution können drei verschiedene Varianten von Daten im Sinn ihrer Herkunft zur Archivierung und Bereitstellung geplant werden: a) ab dem Aufbau des FDZ neu entstehende Daten der eigenen Institution, b) Daten der eigenen Institution mit Erhebungszeitpunkt vor der FDZ-Gründung und c) Daten aus anderen Institutionen. Für Daten der Variante a) aus der eigenen Institution gelten die aktuellen gesetzlichen Regelungen des nicht in allzu weit zurückliegenden Entstehungszeitpunkts, die für die Institution relevant sind. Für Daten der Variante b) aus der eigenen Institution ist abhängig von ihrem möglicherweise in weiterer Vergangenheit liegender Entstehungszeitpunkt unter Umständen noch das BDSG vor Geltung der DSGVO zu berücksichtigen.²⁶ Für Daten der Variante c) von Instituts-externen Datengebern würden gegebenenfalls andere gesetzliche Regelungen der die Daten erhebenden Partei gelten (betreffend der Datenerhebung und -weitergabe).

Das Urheberrecht ist zu berücksichtigen, sofern durch das Urheberrecht geschützte Inhalte durch ein FDZ übernommen und/oder an Dritte im Sinn der Datennutzenden weitergegeben werden. Sowohl die Datengebern als auch das FDZ müssen Sorge dafür tragen, dass keine Urheberrechte beziehungsweise davon abgeleitete Nutzungsrechte Anderer beeinträchtigt werden. Dies kann zum Beispiel Texte oder geschützte Erhebungsinstrumente betreffen. Die Rechteinhaber können Individuen, Forschungseinrichtungen oder Verlage sein. Gerade im Bereich der Psychologie sind Erhebungsinstrumente häufig urheberrechtlich geschützt.²⁷

Gleichzeitig bestehen für öffentliche Archive und Bibliotheken Sonderrechte für die digitale Sicherung urheberrechtlich geschützter Inhalte, die unter bestimmten Umständen auch für ein FDZ gelten können (RatSWD 2019). Diese Optionen gilt es bei der Übernahme von Daten (insbesondere größere Textkorpora) zu berücksichtigen.

²⁵ Dies kann bspw. das Sozialgesetzbuch sein, das sich näher zum Forschungsauftrag und der Sekundärnutzung von Daten der Bundesagentur für Arbeitsmarkt und Berufsforschung äußert (§§ 280, 282 SGB III, insbesondere § 282 Abs. 7 SGB III bzgl. der Bereitstellung anonymisierter Daten für die Forschung).

²⁶ Hierzu ist anzumerken, dass Datenerhebungen vor der Gründung eines FDZ in der Institution bzw. generell in der etwas weiter zurückliegenden Vergangenheit häufig ohne Reflektion späterer Sekundärnutzungen durch Dritte erhoben wurden. Daher kann es sein, dass insbesondere Einverständniserklärungen dieser alten Datenerhebungen eine Sekundärnutzung dieser Daten erschweren oder datenschutzrechtlich und forschungsethisch unmöglich machen, da die Sekundärnutzung zum damaligen Zeitpunkt noch keine Anforderung dargestellt haben mag. Hier sind eine frühzeitige Sichtung und rechtliche Prüfung von zentraler Bedeutung.

²⁷ siehe auch <https://www.forschungsdaten-bildung.de/doku-instrumente?la=de>

6.2 Vertragswerke und Nutzungsbedingungen

Der Bedarf an Vertragswerken und den genannten verwandten Dokumenten wird durch das Aufgabenprofil des FDZ und die Art der Umsetzung der Prozesse insbesondere in der Datenbereitstellung definiert.

Für die Datennutzung müssen die Nutzungsbedingungen zwischen den Datennutzenden und dem FDZ schriftlich geklärt werden. Dies kann über unterschiedliche Prozesse und formale Wege erfolgen, die jedoch letztlich alle der Klärung der Nutzungsbedingungen dienen. Die erste Variante stellt eine schriftliche Darstellung der Nutzungsbedingungen auf der Website des FDZ dar. Diese können individuell für die Bedürfnisse des FDZ formuliert sein, gelten dann aber für alle Datennutzenden gleichermaßen. Sofern inhaltlich für den betreffenden Datenbestand und das FDZ adäquat können auch Standardlizenzen genutzt werden, die ebenfalls faktisch Nutzungsbedingungen darstellen. In dieser Variante erfolgt die Datenbereitstellung mit oder ohne vorherige Registrierung der Datennutzenden als Download über die Website des FDZ oder einen Datenkatalog. An einer Stelle im Prozess des Downloads der Daten sollten die Datennutzenden auf die Nutzungsbedingungen hingewiesen und eine Bestätigung der Akzeptanz und Einhaltung der Nutzungsbedingungen über eine Opt-In-Erklärung als Voraussetzung des Downloads festgehalten werden.²⁸ Die zweite Variante stellt einen Datennutzungsvertrag dar, der zwischen den Datennutzenden und dem FDZ geschlossen wird und regelt, welche Nutzenden zu welchen Nutzungszwecken und in welchem Zeitraum einen Datenbestand nutzen dürfen.

Die unterschiedlichen Varianten weisen unterschiedliche Kontrollniveaus des FDZ über die Datennutzung und unterschiedliche wahrgenommene Verbindlichkeiten der Nutzungsbedingungen auf Seiten der Datennutzenden auf. Zudem besteht im Fall eines Vertrags tendenziell eher die Möglichkeit, in gut begründeten Nutzungsfällen den Vertrag für Einzelfälle anzupassen. Allgemein definierte Nutzungsbedingungen werden hingegen nicht für einen einzelnen Nutzungsfall mit einer besonderen Fallkonstellation angepasst. Sie sind besonders dann geeignet, wenn die Regelungen stark standardisiert werden können, zum Beispiel für die Nutzung anonymisierter Daten. Im Fall der Bearbeitung von Einzelfallkonstellationen muss einerseits vermieden werden, den Grundsatz der Gleichbehandlung der Datennutzenden zu verletzen. Andererseits zeigt die Erfahrung des laufenden Betriebs von FDZ, dass es Konstellationen an Institutionen der Datennutzenden geben kann, die eine Nutzung auf Basis des Standardvertrags unmöglich machen würden. Es ist jeweils zu entscheiden, ob eine spezifische Anpassung gangbar ist und ob diese nur für den Einzelfall vorgenommen werden soll oder die Anpassung ggf. auch grundsätzlich für den Standardvertrag sinnvoll wäre.

Nutzungsbedingungen können Regelungen zu den folgenden Aspekten beinhalten: Nutzungsvoraussetzungen in der Regel mit Einschränkung zur wissenschaftlichen Nutzung,

²⁸ Ein Beispiel dafür sind die Nutzungsbedingungen für den Datendownload bei GESIS Leibniz-Institut für Sozialwissenschaften.

Rahmenbedingungen der Datennutzung und Regelung bzw. ggf. Verbot der Datenweitergabe an Dritte, Datenschutzaspekte und De-Anonymisierungsverbot, Umgang mit den Daten bei Beendigung des Vorhabens bzw. ggf. Löschungspflicht, Haftung bei Verstößen gegen die Nutzungsbedingungen, ggf. Entgelt-/Gebührenordnung oder ein Verweis auf diese. Ein Datennutzungsvertrag umfasst ergänzend Regelungen zum konkreten Vertragsgegenstand, der Vertragsdauer, Erhebung und Nutzung personenbezogener Daten der Datennutzenden, Kündigungsmöglichkeiten, ggf. Kosten und Gerichtsstand.²⁹

Sofern das FDZ Daten archivieren soll, die nicht aus der eigenen rechtlichen Institution stammen, ist zudem ein Datenüberlassungsvertrag oder synonym ein Archivierungsvertrag nötig. Dieser regelt die Überlassung der Daten und den Umgang mit diesen. Im Kontext von FDZ übertragen die Datengebenden den FDZ häufig nur das einfache Nutzungsrecht, im Kontext von klassischen Archiven werden häufig ausschließliche Nutzungsrechte übertragen. Die in Datenüberlassungsverträgen zu regelnden Aspekte sind zum Beispiel: Vertragsgegenstand, Rechte und Pflichten der Daten gebenden Partei sowie des FDZ, Übertragung von Nutzungsrechten inklusive der Weitergabe an Dritte, Regelung der Art der Datenbereitstellung, Haftung, Rechtsnachfolge, Vertragsdauer, Gerichtsstand.

7 Erfahrungen aus der Praxis von FDZ-Neugründungen

7.1 Aufbauphase und Anschubfinanzierung

Die Erfahrungen in diesem Abschnitt sind nur als Anhaltspunkt zu verstehen, da die Ansprüche, Aufwände und Kosten eines FDZ-Aufbaus von zahlreichen Faktoren des Gründungsvorhabens abhängen. Dies zeigt sich bereits in Abschnitt 4 zum Personalbedarfs der bereits bestehenden FDZ, der eine hohe Spannweite von 1 bis 36 und durchschnittlich 8,9 VZÄ aufweist. Entsprechend müssen die konkreten Anforderungen, Aufwände und Kosten für den Einzelfall eruiert werden. Grundlagen für diese Überlegungen können die Hinweise dieser Handreichung sein, ebenso die Bedarfe der spezifischen Zielgruppe. Im Rahmen der Institution muss geprüft werden, ob und welche Infrastrukturen und Ressourcen der Institution mit genutzt werden können und welche grundständig neu aufgebaut werden. Teils können auch externe Dienstleistungen einbezogen werden (zum Beispiel für Langzeitarchivierung oder den Aufbau und Betrieb von IT-Infrastrukturen).

In der Konzeptionsphase des FDZ ist zu überlegen, welche Datenbestände über das FDZ archiviert und bereitgestellt werden sollen. Dies wird Auswirkungen auf die Ressourcen, die Ausgestaltung der IT-Infrastruktur und die Personalakquise haben. Zu entscheiden sind etwa

²⁹ In dieser Publikation werden Formulierungshilfen für Datennutzungsverträge und Nutzungsbedingungen gegeben: Pegelow, et al (2020). Ein Beispiel für einen eher generischen Datennutzungsvertrag für die off-site Nutzung ist der Nutzungsvertrag des GESIS Secure Data Center.

diese Fragen: Welche Datentypen werden im FDZ bearbeitet (quantitative Daten wie Surveydaten oder Registerdaten, qualitative Daten wie Interviewtranskripte, Audiomaterial, Videomaterial, Daten aus Social Media Kontexten)? Wie hoch ist der Aufwand für die Aufbereitung der Daten? Werden nur Daten der eigenen Institution oder auch anderer Institutionen aufgenommen? Sollen nur zukünftig neu entstehende Daten der Institution bearbeitet werden oder sollen auch alte bestehende Daten bearbeitet werden? Sollen Daten einer sensiblen Personengruppe oder besonderer Kategorien personenbezogener Daten (vgl. Art. 9 DSGVO) verarbeitet werden? Brauchen die Nutzenden einen Zugang zu pseudonymisierten (also nicht vollständig anonymisierten) Daten?

In der Finanzierung zu berücksichtigen sind insbesondere Personalkosten, Infrastruktur im Sinn von Hardware und Software, ggf. IT-Dienstleistungen, Reisekosten, Fortbildungskosten, Übersetzungsdienstleistungen und (datenschutz-) rechtliche Beratung (etwa zu Aspekten, die im Abschnitt Rechtliche Rahmenbedingungen beschrieben wurden). Die geplanten Zugangswege sind spezifisch hinsichtlich ihrer technischen und organisatorischen Anforderungen und Kosten zu prüfen.

Bei einem Aufbau eines neuen FDZ kann zunächst als Orientierungspunkt von einer etwa dreijährigen Aufbauphase ausgegangen werden. In dieser Zeit können zum einen die grundständigen Infrastrukturen aufgebaut und zum anderen erste besonders relevante Datenbestände aufbereitet werden. Hier sollte die Personalgewinnung (insbesondere der wissenschaftlichen Leitung), der Aufbau der notwendigen technischen Infrastrukturen, der Prozesse und der Gestaltung der rechtlichen Themen (Datenschutzrecht, Vertragswerke, Urheberrecht) im Fokus stehen, da diese die Grundlage für eine effiziente und nutzungsfreundliche FDZ-Arbeit bilden.

Eine Anschubfinanzierung kann projektförmig erfolgen. So kann die Personal- und Sachmittelausstattung in der Aufbauphase von der Ausstattung des späteren laufenden Betriebs abweichen, etwa durch einen initial erhöhten Bedarf an IT-Ressourcen zum Aufbau der Infrastrukturen. Der Aufbau sollte von einem externen Beirat begleitet werden, der die Bedarfe der Zielgruppe in den Aufbauprozess einbringt. Dies kann durch einen spezifisch für das FDZ neu gegründeten oder einem bestehenden Beirat der Institution erfolgen. Es ist jedoch in beiden Fällen von Vorteil, neben Forschenden aus der Zielgruppe auch Personen mit Expertise zur Forschungsdateninfrastruktur zu beteiligen. Es ist ebenfalls hilfreich, im Laufe des Aufbaus Kontakt zu bestehenden FDZ und zum RatSWD / KonsortSWD aufzunehmen, um in Austausch mit der bestehenden Forschungsdateninfrastruktur-Community zu treten. Sofern inhaltlich eine Passung zum RatSWD besteht, kann auch die RatSWD-Akkreditierung angestrebt werden (vgl. Abschnitt 8). In jedem Fall muss eine Planung für die Überführung in eine institutionelle Grundfinanzierung Bestandteil des Arbeitsplans für die Projektphase sein.

7.2 Laufender Betrieb

Im Übergang von der Aufbauphase in den laufenden Betrieb müssen die konzipierten Prozesse und Infrastrukturen den Praxistest bestehen. Für die Finanzplanung sollte von Beginn an mit Anpassungsbedarfen gerechnet werden, zum Beispiel für Prozessoptimierungen, Klärungen von Fragen und möglicherweise Anpassungen der Vertragstexte. Die initiale IT-Infrastruktur, etwa die Website und der Datenkatalog, müssen sich in der Interaktion mit den Datennutzenden und Datengebenden bewähren und vermutlich angepasst werden. Ein Teil der Personalressourcen muss von nun an für die Betreuung der Datennutzenden und Datengebenden sowie Öffentlichkeitsarbeit und – soweit Teil der Aufgaben des FDZ Personals – für die Erstellung von Projektanträgen für Angebotserweiterungen über den grundständigen Betrieb hinaus eingesetzt werden. Zudem ist zu klären, ob und in welchem Umfang das Personal eigene Forschung als Teil des Aufgabenprofils aufweisen soll, wie dies vom RatSWD empfohlen wird (vgl. Abschnitt 8). Mit einem gewissen Forschungsanteil ist es wahrscheinlicher, dass die Arbeit des FDZ stärker an den Bedürfnissen der wissenschaftlichen Nutzenden orientiert bleibt und die Beratung dauerhaft adäquat erfolgen kann. Die selbstständige Forschung des FDZ-Personals sichert die Datenkompetenz im FDZ. Alternativ können die Personalressourcen vollständig auf den Betrieb und die Weiterentwicklung des FDZ als Forschungsdateninfrastruktur fokussiert werden, was zwar initial die Kosten senkt, langfristig aber die Innovationsfähigkeit des FDZ einschränkt.

Im FDZ-Betrieb ist zu berücksichtigen, dass sich die umgebende IT-Landschaft im Laufe der Zeit ändert und eigene IT-Systeme entsprechend an neue technische Standards und Änderungen der gesetzlichen Regelungen angepasst werden müssen (bspw. neue Frameworks und Aktualisierungen, so dass Webanwendungen angepasst werden müssen). In kalkulatorischer Hinsicht sollte für die IT-Infrastruktur daher mit einer Abschreibung der Systeme gerechnet werden, so dass dauerhaft Mittel für die Aktualisierung der IT eingeplant werden müssen.

Eine grundlegende Überlegung betrifft die Finanzierung eines FDZ im laufenden Betrieb. Diese sollte nach der Aufbauphase in eine grundständige Finanzierung überführt werden. Ein FDZ ist für die dauerhafte Archivierung und Bereitstellung von Daten zuständig. Dies ist eine Daueraufgabe ohne eine vorab zu definierende zeitliche Begrenzung. Projektförmig kann keine dauerhafte Archivierung und Datenbereitstellung sowie auf technischer Ebene keine dauerhafte Bereitstellung von Daten mit DOIs und entsprechenden Landing Pages zugesichert werden.

7.3 Entwicklung: Innovation und Angebotserweiterung

Sofern der grundständige Betrieb realisiert ist, können Erweiterungen der Datenangebote, der Infrastruktur und Prozesse angegangen werden. Nur exemplarisch seien die Ergänzung von weiteren Datenbeständen, die Einführung eines zusätzlichen für die Designated Community relevanten Datentyps, Aktivitäten bzgl. Record Linkage, eine Zertifizierung mit dem

CoreTrustSeal zur stetigen Optimierung der Sicherheit und als Zeichen für ein vertrauenswürdigen FDZ bzw. Datenarchiv oder Optimierungen der Software genannt. Überdies sollten für das FDZ relevante Kooperationen mit der Forschungsdateninfrastruktur ausgebaut werden (bspw. RatSWD/KonsortSWD³⁰, EcoSoc-IN³¹ der GO FAIR Initiative³²).

8 Akkreditierung durch den RatSWD

Seit der Gründung der ersten FDZ in den statistischen Ämtern des Bundes und der Länder im Jahr 2001 haben sich 41 FDZ akkreditieren lassen und sich damit der Qualitätskontrolle durch den RatSWD und seiner Gremien unterworfen.

Für die FDZ liegen die Vorteile der Akkreditierung auf der Hand. Nach außen kann dokumentiert werden, dass die Mindestanforderungen an einen transparenten Datenzugang nach den Kriterien des RatSWD durch das FDZ eingehalten werden. Über die Beschwerdestelle beim RatSWD besteht zudem ein unabhängiges Gremium, das im Falle von Konflikten mit Nutzenden vermitteln kann. Eine Akkreditierung durch den RatSWD ist deshalb ein Qualitätssiegel, sowohl gegenüber Zuwendungs- und Drittmittelgebenden als auch gegenüber den Datennutzenden.

Darüber hinaus sind alle FDZ im Ausschuss Forschungsdateninfrastruktur (FDI Ausschuss) vertreten. Der Ausschuss bietet ein Forum zur Vernetzung und Kooperation der FDZ aus den Bereichen der Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften in Deutschland. Darüber hinaus werden FDZ durch eine Akkreditierung sichtbar, denn der FDI Ausschuss bietet beispielsweise eine FDZ-übergreifende Datensuche³³ und präsentiert alle FDZ auf den Webseiten des RatSWD bzw. des Konsortiums für die Sozial-, Wirtschaftsdaten (KonsortSWD) in der Nationalen Forschungsdateninfrastruktur.

Eine Akkreditierung beim RatSWD erfolgt auf Antrag durch das FDZ. Der FDI Ausschuss berät über den Antrag und gibt eine Empfehlung an den RatSWD ab, der dann entscheidet. Für eine Akkreditierung müssen drei Pflichtkriterien erfüllt sein:³⁴

- Mindestens ein Datenzugangsweg
- Bereitstellung von Dokumentationen zu den Daten
- Konzept zur langfristigen Verfügbarmachung der Daten

Darüber hinaus muss das FDZ seit mindestens sechs Monaten bestehen und die Nutzung der bereitgestellten Daten durch mindestens drei externe Datennutzende nachweisen können, beispielsweise über Nutzungsverträge. Detailliertere Spezifikationen zu den

³⁰ <https://www.konsortswd.de/>

³¹ <https://www.go-fair.org/implementation-networks/overview/ecosoc-in/>

³² <https://www.go-fair.org/>

³³ <https://www.konsortswd.de/datenzentren/datensuche-in-den-fdz/>

³⁴ <https://www.konsortswd.de/datenzentren/akkreditierung/>

Akkreditierungskriterien wurden 2017 von einer gemeinsamen Arbeitsgruppe des RatSWD und des FDI Ausschuss publiziert (RatSWD 2017).

Der RatSWD verlangt bei der Erstakkreditierung zudem verschiedene Informationen, um den Umfang und die Qualität des operativen Geschäfts des antragstellenden FDZ beurteilen zu können:

- Datenangebot im Bereich der Sozial-, Verhaltens- und Wirtschaftsdaten
- Konzept für die zeitnahe Verfügbarmachung der Daten
- Bereitstellung weitergehender Arbeitshilfen
- Qualitätssicherung der Datensätze
- Sicherstellung des Datenschutzes unter Berücksichtigung der Forschungsinteressen
- Servicekonzept
- Einheit von Einrichtung und FDZ
- Verfügbarmachung aller forschungsrelevanten Datensätze
- Überschneidung und Abgrenzung zu bestehenden FDI-Einrichtungen
- Forschungsaktivitäten
- Mehrfachangebote des gleichen Datenangebots (Doppelhosting; nicht mehrere Standorte)
- Bearbeitungsdauer der Anträge zur Datennutzung
- Personalbestand
- Weiterentwicklung der Infrastruktur

Wird ein FDZ akkreditiert, so ist mittels eines standardisierten jährlichen Tätigkeitsbericht Auskunft über das operative Geschäft zu geben. Inhalt dieses Berichts sind Kennzahlen zum Personal und zu dessen Forschungstätigkeiten, zu Datenzugängen und Datenbereitstellungen, zur Nutzung der Daten in der Forschung, zu nationalen und internationalen Kooperationen sowie zum Wissenstransfer. Die Teilnahme am jährlichen Monitoring der FDZ sind für die FDZ obligatorisch.

Literaturverzeichnis

Bundeskanzleramt (2021). Datenstrategie der Bundesregierung. Eine Innovationsstrategie für gesellschaftlichen Fortschritt und nachhaltiges Wachstum.

<https://www.bundesregierung.de/breg-de/service/publikationen/datenstrategie-der-bundesregierung-1845632>.

Lauber-Rönsberg, A. (2021). Rechtliche Aspekte des Forschungsdatenmanagements. Praxishandbuch Forschungsdatenmanagement. De Gruyter Saur. 89-114.

Mauer, R. & Recker, J. (2019). Data Sharing: Von der Sicherung zur langfristigen Nutzung der Forschungsdaten. In: Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten: Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten, Verlag Barbara Budrich, Opladen, 115-133. <https://doi.org/10.3224/84742233.08>.

Pegelow, L., Neuendorf, C., Daniel, A. & Buck, D. (2020). Formulierungshilfen für Forschungsdatenzentren zum Thema Nutzungsbedingungen. RatSWD Working Paper 271/2020. Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.53>.

RatSWD [Rat für Sozial- und Wirtschaftsdaten] (2017). Qualitätssicherung der vom Rat für Sozial- und Wirtschaftsdaten (RatSWD) akkreditierten Forschungsdatenzentren (FDZ). RatSWD Output 8 (5). <https://doi.org/10.17620/02671.4>.

RatSWD [Rat für Sozial- und Wirtschaftsdaten] (2019): Big Data in den Sozial-, Verhaltens- und Wirtschaftswissenschaften: Datenzugang und Forschungsdatenmanagement. RatSWD Output 4 (6). <https://doi.org/10.17620/02671.39>.

RatSWD [Rat für Sozial- und Wirtschaftsdaten] (2019). Remote Access zu Daten der amtlichen Statistik und der Sozialversicherungsträger. RatSWD Output 5 (6). <https://doi.org/10.17620/02671.42>.

Science Europe (2021). Practical Guide to the International Alignment of Research Data Management.

Schiller, D. H., Eberle, J., Fuß, D., Goebel, J., Heining, J., Mika, T., Müller, D., Röder, F., Stegmann, M., & Stephan, K. (2017). Standards des sicheren Datenzugangs in den Sozial- und Wirtschaftswissenschaften: Überblick über verschiedene Remote-Access-Verfahren. RatSWD Working Paper 261/2017. Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.15>.

Schiller, D., & Welpton, R. (2015). Distributing Access to Data, not Data. IASSIST Quarterly, 38(3), 6-6.

Watteler, O. & Ebel, T. (2019). Datenschutz im Forschungsdatenmanagement. In: Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten: Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten, Verlag Barbara Budrich, Opladen, 57-80. <https://doi.org/10.3224/84742233.05>.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR Guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.

Impressum

Kontakt:

Rat für Sozial- und Wirtschaftsdaten (RatSWD)

Geschäftsstelle

Am Friedrichshain 22

10407 Berlin

office@ratswd.de

<https://www.ratswd.de>

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

Berlin, Juli 2022



Diese Veröffentlichung ist unter der Creative-Commons-Lizenz (CC BY 4.0) lizenziert:
<https://creativecommons.org/licenses/by/4.0/>

doi: 10.17620/02671.70

Zitationsvorschlag:

Buck, D., Croisier, J., Eder, C., Hoffstätter, U., Jansen, M., Meyermann, A., & Pascal Siegers (2022). *Handreichung: Forschungsdatenzentren gründen*. RatSWD Working Paper 280/2022. Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.70>