

4

Working Paper
2022

KonsortSWD

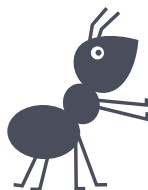


Konsortium für die
Sozial-, Verhaltens-, Bildungs- und
Wirtschaftswissenschaften

Datenzugang

Einführung in das Thema Zugang
zu Daten der Sozial-, Verhaltens-,
Bildungs- und Wirtschaftswissen-
schaften in Forschungsdatenzentren

Ute Hoffstätter und Monika Linne



Dezember 2022

www.konsortswd.de

Datenzugang

Einführung in das Thema Zugang zu Daten der Sozial-,
Verhaltens-, Bildungs- und Wirtschaftswissenschaften
in Forschungsdatenzentren

Ute Hoffstätter,¹ Monika Linne²

Dezember 2022

DOI: 10.5281/zenodo.7347064

¹ Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW)

² RWI – Leibniz-Institut für Wirtschaftsforschung, mehrheitlich zur Zeit der Erstellung des
Papiers GESIS – Leibniz-Institut für Sozialwissenschaften

Wir danken für wertvolle Hinweise (in alphabetischer Reihenfolge): Daniel Buck, Dr. Andreas
Daniel, Alexia Meyermann, Dr. Kati Mozygemba, Dr. Pascal Siegers und Dr. Jonas Recker

Abstract

Zentrale Aufgabe eines Forschungsdatenzentrums (FDZ) ist neben der Archivierung der Forschungsdaten und der damit einhergehenden Dokumentations- und Kurationstätigkeiten die Ermöglichung des Datenzugangs zur Sekundärnutzung.¹ Dieser Artikel soll deshalb (angehenden) FDZ und anderen Forschungsdateninfrastrukturen der Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften grundlegende Informationen dazu liefern, welche verschiedenen Möglichkeiten es gibt, Datenbereitstellungswege zu digitalen Ressourcen anzubieten. Hierfür werden u. a. die verschiedenen Datenbereitstellungswege Download, Varianten des Remote Access, Gastwissenschaftsarbeitsplätze sowie damit einhergehende Eigenschaften oder auch Services zur Archivierung und Veröffentlichung dargestellt. In diesem Zusammenhang werden auch Aspekte aufgezeigt, die für die Auswahl eines Datenzugangs relevant sind. Dazu gehören Kosten, die bei den unterschiedlichen Datenbereitstellungswegen

¹ Oft werden Daten nur eingeschränkt zur Verfügung gestellt, etwa für bestimmte Nutzungsgruppen oder Anwendungszwecke (z. B. wissenschaftliche Forschung).

entstehen können. Das Modell der Five Safes wird herangezogen, um verschiedene Parameter der Datenbereitstellung zu erläutern und die Zusammenhänge dieser Parameter zu illustrieren.

Ebenfalls behandelt werden die Charakteristika der Daten, die bestimmen, wie offen oder geschlossen der Zugang zu den Daten erfolgen kann bzw. ob gegebenenfalls Anonymisierungsmaßnahmen durchgeführt werden müssen. Beim zielgruppenspezifischen Zugang zu Daten helfen Datenkataloge oder Recherchesysteme. Sie können anhand von Zugangskategorien durch die Vergabe von standardisierten Metadaten Auskunft über die Zugriffsberechtigungen der Daten geben.

Darüber hinaus wird in dem vorliegenden Artikel auf formale Regularien hingewiesen, die es zu beachten gilt. Diese definieren, welcher Personenkreis zur Nachnutzung der Forschungsdaten zugelassen sein soll, zu welchen Zwecken die Daten genutzt werden dürfen und wie eine diesbezügliche Überprüfung stattfinden kann. Hier kommen Nutzungsbedingungen, Lizenzen oder auch Datennutzungsverträge zum Tragen, welche den Zweck und die Bedingungen zur Verwendung der Forschungsdaten transparent und eindeutig bestimmen.

Grundlegend für einen nachhaltigen Zugang zu Forschungsdaten ist die Berücksichtigung der FAIR-Prinzipien, deren Anwendung die Auffindbarkeit, Zugänglichkeit, Interoperabilität und Nachnutzbarkeit von digitalen Ressourcen erleichtern. Aus diesem Grunde werden im Verlauf des Papiers verschiedene Tools und Anwendungsbeispiele zur praktischen Umsetzung der FAIR-Prinzipien und somit relevante Maßnahmen für einen nachhaltigen Datenzugang dargestellt, wie z. B. die Möglichkeiten zum Erhalt eines persistenten Identifiers (PID), der Auswahl geeigneter Lizenzen oder der Relevanz standardisierter Metadatenschemata. Die Inhalte dieses Papiers und die aufgeführten (FAIRen) Anwendungsbeispiele beziehen sich konkret auf grundlegende Maßnahmen des Datenzugangs und stellen eine Einführung in das Thema dar.²

Keywords: Datenzugang, Datenbereitstellungswege, FAIR-Prinzipien, FDZ, Five Safes

² Da es sich bei dem vorliegenden Artikel um eine thematische Einführung handelt, können weitere und vertiefende Praktiken des Forschungsdatenmanagements innerhalb von Forschungsdatenzentren nicht berücksichtigt werden.

Inhaltsverzeichnis

1. Einleitung.....	4
2. Datenbereitstellung	4
2.1. Datenbereitstellungswege.....	4
2.1.1. Download	5
2.1.2. Secure Remote Access.....	5
2.1.3. Gastwissenschaftsarbetsplätze (GWAP)	6
2.2. Die Wahl des Datenbereitstellungswegs	8
2.3. Daten	10
2.4. Datenkataloge und Datenzugang	11
2.5. Formale Regularien des Datenzugangs	14
3. FAIRer Datenzugang.....	16
4. Resümee/Ausblick.....	20
Literaturverzeichnis.....	22

1. Einleitung

Forschungsdaten sollen im Sinne der Open Science-Wissenschaftspraxis transparent beschrieben und möglichst offen zur Nachnutzung bereitgestellt werden, um durch Wiederverwendung und Weiterverarbeitung wissenschaftlichen und somit gesellschaftlichen Fortschritt zu ermöglichen. Insbesondere der Datenzugang und die damit einhergehenden Prozesse spielen in der Open Science-Praxis eine wesentliche Rolle. Innerhalb der unterschiedlichen wissenschaftlichen Disziplinen wird der Datenzugang in unterschiedlichem Ausmaß und auf verschiedene Art ermöglicht. Dies begründet sich in unterschiedlichen Charakteristika der Daten, vornehmlich vorhandener Personenbezogenheit der Daten und/oder urheber- bzw. lizenzrechtlicher Bewertung. Der vorliegende Beitrag soll Forschungsdatenzentren (FDZ) als allgemeine Handreichung für den Datenzugang spezifisch in den Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften dienen. Ziel dieser Handreichung ist, allgemein in das Thema einzuführen und zentrale Informationen und Best Practice Beispiele für die Konzeption möglicher Datenbereitstellungswege im Kontext der Gründung eines FDZ zu liefern.

Zu diesem Zweck werden auf konkrete Möglichkeiten für die Umsetzung des Datenzugangs innerhalb von FDZ hingewiesen. Diese Umsetzungsmöglichkeiten werden von den FAIR-Prinzipien³ ergänzt, die für das Forschungsdatenmanagement (FDM) zentrale Grundsätze formulieren, die beim Aufbau von Services bzw. Forschungsdateninfrastrukturen Berücksichtigung finden sollten. Durch diese Vorgehensweise soll ein Grundverständnis für die Thematik geschaffen und konkrete Umsetzungsmöglichkeiten von FDM-Services anhand der FAIR-Prinzipien aufgezeigt werden.

2. Datenbereitstellung

2.1. Datenbereitstellungswege

Es gibt unterschiedliche Möglichkeiten, Forschungsdaten zur Nachnutzung zur Verfügung zu stellen. Sowohl die Datenart als auch der Nachnutzungszweck ist für die Wahl eines Datenbereitstellungsweges entscheidend. Im Folgenden möchten wir auf die häufigsten Datenbereitstellungswege und deren Besonderheiten eingehen. Die Darstellung beginnt mit dem am wenigsten restriktiven Datenbereitstellungsweg. Die im Anschluss dargestellten Verfahren eignen sich für sensible Forschungsdaten, da durch technische und organisatorische Maßnahmen der Datenzugriff und die Weitergabe der Daten leichter kontrolliert werden kann.⁴

³ <https://www.go-fair.org/fair-principles/>

⁴ Einen aktuellen Überblick über die Datenbereitstellungswege der beim RatSWD akkreditierten FDZ gibt auch der Tätigkeitsbericht von 2019 (RatSWD, 2020b, S. 32ff).

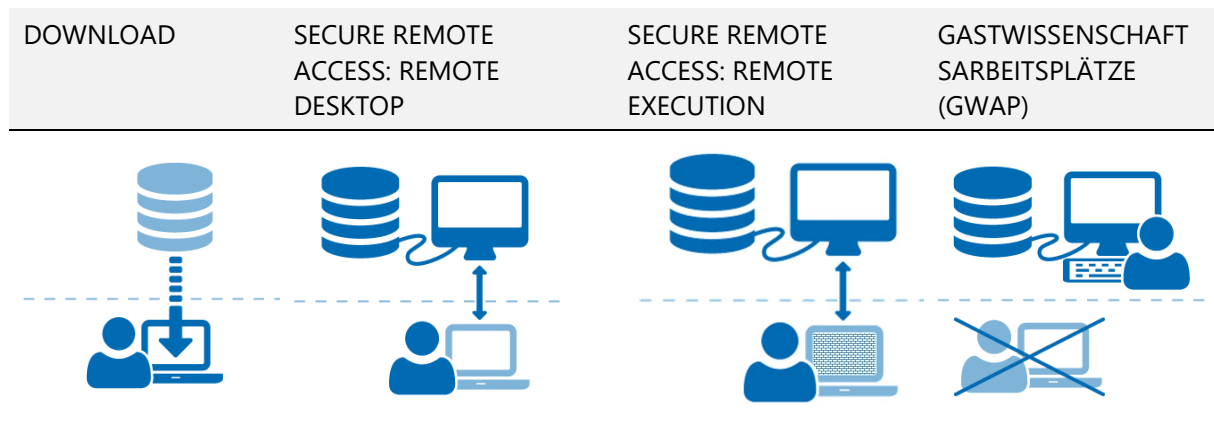


Abbildung 1: Datenbereitstellungswege

Quelle: <https://www.fdz.dzhw.eu/de/datennutzung>, eigene Darstellung

Abbildung 1 stellt häufig genutzte Datenbereitstellungswege dar, beginnend mit dem am wenigsten restriktivsten (Download) bis hin zum restriktivsten (GWAP). Im folgenden Abschnitt werden sie vorgestellt und Auswirkungen für Datenhaltung und Datennutzende beschrieben.

2.1.1. Download

Die offenste Variante des Datenzugangs ist ein *Download* der gewünschten Daten und der dazugehörigen Dokumente aus einem Datenkatalog auf das Endgerät der Datennutzenden. Je nach Datenzugangsstrategie kann dieser Download komplett frei sein oder aber auch eine Registrierung, Authentifizierung⁵ bzw. Abschluss eines Datennutzungsvertrags erfordern. Aktuelle Sicherheitsstandards legen eine verschlüsselte Übermittlung oder eine 2-Faktor-Authentifizierung für einen sicheren Zugriff nahe.⁶

2.1.2. Secure Remote Access

Restriktivere Datenbereitstellungswege belassen die Forschungsdaten entweder auf Servern der FDZ oder Servern von Dienstleistern und gewähren den Datennutzenden Zugriff auf die dort liegenden Daten. Der Oberbegriff *Secure Remote Access*⁷ bzw. *Remote Access* beschreibt verschiedene Verfahren, bei denen ein Zugriff aus der Ferne, d. h. vom Arbeitsplatz der Datennutzenden, vorgenommen wird (Schiller & Welpton, 2015; Schiller et al., 2017, S. 7). Dies bedeutet, dass neben der Datenspeicherung auch die Datenverarbeitung auf Servern des FDZ stattfindet und dort entsprechende Software zur Analyse der Daten (z. B. Stata, SPSS, MAXQDA) bereitgestellt werden muss. Es bedarf für diesen Datenzugang einer besonders geschützten internen IT-Infrastruktur aus einem abgeschirmten Netzwerk von virtuellen Maschinen.

⁵ Diese anzugebenden Daten sollten auf ein Minimum begrenzt werden, die Registrierung selbst sollte elektronisch und für den Nutzenden kostenfrei erfolgen.

⁶ Ein Beispiel hierfür bietet das SOEP https://www.diw.de/de/diw_01.c.601584.de/datenzugang.html#c_695674 oder das Forschungsdatenzentrum des Robert-Koch-Instituts <https://www.rki.de/SharedDocs/FAQ/FDZ/FAQ-Liste.html#FAQId13465420>.

⁷ Eine ausführliche Beschreibung von verschiedenen Verfahren des Remote Access findet sich bei Schiller et al. (2017) und RatSWD (2019). Auch aktuelle Ansätze wie "FDZ-im-FDZ" sind dort beschrieben.

Bei der Datenbereitstellung durch das *Remote Desktop*-Verfahren⁸ können sich Nutzende mittels einer Software auf den Server des FDZ einwählen, die Daten einsehen und mit der auf dem Server bereitgestellten Software analysieren. Nutzende selbst können keine Dateien herunterladen oder importieren. Alle Dateien, die Nutzende im- oder exportieren möchten, werden vorher durch das FDZ geprüft (In-/Outputkontrolle).⁹ Das FDZ-Personal stellt den Datennutzenden nach erfolgter Prüfung die In- und Outputs bereit.

Bei der restriktiveren Variante, dem *Remote Execution*-Verfahren, die auch als *Kontrollierte Datenfernverarbeitung (KDFV)* oder *Fernrechnen* bezeichnet wird, können sich Nutzende ebenfalls mit einer Software auf den Server des FDZ einwählen. Im Unterschied zu Remote Desktop haben sie aber keinen Einblick in die Daten selbst. Auf diese Weise können durch die Datennutzenden besonders sensible Daten analysiert werden, die ihnen jedoch aus Datenschutzgründen nicht offenbart werden dürfen. Skripte oder Syntax zur Modifizierung bzw. Aufbereitung und Analyse der Daten werden ohne Sicht auf die Daten an das FDZ gesendet und anschließend durch das FDZ ausgeführt. Beim Ausführen der Skripte sowie der Übermittlung der Analyseergebnisse sind unterschiedliche Grade an Automatisierung möglich. Alle Dateien, die Nutzende im- oder exportieren möchten, werden vorher durch das FDZ geprüft (In-/Outputkontrolle). Meist werden die Outputkontrollen softwaregestützt durchgeführt, aber in der Regel zusätzlich intellektuell geprüft. Das FDZ-Personal stellt den Datennutzenden anschließend die In- und Outputs bereit.

2.1.3. Gastwissenschaftsarbeitsplätze (GWAP)

Bei dieser Form des Datenzugangs (auch als Gastaufenthalte, On-Site oder Secure Data Center bezeichnet) erfolgt der Datenzugriff vor Ort im FDZ. Dieser Datenbereitstellungsweg sowie Remote Execution stellen damit bezüglich der technischen und organisatorischen Kontrollen die restriktivsten Varianten des Datenzugangs dar. Wie bei Secure Remote Access findet die Datenspeicherung und Datenverarbeitung auf zentralen Servern des FDZ statt, Nutzende arbeiten aber nicht am eigenen Arbeitsplatz, sondern an speziell ausgestatteten Arbeitsplätzen vor Ort im FDZ. Diese sind in der Regel mit einem Rechner ohne Internetzugang und ohne funktionsfähige USB-Anschlüsse, Laufwerke etc. ausgestattet. Überwiegend gibt es weitere Regularien für die Räumlichkeiten wie das Verbot von Fotografien, Abschreiben, Mobiltelefonen, Laptops und ähnliches. Nutzende können die Daten einsehen und mit der bereitgestellten Software analysieren. Es können durch die Nutzenden keine Daten ex- oder importiert werden. Alle Dateien, die Nutzende im- oder exportieren lassen möchten, werden vorher geprüft (In-/Outputkontrolle). Eine nächste Ausbaustufe von GWAP kann die Vernetzung von GWAP verschiedener Institutionen sein.¹⁰ Die im Pilotprojekt RDCnet von KonsortSWD erarbeiteten Mindeststandards für eine Vernetzung von GWAP bezüglich

⁸ Im Angelsächsischen wird eher der Begriff (*virtual*) *data enclave* genutzt.

⁹ Die Outputkontrolle kann unterschiedlich gestaltet sein, für Beispiele siehe RatSWD (2019).

¹⁰ In KonsortSWD wird etwa an einem Pilotprojekt gearbeitet, das bestehende GWAP miteinander vernetzen soll (<https://www.konsortswd.de/konsortswd/das-konsortium/services/rdcnet/>). Die Datenzugänge der Forschungsdatenzentren des IAB und von GESIS sind zudem bereits über das IDAN-Projekt vernetzt (<https://www.gesis.org/angebot/daten-aufbereiten-und-analysieren/forschungsaufenthalte>).

Raumsicherheit und Kriterien hinsichtlich der technischen Umgebung können auch als Richtlinien für einfache GWAP verwendet werden (Murray & Goebel, 2022, S. 11ff).

Generell sollte der am wenigsten restriktive Datenbereitstellungsweg gewählt werden, der für die spezifischen Forschungsdaten bezüglich ihrer datenschutzrechtlichen und forschungsethischen Sensitivität möglich ist. Zusätzlich sind Ressourcenfragen zu berücksichtigen – restriktivere Datenbereitstellungswege gehen oft auch mit höheren Kosten und mehr Bearbeitungsaufwand einher.¹¹ Bei der Planung mehrerer Datenbereitstellungswege sollte eruiert werden, ob dieselben technischen Umgebungen für mehrere Datenbereitstellungswege genutzt werden können. Dies wäre beispielsweise für Remote Desktop und GWAP denkbar, wodurch Kosten eingespart werden könnten. Zu bedenken ist zudem, dass die aufgebaute Infrastruktur für die Zukunft sicher finanziert sein sollte, da Daten entsprechend der Datenbereitstellungsstrategie für einen bestimmten Bereitstellungsweg aufbereitet wurden und ggf. nicht einfach über andere Infrastrukturen zugänglich gemacht werden können.

Aktuell erheben nur einzelne FDZ (nicht kostendeckende) Beträge für die Datennutzung. In der Regel ist die Datennutzung in Deutschland bisher weitestgehend kostenfrei (siehe Schiller et al., 2017). Es gibt jedoch bezogen auf die Datennutzung und in der Datenaufnahme auch Diskussionen und praktische Umsetzungen von Kostenbeteiligungen der Datennutzenden beziehungsweise der Datengebenden. Eine Kostenbeteiligung kann für kostenintensive Datenbereitstellungswege (wie etwa Remote Execution, Remote Access, individuelle aufwändige Datenaufbereitungen speziell für einzelne Datennutzende) oder für aufwändige Datenaufnahmeprozesse und Langzeitarchivierung infrage kommen. Wenn ein FDZ ein transparentes Kostenmodell zur Verfügung stellt, kann die Übernahme dieser Kosten durch die Datengebenden ggf. bei Fördereinrichtungen beantragt werden.¹² Eine Möglichkeit, um die Kosten für die Infrastruktur zur Datenbereitstellung besser kalkulieren zu können, ist die Konsultation externer Dienstleistungsunternehmen.

Einrichtungen ohne geeignete Forschungsdateninfrastruktur, für die ein Aufbau eines eigenen Forschungsdatenzentrums (vorerst) nicht infrage kommt, können alternativ auch bestehende Services zur Archivierung und Veröffentlichung nutzen (z. B. das generische Repositorium RADAR¹³, SowiDataNet für sozial- und wirtschaftswissenschaftliche Daten¹⁴ oder die

¹¹ Für eine Beschreibung der Aufwände für Remote Access Umgebungen siehe Schiller et al. (2017, S. 22).

¹² So übernimmt etwa die DFG ggf. "Kosten, die anfallen, um den Zugang zu Forschungsdaten zu erlangen oder die im Projekt entstandenen Forschungsdaten so zu bearbeiten und aufzubereiten, dass diese von anderen nachgenutzt werden können oder Kosten, die für die Überführung der Daten in ein öffentliches Repositorium anfallen. Diese Kosten können als projektspezifische Kosten beantragt und bewilligt werden. Hierzu zählen beispielsweise personelle Aufwände für die Aufbereitung oder den Transfer von Daten in existierende Repositorien, aber auch Soft- und Hardware, die für diese Zwecke benötigt werden.", siehe Informationen zu beantragbaren Mitteln der DFG:

https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/forschungsdaten/beantragbare_mittel/index.html

¹³ <https://www.radar-service.eu>

¹⁴ <https://data.gesis.org/sharing/#!Home>

Forschungsdatenzentren des RatSWD¹⁵ etc.). Für die Aufnahme von Daten bestehen unterschiedliche Bedingungen, die es bei den betreibenden Einrichtungen zu erfragen gilt.

2.2. Die Wahl des Datenbereitstellungswegs

Auf welchem Weg die Daten bereitgestellt werden können, hängt maßgeblich von den Daten selbst ab, etwa wie sensibel diese sind (z. B. im Sinne von Personenbezogenheit), welche rechtlichen Aspekte mit den Daten einhergehen (z. B. Urheber- und Lizenzrechte, informierte Einwilligung), forschungsethische Aspekte (z. B. mögliche Schädigungen der Befragten in psychischer oder ökonomischer Hinsicht) oder auch die Größe der Datenmenge. Sofern keine rechtlichen oder ethischen Gründe vorliegen, die einen offenen Datenzugriff einschränken, sollten diese als *open data*¹⁶ zur Verfügung gestellt werden.

Daten aus den Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften unterliegen jedoch oft stärkeren Einschränkungen, da Daten mit Personenbezug oft einen höheren Bedarf an technisch-organisatorischer Kontrolle im Datenzugang erfordern.

Der Schutz von sensiblen Forschungsdaten kann durch eine Vielzahl von Sicherungsmechanismen erreicht werden. Eine hilfreiche Übersicht bietet der Portfolioansatz von Desai, Ritchie und Welpton (2016). Dieser Portfolioansatz setzt sich aus den sogenannten Five Safes-Dimensionen zusammen (Schiller et al., 2017, S. 5f):

- *safe people* (geschulte Personen): Kann man den Forschenden vertrauen, dass sie die Daten in angemessener Weise verwenden?
- *safe projects* (geprüfte Projekte): Ist die Verwendung der Daten angemessen?
- *safe settings* (technisches Umfeld): Schränkt die Zugriffsmöglichkeit eine unbefugte Nutzung ein?
- *safe outputs* (kontrollierte Ergebnisse): Sind die Ergebnisse nicht schutzbedürftig?
- *safe data* (anonymisierte Daten): Besteht ein Offenlegungsrisiko bei den Daten selbst?

Forschende haben als geschulte Personen (*safe people*) das Wissen, die Fähigkeiten und die Anreize, die Daten angemessen zu speichern und zu verwenden (Desai et al., 2016, S. 9). So kann etwa die Anstellung an einer wissenschaftlichen Einrichtung als Indikator oder die Akzeptanz von Nutzungsbedingungen ausreichend sein. Je nach Daten kann auch eine Schulung zum Umgang mit den Daten und der technischen Umgebung sinnvoll sein. Geprüfte Projekte (*safe projects*) berücksichtigen rechtliche, moralische und ethische Überlegungen zur Verwendung der Daten (Desai et al., 2016, S. 8). In diesem Fall kann z. B. das Projektvorhaben erfragt werden, für das die Daten benötigt werden. Sichere (technische) Umgebungen (*safe settings*) referenzieren auf die Kontrollmöglichkeiten über die Daten, welche etwa von einem Download ohne Restriktionen bis zu einem Gastaufenthalt oder Remote Execution gehen können. Kontrollierte Ergebnisse (*safe outputs*) beziehen sich auf die Ergebnisse aus

¹⁵ <https://www.ratswd.de/forschungsdaten/fdz>

¹⁶ https://www.forschungsdaten.org/index.php/Open_Access

Datenanalysen. Insbesondere bei weniger anonymisierten Daten ist eine Prüfung der Outputs notwendig. Anonymisierte Daten (*safe data*)¹⁷ zielen auf eine Minimierung des De-Anonymisierungrisikos ab, welches je nach Anonymisierungsgrad variieren kann.

Der Schutz der Daten (und damit der Untersuchungspersonen) kann somit durch die Kontrolle der Personen, der Projekte, der Umgebung, der Ergebnisse und/oder der Daten erreicht werden. Die Five Safes-Dimensionen können dabei helfen, eine adäquate Datenzugangsstrategie zu entwickeln. Der Ansatz bleibt absichtlich vage in der Operationalisierung, gibt aber einen Eindruck davon, wie über das Verändern der einzelnen Dimensionen (analog zum Hoch- und Runterfahren von Reglern) Daten sicher bereitgestellt werden können.

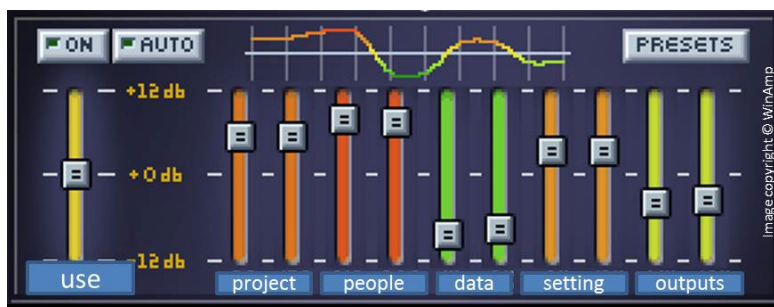


Abbildung 2: Five Safes (Desai et al., 2016, S. 5) basierend auf McEachern, 2015

Abbildung 2 stellt das Zusammenspiel der Dimensionen konzeptuell dar. So können etwa stark anonymisierte Daten (*safe data*) über ein Umfeld mit weniger technischen Einschränkungen bereitgestellt werden. Gleichzeitig können wenig anonymisierte Daten über restriktivere Datenbereitstellungswege (*Safe Settings*) bereitgestellt werden. Sind die Nutzenden vorab geschult worden (*safe people*), können ebenso sensiblere Daten bereitgestellt werden.¹⁸ Je nach Datentyp und Sensibilität der Daten kann eine andere Datenstrategie angemessen sein. Diese Strategie für Daten oder ganze Studienreihen sollte in Zusammenarbeit mit betrieblichen Datenschutzbeauftragten entwickelt werden.

Folgende Tabelle 1 stellt beispielhaft mögliche Ausgestaltungen der Safes dar.¹⁹ Die Vertrauenswürdigkeit des Projekts (*safe project*) könnte jeweils durch die Abfrage von Informationen zum Forschungsvorhaben bei Registrierung oder Antragstellung geprüft werden.

¹⁷ Für den eigentlichen Prozess der Anonymisierung von Daten gibt es eine Vielzahl von Möglichkeiten (Müller, Blien, Knoche und Wirth (1991)); eine der gebräuchlichsten ist, die statistische Aggregation (z. B. Aggregation von geografischen Angaben wie Ortsname zu Kreisname oder Alter zu Altersgruppen). Im Grundsatz gilt: Je stärker aggregiert wird, desto stärker die Anonymisierung und desto geringer das Risiko der Re-Identifikation. Gleichzeitig sinkt jedoch auch der Informationsgehalt der Daten für die wissenschaftliche Nutzung. Konzepte werden in Ebel (2015), Ebel und Meyermann (2015) sowie Eisentraut (2018) beschrieben.

¹⁸ Einzelne FDZ verpflichten die Nutzenden vorab zu einer Schulung, andere sehen die wissenschaftliche Tätigkeit und/oder das Akzeptieren von detaillierten Nutzungsbedingungen als ausreichenden Nachweis an.

¹⁹ Auch die Operationalisierung einer wissenschaftlichen Anstellung (*safe people*) bei den Scientific Use Files ist nur als beispielhaft zu verstehen und soll hier nicht als zwingende Bedingung aufgefasst werden (siehe beispielsweise externe Promovierende mit Anstellungen außerhalb der Wissenschaft).

	SAFE DATA / ANONYMISIERUNGSGRAD	SAFE SETTING / SICHERE (TECHNISCHE) UMGEBUNGEN	SAFE PEOPLE / GESCHULTE PERSONEN	SAFE OUTPUTS / KONTROLLIERTE OUTPUTS	SAFE PROJECT / GEPRÜFTE PROJEKTE
1	Public Use File (PUF)	Download	keine Kontrolle	keine Kontrolle, ggf. Nutzungsbedingungen	keine Kontrolle, ggf. Hinweise
2.1	Scientific Use File (SUF)	Gesicherter Download	Wiss. Anstellung	keine Kontrolle, ggf. Nutzungsbedingungen	Erfragung von Information
2.2	Scientific Use File (SUF)	Remote Access als Remote Desktop	Wiss. Anstellung	In- und Outputkontrolle, ggf. Nutzungsbedingungen	Erfragung von Information
3.1	Secure Use File (SecUF)20	Remote Access als Remote Desktop mit Safe Room (Vernetzte GWAP)	Wiss. Anstellung, Schulung	In- und Outputkontrolle ggf. Nutzungsbedingungen	Erfragung von Information
3.2	Secure Use File (SecUF)	Remote Access als Remote Execution	Wiss. Anstellung, Schulung	In- und Outputkontrolle ggf. Nutzungsbedingungen	Erfragung von Information
3.3	Secure Use File (SecUF)	Gastaufenthalt	Wiss. Anstellung, Schulung	In- und Outputkontrolle ggf. Nutzungsbedingungen	Erfragung von Information

Tabelle 1: Übersicht beispielhafter Kombinationen der Five Safes

Quelle: Eigene Adaption angelehnt an Schiller et al. (2017, S. 5)

Weitere Überlegungen und Beispiele aus der Praxis, insbesondere in Bezug auf Remote Access-Lösungen finden sich in Schiller et al. (2017).

Die Datenzugangsstrategie sollte vorab transparent kommuniziert werden, beispielsweise als Dokument oder als Teil der FDZ-Website. So sollten etwa die Nutzungsbedingungen²¹ bzw. ein Vertrag verfügbar sein, damit die Nutzenden die Nutzungsbedingungen vorab einsehen können.

2.3. Daten

Die in das FDZ aufgenommenen Daten können in der Regel nicht direkt zur Sekundärnutzung bereitgestellt werden. So müssen diese gegebenenfalls aufbereitet, anonymisiert und um standardisierte Metadaten²² angereichert werden. Meist werden die Daten den Nutzenden als standardisierte Datenpakete²³ bereitgestellt. Teilweise werden die Daten aber auch erst *on demand* auf Anfrage der Nutzenden zusammengestellt.²⁴

²⁰ Definition siehe Abschnitt Daten.

²¹ Diese können etwa je nach Datenbereitstellungsweg oder auch für einzelne Datensätze unterschiedlich sein.

²² <https://www.forschungsdaten.org/index.php/Metadaten>

²³ Ein Datenpaket, das für allgemeine Forschungszwecke kuratiert und aufbereitet wurde.

²⁴ Beispielsweise kann das SOEP on demand zusammengestellt werden: <https://paneldata.org/>

Die Daten wiederum können auf unterschiedliche Arten aufbereitet werden - je nach Zielgruppe und Nutzungszweck. Die folgende Beschreibung orientiert sich an der Terminologie zur Unterscheidung von Daten der EU²⁵ (Schiller et al., 2017, S. 4).

- Public Use File (PUF)
- Campus Use File (CUF)
- Scientific Use File (SUF)
- Secure Use File (SecUF)

Public Use Files (PUF) sind Daten, die der gesamten Öffentlichkeit zur Verfügung gestellt werden können, für die es also keinerlei rechtliche oder ethische Bedenken gibt (*open data*). Im Bereich der Sozialwissenschaften sind dies oft stark anonymisierte Daten oder Strukturdateien. *Campus Use File (CUF)* bezeichnet Daten, die nur für die wissenschaftliche Lehre bereitgestellt werden. Oft sind diese Daten ebenfalls stark anonymisiert. *Scientific Use File (SUF)* sowie *Secure Use File (SecUF)* bezeichnen Daten, die nur für die wissenschaftliche Forschung zur Verfügung gestellt werden, wobei SecUF nicht oder schwach anonymisiert oder nur pseudonymisiert sind. In der Praxis ist nicht die gesamte Terminologie verbreitet - oft werden auch SecUF als SUF bezeichnet, da sie auch für die Wissenschaft bereitgestellt werden.

Welche Art von Daten zur Nachnutzung durch das FDZ bereitgestellt werden können bzw. sollen, entscheidet das FDZ mit dem Datengebenden. Zu beachten sind hier auch rechtliche und ethische Aspekte, beispielsweise der *Informed Consent*, also die Einwilligung über die Verarbeitung der erhobenen Daten, die Teilnehmende geben, bevor eine Erhebung, z. B. eine Befragung, stattfindet. Sofern keine rechtlichen oder ethischen Einschränkungen bestehen, bietet es sich an, die Daten möglichst als Public Use File zu veröffentlichen. Oft beinhaltet der *Informed Consent* etwa aber nur die Einwilligung für die Weitergabe von Daten für eine wissenschaftliche Nachnutzung²⁶. Aus einem Datensatz können grundsätzlich auch mehrere Varianten erstellt werden, etwa ein Campus Use File für die Lehre sowie Scientific Use File für die Forschung, um Bedarfe unterschiedlicher Zielgruppen zu erfüllen.

2.4. Datenkataloge und Datenzugang

Die erstellten Daten können beispielsweise über ein Bestellsystem zugänglich gemacht werden – häufig Datenkatalog oder Recherchesystem genannt. In solchen Datenkatalogen werden die Forschungsdaten mittels Metadaten beschrieben. Die Metadaten beinhalten zentrale Informationen über die Forschungsdaten, die einerseits den potenziellen Datennutzenden notwendige Auskünfte über die Forschungsdaten geben und andererseits den automatisierten Austausch von Informationen zwischen Datenkatalogen ermöglichen. In den Sozial-, Verhaltens-, Wirtschafts- und Gesundheitswissenschaften ist die Data Documentation Initiative (DDI) neben anderen ein international häufig verwendeter fachspezifischer

²⁵ COMMISSION REGULATION (EU) No 557/2013

²⁶ Für eine Weitergabe personenbezogener Daten müssen die Befragten explizit durch den Informed Consent aufgeklärt worden sein und zugestimmt haben.

Metadatenstandard, der aufgrund seiner Komplexität in der Lage ist, den vollständigen Datenlebenszyklus mittels XML zu beschreiben.²⁷

Viele FDZ, die größtenteils schon seit Jahren bestehen, betreiben Eigenentwicklungen von Datenkatalogen, um ihre Forschungsdaten anzuzeigen und zu veröffentlichen. Die Beschreibung der Daten sollte auch hier unbedingt einem Metadatenstandard folgen, damit gewährleistet ist, dass die Daten und Metadaten mit anderen Systemen austauschbar sind und somit u. a. den FAIR-Prinzipien genügen. Darüber hinaus ist es wichtig, dass die Metadaten über Schnittstellen (z. B. OAI-PMH, Rest API) abgerufen werden können (siehe auch Abschnitt FAIRer Datenzugang). Daher kann die Nutzung bereits bestehender Softwarelösungen²⁸ sinnvoll sein, die bereits zahlreiche Anforderungen wie Versionierung und Standards (z. B. Metadatenstandards wie Dublin Core, DataCite, DDI, schema.org) erfüllen und zum Teil als Open Source durch eine größere Community weiterentwickelt werden. Diese sind etwa Dataverse²⁹, DSpace³⁰, CKAN³¹ oder Dryad³². Diese Anwendungen stellen die Informationen zu den Forschungsdaten strukturiert und nach etablierten (Metadaten)standards bereit. Insbesondere werden Daten und Metadaten mithilfe von kontrollierten Vokabularen, Ontologien oder Thesauri strukturiert, die entweder generischer Natur sind oder für verschiedene Fachdisziplinen relevant sind. Durch offene Schnittstellen können Informationen darüber hinaus automatisiert und maschinenlesbar ausgetauscht werden. So sind auch schon einige Aspekte der FAIR-Prinzipien (siehe Abschnitt FAIRer Datenzugang) automatisch berücksichtigt.

Um das FAIR-Prinzip der Accessibility (siehe auch Abschnitt FAIRer Datenzugang) zu gewährleisten, empfiehlt sich für die Beschreibung der Zugangskategorien im Datenkatalog ebenfalls die Nutzung eines Standards. Gängige Kategorisierungen und Schemata wie etwa der CESSDA-Standard³³ empfehlen vier Kategorien:

- Open access / Open Access
- Access for registered users (safeguarded) / Zugriff für authentifizierte Nutzende (geschützt)
- Restricted access / Eingeschränkter Zugang
- Embargo / Sperrfrist

Open Access beschreibt dabei die Möglichkeit des direkten Zugriffs ggf. nach Akzeptieren von Nutzungsbedingungen via Opt-In-Verfahren. *Access for registered users (safeguarded)* bedeutet, dass der Datenzugriff nur für Nutzende möglich ist, wenn sie sich im FDZ registriert

²⁷ <https://ddialliance.org/>

²⁸ https://www.forschungsdaten.org/index.php/Repository_Software

²⁹ <https://dataverse.harvard.edu/> Bisher gibt es nur eine englischsprachige Nutzeroberfläche, im Rahmen der EOSC sollen jedoch weitere Sprachen angelegt werden

³⁰ <https://dspace-cris.4science.cloud/>

³¹ <https://ckan.org/>

³² <https://datadryad.org>

³³ <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/6.-Archive-Publish/Publishing-with-CESSDA-archives/Access-categories>

haben. Die Daten dieser beiden Zugangskategorien enthalten keine direkten Identifikatoren, aber es kann ein Risiko der Offenlegung durch die Verknüpfung von indirekten Identifikatoren bestehen, sodass bei diesem offenen Datenzugang besonderes Augenmerk auf dem Datenschutz³⁴ liegen muss. *Restricted Access* bedeutet, dass der Zugriff derart begrenzt ist, dass er nur auf Antrag nach manueller Prüfung gewährt wird. Diese Zugangskategorie gilt in der Regel für sensible Daten, die möglicherweise personenbezogene Informationen enthalten. Forschungsdaten der letzten Kategorie *Embargo* sind mit einem zeitlich begrenzten Embargo versehen. Während des Embargos wird nur die Beschreibung (Metadaten) des Datensatzes veröffentlicht. Die Daten selbst werden erst zu einem späteren fest definierten Zeitpunkt verfügbar sein. Um dem **A**, also der Accessibility innerhalb der FAIR-Prinzipien gerecht zu werden, sollte eine Beschreibung der Zugangskategorien öffentlich zugänglich und mögliche Verfahren zur Prüfung von Zugriffsberechtigungen für die Datennutzenden transparent sein. Auch die mit der Datennutzung einhergehenden Nutzungsbedingungen (siehe Abschnitt Formale Regularien des Datenzugangs) sollten vorab verfügbar sein. Unabhängig davon, welche Zugangskategorien bestehen oder über welchen technischen Datenbereitstellungsweg die Daten zugänglich sind, sollten die Metadaten immer öffentlich zugänglich sein, auch wenn die Daten selbst (z. B. aus datenschutzrechtlichen Gründen) nie oder erst zu einem späteren Zeitpunkt zur Sekundärnutzung bereitgestellt werden dürfen. In der internationalen Community und auch insbesondere in der deutschen Community werden die Bezeichnungen der Zugangskategorien aktuell nicht einheitlich verwendet.

Zur Umsetzung des FAIR-Prinzips der Findability empfiehlt sich unbedingt die Vergabe eines persistenten Identifiers (PID)³⁵, mit dem direkt auf die Quelle der Forschungsdaten verwiesen wird. Hierdurch werden die Daten eindeutig identifizierbar und zitierbar (siehe auch Abschnitt FAIRer Datenzugang). In den Sozialwissenschaften wird häufig ein Digital Object Identifier (DOI) verwendet. Um einen DOI zu erhalten, müssen bestimmte Pflichtangaben gemacht werden, d. h. den Daten werden an dieser Stelle des Dokumentationsprozesses Informationen nach einem strukturierten Schema beigefügt. DataCite ist die zentrale Anlaufstelle zur DOI-Vergabe mit einem für Forschungsdaten spezialisierten Metadatenschema. Beispielsweise da|ra³⁶ ist eine deutsche DOI-Registrierungsagentur für Sozial- und Wirtschaftsdaten, die Forschungsdaten als Mittler bei DataCite registriert. Koch, Akdeniz, Meichsner, Hausstein und Harzenetter (2017, S. 10) definieren folgende Pflichtfelder für die Registrierung eines Datensatzes bei da|ra, die für die Auffindbarkeit unabdingbar sind (siehe Tabelle 2):

³⁴ Als vertiefende Literatur sei die folgende Lektüre empfohlen: Kreuzer und Lahmann (2021), Depping (2021) sowie RatSWD (2020a).

³⁵ https://www.forschungsdaten.org/index.php/Persistent_Identifier

³⁶ <https://www.da-ra.de/>

DAJRA PROPERTY (EN)	BEISPIEL ALLBUS	BEISPIEL NEPS
Resource Type	Dataset	Dataset
Title	Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS 1986 - Non-Response-Studie	NEPS-Startkohorte 6: Erwachsene (SC6 12.1.0)
Creators	GESIS - Leibniz-Institut für Sozialwissenschaften	Artelt, Cordula (Leibniz Institute for Educational Trajectories, Germany); NEPS, National Educational Panel Study, Bamberg (Germany)
DOI	https://doi.org/10.4232/1.1669	https://doi.org/10.5157/NEPS:SC6:12.1.0
dataURL	https://search.gesis.org/research_data/ZA1669?doi=10.4232/1.1669	https://www.neps-data.de/default.aspx?tabid=5247
Publication Date	2004	2021-12-09
Availability	Download	Download

Tabelle 2: da|ra Pflichtfelder für DOI-Registrierung

Diese Pflichtfelder oder die Pflichtfelder des Metadatenschemas der DOI-Registrierungsagentur DataCite³⁷ sollten in jedem Fall zur Beschreibung von Forschungsdaten genutzt werden, da sie einem internationalen Standard folgen und so auch über das eigene Datenportal hinaus interpretiert und verarbeitet werden können - demzufolge gemäß der FAIR-Prinzipien leichter **f**indable und **a**ccessible sind. Darüber hinaus werden weitere Felder (beispielsweise Schlagwörter, Beschreibung, geographische Abdeckung) dringend empfohlen (Koch et al., 2017, S. 11). Eine DOI-Registrierung kann manuell über eine Eingabemaske oder über eine API erfolgen. In einigen Diensten wie Dataverse, CKAN oder Dryad ist eine automatisierte Vergabe von DOI bereits integriert bzw. konfigurierbar. Eine automatisierte Registrierung über eine API erhöht die Integrität der Informationen, weil sie weniger fehleranfällig ist. Im Datenportal sollten die Angaben für die Zitation der Daten transparent dokumentiert sein. Mit einer Registrierung bei da|ra können die Einträge der Forschungsdaten darüber hinaus von verschiedenen Suchportalen³⁸ gecrawlt und dementsprechend dort aufgeführt werden. Wie bereits oben erwähnt bietet es sich daher der Einfachheit halber an, auf bestehende Softwarelösungen zurückzugreifen, die diese Metadatenstandards berücksichtigen.

2.5. Formale Regularien des Datenzugangs

Forschungsdaten in den Sozialwissenschaften werden zur Nachnutzung durch andere oftmals nur für die wissenschaftliche Forschung bereitgestellt. Daher muss in diesen Fällen im ersten Schritt die Wissenschaftlichkeit des Vorhabens vor der Bereitstellung geprüft werden. Dies passiert entweder über Selbstauskunft oder via Prüfung der angegebenen Informationen durch FDZ-Mitarbeitende. Die Wissenschaftlichkeit der Forschung sollte dabei über den Kreis der Nutzenden und über den Zweck der Nutzung definiert werden. Meist wird dieser über die

³⁷ <https://support.datacite.org/docs/datacite-metadata-schema-v44-mandatory-properties>

³⁸ beispielsweise <https://datasearch.gesis.org>, <https://www.base-search.net/>, <https://data.mendeley.com/> oder <https://sociohub-fid.de/>

Zugehörigkeit der Person zu einer wissenschaftlichen Institution (z. B. Hochschule, außeruniversitäre Forschungseinrichtung) operationalisiert (Kreis der Nutzenden), in deren Rahmen das Forschungsvorhaben ausgeübt wird.³⁹ Darüber hinaus sollte vor allem der Nutzungszweck beachtet werden, da prinzipiell auch eine Privatperson wissenschaftlich arbeiten kann, indem sie etwa eine wissenschaftliche Publikation erarbeitet. Bestehen keinerlei rechtliche oder ethische Verpflichtungen und die Daten können für die Gemeinheit zur Verfügung gestellt werden, muss kein wissenschaftlicher Nutzungszweck beachtet werden.

Für die konkrete Datenbereitstellung müssen Nutzungsbedingungen zwischen den Datennutzenden und dem FDZ sowie ggf. mit den Datengebenden vereinbart werden. Bei anonymisierten Daten, die keinerlei Personenbezug (mehr) aufweisen, muss die DSGVO nicht herangezogen werden, allerdings müssen auch hier forschungsethische Aspekte (z. B. mögliche psychische oder rechtliche Schädigungen der Untersuchungsteilnehmenden) berücksichtigt werden. Prinzipiell können z. B. Standardlizenzen⁴⁰ für *open data* oder Lizenzen für die wissenschaftliche Nutzung⁴¹ als Nutzungsbedingungen eingesetzt werden. Die Art und Ausgestaltung der Regularien des Datenzugangs obliegen dabei dem FDZ. Gleichwohl sind bestimmte rechtliche Rahmenbedingungen (z. B. DSGVO bei personenbezogenen Daten, rechtliche Verpflichtungen gegenüber Datengebenden) zu beachten. Bei Daten, die potenziell personenbezogen sind, empfiehlt sich der Abschluss eines Datennutzungsvertrages in Schriftform, der ein vertrauenswürdigen rechtliches Rahmenwerk für den Zugriff auf die Forschungsdaten bildet. Hier können auch die Anforderungen an technische und organisatorische Maßnahmen (z. B. Verschlüsselung von USB-Sticks, Meldepflichten) geregelt werden, die für einen sicheren und verantwortungsvollen Datenzugriff notwendig sind. Ein harmonisierter Datennutzungsvertrag, der auf einer Vertragsharmonisierung von 20 beim RatSWD akkreditierten FDZ basiert, wurde im Rahmen von KonsortSWD entwickelt und kann als Vorlage dienen (Schallaböck, Hoffstätter, Buck & Linne, 2022).

Die Ausgestaltung der Nutzungsbedingungen kann auch innerhalb eines FDZ variieren. So können für bestimmte Datenbestände Standardlizenzen genutzt oder für potenziell personenbezogene Datenbestände wiederum ein Datennutzungsvertrag abgeschlossen werden. Wie auch immer die Nutzungsbedingungen letztendlich ausgestaltet werden, sind sie von besonderer Relevanz für die Erfüllung der **Accessibility**-Kriterien aus den FAIR-Prinzipien. Es muss zu jedem Zeitpunkt transparent und nachvollziehbar sein, wer und unter welchen

³⁹ Im Fall von Qualifikationsarbeiten werden ggf. auch die Anstellungen an wissenschaftlichen Einrichtungen der Betreuenden der sich qualifizierenden Personen herangezogen. Die Anstellung an einer wissenschaftlichen Einrichtung darf aber nicht als zwingendes rechtlich vorgeschriebenes Kriterium interpretiert werden, sondern nur als Indikator für eine wissenschaftliche Nutzung.

⁴⁰ <http://ufal.github.io/public-license-selector/> oder <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/6.-Archive-Publish/Publishing-with-CESSDA-archives/Licensing-your-data>

⁴¹ So könnte geprüft werden, ob etwa die "License for scientific purposes ("Scientific Use License")" des PsychArchives genutzt oder als Vorlage für eine eigene Lizenz verwendet werden kann: <http://doi.org/10.23668/psycharchives.4988>. Auch die Nutzungsbedingungen für den Datendownload bei GESIS Leibniz-Institut für Sozialwissenschaften könnten als Vorlage herangezogen werden: https://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/_bgordnung_bestellen/2018-05-25_Benutzungsordnung_GESIS_DAS.pdf.

Bedingungen Zugriff auf die Daten haben darf. Auch vor dem Hintergrund einer bestmöglichen Interoperabilität verschiedener Datensätze empfiehlt sich die Nutzung eines Standards, der in einem maschinenlesbaren Metadatenformat abgebildet wird. Somit können Computersysteme erkennen, welche Forschungsdaten unter denselben Bedingungen und von welchen Zielgruppen weiterverwendet werden dürfen.

Ein weit verbreitetes offenes Standard-Lizenzmodell wird durch die gemeinnützige Organisation "Creative Commons"⁴² bereitgestellt. Die alleinige Nutzung von Creative Commons-Lizenzen für sozialwissenschaftliche Forschungsdaten in Deutschland ist jedoch nicht empfehlenswert. Eine Einschränkung auf die wissenschaftliche Nutzung der Daten, die in den Sozialwissenschaften und angrenzenden Disziplinen oft Bedingungen für eine Nachnutzung ist, ist beispielsweise nicht möglich. In diesem Fall könnte man beispielsweise zusätzlich den wissenschaftlichen Nutzungszweck erfragen oder auf die obigen Möglichkeiten, wie etwa den Einsatz eines Datennutzungsvertrages, zurückgreifen.

3. FAIRer Datenzugang

Die FAIR-Prinzipien haben sich länder- und disziplinübergreifend als feste Größe zur Beurteilung der Nachnutzbarkeit von Forschungsdaten etabliert. Sie sind Gegenstand des aktuellen Forschungsdiskurses und wurden beispielsweise von der Europäischen Kommission in den Förderrichtlinien⁴³ EU Horizon 2020⁴⁴ aufgenommen und müssen bei einem Antrag innerhalb dieser Förderlinie verpflichtend berücksichtigt werden. Ziel der FAIR-Prinzipien ist, Forschungsdaten für neue Nutzungsszenarien im Rahmen des rechtlich und technisch Möglichen für Menschen und Maschinen optimal zu verwalten und nachnutzbar zu machen. Dies bedeutet allerdings nicht, dass jeder Datensatz uneingeschränkt oder kostenfrei zugänglich ist, sondern lediglich, dass die Voraussetzungen und Datenbereitstellungswege für den Zugriff auf Datenbestände verbessert und transparent dargestellt werden.

Das Akronym FAIR bezieht sich auf den Umgang mit Forschungsdaten. Diese sollen **F**indable (Auffindbar), **A**ccessible (Zugänglich), **I**nteroperable (Interoperabel) und **R**eusable (Wiederverwendbar) sein. In dem hier vorliegenden Artikel soll zwar der Bereich der Accessibility, also der Zugänglichkeit, näher beleuchtet werden. Zum besseren Verständnis und der Verortung von Accessibility in das Gesamtmodell der FAIR-Prinzipien werden allerdings auch die anderen drei Bereiche erläutert. Nicht zuletzt auch deshalb, weil die vier Prinzipien teilweise aufeinander aufbauen, bzw. ineinandergreifen. So ist z. B. eine Vorbedingung für die Zugänglichkeit, dass die Daten überhaupt auffindbar sind.

Findability

Um Zugang zu Forschungsdaten erhalten zu können, müssen sie aufgefunden werden können. Insofern ist das Prinzip der Findability auch wesentlich für den Datenzugang. Bevor

⁴² <https://creativecommons.org/licenses/?lang=de>

⁴³ <https://www.forschungsdaten.info/themen/informieren-und-planen/foerderrichtlinien/>

⁴⁴ <https://ec.europa.eu/programmes/horizon2020/en>

Forschungsdaten oder andere Ressourcen jedoch aufgefunden werden können, müssen zunächst die Bedingungen zur Auffindbarkeit geschaffen werden. Hierbei spielen (maschinenlesbare) Metadaten und persistente Identifikatoren eine tragende Rolle. Metadaten sind eine Grundvoraussetzung für die automatische Erkennung von Datensätzen oder auch Diensten. Durch die Vergabe von persistenten Identifikatoren wie z. B. dem Digital Object Identifier (DOI) ist die dauerhafte Auffindbarkeit und Zitierfähigkeit von Ressourcen gewährleistet. Denn die Identifikatoren verweisen auf die Ressource selbst und nicht auf den Speicherort, sodass Datensätze selbst dann noch auffindbar sind, wenn sich URL-Änderungen ergeben (Brase, 2009, S. 57).

Um die Auffindbarkeit von Ressourcen zu erleichtern und langfristig zu gewährleisten, müssen einige Bedingungen erfüllt sein. Zum einen muss die Ressource mit einem global eindeutigen und persistenten Identifikator (PID) ausgestattet werden. Zur Referenzierung von Daten und digitaler Objekte hat sich der Digital Object Identifier (DOI) als ein Standard für wissenschaftliche Publikationen etabliert⁴⁵. Ein DOI reichert ein zu referenzierendes Objekt mit Metadaten an, die dem Metadatenchema der jeweiligen DOI-Registrierungsagentur⁴⁶ (z. B. Crossref⁴⁷ oder DataCite⁴⁸) entsprechen, welche für die Vergabe von DOIs zuständig sind. Sobald eine Ressource einen persistenten Identifikator erhalten hat, dürfen weder die Ressource noch der Identifikator nachträglich verändert werden. Prinzipiell besteht die Möglichkeit, eine komplette Studie inklusive ihrer dazugehörigen Materialien mit einem PID zu versehen. Eine weitere Möglichkeit ist die PID-Vergabe für jedes einzelne Studienelement (Datensätze, Syntax, Fragebögen, Codebooks, Publikationen etc.).

Um einen PID bzw. DOI zu erhalten, gibt es mehrere Wege. Ressourcen können beispielsweise eigenständig über Repositorien (z. B. datorium⁴⁹ oder RADAR⁵⁰) oder Online-Speicherdienste (z. B. Zenodo⁵¹ oder Figshare⁵²) veröffentlicht werden, wo sie im Zuge der Veröffentlichung mit einem PID ausgestattet werden. Bei Online-Speicherdiensten ist allerdings Vorsicht geboten, denn hochgeladene Ressourcen werden oft weder kontrolliert noch kuratiert. Für FDZ empfiehlt sich deshalb eher der Weg über DataCite bzw. eine Mitgliedschaft im Konsortium einer Registrierungsagentur an (z. B. da|ra Registrierungsagentur für Sozial- und Wirtschaftswissenschaften⁵³), um DOIs für Forschungsdatensätze oder andere Materialien zu erhalten. Über die Registrierung bei DataCite oder einer DOI-Registrierungsagentur werden die Metadaten automatisch in andere Systeme übertragen, sodass sie in vielen Fachportalen auffindbar sind. Daten, die auf diese Weise indexiert sind, können von Internet-Crawlern

⁴⁵ Weitere Persistente Identifikatoren (PID) sind z. B.: Uniform Resource Name (URN), Handle-System (hdl), Persistent Uniform Resource Locator (PURL), Open Researcher and Contributor ID (ORCID), Gemeinsame Normdatendatei (GND)

⁴⁶ Liste aktueller DOI-Registrierungsagenturen: https://www.doi.org/registration_agencies.html

⁴⁷ <https://de.wikipedia.org/wiki/Crossref>

⁴⁸ <https://datacite.org/>

⁴⁹ <https://data.gesis.org/sharing/#!Home>

⁵⁰ <https://www.radar-service.eu/radar/de/home>

⁵¹ <https://zenodo.org>

⁵² <https://figshare.com>

⁵³ <https://www.da-ra.de>

durchsucht werden und werden leichter aufgefunden als Daten, die lediglich auf einer Projektseite erwähnt werden.

Eine weitere Bedingung zur Einhaltung der FAIR-Prinzipien ist die Beschreibung von Ressourcen anhand umfangreicher Metadaten. Präzise und ausführliche Metadaten sind zur Auffindbarkeit von digitalen Ressourcen unabdingbar, da Ressourcen anhand der Informationen, die mithilfe von Metadaten gemacht werden, maschinenlesbar werden und somit von Computern aufgefunden werden können. Je ausführlicher Metadaten zu einer Ressource vorliegen, desto leichter kann auf diese Ressource zugegriffen werden (GO FAIR 2022⁵⁴). Da Inhalte im Web immer automatisierter abgegriffen und verfügbar gemacht werden, ist es wichtig, dass die Metadaten Standards folgen, welche die Informationen strukturieren und die von Computersystemen verstanden werden können. Beispielsweise können die Informationen zu Forschungsdaten auf einer Projektseite in einem großen Textfeld abgelegt sein - das kann zwar von Menschen verstanden werden, aber nicht von Computersystemen. Folgt die Dokumentation einem Standard, der maschinenlesbar ist, können die Informationen maschinell aufgefunden und verbreitet werden. Wie auch bei der Vergabe von PID können Metadaten für ein komplettes Projekt oder für sämtliche Projektelemente vergeben werden.

Zu unterscheiden sind generische und fachspezifische Metadatenstandards. Generische Metadaten können unabhängig von einer Fachdisziplin erstellt werden. Zur Erstellung generischer Metadaten ist beispielsweise der DataCite Metadatengenerator⁵⁵ geeignet. Disziplinspezifische Metadaten sind in der Lage, Ressourcen genauer zu beschreiben, da sie auf die Besonderheiten der jeweiligen Disziplin abgestimmt sind. Die Research Data Alliance (RDA) stellt eine Liste verschiedener Metadatenstandards zur Verfügung.⁵⁶ Diese Liste kann helfen, einen passenden Metadatenstandard für Ressourcen zu finden. Darüber hinaus hat die RDA eine Liste mit Tools zur Erstellung standardisierter Metadaten generiert.⁵⁷

Accessibility

Nachdem die Forschungsdaten aufgefunden wurden, muss bekannt sein, unter welchen Bedingungen auf die Daten zugegriffen werden darf. Im FAIR-Konzept ist "Accessibility" allerdings nicht gleichzusetzen mit offen oder kostenfrei. Es kann gute Gründe dafür geben, dass Forschungsdaten nicht oder nur unter eingeschränkten Bedingungen zugänglich sind. Dies ist insbesondere bei sensiblen oder personenbezogenen Daten der Fall. Die FAIR-Prinzipien beziehen sich allerdings im Gegensatz zu den CARE-Prinzipien⁵⁸ explizit nicht auf moralische oder ethische Sachverhalte hinsichtlich der Offenheit von Forschungsdaten. Die Entscheidung, ob und in welchem Ausmaß Forschungsdaten veröffentlicht werden, liegt bei den Primärforschenden bzw. den Anforderungen der jeweiligen forschungsfördernden Einrichtung. Accessibility bezieht sich dementsprechend nicht auf das „ob“, sondern auf das

⁵⁴ <https://www.go-fair.org/fair-principles/f2-data-described-rich-metadata/>

⁵⁵ <https://dhvlab.gwi.uni-muenchen.de/datacite-generator/>

⁵⁶ <https://rdamsc.bath.ac.uk/>

⁵⁷ <http://rd-alliance.github.io/metadata-directory/tools/>

⁵⁸ <https://www.gida-global.org/care>

„wie“, also die genaue Beschreibung der Zugangsbedingungen zu den Forschungsdaten. Hierzu gehören auch die Konditionen, unter denen die Daten nachgenutzt werden dürfen (Mons et al., 2017, S. 49ff).

Die Zugangsbedingungen zu den Forschungsdaten oder auch einer anderen Ressource sollten deutlich und transparent in den Metadaten beschrieben werden (siehe auch Abschnitt Formale Regularien des Datenzugangs). Der Datenabruf sollte möglich sein, ohne dafür spezielle oder proprietäre Tools verwenden zu müssen. Hierfür kommen standardisierte Kommunikationsprotokolle zum Tragen, mithilfe derer (Meta)-Daten über ihre Identifier abrufbar sind. Die Protokolle müssen offen, frei und universell implementierbar sein (z. B. HTTP, FTP, SMTP). Für den Fall, dass (sensible) Forschungsdaten nur unter bestimmten Bedingungen zugänglich sind, muss das Protokoll die Authentifizierung und Rechteverwaltung unterstützen. Sollten die Forschungsdaten generell nicht verfügbar oder nicht mehr vorhanden sein, müssen die Metadaten dauerhaft verfügbar bleiben (siehe Abschnitt Datenbereitstellungswege).

Interoperability

Daten gelten als interoperabel, wenn sie von Menschen oder Maschinen mit anderen Datensätzen kombiniert werden können. Daten sollten dementsprechend zwischen Computersystemen austauschbar und interpretierbar sein. Dies bedeutet, dass automatisiert erfasst werden kann, ob Daten inhaltlich mit anderen Daten vergleichbar sind, ohne dass spezialisierte oder Ad-hoc-Algorithmen, Übersetzer oder Mappings erforderlich sind. Interoperabilität bedeutet typischerweise, dass ein Computersystem zumindest Kenntnis von den Datenaustauschformaten des anderen Systems hat. Damit dies geschieht und um die automatische Auffindbarkeit und Interoperabilität von Datensätzen zu gewährleisten, ist es entscheidend, dass

- Metadaten verwendet werden, die auf kontrollierten Vokabularen, Ontologien und Thesauri basieren, die in der Fachcommunity etabliert sind.
- Persistente Identifikatoren verwendet werden, mit denen auf andere Datensätze verwiesen werden kann. Hier sollte präzise angegeben werden, ob ein Datensatz auf einem anderen Datensatz aufbaut, ob zusätzliche Datensätze benötigt werden, um die Daten zu vervollständigen, oder ob ergänzende Informationen in einem anderen Datensatz auffindbar sind. Insbesondere die Zusammenhänge zwischen den Datensätzen sollte beschrieben werden, um die intellektuelle Verbindung der Datensätze herstellen zu können.
- Metadatenschemata genutzt werden, die ihrerseits auf den FAIR-Prinzipien basieren.⁵⁹

Reusability

Das ultimative Ziel der FAIR-Prinzipien ist es, die Nachnutzung von Forschungsdaten zu ermöglichen. Zu diesem Zweck müssen sowohl die Daten als auch die Metadaten ausführlich

⁵⁹ <https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/>

beschrieben werden, damit sie zu Replikationszwecken oder auch zur weiteren Verarbeitung genutzt werden können. Auch hierfür sind aussagekräftige Metadaten maßgebend, die neben der Beschreibung der Forschungsdaten selbst auch den Kontext beschreiben, unter denen die Daten erhoben wurden. Dazu gehört beispielsweise eine ausführliche Dokumentation der durchgeführten Methoden oder Geräte, die für die Erstellung der Daten verwendet wurden. Je ausführlicher und aussagekräftiger die Metadaten, desto leichter ist die Nachnutzbarkeit der dazugehörigen Daten.

Des Weiteren müssen die (Meta)daten unter einer eindeutigen Lizenz veröffentlicht werden, welche die Bedingungen für die Nachnutzung sowohl für Menschen als auch Maschinen transparent regelt. Für Metadaten wird eine offene Lizenz empfohlen, etwa die CC0-Lizenz⁶⁰ (Betancort Cabrera et al., 2020, S. 15). Die Nutzungsbedingungen der Forschungsdaten können auf unterschiedliche Art und Weise geregelt werden (siehe Absatz Formale Regularien des Datenzugangs). Um die Nachnutzung der Forschungsdaten eindeutig kenntlich machen zu können, müssen sie zitierfähig sein. Eine ordnungsgemäße Zitation von Forschungsdaten wird durch die oben beschriebene Vergabe von langfristig nachvollziehbaren persistenten Identifikatoren ermöglicht (GO FAIR 2022⁶¹).⁶²

Inzwischen wurden Tools entwickelt, um einzuschätzen, wie FAIR Daten sind. Einen Überblick bietet Pittonet Gaiarin (2020), wobei speziell das (noch in Überarbeitung befindliche) F-UJI Tool⁶³ geeignet ist, um eine Einschätzung über die FAIRness⁶⁴ von Datensätzen zu erlangen.

4. Resümee/Ausblick

Diese Handreichung gibt einen einführenden Überblick über den Datenzugang in den Sozialwissenschaften und angrenzenden Fachdisziplinen. Da Daten in dieser Community oft nicht als open data zur Verfügung gestellt werden können, sind verschiedene Aspekte für einen geeigneten Datenzugang zu bedenken. Es sollte vorab reflektiert werden, welche Infrastrukturen – und speziell welche Datenbereitstellungswege oder Datenkataloge - gewählt werden, da daraus folgende Abhängigkeiten und Strukturen mittel- bzw. langfristige Folgen haben. Insbesondere hinsichtlich der Kosten sollten sorgfältige Kalkulationen durchgeführt werden, möglicherweise auch unter Zuhilfenahme von Erfahrungswerten anderer FDZ. Eine Überprüfung, ob auf bereits vorhandene interne technische Infrastrukturen zurückgegriffen werden kann, sollte ebenfalls stattfinden, um Einsparungsoptionen auszuloten.

Die Five Safes geben Anhaltspunkte, welche Maßnahmen für eine sichere Datenbereitstellung berücksichtigt werden sollten. Damit fließen Überlegungen mit ein, für welchen

⁶⁰ <https://creativecommons.org/publicdomain/zero/1.0/deed.de>

⁶¹ <https://www.go-fair.org/fair-principles/>

⁶² Zur Vertiefung der FAIR-Prinzipien sei die Lektüre von Betancort Cabrera et al. (2020) empfohlen, die die Umsetzung der eher allgemeinen FAIR-Prinzipien in den Sozial-, Verhaltens- und Wirtschaftswissenschaften hervorragend erläutert.

⁶³ <https://www.f-uji.net/>

⁶⁴ Die Operationalisierung basiert auf Devaraju et al. (2020).

Nutzendenkreis und Zweck die Daten aufbereitet und bereitgestellt werden können, etwa für wissenschaftliche Analysen oder ein breiteres Nutzungsspektrum. Digitale Services für Datenkataloge können den Datenzugang erheblich erleichtern. Transparente Nutzungsbedingungen und standardisierte Zugangskategorien sollten dabei den Datenzugang strukturieren und von den Datennutzenden vorab eingesehen werden können. Bereits existierende Open Source-Softwareanwendungen für Datenkataloge können dabei unterstützen, da Informationen nach international standardisierten Schemata aufbereitet und dargestellt werden. Der Aufbau von Datenkatalogen wird durch die Verwendung von Open Source Software, im Vergleich zu vollständigen Eigenentwicklungen, erheblich erleichtert. Nicht zuletzt, weil FDM-Open Source Tools bereits einen Großteil der FAIR-Prinzipien berücksichtigen, sodass der diesbezüglich konzeptionelle Aufwand verringert werden kann. Weitere Möglichkeiten zur Erfüllung der FAIR-Prinzipien hinsichtlich des Datenzugangs werden anhand von Anwendungsbeispielen aufgezeigt.

Um die Qualität der Arbeitsprozesse und Infrastruktur eines FDZ intern zu validieren und extern zu demonstrieren, können weitere Schritte angedacht werden. Beispielsweise werden Datennutzende und Datengebende bei der Suche nach einem geeigneten Datenarchiv oft auf Qualitätssiegel hingewiesen, welche Auskünfte über die Vertrauenswürdigkeit eines Archivs geben. Eine Akkreditierung für ein FDZ durch den RatSWD⁶⁵ kann diesbezüglich ein erster Schritt sein, um eine extern gesicherte Qualitätsprüfung an die wissenschaftliche Community zu signalisieren. Die Akkreditierungskriterien zielen auf die Gleichbehandlung aller berechtigten Datennutzenden beim Datenzugang ab und demonstrieren Datennutzenden die Verlässlichkeit des Datenzugangs. Eine weitere Möglichkeit, die Vertrauenswürdigkeit eines FDZ aus der Perspektive der Datenarchivierung und -nutzung, zu dokumentieren, ist das Core Trust Seal⁶⁶. Zur Erlangung dieses Zertifikats werden die unterschiedlichen Bereiche eines FDZ aus technischer und organisatorischer Perspektive tiefgehend dokumentiert und geprüft. Der Aufwand ist nicht zu unterschätzen (siehe Pegelow, Jansen & Neuendorf, 2021). Jedoch bietet es die Möglichkeit, die eigenen Prozesse kritisch zu hinterfragen, zu optimieren sowie einen höheren Bekanntheitsgrad auf internationaler Ebene zu erreichen.

⁶⁵ <https://www.konsortswd.de/datenzentren/akkreditierung/>

⁶⁶ <https://www.coretrustseal.org>

Literaturverzeichnis⁶⁷

- Betancort Cabrera, N., Bongartz, E. C., Dörrenbächer, N., Goebel, J., Kaluza, H. & Siegers, P. (2020). *White Paper on implementing the FAIR principles for data in the Social, Behavioural, and Economic Sciences*. (RatSWD Working Paper 274/2020). doi:10.17620/02671.60
- Brase, J. (2009). Der Digital Object Identifier (DOI). In H. Neuroth (Hrsg.), *Nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung* (Version 2.0, Juni 2009, S. 57–65). Boizenburg: Hülsbusch; Univ.-Verl. Göttingen.
- Depping, R. (2021). *Rechtliche Aspekte des Forschungsdatenmanagements. Eine Einführung*. <http://kups.ub.uni-koeln.de/id/eprint/45599>
- Desai, T., Ritchie, F. & Welpton, R. (2016). *Five Safes: designing data access for research* (Economics Working Paper Series Nr. 1601). doi:10.13140/RG.2.1.3661.1604
- Devaraju, A., Huber, R., Mokrane, M., Herterich, P., Cepinskas, L., Vries, J. de et al. 2020, 1. January. *FAIRsFAIR Data Object Assessment Metrics*. doi:10.5281/ZENODO.4081213
- Ebel, T. (2015). *Empfehlungen zur Anonymisierung quantitativer Daten*. Mannheim: GESIS – Leibniz-Institut für Sozialwissenschaften.
- Ebel, T. & Meyermann, A. (2015). *Hinweise zur Anonymisierung von quantitativen Daten* (Forschungsdatenzentrum (FDZ) Bildung am DIPF, Hrsg.) (forschungsdaten bildung informiert Nr. 3). Frankfurt am Main: Verbund Forschungsdaten Bildung. <https://www.forschungsdaten-bildung.de/files/fdb-informiert-nr-3.pdf>
- Eisentraut, M. (2018). Data Anonymization. In S. Netscher & C. Eder (Hrsg.), *Data Processing and Documentation. Generating High Quality Research Data in Quantitative Social Science Research* (GESIS Papers, S. 34–36). Köln.
- Koch, U., Akdeniz, E., Meichsner, J., Hausstein, B. & Harzenetter, K. 2017, 1. January. *da|ra Metadata Schema*. doi:10.4232/10.mdsdoc.4.0
- Kreutzer, T. & Lahmann, H. (2021). *Rechtsfragen bei Open Science - Ein Leitfaden* (2. Aufl.) (Behörde für Wissenschaft, Forschung und Gleichstellung (BWFG), Hrsg.). Hamburg: Staats- und Universitätsbibliothek Hamburg Carl von Ossietzky. doi:10.15460/HUP.211
- McEachern, S. (2015). *Implementation of the Trusted Access Model ASSA Policy Roundtable*. http://rsss.anu.edu.au/sites/default/files/SMcEachern_researcherperspective_ASSANov2015.pdf
- Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B. & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*, 37(1), 49–56. doi:10.3233/ISU-170824

⁶⁷ Alle Internetquellen wurden zuletzt geprüft am 20.10.2022

- Müller, W., Blien, U., Knoche, P. & Wirth, H. (1991). *Die faktische Anonymität von Mikrodaten* (Forum der Bundesstatistik, Bd. 19). Stuttgart: Metzler-Poeschel.
- Murray, N. & Goebel, J. (2022). *Vertragliche Grundlagen zur Teilnahme am RDCnet*. (KonsortSWD Working Paper 1/2022). doi:10.5281/zenodo.6358334
- Pegelow, L., Jansen, M. & Neuendorf, C. (2021). Erwerb des Zertifikats CoreTrustSeal (CTS) durch ein Forschungsdatenzentrum im Bildungsbereich – Motivation, Umsetzung und Lessons learned. doi:10.17192/BFDM.2021.1.8310
- Pittonet Gaiarin, S. 2020, 1. January. *FAIR Assessment and Certification in the EOSC region* (European Open Science Cloud, ed.). European Open Science Cloud. doi:10.5281/zenodo.4486280
- RatSWD. (2019). *Remote Access zu Daten der amtlichen Statistik und der Sozialversicherungsträger* (RatSWD Output 5 Nr. 6). doi:10.17620/02671.42
- RatSWD. (2020). *Handreichung Datenschutz. 2. vollständig überarbeitete Auflage*. (RatSWD Output 8 Nr. 6). doi:10.17620/02671.50
- RatSWD. (2020). *Tätigkeitsbericht 2019 der vom RatSWD akkreditierten Forschungsdatenzentren (FDZ)*. doi:10.17620/02671.56
- Schallaböck, J., Hoffstätter, U., Buck, D. & Linne, M. (2022). *Mustervertrag Datennutzung KonsortSWD* (KonsortSWD, Hrsg.). doi:10.5281/zenodo.5828114
- Schiller, D. & Welpton, R. (2015). Distributing Access to Data, not Data. *IASSIST Quarterly*, 38(3), 6. doi:10.29173/iq122
- Schiller, D. H., Eberle, J., Fuß, D., Goebel, J., Heining, J., Mika, T. et al. (2017). *Standards des sicheren Datenzugangs in den Sozial- und Wirtschaftswissenschaften*. (RatSWD Working Paper 261/2017). doi:10.17620/02671.15

Impressum

Kontakt:

Ute Hoffstätter

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW)

Lange Laube 12

30159 Hannover

Tel.: +49 511 450670-404

hoffstaetter@dzhw.eu

Monika Linne

RWI – Leibniz-Institut für Wirtschaftsforschung

Hohenzollernstraße 1-3

45128 Essen

Tel.: +49 201 8149-267

monika.linne@rwi-essen.de

Dezember 2022

KonsortSWD Working Paper:

KonsortSWD baut als Teil der Nationalen Forschungsdateninfrastruktur Angebote zur Unterstützung von Forschung mit Daten in den Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften aus. Unsere Mission ist es, die Forschungsdateninfrastruktur zur Beforschung der Gesellschaft zu stärken, zu erweitern und zu vertiefen. Sie soll nutzungsorientiert ausgestaltet sein und die Bedürfnisse der Forschungscommunities berücksichtigen. Wichtiger Grundstein ist dabei das seit über zwei Jahrzehnten durch den Rat für Sozial- und Wirtschaftsdaten (RatSWD) aufgebaute Netzwerk von Forschungsdatenzentren.

In dieser Reihe erscheinen Beiträge rund um das Forschungsdatenmanagement, die im Kontext von KonsortSWD entstehen. Beiträge, die extern und doppelblind begutachtet wurden sind entsprechend gekennzeichnet.

KonsortSWD wird 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/>

DOI: 10.5281/zenodo.7347064

Zitationsvorschlag:

Hoffstätter, U. & Linne, M. (2022). *Datenzugang. Einführung in das Thema Zugang zu Daten der Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften in Forschungsdatenzentren*. KonsortSWD Working Paper 4/2022. Konsortium für die Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften (KonsortSWD). <https://doi.org/10.5281/zenodo.7347064>