

Auszug aus dem Buch:

Uwe Jensen
Sebastian Netscher
Katrin Weller (Hrsg.)

Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten

Grundlagen und praktische Lösungen
für den Umgang mit
quantitativen Forschungsdaten

Verlag Barbara Budrich
Opladen • Berlin • Toronto 2019

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;
detaillierte bibliografische Daten sind im Internet über
<http://dnb.d-nb.de> abrufbar.

© 2019 Dieses Werk ist beim Verlag Barbara Budrich erschienen und steht unter der Creative Commons Lizenz Attribution-ShareAlike 4.0 International (CC BY-SA 4.0):

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

Diese Lizenz erlaubt die Verbreitung, Speicherung, Vervielfältigung und Bearbeitung bei Verwendung der gleichen CC-BY-SA 4.0-Lizenz und unter Angabe der UrheberInnen, Rechte, Änderungen und verwendeten Lizenz.



Dieses Buch steht im Open-Access-Bereich der Verlagsseite zum kostenlosen Download bereit (<https://doi.org/10.3224/84742233>).

Eine kostenpflichtige Druckversion (Print on Demand) kann über den Verlag bezogen werden. Die Seitenzahlen in der Druck- und Onlineversion sind identisch.

ISBN 978-3-8474-2233-4 (Paperback)

eISBN 978-3-8474-1260-1 (eBook)

DOI 10.3224/84742233

Umschlaggestaltung: Bettina Lehfeldt, Kleinmachnow – www.lehfeldtgraphic.de

Lektorat: Nadine Jenke, Potsdam

Satz: Anja Borkam, Jena – kontakt@lektorat-borkam.de

Titelbildnachweis: Foto: Florian Losch

Druck: paper & tinta, Warschau

Printed in Europe

2. Forschungsdaten und Forschungsdatenmanagement in den Sozialwissenschaften

Uwe Jensen

Die Verfügbarkeit von Daten ist für Forschende in den empirischen Sozialwissenschaften eine notwendige Voraussetzung ihres wissenschaftlichen Arbeitens. In den letzten Jahrzehnten hat sich die empirische Basis für wissenschaftliche Analysen und damit für das Verständnis sozialer Entwicklungen kontinuierlich ausgeweitet (King 2011). Dieser stetige Zuwachs an verfügbaren Daten beruht auf unterschiedlichen Faktoren und verschiedenen Entwicklungen. Hierzu zählen vor allem methodische und technische Neuerungen, wie Statistiksoftware, verfügbare Speicherkapazitäten oder das Internet. Zu nennen sind aber auch neue Datentypen, wie digitale Verhaltens- und Transaktionsdaten, Social-Media-Daten ebenso wie Prozessdaten öffentlicher Einrichtungen. Parallel zum Zuwachs an empirischen Daten ist auch deren Komplexität kontinuierlich gestiegen. Dies betrifft sowohl die Informationen in den Daten als auch die Datenquellen und Methoden der Datenproduktion und Analyse (Ludwig/Enke 2013: 13). Insbesondere die Verknüpfung von neuen Datentypen oder von Daten anderer Disziplinen, wie z.B. Geo- oder Gesundheitsdaten, mit traditionellen Umfragedaten der empirischen Sozialforschung ist dabei mit der Erwartung verbunden, soziale Entwicklungen noch besser verstehen zu können (Jensen et al. 2015: 12f.; OECD 2013: 12).

Mit dem stetigen Zuwachs an Daten und deren Komplexität sind Forschende und Forschungsdatendateninfrastrukturen mit neuen Fragen und Herausforderungen beim Umgang mit diesen (neuen) Daten konfrontiert. Dies betrifft etwa Maßnahmen zum Erhalt der Daten ebenso wie zur Sicherung ihrer Qualität und der Gewährleistung ihrer Verständlichkeit. Forschende, die im Rahmen ihres Forschungsvorhabens Daten produzieren, müssen aus rein intrinsischen Motiven sicherstellen, dass sie diese Daten zur Beantwortung der Forschungsfrage nutzen können. Im Sinne guter wissenschaftlicher Praxis sollten diese Forschenden aber auch gewährleisten, dass die Daten über das Forschungsvorhaben hinaus erhalten bleiben, um produzierte Forschungsergebnisse replizieren und somit überprüfen zu können. In diesem Zusammenhang forderte etwa die Deutsche Forschungsgemeinschaft (DFG) bereits 1988, dass „Primärdaten als Grundlagen für Veröffentlichungen [...] auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, zehn Jahre lang aufbewahrt werden“ (DFG 2013 [1988]: 21) sollen.

In den nachfolgenden Jahren forcierten Initiativen von Forschenden und Forschungsgemeinschaften Überlegungen, öffentlich finanzierte Daten der empirischen Sozialforschung auch für Dritte zur Nachnutzung in neuen Forschungskontexten (*Data Sharing*) verfügbar zu machen (vgl. OECD 2007; Berliner Erklärung 2003). Die Forderung, Daten für Dritte breit-zustellen, wurde auch in den „Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten“ der DFG (2009) aufgegriffen und u.a. hinsichtlich der Qualitätssicherung, fachspezifischer Organisationskonzepte, Standards zur Datenspeicherung und zur Beschreibung durch Metadaten präzisiert. Schließlich formulierte die Allianz der deutschen Wissenschaftsorganisationen 2010 die *Grundsätze zum Umgang mit Forschungsdaten*. Sie verweisen auf den disziplinspezifischen Charakter der Daten und entsprechende fachspezifische Regeln und Standards, etwa in Bezug auf den Datenschutz oder den Zugang zu Daten. In den letzten Jahren implementierten schließlich mehr und mehr Förderer in ihren Ausschreibungen die Verpflichtung zum Bereitstellen von Daten, deren Erstellung

durch das Forschungsvorhaben finanziert wurde. Dazu zählen z.B. das Horizon 2020 Programme (o.J.) der Europäischen Kommission, die Förderung der Forschung zur Digitalisierung im Bildungsbereich des Bundesministeriums für Bildung und Forschung (BMBF 2017) und nicht zuletzt die DFG mit ihren Leitlinien zum Umgang mit Forschungsdaten (DFG 2015) und Leitfäden für Antragstellende (DFG 2018).

Alle im Rahmen eines empirischen Forschungsvorhabens (und darüber hinaus) ergriffenen Maßnahmen zur Sicherung der Daten, zu ihrem längerfristigen Erhalt, zur Gewährleistung ihrer Qualität und ihrer Verständlichkeit ebenso wie zu ihrer Bereitstellung für Dritte werden zumeist unter dem Begriff des *Forschungsdatenmanagements* subsumiert. Zur Einführung in das Thema Forschungsdatenmanagement stellt Abschnitt 2.1 zunächst übergreifende Begriffe und Konzepte eines sozialwissenschaftlich orientierten Forschungsdatenmanagements vor, auf die sich die weiteren Kapitel explizit oder implizit beziehen. Dazu wird insbesondere auf allgemeine Charakteristika von Umfragedaten in den Sozialwissenschaften eingegangen, um ein Verständnis derartiger Daten quer zu allen Kapiteln dieses Buches zu erleichtern. Abschnitt 2.2 thematisiert einleitend, wie sich Forschungsprozesse in den Lebenszyklusmodellen von Forschungsdaten einordnen lassen. Der spezifischere Zusammenhang des Forschungsdatenmanagements bei der Produktion von Daten und Metadaten wird danach entlang des DDI-Data-Lifecycle-Modells beschrieben. Gleichzeitig wird die Rolle des gleichnamigen DDI-Metadatenstandards zur Dokumentation von Umfragedaten erläutert, auf den in einigen Kapiteln Bezug genommen wird. Abschnitt 2.3 dient schließlich der Zusammenschau des sozialwissenschaftlichen Forschungsdatenmanagements aus drei verschiedenen Perspektiven. Zuerst werden projektbezogene Maßnahmen und Aktivitäten des Forschungsdatenmanagements quer zu den Phasen des Lebenszyklus von Daten resümiert. Anschließend geht der Abschnitt kurz auf die Rolle und Dienstleistungen von Dateninfrastruktureinrichtungen im Rahmen des Forschungsdatenmanagements ein und thematisiert neue Herausforderungen im Umgang mit neuen Datenformen und Datenquellen. Abschnitt 2.4 fasst zentrale Aspekte des Forschungsdatenmanagements zusammen, die die weiteren Kapitel dieses Buches vertiefen.

2.1 Besonderheiten sozialwissenschaftlicher Daten

Im Mittelpunkt dieses Buches steht das Management quantitativer sozialwissenschaftlicher Umfrage- bzw. Forschungsdaten, die durch standardisierte Befragungen erhoben werden. Gegenstand ist der Umgang mit eben diesen Forschungsdaten, die durch die disziplin- oder fachspezifische Methoden und Verfahren der empirischen Sozialforschung erzeugt werden. Konkrete Strategien, Maßnahmen und Regeln des Forschungsdatenmanagements in Forschungsvorhaben sind in starkem Maße von den Charakteristika der Daten und von den Besonderheiten der Datensatzstrukturen abhängig, mit denen eine Disziplin forscht. Bevor diese Aspekte im vorliegenden Kapitel vertieft werden, soll jedoch zunächst auf das Verständnis der Nutzung einiger Begriffe in diesem Buch eingegangen werden.

2.1.1 *Begriffsverständnis: Forschungsprojekt, Forschungsziel, Replikation und Nachnutzung*

Einige Begriffe der empirischen Sozialforschung, die in allen Kapiteln dieses Buches wie selbstverständlich verwendet werden, sind durchaus mit unterschiedlichen Bedeutungen

belegt. So bezeichnet im Kontext dieses Buches der Begriff *Forschungsprojekt* ein von einer oder mehreren Forschenden durchgeführtes empirisches Forschungsvorhaben, in dem sozialwissenschaftliche Daten erstellt, aufbereitet und/oder genutzt werden. In diesem Zusammenhang sind auch die Begriffe *Projektalltag*, d.h. die tägliche Arbeit am Forschungsvorhaben, *Projektziele*, d.h. die mit dem Forschungsvorhaben originär verknüpften Ziele, wie beispielsweise die Beantwortung einer konkreten Forschungsfrage, *Projektworkflow*, d.h. der Arbeitsablauf zur Umsetzung des Forschungsprojekts oder der *Lebenszyklus* des Projektes als systematische Abfolge des Projektworkflows, etc. zu verstehen.

Gleichzeitig werden bei der Verwendung dieser Begriffe keine spezifischen formalen Organisations- oder Kooperationsformen unterstellt. Dementsprechend ist auch der Begriff der *kleinen und mittleren Forschungsprojekte* im Verlauf dieses Buches bewusst unscharf verwendet und wird nicht mit einem bestimmten Umfang des Forschungsvorhabens, etwa in Form einer bestimmten Anzahl von im Projekt beteiligten Forschenden verbunden. Dies liegt zunächst daran, dass im hiesigen Zusammenhang der Begriff des Forschungsprojekts losgelöst von jeglicher zeitlichen oder finanziellen Dimension zu verstehen ist, d.h. es kann sich dabei um ein bereits beendetes, noch laufendes oder in Planung befindliches Projekt handeln, dass institutionell finanziert oder durch Drittmittelgeber gefördert ist.

Darüber hinaus fehlt es an entsprechender Trennschärfe, was genau als Forschungsprojekt zu definieren ist. Gerade in internationalen Umfragen können einzelne Erhebungen in unterschiedlichen Ländern getrennt betrachtet werden und so das große, gesamte Forschungsprojekt in viele Teilprojekte zerlegt werden. Zum anderen liefert beispielsweise die Anzahl an beteiligten Forschenden im Projekt nur eine vage Einschätzung der Projektgröße. So können ggf. große internationale Umfrageprogramme mit vergleichsweise wenigen, aber sehr gut geschulten Mitarbeitenden die Aufbereitung und Dokumentation der erstellten Daten bewerkstelligen. Umgekehrt können relativ kleine Forschungsprojekte mit einer relativ hohen Anzahl an Projektbeteiligten ausgestattet sein, z.B., wenn im Rahmen des Forschungsvorhabens von einem Teil der Forschenden weiterführende Qualifikationen, wie etwa Promotionen, angestrebt werden.

Forschungsziele sind in den Sozialwissenschaften zumeist verbunden mit der Beantwortung konkreter Forschungsfragen, wie etwa zum Wahlverhalten in einem oder mehreren Ländern bzw. zu einem oder mehreren Zeitpunkten. Forschungsziele können aber viel weiter gefasst werden. So kann beispielsweise auch die Erstellung eines Datensatzes Ziel eines Projektes sein oder zumindest ein wichtiges Produkt der eigentlichen Projektarbeit darstellen. Doch auch unter der Annahme, dass die Beantwortung einer konkreten Forschungsfrage im Vordergrund des Forschungsvorhabens steht, können projektintern verschiedene Ziele im Umgang mit den erstellten Forschungsdaten definiert werden. Dies betrifft in erster Linie natürlich die erfolgreiche Umsetzung der eigentlichen Projektziele und somit den Erhalt qualitativ hochwertiger und verständlicher Forschungsdaten im Projektverlauf.

Im Sinne guter wissenschaftlicher Praxis sollten Forschende die Daten zu Replikationszwecken über das eigentliche Projektende hinaus aufbewahren. Zumindest wenn diese Daten zur Generierung publizierter Forschungsergebnisse genutzt wurden, muss sichergestellt sein, dass diese Ergebnisse erneut erzeugt werden können. In diesem Buch verstehen wir unter dem Begriff der *Replikation* die Möglichkeit, Analysen und die zugrunde liegenden Daten aus publizierten Forschungsergebnissen systematisch zu überprüfen. Dies betrifft sowohl die Überprüfung der Ergebnisse an sich, d.h. deren Verifikation durch die wiederholte Datenanalyse, als auch die wiederholte Erstellung der Daten auf Basis des ehemaligen Studiendesigns, Messinstruments etc.

Der Begriff der *Nachnutzung* bzw. des *Data Sharing* beschreibt hingegen den Sachverhalt, dass relevante Forschungsdaten von den Primärforschenden (nach Projektende) Dritten zur weiteren Nutzung bereitgestellt und verfügbar gemacht werden (vergl. für die Sozial-

wissenschaften etwa den Beitrag von Huschka et al. 2011: 37f.). Die Gruppe an Nachnutzenden kann dabei auf bestimmte Personen oder Personengruppen, wie z.B. Forschende der Sozialwissenschaften, begrenzt sein, oder aber eine breitere Öffentlichkeit, wie z.B. Forschende aus anderen Disziplinen, der Politik oder der Presse, umfassen. Gleiches gilt für den Zweck der Nachnutzung. Auch hier kann die Datennutzung auf bestimmte Zwecke, wie die sozialwissenschaftliche Forschung, begrenzt oder aber für alle möglichen Zwecke, also beispielsweise auch zur kommerziellen Nutzung, freigegeben werden.

2.1.2 Forschungsdaten in den Sozialwissenschaften

Im Zusammenhang mit Forschungsdaten wird oftmals von Rohdaten, Primärdaten, empirischen Daten, Ausgangdaten etc. gesprochen, die erhoben, verarbeitet, harmonisiert und analysiert werden sollen. Eine einheitliche oder universelle Definition von Forschungsdaten, die im Detail für alle Disziplinen zutreffen würde, existiert jedoch nicht. Vielmehr herrscht in vielen Schriften und Leitlinien zum Umgang mit Daten Einigkeit, dass eine fachspezifische Betrachtung erforderlich ist. So heißt es etwa im *Positionspapier Forschungsdaten* der Deutschen Initiative für Netzwerkinformation e.V. (DINI 2009: 7):

Forschungsdaten variieren nach Disziplin. Anders als beim Umgang mit klassischen Textpublikationen ist beim Umgang mit Forschungsdaten häufig ein umfassendes Verständnis der jeweiligen Daten vonnöten, um den vielschichtigen disziplinspezifischen Charakteristika der Daten gerecht zu werden.

Demnach definiert die jeweilige Forschungsgemeinschaft, was als Forschungsdaten im vorliegenden Kapitel betrachtet wird und welche Anforderungen von Datenproduzierenden und -nutzenden an den Umgang mit ihnen zu berücksichtigen sind. Dabei ist es nach Ludwig (2012) nicht sinnvoll, Inhalte oder Quellen auszuschließen, die die Daten behandeln können oder denen sie entstammen, „da prinzipiell alles Untersuchungsgegenstand der Wissenschaft werden kann“. Stattdessen ist die Frage nach der Definition des Begriffs Daten „eher eine Aussage über ihre methodische Verwendung in einem bestimmten wissenschaftlichen Kontext“ (ebd.: 300).

Demnach muss die Frage nach der Definition von Forschungsdaten im vorliegenden Kontext aus sozialwissenschaftlicher Perspektive betrachtet werden. Die Disziplin *Sozialwissenschaften* ist ein Sammelbegriff für eine Vielfalt akademischer Fächer, von Anthropologie über die Soziologie und Politologie bis hin zur Ökonomie. Allen diesen Fächer gemein ist ihr Bezug zum sozialen Handeln und die Frage nach dem gesellschaftlichen Zusammenhalt. Die im Rahmen der Disziplin genutzten Methoden und Verfahren werden unter dem Begriff *empirische Sozialforschung* zusammengefasst. Die unterschiedlichen Fächer der Sozialwissenschaften haben dabei ein gemeinsames Verständnis von Forschung und wissenschaftlichem Arbeiten im Umgang mit den verwendeten Methoden (Quandt/Mauer 2012: 61). Zu diesen Methoden zählen im Wesentlichen die Befragung, die Beobachtung, das Experiment und die Inhaltsanalyse, die in sich weiter differenzierbar sind. Bei der Produktion bzw. der Erhebung empirischer Informationen werden unterschiedliche Instrumente eingesetzt, wie etwa Fragebögen, leitfadengestützte Interviews, Beobachtungsschemata, inhaltsanalytische Kategorienschemata usw.

Im Sinne eines methodenbasierten Verständnisses (Ludwig 2012: 299) werden daher im Folgenden alle Informationen als *sozialwissenschaftliche (Forschungs-)Daten* definiert, die in wissenschaftlichen Kontexten durch fach- bzw. disziplinspezifische Methoden und Verfahren zur Beantwortung von Forschungsfragen und Gegenständen der Sozialwissenschaften erzeugt werden oder aus deren Bearbeitung entstehen. Dies umfasst sowohl die traditionellen Umfragedaten der empirischen Sozialforschung als auch neue Datentypen, wie etwa

prozessgenerierte Daten, oder Daten aus (relativ) neuen Quellen, wie etwa von Social-Media-Plattformen im Internet. D.h. auch, dass alle verfügbaren Informationen über die Entstehung und Bearbeitung dieser Daten, immer – zumindest implizit – in eine solche Definition von Forschungsdaten einbezogen werden müssen. Aus praktischer Sicht sind die rohen Messwerte, z.B. in einer Datenmatrix, ohne eine entsprechende Dokumentation nicht verständlich und somit auch nicht nutzbar.

Im Rahmen des Buches stehen vor allem Umfragedaten im Vordergrund, die durch standardisierte Fragebögen erhoben werden. Sie werden im nächsten Abschnitt thematisiert. Darüber hinaus wird der Umgang mit Daten aus neuen Datenquellen im Kontext neuer Herausforderungen an das Forschungsdatenmanagement behandelt.

2.1.3 Studiendesign und Datenstruktur in der empirischen Sozialforschung

Forschungsdaten sind die Basis empirischer Untersuchungen in den Sozialwissenschaften und liegen zumeist in digitaler Form vor. Sie bilden die zu untersuchenden Ausschnitte gesellschaftlicher Wirklichkeit ab.

Diese können im Rückblick ohnehin schon neben ihrem Wert für die ursprüngliche Forschungsfrage sehr oft eine historisch beschreibende Funktion bekommen. [...] Sowohl historische amtliche Statistiken wie auch Daten aus Umfragen oder Beobachtungen können so einen analytischen Wert erhalten, der zum Zeitpunkt ihrer Erhebung noch nicht vorauszusehen war. (Quandt/Mauer 2012: 62)

Der Begriff der *Primärdatenerhebung* bzw. der damit einhergehenden *Primäranalyse* beschreibt dabei Daten, die zur Beantwortung einer spezifischen Forschungsfrage neu erhoben, aufbereitet und analysiert werden. Entsprechend der Forschungsfrage und dem Analyseziel eines Forschungsvorhabens müssen die notwendigen Informationen mit Hilfe eines geeigneten Studiendesigns und entsprechender Verfahren erhoben werden. So schreibt z.B. Mochmann (2014: 233):

Soll zu einem bestimmten Zeitpunkt die Verteilung von Merkmalen, wie z.B. Einstellungen zu aktuellen politischen Themen und Wahlpräferenzen analysiert werden, dann ist die Querschnittstudie bei einer Stichprobe der Wahlberechtigten das geeignete Verfahren. [...] So können zum jeweiligen Zeitpunkt Einstellungen zu den aktuellen politischen Fragen in Abhängigkeit von Merkmalen wie Alter, Schulbildung oder Einkommen erforscht werden. Die Ergebnisse haben überwiegend deskriptiven Charakter.

Sollen demgegenüber Entwicklungen und Veränderungen von Einstellungen und Verhalten zu bestimmten Themen im Zeitverlauf erfasst werden,

sind Trendstudien das zielführende Verfahren. [...] So werden z.B. zur Analyse des Wahlverhaltens über Zeit jeweils nach den Wahlen die gleichen Fragen erneut gestellt und die Antwortverteilungen mit den Ergebnissen früherer Studien verglichen. (Ebd.: 234)

Eine dritte Form der vergleichenden empirischen Sozialforschung untersucht Entwicklungen nicht nur zu verschiedenen Zeitpunkten, sondern bezieht in den Vergleich auch unterschiedliche (geographische) Grundgesamtheiten mit ein. So sind etwa international vergleichende Studien eine geeignete Form, um globale Entwicklungen von Demokratien und Staaten zu erfassen, indem etwa das Wahlverhalten zu unterschiedlichen Zeitpunkten bei Wählerinnen und Wählern in verschiedenen Ländern untersucht wird (vgl. CSES o.J.).

Entgegen der Primärerhebung nutzt die sogenannte *Sekundäranalyse* bereits bestehende Daten. Sie wertet die Daten hinsichtlich der eigenen Forschungsfrage aus, unabhängig vom ursprünglichen Zweck ihrer Sammlung. Baur und Blasius (2014: 48f.) schreiben in diesem Zusammenhang:

Da die Datenerhebung meistens sehr aufwändig und teuer ist, greifen Sozialforscherinnen und Sozialforscher zunehmend auf vorhandene Datenbestände zurück, um diese einer Sekundäranalyse zu unterziehen. Für die Analyse von sozialen Prozessen ist dies teilweise sogar erforderlich, weil sich diese erst über Zeit entfalten [...]. So können einige der Daten, die bereits vor vielen Jahren oder gar vor mehreren Jahrzehnten erhoben wurden, erst heute in ihrem vollen Umfang ausgewertet werden. Des Weiteren ist oft auch im Sinne der Nachvollziehbarkeit und Überprüfung von früheren Forschungsergebnissen eine Re-Analyse von Daten wünschenswert.

Je nach konkreter Forschungsfrage bzw. je nach Forschungsvorhaben kann das Studiendesign unterschiedlich komplex ausfallen. Entsprechend umfangreich und vielseitig sind die von einem Projekt erzeugten bzw. analysierten Forschungsdaten. Die grundlegende Struktur sozialwissenschaftlicher Forschungsdaten, die im Kontext des Forschungsdatenmanagements relevant werden, ergibt sich dabei aus drei Dimensionen:

- Anzahl der Stichprobe(n) der zu untersuchenden Grundgesamtheit(en) (*Sample*),
- Zeitpunkt(e) der Datenerhebung(en) (*Time*),
- geographischer Bezug der Daten (*Space*).

Je nach Kombination dieser drei Dimensionen (*Sample x Time x Space*) lassen sich die in Schaukasten 2.1 dargestellten Formen des Studiendesigns bzw. die wachsende Komplexität der Datenstrukturen unterscheiden. Dabei handelt es sich lediglich um einige Beispiele, d.h. die im Schaukasten gelisteten Studiendesigns sind keineswegs vollständig. Darüber hinaus werden in den Sozialwissenschaften Daten aus amtlichen Statistiken genutzt oder auch prozessproduzierte Daten, die z.B. im Rahmen von Verwaltungsprozessen entstehen.

Die jeweiligen Studiendesigns implizieren unterschiedlich komplexe Strukturen von Daten. Diese können auf eine grundlegende Form der Datenstruktur, die sogenannte *Datenmatrix*, zurückgeführt werden. Sie besteht in den quantitativen Sozialwissenschaften in der Regel aus rechteckigen Tabellen, in denen eine Liste von Subjekten in den Zeilen und deren Merkmale in den Spalten eingetragen werden. *Merkmalsträger* bzw. Subjekte können Personen, Haushalte, Körperschaften, staatliche Entitäten, Güter etc. sein. Die Datenstrukturen können unterschiedliche Datentypen beinhalten, wie etwa Individualdaten oder Aggregatdaten, Daten der amtlichen Statistik oder prozessproduzierte Mikrodaten öffentlicher Verwaltungen. So umfassen z.B. Individualdaten einer Umfrage je Zeile eine befragte Person und pro Spalte je ein *Merkmal* (z.B. Alter der Person), das mit einem Frageitem des Fragebogens (z.B. durch die Frage „In welchem Jahr wurden Sie geboren?“) erhoben wurde. Die Antworten jeder befragten Person werden dann in die entsprechende Zelle der Matrix als Wert eingetragen.

Für die statistische Auswertung werden alle Informationen bezüglich eines Frageitems für alle befragten Personen dazu in der gleichen Spalte als Variable mit zulässigen Variablenwerten abgelegt. Dementsprechend bilden Variablen Merkmale von Merkmalsträgern ab. Die Begriffe Variable und Merkmal werden dabei oftmals synonym verwendet. Die Menge aller Merkmalsträger – auch oft als Fälle bezeichnet – heißt Untersuchungseinheit. Variablenwerte repräsentieren – in der Regel durch numerische Werte – die *Ausprägungen* oder *Kategorien der Merkmale* (*value; category*) eines jeden Merkmalsträgers, wie etwa das konkrete Alter einer befragten Person oder die Arbeitslosenquote in einem bestimmten Staat bzw. einer bestimmten Region.

Diese Informationen müssen in der Primärerhebung zunächst gesammelt werden, z.B. durch die Befragung von Personen oder das Abrufen amtlicher Statistiken, die z.B. vom Statistischen Bundesamt (Destatis) oder vom Statistischen Amt der Europäischen Union (Eurostat) bereitgestellt werden. Die hierbei entstehende Informationssammlung wird häufig auch als Rohdaten bezeichnet, die ggf. zunächst in eine rechteckige Datenmatrix überführt werden muss. Dazu sind die quantitativen Daten der empirischen Erhebung zunächst auf

Basis des Messinstruments, z.B. eines Fragebogens, zu definieren. Die *Datendefinition* umfasst die Erstellung von Variablendefinitionen und Kodierungsschemata eines Datensatzes. Dazu wird ein erhobenes Merkmal als Variable im Analyseprogramm abgebildet und durch Variablenattribute beschrieben.

Schaukasten 2.1: Beispiele für Studiendesigns in den Sozialwissenschaften anhand der Dimensionen Stichprobe (*Sample*), Zeitpunkt (*Time*) und geographischer Bezug (*Space*)

- Querschnittsstudien:
 - Merkmale werden für eine Stichprobe (Sample = 1) einer geographischen Einheit (Space = 1) zu einem einzelnen Zeitpunkt (Time = 1) untersucht, z.B. werden alle 18-jährigen Personen in Deutschland im Jahr 2018 zu einem Thema befragt. Es entsteht ein Datensatz mit n Befragten und z Variablen, die die befragten Merkmale aus dem Fragebogen erfassen.
 - In vergleichender Perspektive werden je ein Sample (> 1) mit unterschiedlichem geographischen Bezug (Space > 1) zu einem Zeitpunkt (Time = 1) untersucht, z.B. das Wahlverhalten von Wahlberechtigten in unterschiedlichen Staaten. Es wird ein Datensatz je Sample mit n Befragten und z Variablen erzeugt und für Analyse Zwecke während der Datenaufbereitung und Dokumentation in einen Datensatz integriert; es entsteht ein sogenannter *integrierter Datensatz*.
- Vergleichende Studien im Zeitverlauf – Trendstudien:
 - Hier werden Merkmale anhand von unterschiedlichen Stichproben der gleichen Grundgesamt (Sample > 1) und gleichem geographischen Bezug (Space = 1) zu verschiedenen Zeitpunkten (Time > 1) untersucht, z.B. indem alle 18-Jährigen eines Landes im Jahr 2016, 2017 und 2018 befragt werden. Hieraus entstehen, je nach Anzahl der Erhebung, x einzelne Datensätze mit der den Frageitems entsprechenden Anzahl von Variablen, die ab der zweiten Erhebung ebenfalls integriert werden können. Da sich die Antworten mit den Erhebungszeitpunkten anhäufen, spricht man auch von einem *kumulierten Datensatz*.
 - Eine besondere Form des Untersuchungsdesigns stellen Panelstudien dar. Hier wird eine Stichprobe (Sample = 1) mit dem gleichen geographischen Bezug (Space = 1) wiederholt zu verschiedenen Zeitpunkten (Time > 1) befragt.
- Vergleichende Studien unterschiedlicher geographischer Räume im Zeitverlauf:
 - Dieses komplexe Studiendesign untersucht je ein Sample der gleichen Grundgesamt (Sample = 1) in verschiedenen Ländern (Space > 1) zu unterschiedlichen Zeitpunkten (Time > 1). Mit diesem Studiendesign wird eine große Anzahl einzelner länderspezifischer Datensätze je Erhebungszeitpunkt generiert, die mit jeder neuen Erhebungswelle zunehmen. Entsprechend aufwendig sind die Erstellung eines integrierten Datensatzes und seine Erweiterung, wenn neue Daten einer Welle in diesen Datensatz aufgenommen werden sollen.

Quelle: Eigene Darstellung

Somit müssen die einzelnen Variablen zunächst bestimmt, ggf. transformiert und eindeutig benannt werden. So kann beispielsweise die Frage nach dem Geburtsjahr einer Person zunächst in das approximierte Alter der Person transformiert (Jahr der Datenerhebung minus Geburtsjahr) und als neue Variable abgelegt werden. Diese sollte dann, ebenso wie die Ausgangsvariable (Geburtsjahr), mit einem sprechenden Namen versehen werden, wie z.B. Alter der befragten Person. Daran anschließend müssen die Variablen gelabelt, d.h. näher beschrieben werden. Label steigern die Verständlichkeit der Variable durch zusätzliche Informationen, die z.B. ersichtlich machen, dass es sich bei der Variable Alter um das Alter der jeweiligen Person handelt. Darüber hinaus kann das Label auch Informationen dazu liefern, dass die Variable Alter aus der Angabe zum Geburtsjahr abgeleitet wurde.

Analog müssen die einzelnen Werte jeder Variable mit eindeutigen Wertennamen versehen werden. Dies mag bei Angaben zum Alter trivial wirken, doch auch hier muss klar sein, ob die Werte der Variable z.B. in Jahren, Monaten oder Tagen angegeben sind. Die Relevanz eindeutiger Wertelabels lässt sich am Beispiel einer (dichotomen) Variable (0/1 kodiert), etwa zum Geschlecht der Person, verdeutlichen. Hier muss schriftlich fixiert werden, welcher Wert (0 oder 1) welches Geschlecht (weiblich oder männlich) wiedergibt. Dieses betrifft

nicht nur gültige Angaben, wie weiblich oder männlich, sondern z.B. auch fehlende Werte oder Antwortverweigerungen.

Im Verlauf der weiteren Datenaufbereitung bis hin zur Erstellung eines vollständigen Forschungsdatensatzes für die Analyse können dann schrittweise zusätzliche Variablen z.B. für Befragte, Länder oder Zeitpunkte (*administrative Variablen*) oder zur Gruppierung von Merkmalen bzw. zur Generierung weiterer Indizes (*inhaltliche Variablen*) definiert werden. Im Verlauf der Datenanalysen können Variablen ergänzt, modifiziert oder harmonisiert werden. Darüber hinaus müssen die Daten in der Aufbereitung auf eventuelle logische Fehler – z.B. die befragte Person ist 273 Jahre alt –, inkonsistentes Antwortverhalten – z.B. eine befragte Person in einem Singlehaushalt macht Angaben zu anderen (Mit-)Bewohner/innen – etc. kontrolliert und der Prozess der Datenerhebung und -aufbereitung entsprechend dokumentiert werden.

Ein vorausschauendes Forschungsdatenmanagement berücksichtigt dabei die Prozesse der Datenerhebung und -aufbereitung und seine konkreten zeitlichen Abläufe und Einzelschritte, die in den gesamten Forschungsprozess integriert werden müssen. Die Grundlage, wie die Daten dann im Prozessverlauf im Detail definiert, aufbereitet und bereinigt werden, wird durch entsprechend dokumentierte Standardregeln und Konventionen festgelegt (vgl. Netscher/Eder 2018; Ebel/Trixa 2015; Lück/Landrock 2014; Jensen 2012).

2.2 Forschungsprozess und der Lebenszyklus von Forschungsdaten

Im Umgang mit sozialwissenschaftlichen Forschungsdaten lassen sich unterschiedliche Akteure in verschiedenen Rollen definieren. So können Forschende als Datenproduzierende in der Primärerhebung auftreten oder aber im Rahmen der Sekundäranalyse Nachnutzende bereits existierender Daten sein (um ggf. selbst wieder Produzent neuer Daten zu werden). Darüber hinaus bestehen verschiedene Einrichtungen und Organisationen, die u.U. Einfluss auf die Erzeugung bzw. die (Nach-)Nutzung der Forschungsdaten haben, wie etwa Wissenschaftsorganisationen, Fachgesellschaften, Forschungsförderer oder Universitäten. Derartige Einrichtungen bringen ihre jeweiligen Perspektiven im Umgang mit den Forschungsdaten ein, erlassen (institutseigene) Richtlinien und Verpflichtungen, setzen Rahmenbedingungen für Forschungsvorhaben oder erstellen Empfehlungen zur richtigen Handhabung von Forschungsdaten. Zu diesen Organisationen zählen schließlich auch Dateninfrastruktureinrichtungen, wie Datenarchive oder Repositorien, die Forschungsdaten ggf. weiter aufbereiten und dokumentieren, vor allem aber archivieren, kuratieren und Dritten zur Nachnutzung bereitstellen (vgl. Abschnitt 2.3.2).

Die Themen in diesem Buch orientieren sich vor diesem Hintergrund an den konkreten Herausforderungen und Zielen von Datenproduzierenden und Datennutzenden, die mit der praktischen Planung von Forschungsvorhaben und ihre Umsetzung im laufenden Projekt verbunden sind. Einzelne Einrichtungen und Organisationen werden dabei in die Diskussionen mit einbezogen, soweit dies für die jeweilige Thematik zielführend ist. Der vorliegende Abschnitt fokussiert daher zunächst grundlegende Begriffe und Konzepte, deren Verständnis in den weiteren Kapiteln des Buches vorausgesetzt wird.

2.2.1 Die praktische Rolle von Lebenszyklusmodellen im Forschungsdatenmanagement

Die Nutzung von Phasen- bzw. *Lebenszyklusmodellen* ist in der Forschung und im Forschungsdatenmanagement weitverbreitet und wird je nach Gegenstand und Zielsetzung unterschiedlich stark differenziert. Solche Modelle beziehen sich auf ganz unterschiedliche Aspekte der empirischen Forschungstätigkeit. Aktivitäten und Abläufe des Forschungsprozesses reichen von der Formulierung des Forschungsproblems über die Planung des Studiendesigns bis zur Erhebung, Aufbereitung, Auswertung der Daten und der Publikation von Forschungsergebnissen. Zur detaillierten Beschreibung der Phasen des Forschungsprozesses sei hier auf die Darstellungen in Lehrbüchern zu den Methoden der empirischen Sozialforschung verwiesen (vgl. u.a. Baur/Blasius 2014; Schnell/Hill/Esser 2013; Diekmann 2007).

Im Kontext des Forschungsdatenmanagements und der Kultur des Data Sharing haben Modelle zum Lebenszyklus von Forschungsdaten eine spezifische Popularität erlangt. Sie sind ein hilfreiches Werkzeug zur Planung und Umsetzung komplexer Arbeitsabläufe, die speziell auf den praktischen Umgang mit den Daten im Forschungsprozess – im weitesten Sinne – fokussiert sind. Sie ergänzen die Phasen des Forschungsprozesses und typische Projektabläufe, indem sie explizit die Sicherung, Archivierung, Erschließung und Nachnutzung von Forschungsdaten nach Projektende thematisieren.

In einer sehr einfachen Form lassen sich fünf grundlegende Phasen des Lebenszyklus von Forschungsdaten definieren:

1. Die *Recherche* nach bereits vorhandenen Daten dient zum einen der Kontrolle, ob zur Bearbeitung des Forschungsvorhabens nicht bereits existierende Forschungsdaten genutzt werden können, zum anderen ist sie Ausgangspunkt der Planung des eigenen Messinstruments.
2. In der Phase der *Studienplanung* wird die Sammlung der notwendigen Informationen konkret geplant. Dies betrifft die Definition der Untersuchungsobjekte bzw. Merkmalsträger, das Studiendesign, die Erstellung des Messinstruments etc.
3. Darauf aufbauend werden in der Phase der *Datenerhebung* die Informationen gesammelt, aufbereitet und in eine entsprechende Datenstruktur überführt.
4. In der daran anschließenden Phase der *Datenanalyse* liegt zumeist das eigentliche Interesse der Forschenden. Hier werden die zuvor erstellten Forschungsdaten mit Blick auf die zugrunde liegende Forschungsfrage ausgewertet.
5. Die letzte Phase umfasst die *langfristige Sicherung* der genutzten Forschungsdaten zu Replikationszwecken bzw. deren *Bereitstellung* zur Nachnutzung durch Dritte.

Diese Phasen schaffen einen ersten groben Überblick über die wichtigsten Ereignisse, die beim Umgang mit Forschungsdaten und damit im Kontext des Forschungsdatenmanagements von Bedeutung sind. Um Anforderungen an Arbeitsabläufe, Regeln und Konzepte des jeweiligen Forschungsvorhabens konkret zu planen, sind die genannten Phasen ggf. durch weitere zu ergänzen und die jeweiligen Datenergebnisse im Lebenszyklus zu definieren. Dazu wurde mittlerweile – je nach Typ und Herkunft der Daten und disziplinspezifischer Fragestellung des Forschungsdatenmanagements – eine Vielzahl von Lebenszyklusmodellen und -typologien entwickelt (vgl. Cox/Tam 2018; Carlson 2014; Ball 2012).

Praxisbezogene Publikationen im sozialwissenschaftlichen Kontext nutzen Lebenszyklen zur Beschreibung von Leitlinien und Best-Practice-Empfehlungen zur Gestaltung von Arbeitsabläufen, Regeln und Konzepten im Rahmen des Managements von disziplinspezifischen Forschungsdaten (vgl. Corti et al. 2014; ICPSR 2012; Jensen 2012). Andere Modelle wiederum sind auf bestimmte Studientypen ausgelegt, wie z.B. Längsschnittstudien oder Kulturvergleiche (vgl. CCSG – Cross-Cultural Survey Guidelines 2018). Solche praxisnahen Modelle und Empfehlungen lassen sich an die konkreten Bedürfnisse und Anforderungen des jeweiligen Forschungsvorhabens anpassen. Sie dienen insofern als Blaupause zur projektspezifischen Planung notwendiger Arbeitsschritte und Maßnahmen entlang der Phasen

mit datenrelevanten Ereignissen. Ausgehend von den konkreten datenbezogenen Anforderungen wird für jede einzelne Phase festgelegt, welche spezifischen Forschungsdatenmanagementkonzepte und Arbeitsschritte wann, wie und von wem umgesetzt werden sollen. Auf dieser Grundlage kann auch der Bedarf an Ressourcen für die unterschiedlichen Maßnahmen und Aktivitäten des Forschungsdatenmanagements abgeschätzt werden. In diesem Zusammenhang wird im Folgenden das *Data Lifecycle Model der Data Documentation Initiative (DDI)* als Standard zur Dokumentation sozialwissenschaftlicher Daten vorgestellt.

2.2.2 Daten und Metadaten entlang des DDI Data Lifecycle

Der DDI-Standard ist ein internationaler De-Facto-Standard zur Beschreibung von Daten aus Umfragen und anderen Beobachtungsmethoden in den Sozial-, Verhaltens-, Wirtschafts- und Gesundheitswissenschaften, der von der DDI-Alliance entwickelt wird. Der Fokus liegt auf Mikrodaten aus unterschiedlich komplexen Umfragen oder aus administrativen Quellen und schließt auch die Beschreibung aggregierter Daten ein. Zur systematischen Strukturierung, Erfassung und Organisation der Informationen einer Studie (Daten und Metadaten) hat die Alliance einen disziplinspezifischen DDI Data Lifecycle (2018) entwickelt. Dieser identifiziert acht abgrenzbare Phasen, die den Lebenszyklus der Produktion von Forschungsdaten systematisch strukturieren. Das Modell bildet die konzeptuelle Basis zur Erfassung von Informationen über Daten (d.h. *Metadaten*) entlang der Phasen ihres Lebenszyklus. Gleichzeitig liefert das Modell eine anschauliche Grundlage zur Planung von datenbezogenen Ereignissen und Anforderungen an die Datendokumentation im Rahmen des Forschungsdatenmanagements.

Zum grundlegenden Verständnis für die folgenden Kapitel dieses Buches, die sich implizit oder explizit auf den DDI-Standard und dessen Lebenszyklusmodell beziehen, wird der DDI Data Lifecycle, wie in Abbildung 2.1 zu sehen, kurz vorgestellt. Die Phasen müssen nicht notwendigerweise linear, d.h. von Anfang bis Ende, durchlaufen werden. Vielmehr ermöglicht das Modell, jede einzelne Phase unabhängig von den anderen Phasen hinsichtlich der spezifischen datenbezogenen Ereignisse und Anforderungen an die Dokumentation der Daten zu betrachten. Dabei geht das Lifecycle-Konzept davon aus, dass die jeweils spezifischen Informationen über Daten an den Stellen erfasst werden, an denen sie entstehen und für besondere Zwecke erforderlich sind.

In der Phase *Studienplanung (Concept)* wird die Studienkonzeption entwickelt, indem das Projekt Forschungsfrage und Methodik der Untersuchung und Verfahren zur Erhebung der Daten festlegt. Teil der Studienplanung ist neben der klassischen Literaturrecherche zum Forschungsstand auch die Erschießung und Evaluierung vorhandener Daten, die zur eigenen Forschungsfrage vorliegen. Bei der Festlegung des generellen Studiendesigns in Abhängigkeit von der Fragestellung (z.B. Querschnittsstudie; Vergleich über Raum und Zeit; Experiment etc.) sind methodische Fragen der zu untersuchenden Population, wie etwa zur Grundgesamtheit, Stichprobe oder zum Auswahlverfahren, zu klären. Wird ein standardisierter Fragebogen genutzt, sind konkrete Beobachtungen (wie z.B. Alter) und theoretische Konstrukte (wie z.B. Einkommen) in geeignete Fragetypen und Frageitems auf Grundlage von (bereits vorhandenen) Messkonzepten, Skalen und Indizes umzusetzen (Operationalisierung) und zu testen (Validität; Reliabilität). Mit der Entscheidung über den Aufbau des Messinstruments wird zugleich auch die Grundlage für die Definition der Daten des späteren Datensatzes und seiner inhaltlichen Variablen gelegt. Die Dokumentation, welche Methoden der empirischen Sozialforschung zur Produktion empirischer Forschungsdaten genutzt wurden, ist ein zentraler Inhalt der Informationen über Daten in dieser Phase. Diese Kontextinformationen über

die Art und Weise der Produktion von Forschungsdaten sind über ein Vorhaben hinaus unverzichtbar, wenn Daten replizierbar und nachnutzbar bleiben sollen.

Abbildung 2.1: DDI Data Lifecycle



Quelle: DDI Data Lifecycle (2018)

In der Phase der *Datenerhebung (Collection)* wird die Befragung durchgeführt, ggf. in Kooperation mit einem Erhebungsinstitut. In dieser Phase fallen – über die erhobenen Daten hinaus – weitere Informationen über die Produktion der Daten an. Dies betrifft insbesondere den tatsächlich eingesetzten Fragebogen und die Form der Befragung, wie z.B. persönlich-mündlich, telefonisch, schriftlich, mit oder ohne Interviewer bzw. Computereinsatz etc. Dabei ist es aus Sicht des Forschungsdatenmanagements durchaus eine Herausforderung, diese Erhebungsformen im Interesse der späteren Datenarchivierung und -nachnutzung transparent und nachvollziehbar zu dokumentieren. Gleiches gilt für Informationen, die – neben den Daten – durch deren Erhebungsprozess entstehen. Diese sogenannten Paradata dienen der Prozess- und Qualitätskontrolle der Durchführung einer Befragung (vgl. Felderer/Birg/Kreuter 2014: 357).

Die Phase *Datenaufbereitung und -analyse (Processing)* ist der nächste zentrale Abschnitt des Lebenszyklusmodells. In dieser Phase werden die erhobenen Daten im Forschungsprojekt durch inhaltliche bzw. administrative Variablen ergänzt, die Datensatzstruktur endgültig definiert und die erhobenen Daten aufbereitet. Dabei entstehen Kontextinformationen, die für das Verständnis der Daten und für das Arbeiten mit ihnen notwendig sind. Gleichzeitig werden in dieser Phase die Daten systematisch auf Fehler geprüft und bereinigt. Erst auf dieser Grundlage kann der analysefähige Datensatz zur eigentlichen Auswertung erstellt werden. Die zugrunde liegenden Maßnahmen und Regeln der Datendefinition und der Datenbereinigung sind relevante Informationen über den Umgang mit den Daten, die zur Sicherung der Datentransparenz und Datenqualität in Codebüchern oder Methodenberichten dokumentiert werden sollten. Die Prozesse der wissenschaftlichen Datenanalyse der zuvor erhobenen Primärdaten sind nicht Gegenstand des gezeigten DDI-Data-Lifecycle-Modells.

Die Datenprodukte aus der Datenaufbereitung können dann entweder in die Phase der *Datenarchivierung* oder der *Datenbereitstellung* übergehen, wenn ein z.B. ein Forschungsprojekt im Rahmen eigener Ressourcen die Daten bereitstellt. In der Phase der *Datenarchivierung* (*Archiving*) werden die Forschungsdaten in einem Archiv oder einem Datenrepositorium zu Replikationszwecken oder zur langfristigen Nachnutzung gesichert. Die Informationen über Daten werden in diesem Kontext durch unterschiedliche Arten von Metadaten für spezifische Zwecke neu erstellt, die u.a. der Aufnahme und Archivierung der Daten unter Berücksichtigung des Datenschutzes dienen und den Datensatz als Ganzes sowie die darin enthaltenen Variablen beschreiben.

In der Phase der *Datenbereitstellung* (*Distribution*) werden vor allem Regelungen festgelegt, wie, von wem und zu welchem Zweck die Daten datenschutzkonform genutzt werden können und wie die Daten zitiert werden können. Weitere Informationen beschreiben außerdem u.a. den Umfang und technische Formate der bereitgestellten Daten und der dazugehörigen Dokumentationen.

Eng mit der Datenbereitstellung verbunden ist die darauffolgende Phase der *Datensuche* (*Discovery*). Daten, die nachnutzbar sein sollen, müssen auffindbar und zugänglich sein. Dazu werden in Archiven oder Datenrepositorien gespeicherte Daten in Datenkatalogen publiziert. Damit Forschende diese Bestände gezielt nach für sie relevante Daten durchsuchen können, informieren indexierte Metadaten über die wesentlichen Aspekte wie Inhalt, Methodik, Verfahren, Untersuchungseinheit, Gebiet und Zeitraum der Datenerhebung. Sie beinhalten Angaben zu den Datenproduzierenden, zur Zitation der Daten sowie zu bereits bestehenden Publikationen auf Basis dieser Daten. Die Zusammenstellung dieser Informationen in einem Datenkatalog wird *Studienbeschreibung* genannt. Im Rahmen vorher festgelegter Zugangs- und Vertriebsbedingungen sind die Daten so zur weiteren *Analyse* (*Analysis*) verfügbar.

Die letzte Phase des DDI-Lifecycles thematisiert daher auch direkt die *Datennachnutzung* (*Repurposing*). Im Sinne einer Sekundärnutzung behandelt das Modell Informationen über Daten, die im Zuge der Harmonisierung von (großen) Datenbeständen und entsprechenden Datentransformationen durch nachgelagerte Projekte entstehen. Die Ergebnisse solcher Harmonisierungsprojekte können dann wieder zu neuen Informationen über Daten in der Phase der *Datenaufbereitung* (*Processing*) führen. Um die Vergleichbarkeit von Daten eines Umfrageprogramms zu ermöglichen, werden z.B. vergleichbare Variablen aus unterschiedlichen Erhebungszeitpunkten harmonisiert, z.B. um dadurch eine Zeitreihe zu erstellen. Diese können ihrerseits als neue Daten die nachfolgenden Phasen des Data-Lifecycles durchlaufen.

2.3 Forschungsdatenmanagement in den Sozialwissenschaften

Lebenszyklusmodelle wie der DDI Data Lifecycle organisieren die Datenproduktion und die dabei entstehenden Informationen von Planung bis zur Bereitstellung und Nachnutzung der Daten, um Forschende bei der Erstellung qualitativ hochwertiger Daten zu unterstützen. Zu bedenken ist jedoch, dass Lebenszyklusmodelle nicht unbedingt dafür entwickelt wurden, den Umgang mit Daten im Sinne des Forschungsdatenmanagements insgesamt zu planen. *Forschungsdatenmanagement*, verstanden als Summe aller Strategien, Maßnahmen, Konventionen, Arbeitsabläufe etc., die den Umgang mit Forschungsdaten im Projektverlauf und darüber hinaus regeln, beinhaltet auch organisatorische, technische und administrative Aspekte, die nicht explizit Gegenstand eines datenzentrierten Lebenszyklusmodells sind.

2.3.1 *Forschungsdatenmanagement entlang des Lebenszyklus*

Die einzelnen Maßnahmen des Forschungsdatenmanagements werden entlang des Lebenszyklus der Forschungsdaten geplant und die entsprechenden Aktivitäten in der jeweiligen Phase umgesetzt. Dabei ist zu berücksichtigen, dass verschiedene Aktivitäten und Maßnahmen quer zum Lebenszyklus liegen und mehrere, wenn nicht sogar alle Phasen betreffen.

Dies gilt zuallererst für eine systematische Vorbereitung des Forschungsdatenmanagements im Rahmen der eigentlichen Projektplanung. Diese muss über den gesamten Projektverlauf stetig aktualisiert und an sich ändernde Gegebenheiten angepasst werden. Nur durch vorausschauendes Handeln und eine planvolle Herangehensweise kann sichergestellt werden, dass im Laufe des Forschungsprojekts qualitativ hochwertige Forschungsdaten erstellt und für die Analyse genutzt werden können. Kapitel 3 dieses Buches befasst sich daher näher mit der Planung und Umsetzung des Forschungsdatenmanagements.

Zweitens läuft die Frage der Organisation der Forschungsdaten quer zu ihrem Lebenszyklus. Über den gesamten Projektverlauf und darüber hinaus ist sicherzustellen, dass die richtigen Daten in der richtigen Version weiterverarbeitet, dokumentiert und analysiert werden. Die Daten müssen dabei in ihren unterschiedlichen Stadien, wie etwa Rohdaten und aufbereitete Forschungsdaten, ausreichend gegen Verlust und unautorisierten Zugriff geschützt und ggf. zwischen den Projektbeteiligten kommuniziert bzw. transferiert werden, wie in Kapitel 5 näher ausgeführt wird.

Drittens werden in den Sozialwissenschaften häufig Informationen über Menschen gesammelt und analysiert. Dabei kommen datenschutzrechtliche Maßnahmen zum Einsatz, die überprüfen, ob die Daten überhaupt erhoben, gespeichert und verarbeitet werden dürfen. Dies betrifft zunächst einmal die informierte Einwilligung der befragten bzw. beobachteten Personen. Die Einwilligung muss spätestens in der Phase der Studienplanung vorbereitet, in der eigentlichen Datenerhebung eingeholt und bis zu der endgültigen Vernichtung der Daten über alle weiteren Lebensphasen hinweg verwaltet werden. Darüber hinaus müssen die betroffenen Personen vor negativen Folgen aufgrund der Befragung oder Beobachtung geschützt werden. Kapitel 4 befasst sich in diesem Zusammenhang mit der Anonymisierung von Umfragedaten und liefert Ansätze und Beispiele für den rechtskonformen Umgang mit Forschungsdaten.

Demgegenüber lassen sich andere Maßnahmen des Forschungsdatenmanagements augenscheinlich konkreten Phasen im Lebenszyklus der Daten zuweisen. So ist das inhaltliche Füllen einer Datenmatrix zwar erst nach oder frühestens im Laufe der eigentlichen Datenerhebung möglich. Dennoch sollte die Datenmatrix bereits vor dieser Phase als Blaupause aufgesetzt werden. Dies dient neben der Planung der eigentlichen Datenerhebung und ihrer Dokumentation zudem der Implementierung geeigneter Maßnahmen zur Kontrolle der Datenqualität sowie zur Entwicklung von Richtlinien zur Datenaufbereitung, wie in Kapitel 6 näher beschrieben. Zwar setzen die meisten dieser Maßnahmen und Arbeitsabläufe erst in bzw. nach der Phase der Datenerhebung an, sie müssen aber bereits in der Projektplanung und der Erstellung des Studiendesigns bzw. des Messinstruments angegangen werden.

Spezifische Metadaten und Standards sind – implizit oder explizit – in alle Phasen des Lebenszyklus eingebunden. Metadatenstandards unterstützen die Zitierbarkeit und Wieder auffindbarkeit von Forschungsdaten und liefern damit einen zentralen Beitrag zum guten wissenschaftlichen Arbeiten und zur Anerkennung der Leistung von Forschenden im Kontext des Forschungsdatenmanagements. Kapitel 10 befasst sich ausführlich mit grundlegenden Konzepten und Services, die das Zitieren von Forschungsdaten in wissenschaftlichen Arbeiten ermöglichen. Darüber hinaus sind Informationen über die Daten entlang des gesamten Lebenszyklus von zentraler Bedeutung, um Daten zu finden, mit ihnen zu arbeiten und sie zu analysieren, wie in Kapitel 9 ausführlich erörtert wird.

Analog zur Phase der Aufbereitung ist auch die Archivierung der Forschungsdaten im DDI Data Lifecycle als einzelner Schritt dargestellt. Doch bereits die Vorbereitung der Archivierung, die Übergabe der Daten an ein Datenarchiv oder Repositorium und die Sicherstellung der rechtskonformen Nachnutzbarkeit der Daten durch Dritte ist keineswegs ein einzelner kleiner Schritt im Leben der Forschungsdaten. Vielmehr setzt auch hier ein planvolles Forschungsdatenmanagement frühzeitig im Lebenszyklus der Daten an und bereitet die Erstellung und Übergabe zu archivierender Daten und Dateien systematisch vor. Kapitel 7 befasst sich näher mit den verschiedenen Möglichkeiten der längerfristigen Sicherung und Bereitstellung von Forschungsdaten.

Jenseits der Phase der Datenarchivierung wurden am Beispiel des DDI-Data-Lifecycles Abläufe im Kontext der Harmonisierung von großen Datenbeständen aufgezeigt, um den Bestand vergleichbarer Daten zu erweitern. Dem erfahrenen Datennutzenden wird aber aufgefallen sein, dass die Nachnutzung von Daten im Kontext von Projekten zu Sekundäranalysen in diesem Modell nicht als eigene – spezialisierte – Phase ausgewiesen ist. Explizit wird nur auf die Erstellung von Codes und Routinen zur Replikation sowie auf Aspekte der Harmonisierung bereitgestellter Daten hingewiesen. Deshalb erörtert Kapitel 8 einige bislang vernachlässigte Fragen im Forschungsdatenmanagement in der Sekundäranalyse.

2.3.2 Die Rolle von Forschungsdateninfrastrukturen im Forschungsdatenmanagement

In diesem Kapitel wurde der Umgang mit Forschungsdaten bislang hauptsächlich aus der Sicht von Datenproduzierenden bzw. Datennutzenden thematisiert. Dabei wurde wiederholt auf die Nutzung von Datenarchiven und Repositorium hingewiesen. Wie solche Infrastrukturen Forschende und ihre Forschungsvorhaben beim Forschungsdatenmanagement unterstützen können, soll im Folgenden kurz erörtert werden.

Forschungsinfrastrukturen stellen grundlegende Einrichtungen, Ressourcen und Dienstleistungen im Wissenschaftssystem bereit, um die Wissenschaftsgemeinschaft bei ihrer Forschung langfristig zu unterstützen. Seit den 1920er Jahren wird der Begriff zunächst nur für die naturwissenschaftliche Forschung mit großen technischen Anlagen benutzt. Im Zuge neuer digitaler Informations- und Kommunikationstechnologien wurden seit den 1970er Jahren auch in den Geistes- und Sozialwissenschaften eigene Forschungsinfrastrukturen aufgebaut oder weiterentwickelt. Die neuen Technologien erlaubten es, Informationen und empirische Forschungsdaten schrittweise immer effizienter und qualifizierter (digital) zu organisieren und bereitzustellen. Der Typus *Informationsinfrastrukturen* kennzeichnet dabei Einrichtungen wie Bibliotheken, Sammlungen und Archive (RfII – Rat für Informationsinfrastrukturen 2016; Wissenschaftsrat 2011). Im Zuge von nationalen wie internationalen Initiativen zur Verbesserung und Weiterentwicklung von Infrastrukturen für die Erhebung, Archivierung und Bereitstellung von Forschungsdaten etablierte sich in den Sozial- und Wirtschaftswissenschaften der Begriff *Forschungsdateninfrastrukturen* (vgl. RfII 2016:2).

In Deutschland werden die meisten Forschungsdateninfrastrukturen durch Bund und Länder, das Bundesministerium für Bildung und Forschung (BMBF) oder die DFG gefördert. So wurde beispielsweise im Zuge der Empfehlungen zur Weiterentwicklung der Dateninfrastruktur der Rat für Sozial- und Wirtschaftsdaten (RatSWD) 2004 gegründet (vgl. KVI 2001). Seitdem hat der RatSWD im Rahmen seiner Aufgaben zur Förderung der disziplinspezifischen Dateninfrastruktur über dreißig datenproduzierende Forschungsdatenzentren (FDZ) aus unterschiedlichsten Fachdisziplinen akkreditiert. Dadurch verpflichten sich die Forschenden, ihre Daten qualitätsgesichert und datenschutzkonform sowie möglichst einfach zur Verfügung zu stellen (vgl. RatSWD Arbeitsprogramm 2017–2020).

Teil der sozialwissenschaftlichen Forschungsdateninfrastruktur sind forschungsbasierte Einrichtungen der Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz e.V. (kurz Leibniz-Gemeinschaft). Dazu gehören u.a. das Leibniz-Informationszentrum Wirtschaft (ZBW) und das GESIS – Leibniz-Institut für Sozialwissenschaften. Die Forschungsdateninfrastrukturen sind ihrerseits wiederum international vernetzt. So ist z.B. das GESIS Datenarchiv für Sozialwissenschaften auf europäischer Ebene eng mit den Dienstleistungen und Ressourcen des Consortium of European Social Science Data Archives (CESSDA ERIC) verzahnt. Kernbestandteil der Forschungsdateninfrastrukturen sind die großen sozialwissenschaftlichen Umfrageprogramme, die auf nationaler, europäischer und transnationaler Ebene im Rahmen der Forschungsförderung etabliert wurden, wie etwa das Sozio-ökonomische Panel (SOEP), die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) oder das European Social Survey (ESS) etc. (vgl. Wissenschaftsrat 2011: Anhang 1).

Aus Sicht der Datenproduzierenden bzw. Datennutzenden stellt sich die Frage, welche praktischen Leistungen und Ressourcen Forschungsdateninfrastrukturen für ihr individuelles Forschungsvorhaben anbieten können. Zu nennen sind hier natürlich zuallererst die vielfältigen Informations-, Beratungs- und Schulungsangebote, die an den spezifischen Bedarf von Forschenden und Disziplinen angepasst sind. Diese reichen von praktischen Handreichungen zum Forschungsdatenmanagement bis hin zu spezialisierten Informationsveranstaltungen, Workshops, Konferenzen oder Summer Schools.

Betrachtet man die Leistungen und Ressourcen der Forschungsdateninfrastrukturen im Kontext des Lebenszyklus von Forschungsdaten, unterstützen Datenkataloge von Datenarchiven, Repositorien oder Datenzentren Forschende in der Phase der Studienplanung bei ihren Recherchen nach bereits existierenden Datenbeständen und zugrundeliegenden Messinstrumenten. Diese Informationsbestände unterstützen – neben klassischen Literaturrecherchen zum Forschungsthema – die weitere Planung des jeweiligen Forschungsvorhabens. So kann die Nachnutzung bereits bestehender Daten angesichts von deren Menge und Vielfalt eine effiziente Alternative zur eigenen Datenerhebung darstellen oder das Verknüpfen neu erstellter Daten mit bereits bestehenden Beständen ermöglichen.

In den nachfolgenden Phasen der Datenerhebung und Datenaufbereitung geht es für Forschende zumeist um Fragen der Qualitätssicherung – sowohl mit Blick auf die erstellten Daten als auch in Bezug auf die z.B. im Rahmen der Feldarbeit anfallenden Kontextinformationen. Während Erhebungsinstitute Forschenden die Datenerhebung an sich abnehmen, fokussieren sich Forschungsdateninfrastrukturen in diesen Phasen vor allem auf Best-Practice-Empfehlungen, Richtlinien und Handreichungen. Zu nennen ist hier beispielsweise abermals der DDI Standard, der, wie in Kapitel 9 näher erörtert, auf Initiative von Datenarchiven implementiert wurde und auf deren Betreiben bis heute aktiv weiterentwickelt wird. Im Bereich der Datenaufbereitung entstehen auf Seiten der Forschungsdateninfrastrukturen zudem mehr und mehr Dienstleistungsangebote, etwa zur Datendokumentation auf Basis internationaler Standards, wie DDI. Diese sind z.T. zwar kostenpflichtig, nehmen den Datenproduzierenden jedoch erhebliche Arbeiten ab und schaffen so Freiräume für die eigentliche Forschung (vgl. GESIS Datenservices o.J.).

Für Forschende ist die Phase der Datenanalyse natürlich von zentralem Interesse, da sie der Überprüfung der zugrunde liegenden Forschungsfrage(n) dient und einen wesentlichen Bestandteil späterer Publikationen der erzielten Forschungsergebnisse darstellt. Dabei gilt der bereits mehrfach erwähnte Grundsatz guter wissenschaftlicher Praxis, nachdem diese Ergebnisse über das eigentliche Forschungsprojekt hinaus replizierbar sein müssen. Demnach sollten Forschungsdaten, auf deren Basis Forschungsergebnisse publiziert wurden, auch nach dem Projektende aufbewahrt werden. Gerade qualitativ hochwertige Forschungsdaten sollten des Weiteren zur langfristigen Nachnutzung durch Dritte bereitgestellt werden, indem sie an Einrichtungen der Dateninfrastruktur übergeben werden. Forschungsdateninfrastrukturen

unterstützen dazu Forschende durch zweckorientierte Angebote. Beispielsweise verfügen viele Hochschulen und Forschungseinrichtungen heute über unterschiedliche organisatorische und technische Dienste zur Sicherung, Archivierung und Bereitstellung von Daten, die in der eigenen Einrichtung produziert wurden. Die Nutzung solcher Dienste kann z.T. verbindlich durch Empfehlungen der Einrichtung zum Umgang mit Forschungsdaten geregelt sein. Dabei ist die jeweilige datenspezifische Ausrichtung einer Einrichtung oder eines Datenarchivs zu berücksichtigen. So stellt das GESIS Datenarchiv vorrangig Dienste zur Selbst-Archivierung (vgl. GESIS datorium) sowie umfangreiche Services im Rahmen der Langzeitarchivierung für quantitative Umfragedaten (vgl. Quandt/Mauer 2012: 67) bereit. Weiterhin sind auch Entwicklungen wissenschaftlicher Verlage zu berücksichtigen. Diese gehen mehr und mehr dazu über, Datensätze und Materialien, auf deren Basis Artikel in ihren Journalen veröffentlicht werden, zu sichern, um so deren Replizierbarkeit zu ermöglichen. Entsprechende Dienste werden z.T. gemeinsam mit Forschungsdateninfrastrukturen (vgl. ZBW Journal Data Archive) angeboten. Das Themenspektrum von der Sicherung bis zur langfristigen Nutzung von Forschungsdaten wird ausführlich in Kapitel 7 behandelt.

Ein spezieller Aspekt des Forschungsdatenmanagements stellt die Ermöglichung der *Zitation von Forschungsdaten* dar. Dem liegt die Forderung vieler Data Policies zugrunde, die Erzeugung und Bereitstellung von Forschungsdaten als Forschungsprodukt und somit als eigenständige Forschungsleistung anzuerkennen. Um entsprechende *credits* zu ermöglichen, haben Forschungsdateninfrastrukturen Systeme mit *dauerhaften Identifikatoren (Persistent Identifier)* für Datensätze und zugehörige Informationstypen entwickelt. Persistent-Identifier-Systeme, wie URN, DOI oder Handle, dienen der dauerhaften Referenzierung digitaler Ressourcen, wie z.B. der Publikationen und Datensätze. In den sozialwissenschaftlichen Forschungsdateninfrastrukturen wird heute zumeist das DOI-System genutzt. Die Entwicklung solcher Systeme erfolgt initial durch entsprechend geförderte Infrastrukturprojekte mit mehreren Partnern, die dann als Regelbetrieb verstetigt werden. Die Vergabe einer DOI ist in der Regel mit der Archivierung bzw. Publikation eines Forschungsdatensatzes über ein Datenarchiv, Repositorium, Forschungsdatenzentrum oder einen Verlag verbunden, die an das DOI-System angeschlossen sind. Die DOI erfüllt dabei mehrere Zwecke. Mit ihrer Hilfe können die Daten zitiert werden und sind so dauerhaft und weltweit identifizierbar sowie im Internet auffindbar. Dadurch wird zugleich der Datenzugang für die Nachnutzung und Replizierbarkeit der Daten erheblich vereinfacht. Weiterhin wird die Messung der Datennutzung durch die Aufnahme in Zitationsindizes gefördert. Somit erhöht sich die Reputation der Forschenden ebenso wie die Sichtbarkeit der Daten. Erste empirische Untersuchungen zeigen, dass Publikationen, die auch die zugrunde liegenden Daten veröffentlichen, häufiger zitiert werden, als Veröffentlichungen ohne Datennachweis (vgl. Piwowar/Vision 2013; Piwowar/Day/Frisdman 2007). Die Anwendung und Organisation des DOI-Systems ist Thema von Kapitel 10 in diesem Buch.

Wie an verschiedenen Stellen gezeigt, erfassen standardisierte Metadaten komplexe Informationen über die Daten und ihre Produzenten. Sie informieren entlang des Lebenszyklus von Daten über die Genese ihrer Entstehung bis hin zur Nachnutzung in neuen Forschungskontexten. Je nach Anforderungen und Zweck der Informationen über Daten, Personen, relevante Quellen und Objekte der Forschung werden verschiedene Metadatenstandards (vgl. Jensen/Katsanidou/Zenk-Möltgen 2011) angewendet. Sie müssen zugleich zwischen verschiedenen technischen Systemen austauschbar, d.h. interoperabel, sein, damit Forschende z.B. in Suchportalen gebündelt auf Informationen zugreifen können, die aus verschiedenen Quellen stammen. Diese Aspekte werden in Kapitel 9 näher erörtert. Die Entwicklung und Implementierung von Metadatenstandards, technischen Systemen und Werkzeugen für die Erfassung und den Austausch von Metadaten erfolgt u.a. im Rahmen nationaler oder internationaler Infrastrukturprojekte.

Der Rat für Informationsinfrastrukturen hat 2016 das Positionspapier „Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland“ (RfII 2016) veröffentlicht, das facettenreich und informativ die vielfältigen Herausforderungen und Lösungsansätze in diesem Umfeld thematisiert.

2.3.3. *Neue Datentypen und Herausforderungen an das Forschungsdatenmanagement*

In den bisherigen Abschnitten sind Fragen des sozialwissenschaftlichen Forschungsdatenmanagements mit Bezug auf die Erhebung und Nachnutzung von Umfragedaten thematisiert worden. Gleichzeitig wird diese Datenbasis verstärkt durch (digitale) Daten aus (neu verfügbaren) Quellen erweitert, die ursprünglich nicht für Forschungszwecke entwickelt wurden. So stellt beispielsweise die Organisation for Economic Co-operation and Development (OECD 2013: 12) zur Bedeutung derartiger Daten fest:

[...] forms of social science data not specifically designed for re-search purposes are emerging as important alternatives and additions to more standard sources. Various types of administrative data, while not new, have become newly accessible in the form of electronic records, while entirely new forms of social science data have emerged as a consequence of the internet revolution.

Um welche Typen und neue Formen von Daten es sich dabei handelt, zeigt beispielhaft der Schaukasten 2.2 (vgl. OECD 2013; zu deutschen Datenquellen vgl. z.B. ZBW/GESIS/RatSWD 2018). Die Daten der Kategorie A und B sind nur insofern neu, weil sie z.T. erst durch die schrittweise Öffnung dieser Datenbestände für Forschungszwecke erschlossen werden können. Relativ neu ist in den Sozialwissenschaften die Nutzung von vorhandenen Daten aus anderen Disziplinen, wie z.B. der Medizin oder den Geowissenschaften. Demgegenüber stellen die Daten der Kategorien C bis E eine Flut neuer Datentypen dar, die in unterschiedlichsten Dateiformaten vorliegen können (Text, Bild, Audio) (vgl. OECD 2013; RatSWD 2012).

Schaukasten 2.2: Beispiele für digital verfügbare neue Datentypen und neuartige Datenformen

- A Prozessproduzierte Daten (Mikrodaten) der öffentlichen Verwaltung, z.B. Meldesysteme, Steuern, Sozialversicherung, Arbeitsmarkt, Umwelt, Haus- und Landnutzung, Gesundheit, Wirtschaft
- B Daten der Amtlichen Statistik, wie Aggregat- oder Makrodaten, auf nationaler und internationaler Ebene, z.B. Destatis, Eurostat, OECD
- C Transaktionsdaten im Finanz-, Versicherungs- und Konsumsektor, z.B. Kaufverhalten, Mobilfunknutzung, Kundenkarten
- D Daten zur Internetnutzung, z.B. Suchbegriffe, Downloads, Weblogs, soziale Netzwerke, News
- E Trackingdaten, z.B. GPS, Verkehr, Internutzug, Kameraüberwachung, Satellitenbeobachtung

Quelle: Eigene Darstellung

Forschende in den Sozialwissenschaften nutzen diese neuen Typen und Formen ergänzend zu klassischen Umfragedaten. So werden z.B. Umfragedaten gemeinsam mit medizinischen Parametern (Biomarker) erhoben oder im Rahmen eines Forschungsvorhabens verknüpft (vgl. Jensen et al. 2015: 17f). Dadurch entstehen neue Forschungspotentiale und Erkenntnis-möglichkeiten. So hinterlassen etwa die Aktivitäten von Nutzenden auf Verkaufsportalen, Suchmaschinen oder sozialen Netzwerken digitale Verhaltensspuren – auch digitale Spurendaten genannt – die es ermöglichen, Einstellungen und Verhaltensweisen zu analysieren. Derartige Daten können bewusst erzeugt sein, z.B. durch Beiträge auf Facebook, oder aber unbewusst und nicht direkt sichtbar gesammelt werden, z.B. wenn Metadaten über ein Kommunikationsgeschehen den Zeitpunkt, das Gerät und den Standort der Kommunikation erfassen.

Solche neuen Datentypen und Datenformen sowie die zugrunde liegenden Datenquellen stellen das Forschungsdatenmanagement und die handelnden Akteure auf verschiedenen Ebenen vor neue Herausforderungen. Einige Aspekte sind im Schaukasten 2.3 beispielhaft ohne Anspruch auf Vollständigkeit zusammengestellt.

Schaukasten 2.3: Herausforderungen im Umgang mit neuen Daten

- Forschungsprozess:
 - Entwicklung gemeinsamer Forschungsmethoden und -standards bei der Analyse von internet-basierten Kommunikationsformen; z.B. bei der Untersuchung politischer Kommunikation auf Social-Media-Plattformen
 - übergreifende Anforderungen an die Vergleichbarkeit von Studien und Ergebnissen aufgrund unterschiedlicher Konzepte sowie neuartiger Fragestellungen und Untersuchungsdesigns
 - Grundanforderungen an die Dokumentation von Daten und Methoden zur Nachvollziehbarkeit von Ergebnissen. Sind Inhalt, Datengrundlage, Untersuchungseinheit, Methode der Datenerhebung sowie die Datenaufbereitung und -bereinigung unzureichend definiert und dokumentiert, wird dadurch das Verständnis von Daten und Analyseergebnisse erschwert.
- Rechtliche und ethische Aspekte:
 - fehlende Zugangsmöglichkeiten zu privatwirtschaftlichen Internetdaten für Sekundäranalysen
 - datenschutzrechtliche Anforderungen und Verfahren zur Wahrung der Anonymität von untersuchten Personen und größtmöglichen Erhalt von Analysepotentialen für Forschungszwecke
 - ethische Fragen im Rahmen verantwortungsvoller Forschung beim Umgang mit medizinischen Informationen (Biomarker), internetbasierten Verhaltensspuren und prozessproduzierter administrativer Daten
- Dateninfrastrukturen:
 - technische Konzepte und Tools zum Umgang mit den Datenstrukturen neuer Datenformen
 - administrative Konzepte zum Umgang mit neuen Datenformen, Dateiformaten und Datenmengen im Rahmen der Arbeitsabläufe von Dateninfrastrukturen
 - Entwicklung von Strategien zur Datenanonymisierung neuer Datenformen
 - Anpassung inhaltlicher Standards zur Dokumentation neuer sozialwissenschaftlicher Datenstrukturen und ihrer methodischen Grundlagen
 - Modifikation und Anpassung vorhandener Metadatenstandards wie DDI und Sicherung der Interoperabilität mit Metadatenstandards anderer Disziplinen
 - technische Sicherung und Erhaltung neuer digitaler Datenformen, wie Social-Media-Daten
 - Strategien und Kooperationen mit Anbietern bisher nicht zugänglicher Datenquellen
 - Konzepte zum rechtlicheren Zugang und der Bereitstellung neuer Datenformen im Rahmen der aktuellen Datenschutzgesetze

Quelle: Eigene Darstellung

Diese Herausforderungen beim Umgang mit neuen Datenquellen und Datenformen beleuchten zwei Beiträge in diesem Buch. Grundsätzliche Fragen von Social-Media-Daten als Forschungsgrundlage thematisiert Kapitel 11. In diesem Kontext wird die prinzipielle Frage behandelt, ob und unter welchen Bedingungen Daten von entsprechenden Kommunikationsplattformen überhaupt für wissenschaftliche Zwecke zugänglich sind. Damit einhergehend ist die Frage der Qualität entsprechender Daten im Licht unterschiedlicher Nutzungsaspekte von Forschenden zu prüfen. Schließlich sind im Kontext des Forschungsdatenmanagements sowohl die aktuelle Praxis als auch die Herausforderungen bei der Archivierung und der Ermöglichung der Nachnutzung von Social-Media-Daten gemeinsam mit Infrastruktureinrichtungen kritisch zu diskutieren. Fehlende Standards und rechtliche Hürden verhindern es bislang, diesen Datentypus als robuste Forschungsgrundlage zu nutzen.

Stellen Social-Media-Daten noch ein *Moving Target* (Weller 2015) in den Sozialwissenschaften und dem sozialwissenschaftlichen Forschungsdatenmanagement dar, ist die Verknüpfung von Umfrage- und Geodaten ein einfacheres Unterfangen. Beide Datentypen sind nicht neu, ihre Verknüpfung stellt allerdings besondere Herausforderungen an Forschende. Auch hier ist die Kenntnis von bestehenden Schwierigkeiten eine wichtige Grundlage zur erfolgreichen Umsetzung eines entsprechenden Forschungsvorhabens. Kapitel 12 thematisiert

in diesem Zusammenhang drei zentrale Herausforderungen an das Forschungsdatenmanagement. Erstens müssen sich Forschende ein entsprechendes Fachwissen zu Verfahren und Techniken der Georeferenzierung aneignen, um Umfrage- und Geodaten überhaupt verknüpfen zu können. Damit gehen Anforderungen an die technische Ausstattung entsprechender Forschungsprojekte einher, z.B. in Bezug auf leistungsfähige Hardware und geeignete Software. Zudem sind organisatorische und lizenzrechtliche Voraussetzungen zu erfüllen, um Zugangs- und Nutzungsrechte zu den erforderlichen Daten zu erlangen. Ein zweiter Themenkreis betrifft Herausforderungen des Datenschutzes und Risiken der Re-Identifikation bei der Produktion georeferenzierter Daten im Projektverlauf. Schließlich müssen drittens Geo- und Umfragedaten nachvollziehbar und transparent dokumentiert werden, um die Replizierbarkeit publizierter Ergebnisse ebenso wie die Nachnutzung der georeferenzierten Daten zu ermöglichen.

Anhand dieser beiden Themenbereiche, d.h. die Nutzung von Social-Media-Daten bzw. die Verknüpfung von Umfrage- und Geodaten, werden exemplarisch einige neue Herausforderungen im Forschungsdatenmanagement aufgezeigt. Dabei sind die dort angesprochenen Probleme und die damit einhergehenden Entwicklungen im Forschungsdatenmanagement in keiner Weise erschöpfend. Die Diskussion zeigt jedoch eindringlich, dass es sich auch beim Forschungsdatenmanagement um keinen statischen Bereich der Sozialforschung handelt. Vielmehr ist es einem stetigen Wandel unterworfen. Das Forschungsdatenmanagement muss sich an neue Gegebenheiten, etwa im Bereich des Datenschutzes oder der Weiterentwicklung von Dokumentationsstandards, genauso anpassen wie an die Anforderungen neuer Datentypen.

2.4 Zusammenfassung

Ein professionelles Management von Forschungsdaten gewinnt in vielen Disziplinen eine immer größere Bedeutung. Dies geht mit der Etablierung des Forschungsdatenmanagements als Teil guter wissenschaftlicher Praxis einher, das die Replikation der Forschungsergebnisse ebenso wie der genutzten Daten erlaubt und es Dritten ermöglicht, mit den erstellten Forschungsdaten in neuen Kontexten weiterzuarbeiten. Daneben trägt die zunehmende Anerkennung von (digitalen) Forschungsdaten als wertvolle Wissensquelle und eigenständiges Forschungsprodukt zu dieser gesteigerten Bedeutung weiter bei. In den Sozialwissenschaften zeigen sich seit der Jahrtausendwende vielfältige Bemühungen, das Management von Forschungsdaten systematisch und nachhaltig zu organisieren.

Die Entwicklung eines professionellen Umgangs mit (sozialwissenschaftlichen) Forschungsdaten erfordert es, die Wertschöpfungskette entlang des gesamten Lebenszyklus dieser Daten zu berücksichtigen. Nur so können Forschende sicherstellen, dass die im Projekt erstellten Daten zum Erreichen der Projektziele genutzt werden können, dass die produzierten Forschungsergebnisse replizierbar bleiben und die Daten durch Dritte für die unterschiedlichsten Zwecke nachgenutzt werden können. Das Forschungsdatenmanagement ist insgesamt mit komplexen Aufgaben und Zielen verbunden und stellt damit die Akteure des Wissenschaftssystems im Rahmen ihrer jeweiligen Rolle vor unterschiedliche Herausforderungen. Diese Herausforderungen können entweder disziplinübergreifend bestehen oder sich auf Besonderheiten einzelner Disziplinen beziehen. So ist das Forschungsdatenmanagement in den Sozialwissenschaften einerseits in die übergreifenden Empfehlungen von Wissenschaftsorganisationen und in die Leitlinien der Forschungsförderung eingebunden. Andererseits leisten spezialisierte Dateninfrastrukturen ihren Beitrag durch fachspezifische Dienstleistungen,

die Forschende etwa bei der Erschließung, Archivierung, Nachnutzung und Zitation von sozialwissenschaftlichen Forschungsdaten unterstützen. Dazu stehen praxiserprobte Grundsätze und etablierte Lösungsansätze für Forschende zur Verfügung. Gleichzeitig stellen neu erschlossene Datenquellen, etwa aus fachfremden Disziplinen wie Medizin oder Geographie, sowie neue Datentypen, wie z.B. Social-Media-Daten, und deren Verknüpfung mit den ‚klassischen‘ sozialwissenschaftlichen Umfragedaten ein systematisches und nachhaltiges Management der Daten vor neue Herausforderungen.

Betrachtet man die Anforderungen des Forschungsdatenmanagements aus Sicht der Forschenden, so gehen die umzusetzenden Maßnahmen und Aktivitäten über die eigentliche (sozial-)wissenschaftliche Forschungsarbeit im engeren Sinne hinaus. Während Forschende bei der Erstellung von Daten im Rahmen von Forschungsprojekten zumeist die Analyse der Daten mit Bezug auf eine oder mehrere konkrete Forschungsfragen verbinden, erfordert beispielsweise die Dokumentation von Daten zur Archivierung und Nachnutzung einen teils erheblichen zusätzlichen Aufwand an Zeit und Ressourcen. Diesen spezifischen Aufwand zu berücksichtigen, ist ein wichtiger Bestandteil der Planung eines Forschungsprojekts und damit einhergehend der Vorbereitung des Forschungsdatenmanagements.

Der sogenannte *Datenmanagementplan* (vgl. CESSDA 2017; Jensen 2011) ist hierbei ein wichtiges Werkzeug. Er beinhaltet die systematische Dokumentation des Forschungsdatenmanagements und beschreibt, welche Maßnahmen von wem wie und warum gesetzt wurden bzw. umgesetzt werden müssen und wird so zu einem Instrument, welches das Forschungsdatenmanagement auf Projektebene planvoll und nachvollziehbar beschreibt. Forschende sollten daher bereits im Rahmen der Projektplanung einen projektspezifischen Datenmanagementplan erstellen und darin alle geplanten Strategien, Maßnahmen und Aktivitäten im Umgang mit den Forschungsdaten festlegen und im Laufe des Forschungsprojekts systematisch als Leitfaden anwenden.

Zu bedenken ist in diesem Zusammenhang, dass Forschungsprojekte kein statisches Unterfangen sind. Forschungsvorhaben und Forschungsziele können sich im Projektverlauf verändern. Damit gehen oftmals Veränderungen in den (zu erstellenden) Forschungsdaten und somit konsequenterweise im projektspezifischen Forschungsdatenmanagement einher. Dementsprechend müssen Maßnahmen und Aktivitäten angepasst und im Datenmanagementplan entsprechend dokumentiert werden. Mit anderen Worten: Analog zu Forschungsvorhaben ist auch das Forschungsdatenmanagement ein dynamischer Prozess und seine Dokumentation im Datenmanagementplan stetigen Anpassungen unterworfen. Als Instrument der Qualitätssicherung können mit seiner Hilfe die Vorgehensweisen und Arbeitsabläufe im Projekt stetig spezifiziert, überprüft und angepasst werden.

Literaturverzeichnis

- Allianz der deutschen Wissenschaftsorganisationen (2010): Grundsätze zum Umgang mit Forschungsdaten. <https://doi.org/10.2312/ALLIANZOA.019>.
- Ball, Alexander (2012): Review of Data Management Lifecycle Models (version 1.0). REDm-MED Project Document. University of Bath. <http://opus.bath.ac.uk/28587/1/redm1rep120110ab10.pdf> [Zugriff: 12.07.2018].
- Baur, Nina/Blasius, Jörg (2014): Methoden der empirischen Sozialforschung. Ein Überblick. In: Baur, Nina/Blasius, Jörg (Hrsg.): Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer, S. 41-66. <https://doi.org/10.1007/978-3-531-18939-0>.
- Berliner Erklärung (2003): Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen. Deutsche Version. https://openaccess.mpg.de/68053/Berliner_Erklaerung_dt_Version_07-2006.pdf [Zugriff: 12.07.2018].

- BMBF – Bundesministerium für Bildung und Forschung (2017): Richtlinie zur Förderung von Forschung zu „Digitalisierung im Bildungsbereich – Grundsatzfragen und Gelingensbedingungen“. <https://www.bmbf.de/foederungen/bekanntmachung-1420.html>.
- Carlson, Jake (2014): The Use of Life Cycle Models in Developing and Supporting Data Services. In: Ray, Joyce M. (Hrsg.): *Research Data Management. Practical Strategies for Information Professionals*, S. 63-86. West Lafayette: Purdue University Press. <https://www.jstor.org/stable/j.ctt6wq34t> [Zugriff: 12.07.2018].
- CCSG – Cross-Cultural Survey Guidelines (2018). <http://www.ccsq.isr.umich.edu/index.php/> [Zugriff: 12.07.2018].
- CESSDA (2017): Expert Tour Guide on Data Management. <https://www.cessda.eu/dmguide> [Zugriff: 12.07.2018].
- Corti, Louise/Van den Eynden, Veerle/Bishop, Libby/Woollard, Matthew (2014): *Managing and Sharing Research Data. A Guide to Good Practice*. London: Sage Publications.
- Cox, Andrew Martin/ Tam, Winnie Wan Ting (2018): A Critical Analysis of Lifecycle Models of the Research Process and Research Data Management. In: *Aslib Journal of Information Management* 70, 2, S. 142-157. <https://doi.org/10.1108/AJIM-11-2017-0251>.
- DDI Data Lifecycle(2018). <http://www.ddialliance.org/training/why-use-ddi> [Zugriff: 12.07.2018].
- Diekmann, Andreas (2007): *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. Reinbek: Rowohlt.
- DINI – Deutsche Initiative für Netzwerkinformation e.V. (2009): *Positionspapier Forschungsdaten*. <https://doi.org/10.18452/1489> [Zugriff: 12.07.2018].
- DFG – Deutsche Forschungsgemeinschaft (2018): *Leitfaden für die Antragstellung. Projektanträge [03/18]*. http://www.dfg.de/formulare/54_01/54_01_de.pdf [Zugriff: 12.07.2018].
- DFG (2015): *Leitlinien zum Umgang mit Forschungsdaten*. http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf [Zugriff: 12.07.2018].
- DFG (2009): *Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme. Unterausschuss für Informationsmanagement. Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten*. http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf [Zugriff: 12.07.2018].
- DFG (2013) [1988]: *Empfehlung 7*. In: *DFG Denkschrift: Sicherung guter wissenschaftlicher Praxis*. Weinheim: WILEY-VCH, S. 21. http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf [Zugriff: 12.07.2018].
- Ebel, Thomas/Trixa, Jessica (2015): *Hinweise zur Aufbereitung quantitativer Daten*. *GESIS Papers 2015/09*. <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-432235> [Zugriff: 12.07.2018].
- Felderer, Barbara/Birg, Alexandra/Kreuter, Frauke (2014): *Paradaten*. In: Baur, Nina/Blasius, Jörg (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer, S. 357-365. <https://doi.org/10.1007/978-3-531-18939-0>.
- Horizon 2020 Programme (o.J.): *Open Access & Data Management*. https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm [Zugriff: 12.07.2018].
- Huschka, Denis/Oellers, Claudia/Ott, Notburga/Wagner, Gert G. (2011): *Datenmanagement und Data Sharing. Erfahrungen in den Sozial- und Wirtschaftswissenschaften*. In: Büttner, Stephan/Hobohm, Hans-Christoph/Müller, Lars (Hrsg.): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen, S. 35-48. <https://opus4.kobv.de/opus4-fhpotsdam/frontdoor/index/index/docId/194> [Zugriff: 12.07.2018].
- ICPSR – Inter-university Consortium for Political and Social Research (2012): *Guide to Social Science Data Preparation and Archiving. Best Practice Throughout the Data Life Cycle (5th ed.)*. Ann Arbor, MI. <https://www.icpsr.umich.edu/files/deposit/dataprep.pdf> [Zugriff: 12.07.2018].
- Jensen, Uwe/Ekman, Stefan/Hjelm, Claus-Göran/Irebäck, Hans/Schweers, Stefan (2015): *DELIVERABLE D7.6 Metadata Standards and Practices in Related Disciplines and Standards for Linking Different Sources. Data without Boundaries (DwB) WORK PACKAGE 7 – Standards Development*. www.dwbproject.org/export/sites/default/about/public_deliverables/dwb_d7-6_metadata-standards-practices-related-disciplines_report-final.pdf [Zugriff: 12.07.2018].
- Jensen, Uwe (2012): *Leitlinien zum Management von Forschungsdaten. Sozialwissenschaftliche Umfragedaten*. *GESIS-Technical Reports 2012/07*. <https://www.ssoar.info/ssoar/handle/document/32065> [Zugriff: 12.07.2018].
- Jensen, Uwe (2011): *Datenmanagementpläne*. In: Büttner, Stephan/Hobohm, Hans-Christoph/Müller, Lars (Hrsg.): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen, S. 71-82. <http://opus.kobv.de/fhpotsdam/volltexte/2011/230/> [Zugriff: 12.07.2018].

- Jensen, Uwe/Katsanidou, Alexia/Zenk-Möltgen, Wolfgang (2011): Metadaten und Standards. In: Büttner, Stephan/Hobohm, Hans-Christoph/Müller, Lars (Hrsg.): Handbuch Forschungsdatenmanagement. Bad Honnef: Bock + Herchen, S.83-100.
https://opus4.kobv.de/opus4-fhpotsdam/files/198/2.4_Metadaten_und_Standards.pdf [Zugriff: 12.07.2018].
- King, Garry (2011): Ensuring the Data Rich Future of the Social Sciences. In: *Science* 331, 6018, S. 719-721.
<http://dx.doi.org/10.1126/science.1197872> [Zugriff: 12.07.2018].
- KVI – Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (Hrsg.) (2001): Wege zu einer besseren informationellen Infrastruktur. Baden-Baden: Nomos.
- Ludwig, Jens (2012): Zusammenfassung und Interpretation. In: Neuroth, Heike/Strathmann, Stefan/Oßwald, Achim/Scheffel, Regine/Klump, Jens/Ludwig, Jens (Hrsg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg: Werner Hülsbusch, S. 295-310.
http://www.nestor.sub.uni-goettingen.de/bestandsaufnahme/nestor_lza_forschungsdaten_bestandsaufnahme.pdf [Zugriff: 12.07.2018].
- Ludwig, Jens/Enke, Harry (2013): Einleitung. In: Ludwig, Jens/Enke, Harry (Hrsg.): Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt. Glückstadt: Werner Hülsbusch, S. 13f.
https://www.univerlag.uni-goettingen.de/bitstream/handle/3/isbn-978-3-86488-032-2/leitfaden_DGRID.pdf [Zugriff: 12.07.2018].
- Lück, Detlev/Landrock, Uta (2014): Datenaufbereitung und Datenbereinigung in der quantitativen Sozialforschung. In: Baur, Nina/Blasius, Jörg (Hrsg.): Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer, S. 233-244.
<https://doi.org/10.1007/978-3-531-18939-0>.
- Mochmann, Ekkehard (2014): Quantitative Daten für die Sekundäranalyse. In: Baur, Nina/Blasius, Jörg (Hrsg.): Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer, S. 233-244.
<https://doi.org/10.1007/978-3-531-18939-0>.
- Netscher, Sebastian/Eder Christina (Hrsg.) (2018): Data processing and Documentation: Generating High Quality Research Data in Quantitative Social Science research. GESIS Papers, 2018/22.
<https://nbn-resolving.org/urn:nbn:de:0168-ssoar-59492-3> [Zugriff: 26.10.2018].
- OECD – Organisation for Economic Co-operation and Development (2013): New Data for Understanding the Human Condition. OECD Global Science Forum Report on Data and Research Infrastructure for the Social Sciences. <http://www.oecd.org/std/microdata-access-final-report-OECD-2014.pdf> [Zugriff: 12.07.2018].
- OECD (2007): OECD Principles and Guidelines for Access to Research Data from Public Funding.
<http://www.oecd.org/science/sci-tech/38500813.pdf> [Zugriff: 12.07.2018].
- Piwowar, Heather A./Vision, Todd J. (2013): Data reuse and the open data citation advantage. *PeerJ* 1:e175 <https://doi.org/10.7717/peerj.175> [Zugriff: 12.07.2018].
- Piwowar, Heather A./Day, Roger S./Fridsma, Douglas B. (2007): Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE* 2(3): e308. <https://doi.org/10.1371/journal.pone.0000308> [Zugriff: 12.07.2018].
- Quandt, Markus/Mauer, Reiner (2012): Sozialwissenschaften. In: Neuroth, Heike/Strathmann, Stefan/Oßwald, Achim/Scheffel, Regine/Klump, Jens/Ludwig, Jens (Hrsg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg: Werner Hülsbusch, S. 61-81. http://www.nestor.sub.uni-goettingen.de/bestandsaufnahme/nestor_lza_forschungsdaten_bestandsaufnahme.pdf [Zugriff: 12.07.2018].
- Rat für Informationsinfrastrukturen – RfII (2016): Leistung aus Vielfalt. <http://nbn-resolving.de/urn:nbn:de:101:1-201606229098> [Zugriff: 12.07.2018].
- Rat für Sozial- und Wirtschaftsdaten – RatSWD (2017): Arbeitsprogramm der 6. Berufenungsperiode 2017–2020. https://www.ratswd.de/dl/downloads/Arbeitsprogramm_RatSWD_2017-20.pdf [Zugriff: 12.07.2018].
- RatSWD (2012): Georeferenzierung von Daten. Situation und Zukunft der Geodatenlandschaft in Deutschland. *Sci-vero*. https://www.ratswd.de/dl/downloads/RatSWD_Geodatenbericht.pdf [Zugriff: 12.07.2018].
- Schnell, Rainer/Hill, Paul B./Esser, Elke (2013): Methoden der empirischen Sozialforschung. München: Oldenbourg.
- Weller, Katrin (2015): Accepting the Challenges of Social Media Research. In: *Online Information Review* 39, 3, S. 281-289. <https://doi.org/10.1108/OIR-03-2015-0069>.
- WR – Wissenschaftsrat (2011): Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften. <https://www.wissenschaftsrat.de/download/archiv/10465-11.pdf> [Zugriff: 12.07.2018].
- ZBW – Leibniz-Informationzentrum Wirtschaft/GESIS- Leibniz-Institut für Sozialwissenschaften/RatSWD - Rat für Sozial- und Wirtschaftsdaten (2015): Auffinden, Zitieren, Dokumentieren. Forschungsdaten in den Sozial- und Wirtschaftswissenschaften. <https://dx.doi.org/10.4232/10.fisuzida2015.2>.

Linkverzeichnis

- ALLBUS – Allgemeine Bevölkerungsumfrage der Sozialwissenschaften: <https://www.gesis.org/allbus/> [Zugriff: 12.07.2018].
- CESSDA – Consortium of European Social Science Data Archives: <https://www.cesda.eu/> [Zugriff: 12.07.2018].
- CSES – Comparative Study of Electoral System (o.J.): <http://www.cses.org/> [Zugriff: 12.07.2018].
- DDI – Data Documentation Initiative / DDI Alliance: <https://www.ddialliance.org/> [Zugriff: 12.07.2018].
- Destatis – Statistisches Bundesamt (2018): <https://www.destatis.de/DE/Startseite.html> [Zugriff: 12.07.2018].
- DOI – Digital Object Identifier System: <https://www.doi.org/> [Zugriff: 12.07.2018].
- European Social Survey: <http://www.europeansocialsurvey.org/> [Zugriff: 12.07.2018].
- Eurostat – Statistisches Amt der Europäischen Union (o.J.): <http://ec.europa.eu/eurostat/de/home> [Zugriff: 12.07.2018].
- GESIS – Leibniz-Institut für Sozialwissenschaften: <https://www.gesis.org/> [Zugriff: 12.07.2018].
- GESIS Datenservices (o.J.): <https://www.gesis.org/datenservices/home/> [Zugriff: 12.07.2018].
- GESIS datorium: <https://datorium.gesis.org/> [Zugriff: 12.07.2018].
- Gottfried Wilhelm Leibniz e. V / Leibniz-Gemeinschaft: <https://www.leibniz-gemeinschaft.de/start/> [Zugriff: 12.07.2018].
- Handle System: <https://www.dona.net/handle-system> [Zugriff: 12.07.2018].
- URN – Uniform Resource Name: http://www.dnb.de/DE/Netzpublikationen/URNService/urnservice_node.html [Zugriff: 12.07.2018].
- OECD – Organisation for Economic Co-operation and Development: <https://www.oecd.org/> [Zugriff: 12.07.2018].
- SOEP – Sozio-oekonomische Panel: <https://www.diw.de/soep> [Zugriff: 12.07.2018].
- ZBW – Leibniz-Informationszentrum Wirtschaft. <http://www.zbw.eu/de/> [Zugriff: 12.07.2018].
- ZBW Journal Data Archive. <http://www.journaldata.zbw.eu/> [Zugriff: 12.07.2018].

