

Forschungsdatenmanagement systematisch planen und umsetzen

Netscher, Sebastian; Jensen, Uwe

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Netscher, S., & Jensen, U. (2019). Forschungsdatenmanagement systematisch planen und umsetzen. In U. Jensen, S. Netscher, & K. Weller (Hrsg.), *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten: Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten* (S. 37-55). Opladen: Verlag Barbara Budrich.
<https://doi.org/10.3224/84742233.04>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-SA Lizenz (Namensnennung-Weitergabe unter gleichen Bedingungen) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by-sa/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-SA Licence (Attribution-ShareAlike). For more information see:
<https://creativecommons.org/licenses/by-sa/4.0>

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

3. Forschungsdatenmanagement systematisch planen und umsetzen

Sebastian Netscher und Uwe Jensen

Ein systematisches Forschungsdatenmanagement ist grundlegend für die erfolgreiche Umsetzung empirischer Forschungsprojekte, d.h. die planvolle Umsetzung eines Forschungsvorhabens, und fixer Bestandteil guten wissenschaftlichen Arbeitens. Es bezieht u.a. Fragen der Datengenerierung, Dokumentation, Auswertung und längerfristigen Sicherung von Forschungsdaten in die Projektplanung ein. Das Forschungsdatenmanagement dient so nicht nur der erfolgreichen Umsetzung des eigentlichen Forschungsvorhabens, sondern unterstützt auch weiterführende Ziele und Anforderungen im Umgang mit den Daten über das originäre Forschungsprojekt hinaus. Dazu zählen die Aufbewahrung, Sicherung und Nachnutzung der Daten ebenso wie die Replizierbarkeit der publizierten Ergebnisse.

Gegenstand dieses Kapitels ist die Beschreibung von Grundzügen des Forschungsdatenmanagements. Dabei steht vor allem der Umgang mit quantitativen Daten der empirischen Sozialforschung im Vordergrund. Ausgehend von drei zentralen Zielen wird die planvolle Implementierung des Forschungsdatenmanagements im Projekt und darüber hinaus vertieft. Der Abschnitt 3.1 fokussiert als erstes Ziel ein planvolles projektinternes Forschungsdatenmanagement. Hierbei wird der Umgang mit den Forschungsdaten im Projektverlauf mit dem Ziel festgelegt, deren Verständlichkeit, Interpretierbarkeit und Nachvollziehbarkeit sicherzustellen. Dazu werden im Folgenden wesentliche disziplinspezifische Strategien, Prozesse und Maßnahmen im Rahmen des Forschungsdatenmanagements kurz erörtert.

Daran anschließend werden in Abschnitt 3.2 zwei weiterführende Ziele des Forschungsdatenmanagements diskutiert. Hierzu unterscheiden wir grob zwischen dem Erhalt der Daten zu Replikationszwecken und der Bereitstellung der Daten zur Nachnutzung durch Dritte. Der Erhalt der Daten zu Replikationszwecken dient vor allem der Verifizierbarkeit von Forschungsergebnissen ebenso wie der Reproduzierbarkeit der dabei verwendeten Daten. Im Sinne guter wissenschaftlicher Praxis unterstützt der Datenerhalt damit die Transparenz und Überprüfbarkeit des originären Forschungsvorhabens. Die Bereitstellung der Daten zur Nachnutzung (*Data Sharing*) geht darüber hinaus und ermöglicht es Dritten, diese Daten in neuen Forschungskontexten und für neue Forschungsfragen nachzunutzen. Die Datenbereitstellung ist damit nicht nur Teil guter wissenschaftlicher Praxis, sondern fördert auch den Erkenntnisfortschritt in der Forschung. *Data Sharing* ermöglicht den effizienten Einsatz von Forschungsressourcen in der Forschungsförderung und unterstützt die Ziele eines offenen Zugangs zu Forschungsdaten (*Open Data*). Schließlich steigert die Datenbereitstellung die Reputation der Forschenden, indem ihre Daten an Sichtbarkeit gewinnen sowie dauerhaft auffindbar und zitierbar sind (ZBW/GESIS/RatSWD 2015; Allianz der deutschen Wissenschaftsorganisationen 2010).

3.1 Die Planung des projektinternen Forschungsdatenmanagements

Um Forschungsdatenmanagement im Rahmen eines empirischen Forschungsprojekts gezielt betreiben zu können, gilt es zu klären, was, wie, wann, von wem und mit welchem Aufwand

im Umgang mit den Forschungsdaten gemacht werden soll bzw. bereits gemacht wurde. Entsprechende Aktivitäten müssen geplant, entwickelt, untereinander abgestimmt und in den gesamten Projektablauf integriert werden. Sie machen den Status und Entstehungskontext der Daten in jeder Projektphase transparent und nachvollziehbar und dienen so der Qualität, Verständlichkeit und Interpretierbarkeit der Daten.

Zentrales Werkzeug eines planvollen Forschungsdatenmanagements ist der Datenmanagementplan (vgl. Jensen 2011). In ihm werden alle notwendigen Aktivitäten definiert und systematisch dokumentiert. Der Datenmanagementplan beschreibt den (geplanten) Umgang mit den Daten im Projektverlauf ebenso wie über das Projektende hinaus. Dabei verfolgt er einen doppelten Zweck: Zum einen fungiert der Datenmanagementplan als Leitfaden, der die notwendigen Aktivitäten des Forschungsdatenmanagements in konkrete Arbeitsabläufe, Aufgaben, Regeln und Verantwortlichkeiten strukturiert. Zum anderen ist er ein Instrument der Qualitätssicherung, das die Transparenz im Projekt erhöht und die Grundlagen für den längerfristigen Erhalt der Daten über das Projektende hinaus schafft.

Bei der Umsetzung des Forschungsdatenmanagements sollten Forschende immer ihre eigentlichen Ziele im Umgang mit den Daten im Auge behalten. So muss das projektinterne Forschungsdatenmanagement den Umgang mit den Daten im Forschungsprojekt regeln. Separat zu betrachten sind Pläne zum Umgang mit den Daten nach Projektende, wie etwa der längerfristige Datenerhalt zu Replikationszwecken oder die Bereitstellung der Daten zur Nachnutzung, wie in Abschnitt 3.2 näher erläutert. Im vorliegenden Abschnitt konzentrieren wir uns zunächst auf die Planung des projektinternen Forschungsdatenmanagements und die dabei zu berücksichtigenden Aktivitäten. Um die Qualität und Nutzbarkeit der Daten im Forschungsprojekt sicherzustellen, müssen diese adäquat aufbereitet und dokumentiert sein, rechtliche Aspekte, z.B. in Bezug auf den Datenschutz oder das Urheberrecht, berücksichtigt und die Daten systematisch organisiert werden. Aufbauend auf dem projektinternen Forschungsdatenmanagement können dann weiterführende Ziele im Umgang mit den Daten definiert und implementiert werden.

3.1.1 Datenaufbereitung und Datendokumentation

Je nach zugrunde liegender Forschungsfrage müssen Forschende zunächst definieren, welche Daten zur Realisierung des Forschungsvorhabens notwendig sind. Das bedeutet, dass die Forschenden zuallererst klären müssen, welche Informationen, über wen oder was, in welcher Form und wie erhoben werden sollen. Derartige Maßnahmen zur Erstellung des Frageprogramms, zur Definition der Grundgesamtheit, zur Stichprobenziehung, zur Methode der Datenerhebung etc. sind spezifisch für das jeweilige Forschungsprojekt festzulegen. In der Regel sind sie Teil der eigentlichen Projektplanung und weniger des projektinternen Forschungsdatenmanagements. Für die vorausschauende Planung des Forschungsdatenmanagements ist die detaillierte Kenntnis u.a. des Frageprogramms oder der Untersuchungspopulation aber unerlässlich. Sie ist die Grundlage zur systematischen Vorbereitung aller Aktivitäten des Forschungsdatenmanagements.

Zur Gewährleistung ihrer Qualität und Nutzbarkeit im Projektverlauf müssen die einmal erhobenen Daten aufbereitet, validiert und dokumentiert werden. Um einen reibungslosen Umgang mit den Daten im Projekt sicherzustellen, gilt es daher bereits in der Planung der Datenerhebung Konzepte zur Aufbereitung und Dokumentation zu entwickeln. Ein Konzept zur Aufbereitung sollte von den erhobenen Rohdaten ausgehen. Je nach Art und Umfang müssen diese zunächst strukturiert und in ein für die Datenanalyse geeignetes Format gebracht werden. Dies betrifft sowohl den Aufbau des Datensatzes als auch die dazu verwendete Software und die genutzten Dateiformate. Bei der Planung des Datensatzaufbaus gilt es

festzulegen, wie die Variablen strukturiert werden. Dies betrifft neben deren thematischen Gliederung, etwa nach administrativen und inhaltlichen Variablen, auch die Sicherstellung ihres Bezugs zum originären Frageprogramm, inklusive darin enthaltener Filter, Verzweigungen oder Variablenabfolgen. In Bezug auf die zu verwendenden Dateiformate empfiehlt sich generell die Verwendung quelloffener Formate, wie z.B. das grundlegende *ASCII*-Format bzw. die Nutzung nicht-proprietärer Software, wie z.B. das Statistikprogramm *R*. In jedem Fall muss gerade in Forschungsk Kooperationen sichergestellt werden, dass alle Projektbeteiligten jederzeit Zugang zu den Daten haben und mit diesen arbeiten können.

Nach der Definition der Datensatzstruktur sind alle darin enthaltenen Variablen eindeutig zu definieren. Dazu gehört, dass ihr Bezug zum originären Frageprogramm sichergestellt, die Variable eindeutig benannt und ihre Bedeutung durch ein aussagekräftiges Label beschrieben (*labeln*) wird. Analog müssen alle in den Variablen enthaltenen Werte (*Kodes*) gemäß den Antwortvorgaben des Frageprogramms definiert und ihre Bedeutung beschrieben werden. Gleichfalls müssen offene Nennungen in einzelnen Fragen konsistent strukturiert und kodiert werden. Dies gilt auch für den Umgang mit fehlenden Werten. Hierbei gilt es zunächst zu klären, ob die unterschiedlichen Gründe für fehlende Angaben, wie etwa *weiß nicht* oder *sage ich nicht*, für das Forschungsvorhaben relevant sind. Dementsprechend müssen diese detailliert erhoben und kodiert werden oder können pauschal erfasst werden. Schließlich müssen abgeleitete und konstruierte Variablen, wie Gewichte, Skalen und Indizes, definiert, gelabelt und in die bestehende Datensatzstruktur integriert werden (Jensen 2012: 23f.).

Daneben müssen im Rahmen des Aufbereitungskonzepts Validierungsmaßnahmen entwickelt und implementiert werden. Die frühzeitige Planung entsprechender Kontrollmaßnahmen dient sowohl ihrer konsistenten Umsetzung, und damit der Datenqualität, als auch der Datenerhebung, etwa durch die erneute Kontrolle der geplanten Filterführung im Frageprogramm. Die zu entwickelnden Validierungsmaßnahmen umfassen dabei u.a. die Überprüfung der Repräsentativität der Daten ebenso wie Maßnahmen zur Kontrolle aller Variablen bzw. Variablengruppen, z.B. in Bezug auf die Filterführung oder auf etwaige Verzerrungen und Ausreißer, Häufungen fehlender Antworten, systematisches oder inkonsistentes Antwortverhalten. Die Planung und Umsetzung entsprechender Prozesse und Arbeitsschritte wird in Kapitel 6 Schritt für Schritt behandelt.

Parallel zur Planung der Datenaufbereitung gilt es die Daten zu dokumentieren. Ein entsprechendes Konzept zur Datendokumentation muss gewährleisten, dass alle relevanten Schritte der Datenerhebung und Aufbereitung ebenso wie die Forschungsdaten selbst ausreichend beschrieben sind. Die Dokumentation dient der Qualitätssicherung und ermöglicht es, den Zustand der Daten in unterschiedlichen Projektphasen nachzuvollziehen und reversibel zu kontrollieren. Sie stellt somit die Transparenz der Daten, deren Verständlichkeit und Interpretierbarkeit im Rahmen des Forschungsprojekts sicher.

Das Dokumentationskonzept muss Antworten auf zwei zentrale Fragen liefern. Erstens: Was muss dokumentiert werden? D.h., welche Informationen zu den Daten und ihrem Entstehungsprozess müssen erhalten bleiben, um deren Transparenz und Verständlichkeit zu gewährleisten? Zweitens: Wie soll dokumentiert werden? D.h., welche Standards und Best-Practice-Empfehlungen lassen sich hierfür nutzen (zur Einführung vgl. z.B. Corti et al. 2014; Jensen/Katsanidou/Zenk-Möltgen 2011)? In Bezug auf die erste Frage nach dem Inhalt der Dokumentation empfiehlt sich die Unterscheidung von Informationen auf der Studien- und der Variablenebene. Auf Studienebene müssen der übergreifende Kontext der Datenerhebung und deren methodische Grundlagen beschrieben werden, etwa in Form eines Methodenberichts (Watteler 2010). Dieser sollte auch Informationen über rechtliche Aspekte, wie zum Datenschutz oder zum Urheberrecht (s.u.), beinhalten und das Aufbereitungskonzept beschreiben. Auf Variablenebene sind Fragen und Antwortvorgaben des originären Frageprogramms zu dokumentieren.

Neben der Frage nach den Inhalten der Dokumentation muss das Konzept auch die Struktur der dargebotenen Informationen thematisieren. Es gilt zu klären, welche Metadaten und Metadatenstandards zur Datendokumentation genutzt werden und auf welches kontrollierte Vokabular dabei zurückgegriffen wird. Vereinfacht ausgedrückt sind Metadaten nichts anderes als Informationen über die Daten, die diese beschreiben und so deren Transparenz, Verständlichkeit und Interpretierbarkeit sicherstellen. Wallace (2001: 262) unterscheidet dabei zwischen „structured or semi-structured information which enables the creation, management, and use of records through time and within and across domains“.

Schaukasten 3.1: Datenaufbereitung und Datendokumentation

Datensatzstruktur:

- Ist der Datensatz sinnvoll strukturiert und spiegelt den Aufbau des Frageprogramms wider?
- Sind alle Filter, Verzweigungen, Variablen analog zum Frageprogramm spezifiziert?

Variablen, Werte und Labels:

- Sind alle Variablen strukturiert sowie einheitlich definiert, benannt und gelabelt?
- Sind alle Werte der Variablen gelabelt und den Antwortvorgaben des Frageprogramms zuordenbar?
- Sind alle offenen Antwortmöglichkeiten und fehlende Werte konsistent kodiert und gelabelt?
- Sind abgeleitete Variablen inhaltlich definiert, gelabelt und in die Datensatzstruktur integriert?

Qualitätssicherung:

- Welche Validierungs- und Qualitätssicherungsmaßnahmen sind geplant?

Dokumentation und Metadaten:

- In welcher Form werden die Daten dokumentiert?
- Wie werden Veränderungen der Ursprungsdaten, z.B. durch die Datenaufbereitung, dokumentiert?
- Wie werden abgeleitete Variablen, Skalen, Indizes, Gewichte etc. dokumentiert?
- Welche Metadatenstandards werden genutzt?
- Wird ein kontrolliertes Vokabular zur Datendokumentation genutzt?
- Werden Länder, Berufe, Bildung etc. mittels Standardklassifikationen dokumentiert?

Quelle: Eigene Darstellung

Semi-strukturierte Metadaten beziehen sich auf alle Formen zusammenhängender textbasierter Dokumente, wie Methodenberichte, originäre Fragebögen oder Codebücher. Derartige Dokumente enthalten alle notwendigen Informationen, um die Transparenz und Verständlichkeit der Daten zu gewährleisten, die in Aufbau und Struktur keinen (internationalen) Standards oder Konventionen folgen. Strukturierte bzw. standardisierte Metadaten dienen u.a. der methodischen Beschreibung einzelner Variablen bzw. Messkonzepte, z.B. die International Standard Classification of Education (ISCED 2011), oder der adäquaten Zitation von Daten, wie am Beispiel von DataCite (2016) in Kapitel 10 erörtert. Die Standards der Data Documentation Initiative (DDI 2017) ermöglichen es darüber hinaus, sozialwissenschaftliche Daten systematisch auf Studien- und Variablenebene zu dokumentieren (vgl. Kapitel 9). Die Verwendung standardisierter Metadaten zur Datendokumentation ist zwar etwas zeitaufwendiger als die Beschreibung der Daten mit semi-strukturierten Metadaten. Dieser höhere Aufwand zahlt sich jedoch bereits in der Datenanalyse aus. Eine vollständige Dokumentation anhand standardisierter Metadaten steigert die Transparenz der Daten, erhöht ihre Verständlichkeit und erleichtert die Interpretation von Forschungsergebnissen über einen längerfristigen Zeitraum hinweg. Schaukasten 3.1 fasst relevante Leitfragen zur Planung der Aufbereitung und Dokumentation von Forschungsdaten zusammen.

3.1.2 *Datenschutzkonzept*

Zumeist beinhalten sozialwissenschaftliche Daten Informationen zu natürlichen Personen, menschlichem Verhalten oder Einstellungen. Dementsprechend sind der „Schutz der Untersuchungsperson sowie der Schutz vor einer missbräuchlichen Verwendung ihrer personenbezogenen Informationen [...] Kernaspekte des verantwortlichen Forschens im sozialwissenschaftlichen Umfeld“ (Jensen 2012: 13). Mit anderen Worten liegt es in der Verantwortung der Forschenden, geeignete Maßnahmen zu planen und wirksam umzusetzen, um Untersuchungspersonen vor möglichen negativen Konsequenzen ihrer Teilnahme am Frageprogramm zu schützen. Dabei handelt es sich nicht nur um eine forschungsethische Verantwortung. Vielmehr schreiben die Regelungen der EU-Datenschutz-Grundverordnung (DSGVO) entsprechende Maßnahmen im Umgang mit sogenannten personenbezogenen Informationen vor. Derartige Informationen umfassen zum einen alle Angaben, die zur direkten oder indirekten Identifikation von natürlichen Personen geeignet sind, wie z.B. deren Klarnamen. Zum anderen geht es aber auch um sogenannte besondere Kategorien personenbezogener Daten, d.h. Angaben etwa zur ethnischen Herkunft, zu religiösen Überzeugungen oder zu sexuellen Orientierungen (Art. 9 DSGVO).

Sollen in einem Forschungsprojekt personenbezogene Informationen gemäß Art. 4 DSGVO erhoben, gespeichert oder verarbeitet werden, müssen Forschende bereits zu Beginn des Forschungsprojekts ein umfassendes Datenschutzkonzept entwickeln. Dieses muss zunächst etwaige Risiken für die Untersuchungspersonen bei Bekanntwerden entsprechender Informationen ebenso wie deren Eintrittswahrscheinlichkeit abschätzen (Art. 25 DSGVO). Um derartige Risiken zu minimieren bzw. ihr Eintreten auszuschließen, ist es notwendig, im Forschungsprojekt entsprechende Vorkehrungen zum Schutz der Untersuchungspersonen und ihrer Identität zu planen und umzusetzen. Hierzu zählen u.a. die sogenannte informierte Einwilligungserklärung, die Pseudonymisierung bzw. Anonymisierung der Daten ebenso wie das geschützte Speichern von sensiblen Informationen oder deren letztendliche Vernichtung.

Das Datenschutzkonzept muss somit zunächst sicherstellen, dass die informierten Einwilligungen aller Untersuchungspersonen vorliegen und dass diese ihre Rechte im Rahmen der Erhebung, Speicherung und Verarbeitung der personenbezogenen Informationen geltend machen können. Analog erhalten die Forschenden durch die Einwilligungserklärung die rechtliche Möglichkeit, personenbezogene Informationen im Rahmen ihres Forschungsprojekts überhaupt nutzen zu dürfen.

Darüber hinaus sind im Rahmen der Speicherung und Verarbeitung personenbezogener Informationen zwei primäre datenschutzrechtliche Maßnahmen zu berücksichtigen. Erstens sollten personenbezogene Daten in der Forschung technisch und organisatorisch so bearbeitet werden, dass die Rechte der Untersuchungspersonen geschützt sind. Hierzu können die Daten pseudonymisiert oder anonymisiert werden. Ein frühzeitiges Auseinandersetzen mit entsprechenden Maßnahmen ist aus unterschiedlichen Gründen relevant. Zunächst sind es die konkret erhobenen Informationen und die Art der Daten, die die Anonymisierungsstrategie leiten. Eine frühzeitige Planung hilft, Datenschutzprobleme zu erkennen und bereits in der Phase der Datenerhebung Risiken für die Untersuchungspersonen zu minimieren. So wird zudem sichergestellt, dass Anonymisierungsmaßnahmen bei der Datenaufbereitung konsistent umgesetzt und dokumentiert werden können.

Zweitens müssen personenbezogene Informationen geschützt vor dem unautorisierten Zugriff Dritter gespeichert und innerhalb des Forschungsprojekts entsprechend kommuniziert werden (Art. 25 DSGVO). Im Datenschutzkonzept ist zu klären, wer wie Zugang zu welchen Daten hat und welche technischen Schutzmaßnahmen, wie z.B. die Verwendung von Passwörtern, die Nutzung geschützter Serverbereiche oder die Verschlüsselung der

Daten, zu ergreifen sind (zum Thema Anonymisierung und Risikomanagement vgl. die ausführliche Darstellung von Elliot et al. 2016: 1f.). Letztendlich ist auch die Vernichtung nicht mehr benötigter Informationen Teil des Datenschutzkonzepts. Im Laufe des Forschungsprojektes werden verschiedene schützenswerte Daten und deren Back-ups obsolet, z.B. die Kontaktinformationen der Untersuchungspersonen, oder müssen aufgrund rechtlicher Vorgaben oder Auflagen Dritter, z.B. von Behörden, vernichtet werden. Auch diese Vernichtung muss im Rahmen des Datenschutzkonzepts genau geplant werden, sodass beispielsweise etwaige Löschrufen im Projektverlauf gewährleistet sind.

Schaukasten 3.2: Datenschutzkonzept

Risikoabschätzung:

- Welche Risiken bergen die Daten bei Bekanntwerden für die Untersuchungspersonen?
- Mit welcher (Eintritts-)Wahrscheinlichkeit werden die Daten bekannt?

Einwilligung der Untersuchungspersonen:

- Liegen die informierten Einwilligungen vor, sodass personenbezogene Daten erhoben, gespeichert und verarbeitet werden können?
- Wie werden das Einholen der Einwilligung und die Wahrnehmung von Rechten organisiert?

Anonymisierung:

- Welche Datenschutzmaßnahmen sind vorgesehen?
- Welche Maßnahmen zum Schutz schwach- bzw. nicht-anonymer Daten sind vorgesehen?

geschütztes Speichern:

- Bestehen geschützte Serverbereiche und wie werden ggf. die Zugriffsrechte organisiert?
- Bestehen Konventionen zum Passwortschutz und zum Verschlüsseln bestimmter Daten und Dateien?

Vernichtung:

- Besteht ein Konzept zur Vernichtung obsoleter Daten und Dateien?
- Wie werden ggf. Löschrufen für einzelne Daten oder Dateien im Projektverlauf gewährleistet?

Quelle: Eigene Darstellung

Schaukasten 3.2 listet zentrale Leitfragen zur projektinternen Planung des Datenschutzkonzepts. Kapitel 4 erörtert darüber hinaus ausführlich Themen und Anforderungen der DSGVO zum Umgang mit personenbezogenen Daten in Forschungsprojekten. Dabei stehen Aspekte der Datenanonymisierung ebenso wie Vorgaben zur informierten Einwilligungserklärung im Fokus.

3.1.3 Verwaltung von Urheberrechten

Neben datenschutzrechtlichen Aspekten spielen in der Planung des Forschungsdatenmanagements Urheberrechte eine zentrale Rolle. Sie sichern die rechtskonforme Verwendung urheberrechtlich geschützter Materialien im Forschungsprojekt. Dementsprechend muss bei der Planung des projektinternen Forschungsdatenmanagements auch ein Konzept zur Organisation und Verwaltung von Urheberrechten entwickelt werden. Dies betrifft zum einen die im Forschungsprojekt generierten Materialien, wie z.B. Forschungsergebnisse oder Daten. Die Urheber- und Verwertungsrechte an derartigen Materialien sollten frühzeitig, d.h. bereits zu Projektbeginn, schriftlich unter den Projektbeteiligten geregelt werden. Nur so kann eine rechtskonforme Verwendung der Daten im Projekt gewährleistet und sichergestellt werden, dass die Verwertungsrechte an Forschungsergebnissen und den damit verbundenen

Publikationen geklärt sind. Sollen generierte Materialien, wie Daten und deren Dokumentationen, nach Projektende auch zur Nachnutzung bereitgestellt werden, empfiehlt es sich, Nutzungsrechte durch eine Lizenz zu regeln.

Schaukasten 3.3: Verwaltung von Urheberrechten

Urheberrechte an im Forschungsprojekt generierten Daten und Dateien:

- Wer hat Urheberrechte an im Forschungsprojekt generierten Daten und Dateien?
- Wie werden die Verwertungsrechte an den Daten und Dateien geregelt?

Urheber- und Verwertungsrechte Dritter:

- Bestehen Urheberrechte Dritter an im Forschungsprojekt verwendeten Materialien?
- Wie erlangt das Forschungsprojekt die Rechte zur Nutzung dieser Materialien im Projekt?

Quelle: Eigene Darstellung

Zum anderen geht es um Urheberrechte Dritter, etwa an Methoden und Standards, wie z.B. Skalen, die im Rahmen der Datenerhebung und Aufbereitung genutzt werden sollen. Werden derartige urheberrechtlich geschützte Materialien eingesetzt, so ist zu klären, welche Nutzung die entsprechenden Verwertungsrechte erlauben. Ein Mangel an entsprechenden Rechten zur Nachnutzung kann sich negativ auf die Möglichkeiten der Datenanalyse oder deren Verwertung im Rahmen wissenschaftlicher Publikationen auswirken. Gerade deshalb ist es zentral, bereits während der Projektplanung und vor der eigentlichen Datenerhebung entsprechende Urheber- und Verwertungsrechte adäquat zu verwalten und zu dokumentieren. In Schaukasten 3.3 sind Leitfragen zur Verwaltung von Urheberrechten für die Projektplanungen zusammengestellt.

3.1.4 Organisation der Forschungsdateien

Zur Planung des projektinternen Forschungsdatenmanagements gehört auch die logische und technische Organisation der Daten und Dateien im Forschungsprojekt. Die Entwicklung und konsistente Umsetzung eines entsprechenden Organisationkonzepts ist gerade in Forschungsk Kooperationen unerlässlich. Sie dient primär der Arbeitserleichterung und der Sicherung der Arbeit als solche. Eine gute Organisation ermöglicht es allen Projektbeteiligten, Daten und Dateien schnell zu identifizieren, zu bearbeiten und zu sichern.

Ein entsprechendes Organisationkonzept sollte zunächst eine Ordnerstruktur für das Forschungsprojekt festlegen. Dies kann beispielsweise anhand unterschiedlicher Daten- und Dateitypen oder anhand verschiedener Themenkomplexe geschehen. Hierbei ist der Umgang mit personenbezogenen Informationen und datenschutzrechtlich relevanten Daten und Dateien zu berücksichtigen. Zudem sind etwaige Regelungen und Maßnahmen des projektinternen Datenschutzkonzepts im Auge zu behalten.

Analog muss das Konzept klären, wie Daten und Dateien im Forschungsprojekt benannt und versioniert werden. So können Dateinamen eine Reihe von Informationen, wie etwa den Dateityp, die Zugehörigkeit zu anderen Dateien, die erstellende bzw. bearbeitende Person oder die Version der Datei enthalten. Denkbar ist es auch, datenschutzrechtlich relevante Dateien durch eine spezielle Kennzeichnung im Dateinamen zu markieren. Analog ist bei der Entwicklung von Versionierungsregeln darauf zu achten, dass Beziehungen zwischen unterschiedlichen Versionen zusammengehöriger Dateien ersichtlich bleiben, um beispielsweise einer bestimmten Version des Datensatzes eine bestimmte Version des Codebuchs dauerhaft zuordnen zu können.

Schließlich muss im Rahmen des Organisationskonzepts sichergestellt werden, dass die Daten durch eine geeignete Back-up-Strategie ausreichend vor versehentlichem Verlust geschützt sind. Sie legt fest, welche Dateien wo, wie oft, von wem gesichert werden und wie der Erhalt sowie die Lesbarkeit der Sicherungskopien gewährleistet werden. In der Praxis besitzen heute die meisten Forschungseinrichtungen institutseigene Back-up-Strategien, basierend auf automatisierten Sicherungsprozessen. Entsprechend ist zu klären, ob am jeweiligen Institut derartige automatisierte Prozesse existieren und wie sie funktionieren. Dabei müssen Maßnahmen des Datenschutzkonzepts auch in Bezug auf die Sicherungskopien umgesetzt werden. Diese sind ggf. zu verschlüsseln, mit einem Passwort zu schützen und vor dem unautorisierten Zugriff Dritter zu sichern. Die Leitfragen in Schaukasten 3.4 dienen der Orientierung bei der Projektplanung. Kapitel 5 erörtert darüber hinaus ausführlich Strategien und Maßnahmen zum systematischen und rechtkonformen Umgang mit Forschungsdateien in Projekten.

Schaukasten 3.4: Organisation der Forschungsdateien

Ordnerstruktur:

- Wie werden die Daten und Dateien im Rahmen der Ordnerstruktur organisiert?
- Wie werden die Zugriffsrechte der Projektbeteiligten organisiert?

Datenbenennung und -versionierung:

- Bestehen Regeln zur eindeutigen Benennung von Daten und zugehörigen Dateien?
- Bestehen Regeln zur eindeutigen Versionierung von Daten und zugehörigen Dateien?

Sicherung gegen Verlust (Back-ups):

- Wie werden die Daten und Dateien technisch gegen Verlust gesichert (Back-ups)?
- Bestehen in der jeweiligen Forschungseinrichtung automatisierte Back-up-Prozesse?
- Wie werden Datenschutzmaßnahmen im Kontext der Back-ups berücksichtigt?

Quelle: Eigene Darstellung

3.1.5 Ressourcen und Verantwortlichkeiten

Die Planung und Umsetzung des projektinternen Forschungsdatenmanagements erfordert letztendlich die vorausschauende Berücksichtigung notwendiger Ressourcen und Verantwortlichkeiten. Gerade in Forschungsk Kooperationen sollten Verantwortlichkeiten für die unterschiedlichen Aktivitäten des Forschungsdatenmanagements von Beginn an klar definiert und zwischen den Kooperationspartnern schriftlich vereinbart werden. Generell empfiehlt es sich, die Aktivitäten des Forschungsdatenmanagements als festen Bestandteil der Projektorganisation zu betrachten und dort zu verorten, wo die jeweilige Tätigkeit auch tatsächlich ausgeführt werden soll. So lässt sich beispielsweise die Dokumentation des Frageprogramms an die Entwicklung des Forschungsdesigns knüpfen. Entsprechend sind jene Projektbeteiligten, die für die Entwicklung des Frageprogramms verantwortlich sind, auch für dessen adäquate Dokumentation zuständig.

Bei der Ressourcenplanung bzw. der Ausarbeitung eines Budgetplans ist zu berücksichtigen, dass auch das Forschungsdatenmanagement selbst Kosten verursacht. Diese sind pauschal nur schwer definierbar. Sie hängen primär von der konkreten wissenschaftlichen und organisatorischen Gestaltung des Forschungsprojektes und den zu generierenden Daten ab. Bei der Kostenkalkulation sollten Forschende daher zunächst ihr Forschungsvorhaben hinsichtlich aller Kostenaspekte reflektieren. Darauf aufbauend können die unterschiedlichen Aktivitäten des projektinternen Forschungsdatenmanagements definiert und budgetiert

werden. Zur ersten Orientierung lassen sich die dabei entstehenden Kosten in Sach- und Personalmitteln differenzieren. Unter Personalmitteln sind sämtliche personelle Aufwendungen zu subsummieren, z.B. für die Koordination des Forschungsdatenmanagements oder für die relativ zeitaufwendige Aufbereitung und Dokumentation der Daten. Sachmittel beinhalten u.a. Kosten für Softwarelizenzen oder für ausreichende Speicherkapazitäten etc. (Jensen 2012:16f.). Wesentliche Fragestellungen bei der Planung von Verantwortlichkeit und Ressourcen im Rahmen des Forschungsdatenmanagements zeigt Schaukasten 3.5.

Schaukasten 3.5: Ressourcen und Verantwortlichkeiten

Verantwortlichkeiten:

- Welche Aufgaben und Anforderungen sind im Forschungsdatenmanagement zu berücksichtigen?
- Wie werden die entsprechenden Verantwortlichkeiten festgelegt und wo werden diese verortet?
- Sind besondere Anforderungen, z.B. im Rahmen von Forschungsk Kooperationen, zu berücksichtigen?

Ressourcen:

- Wie beeinflussen das Forschungsprojekt und die zu generierenden Daten die Kosten des Forschungsdatenmanagements?
- Welche Personalkosten fallen im Rahmen des Forschungsdatenmanagements an?
- Welche Sachkosten fallen im Rahmen des Forschungsdatenmanagements an?

Quelle: Eigene Darstellung

3.2 Forschungsdatenmanagement über das Projektende hinaus planen

Die systematische Planung des projektinternen Forschungsdatenmanagements dient der Sicherung der guten wissenschaftlichen Praxis im Projektverlauf. Sie unterstützt die Umsetzung der eigentlichen Forschungsziele und sichert die Qualität, Verständlichkeit und Interpretierbarkeit der zu verwendenden Daten. Ebenso wird durch die systematische Planung der datenschutz- und urheberrechtskonforme Umgang anhand geeigneter Konzepte gewährleistet. Ein vorausschauend geplantes, projektinternes Forschungsdatenmanagement schafft aber auch die notwendigen Grundlagen, um die Daten über das eigentliche Forschungsprojekt hinaus zu erhalten:

Good research data management is not a goal in itself, but rather the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse. (Horizon 2020 Programme 2016: 3)

Im Rahmen der Projektplanung muss daher nicht nur das projektinterne Forschungsdatenmanagement berücksichtigt werden, sondern es bedarf auch der Entwicklung geeigneter Konzepte zur langfristigen Sicherung der Daten nach Projektende. Forschende sollten dabei Aktivitäten des projektinternen Forschungsdatenmanagements von den weiterführenden Maßnahmen zur Sicherung der Daten über das Projekt hinaus abgrenzen. Dies erleichtert die Planung, dient vor allem aber einer gezielten Umsetzung weiterführender Maßnahmen. Schließlich müssen derartige weiterführende Aktivitäten im Projektverlauf implementiert und umgesetzt werden, auch wenn ihre Maßnahmen weiter über das Projekt hinausreichen. Letztendlich unterstützt die Unterscheidung auch die Erfüllung etwaiger Auflagen und Anforderungen Dritter, wie z.B. von akademischen Journals oder Drittmittelgebern. Dabei lassen sich primär zwei Auflagen Dritter erkennen, denen zwei unterschiedliche wissenschaftliche Nachnutzungszwecke zugrunde liegen: a) die Replikation publizierter Forschungsergebnisse

zur Verifizierung und b) die Bereitstellung der Daten zur Nachnutzung durch Dritte in neuen Forschungskontexten (*Data Sharing*).

3.2.1 Die Replizierbarkeit von Forschungsergebnissen sicherstellen

Ein Grundsatz guter wissenschaftlicher Praxis ist die Transparenz und Überprüfbarkeit von publizierten Forschungsergebnissen ebenso wie von verwendeten Daten. So fordert die Deutsche Forschungsgemeinschaft in ihren Leitlinien zum Umgang mit Forschungsdaten, dass diese „in der eigenen Einrichtung oder in einer fachlich einschlägigen, überregionalen Infrastruktur für mindestens 10 Jahre archiviert werden“ (DFG 2015: 1) sollen, um u.a. die Replizierbarkeit zu gewährleisten (vgl. dazu auch Huschka/Oellers 2013; Allianz der deutschen Wissenschaftsorganisationen 2010). Unter Replikation wird dabei zum einen die Prüfung und Verifikation von Analyseergebnissen verstanden, indem diese mit den vorhandenen Daten unter Nutzung der gleichen Methoden reproduziert werden. Zum anderen meint Replikation den Versuch, Ergebnisse einer wissenschaftlichen Studie durch Erhebung neuer Daten mit gleichen Analysemethoden zu verifizieren (Freese 2007: 178). Analog fordern heute mehr und mehr Hochschulen und Forschungseinrichtungen bzw. -verbände, z.B. die Leibniz-Gemeinschaft (2015) oder die Max-Planck-Gesellschaft (2000), aber auch akademische Journals, etwa die Politische Vierteljahresschrift (2011), die Replizierbarkeit publizierter Forschungsergebnisse.

Die Planung weiterführender Aktivitäten des Forschungsdatenmanagements muss daher auch die Entwicklung eines Konzepts zum längerfristigen Datenerhalt berücksichtigen. Forschende sollten dazu zunächst ihr Forschungsprojekt und die potentiell zu replizierenden Daten reflektieren. Darauf aufbauend gilt es zu klären, welche Anforderungen und Standards zur Replikation zu berücksichtigen sind. So schreibt etwa King (1995: 444):

The replication standard holds that sufficient information exists with which to understand, evaluate, and build upon a prior work if a third party could replicate the results without any additional information from the author.

Wann derartige *hinreichende* Informationen gegeben sind, sodass die Daten für Dritte zugänglich, transparent, verständlich und interpretierbar bleiben, beschreiben mittlerweile vereinzelte Leitlinien von Datenrepositorien, u.a. von Dataverse (o.J.). Entsprechende disziplinspezifische Leitlinien von Journalen, etwa im Bereich der Politik- oder der Wirtschaftswissenschaften, diskutieren beispielsweise Gherghina und Katsanidou (2013) oder Vlaeminck und Siegert (2012).

Um die Nachvollziehbarkeit der durchgeführten Datenanalysen und ihrer Ergebnisse praktisch zu ermöglichen, müssen neben den verwendeten Daten alle zur Replikation notwendigen Skripte, beispielsweise zur Datenaufbereitung und -analyse, in verständlicher und nachvollziehbarer Form erhalten werden. Weitergehende Strategien zur Sicherung von Aufbereitungsskripten beschreibt der dritte Abschnitt in Kapitel 8. Im Folgenden stehen generelle Aspekte zum Erhalt von Forschungsdaten zu Replikationszwecken im Vordergrund.

In der Planung des Forschungsdatenmanagements erfordert z.B. der zehnjährige Erhalt der Daten, dass Konzepte und Maßnahmen während der Projektlaufzeit und nach Projektende wirksam ineinandergreifen. Einzelne Aktivitäten des projektinternen Forschungsdatenmanagements müssen dazu an das Ziel des längerfristigen Datenerhalts angepasst werden. Dies betrifft insbesondere die Aufbereitung und Dokumentation der Daten. Um über die nächsten zehn Jahre transparent und nachvollziehbar zu bleiben, müssen die Daten ebenso wie genutzte Methoden und Konzepte hinreichend dokumentiert sein. Dabei sollten soweit als

möglich Metadatenstandards verwendet werden, um die Verständlichkeit der Dokumentation dauerhaft zu erhöhen.

Darüber hinaus ist der zehnjährige Erhalt der Daten ggf. aber auch an rechtliche Voraussetzungen gebunden. Hierzu zählen einerseits Vorgaben des Datenschutzes. So darf beispielsweise die informierte Einwilligung der Untersuchungspersonen den längerfristigen Datenerhalt nicht ausschließen. Sollen personenbezogene Informationen über zehn Jahre hinweg erhalten bleiben, so muss auch deren Vertraulichkeit gewährleistet werden. Entsprechend sind Maßnahmen zu ergreifen, die den Schutz der Daten für diesen Zeitraum sicherstellen. Analog gilt es, die Urheberrechte an allen zu erhaltenden Daten und Begleitmaterialien im Auge zu behalten und sicherzustellen, dass entsprechende Verwertungsrechte bestehen. Dies betrifft sowohl die Urheberrechte an den im Forschungsprojekts generierten Daten und Dateien als auch eventuelle Urheberrechte Dritter, etwa an verwendeten und daher zu erhaltenden Materialien.

Schließlich gilt es zu klären, in welchen Dateiformaten die Daten und Begleitmaterialien, wie z.B. die Datendokumentation, gespeichert werden sollen. Zu bedenken ist in diesem Zusammenhang, dass die Forschenden die Zugänglichkeit zu den Daten und Dateien längerfristig sicherstellen müssen. Die zur Speicherung verwendeten Formate sollten demnach so gewählt werden, dass die einzelnen Dateien nutzbar bleiben. Analog zum projektinternen Forschungsdatenmanagement empfiehlt sich daher auch beim längerfristigen Datenerhalt die Verwendung nicht-proprietärer Formate, deren Quellcodes ggf. Teil der zu erhaltenden Begleitmaterialien sein können.

Der eigentliche physische Datenerhalt kann dann durch das Forschungsprojekt selbst organisiert und umgesetzt werden. Eine (mindestens) zehnjährige Sicherung von Forschungsdaten zu Replikationszwecken setzt jedoch voraus, dass im Rahmen des Forschungsprojektes eine entsprechende Strategie entwickelt wird. Diese umfasst zunächst die Definition von Verantwortlichkeiten für den Erhalt. Daran anschließend muss festgelegt werden, welche Daten und Dateien erhalten und in welchen Formaten diese gespeichert werden sollen. Dies betrifft neben der Klärung der Bitstream-Sicherung, d.h. den „physische[n] Erhalt des gespeicherten Datenobjekts (Bitstreams)“ (Ulrich 2010: Kap. 8:3), das Einhalten eventueller datenschutzrechtlicher Vorgaben. Dabei muss der Zugang zu den Daten für die Projektbeteiligten ebenso wie für Dritte zu Replikationszwecken über die Jahre hinweg gewährleistet werden. Schließlich sind die Kosten der Sicherung und eventuelle Mehraufwendungen, z.B. im Rahmen einer standardisierten Dokumentation, zu berücksichtigen und in die Projektplanung aufzunehmen (zu Kosten bzw. Modellen der Kostenermittlung vgl. Kimpton/Morris 2014; Beagrie 2011).

Alternativ zum längerfristigen Erhalt durch das Forschungsprojekt lassen sich die Daten auch in einem Datenrepository archivieren. Repositorien ermöglichen die längerfristige Sicherung der Daten unter Einhaltung rechtlicher Vorgaben und gewährleisten deren Zugänglichkeit für die Projektbeteiligten ebenso wie für Dritte zu Replikationszwecken. Im Rahmen der Projektplanung sollten Forschende frühzeitig die sachgerechte Sicherung oder Archivierung in einer Dateninfrastruktureinrichtung thematisieren. In der Entwicklung eines Konzepts zum längerfristigen Datenerhalt gilt es, mögliche Anforderungen an die zu archivierenden Daten, die Abläufe der Archivierung sowie die Nutzungsbedingungen des jeweiligen Repositoriums rechtzeitig zu beachten und entsprechend im Projektverlauf umzusetzen. Dabei ist auch zu klären, was nach Ablauf der vertraglich geregelten Archivierungsfrist mit den Daten im Repository geschieht. Letztendlich müssen eventuelle Archivierungskosten des jeweiligen Repositoriums berücksichtigt und im Rahmen der Projektplanung budgetiert werden. Schaukasten 3.6 fasst die Themen und Fragestellungen zusammen, die bei der Planung zur Sicherstellung replizierbarer Forschungsergebnisse berücksichtigt werden sollten.

Schaukasten 3.6: Die Replizierbarkeit von Forschungsergebnissen sicherstellen

Vorüberlegungen:

- Zu welchem Zweck sollen die Daten längerfristig erhalten werden?
- Welche Standards und Vorgaben Dritter müssen beim Datenerhalt beachtet werden?

Organisatorischer und technischer Datenerhalt:

- Welche Daten und Begleitdateien sollen erhalten werden?
- Wo und wie lange erfolgt die Sicherung nach Projektende?
- Wie wird der Datenzugang für Projektbeteiligte und Dritte zu Replikationszwecken geregelt?
- Wie erfolgt die technische Sicherung, d.h. Bit-Stream-Sicherung, Prüfung der Lesbarkeit, Back-ups etc.?
- In welchen Dateiformaten werden die Daten und Begleitdateien gespeichert?
- Was geschieht mit den Daten nach Ablauf der Sicherungsfrist?
- Welche Kosten fallen zusätzlich an und wie werden diese durch das Projektbudget gedeckt?

Dokumentation:

- Sind alle Daten ausreichend dokumentiert?
- Welche Metadaten(-standards) werden genutzt?
- Wird ein kontrolliertes Vokabular genutzt?

Datenschutz:

- Ermöglicht die informierte Einwilligung den längerfristigen Datenerhalt?
- Sind die Daten ausreichend anonymisiert?
- Wie werden schwach bzw. nicht anonymisierte Daten vor dem unautorisierten Zugriff Dritter längerfristig geschützt?

Urheberrechte:

- Liegt die Genehmigung aller Urhebenden zum längerfristigen Datenerhalt vor?
- Sind Urheberrechte Dritter vom Datenerhalt betroffen?
- Wie werden ggf. Urheberrechte Dritter berücksichtigt und Verwertungsrechte sichergestellt?

Ressourcen:

- Sind die notwendigen Ressourcen zum längerfristigen Datenerhalt im Projektbudget berücksichtigt?

Quelle: Eigene Darstellung

3.2.2 Die Bereitstellung von Forschungsdaten zur Nachnutzung

In Abgrenzung zum Datenerhalt zu Replikationszwecken geht es bei der Bereitstellung der Daten (*Data Sharing*) um die Möglichkeit Dritter, diese in neuen Kontexten nachnutzen zu können. Die sogenannte Sekundäranalyse von bestehenden Datenbeständen ist eine anerkannte Forschungsstrategie, die von Forschenden der eigenen Disziplin bzw. von Nachbardisziplinen bis hin zu transdisziplinär Forschenden genutzt wird. Sie umfasst die Nachnutzung bereits existierender Daten zum Vergleich thematisch und methodisch ähnlicher Datensätze, zur Bearbeitung neuer Forschungsfragen ebenso wie z.B. in der akademischen Lehre, ohne dabei selbst Daten generieren zu müssen.

Analog zum Datenerhalt zu Replikationszwecken reicht das Ziel der Datenbereitstellung weit über das eigentliche Forschungsprojekt hinaus. Die Nachnutzung der Daten durch Dritte findet häufig erst statt, wenn das Forschungsprojekt als solches bereits beendet und das Analysepotential der Daten im Rahmen des originären Frageprogramms weitestgehend ausgeschöpft ist. Die Bereitstellung der Daten muss aber im Verlauf des eigentlichen Forschungsprojekts geplant und vorbereitet werden, auch wenn diese erst zum Projektende Dritten verfügbar gemacht werden.

Dabei sollten sich die Forschenden zunächst mit etwaigen Auflagen zum *Data Sharing* auseinandersetzen. Diese können sich z.B. aus institutseigenen Verpflichtungen, etwa zum *Open Access* im Rahmen der Berliner Erklärung (2003), oder aus Regelungen z.B. von Drittmittelgebern ableiten. So fordert beispielsweise die DFG (2015: 1), dass bereits bei der Planung des zu beantragenden Forschungsprojektes Überlegungen angestellt werden,

ob und welche der aus einem Vorhaben resultierenden Forschungsdaten für andere Forschungskontexte relevant sein können und in welcher Weise diese Forschungsdaten anderen Wissenschaftlerinnen und Wissenschaftlern zur Nachnutzung zur Verfügung gestellt werden können.

Diese sollten „so zeitnah wie möglich verfügbar gemacht werden“ (ebd.) und Dritten die weitere (Nach-)Nutzung ermöglichen.

Mit derartigen Auflagen einher gehen zumeist Forderungen zum Betreiben eines geeigneten Forschungsdatenmanagements und dessen Dokumentation, etwa in Form des Datenmanagementplans. Damit soll sichergestellt werden, dass die im Projektverlauf generierten Daten auch tatsächlich bereitgestellt werden können (vgl. etwa die Anforderungen in Horizon 2020 Programme – European Commission 2016). Gerade Forschende in Drittmittelprojekten sollten sich mit derartigen Auflagen zur Bereitstellung ihrer Daten und zur Erstellung eines Datenmanagementplans vertraut machen. Generell gilt, dass sich dieser Datenmanagementplan für Drittmittelgeber von der projektinternen Dokumentation des Forschungsdatenmanagements unterscheidet. Im Gegensatz zum projektinternen Datenmanagementplan sollte sich das Berichtswesen gegenüber Mittelgebern auf die wesentlichen Aspekte der Generierung nachnutzbarer und weitergabefähiger Daten fokussieren und bestehende Anforderungen und Auflagen berücksichtigen.

Analog zum Datenerhalt zu Replikationszwecken müssen Forschende im Rahmen ihrer Projektplanung die Entwicklung eines Konzepts zur Bereitstellung ihrer Daten angehen. Dieses Konzept muss, unter Berücksichtigung etwaiger spezifischer Förderauflagen, zunächst klären, welche Daten und Dateien, wem, unter welchen Bedingungen, wie bereitgestellt werden sollen. Im Gegensatz zum Datenerhalt zu Replikationszwecken werden zur Sekundäranalyse zumeist nur einige wenige Dateien des ursprünglichen Forschungsprojekts zur Verfügung gestellt. So werden in der Regel die ursprünglich erhobenen und noch nicht aufbereiteten (Roh-)Daten des Forschungsprojekts nicht für Dritte verfügbar gemacht, sondern ggf. nur längerfristig zu Replikationszwecken gesichert. Dementgegen umfasst die Bereitstellung zur Nachnutzung zumeist nur die Forschungsdaten sowie notwendige Begleitmaterialien zur Sicherung ihrer Nutzbarkeit, wie beispielsweise die Datendokumentation. Im Vergleich zum Datenerhalt zu Replikationszwecken sollten Forschende bei der Datenbereitstellung zur Nachnutzung aber nicht nur jene Daten zugänglich machen, die zur Erstellung der unterschiedlichen Projektergebnisse verwendet wurden. Vielmehr sollten alle im Forschungsprojekt generierten Daten – wenn möglich in einem integrierten Datensatz – bereitgestellt werden. Dieser kann auch Informationen enthalten, die im Forschungsprojekt zwar erhoben wurden, in der eigentlichen Projektarbeit aber ungenutzt blieben. Die Forschenden steigern damit das Analysepotential ihrer Daten und erhöhen deren Nachnutzbarkeit durch Dritte.

Neben der Frage, welche Daten verfügbar gemacht werden sollen, muss für das Konzept zur Erstellung nachnutzbarer Daten auch berücksichtigt werden, was die Bereitstellung überhaupt bedeutet. Um dies besser verstehen zu können, empfiehlt sich ein kurzer Blick auf die sogenannten *FAIR Data Principles* (Force11 2017; Wilkinson et al. 2016; Recker et al. 2015). Das Akronym FAIR steht für vier grundlegende Prinzipien – *findable, accessible, interoperable* und *re-usable* –, die die Daten erfüllen müssen, um als nachnutzbar zu gelten.

Demzufolge müssen nachnutzbare Daten erstens auffindbar (*findable*) sein, d.h. Dritte müssen von den Daten Kenntnis erlangen können. Für die Forschenden impliziert dies, die Metadaten ihrer Daten möglichst publik zu machen, etwa über die eigene Projektwebseite

oder durch die Registrierung der Daten in Datenkatalogen wie DataCite (2016). Zweitens müssen die Daten für Dritte sowohl physisch als auch rechtlich zugänglich (*accessible*) sein. Dritte müssen tatsächlich auf die Daten zugreifen können, was sowohl die Formate der einzelnen Dateien als auch die damit verbundenen Zugangsrechte betrifft.

Schaukasten 3.7: *datorium* – Datenrepositorium für die Sozial- und Wirtschaftswissenschaften

<https://datorium.gesis.org/xmlui/>

datorium ist ein Service für die eigenständige Dokumentation, Sicherung und Veröffentlichung von Forschungsdaten, der durch eine Nutzungsbedingung abgesichert wird.

Registrierung der Daten:

- Einspielen von Metadaten in Online-Katalogen zur freien Recherche im Internet;
- Vergabe eines Digital Objective Identifiers (DOI 2018) zur eindeutigen Zitier- und Referenzierbarkeit.

Nachnutzung:

- Lizenzierung der Daten zur Nachnutzung für Dritte;
- physische und rechtskonforme Zugänglichkeit für mindestens zehn Jahre.

datorium bietet darüber hinaus Unterstützung bei der Datenbereitstellung, etwa in Form von Handreichungen und Best-Practise-Empfehlungen, z.B. zum Umgang mit Syntaxen oder zur Aufbereitung und Anonymisierung quantitativer Daten (*datorium* o.J.b)

Quelle: Eigene Darstellung

Die beiden Prinzipien der Auffindbarkeit und Zugänglichkeit hängen dabei primär von zwei Fragen ab: Wie und wo werden die Daten archiviert und für Dritte bereitgestellt? Forschende, die ihre Daten bereitstellen wollen, sollten daher bereits bei der Projektplanung überlegen, ob die Daten durch das Forschungsprojekt selbst – z.B. über die Projektwebseite¹ – oder via externer Dateninfrastruktureinrichtungen – z.B. ein institutionelles oder disziplinspezifisches überregionales Datenrepositorium – verfügbar gemacht werden sollen (ZBW/GESIS/RatSWD 2015: 22; vgl. die ausführliche Darstellung in Kapitel 7). Repositorien sichern die physische und rechtskonforme Zugänglichkeit der Daten und steigern deren Auffindbarkeit, u.a. durch die Registrierung der Daten in Datenkatalogen und die Vergabe persistenter Identifikatoren (Corti et al. 2014; Huschka et al. 2011; Jensen/Katsanidou/Zenk-Möltgen 2011; Doorn 2010). Schaukasten 3.7 zeigt ein Beispiel für ein derartiges disziplinspezifisches, überregionales Datenrepositorium: *datorium* (o.J.a), ein Service von GESIS – Leibniz-Institut für Sozialwissenschaften.

Das dritte FAIR-Prinzip bezieht sich auf die *Interoperabilität* der Daten. Dritte müssen die bereitgestellten Daten technisch nachnutzen können. Dies betrifft zum einen die verwendeten Dateiformate. Forschende sollten sich an disziplinspezifischen Standards orientieren und möglichst auf nicht-proprietäre Formate zum Speichern der Daten und Begleitmaterialien zurückgreifen. Zum anderen umfasst Interoperabilität die Datendokumentation. Diese sollte auf einem kontrollierten Vokabular basieren und disziplinspezifische (Metadaten-)Standards nutzen, um die Transparenz der Daten und ihre Verständlichkeit dauerhaft zu erhöhen.

Das letzte FAIR-Prinzip benennt schließlich die analytische Nachnutzbarkeit (*re-usable*) der Daten. Um mit den Daten in neuen (Forschungs-)Kontexten weiterarbeiten zu können, müssen Dritte diese inhaltlich und methodisch verstehen, interpretieren und nachvollziehen können. Dementsprechend ist es erforderlich, dass alle für die Nachnutzung relevanten

1 Die Bereitstellung der Daten über das Projektende hinaus ist allein durch das eigene Forschungsprojekt zumeist schwer zu realisieren. So fehlen zumeist die notwendigen technischen und personellen Ressourcen, um den Zugang zu den Daten längerfristig sicherzustellen.

Informationen vorliegen. Nur so können das Analysepotential der Daten und deren Qualität beurteilt werden. Dies umfasst die Daten als solche ebenso wie ihren Entstehungsprozess, die zugrunde liegenden methodischen Aspekte des Studiendesigns, die Datenerhebung etc. Insgesamt gilt, dass die Daten in Kombination mit ihrer Dokumentation selbsterklärend sein müssen und Dritte diese nachnutzen können, ohne weitere Informationen von den ursprünglichen Forschenden einzuholen. Haben die Forschenden bei der Aufbereitung und Dokumentation ihrer Daten im Rahmen des projektinternen Forschungsdatenmanagements bereits ein besonderes Augenmerk auf die Vollständigkeit der Informationen und die Verwendung von (Metadaten-)Standards gelegt, werden wesentliche Anforderungen für die analytische Nachnutzbarkeit der Daten bereits erfüllt sein.

Um Daten entsprechend der *FAIR*-Prinzipien für Dritten nachnutzbar zu machen, müssen abermals rechtliche Aspekte berücksichtigt und in ein entsprechendes Bereitstellungs-konzept integriert werden. Dies betrifft zum einen den Datenschutz. Analog zum Datenerhalt zu Replikationszwecken darf auch die Bereitstellung der Daten und deren Nachnutzung nicht per se durch die Einwilligungserklärung ausgeschlossen werden und die Daten müssen zur Weitergabe an Dritte ausreichend anonymisiert sein. Daten, die nicht oder nicht ausreichend anonymisiert werden können, z.B. um deren Analysepotential zu erhalten, müssen ggf. mit speziellen Zugangs- und Nachnutzungsrechten versehen und unter bestimmten organisatorischen und rechtlichen Bedingungen verfügbar gemacht werden, wie im Abschnitt 3 des vierten Kapitels näher ausgeführt wird. Zum anderen müssen die notwendigen Urheberrechte vorliegen, um entsprechende Verwertungsrechte, etwa zur Nachnutzung, an Dritte übertragen zu können, wie in Abschnitt 3.1.3 erörtert.

Letztendlich müssen im Konzept zur Bereitstellung der Daten die dabei anfallenden Kosten eingeplant und budgetiert werden. Ebenso wie beim Datenerhalt zu Replikationszwecken betrifft dies zunächst die Mehraufwendungen zur Erstellung nachnutzbarer Daten und deren Begleitmaterialien, wie z.B. die Datendokumentation. Darüber hinaus müssen ggf. aber auch Aufwendungen zur Gewährleistung der Auffindbarkeit und Zugänglichkeit der Daten berücksichtigt werden (vgl. zu Kosten bzw. zur Kostenermittlung Kimpton/Morris 2014; Beagrie 2011). Ist die Bereitstellung der Daten mit einer Auflage durch Drittmittelgeber verbunden, sollten die Forschenden klären, ob und wie sie entsprechende Mehraufwendungen beantragen können. So folgen beispielsweise das Horizon 2020 Programme (2016) ebenso wie die DFG (2015: 1) dem Förderansatz, dass

projektspezifische Kosten, die im Rahmen eines wissenschaftlichen Projekts bei der Aufbereitung von Forschungsdaten für eine Anschlussnutzung bzw. für die Überführung von Forschungsdaten in existierende Infrastrukturen entstehen, [...] eingeworben werden [können].

Um diese Kosten adäquat geltend zu machen, ist eine frühzeitige Planung im Rahmen des Projektantrags unabdingbar.

In diesem Zusammenhang sollten Forschende auch prüfen, ob verschiedene Aktivitäten zur Generierung nachnutzbarer Daten nicht durch spezialisierte Dienstleister erbracht werden können. So bietet beispielsweise das Datenarchiv für Sozialwissenschaften der GESIS im Rahmen der Datenarchivierung (GESIS 2018a) spezifische Datenservices an, z.B. zur Aufbereitung und Dokumentation von Forschungsdaten (GESIS 2018b). Entsprechenden Aktivitäten werden damit an ein speziell geschultes Fachpersonal übertragen, deren Kompetenzen in der Aufbereitung und Dokumentation von Daten nach (internationalen) Standards liegen. Umgekehrt schaffen sich Forschende so den Freiraum, sich primär auf ihre Forschung und weniger auf die Generierung und Bereitstellung ihrer Daten zu konzentrieren. Die entstehenden Kosten zur Auslagerung derartiger Aktivitäten des Forschungsdatenmanagements an Serviceeinrichtungen müssen dabei im Projektbudget berücksichtigt und ggf. gegenüber

den Drittmittelgebern geltend gemacht werden. Schaukasten 3.8 fasst die wesentlichen Aspekte und Fragen zur Planung der Bereitstellung von Forschungsdaten zusammen.

Schaukasten 3.8: Die Bereitstellung von Forschungsdaten zur Nachnutzung

Vorüberlegungen:

- Zu welchem Zweck sollen die Daten bereitgestellt werden?
- Welche Standards und Vorgaben Dritter, wie z.B. Forschungsförderer, sind zu beachten?
- Welche Richtlinien und Empfehlungen von Datenrepositorien sind ggf. zu beachten?

organisatorische und technische Bereitstellung:

- Welche Daten und Begleitdateien sollen bereitgestellt werden?
- Wie und wo werden die Daten und Begleitdateien bereitgestellt?
- Wie wird die Auffindbarkeit und Zugänglichkeit zu den Daten und Begleitdateien sichergestellt?
- In welchen Dateiformaten werden die Daten und Begleitdateien bereitgestellt?

Dokumentation:

- Sind die Daten selbsterklärend dokumentiert?
- Welche zusätzliche Dokumentation ist erforderlich, um die Transparenz und Verständlichkeit der Daten zu gewährleisten?
- Welche Metadaten(-standards) werden genutzt?
- Wird ein kontrolliertes Vokabular genutzt?
- Werden die Daten zu besserer Auffindbarkeit und Zitierbarkeit registriert?

Datenschutz:

- Ermöglicht die informierte Einwilligung die Nachnutzung der Daten durch Dritte?
- Sind die Forschungsdaten ausreichend anonymisiert?
- Wie wird die Nachnutzung schwach- bzw. nicht-anonymisierter Daten geregelt?

Urheber- und Verwertungsrechte:

- Liegt die Genehmigung aller Urhebenden zur Bereitstellung der Daten vor?
- Bestehen Urheberrechte Dritter an bereitzustellenden Materialien?
- Wie wird das Übertragen von Verwertungsrechten zur Nachnutzung geregelt, z.B. durch Lizenzen?

Ressourcen:

- Entstehen durch die Bereitstellung zusätzliche Kosten, z.B. durch die Archivierung oder durch eine granulare Datendokumentation?
- Sind zusätzliche Kosten im Projektbudget eingeplant?
- Können zusätzliche Kosten ggf. gegenüber Drittmittelgebern geltend gemacht werden?

Quelle: Eigene Darstellung

3.3 Abschließende Bemerkung

Forschungsdatenmanagement ist ein Muss für jedes empirische Forschungsprojekt. Eine systematische und frühzeitige Planung des Forschungsdatenmanagements ist eine der zentralen Voraussetzungen zur erfolgreichen Umsetzung des eigentlichen Forschungsvorhabens. Nur durch entsprechende Konzepte im Rahmen der Projektplanung kann sichergestellt werden, dass in der angestrebten Datenanalyse qualitativ hochwertige Forschungsdaten vorliegen, die auf Basis ihrer Dokumentation transparent, verständlich und nachvollziehbar sind und eine rechtskonforme Nutzung erlauben. Dabei ist das Forschungsdatenmanagement auch ein Planungsinstrument zum Umgang mit den Daten über das Projektende hinaus. Die Entwicklung und Implementierung von Konzepten zum längerfristigen Datenerhalt bzw. zur

Bereitstellung der Daten zur Nachnutzung durch Dritte macht das Forschungsdatenmanagement zu einem elementaren Bestandteil guter wissenschaftlicher Praxis. Es sichert Transparenz im Forschungsprozess, ermöglicht die Replikation von Forschungsergebnissen ebenso wie von verwendeten Daten und schafft die notwendigen Voraussetzungen zur weiteren Nutzung der Daten in neuen Kontexten.

Abschließend bleibt hervorzuheben, dass jedes Forschungsprojekt ein Unikat ist und die dort generierten Forschungsdaten einzigartig sind. Eine systematische Planung des Forschungsdatenmanagements muss dieser Einzigartigkeit gerecht werden und sich am jeweiligen Forschungsprojekt und den zu generierenden Daten orientieren. Zur Planung des projektinternen Forschungsdatenmanagements ebenso wie zur Implementierung weiterführender Maßnahmen sollten Forschende daher zunächst ihr eigentliches Forschungsvorhaben reflektieren. Dabei gilt es klar zu definieren, welche weiterführenden Ziele die Forschenden im Umgang mit den Daten über das Projektende hinaus anstreben. Aufbauend auf dem Projektvorhaben, den zu generierenden Daten und den damit verbundenen weiterführenden Zielen lassen sich dann das projektinterne Forschungsdatenmanagement und die notwendigen Maßnahmen zum Datenerhalt bzw. zur Datenbereitstellung gezielt planen, in die Projektablaufe integrieren und im Projektverlauf systematisch umsetzen.

Literaturverzeichnis

- Allianz der deutschen Wissenschaftsorganisationen (2010): Grundsätze zum Umgang mit Forschungsdaten. <https://doi.org/10.2312/ALLIANZOA.019>.
- Beagrie, Charles (2011): User Guide for Keeping Research Data Safe. Assessing Costs/Benefits of Research Data Management, Preservation and Re-Use. Version 2.0. https://beagrie.com/KeepingResearchDataSafe_UserGuide_v2.pdf [Zugriff: 05.06.2018].
- Berliner Erklärung (2003): Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen. https://openaccess.mpg.de/68053/Berliner_Erklarung_dt_Version_07-2006.pdf [Zugriff: 05.06.2018].
- Corti, Louise/Van den Eynden, Veerle/Bishop, Libby/Woollard, Matthew (2014): Managing and Sharing Research Data. A Guide to Good Practice. London: Sage Publications.
- DataCite (2016): DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.0. <http://doi.org/10.5438/0012>.
- DFG – Deutsche Forschungsgemeinschaft (2015): Leitlinien zum Umgang mit Forschungsdaten. http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf [Zugriff: 05.06.2018].
- Doorn, Peter K. (2010): Preparing Data for Sharing. DANS Data Guide 8, 48. Den Haag: Amsterdam University Press. <https://dans.knaw.nl/nl/over/organisatie-beleid/publicaties/DANSpreparingdataforsharing.pdf> [Zugriff: 05.06.2018].
- DSGVO – Datenschutz-Grundverordnung (2016): Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung). <http://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32016R0679&from=DE> [Zugriff: 05.06.2018].
- Elliot, Mark/Mackey, Elaine/O'Hara, Kieron/Tudor, Caroline (2016): The Anonymization Decision-Making Framework. Manchester: University of Manchester. <http://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf> [Zugriff: 05.06.2018].
- Freese, Jeremy (2007): Replication Standards for Quantitative Social Science. *Sociological Methods & Research* 36, S. 153-172. <http://journals.sagepub.com/doi/10.1177/0049124107306659> [Zugriff: 05.06.2018].
- Gherghina, Sergiu/Katsanidou, Alexia (2013): Data Availability in Political Science Journals. *European Political Science* 12, S. 333-349.
- Horizon 2020 Programme – European Commission (2016): H2020 Programme Guidelines on FAIR Data Management in Horizon 2020. Version 3.0. 26 July 2016.

- http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf [Zugriff: 05.06.2018].
- Huschka, Denis/Oellers, Claudia (2013): Einführung. Warum qualitative Daten und ihre Sekundäranalyse wichtig sind. In: Huschka, Denis/Knoblach, Hubert/Oellers, Claudia/Solga, Heike (Hrsg.): Forschungsinfrastrukturen für die qualitative Sozialforschung. Berlin: Scivero, S. 9-18.
<https://www.ratswd.de/publikationen/forschungsinfrastrukturen-qualitative-sozialforschung> [Zugriff: 05.06.2018].
- Huschka, Denis/Oellers, Claudia/Ott, Notburga/Wagner, Gert G. (2011): Datenmanagement und Data-Sharing. Erfahrungen in den Sozial- und Wirtschaftswissenschaften.
https://www.ratswd.de/download/RatSWD_WP_2011/RatSWD_WP_184.pdf [Zugriff: 05.06.2018].
- Jensen, Uwe (2011): Datenmanagementpläne. In: Büttner, Stephan/Hobohm, Hans-Christoph/Müller, Lars (Hrsg.): Handbuch Forschungsdatenmanagement. Bad Honnef: Bock + Herchen, S. 71-82.
https://opus4.kobv.de/opus4-fhpotdam/files/197/2.3_DatenmanagementplAne.pdf [Zugriff: 05.06.2018].
- 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-fhpotdam/files/198/2.4_Metadaten_und_Standards.pdf [Zugriff: 05.06.2018].
- Jensen, Uwe (2012): Leitlinien zum Management von Forschungsdaten. Sozialwissenschaftliche Umfragedaten. GESIS-Technical Reports 2012/07. <https://www.ssoar.info/ssoar/handle/document/32065> [Zugriff: 05.06.2018].
- Kimpton, Michele/Minton Morris, Carol (2014): Managing and Archiving Research Data. Local Repositories and Cloud-Based Practices. In: Ray, Joyce M. (Hrsg.): Research Data Management. Practical Strategies for Information Professionals. West Lafayette: Purdue University Press, S. 223-238.
- King, Gary (1995): Replication, Replication. In: Political Science and Politics 28, 3, S. 443-499.
- Leibniz Gemeinschaft (2015): Leitlinie. Empfehlungen der Leibniz-Gemeinschaft zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit Vorwürfen wissenschaftlichen Fehlverhaltens.
<https://www.leibniz-gemeinschaft.de/forschung/gute-wissenschaftliche-praxis/> [Zugriff: 05.06.2018].
- Max Planck Gesellschaft (2000): Regeln zur Sicherung Guter Wissenschaftlicher Praxis. Verfahrensordnung bei Verdacht auf Wissenschaftliches Fehlverhalten. https://www.mpg.de/ueber_uns/verfahren [Zugriff: 05.06.2018].
- Politische Vierteljahresschrift (2011): Autorenhinweise 2011. www.kj.nomos.de/fileadmin/pvs/doc/PVS_Autorenhinweise_2011_deutsch.pdf [Zugriff: 05.06.2018].
- Recker, Astrid/Müller, Stefan/Trixa, Jessica/Schumann, Natascha (2015): Paving the Way For Data-Centric, Open Science. An Example From the Social Sciences. In: Journal of Librarianship and Scholarly Communication 3, 2, eP1227. <http://dx.doi.org/10.7710/2162-3309.1227>.
- Ulrich, Dagmar (2010): Bitstream Preservation. In: Neuroth, Heike/Oßwald, Achim/Scheffel, Regine/Strathmann, Stefan/Huth Karsten (Hrsg.): nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Kap. 8:3-8:8. Version 2.3. <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2010071949> [Zugriff: 05.06.2018].
- Vlaeminck, Sven/Siegert, Olaf (2012): Welche Rolle spielen Forschungsdaten eigentlich für Fachzeitschriften? Eine Analyse mit Fokus auf die Wirtschaftswissenschaften. Working Paper Series des Rates für Sozial- und Wirtschaftsdaten 210. <https://hdl.handle.net/11108/71> [Zugriff: 05.06.2018].
- Wallace, David A (2001): Archiving Metadata Forum. Report from the Recordkeeping Metadata Working Meeting. June 2000. In: Archival Science 1, S. 253-269.
- Watteler, Oliver (2010): Erstellung von Methodenberichten für die Archivierung von Forschungsdaten.
https://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Aufbau_Methodenbericht_v1_2010-07.pdf [Zugriff: 05.06.2018].
- Wilkinson, Mark D. et al. (2016): The FAIR Guiding Principles for Scientific Data Management and Stewardship. In: Scientific Data 3, 160018. <https://dx.doi.org/10.1038/sdata.2016.18>.
- ZBW - Leibniz-Informationszentrum 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

- Dataverse Online (o.J.): <https://dataverse.org/best-practices/replication-dataset> [Zugriff: 27.06.2018].
- datorium (o.J.a): <https://datorium.gesis.org/xmlui/> [Zugriff: 27.06.2018].
- datorium (o.J.b): Empfehlungen zur Aufbereitung. <https://datorium.gesis.org/xmlui/?publication> [Zugriff: 27.06.2018].
- DDI Alliance (2017): Data Documentation Initiative: <http://www.ddialliance.org/> [Zugriff: 27.06.2018].
- DOI-System (2018): Digital Object Identifier: <https://www.doi.org/> [Zugriff: 27.06.2018].
- Force11 (2017): The FAIR Data Principles: <https://www.force11.org/group/fairgroup/fairprinciples> [Zugriff: 27.06.2018].
- GESIS – Leibniz-Institut für Sozialwissenschaften (2018a): Datenarchivierung: <https://www.gesis.org/angebot/archivieren-und-registrieren/datenarchivierung> [Zugriff: 27.06.2018].
- GESIS – Leibniz-Institut für Sozialwissenschaften (2018b): Datenservices: <https://www.gesis.org/datenservices>.
- ISCED (2011): International Standard Classification of Education. UNESCO Institute for Statistics: <http://uis.unesco.org/en/topic/international-standard-classification-education-isced> [Zugriff: 27.06.2018].

