

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

1. Einleitung

Uwe Jensen, Sebastian Netscher und Katrin Weller

Forschungsdatenmanagement gewinnt in den Sozialwissenschaften in den letzten Jahren zunehmend an Bedeutung. Hier zunächst verstanden als Sammelbegriff für alle Aktivitäten und Maßnahmen im Umgang mit Forschungsdaten, hat sich das Forschungsdatenmanagement als zentraler Bestandteil guter wissenschaftlicher Praxis in empirischen Forschungsvorhaben etabliert. Forschungsdatenmanagement erzeugt die notwendige Transparenz in der Erstellung der Forschungsdaten, gewährleistet deren Nachvollziehbarkeit und ermöglicht so die Replikation von Forschungsergebnissen. Die Forderung nach der Bereitstellung von Forschungsdaten zur weiteren Nutzung im Sinne von *Open Science* bzw. *Open Data* durch Wissenschaftsorganisationen und Forschungsförderer unterstreicht die Relevanz eines systematischen Umgangs mit Forschungsdaten und die Anforderungen an ein nachhaltiges Forschungsdatenmanagement.

Um Forschenden in den Sozialwissenschaften den Umgang mit quantitativen Forschungsdaten zu erleichtern, nimmt der vorliegende Sammelband praktische Herausforderungen des Forschungsdatenmanagements in den Blick. Mit Hilfe von Anwendungsfällen aus der empirischen Sozialforschung vermittelt das Buch Grundlagen des Forschungsdatenmanagements und bietet anschauliche Beispiele für die Praxis. Es vermittelt Lösungsansätze und erlaubt, diese unmittelbar in die tägliche Arbeit zu übertragen. Dabei bezieht der Sammelband Empfehlungen, Richtlinien und Werkzeuge zum Forschungsdatenmanagement systematisch ein und erörtert spezifische Gegebenheiten der deutschen Forschungslandschaft. Die Beiträge sind dabei von Publikationen zu den Methoden der empirischen Sozialforschung und entsprechenden Forschungsaktivitäten abzugrenzen. Stattdessen fokussiert das Buch die Anwendung von Maßnahmen und Verfahren zur Erstellung, Nutzung und Sicherung sozialwissenschaftlicher Forschungsdaten.

Die einzelnen Kapitel des Sammelbandes wurden von (ehemaligen) Mitarbeitenden des Datenarchivs für Sozialwissenschaften bei GESIS – Leibniz-Institut für Sozialwissenschaften verfasst. Diese besitzen langjährige Praxiserfahrung im Umgang mit Forschungsdaten und (inter-)nationalen Forschungsvorhaben. Die Autorinnen und Autoren richten sich mit dem Buch zum einen an Forschende in den Sozialwissenschaften und verwandten Bereichen, die quantitative Umfragedaten erheben, aufbereiten oder analysieren. Als Leitfaden soll der Sammelband diese mit den notwendigen Kenntnissen im Umgang mit Forschungsdaten ausrüsten und sie in der Planung und Umsetzung ihres eigenen Forschungsdatenmanagements unterstützen. Zur Zielgruppe des Buchs zählen zum anderen Mitarbeitende in Forschungsprojekten und Personen, die beruflich mit entsprechenden Daten umgehen, wie z.B. Archivar/innen oder Projekt- bzw. Drittmittelmanager/innen. Diese Zielgruppe soll das Buch dabei unterstützen, ihr vorhandenes Wissen in der Handhabung von Forschungsdaten zu erweitern und zu vertiefen.

Der Sammelband behandelt das Thema Forschungsdatenmanagement aus drei Perspektiven. Der erste Teil liefert eine generelle Einführung in das Thema und erörtert entsprechende Aktivitäten und Maßnahmen im Rahmen eines empirischen Forschungsprojekts. Im Zentrum des zweiten Teils steht die spezielle Rolle von Forschungsdateninfrastrukturen bei der Etablierung und Fortentwicklung von Metadatenstandards ebenso wie von Standards zu Datenzitation. Der Sammelband schließt im dritten Teil mit einem Ausblick auf neue Heraus-

forderungen des Forschungsdatenmanagements am Beispiel neuer Datentypen bzw. neuer Datenquellen, wie Social-Media-Daten oder georeferenzierte Umfragedaten.

Der erste Teil des Buches umfasst die Kapitel 2 bis 8. Kapitel 2 führt zunächst allgemein in die Themen *Forschungsdaten und Forschungsdatenmanagement in den Sozialwissenschaften* ein. Es schafft ein grundlegendes Verständnis für zentrale Begriffe des Buches und beschreibt die besonderen Eigenarten sozialwissenschaftlicher Umfragedaten. Daran anschließend wird der Umgang mit diesen Daten im Kontext von Lebenszyklusmodellen anhand des *DDI Data Lifecycles* von der Planung bis zur Nachnutzung erörtert. Darauf aufbauend wird das Forschungsdatenmanagement in den Sozialwissenschaften im Kontext des Lebenszyklus von Daten, der Rolle von Forschungsdateninfrastrukturen und der neuen Herausforderungen beim Management von Daten diskutiert.

Kapitel 3 thematisiert, wie Forschende ihr *Forschungsdatenmanagement systematisch planen und umsetzen* können. Ein systematisches Forschungsdatenmanagement beginnt mit einer vorausschauenden Vorbereitung im Rahmen der eigentlichen Projektplanung. Dabei müssen Forschende zunächst überlegen, welche Ziele sie mit welchen Maßnahmen des Forschungsdatenmanagements verbinden. So muss jede empirische Forschung durch ein projektinternes Forschungsdatenmanagement zuallererst sicherstellen, dass die eigentlichen Forschungsziele erreicht werden können. Im Sinne guter wissenschaftlicher Praxis sollten die im Forschungsprojekt erstellten Daten aber über das Forschungsprojekt hinaus erhalten bleiben und so eine längerfristige Replikation der Forschungsergebnisse bzw. eine weitere Nutzung durch andere Forschende in neuen Forschungskontexten ermöglichen.

Die folgenden Kapitel 4 bis 8 adressieren spezifische Themenkomplexe des Forschungsdatenmanagements. Kapitel 4 beginnt mit dem *Datenschutz im Forschungsdatenmanagement*. Die empirische Sozialforschung greift zumeist auf personenbezogene und eventuell sensible Daten, z.B. von Studienteilnehmer/innen, zurück. Die Erhebung, Verarbeitung und Nutzung derartiger Informationen unterliegen datenschutzrechtlichen und forschungsethischen Bestimmungen. Das Kapitel zeigt grundlegende Regelungen im Bereich des aktuellen Datenschutzrechts auf und bietet Anleitungen, wie diese Regelungen in der Praxis umgesetzt werden können. Dazu erörtert das Kapitel neben forschungsethischen und datenschutzrechtlichen Aspekte vor allem das Thema Anonymisierung von quantitativen Daten. Es beleuchtet anhand von Fallbeispielen gängige Verfahren ebenso wie häufig wiederkehrende Fehler bei der Umsetzung datenschutzrechtlicher Bestimmungen.

Kapitel 5 befasst sich mit der *Dateiorganisation in empirischen Forschungsprojekten*. Der Erfolg von empirischen Forschungsprojekten fußt vor allem darauf, dass die im Projekt generierten oder weiterverwendeten Informationen zum richtigen Zeitpunkt für den richtigen Personenkreis auffindbar und zugänglich sind. Das gilt insbesondere für Daten und ihre Dokumentation, aber auch für Dokumente, die Prozesse und Entscheidungen transparent und replizierbar machen. Das Kapitel zeigt anhand von Beispielen erstens, wie Informationsflüsse im Projekt mit Hilfe eines Modells zur Beschreibung des Lebenszyklus von Forschungsdaten effizient geplant und gesteuert werden können. Zweitens beschreibt es, wie Dateiverzeichnisse und Dateien so benannt und strukturiert werden können, dass sie die Auffindbarkeit, Authentizität und Integrität aller im Projekt anfallenden Informationen unterstützen. Zusammengefasst helfen diese Maßnahmen, Informationsverluste zu vermeiden und den Forschungsprozess mit allen getroffenen Entscheidungen, die Auswirkungen auf die Forschungsergebnisse haben, nachvollziehbar zu machen.

Die *Datenaufbereitung und Dokumentation* wird in Kapitel 6 thematisiert. Als Phase im Lebenszyklus der Forschungsdaten hat sie die Aufgabe, die Daten für die Forschung nutzbar zu machen. Das stellt Forschungsprojekte vor die Herausforderung, einen Workflow zu entwickeln, der die Projektziele in unmittelbare Arbeitsschritte übersetzt und diese für alle Beteiligten transparent und verständlich macht. Ausgehend hiervon fokussiert das Kapitel die

Planung und Organisation der Datenaufbereitung im Projektverlauf. Sie soll zu einem möglichst effizienten Workflow führen, der auf eine hohe Qualität und umfassende Dokumentation der Daten gerichtet ist.

Kapitel 7, *Data Sharing: Von der Sicherung zur langfristigen Nutzung der Forschungsdaten*, greift die abschließende Publikation von Daten als wichtiges Ergebnis von Forschungsvorhaben auf. Die dauerhafte Verfügbarkeit von Forschungsdaten – und damit auch ihre Archivierung als notwendige Voraussetzung – leistet einen wesentlichen Beitrag zu Open Science: Sie macht Ergebnisse empirischer Forschung nicht nur nachvollziehbar und replizierbar, sondern auch anschlussfähig. Das Kapitel liefert einen Überblick über verschiedene Möglichkeiten, Forschungsdaten nach Projektende zu sichern und Dritten zur Nachnutzung zur Verfügung zu stellen.

Unter dem Titel *Forschungsdatenmanagement in der Sekundäranalyse* schlägt Kapitel 8 die Brücke zwischen dem Forschungsdatenmanagement datengenerierender Projekte und der Nachnutzung von Forschungsdaten im Rahmen neuer Forschungsvorhaben. Es beschreibt im Sinne guter wissenschaftlicher Praxis die Notwendigkeit replizierbarer Forschungsergebnisse bei der Nachnutzung bereits existierender Daten. Der Idee von Open Science folgend, können darüber hinaus im Forschungsprojekt erstellte Skripte, etwa in Form von Datenharmonisierungskonzepten, entwickelt werden, die ihrerseits für Dritte zur Nachnutzung interessant sein können und folglich verfügbar gemacht werden sollten. Das Kapitel zeigt anhand eines Beispiels die relevanten Schritte auf, um die Replizierbarkeit und Nachnutzbarkeit verwendeter Forschungsdaten und Aufbereitungskonzepte sicherzustellen. Es liefert somit auch einen Beitrag zur Fortentwicklung des Forschungsdatenmanagements in der Sekundäranalyse.

Der zweite Teil des Sammelbandes umfasst die Kapitel 9 und 10 und ist den spezifischen Themen Metadaten und Datenzitation im Forschungsdatenmanagement gewidmet. Die Kapitel thematisieren Dienstleistungen und technische Entwicklungen von sozialwissenschaftlichen Dateninfrastrukturen, die den konkreten Projektalltag mittelbar oder unmittelbar unterstützen. Die beiden Beiträge richten sich an Interessierte, die diese Themen bei ihrer Arbeit mit Forschungsdaten vertiefen und praktische Nutzungsmöglichkeiten besser verstehen wollen.

Kapitel 9 behandelt dazu zunächst *Metadatenstandards im Kontext sozialwissenschaftlicher Daten* und deren Rolle für die Dokumentation und Erschließung von Forschungsdaten. Es stellt gängige Metadatenstandards vor und beschreibt u.a. ihre Relevanz für Forschende beim Auffinden von Studien in nationalen wie internationalen Datenkatalogen ebenso wie bei der Dokumentation von Variablen und Fragen in sozialwissenschaftlichen Datensätzen. Darüber hinaus erörtert der Beitrag Standards und Tools zur Bearbeitung von Metadaten und stellt insbesondere den Metadatenstandard der Data Documentation Initiative (DDI) sowie seine Möglichkeiten zur Dokumentation von Forschungsdaten vor.

Kapitel 10 thematisiert anschließend die Aspekte der *Zitierbarmachung und Zitation von Forschungsdaten*. Nicht nur Forschungspublikationen, sondern auch die im Forschungsprozess entstandenen Daten sollen gemäß anerkannter Grundsätze zum Umgang mit Forschungsdaten zitierbar sein und zitiert werden. Die Datenzitation fördert die Anerkennung der Produktion von Forschungsdaten als eine primäre Wissenschaftsleistung. Das Kapitel erörtert die Möglichkeiten und Herausforderungen bei der Zitierbarmachung von Forschungsdaten. Es werden vorhandene Services zur Unterstützung der Datenproduzierenden beschrieben und anhand eines ausgewählten Dienstes – des DOI[®]-Systems – konkrete Empfehlungen ausgesprochen.

Der letzte Teil des Buches befasst sich schließlich in den Kapiteln 11 und 12 mit neuen Herausforderungen und Entwicklungen im Forschungsdatenmanagement am Beispiel neuer Datentypen bzw. neuer Datenquellen. Kapitel 11 gibt einen Ausblick auf die Heraus-

forderungen im Umgang mit Social Media-Inhalten als neue Art von Forschungsdaten. Die Nutzung neuartiger, großer Datenbestände etwa aus Suchmaschinen oder Social-Networking-Plattformen wird seit einiger Zeit als Grundlage zum besseren Verständnis zahlreicher Lebensbereiche systematisch erprobt. Die unterschiedlichen Herausforderungen beim Umgang mit diesen als *New Data* oder *Big Data* bezeichneten Datenquellen stehen im Mittelpunkt des Beitrags. Ausgehend von Social-Media-Daten als Forschungsgrundlage werden aktuelle Möglichkeiten und Grenzen bei der Sammlung, der Qualitätssicherung sowie die Archivierung und Nachnutzung dieses Datentyps diskutiert.

Anhand der *räumlichen Verknüpfung georeferenzierter Umfragedaten mit Geodaten* wird in dem abschließenden Kapitel 12 die Verbindung verschiedener Datenquellen für die integrierte Analyse exemplarisch vorgestellt. Ausgehend vom inhaltlichen Mehrwert einer räumlichen Verknüpfung diskutiert das Kapitel technische, rechtliche und dokumentarische Herausforderungen ebenso wie praktische Empfehlungen zur Datenorganisation.

Insgesamt hoffen wir, mit diesem Sammelband Forschenden, Lehrenden und Personen, die im beruflichen Alltag mit quantitativen sozialwissenschaftlichen Daten arbeiten, einen praktischen Leitfaden an die Hand zu geben, der es ihnen ermöglicht, die unterschiedlichen Facetten des Forschungsdatenmanagements besser zu verstehen und den einen oder anderen Lösungsansatz in ihre tägliche (Forschungs-)Arbeit erfolgreich zu integrieren. Gleichzeitig hoffen wir als Autorinnen und Autoren auf vielfältige kritische Rückmeldungen und Verbesserungsvorschläge, die dazu anregen, Grundsätze und Lösungen zum Forschungsdatenmanagement sozialwissenschaftlicher Daten in Theorie und Praxis weiterzuentwickeln.

Danksagung

Wir möchten uns an dieser Stelle recht herzlich bei den Autorinnen und Autoren dieses Sammelbandes für ihre Beiträge, ihren fortwährenden Einsatz und die konstruktive Zusammenarbeit bei der Entwicklung erster Ideen bis hin zur Fertigstellung des Buches bedanken. Unser Dank gilt all jenen Kolleginnen und Kollegen, die uns und den Autorinnen und Autoren den erforderlichen Freiraum für dieses Projekt einräumten und uns fachlich mit Rat zur Seite standen. Ebenso bedanken wir uns für die Unterstützung durch GESIS – Leibniz-Institut für Sozialwissenschaften und die Möglichkeit, dieses Buch sowohl als Printausgabe als auch Online unter einer Creative-Commons-Lizenz zu publizieren. Bei der redaktionellen Bearbeitung unterstützten uns unsere studentischen Mitarbeiterinnen Livia Leonhardt und Rabea Lukies, denen wir für ihre kompetente und sorgfältige Arbeit danken. Nicht zuletzt gilt unser herzlicher Dank dem Verlag Barbara Budrich und seinen Mitarbeitenden. Frau Lotz, Frau Blinkert und Frau Budrich haben unsere Nachfragen in allen Phasen des Buchprojektes geduldig und gelassen beantwortet und sind stets flexibel auf unsere Gestaltungswünsche zur Publikation des Buches eingegangen.