

Die Kuratierung sozialwissenschaftlicher Forschungsdaten – Praxisfragen und Beispiellösungen

Patrick Droß, Julian Naujoks

Wissenschaftszentrum Berlin für Sozialforschung (WZB)

Zusammenfassung

Der Beitrag bietet einen Blick in die Praxis der Veröffentlichung sozialwissenschaftlicher Forschungsdaten. Dabei beschreibt er den Arbeitsschritt der Datenkuratierung als wesentliche Komponente, um Forschungsdaten nachhaltig verfügbar zu machen. Anhand der Erfahrungen des institutionellen Forschungsdatenmanagements am Wissenschaftszentrum Berlin für Sozialforschung (WZB) werden einzelne Problemfelder und Praxislösungen dargestellt, die im Laufe des Kuratierungsprozesses auftreten. Dabei zeigt sich, dass die Datenkuratierung ein komplexer Vorgang ist, bei dem sich fachspezifische, aber auch (arbeits-)organisatorische Herausforderungen stellen. Diese werden anhand von drei exemplarischen Themenkomplexen skizziert: Datenschutzfragen müssen im Forschungsprozess frühzeitig mitgedacht werden. Dazu gehört, dass in informierten Einwilligungen die Rechte der Befragten geschützt, gleichzeitig aber spätere Nachnutzungsmöglichkeiten nicht durch zu restriktive Formulierungen ausgeschlossen werden. In Bezug auf die Verwertungs- und Nutzungsrechte muss gerade in Projektkonstellationen mit mehreren Beteiligten frühzeitig eine Abstimmung über die Datenveröffentlichung erfolgen. Schließlich verdeutlicht die Verknüpfung von Daten- und Textpublikation, wie sich Veröffentlichungsroutinen und Suchgewohnheiten der Wissenschaftler*innen auf die Auffindbarkeit der Forschungsdaten auswirken.

Abstract

This article offers an insight into the practice of publishing social science research data. It describes the step of data curation as an essential component to make research data sustainably available. Based on the experiences of institutional research data management at the WZB Berlin Social Science Center, particular problem areas and practical solutions that arise during the curation process are presented. This shows that data curation is a complex process in which both subject-specific and (work) organisational challenges arise. These challenges are outlined using three exemplary

topics: Data protection issues must be considered at an early stage in the research process. This includes protecting the rights of the respondents in informed consents, but at the same time not excluding the possibility of subsequent use by overly restrictive wording. With regard to the rights of exploitation and use, especially in project constellations with several participants, the publication of data must be coordinated at an early stage. Finally, the linking of data and text publication illustrates how publication routines and search habits of scientists affect the findability of research data.

1. Einleitung: Datenqualität, Kuratierung und sozialwissenschaftliche Forschungsdaten

Der Rat für Informationsinfrastrukturen (RfII) fordert in einem jüngst veröffentlichten Strategiepapier zum Thema Open Data, den Fokus vermehrt auf die Qualität der veröffentlichten Forschungsdaten zu legen. So soll die Entwicklung „anstelle eines quantitativen Wachstums vor allem das qualitative Ziel hochwertiger Datenbestände und Datendienste verfolgen“ (RfII 2019: 2). „Open“, so die Autor*innen, werde noch zu oft mit „online“ gleichgesetzt, ohne dass von einer langfristigen Nachnutzbarkeit auszugehen sei (ebd.: 5). Der Mehrwert, den die Nachnutzbarkeit der Forschungsdaten bieten könne, wäre jedoch gerade dann gewährleistet, wenn ihre Verfügbarmachung mit einer Qualitätssicherung entlang fachlicher Standards verbunden werde.

In der Open-Data-Praxis gewinnt aufgrund dieser vermuteten Abhängigkeit zwischen der tatsächlichen Nachnutzbarkeit von Forschungsdaten sowie Art und Umfang der Qualitätssicherung das Aufgabengebiet der Datenkuratierung an Bedeutung. Doch was meint die Tätigkeit der „Kuratierung“ in Bezug auf Forschungsdaten eigentlich genau? Büttner et al. rekurrieren auf die alltagsgebräuchliche Verwendung des Begriffs der Kuratierung (z.B. in der Konservierung und Ausstellung von Museumsobjekten), um damit das Vorhaben zu charakterisieren, Wissen für zukünftige Forschung zu erhalten und gleichzeitig zu schaffen (Büttner et al. 2011: 17). Ulrich Herb folgend liegt die Aufgabe der Datenkuratierung darin, die „inhaltlich-logische oder wissenschaftliche Nutzbarkeit“ (Herb 2015: 134) der Forschungsdaten zu sichern. Der Vorgang der Kuratierung von Forschungsdaten wird demnach als

Prozess verstanden, der Forschungsdaten aufbereitet, sie verfügbar und vor allem qualitätsgesichert nachnutzbar macht.

Der vorliegende Beitrag nimmt die fachspezifischen Anforderungen der Kuratierung von sozialwissenschaftlichen Forschungsdaten in den Blick (vgl. Droß et al 2017). Disziplinär sind die Sozialwissenschaften durch eine erhebliche Vielfalt in den Wegen der Datengenerierung zur Beantwortung der Forschungsfragen gekennzeichnet. So werden in klassischen quantitativen Surveys unter Verwendung standardisierter Erhebungsinstrumente Mikrodaten erhoben. In experimentellen Laborsettings wird der Einfluss gezielter Stimuli auf Proband*innen untersucht oder in Feldexperimenten das Verhalten ganzer Untersuchungsgruppen in den Blick genommen. Vielfach werden jedoch auch prozessproduzierte bzw. Sekundärdaten aufbereitet (z.B. Text und Data Mining), weiterverarbeitet und mit neuen Informationen angereichert. Schließlich umfasst die qualitative Forschung ein ganzes Spektrum unterschiedlicher Erhebungsmethoden: Dies reicht von klassischen Expert*inneninterviews über ethnografische Forschungsansätze (teilnehmende Beobachtung, Feldnotizen, Fotografie), der Codierung von Sekundärtexten (Zeitungsartikel, Parteiprogramme) bis hin zu Videoaufzeichnungen.

2. Die Kuratierung von Forschungsdaten am Wissenschaftszentrum Berlin für Sozialforschung (WZB)

Das WZB ist Mitglied der Leibniz-Gemeinschaft und betreibt sozialwissenschaftliche Grundlagenforschung. Das empirisch und interdisziplinär ausgerichtete Forschungsprofil des WZB steht exemplarisch für die erwähnte methodische Vielfalt der Datenerhebung und -generierung in den Sozialwissenschaften. Den sowohl quantitativ als auch qualitativ forschenden Wissenschaftler*innen am Institut steht dabei eine forschungsnahe und bedarfsgerechte Unterstützung durch die Infrastrukturabteilung „Wissenschaftliche Information“ zur Seite, in der die Bereiche Bibliothek, Forschungsdatenmanagement, Institutsarchiv sowie Open Access zusammengefasst sind.

Seit 2017 existiert am WZB eine Leitlinie zum Umgang mit Forschungsdaten.¹ Das WZB fordert darin seine Forscher*innen nicht nur zu einem nachhaltigen Umgang mit den von ihnen produzierten und verwendeten Forschungsdaten auf. Es empfiehlt ausdrücklich die Veröffentlichung von am WZB erstellten Forschungsdaten auf einem geeigneten Repository. Hier weist die Leitlinie dem institutionellen Forschungsdatenmanagement eine besondere Rolle zu: Es soll die Forscher*innen bei der Veröffentlichung aktiv unterstützen und die Qualitätssicherung nach fachlichen Standards maßgeblich gewährleisten. Die Intention ist hierbei, dass ein zentral und institutionell angebotener Service der Kuratierung die Verfügbarkeit von WZB-Forschungsdaten als Ganzes fördert und diese darüber hinaus den oben genannten Qualitätsansprüchen genügen.

Die Kuratierung von sozialwissenschaftlichen Daten ist ein äußerst komplexer Vorgang, bei dem diverse praktische, organisatorische und rechtliche Aspekte zu beachten sind (u.a. Datenschutz der Proband*innen, die Rechte von Dritten, etc.). Ein zentraler Kuratierungsservice ist also aus institutioneller Perspektive sinnvoll: In Einzelfällen erlangte Expertise wird gebündelt und kann in verschiedenen Beratungsangeboten für weitere Nutzer*innen zur Verfügung gestellt werden. Zudem muss immer auch zwischen Aufwand und Nutzen der Datenaufbereitung abgewogen werden, bspw. wenn durch Anonymisierungsmaßnahmen Kontextinformationen entfernt werden (vgl. DGS 2019: 5f.). Zwar sind Forscher*innen die Expert*innen ihrer Daten, dennoch fehlt es ihnen häufig an zeitlichen und personellen Ressourcen, die Daten für eine Veröffentlichung vorzubereiten. Der institutionelle Service der Datenkuratierung entlastet die Forscher*innen somit auch erheblich in ihrer Arbeit.

Fachlich und personell ist das Forschungsdatenmanagement am WZB für die Kuratierung der Daten gut aufgestellt: Zu dem Bereich gehören ehemalige Wissenschaftler*innen, der Datenschutzbeauftragte des WZB sowie Fachangestellte für Markt- und Sozialforschung (FAMS), die die Abteilung auch ausbildet. Gleichzeitig war das WZB gemeinsam mit dem GESIS-Leibniz-Institut für Sozialwissenschaften, dem ZBW-Leibniz-Informationszentrum Wirtschaft und dem Deutschen Institut für Wirtschaftsforschung (DIW) Teil des SowiDataNet-Projektconsortiums, welches ein Repository für Forschungsdaten aus den Sozial- und Wirtschaftswissenschaften

¹ Vgl. „Leitlinie zum Umgang mit Forschungsdaten am Wissenschaftszentrum Berlin für Sozialforschung“. Online verfügbar unter: www.wzb.eu/system/files/docs/sv/win/forschungsdaten_leitlinien.de_.pdf (zuletzt geprüft am 11.04.2019).

entwickelt hat (vgl. Droß/Linne 2016). Zurzeit nutzt das WZB für die Veröffentlichung seiner Forschungsdaten SowiDataNet in der Betaphase.

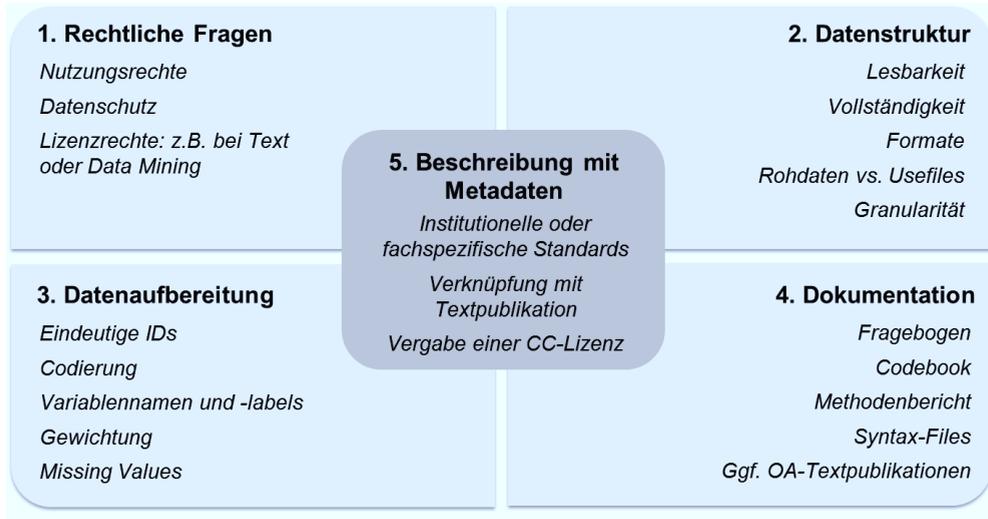


Abbildung 1: Praktische Schritte der Datenkuratierung am WZB

Der gegenwärtige Ablauf des zentralen Kuratierungsservices am WZB lässt sich grob in drei Schritten beschreiben: (1) Zu Beginn steht die Kontaktaufnahme, die beidseitig sowohl durch Forscher*innen aber auch proaktiv durch Mitarbeiter*innen des Forschungsdatenmanagements erfolgt. In einem Beratungsgespräch werden zunächst der Stand der Datenaufbereitung, rechtliche Fragen und selbstverständlich auch die prinzipielle Eignung und Nachnutzbarkeit der Forschungsdaten geklärt. (2) Anschließend stellen die Wissenschaftler*innen dem Forschungsdatenmanagement ihre Daten zur Verfügung. Hier beginnt der eigentliche Prozess der Kuratierung: Die Mitarbeiter*innen des Forschungsdatenmanagements prüfen Datenqualität, -struktur und -aufbereitung (siehe Abbildung 1), sichten die Begleitdokumentation und übernehmen die Metadaten-Beschreibung. Änderungen an den Daten erfolgen möglichst durch die Forscher*innen selbst oder nach Rücksprache mit den Forscher*innen durch das Forschungsdatenmanagement und werden nachvollziehbar dokumentiert. (3) Nach einer finalen Zustimmung durch die Forscher*innen werden die Forschungsdaten schließlich inkl. DOI-Registrierung über da|ra im SowiDataNet-Repository veröffentlicht.

3. Praxisfragen und Beispiellösungen

Am WZB ist – wie vermutlich in der gesamten Forschungsdatenlandschaft – die Implementierung von Best Practices des Forschungsdatenmanagements noch nicht abgeschlossen. Im Gegenteil: Im Rahmen der Kuratierung der Forschungsdaten stoßen wir immer wieder auf praktische Probleme, Fragen und komplexe Einzelfälle, für die noch keine Standards der Bearbeitung vorliegen, für die aber gleichwohl ein Umgang gefunden werden muss. Drei solcher „Problembereiche“ mitsamt den praktischen Lösungsansätzen aus unserer alltäglichen Arbeit wollen wir im Folgenden vorstellen. Die Darstellungen haben dabei nicht den Anspruch abschließender Empfehlungen, sondern eher den Charakter eines Werkstattberichts. Keinesfalls ersetzen sie die Rechtsberatung im Einzelfall. Wir wollen mit ihnen vielmehr zu einer Diskussion beitragen und anregen, in der jenseits der allgemeinen Empfehlungen, inzwischen zahlreich vorliegenden Guidelines und Forschungsdaten-Richtlinien verstärkt die Praxisfragen der Datenkuratierung thematisiert werden, die uns Praktiker*innen in zunehmendem Maße beschäftigen.

3.1 Datenschutz: Die Bedeutung der informierten Einwilligung

In einer internen Befragung am WZB gaben 2015 rund 66 Prozent der Forscher*innen an, dass in ihren aktuellen Projekten neue Daten generiert werden. Da es sich im sozialwissenschaftlichen Kontext zumeist um Daten mit Personenbezug handelt, unterliegt der gesamte Verarbeitungsprozess dieser Forschungsdaten datenschutzrechtlichen Vorschriften. Diese wurden bis Mai 2018 durch das alte Bundesdatenschutzgesetz (BDSG) definiert und werden seither durch die Europäische Datenschutz-Grundverordnung (DSGVO) und das BDSG-neu (2018), ggf. auch durch die jeweiligen Landesdatenschutzgesetze geregelt.²

Wurden Datenschutz-Vorschriften in der Vergangenheit im Einzelfall nicht immer ganz exakt umgesetzt, stellte dies zumeist kein gravierendes Problem dar. Die Daten wurden in erster Linie ohnehin nur in der Forschung genutzt und, wenn überhaupt, nur

² Die Komplexität des Themas „Forschungsdatenmanagement und Datenschutz“ kann im Rahmen dieses Beitrags nicht im Detail behandelt werden. Einen hervorragenden und v.a. aktuellen Überblick bieten Watteler/Ebel 2019.

in einem engen Forschungskontext unter Kolleg*innen geteilt. Ein Zugriff durch unbefugte Dritte war damit weitgehend ausgeschlossen. Mit der weltweiten Verfügbarkeit von in Repositorien veröffentlichten Datensätzen wird der datenschutzkonforme Umgang mit Forschungsdaten für Sozialwissenschaftler*innen jedoch wichtiger denn je. Der Aufbau von Datenschutz-Expertise im Forschungsdatenmanagement ist daher unerlässlich. Doch auch für die Forschung ist die Sensibilisierung notwendig, denn wie im Forschungsdatenmanagement insgesamt gilt: Je früher im Forschungsprozess über den Datenschutz nachgedacht wird, je einfacher ist die Datenveröffentlichung.

Die datenschutzrechtlichen Vorgaben der DSGVO sind im Grundsatz nicht neu. Maßgeblich bleibt, dass für die Verarbeitung personenbezogener Befragungsdaten in sozialwissenschaftlichen Projekten in den allermeisten Fällen eine informierte Einwilligung (informed consent) der Untersuchungspersonen eingeholt werden muss. In dieser stimmen die Befragten auf Basis ausreichender Angaben zum Forschungsvorhaben, der Erhebung und Verarbeitung ihrer Daten zu. Dabei sind nun jedoch auch die neuen Regelungen der DSGVO hinsichtlich der Informationspflichten zu beachten. So muss bspw. über die Betroffenenrechte, die Rechtsgrundlage und eine Kontaktmöglichkeit zum Datenschutzbeauftragten aufgeklärt werden. Insbesondere für die Betroffenenrechte enthalten die neuen gesetzlichen Vorschriften vielfältige Ausnahmeregelungen für die Forschung, die bislang jedoch nicht systematisch aufbereitet wurden. Als genereller Trend lässt sich zudem beobachten, dass im Zuge der Einführung der DSGVO das Thema Datenschutz eine deutlich größere Aufmerksamkeit seitens der Betroffenen, aber auch der Aufsichtsbehörden erfahren hat. Für Wissenschaftseinrichtungen ergibt sich daher die Notwendigkeit, sich verstärkt mit dem Thema Datenschutz zu beschäftigen – nicht zuletzt, weil das Fehlverhalten einzelner Forscher*innen in Form von Bußgeldern oder Reputationsschäden auf die gesamte Einrichtung zurückfallen kann.

Bei der Einholung des datenschutzrechtlichen informed consent bestehen in der Praxis nach unseren Erfahrungen erhebliche Unsicherheiten. Bei der Formulierung kommt es häufig zu zwei typischen Fehlern, die später zu unnötigen Barrieren für eine Nutzbarkeit der Forschungsdaten werden können: 1) Erklärungen sind zu knapp gehalten und genügen nicht den formalen gesetzlichen Anforderungen.

Hier besteht die Gefahr, dass die Gültigkeit der Einwilligung im Nachhinein angezweifelt werden kann. Eine Datenveröffentlichung wäre in diesen Fällen stets mit gewissen Unsicherheiten verbunden. 2) Erklärungen sind formal korrekt, es werden aber überzogene Einschränkungen aufgenommen: So ist bspw. die Zusage, dass „*sämtliche Daten* zum Ende des Projektes gelöscht werden“ oder, dass „*die Daten* keinesfalls an Dritte weitergeben werden“ für eine geplante Datenveröffentlichung problematisch, da sie eine Nachnutzung pauschal ausschließen. Hier steckt der Teufel quasi im Formulierungsdetail und es ist stets eine Differenzierung zwischen „den Forschungsdaten“ und „den personenbezogenen Merkmalen“ notwendig. Letztere sind zu löschen, wenn der Forschungszweck erreicht ist und dies sollte den Untersuchungsteilnehmer*innen auch zugesichert werden. Die durch die Löschung personenbezogener Identifikatoren anonymisierten Daten können jedoch zur Nachnutzung veröffentlicht werden. Idealerweise wird daher im Rahmen der Formulierung eines informed consent ein dreistufiges Vorgehen beschrieben, welches sowohl den Proband*innen Datenschutzkonformität garantiert, gleichzeitig aber auch die Veröffentlichung der Daten ermöglicht:

- 1) Die Trennung der Forschungsdaten von den personenbezogenen Daten während des Forschungsprozesses (Pseudonymisierung).
- 2) Die Löschung personenbezogener Merkmale spätestens mit Abschluss des Forschungsprozesses (also eine Anonymisierung der Daten) und
- 3) der Hinweis, dass anonymisierte Forschungsdaten archiviert und auch nach Abschluss des Forschungsvorhabens zur Nachnutzung bereitgestellt werden können. Dies sorgt für umfängliche Transparenz und rechtliche Absicherung.

Wie eng informierte Einwilligung und die spätere Nachnutzbarkeit der Forschungsdaten in Zusammenhang stehen, verdeutlicht auch ein weiteres Praxisproblem: Häufig wird den Untersuchungsteilnehmer*innen zugesichert, dass die Ergebnisse der Studie nur für wissenschaftliche Zwecke verwendet werden. Diese Einschränkung der Nachnutzung ist dabei vielfach auch ein expliziter Wunsch der Forscher*innen selbst. Denn: Aus ihrer Sicht besteht die Gefahr, „dass die Teilnahmebereitschaft an Umfragen/wissenschaftlichen Surveys und das Vertrauen von Versuchspersonen sich verändern, wenn Daten nicht allein wissenschaftlich genutzt werden“ (Rfll 2019: 3). Doch so verständlich der Wunsch nach einer ausschließlich wissenschaftlichen Verwendung der Forschungsergebnisse auch sein mag, so sehr bereitet sie für eine spätere Nachnutzung der Daten Probleme.

Repositorien erlauben nämlich vielfach als Default einen Download auch für nichtwissenschaftliche Zwecke, was dann im Widerspruch zu Formulierung der Einwilligung steht.

Eine Einschränkung der Nachnutzung auf rein wissenschaftliche Zwecke lässt sich in der Praxis nur schwer umsetzen: Technisch hängt dies zunächst freilich von den Möglichkeiten des Repositoriums ab, in welchem die Daten abgelegt werden sollen. Theoretisch könnten hier Identitätsmanagementsysteme wie Shibboleth Abhilfe schaffen, deren Einsatz ist aber bislang kein Standard in der sozialwissenschaftlichen Forschungslandschaft. Einzelprüfungen der Nachnutzungszwecke sind im Rahmen eingeschränkter Datenzugänge in Repositorien zwar vielfach möglich, jedoch stets mit erhöhtem Arbeits- und Organisationsaufwand verbunden. Hier müsste festgelegt werden, wer im Einzelfall den Nachnutzungszweck kontrollieren (z.B. die Forscher*innen, das institutionelle Forschungsdatenmanagement) und wie die Prüfung erfolgen soll (E-Mail-Adresse, Zugehörigkeit zu einer Forschungseinrichtung, usw.). Letztlich steht die Beschränkung auf eine wissenschaftliche Nachnutzung auch der Idee einer generellen Offenheit entgegen, die mit dem Open-Science-Paradigma für Forschungsergebnisse angestrebt wird. Hierzu zählt auch, dass spätestens mit der Vergabe von offenen Lizenzen die Einschränkung auf einen bestimmten Kreis von Nachnutzer*innen nicht mehr vereinbar wäre.

Zusammenfassend lässt sich festhalten, dass dem Thema Datenschutz eine große Bedeutung im Prozess der Kuratierung sozialwissenschaftlicher Forschungsdaten eingeräumt werden muss. Das Fallbeispiel der informierten Einwilligung zeigt, dass es trotzdem möglich ist, die Forschungsdaten unter Wahrung der rechtlichen Schutzvorgaben nachnutzbar zu machen, wenn die Anforderungen von Beginn an im Forschungsprozess mitgedacht werden. Dies lässt sich am besten durch eine frühzeitige Beratung durch das Forschungsdatenmanagement sicherstellen. Gleichzeitig hat das Beispiel der Beschränkung der Nachnutzung auf wissenschaftliche Zwecke gezeigt, dass bei der Erstellung der informierten Einwilligung darauf geachtet werden muss, dass keine zu restriktiven Formulierungen verwendet werden, die eine Veröffentlichung der Forschungsdaten ausschließen. Gleichwohl muss hier festgehalten werden, dass gerade die Sozialwissenschaften eine Disziplin sind, in der nicht alle Forschungsdaten veröffentlicht werden können und in der insbesondere der Umfang der datenschutzkonformen Kuratierung stets in einem Nutzen-Aufwand-Verhältnis gesehen werden sollte (DGS 2019: 5).

3.2 Die Beachtung von Verwertungs- und Nutzungsrechten bei der Datenveröffentlichung

Neben den datenschutzrechtlichen Vorgaben ist innerhalb des Kuratierungsprozesses außerdem die Frage zu klären, wem die Forschungsdaten rechtlich zuzuordnen sind, d.h. wer letztlich über die Veröffentlichung und die Modalitäten der Nachnutzung durch Dritte entscheiden kann. Auch im sozialwissenschaftlichen Kontext wird daher aktuell diskutiert, ob Forschungsdaten als urheberrechtlich geschützte Werke zu betrachten sind. Wäre dies der Fall, läge die Entscheidungsbefugnis über die Datenveröffentlichung klar bei den jeweiligen Primärforscher*innen. Für die Sozialwissenschaften zeichnet sich gemäß eines jüngst veröffentlichten Kurzgutachtens jedoch die Einschätzung ab, dass ein urheberrechtlicher Schutz für Forschungsdaten, die durch standardisierte fachtypische Erhebungsmethoden, also bspw. mittels schriftlicher, Online- oder Telefonbefragung generiert werden, in der Regel nicht anzunehmen ist (vgl. Lauber-Rönsberg et al. 2018). Für qualitative Forschungsdaten könnte ggf. ein Urheberrechtsschutz bestehen, die Hürden sind jedoch in Hinblick auf die notwendige „geistige Schöpfung“ sehr hoch. Für den Einsatz fachüblicher Methoden, wie etwa der Verwendung eines entlang der Forschungsfragen entwickelten Interviewleitfadens, gehen wir daher aktuell ebenfalls eher nicht von einem Urheberrechtsschutz aus. Auch wenn diese Einschätzungen als allgemeine Richtschnur gelten können, kommen die Fachjurist*innen in ihrem Gutachten letztlich zu dem für Forschungsdatenpraktiker*innen nicht ganz befriedigendem Ergebnis, dass „die Schutzfähigkeit einzelner Forschungsdaten in der Regel nur im Einzelfall und selbst dann nicht mit hinreichender Rechtssicherheit beurteilt werden kann“ (ebd.: 3). Selbstverständlich hat dies auch Auswirkungen für die Vergabe von offenen Lizenzen. Zwar wird für die Veröffentlichung von Forschungsdaten aktuell zu der Verwendung der Creative-Commons-Lizenzen „CC0“ oder „CC-BY 4.0“ geraten. Jedoch stellt sich hier das Problem, dass, wenn für sozialwissenschaftliche Forschungsdaten überwiegend kein Urheberrechtsschutz besteht, sich letztlich aus den vergebenen Lizenzen keine rechtlich durchsetzbaren Pflichten, bspw. zur Namensnennung, ableiten lassen. Für unsere Praxis halten wir die angestrebte Vergabe von CC-Lizenzen dennoch für sinnvoll, da sie einerseits einen symbolischen Wert besitzen und zum anderen eine Absicherung für den (vermutlich seltenen) Einzelfall darstellen, in welchem es sich doch um urheberrechtlich geschützte Daten handelt.

Auch unabhängig vom Urheberrechtsschutz stellt sich in der Praxis des Forschungsdatenmanagements jedoch die ganz grundsätzliche Frage, wer über die Veröffentlichung der Daten entscheiden kann. Auch für nicht urheberrechtlich als Werk geschützte Forschungsdaten empfiehlt das Gutachten auf Grundlage des allgemeinen Persönlichkeitsrechts der Wissenschaftler*innen sowie aufgrund der forschungsinternen Normen der guten wissenschaftlichen Praxis, dass die Entscheidungsbefugnis über die Veröffentlichung der Daten bei den Primärforscher*innen liegen sollte. Zwar können sich im Rahmen von arbeits- bzw. dienstrechtlichen Regelungen Forschungsinstitute Nutzungsrechte an den Arbeitsergebnissen der Forscher*innen als Arbeitgeber sichern und die Daten auch ohne Zustimmung der Forscher*innen veröffentlichen. Für die aktuelle Praxis halten wir es jedoch grundsätzlich für empfehlenswert, Daten nicht über die Köpfe der Primärforscher*innen hinweg zu veröffentlichen und die datenproduzierenden Wissenschaftler*innen stets selbst darüber entscheiden zu lassen, ob und unter welchen Bedingungen ihre Daten für Dritte zugänglich gemacht werden sollen. Statt quasi-verpflichtender Bestimmungen sollte u.E. eher auf Informationsangebote, gute Argumente über die Vorzüge einer Datenveröffentlichung oder den Verweis auf die Empfehlungen der institutsinternen Forschungsdatenleitlinie zurückgegriffen werden. Veröffentlichen die Forscher*innen ihre Daten auf eigenen Wegen, stellt sich die Frage der Zustimmung nicht. Veröffentlicht aber – wie am WZB – ein institutionelles Forschungsdatenmanagement die Daten für oder in Kooperation mit den jeweiligen Wissenschaftler*innen, empfiehlt es sich in jedem Fall, deren Zustimmung frühzeitig einzuholen. Dies ist auch gerade dann wichtig, wenn die Daten noch am Institut vorliegen, die Forscher*innen jedoch bereits an einer neuen Einrichtung tätig sind. Denn Daten werden häufig als Kopie mitgenommen und teils auch in die Forschungsvorhaben am neuen Institut eingespeist – bzw. können sie auch bereits am neuen Institut veröffentlicht worden sein.

Ein weiteres komplexes Szenario liegt für uns in der Klärung der Entscheidungsbefugnisse über Datenveröffentlichung und -verwendung in Kooperationsprojekten. Die Forschung in Verbänden nimmt zu und ist seitens der Forschungsförderer auch vielfach explizit erwünscht. Dieses Problem betrifft deshalb sowohl kleinere, oft nur halb formalisierte aber vor allem auch große Projektkooperationen.

Würde ein urheberrechtlicher Schutz bestehen, wären die Rechte aller Miturheber*innen zu berücksichtigen und zu schützen. Selbst wenn jedoch „nur“ von einem Persönlichkeitsrecht der Forscher*innen und den Normen der guten wissenschaftlichen Praxis auszugehen ist, sollte die gemeinsame Entscheidung und Abstimmung der Projektpartner*innen stets Voraussetzung für die Veröffentlichung der Daten sein. Diese erfolgt im optimalen Fall bereits vor Projektbeginn – bspw. über das Instrument einer Kooperationsvereinbarung – und wird im Detail durch einen Datenmanagementplan ergänzt, woran bereits zum Projektstart zu denken ist. Darin können einer Partnereinrichtung die Aufgaben der Datenaufbereitung und -veröffentlichung innerhalb der Projektkonstellation übertragen und die jeweiligen Bedingungen festgehalten werden. Hierbei wäre zum Beispiel zu denken an: Zeitpunkt und Ort der Veröffentlichung, die Dauer einer Embargo-Frist, die Einhaltung der Datenschutzvorschriften, die Vergabe einer CC-Lizenz und die Art der Nennung aller beteiligten Primärforscher*innen und Institute.

Findet die Klärung der Frage der Datenveröffentlichung zum Ende oder erst nach Abschluss eines Projektes statt – was in den Sozialwissenschaften häufig (noch) der Fall ist –, sollte ebenfalls ein Einverständnis aller Projektpartner*innen eingeholt werden. Oft sind zu diesem Zeitpunkt eine Vereinbarung oder ein Datenmanagementplan nicht mehr realisierbar. Am WZB sind wir daher darum bemüht, eine niedrighschwellige Zustimmung einzuholen, für die wir eine Vorlage bereithalten. Auch wenn uns in der Praxis noch kein Fall bekannt ist, bei dem es zwischen beteiligten Kooperationspartner*innen Interessenskonflikte bei der Veröffentlichung von Forschungsdaten gegeben hat, bietet diese Zustimmung doch diverse Vorteile: Sie gewährleistet ein Mindestmaß an Rechtssicherheit und Transparenz und verspricht, unkoordinierte Mehrfachveröffentlichungen derselben Forschungsdaten zu vermeiden. Gewiss bleiben frühzeitig auf Projektebene erstellte Datenmanagementpläne ohne Frage das Instrument der Wahl, um solche Szenarien zu verhindern, in den Sozialwissenschaften sind sie jedoch bislang eher die Ausnahme als die Regel.

3.3 Die Verknüpfung von Text- und Datenveröffentlichung

Ein Meilenstein für die Anerkennung von Forschungsdaten als eigenständiger wissenschaftlicher Output liegt in ihrer Zitierbarkeit. Vergleichbar zu den im Forschungsprojekt entstandenen Textpublikationen wird mit der Zitation von Forschungsdaten die Arbeit der Datenproduzent*innen als Forschungsleistung sichtbar. Auch wenn sich bereits verschiedene Initiativen und wissenschaftliche Communities damit beschäftigt haben, Empfehlungen zur Zitation von Forschungsdaten und bibliografischen Regeln zu erstellen, gibt es (noch) keinen einheitlichen Standard (vgl. Hausstein 2019). Weitgehend Einigkeit besteht jedoch darüber, dass aufgrund des dynamischen und fragilen Charakters digitaler Forschungsdaten Speicherort, eindeutige Version und Metadaten mittels persistenter IDs (z.B. Digital Object Identifier (DOI)) identifizierbar und auffindbar sein müssen. Dadurch wird es prinzipiell möglich, mittels der Datenzitation in den sich anschließenden Textveröffentlichungen auf die Forschungsdaten eindeutig und transparent zu verweisen.

Spätestens hier stellt sich die Frage nach der wechselseitigen Verknüpfung von Daten- und Textpublikationen. In der Praxis stößt man schnell auf die Herausforderung, dass die Aufnahme der Datenzitation in die erste Textpublikation häufig an ein spezifisches Zeitfenster gebunden ist. Die Option, einen Verweis auf die veröffentlichten Forschungsdaten inkl. DOI in einem Zeitschriftenaufsatz unterzubringen, besteht eigentlich nur vor der Einreichung des Artikels beim Verlag und ggf. noch im Rahmen des Review-Prozesses. Ab der finalen Veröffentlichung ist die Aufnahme im Aufsatz selbst de facto nicht mehr möglich. Wir bemühen uns daher in der Praxis, Forscher*innen diesbezüglich rechtzeitig zu informieren. Als ein nützliches Instrument hat sich zudem erwiesen, für die Forschungsdaten im Repositorium einen DOI reservieren zu lassen. Dadurch wird es möglich, die Datenzitation und -verlinkung in die Textpublikation aufzunehmen, auch wenn die tatsächliche Veröffentlichung der Forschungsdaten noch nicht final abgeschlossen ist.

2016 hat eine Studie die Zitationsrate von Forschungsdaten anhand des „Data Citation Index“ sowie dem heterogenen Feld der Altmetrics untersucht. Sie kam zu dem Ergebnis, dass 85 % Prozent der Forschungsdaten unzitiert bleiben (Peters et al. 2016: 740). Gewiss müsste hier tiefer in die Analyse der Kontextbedingungen eingestiegen

werden. Fachspezifische Standards wären zu berücksichtigen und es ist davon auszugehen, dass mit dem zunehmenden Bedeutungsgewinn von Open Data in den letzten Jahren die Tendenz steigend sein dürfte. Dennoch lässt sich ohne weiteres festhalten, dass die Zitationsmechanismen bei Forschungsdaten noch nicht analog zu den Textpublikationen funktionieren. Technisch wird es in Zukunft mit Sicherheit möglich werden, mittels automatisierter Prozesse Zitationen von Forschungsdaten zu erfassen und sichtbar zu machen. Vollintegrierte Forschungsinformationssysteme, die Publikationen, Daten, Projektinformationen u.ä. verbinden, könnten eine optimale Lösung darstellen. Sie sind jedoch mit hohem Entwicklungs- und Ressourcenaufwand verbunden und daher nicht an jedem Institut realisierbar. Zwar ist davon auszugehen, dass die Suche nach und Auffindbarkeit von Forschungsdaten mittelfristig über Fachrepositorien oder Meta-Suchen stattfinden wird. Unsere Erfahrung für die Sozialwissenschaften zeigt jedoch, dass bislang wenig nach Daten in Repositorien recherchiert wird – und wenn dann vor allem durch Forschungsdatenmanager*innen, die Datenrecherchen als Service anbieten.

Es ist folglich davon auszugehen, dass die Forscher*innen einstweilen über die klassische Suche nach Textpublikationen indirekt auf die dazugehörigen Forschungsdaten stoßen. Am WZB haben wir uns daher zunächst für ein pragmatisches Vorgehen entschieden. Sofern Forscher*innen Textpublikationen auf Basis ihrer am WZB veröffentlichten Forschungsdaten erstellen, bitten wir sie, einen Hinweis zur Datenveröffentlichung inkl. DOI im Artikel zu ergänzen. Wird die Textpublikation nach Veröffentlichung im Bibliothekskatalog aufgenommen, ergänzen wir im Katalog zudem den Verweis ins Forschungsdaten-Repository. Im Falle von Zweitveröffentlichungen im Open-Access-Repository wird auch hier der Ablageort der dazugehörigen Forschungsdaten angeführt. Schließlich nehmen wir zugehörige Textpublikationen der Forscher*innen im Datenrepository auf. Denn auch für potentielle Nachnutzer*innen der Daten ist es sicher von großem Interesse, welche Primär- bzw. Sekundärpublikationen bereits auf Grundlage der Daten erstellt wurden. Doch hier ergibt sich ein weiteres zeitliches Problem: Werden die Daten am Ende eines Forschungsprojektes veröffentlicht, liegen die Textpublikationen noch nicht vor. Selbst bei kleineren Projekten kann sich in den Sozialwissenschaften die Publikationsphase über mehrere Jahre erstrecken. Hier wäre die Anforderung entsprechend hoch, die Literatur im Repository nachträglich zu ergänzen.

In der Praxis versuchen wir als institutionelles Forschungsdatenmanagement daher zumindest, dies für die Hauptpublikationen der Datenproduzent*innen zu realisieren und bitten die Forscher*innen nach der Datenveröffentlichung entsprechend, uns über aktuelle Publikationen zu informieren.

4. Fazit

Für das Ziel, Datenbestände langfristig verfügbar und nachnutzbar zu machen, ist eine Qualitätssicherung entlang fachlicher Standards unverzichtbar. Der Weg zur Entwicklung dieser Standards führt unseres Erachtens über die alltägliche Praxis der Datenkuratierung und den Austausch über die jeweils fachspezifischen Probleme und bereits vorliegenden Lösungsvorschläge. Es konnte mit den drei vorgestellten Praxisbeispielen gezeigt werden, dass eine frühzeitige Berücksichtigung von zentralen Aspekten der Kuratierung innerhalb des Forschungsprozesses die Nachnutzbarkeit von Forschungsdaten fördert. Sowohl die Thematik der informierten Einwilligung aus dem Bereich Datenschutz, die Verständigung über Verwertungs- und Nutzungsrechte bei der Veröffentlichung, aber auch die Verknüpfung von Text- und Datenpublikation haben zudem verdeutlicht, dass die forschungsnahe Infrastruktur hierbei eine zentrale Rolle einnimmt. Es muss sich noch zeigen, wie viel Standardisierung die Kuratierung letztlich zulassen wird. Angesichts der Vielfalt und Komplexität der hier aufgerufenen Themen steht bereits jetzt schon außer Frage, wie groß hierbei der Bedarf an Expertise und Unterstützungsangeboten durch ein professionelles institutionelles Forschungsdatenmanagement ist.

5. Literatur

Büttner, S., Hobohm, H.-C. & Müller, L. (2011): Research Data Management. In: dies. (Hg.): Handbuch Forschungsdatenmanagement. Bad Honnef: Bock + Herchen, S. 13–24.

Deutsche Gesellschaft für Soziologie (DGS) (2019): Bereitstellung und Nachnutzung von Forschungsdaten in der Soziologie. Stellungnahme des Vorstands und Konzils der DGS. Online verfügbar unter https://www.sozioogie.de/uploads/media/DGS-Stellungnahme_zum_Forschungsdatenmanagement_08.01.2019_01.pdf, zuletzt geprüft am 26.04.2019.

Droß, P. J. & Linne, M. (2016): Sicheres und einfaches Data Sharing mit SowiDataNet. Dokumentieren – veröffentlichen – nachnutzen. In: Bibliotheksdienst, Vol. 50, H. 7, S. 649–660. Online verfügbar unter <https://doi.org/10.1515/bd-2016-0079>.

Droß, P., Fräßdorf, M., Kubaty, P. & Naujoks, J. (2017): Open Data in den Sozial- und Wirtschaftswissenschaften: das Forschungsdatenrepositorium SowiDataNet, In: Kratzke, J. & Heuveline, V. (Hg.): E-Science-Tage 2017: Forschungsdaten managen. Heidelberg: heiBOOKS, Heidelberg, S. 31-42. Online verfügbar unter <http://hdl.handle.net/10419/172495>.

Hausstein, B. (2019): Zitierbarmachung und Zitation von Forschungsdaten. In: U. Jensen, S. Netscher und K. Weller (Hg.): Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten. Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten. Opladen, Berlin, Toronto: Barbara Budrich, S. 177–192. Online verfügbar unter <https://doi.org/10.3224/84742233.11>.

Herb, U. (2015): Open Science in der Soziologie. Eine interdisziplinäre Bestandsaufnahme zur offenen Wissenschaft und eine Untersuchung ihrer Verbreitung in der Soziologie. Glückstadt: VWH. <https://zenodo.org/record/31234#.WXSd4eLdPY>.

Lauber-Rönsberg, A., Krahn, P. & Baumann, P. (2018): Gutachten zu den rechtlichen Rahmenbedingungen des Forschungsdatenmanagements. Kurzfassung. Im Rahmen des DataJus-Projektes. Online verfügbar unter https://tu-dresden.de/gsw/jura/igetem/ifbimd13/ressourcen/dateien/publikationen/DataJus_Zusammenfassung_Gutachten_12-07-18.pdf, zuletzt geprüft am 26.04.2019.

Peters, I., Kraker, P., Lex, E., Gumpenberger, C., Gorraiz, J. (2016): Research data explored: an extended analysis of citations and altmetrics. In: *Scientometrics* 107, S. 723–744. Online verfügbar unter <https://doi.org/10.1007/s11192-016-1887-4>.

Rat für Informationsinfrastrukturen (RfII) (2019): Stellungnahme des Rates für Informationsinfrastrukturen (RfII) zu den aktuellen Entwicklungen rund um Open Data und Open Access. Online verfügbar unter <http://www.rfii.de/?p=3748>, zuletzt geprüft am 26.04.2019.

Watteler, O., Ebel, T. (2019): Datenschutz im Forschungsdatenmanagement. In: U. Jensen, S. Netscher und K. Weller (Hg.): Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten. Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten. Opladen, Berlin, Toronto: Barbara Budrich, S. 57–80. Online verfügbar unter <https://doi.org/10.3224/84742233.05>.