

Droß, Patrick J.; Linne, Monika

Article — Published Version

Sicheres und einfaches Data Sharing mit SowiDataNet: dokumentieren – veröffentlichen – nachnutzen

Bibliotheksdienst

Provided in Cooperation with:

WZB Berlin Social Science Center

Suggested Citation: Droß, Patrick J.; Linne, Monika (2016) : Sicheres und einfaches Data Sharing mit SowiDataNet: dokumentieren – veröffentlichen – nachnutzen, Bibliotheksdienst, ISSN 2194-9646, de Gruyter, Berlin, Vol. 50, Iss. 7, pp. 649-660, <https://doi.org/10.1515/bd-2016-0079>

This Version is available at:

<https://hdl.handle.net/10419/162875>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



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

Patrick J. Droß, Monika Linne

Sicheres und einfaches Data Sharing mit SowiDataNet: dokumentieren – veröffentlichen – nachnutzen

DOI 10.1515/bd-2016-0079

Zusammenfassung: SowiDataNet ist eine im Aufbau befindliche Forschungsdateninfrastruktur für die Sozial- und Wirtschaftswissenschaften. Zentraler Baustein ist die Entwicklung eines web-basierten Forschungsdatenrepositoriums, welches es Forscherinnen¹ erlaubt, ihre Daten sicher und dauerhaft zu dokumentieren, zu publizieren und mit anderen Forschern zu teilen. Die Entwicklung dieses neuen Services orientiert sich am konkreten Bedarf der Fachcommunity, was sich u. a. in der Durchführung einer detaillierten Anforderungsanalyse widerspiegelt. Ein spezieller Fokus liegt zudem auf der flexiblen Einbindung des Repositoriums in die praktischen Workflows des institutionellen Forschungsdatenmanagements.

Schlüsselwörter: Forschungsdaten, Forschungsdatenmanagement, Forschungsdaten-Repositorium, Data Sharing, Datenarchivierung, Datenkuration

Secure and simple Data Sharing with SowiDataNet: documenting – publishing – reusing

Abstract: SowiDataNet is a research data infrastructure under construction for social sciences and economics. Its central element is the development of a web-based research data repository which allows researchers to document, publish and share their data with other researchers in a secure and lasting way. The development of this new service is orientated towards the community's real needs which, amongst other things, is mirrored in the conduct of a detailed demand analysis. It focusses especially on the repository's flexible integration into the practical workflows of the institutional research data management.

¹ Aus Gründen der besseren Lesbarkeit wird im Text abwechselnd die weibliche und die männliche Form verwendet. Gemeint sind stets alle Geschlechter.

Monika Linne: Monika.Linne@gesis.org

Patrick J. Droß: patrick.dross@wzb.eu

Keywords: research data, research data management, research data repository, Data Sharing, data archiving, data curation

1 Einleitung

Forschungsdaten sind in vielen Disziplinen die Grundlage für neue wissenschaftliche Erkenntnisse und bilden die Basis der daraus resultierenden Veröffentlichungen. Gerade in Zeiten einer immer stärkeren Digitalisierung, in der die überwiegende Anzahl von Forschungsdaten in elektronischer Form vorliegt, kommt der flexiblen Datendistribution und der Nachnutzung von Forschungsdaten eine herausragende Bedeutung zu. Ihre freie und langfristige Verfügbarkeit sichert den Erkenntnisgewinn und stellt die notwendige Transparenz und Nachvollziehbarkeit im Forschungsprozess sicher². Die Möglichkeit zur Replikation trägt zur Vermeidung redundanter Untersuchungen bei, sorgt für eine fachliche Qualitätskontrolle und ist darüber hinaus eine nützliche Basis für die Ausbildung in der Wissenschaft³. Ferner trägt die Verfügbarkeit von Forschungsdaten auch dazu bei, dass die Primärforscherinnen öfter zitiert werden⁴, womit ein wesentlicher Anreiz für die teils durchaus zeitaufwändige Dokumentation der Daten gegeben wird.

Viele Wissenschaftsorganisationen haben Empfehlungen für die Archivierung von Forschungsdaten ausgesprochen. So heißt es z. B. in den aktuellen DFG-Leitlinien zum Umgang mit Forschungsdaten: „Den Regeln der guten wissenschaftlichen Praxis folgend sollen Forschungsdaten in der eigenen Einrichtung oder in einer fachlich einschlägigen, überregionalen Infrastruktur für mindestens 10 Jahre archiviert werden“⁵. Auch der Wissenschaftsrat hat in seinen Empfehlungen neben der Datendokumentation und -aufbereitung vor allem die Forderung

² Vgl. Jasny, Barbara R.; Chin, Gilbert; Chong, Lisa; Vignieri, Sacha: Again, and Again, and Again ... In: *Science* 6060 (2011), S. 1225.

³ Vgl. King, Gary: Replication, Replication. In: *Political Science and Politics* 3 (1995), S. 443–499.

⁴ Vgl. Piwowar, Heather A.; Day, Roger S., Fridsma, Douglas B.: Sharing Detailed Research Data Is Associated with Increased Citation Rate. In: *PLoS ONE* 3 (2007), e308.

⁵ Deutsche Forschungsgemeinschaft: Leitlinien zum Umgang mit Forschungsdaten. http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf [Zugriff: 22.4.2016], S. 1.

nach einer besseren Zugänglichkeit von Forschungsdaten erhoben⁶. Der Rat für Sozial- und Wirtschaftsdaten hat schließlich in seinem Bericht „Auf Erfolgen aufbauend“ darauf hingewiesen, dass die Verwendung von Metadaten-Standards bei der Beschreibung von Forschungsdaten von entscheidender Bedeutung ist⁷.

Für eine Umsetzung dieser Empfehlungen der Wissenschaftsorganisationen existieren jedoch bislang für die Sozial- und Wirtschaftswissenschaften kaum adäquate Infrastrukturen. An dieser Stelle setzt SowiDataNet⁸ an, ein Projektverbund, zu welchem sich GESIS – Leibniz-Institut für Sozialwissenschaften (Projektleitung), WZB – Wissenschaftszentrum Berlin für Sozialforschung, DIW Berlin – Deutsches Institut für Wirtschaftsforschung e. V. und ZBW – Deutsche Zentralbibliothek für Wirtschaftswissenschaften zusammengeschlossen haben. Die Partnerinstitute stellten im Wettbewerbsverfahren der Leibniz Gemeinschaft einen entsprechenden Antrag, welcher mit einer Laufzeit vom April 2014 bis März 2017 bewilligt wurde.

Eines der übergeordneten Projektziele ist die Integration dezentral erzeugter Forschungsdaten unter Nutzung einer gemeinsamen Infrastruktur. Dies hat den Vorteil, dass Forschungsdaten der Sozial- und Wirtschaftswissenschaften in Deutschland nach einheitlichen Standards archiviert und dokumentiert werden, zitierbar und zugänglich gemacht werden können. Die Zusammenführung von Forschungsdaten aus unterschiedlichen Quellen führt zudem zu innovativen Recherchemöglichkeiten, bspw. nach thematischen Schlagworten oder nach unterschiedlichen Erhebungsmethoden. Mit dem Projekt werden darüber hinaus die Möglichkeiten einer effektiven Langzeitarchivierung der Forschungsdaten aus den Sozial- und Wirtschaftswissenschaften substanziell verbessert (durch eine perspektivische Anbindung an das Datenarchiv der GESIS). Da der Kontakt zu den Primärforschern nur selten dauerhaft sicher gestellt werden kann, ist es schließlich Ziel des Projekts, Forschungsdaten über SowiDataNet in einer Form zu dokumentieren, die es erlaubt, sie später auch ohne Unterstützung durch die Datenproduzentinnen nachzuvollziehen und nachnutzen zu können.

Insgesamt zielt das Projekt somit auf den Aufbau und die Etablierung eines digitalen Forschungsdaten-Repositoriums für die Sozial- und Wirtschaftswissen-

6 Wissenschaftsrat: Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020. <http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf> [Zugriff: 22.4.2016], S. 54.

7 Vgl. RatSWD (Hrsg.): Auf Erfolgen aufbauend. Zur Weiterentwicklung der Forschungsinfrastruktur für die Sozial-, Verhaltens- und Wirtschaftswissenschaften. Empfehlungen des Rats für Sozial- und Wirtschaftsdaten. Opladen & Farmington Hills 2011, S. 32.

8 <http://www.sowidatanet.de/>

schaften, welches Wissenschaftlern eine sichere Archivierung und Weitergabe ihrer Forschungsdaten auf nutzungsfreundliche Art und Weise ermöglicht.

2 Projektvorstellung SowiDataNet

Um eine flexible Datendistribution und die wissenschaftliche Nachnutzung von Forschungsdaten in den Sozial- und Wirtschaftswissenschaften zu fördern, startete das Datenarchiv von GESIS bereits Anfang 2012 die Entwicklung von *datorium*, einem digitalen Forschungsdatenrepositorium, in dem Einzelforscherinnen ihre Daten eigenständig dokumentieren können, um sie im Anschluss von GESIS veröffentlichen zu lassen⁹. Seit Januar 2014 ist das Repositorium online und erfährt seitdem eine zunehmende Nutzung. Im Laufe der Betriebsphase von *datorium* wurde jedoch schnell deutlich, dass zusätzlich eine digitale Arbeitsumgebung benötigt wird, die speziell auf die Bedürfnisse eines institutionellen Forschungsumfeldes zugeschnitten ist. Aus diesem Grunde schloss sich der Projektverbund SowiDataNet zusammen, um die spezifischen Anforderungen institutioneller Nutzer hinsichtlich eines Datenrepositoriums zu ermitteln und umzusetzen. So wird SowiDataNet mit besonderer Konzentration auf Forschungseinrichtungen ohne eigene Forschungsdateninfrastruktur weiterentwickelt, während *datorium* vorerst weiterhin auf Einzelforscherinnen fokussiert. Durch diese Vorgehensweise können die Bedarfe aller Nutzergruppen optimal erfasst und berücksichtigt werden. Nach Abschluss der Projektphase und einem produktiven Einsatz von SowiDataNet wird *datorium* und der darin enthaltene Datenbestand in SowiDataNet integriert werden. Dadurch entstehen ein Service und ein System, welches für verschiedene Nutzergruppen zugänglich ist und langfristig eine übergreifende und nutzerfreundliche Recherche ermöglicht. Dies steuert einer weiteren Zerklüftung der Datenlandschaft in Deutschland wirksam entgegen.

In SowiDataNet werden die Wissenschaftler der Partnerinstitute die Möglichkeit erhalten, ihre Daten, Metadaten und relevanten Dokumente eigenständig in das Repositorium einzupflegen sowie ihre Daten mit dazugehörigen Publikationen zu verlinken. Darüber hinaus können sie anhand von Zugangsklassen selbst definieren, unter welchen Bedingungen – z. B. freie Verfügbarkeit bis hin zu Embargofristen – sie den Zugriff auf ihre Daten erlauben. Durch dieses Vorgehen

⁹ Vgl. Linne, Monika: Sustainable data preservation using *datorium*: facilitating the scientific ideal of data sharing in the social sciences. In: Proceedings of the 10th International Conference on Preservation of Digital Objects, edited by José Borbinha, Michael Nelson, and Steve Knight. Lisbon 2013, S. 150–155.

obliegt die Kontrolle über die Datenweitergabe den Forscherinnen. Grundsätzlich ist es jedoch stets das Ziel des Repositoriums, die Daten verfügbar zu machen.

Den Instituten erlaubt SowiDataNet ihre Forschungsergebnisse systematisch zu archivieren und zeitnah der Wissenschaftsgemeinschaft zur Verfügung zu stellen. Die institutsspezifische Datensammlung kann dabei im Corporate Design in den eigenen Webauftritt eingebunden werden. SowiDataNet bietet zudem die vordefinierte Rolle eines institutionellen Kurators, mittels der sich das System in die Workflows des institutionellen Forschungsdatenmanagements integrieren lässt.

Zur sicheren und persistenten Identifizierung von Forschungsdaten nutzt SowiDataNet den Service der Registrierungsagentur für Sozial- und Wirtschaftsdaten da|ra zur DOI-Vergabe. Der DOI-Name identifiziert ein Forschungsprojekt in SowiDataNet unmittelbar und stellt die persistente Verknüpfung mit strukturierten Metadaten sicher. Dies ist eine zentrale Voraussetzung für eine dauerhafte Verbindung von Forschungsergebnissen in Form von Publikationen und den ihnen zugrunde liegenden Primärdaten.

2.1 Art der Forschungsdaten und Zielgruppe

Um ein möglichst großes Spektrum an Daten aufzunehmen, wird der Begriff „Forschungsdaten“ durch das Projekt weit gefasst. So können sowohl Primärdaten in das Repository aufgenommen werden, die aufgrund der Nutzung empirischer Forschungsinstrumente neu produziert werden, als auch Sekundärdaten, die z. B. aus vorhandenen Quellen in neue Datenbanken transformiert werden. Darüber hinaus können ebenfalls Kombinationen von Primär- und Sekundärdaten sowie die für Datentransformationen und -analysen benutzten Routinen oder Skripte berücksichtigt werden. Ergänzend können Dokumente bereitgestellt werden, welche die Nachnutzung der Daten erleichtern (etwa Fragebögen, Codebooks oder Technical Reports).

Im Fokus von SowiDataNet liegen zunächst Projekte mit fehlenden oder geringen personellen und finanziellen Ressourcen für die Dokumentation und Archivierung von Forschungsdaten. Dies sind in der Regel mittlere und kleine Forschungs- oder Promotionsprojekte. Die fachliche Zielgruppe ist die sozial- und wirtschaftswissenschaftliche Forschungsgemeinschaft in Deutschland. Mit den adressierten Disziplinen werden Fächer angesprochen, die zusammen genommen die meisten Studierendenzahlen und dazu entsprechende Lehrstuhl-Strukturen in Deutschland auf sich vereinen.

Bei der Zielgruppe ist zudem zwischen Datengeberinnen und Datennutzern zu differenzieren. Datengeberinnen sollen über SowiDataNet ihre Forschungs-

daten in einfacher Weise dokumentieren, archivieren und anderen zugänglich machen können. Datennutzer sollen mittels der Metadaten recherchieren, Daten für Reanalysen oder neue Forschungsvorhaben nutzen können und mit SowiDataNet einen zentralen Zugang für die Arbeit mit sozial- und wirtschaftswissenschaftlichen Forschungsdaten erhalten. Die Anbindung von Graduiertenkollegs, Sonderforschungsbereichen oder Experimentallaboren wird ebenfalls im Zuge des Projektes eruiert, wobei es sich zumindest bei den ersten beiden Kategorien um Vorhaben handelt, die zum Teil über eigene Infrastrukturen verfügen. Im Verlauf der Betriebsphase soll darüber hinaus geprüft werden, ob und inwieweit Forschungsdaten aus größeren Projekten und Institutionen sowie etablierte Datenbestände in den aufzubauenden Verbund sinnvoll einzubinden sind.

2.2 Stand der Entwicklung

Den Startpunkt des Projektes im Juni 2014 bildete eine nutzerorientierte Bedarfsermittlung, in deren Rahmen ein Auftaktworkshop am Deutschen Institut für Wirtschaftsforschung (DIW), eine Literaturanalyse sowie eine Reihe qualitativer Experteninterviews durchgeführt wurden. Die Ergebnisse und die daraus abgeleiteten Anforderungen ergaben, dass das von GESIS betriebene Repositorium *datorium* bereits grundlegende technische Anforderungen von SowiDataNet erfüllt. Entsprechend wurde auch SowiDataNet mit DSpace aufgesetzt, einer freien Software für die Erfassung, Speicherung und Distribution von digitalen Ressourcen. Daher war das Entwicklungsteam in der vorteilhaften Lage, sehr zügig einen ersten Prototyp von SowiDataNet als eine technische und inhaltliche Weiterentwicklung von *datorium* zu implementieren. Dies hat den Vorteil, dass doppelte Entwicklungen vermieden und Synergien effizient genutzt werden können.

Der erste Prototyp wurde Anfang 2015 fertiggestellt und anhand verschiedener exemplarischer Forschungsdatensätze aus der Praxis der Partnerinstitute getestet. Geprüft wurde die Bedienbarkeit und Eignung für Datensätze unterschiedlichen Formats sowie die Usability der Benutzeroberfläche. Die Erkenntnisse daraus wurden im Detail dokumentiert und fließen in die Weiterentwicklung des ersten Prototyps ein. Im nächsten Schritt wird in diesem Jahr die Testphase auf freiwillige externe Beta-Nutzerinnen erweitert. In dieser Phase ist die Funktionalität des Prototyps bereits gewährleistet. Zu diesem Zeitpunkt ist der Projektverbund besonders auf die Einbindung der wissenschaftlichen Community und deren Hinweise und Verbesserungsvorschläge angewiesen.

Ein wesentlicher Bestandteil im Dokumentationsprozess von Forschungsprojekten ist die Generierung von Metadaten. Metadaten sind nicht nur zur bibliographischen Identifizierung und Erschließung der Daten erforderlich, sondern ganz

wesentlich für ihr Verständnis (Datendefinition), ihre Interpretation (Reichweite möglicher Aussagen, Vergleichbarkeit mit anderen Daten) und ihre Bewertung (Qualität). Sie sind unerlässlich für die Auffindbarkeit sowie die erneute Verwendung der archivierten Daten. Das Metadatenschema von SowiDataNet enthält obligatorische Kernelemente, die zur Beschreibung eines Datensatzes vorhanden sein müssen: Titel, Primärforscher, Institution, Jahr der Veröffentlichung und Verfügbarkeit. Darüber hinaus können optionale Metadatenelemente angegeben werden, um die Daten präziser zu erfassen, z. B. das Fachgebiet, das verwendete Erhebungsverfahren oder die Anzahl der Variablen. Für sämtliche Metadatenelemente werden die jeweiligen Namen, Definitionen, Attribute, ggf. Bedingungen, Kardinalitäten (Anzahl), sowie Wertebereiche definiert, stellenweise wird auf weitere anzuwendende Standards verwiesen. Das Metadatenschema von SowiDataNet ist kompatibel zum Codebook-Standard der Data Documentation Initiative, sowie zu den Metadatenschemata der Datenregistrierungsagentur da|ra und DataCite. Soweit möglich, werden Elemente der Dublin Core Metadata Initiative verwendet. Die Erkenntnisse des ersten Prototypentests hinsichtlich des Metadatenschemas werden aktuell implementiert und in kommenden Tests weiter geprüft, um den Ansprüchen eines nutzerorientierten Repositoriums gerecht zu werden.

Darüber hinaus wird an einem Versionierungskonzept für Forschungsprojekte gearbeitet, welches insbesondere für Forschungsdaten von institutionellen Datengeberinnen von Bedeutung ist. Hierbei ist zu beachten, dass im System auf die jeweiligen Vorgänger- bzw. Nachfolgeversionen referenziert wird, damit stets zur gewünschten Version navigiert werden kann und Irritationen bei nachnutzenden Wissenschaftlern vermieden werden.

Eine spezielle Herausforderung liegt schließlich im Bereich der Anwendung von Thesauri und Klassifikationen, welche sowohl für die Sozial- als auch Wirtschaftswissenschaften geeignet sein müssen. Deren Eignung ist besonders wichtig, um adäquate Funktionen zur Strukturierung und inhaltlichen Erschließung von Forschungsprojekten anzubieten. Insbesondere für die Nachnutzung und Auffindbarkeit von Forschungsdaten ist dies von zentraler Bedeutung.

3 Zentrale Erkenntnisse der Anforderungsanalyse

Anlässlich des Auftaktworkshops waren mehr als 30 Teilnehmerinnen aus Forschung und Infrastruktur eingeladen, gemeinsam mit dem Projektteam über Erfahrungen in der Datenerzeugung und im Datenmanagement, Standards der Beschreibung von Forschungsdaten (Metadaten), Zugriffsrechte und Recherchemöglichkeiten zu diskutieren. Mögliche Bedenken seitens der Forscher in Hin-

blick auf die offene Verfügbarkeit ihrer Daten waren dabei ebenfalls ein wichtiges Thema. Ziel des Workshops war es, fachspezifische Bedarfe und konkrete Anforderungen aus Sicht der Praxis an die zu entwickelnde Forschungsdateninfrastruktur zu identifizieren. Hierzu waren sich die Teilnehmerinnen im Grundsatz einig, dass seitens der Forschungsförderer (bspw. DFG oder EU) aktuell zwei entscheidende Aufgabenstellungen formuliert werden: Forschungsdaten sollen über einen längeren Zeitraum als bisher aufbewahrt und zudem der Community zur Nachnutzung bereitgestellt werden. Ein neues Infrastrukturangebot sollte es den Forschenden somit in jedem Fall ermöglichen, diesen beiden Anforderungen zeitnah und möglichst ohne großen Arbeitsaufwand gerecht werden zu können.

Die Themen des Auftaktworkshops konnten in der Folge in einer Reihe qualitativer Experteninterviews mit empirisch arbeitenden Forscherinnen aus den Sozial- und Wirtschaftswissenschaften weiter vertieft werden. Dabei wurden vielfach Anforderungen formuliert, die als Spannungsverhältnisse zu beschreiben sind: Einerseits der Wunsch, dass Forschungsdaten frei zugänglich gemacht werden – andererseits das Anliegen, Zugriffsmöglichkeiten auf archivierte Daten zu begrenzen (z. B. durch Embargofristen). Einerseits hohe Anforderungen an eine qualitativ hochwertige Dokumentation der Forschungsdaten – andererseits ein möglichst geringer Arbeitsaufwand bei der Kuratierung und Archivierung durch die Datengeber. Zum einen ein großes Interesse und eine grundsätzliche Bereitschaft, Daten mit anderen Forscherinnen zu teilen – zum anderen vielfach Unsicherheiten in Bezug auf Fragen der Anonymisierung von Forschungsdaten oder in Bezug auf lizenz- oder urheberrechtliche Aspekte. In diesen Ambivalenzen spiegelt sich z. T. die Doppelrolle der Forscher als Datenproduzenten einer- und als Datennutzer andererseits. Jedoch zeigten sich auch Unterschiede in Abhängigkeit der verwendeten Forschungsdesigns, welche selbst innerhalb der sozial- u. wirtschaftswissenschaftlichen empirischen Forschungspraxis sehr heterogen ausfallen können. So stellen sich etwa Fragen in Bezug auf die Anonymisierung der Daten vermehrt dann, wenn neben quantitativen auch qualitative Erhebungsmethoden zum Einsatz kommen, bspw. im wachsenden Bereich der Mixed-Methods-Forschung. Unklarheiten in Bezug auf die Nutzungsrechte an den Daten werden hingegen häufiger zum Thema, wenn das Schlagwort Big Data fällt oder wenn Sekundärdaten aufbereitet und mit neuen Informationen angereichert werden.

Die Resultate der Anforderungsanalyse im SowiDataNet-Projekt bestätigen somit grundlegende Erkenntnisse aus bereits vorliegenden Studien zum Themenfeld Data-Sharing. Ein Großteil der formulierten Bedarfe aber auch die bestehenden Bedenken sind zumeist weniger technischer sondern v. a. nicht-technischer

bzw. organisatorischer Natur¹⁰. Demnach ist es kaum verwunderlich, dass seitens der Forscher mit Abstand am häufigsten nach dem Aufwand gefragt wird, der mit der Forschungsdatenkuratierung verbunden ist und danach, ob und welche Unterstützungsangebote hierzu für die Forscherinnen angedacht sind. So ist die Datendokumentation derzeit oftmals nur für den internen Projektkontext konzipiert, bspw. um Informationen für Kollegen oder die eigene Nachnutzung festzuhalten. Hierbei wird häufig nicht nach einheitlichen Standards vorgegangen. So berichtet etwa eine empirische Sozialforscherin aus ihrem Arbeitsalltag: „*Ich weiß nicht, wie es woanders ist, aber bei uns gibt's das im eigentlichen Sinne nicht, so eine strukturierte Dokumentation, jeder macht das irgendwie auf seine Art, [...] wo man wahrscheinlich sagen müsste, wenn man Daten teilt, wird es noch mal mehr Arbeit machen*“¹¹.“ Diese Feststellung trifft umso häufiger zu, wenn Daten aus kleineren und mittleren Forschungsprojekten stammen, in denen die Ressourcen knapp bemessen sind und die Zeit zur systematischen Aufarbeitung und Dokumentation der Datenerhebungen oftmals fehlt.

Die bisherigen Bestrebungen, den Forscherinnen allgemeine Informationen bspw. in Form von Guidelines an die Hand zu geben, werden in der Community äußerst positiv aufgenommen¹². Sie haben entschieden dazu beigetragen die Themen Daten-Archivierung und Data-Sharing in den Forschungsinstituten auf die Tagesordnung zu setzen. Wenn sich Institute allerdings dazu entschließen, diese Themen in der Praxis aufzugreifen, stellen sich schnell eine Reihe ganz praktischer Fragen: Welche Arten von Forschungsdaten werden an unserem Institut produziert? Wie können wir die Forscher für die Idee des Data-Sharing gewinnen? Welche Workflows sind dafür zu etablieren? Welche Infrastrukturangebote können wir nutzen? Schnell wird man feststellen, dass die Forscherinnen in Bezug auf ihre konkreten Daten vielfach Fragen haben, die sich letztlich nur im persönlichen Kontakt klären lassen. Für das SowiDataNet-Projekt stellte sich somit die Herausforderung, in der Entwicklung des neuen Angebots nicht nur zu bedenken, wie sozial- und wirtschaftswissenschaftliche Forschungsdaten nach einheitlichen Standards beschrieben und veröffentlicht werden können. Es galt ebenso zu ermitteln, welchen Beitrag ein institutsübergreifendes Infrastrukturangebot zur Unterstützung des lokalen Forschungsdatenmanagement leisten

¹⁰ Vgl. Feijen, Martin: What researchers want. Utrecht 2011. https://www.surf.nl/binaries/content/assets/surf/en/knowledgebase/2011/What_researchers_want.pdf [Zugriff: 22.4.2016].

¹¹ Droß, Patrick J.: Anforderungen an die Archivierung sozial- und wirtschaftswissenschaftlicher Forschungsdaten. Berlin 2015, S. 18. https://sowidatanet.de/images/pdfs/Kurzstudie_Qua_litative_Interviews.pdf [Zugriff: 22.4.2016].

¹² Vgl. z. B. ZBW, GESIS & RatSWD: Auffinden, Zitieren, Dokumentieren: Forschungsdaten in den Sozial- und Wirtschaftswissenschaften. Hamburg 2015.

kann. Wie lassen sich Funktionalitäten eines Forschungsdatenrepositoriums in den praktischen Arbeitsfluss der Forscher und des wissenschaftsunterstützenden Personals am Institut integrieren?

4 Institutionelle Kuratorinnen als Mittler zwischen Forschung und Forschungsinfrastruktur

Empirische Sozial- und Wirtschaftsforschung findet auch weiterhin überwiegend in institutionellen Kontexten statt, universitär wie außeruniversitär. Nur den wenigsten Forschungsinstituten ist es allerdings möglich auf eigener Ressourcenbasis Forschungsdateninfrastrukturen aufzubauen und dauerhaft zu betreiben, um den beiden Grundanforderungen (längere Aufbewahrung und Zugänglichkeit der Forschungsdaten für die Community) entsprechen zu können. Bisherige Erfahrungen aus institutsübergreifenden Infrastrukturen (bspw. dem GESIS Datenarchiv) zeigen wiederum, dass eine qualitativ hochwertige Kuratierung der Forschungsdaten sehr arbeitsaufwendig ist und insbesondere durch den großen Abstand zur Forschungspraxis eine koordinative Herausforderung darstellt. Im Ergebnis sind die Kapazitäten institutsübergreifender Infrastrukturen entweder begrenzt oder die qualitativen Standards der Datenerschließung müssen stark reduziert werden. Den letzteren Weg geht bereits eine Vielzahl internationaler Angebote zur fachübergreifenden webbasierten Selbstarchivierung von Forschungsdaten. Sie ermöglichen eine schnelle Datenarchivierung ohne differenzierte Datenkuratierung. Der Vorteil ist zweifellos, dass die Daten unmittelbar über einen persistenten Identifikator auffindbar und zitierbar werden. Der Nachteil ist eine auf das absolut Notwendige reduzierte und fachunspezifische Erschließung der Daten. Zur direkten Nachnutzung sind diese Daten somit leider häufig nicht geeignet. Fehlt dann noch eine Möglichkeit die Primärforscher zu kontaktieren, wird eine Nachnutzung u. U. sogar ganz und gar unmöglich. Eine Lösung für dieses Dilemma liegt in einer intermediären Instanz, die zwischen Infrastrukturangebot und Forscherin vermittelt und letztlich am konkreten Institut angesiedelt sein sollte: Dem institutionellen Datenkurator¹³.

13 Bzw. eines Data Librarian oder Forschungsdatenmanagers – Bezeichnung und Zuschnitt dieses Aufgabenbereichs variieren je nach Institut, organisatorischer Aufgabenteilung und Ressourcenausstattung.

SowiDataNet wird daher Funktionen implementieren, die sich direkt an die Datenkuratorinnen richten und in deren organisatorische Workflows integrieren lassen. In einem *ersten Schritt* können Forscher neue Datenprojekte anlegen, Forschungsdaten hochladen und mit standardisierten Metadaten beschreiben. Fragebögen, Codebooks oder Syntaxfiles lassen sich direkt ergänzen. Bereits während der Eingabe können Kommentarfunktionen genutzt werden um offene Fragen festzuhalten und direkt an den Kurator weiterzuleiten. Ist die Eingabe seitens der Forscherinnen beendet, wird das Datenprojekt in einen institutionellen Projektpool übergeben. Auf diesen greift in einem *zweiten Schritt* der Institutskurator zu, wählt ein Projekt aus und startet einen inhaltlichen Review. Dabei prüft er die Daten, Metadaten und Dokumente nach formalen Kriterien, auf Lesbarkeit, Vollständigkeit, Datenschutzerfordernungen und korrekte Beschreibung. Bei Bedarf kann er in Abstimmung mit den Forscherinnen Angaben ergänzen bzw. korrigieren. Als Angebot in Richtung Standardisierung und als Arbeitshilfe für den Kurator wird durch SowiDataNet systemseitig eine Checkliste bereitgestellt, entlang derer die eingereichten Datenprojekte geprüft werden können. Diese Checkliste soll sich künftig an die jeweiligen Bedarfe des Instituts anpassen lassen und auch nach dem offiziellen Projektstart in Zusammenarbeit der Nutzerinnen weiterentwickelt werden. Ist der institutionelle Review abgeschlossen, übermittelt die Kuratorin in einem *dritten Schritt* das Datenprojekt an den technischen Betreiber des Repositoriums (GESIS). Hier erfolgen letzte technische Kontrollen bevor das Projekt über den Registrierungsservice da|ra einen persistenten Identifikator (DOI) erhält und über SowiDataNet veröffentlicht wird.

Es empfiehlt sich, dass die Kuratorinnen die Forscher bereits zu Beginn ihrer Projekte darüber informieren, wie sich schon während der laufenden Projektarbeiten Datenkuratierung und Veröffentlichung vorbereiten lassen. Die Kuratorin kann bspw. mögliche Embargo-Fristen, formale Anforderungen an die Datenaufbereitung oder sinnvolle Schritte der Datendokumentation erläutern. So können zusätzliche Belastungen während der oftmals arbeitsintensiven Abschlussphase im Projekt vermieden werden. Mittelfristig kann SowiDataNet zudem genutzt werden, um die institutionelle Datensammlung zu verwalten, bei der Datenversionierung zu unterstützen und Reporting-Informationen bereitzustellen.

Indem SowiDataNet ein Angebot schafft, welches institutionellen Forschungsdatenkuratorinnen ermöglicht, die Funktionen des Repositoriums flexibel in ihren organisatorischen Workflow zu integrieren, trägt das Projekt auch dazu bei, die Rolle des institutionellen Datenmanagements insgesamt zu stärken.

5 Resümee

Wissensproduktion ist ein Prozess, der darauf basiert, dass neue Erkenntnisse schnell verbreitet werden, damit sie von anderen Forschern weiterverwendet werden können. Wesentlich ist deshalb gerade nicht eine private oder institutionelle Aneignung und Verwertung von Forschungsdaten, sondern deren Weitergabe in der Forschungsgemeinschaft zum Zweck des Erkenntnis- aber auch Reputationsgewinns. Darüber hinaus erleichtert die Veröffentlichung von Forschungsdaten der wissenschaftlichen Gemeinde das Auffinden und die Zitation von relevanten wissenschaftlichen Forschungsergebnissen. In diesem Sinne eignet sich ein Infrastrukturangebot in der Form von SowiDataNet besonders gut, um die Verbreitung und Nutzung von wissenschaftlichen Ergebnissen in einem größeren Spektrum als bisher zu unterstützen und zu erleichtern. Die Anbindung an die institutionellen Workflows ist hierfür unerlässlich. Werden die Bedingungen für eine Entwicklung hin zu einer erfolgreichen Praxis der Forschungsdatendokumentation, -veröffentlichung und -nachnutzung doch nicht zuletzt auch durch ein gutes Zusammenspiel von Forschung und Infrastruktur auf Institutsebene gestaltet.



Patrick J. Droß

Wissenschaftszentrum Berlin für Sozialforschung (WZB)
Bibliothek und wissenschaftliche Information
Projekt SowiDataNet
Reichpietschufer 50
10785 Berlin
E-Mail: patrick.dross@wzb.eu



Monika Linne

GESIS Leibniz-Institut für Sozialwissenschaften
Unter Sachsenhausen 6–8
50667 Köln
E-Mail: Monika.Linne@gesis.org