
Open Data in den Sozial- und Wirtschaftswissenschaften: Das Forschungsdatenrepositorium SowiDataNet

Patrick J. Droß¹, Mathis Fräßdorf², Paul Kubaty³, Julian Naujoks⁴

1,2,3,4 Wissenschaftszentrum Berlin für Sozialforschung

Zusammenfassung. Diverse Empfehlungen und Richtlinien von Forschungsförder- und Wissenschaftsorganisationen verdeutlichen die Bedeutung der Nachnutzung von Forschungsdaten. Die Forderung nach Open Data – also der Archivierung, Veröffentlichung und Nachnutzung von Forschungsdaten – wird immer lauter. Stellvertretend für die Sozial- und Wirtschaftswissenschaften rekonstruiert der Beitrag am Beispiel des Repositoriums SowiDataNet zentrale Anforderungen für den Aufbau einer neuen Forschungsdateninfrastruktur mit dem Ziel des Data Sharings. Drei Aspekte müssen besonders berücksichtigt werden: Fachspezifische Besonderheiten, institutionelle Bedürfnisse und nicht zuletzt die subjektiven Bedenken aus dem Forschungsalltag der Wissenschaftler/innen.

Schlagwörter. Forschungsdaten, Open Data, Repositorium, Sozialwissenschaften, Wirtschaftswissenschaften

Open Data: Ein neues Arbeitsgebiet für Wissenschaft und Infrastruktur

Forschungsdaten nehmen in den empirischen Wissenschaften, insbesondere in den Sozial- und Wirtschaftswissenschaften, einen immer größeren Raum ein. Sie sind nicht nur ein wichtiger Ausgangspunkt wissenschaftlicher Arbeit, sondern werden auch zunehmend als Teil des wissenschaftlichen Outputs gesehen. Wo es früher in Veröffentlichungen nur Beschreibungen von Datenerhebungen gab, wird es heute immer häufiger als gute Praxis angesehen, auch die den Veröffentlichungen zugrundeliegenden Daten bereit zu stellen.

Diese Entwicklung hat unterschiedliche Ursachen: Zunächst wurden im Sinne der guten wissenschaftlichen Praxis Forderungen nach mehr Transparenz und besseren Möglichkeiten der Nachvollziehbarkeit wissenschaftlicher Ergebnisse erhoben. Bereits 1998 formulierte die DFG entsprechende Grundsätze in ihren Empfehlungen (vgl. DFG 1998). Zu diesem Zeitpunkt war an eine entsprechende Umsetzung – insbesondere an die Verfügbarkeit von Daten – allerdings noch nicht zu denken. Die technischen Voraussetzungen waren schlichtweg nicht gegeben. In der Tat rückte erst die Open-Access-Bewegung den Zusammenhang zwischen den Potenzialen des Internets und den Möglichkeiten der Zugänglichmachung wissenschaftlicher Ergebnisse in den Fokus der Diskussion.

Insbesondere durch die Berliner Erklärung aus dem Jahr 2003 wurden diese Möglichkeiten mit einem explizit politischen Anspruch verbunden: Es ging um den Abbau von Schranken zum Wissen, ganz gleich, ob sie technischer, sozialer oder wirtschaftlicher Natur sind. Wissenschaftlicher Output wurde als Teil eines digitalen Gemeinguts betrachtet (vgl. Blasetti et al. 2017). Die nachvollziehbare Logik dahinter: Resultate aus öffentlich finanzierter Forschung sollten auch der

Allgemeinheit zur Verfügung stehen. Das umfasst sämtliche im Forschungsprozess erzielten Ergebnisse, also nicht nur den fertigen Journal-Artikel, sondern zugleich die Forschungsdaten, möglicherweise auch Syntaxfiles und Fragebögen. Auch vor dem Hintergrund einer effizienten Mittelvergabe wird diese Forderung verständlich: Wenn Daten verfügbar gemacht werden, können andere diese Daten möglicherweise nachnutzen und müssen keine eigenen Erhebungen durchführen. In der Praxis gibt es mittlerweile zahlreiche Forschungsförderer, die entlang dieser Argumentation eine Veröffentlichung von Forschungsdaten verbindlich einfordern: Bspw. erwartet die EU in ihrem Rahmenprogramm „Horizon 2020“ explizit die Bereitstellung von Daten aus Forschungsprojekten, die über ihre Förderlinie erstellt werden.

Unabhängig von den eher (wissenschafts-)politischen Argumenten gibt es aber auch Anreize für Wissenschaftler/innen, die Forschungsdaten produzieren, diese auch zu veröffentlichen. Zunächst ist hier eine erhöhte Sichtbarkeit der eigenen Arbeit hervorzuheben. Man wird nicht nur als Autor einer Textpublikation wahrgenommen, sondern auch als Datenproduzent, der anderen eine Nachnutzung ermöglicht und sich außerdem nicht scheut, dass die eigene Arbeit überprüfbar wird. Überdies gibt es Untersuchungen, die nahelegen, dass durch die Verfügbarkeit von Forschungsdaten auch die Zitation der entsprechenden Textpublikationen steigt und sich so die Bewertung der Autorin über die gängigen Indices ebenfalls erhöht (vgl. Piwowar et al. 2017). Langfristig ist zudem damit zu rechnen, dass die Produktion von Forschungsdaten in die Beurteilung der wissenschaftlichen Leistung einfließen wird.

SowiDataNet – Kurzportrait

Im Rahmen des SowiDataNet-Projektverbundes arbeiten GESIS – Leibniz-Institut für Sozialwissenschaften, die Deutsche Zentralbibliothek für Wirtschaftswissenschaften (ZBW) das Wissenschaftszentrum Berlin für Sozialforschung (WZB) und das Deutsche Institut für Wirtschaftsforschung (DIW Berlin) am Aufbau einer neuen Forschungsdateninfrastruktur für die Sozial- und Wirtschaftswissenschaften. Das Projekt ist gefördert durch die Leibniz-Gemeinschaft. Zentraler Baustein ist die Entwicklung eines web-basierten Forschungsdatenrepositoriums, welches es Wissenschaftler/innen in einem institutionellen Arbeitsumfeld erlaubt, ihre Daten nachhaltig zu dokumentieren, sicher und dauerhaft zu archivieren, zu veröffentlichen und damit anderen Forscher/innen zur Nachnutzung zur Verfügung zu stellen. Einen besonderen Fokus legt das Projekt auf die speziellen Bedarfe der beiden Fachcommunities sowie auf eine möglichst flexible und praxisnahe Einbindung des Repositoriums in die Workflows des institutionellen Forschungsdatenmanagements.

Den Instituten soll die Möglichkeit gegeben werden, für die Veröffentlichung ihrer Daten, die in der Regel in mittleren und kleinen Forschungs- oder auch Promotionsprojekten produziert werden, einen zentralen Infrastrukturservice zu nutzen, ohne in eine eigene Infrastruktur zur Datenpublikation investieren zu müssen. Dies hat zudem den Vorteil, dass Forschungsdaten der Sozial- und Wirtschaftswissenschaften in Deutschland nach einheitlichen Standards archiviert und dokumentiert, zitiert und zugänglich gemacht werden können. Die Zusammenführung von Forschungsdaten aus unterschiedlichen Instituten führt zu innovativen Recherche-Möglichkeiten, bspw. nach thematischen Schlagworten oder nach unterschiedlichen Erhebungsmethoden. Dabei verbessert eine zentrale Anbindung an das GESIS-Datenarchiv substantziell die Möglichkeiten einer effektiven Langzeitarchivierung von Forschungsdaten und stellt trotz der Fluktuation der

Forscher/innen über Institutsgrenzen hinweg sicher, dass Forschungsdaten zentral aufbewahrt und in einer Form dokumentiert werden, die es erlaubt, sie später auch ohne Unterstützung durch die Datenproduzent/innen nachvollziehen und nachnutzen zu können.

In SowiDataNet werden Forschende und Forschungsdatenmanager/innen daher die Möglichkeit erhalten, ihre Datensätze sowie relevante Begleitdokumente (etwa Fragebögen, Codebooks, Skripte oder Technical Reports) eigenständig in das Repositorium einzupflegen und mit detaillierten Metadaten zu beschreiben. Forscher/innen können hierbei anhand von frei wählbaren Zugangsklassen selbst definieren, unter welchen Bedingungen – z.B. freie Verfügbarkeit bis hin zu Embargofristen – sie den Zugriff auf ihre Daten erlauben. Zur sicheren und persistenten Identifizierung von Forschungsdaten nutzt SowiDataNet den Service der Registrierungsagentur für Sozial- und Wirtschaftsdaten da|ra zur DOI-Vergabe.

Die Generierung von Metadaten ist ein wesentlicher Bestandteil im Dokumentationsprozess von Forschungsdaten. Metadaten sind unerlässlich für die Auffindbarkeit sowie die erneute Verwendung der archivierten Daten. Das Metadatenschema von SowiDataNet enthält daher die obligatorischen Kernelemente, die zur Beschreibung der Forschungsdaten notwendig sind, erlaubt aber darüber hinaus die Angabe einer Vielzahl fachspezifischer Metadaten-Elemente. Um die Auffindbarkeit und Nachnutzung der Daten zu vereinfachen, ist zudem eine inhaltliche Erschließung von Forschungsdaten mittels fachspezifischer Thesauri vorgesehen.

Im Folgenden werden einige zentrale Punkte der im Rahmen des Projektes durchgeführten Anforderungsanalyse beleuchtet und anschließend die konkreten Umsetzungsschritte dargestellt.

Anforderungen an das Data Sharing

Viele Forschungsförder- und Wissenschaftsorganisationen empfehlen und fordern die Sicherung und Veröffentlichung von Forschungsdaten und begründen dies mit dem Argument, dass dadurch Anknüpfungspunkte für weitere darauf aufbauende Forschung geschaffen werden (vgl. DFG 2015, Allianz der deutschen Wissenschaftsorganisationen 2010). Aus dieser Perspektive ist eine ausschließliche Archivierung bzw. geschlossene Aufbewahrung von Forschungsdaten nicht erstrebenswert. Das Ziel liegt vielmehr in der Nachnutzung der Daten – doch wie kann das „Data Sharing“ in der Praxis konkret aussehen?

Der Rat für Sozial- und Wirtschaftsdaten (RatSWD) hat in seinem jüngst verfassten Ratgeber „Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften“ darauf hingewiesen, dass die „Dokumentation [...] so ausgestaltet sein [sollte], dass eine Nachnutzung der Daten möglich ist“ (RatSWD 2016, 10). Zwar könne die reine Sicherung von Daten bereits mit einer „Minimaldokumentation“ (ebd.) erfolgen, die Möglichkeit der Nachnutzung sei allerdings erst dann gegeben, wenn die Forschungsdaten nutzerfreundlich dokumentiert werden. Dies deckt sich mit Ergebnissen weiterer Untersuchungen, nach denen die Qualität der Datendokumentation ausschlaggebend für eine effiziente Nachnutzung ist (vgl. Fecher/Puschmann 2015, Fecher et al. 2015). Forschungsdaten sollten also mit umfangreichen Metadaten beschrieben und qualitätsgesichert kuratiert werden. Da sich sowohl die Wege der Datengenerierung als auch die Formen der Datennutzung je nach wissenschaftlicher Disziplin stark unterscheiden, sind fachspezifische Standards der Datendokumentation unerlässlich: Bereits 2010 hat die Allianz der deutschen Wissenschaftsorganisationen angeführt, dass die Möglichkeiten der Nachnutzung von Forschungsdaten von der fachspezifischen Nachvollziehbarkeit „der Art und Weise der Datenerhe-

bung, des Umfangs und der Vernetzbarkeit des Datenmaterials sowie der praktischen Brauchbarkeit der Daten“ (Allianz der deutschen Wissenschaftsorganisationen 2010) abhängig sind. Ähnlich lautet der Aufruf der DFG, „angemessene Regularien zur disziplinspezifischen Nutzung und ggf. offenen Bereitstellung von Forschungsdaten zu entwickeln“ (DFG 2015, 2).

Im Projektaufbau von SowiDataNet wurden diese Aspekte von Beginn an integriert: Dementsprechend galt bei der Konzeption die Prämisse, dass die Entwicklung eines Forschungsdaten-repositories im Zusammenspiel von Infrastruktur und Forschung erfolgen muss. Dies war ausschlaggebend dafür, dass mit dem GESIS – Leibniz-Institut für Sozialwissenschaften und der Deutschen Zentralbibliothek für Wirtschaftswissenschaften (ZBW) bzw. dem Wissenschaftszentrum Berlin für Sozialforschung (WZB) und dem Deutschen Institut für Wirtschaftsforschung (DIW Berlin) zwei etablierte Einrichtungen der Leibniz-Gemeinschaft jeweils aus der Infrastruktur und aus der sozial- und wirtschaftswissenschaftlichen Forschung beteiligt waren.

Die disziplinäre Begrenzung auf die Sozial- und Wirtschaftswissenschaften folgte dem Anspruch, die notwendigen detaillierten fachspezifischen Standards für die Datendokumentation aufzustellen und damit das Repository an den Bedürfnissen der Fachcommunities auszurichten. Im Zeitraum von 2014 bis 2015 wurden deshalb im Rahmen einer Anforderungsanalyse Projektworkshops durchgeführt und leitfadengestützte Experteninterviews mit empirisch arbeitenden Forscher/innen aus den Sozial- und Wirtschaftswissenschaften geführt. Diese Interviews sollten Einblicke in die jeweiligen Arbeitsweisen und Bedürfnisse der Wissenschaftler/innen geben, um diese in die aufzubauende Forschungsdateninfrastruktur einfließen zu lassen. Der Gedanke war hierbei, dass dabei letztlich auch die Bereitschaft der Forscher/innen steigen dürfte, das Repository zu nutzen. In den folgenden Abschnitten werden zentrale Erkenntnisse aus dieser Anforderungsanalyse zusammengefasst: Auf der fachspezifischen Ebene müssen zunächst disziplinäre Eigenheiten berücksichtigt werden. Institutionell kommen konkrete Anforderungen der Forschungseinrichtungen zum Ausdruck und auf der forschungspraktischen Ebene ist es von Bedeutung, den (Arbeits-)Alltag und die subjektiven Erwartungen der Forscher/innen miteinzubeziehen.

Fachspezifische Anforderungen

Die fachliche Zielgruppe des SowiDataNet-Projektverbunds ist die sozial- und wirtschaftswissenschaftliche Forschungsgemeinschaft in Deutschland. Die beiden Disziplinen eint nicht nur, dass sie im beachtlichen Maße empirisch arbeiten, sie kennzeichnen sich vor allem durch eine erhebliche Vielfalt in den Wegen und Methoden der Datengenerierung zur Beantwortung ihrer Forschungsfragen. So werden in klassischen quantitativen Surveys unter Verwendung standardisierter Erhebungsinstrumente Mikrodaten erhoben. In experimentellen Laborsettings wird der Einfluss gezielter Stimuli auf Probanden untersucht oder in Feldexperimenten das Verhalten ganzer Untersuchungsgruppen in den Blick genommen. Vielfach werden jedoch auch prozessproduzierte bzw. Sekundärdaten aufbereitet, weiterverarbeitet und mit neuen Informationen angereichert. Schließlich umfasst die qualitative Forschung ein ganzes Spektrum unterschiedlicher Erhebungsmethoden: Dies reicht von klassischen Experteninterviews über ethnografische Forschungsansätze (teilnehmende Beobachtung, Feldnotizen, Fotografie), der Codierung von Sekundärtexten (Zeitungsartikel, Parteiprogramme) bis hin zu Videoaufzeichnungen. Diese methodischen Zugänge werden immer häufiger im Rahmen von Mixed-Methods-Studien miteinander verwoben, um die Vorteile der einzelnen Ansätze zu kombinieren.

Das Feld der empirischen Forschungspraxis in den Sozial- und Wirtschaftswissenschaften erweist sich somit als äußerst heterogen. Die unterschiedlichen Forschungsansätze bringen teils unterschiedliche Prioritäten im Hinblick auf Datenmanagement und Datenarchivierung mit sich. Geringe Fallzahlen (z.B. in Experimenten) oder nicht standardisierbare Methoden (z.B. qualitative Interviews) führen bspw. zu höheren Anforderungen an den Datenschutz, als dies bei anonymisierten Umfragen der Fall ist. Bei der Verwendung von prozessproduzierten bzw. Sekundärdaten muss hingegen die Frage der Nutzungsrechte hinreichend geklärt sein, was wiederum bei Erhebungen durch Forschende selbst meist keine große Rolle spielt. Aus diesen fachspezifischen Besonderheiten lassen sich infolgedessen zwei wesentliche Anforderungen an ein Repositorium für sozial- und wirtschaftswissenschaftliche Forschungsdaten ableiten: Erstens ist es für die potenzielle Nachnutzung der Daten unerlässlich, durch die Erfassung fachspezifischer Metadaten den Entstehungskontext der Daten nachvollziehbar zu machen. Zweitens ist es für die Nachvollziehbarkeit der Forschungsdaten notwendig, dass sich die teils komplexen methodischen Designs im Repositorium abbilden lassen und auch für einzelne Datensätze Angaben zur Methodik erfassen lassen.

Institutionelle Anforderungen

Die Empfehlungen der Förder- und Wissenschaftsorganisationen für einen geregelten Umgang mit Forschungsdaten haben nicht nur Auswirkungen auf die Forscher/innen, sondern nehmen selbstverständlich auch deren Arbeitgeber, Universitäten und außeruniversitäre Forschungsinstitute, in die Pflicht. Denn als einzelne/r Forscher/in sind die Anforderungen nur mit großem Aufwand leistbar, was wiederum Effizienzverluste und damit negative Auswirkungen für den Arbeitgeber hätte. Entsprechend stehen wissenschaftliche Institutionen vor der Herausforderung, Lösungen zu präsentieren, die es den Mitarbeiter/innen erlauben, die Kriterien der guten wissenschaftlichen Praxis zu erfüllen.

Gleichzeitig wird die Abfrage institutioneller Forschungsleistungen mittelfristig zunehmen – z.B. im Rahmen von Evaluierungen – und dabei Forschungsdaten nicht ausklammern. Ähnlich wie bei Text-Publikationen haben Institutionen ein Interesse, dass Informationen über die Forschungsdatenproduktion zentral verfügbar sind und Forscher/innen nicht in Eigeninitiative ihre Forschungsdaten z.B. in sozialen Netzwerken, bei kommerziellen Anbietern, auf Projektwebseiten oder auf privaten Homepages zur Verfügung stellen.

Auch die Anerkennung von Forschungsdaten als wissenschaftlicher Output liegt im Interesse der Institutionen, deren Mitarbeiter/innen in ihrer Forschung Daten produzieren. Hierzu gilt es, zunächst infrastrukturelle Voraussetzungen zu schaffen, die es überhaupt erst ermöglichen, dass diese eigenständige Forschungsleistung als solche wahrgenommen werden kann.

Forschungspraktische Anforderungen

Um Perspektiven aus dem praktischen Forschungsalltag in die Entwicklung von SowiDataNet miteinzubeziehen, wurden im Rahmen der Anforderungsanalyse zehn Forscher/innen zu ihren Erwartungen an ein Forschungsdatenrepositorium mit dem Ziel der Nachnutzung befragt. In der Auswertung ließen sich unterschiedliche Bedenken – vor allem nicht-technischer Natur – heraus-

stellen, die in der folgenden Abbildung aufgelistet sind. Mehrfachnennungen in einem einzelnen Interview wurden einfach gewertet.



Abbildung 1. Bedenken der Forschenden sortiert nach Anzahl der Interviews mit entsprechender Nennung, N=10 (Patrick J. Droß)

Fast alle interviewten Personen teilten die Befürchtung, dass aus der Archivierung und der damit verbundenen Beschreibung und Aufbereitung der Daten ein zusätzlicher Arbeitsaufwand resultiert. Die folgende Aussage kann hierfür als exemplarisch angesehen werden: „Die Frage ist natürlich, mit welchem Aufwand das verbunden ist. Also ich wäre bereit, die Daten alle ins Netz zu stellen, hätte allerdings keine Lust, daran einen Monat zu arbeiten, um die einzugeben, dann sage ich nee, also keinen Bock drauf“. Häufig wiesen die befragten Personen auf ihre geringen Spielräume in Bezug auf die finanziellen und zeitlichen Ressourcen hin. Verschiedentlich wurde die knappe Kalkulation von Forschungsprojekten angeführt, aus der nicht selten eine Aus- und Überlastung abzulesen waren – ein grundlegendes Hemmnis, was die Bereitschaft angeht, für eine Dokumentation von Forschungsdaten zusätzliche Mehrarbeit in Kauf zu nehmen.

Die Mehrzahl der Interviewten gab zudem an, dass die bisherige Datendokumentation häufig nur für den internen Gebrauch konzipiert ist, um beispielsweise Informationen für die eigene Weiterverwendung oder für Kollegen/innen festzuhalten. Eine strukturierte Datendokumentation mit dem Ziel der Veröffentlichung müsste ganz andere Arbeitsweisen innerhalb von Forschungsgruppen nach sich ziehen: „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. [...] Wenn man es sozusagen mal für die Allgemeinheit vernünftig aufbereiten wollte, würde man das wahrscheinlich ganz anders konzipieren“. Meist sind die Daten demnach nicht in solch einem Ausmaß dokumentiert, dass sie ohne weiteres der wissenschaftlichen Gemeinschaft zur Nachnutzung bereitgestellt werden könnten.

Eine weitere Befürchtung betrifft den Besitz- bzw. Erstnutzungsanspruch. Sechs von zehn Personen haben den Wunsch geäußert, ihre erhobenen Daten zunächst selbst verwenden zu wollen, bevor sie sie zur Nachnutzung bereitstellen: „Also speziell die aufbereiteten Sekundärdaten, da haben wir dann manchmal Anfragen, da sind wir aber sehr zurückhaltend, einfach, weil es wahnsinnig viel Arbeit war und weil man jetzt erstmal nichts davon hat, wenn man den Datensatz rausgibt. Also eigentlich müssten wir selber noch erstmal mehr davon veröffentlicht haben, wenn

ich das mal so frank und frei sagen kann“. Die Sorge besteht darin, dass eine zu frühe Publikation der Forschungsdaten dazu führen könnte, dass andere Forschende die Ergebnisse schneller veröffentlichen. In diesem Verständnis wird den Forschungsdaten eine gewisse Exklusivität eingeräumt. Als intellektuelles Kapital können sie zum Beispiel für neue Projektanträge verwendet werden, wie es eine befragte Person klar zum Ausdruck bringt: „Also wir haben uns auch nicht so sehr bemüht [die Daten zugänglich zu machen], ich würde auch immer sehen, dass es ja auch ein bisschen gefährlich ist, gerade solche originären Daten, die wenig vorhanden sind, breit zu streuen, weil das natürlich in unserer Hand auch ein Pfund war. Wenn ich das gleich auf den Markt schmeiße, dann ist das natürlich weg, und wir wussten ja auch nicht, inwieweit wir selber noch weitere Analysen damit durchführen“.

SowiDataNet: dokumentieren – veröffentlichen – nachnutzen

Ohne den Prozess der Datenkuratierung und -veröffentlichung mit SowiDataNet hier in seiner gesamten Breite darstellen zu können, soll im Folgenden zumindest auf drei Spezifika des Repositoriums eingegangen werden, auf deren Entwicklung in Folge der Anforderungsanalyse ein besonderer Schwerpunkt lag: Die Unterscheidung von Projekt- und Objektebene bei der Datenbeschreibung, die spezifischen Funktionen für institutionelle Datenkuratoren sowie die Darstellung der institutionellen Sammlung in einer „Vitrine“.

Datenbeschreibung auf Projekt- und Objektebene

Wie die Anforderungsanalyse im Projekt SowiDataNet zeigte, stehen Forscher/innen bei der Veröffentlichung ihrer Daten vor der Frage, wie sich die teils komplexen Forschungsdesigns in einer Datenpublikation abbilden lassen. Unsicherheiten bestehen daher vor allem in Bezug auf die möglichen Granularitätsstufen einer Datenpublikation: Sollen bspw. zehn experimentelle Untersuchungsreihen besser gemeinsam oder in zehn getrennten Datenprojekten veröffentlicht werden? Wie können die methodischen Unterschiede einer Mixed-Methods-Studie beschrieben werden, ohne dass der für das Verständnis der Daten unerlässliche Gesamtkontext des Studiendesigns verloren geht? Auf diese Fragen kann es aus informationstechnischer Sicht kaum pauschale Antworten geben. Es bedarf vielmehr flexibler und pragmatischer Lösungen im Einzelfall sowie der Unterstützung durch erfahrene Datenkurator/innen.

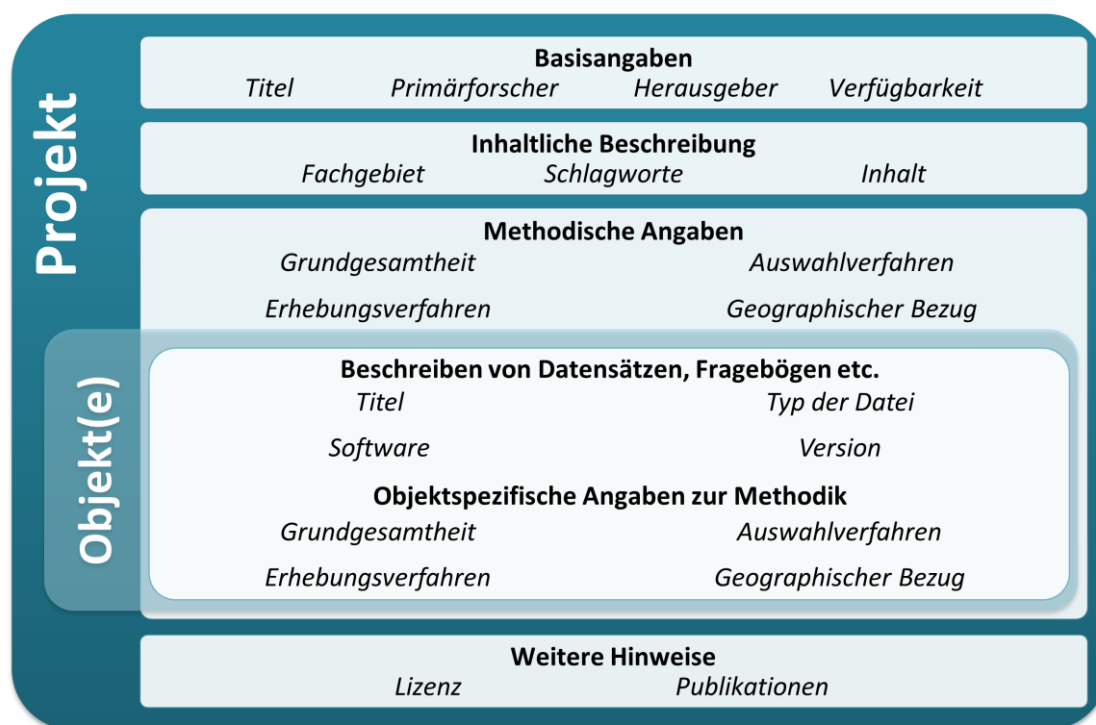


Abbildung 2. Die Projekt-Objekt-Ebene in SowiDataNet (Paul Kubaty)

Um flexible Lösungen auf unterschiedlichen Granularitätsstufen zu ermöglichen, unterscheidet das SowiDataNet-Repository bei der Dokumentation der Forschungsdaten zwischen einer Projekt- und einer Objektebene. Unter der Projektebene wird das zu veröffentlichende Datenprojekt als Ganzes verstanden, während sich die Objektebene auf die einzelnen Dateien (Datensätze und Begleitdokumente) bezieht. Wird im Repository ein neues Datenprojekt angelegt, können zunächst Basisangaben gemacht sowie inhaltliche Beschreibungen auf der Projektebene vorgenommen werden (s. Abb. 2). Auch Angaben zur Methodik lassen sich auf dieser übergeordneten Ebene erfassen. Beim Upload einzelner Datensätze lässt es das Repository jedoch ebenfalls zu, dass zusätzliche Metadaten auf der Ebene einzelner Objekte erfasst werden. Dies führt in der Konsequenz dazu, dass unterschiedliche Datenobjekte in einer Datenpublikation zusammengefasst werden können, ohne dass auf die differenzierte methodische Beschreibung einzelner Objekte verzichtet werden muss.

Spezielle Funktionen für institutionelle Datenkuratoren

Forscher/innen aus den Sozial- und Wirtschaftswissenschaften äußern vielfach Bedenken hinsichtlich des Arbeitsaufwands, der mit der Veröffentlichung ihrer Forschungsdaten verbunden ist. Institutionellen Beratungs- und Unterstützungsangeboten kommen daher bei der weiteren Verbreitung des Open-Data-Gedankens eine Schlüsselrolle zu. Datenkurator/innen können frühzeitig über die formalen Anforderungen an die Datenaufbereitung, sinnvolle Schritte der Datendokumentation sowie über mögliche Embargofristen und Lizenzen informieren. In den oftmals arbeitsintensiven Abschlussphasen der Forschungsprojekte können die Forscher/innen jedoch auch von einer

aktiven Unterstützung bei der Datenkuratierung profitieren. Neue Infrastrukturangebote sollten daher idealerweise die Arbeit des institutionellen Forschungsdatenmanagements unterstützen und sich flexibel in dessen Workflows integrieren lassen.

Suche

Meine Forschungsdaten

Forschungsdaten hinzufügen

Weitere Informationen

Nutzungsbedingungen

Mitglied der
Leibniz
Leibniz-Gemeinschaft

WZB Pool | GESIS Pool | Veröffentlichungen

Hier finden Sie alle Projekte, die sich noch im institutionellen Review befinden. Es gibt zwei Bearbeitungskategorien, eine für Aufgaben, die Sie persönlich übernommen haben, eine andere für Aufgaben, die noch kein Bearbeiter übernommen hat.

Ihre eigenen Aufgaben

Titel	Erstellt von	Erstellt am	Zuletzt bearbeitet am
<input type="checkbox"/> Nachbarschaftliches Zusammenleben in Berlin	WZB Nutzer	02.02.2017	17.02.2017
<input type="checkbox"/> Bildungsperspektiven und Vorurteile	WZB Nutzer	02.02.2017	08.02.2017
<input type="checkbox"/> Wissen schafft Infrastruktur	WZB Nutzer	09.03.2017	09.03.2017
<input type="checkbox"/> Insiderhandel und Informationsaustausch	WZB Nutzer	02.02.2017	09.03.2017

Zurück in den Bearbeitungspool

Aufgaben im Bearbeitungspool

Titel	Erstellt von	Erstellt am	Zuletzt bearbeitet am
<input type="checkbox"/> Erst die Akademie, dann die Berufsausbildung	WZB Nutzer	03.02.2017	09.03.2017
<input type="checkbox"/> Occupational Closure and Women's Timing of Family Formation in Young Adulthood	WZB Nutzer	09.03.2017	09.03.2017
<input type="checkbox"/> Die Entwicklung sozialer Bildungsgerechtigkeiten in der Bundesrepublik	WZB Nutzer	09.03.2017	09.03.2017

Übernehmen Sie die ausgewählten Forschungsprojekte zur Bearbeitung

Abbildung 3. Der institutseigene Pool in SowiDataNet (Gesis)

SowiDataNet implementiert daher Funktionen, die sich speziell an die Datenkurator/innen des jeweiligen Forschungsinstitutes richten (s. Abb. 3). So können Forscher/innen in einem ersten Schritt neue Datenprojekte selbst anlegen, Forschungsdaten hochladen und mittels standardisierter Metadaten beschreiben. Bereits während der Bearbeitung können sie Kommentarfunktionen nutzen, um offene Fragen festzuhalten. Ist die Eingabe seitens der Forscher/innen beendet, wird das Datenprojekt in einen institutionellen Projektpool übergeben. Auf diesen kann in einem zweiten Schritt der/ die Institutskurator/in zugreifen und für ein ausgewähltes Projekt einen inhaltlichen Reviewprozess starten. Dabei prüft er die Daten, Metadaten und Begleitdokumente nach formalen Kriterien, auf Lesbarkeit, Vollständigkeit und korrekte Beschreibung. Er kann auf Fragen der Forscher/innen eingehen und bei Bedarf in Abstimmung mit den Forscher/innen Informationen ergänzen bzw. direkt selbst Korrekturen vornehmen. Wenn erforderlich, kann ein Datenprojekt auch an die Forscher/innen zurückgegeben werden. Als ein Angebot zur Standardisierung und als Arbeitshilfe für den/ die Kurator/in wird von SowiDataNet systemseitig eine Checkliste bereitgestellt, entlang derer die eingereichten Datenprojekte überprüft werden können. Diese Checkliste soll sich künftig an die jeweiligen Bedarfe des Instituts anpassen lassen und auch nach dem offiziellen Projektstart in Zusammenarbeit mit den Nutzern weiterentwickelt werden. Ist der institutionelle Review abgeschlossen, übermittelt der/ die Kurator/in das Datenprojekt in einem dritten Schritt an GESIS, den technischen Betreiber des Repositoriums. Hier erfolgen letzte technische Kontrollen, bevor das Projekt über den Registrierungsservice *data* mit der Vergabe einer DOI veröffentlicht wird.

Darstellung der institutionellen Sammlung

Wie eingangs beschrieben, wird es zunehmend als gute wissenschaftliche Praxis angesehen, nicht nur die Auswertungsergebnisse, sondern auch die den Auswertungen zugrunde liegenden Forschungsdaten zu veröffentlichen. Auch wenn sich die Institute dazu entscheiden, einen zentralen Service zur Datenpublikation zu nutzen, wird dennoch das Interesse bestehen, die eigenen Forschungsdaten lokal sichtbar zu machen und als institutionelle Forschungsleistungen zu präsentieren. SowiDataNet bietet daher in einem Zusatzmodul die Möglichkeit, die institutionelle Datensammlung innerhalb des eigenen Webauftritts darzustellen und durchsuchbar zu machen. Diese institutionelle Vitrine kann an das jeweilige Corporate Design des Instituts angepasst werden (s. Abb. 4).

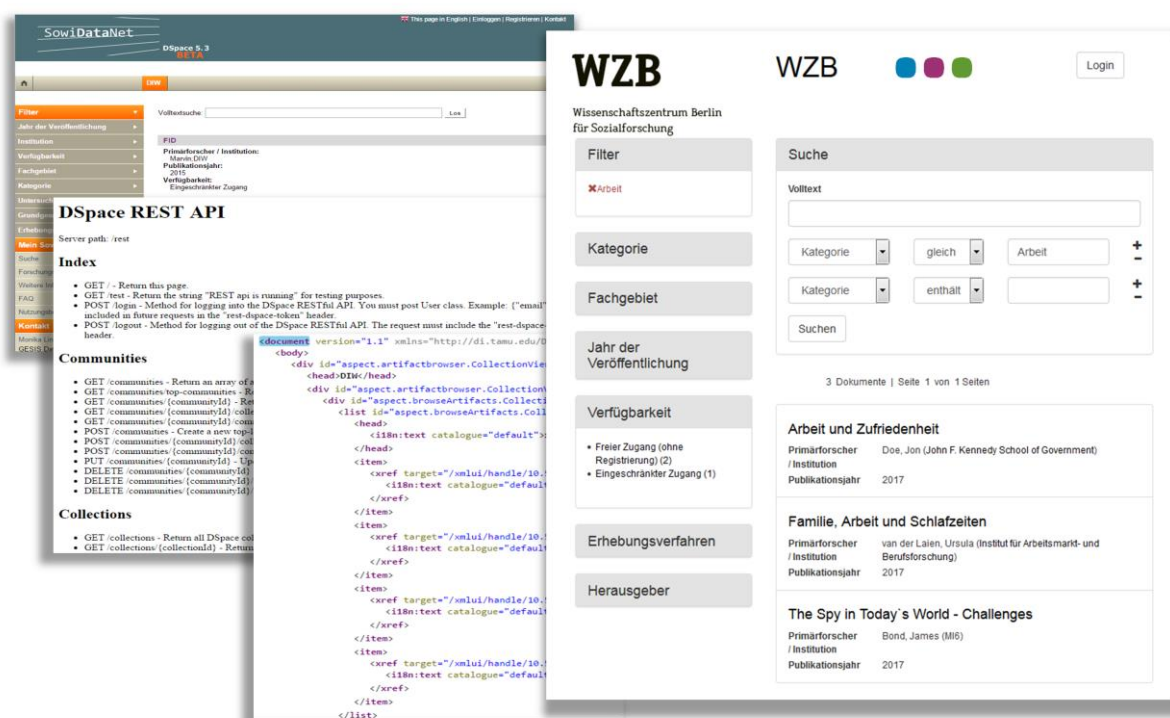


Abbildung 4. Die institutionelle Ansicht in SowiDataNet (Timo Borst, Patrick Droß)

Mit der institutionellen Vitrine stellt SowiDataNet ein eigenes Frontend für die Institute bereit. Die Suchfunktionen entsprechen dabei eins zu eins der Suche auf der SowiDataNet-Startseite (Freitextsuche sowie Filtern über Facetten). Lediglich der Suchraum wird auf die jeweilige institutionelle Datensammlung begrenzt. Der Informationsaustausch zwischen Vitrine und SowiDataNet erfolgt über eine REST API. Daher sind sowohl die Metadaten als auch die Datensätze selbst direkt über die Vitrine verfügbar.

Zur Nutzung der Vitrine besteht zum einen die Möglichkeit diese durch SowiDataNet als zentralen Webservice hosten zu lassen. In diesem Fall kann die Vitrine entweder über einen iFrame in die Institutswebseiten integriert oder als eigenständige Unterseite der Instituts-Homepage aufgerufen werden. Für beide Varianten kann die Ansicht durch Anpassung einer institutsspezifischen CSS-Datei individualisiert werden, z.B. durch die Einbindung eines Logos oder die Anpassung von Farbe und Schriftart. Alternativ besteht die Möglichkeit eine lokale Installation auf den Servern des Instituts vorzunehmen und den Quellcode an die eigenen Bedarfe anzupassen.

Zusammenfassung und Schluss

Am Beispiel von SowiDataNet wurden im vorliegenden Beitrag zentrale Herausforderungen und mögliche Lösungsansätze auf dem Weg zur verstärkten Nachnutzung von Forschungsdaten in den Sozial- und Wirtschaftswissenschaften aufgezeigt. Zentraler Aspekt war dabei die Durchführung einer Anforderungsanalyse, um die Bedarfe der Forscher/innen zu erfragen und diese in die Entwicklung einfließen zu lassen. Dabei wurde auf fachspezifische, institutionelle und forschungspraktische Anforderungen eingegangen. SowiDataNet soll künftig als zentrale Anlaufstelle dienen, über die Forschungsdaten aus den Sozial- und Wirtschaftswissenschaften gesucht und nachgenutzt werden können. Durch die explizite Beteiligung der Fachcommunities und den Fokus auf die praktischen Aufgaben der Datenkuratierung wird zugleich die Rolle des institutionellen Forschungsdatenmanagements gestärkt und der Kulturwandel hin zum Data Sharing befördert.

Literaturangaben

Allianz der deutschen Wissenschaftsorganisationen. 2010. „Grundsätze zum Umgang mit Forschungsdaten.“ Online verfügbar unter <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze.html>. zuletzt geprüft am 13.03.2017.

Blasetti, Alessandro, Mathis Fräßdorf, Patrick J. Droß und Julian Naujoks. 2017. „Digital ist teilbar: Potenziale und Erfolgsbedingungen von Open Access und Open Data.“ *WZB-Mitteilungen* (155): 34-37. Online verfügbar unter https://wzb.eu/sites/default/files/publikationen/wzb_mitteilungen/34-37winwm155web.pdf. zuletzt geprüft am 13.03.2017.

Deutsche Forschungsgemeinschaft (DFG). 1998. „Sicherung guter wissenschaftlicher Praxis – Denkschrift.“ Erste Auflage. Weinheim: Wiley-VCH. Online verfügbar unter http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf. zuletzt geprüft am 13.03.2017.

Deutsche Forschungsgemeinschaft (DFG). 2015. „Leitlinien zum Umgang mit Forschungsdaten: Verabschiedet durch den Senat der DFG am 30. September 2015.“ Online verfügbar unter http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf. zuletzt geprüft am 13.03.2017.

Droß, Patrick J. und Monika Linne. 2016. „Sicheres und einfaches Data Sharing mit SowiDataNet: Dokumentieren – veröffentlichen – nachnutzen.“ *Bibliotheksdienst* 50 (7): 649–60. doi:10.1515/bd-2016-0079.

Fecher, Benedikt, Sascha Friesike, Marcel Hebing, Stephanie Linek und Armin Saueremann. 2015. „A Reputation Economy: Results from an Empirical Survey on Academic Data Sharing.“ *DIW Discussion Papers* 1454. Online verfügbar unter <http://hdl.handle.net/10419/107687>. zuletzt geprüft am 13.03.2017.

- Fecher, Benedikt und Cornelius Puschmann. 2015. „Über die Grenzen der Offenheit in der Wissenschaft – Anspruch und Wirklichkeit bei der Bereitstellung und Nachnutzung von Forschungsdaten.“ *Information - Wissenschaft & Praxis* 66 (2-3): 146–50. doi:10.1515/iwp-2015-0026.
- Piwowar, Heather A., Roger S. Day und Douglas B. Fridsma. 2007. „Sharing detailed research data is associated with increased citation rate.“ *PloS one* 2 (3): e308. doi:10.1371/journal.pone.0000308.
- Rat für Sozial- und Wirtschaftsdaten (RatSWD). 2016. „Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften: Orientierungshilfen für die Beantragung und Begutachtung datengenerierender und datennutzender Forschungsprojekte.“ RatSWD Output Series 3. Online verfügbar unter http://www.ratswd.de/dl/RatSWD_Output3_Forschungsdatenmanagement.pdf. zuletzt geprüft am 13.03.2017.