

Bericht zum Projektworkshop SowiDataNet

23.06.2014, DIW Berlin

Monika Linne (GESIS)

Claus-Peter Klas (GESIS)

Timo Borst (ZBW)

Ralf Toepfer (ZBW)

Mathis Schröder (DIW)

Rainer Siegers (DIW)

Patrick J. Droß (WZB)

Berlin, 22.07.2014

Inhalt

1.	Projektworkshop SowiDataNet – 23. Juni 2014	2
2.	Erfahrungen in der Datenerzeugung und im Datenmanagement.....	3
	Begriffsklärung „Forschungsdaten“	3
	Stand der Erfassung von Forschungsdaten.....	5
	Bisherige Nutzung externer Infrastrukturangebote	5
3.	Standards der Beschreibung (Metadaten), Zugriffsrechte und Recherchemöglichkeiten	7
	Welche Metadaten(-schemata) sollen verwendet werden?	7
	Wie könnten Zugriffsrechte für SowiDataNet aussehen?	7
	Welche Recherchemöglichkeiten sollte SowiDataNet bieten?	8
	Sonstiges	8
4.	Anforderungen an eine zukünftige Forschungsdateninfrastruktur	9
	Provenance	9
	Versionierung.....	10
	Suchbarkeit/Nutzungsmöglichkeiten.....	10
	Ease of use	10
	Rechtemanagement und Datenschutzaspekte.....	10
	Qualitätsprüfung.....	10
	Wissenstransfer	11
	DOI	11
	Sonstiges	11
5.	Hemmnisse für die Nutzung seitens der Forscher/innen	12
	Nutzungsbarrieren seitens der Forschenden als Datenproduzent/innen	12
	Perspektiven zum Abbau der Nutzungsbarrieren.....	13
	Anforderungen für das Projekt	15
6.	Abschlussrunde und offene Fragen für das Projekt.....	16
	Anhang: Liste der Teilnehmenden.....	Fehler! Textmarke nicht definiert.

1. Projektworkshop SowiDataNet – 23. Juni 2014

Werden in Ihrer Institution derzeit schon Forschungsdaten in einer bestehenden Infrastruktur erfasst? Welche Metadatenstandards werden dabei genutzt? Was wären Anforderungen an eine künftige Forschungsdateninfrastruktur? Welche Hemmnisse können der Nutzung einer solchen Infrastruktur seitens der Datengeber/innen entgegenstehen? Dies waren zentrale Fragen des Projektworkshops SowiDataNet, welcher am 23. Juni 2014 am Deutschen Institut für Wirtschaftsforschung (DIW Berlin) stattfand. Der Workshop war zentraler Bestandteil des Arbeitspakets „Anforderungsanalyse“ (AP1) und hatte die doppelte Zielstellung, mehr über den derzeitigen Ist-Zustand im Umgang mit Forschungsdaten in den verschiedenen Instituten in Erfahrung zu bringen sowie Bedarfe und konkrete Anforderungen der Teilnehmenden im Hinblick auf den operativen Aufbau einer künftigen Forschungsdateninfrastruktur im Rahmen des SowiDataNet-Projekts zu ermitteln.

Begrüßt wurden die mehr als 30 Teilnehmenden aus Infrastruktureinrichtungen und Wissenschaft von Mathis Schröder (DIW Berlin). Anschließend stellte Monika Linne seitens GESIS als federführendem Institut im Projektverbund Ausgangspunkte und Zielstellungen des Projekts vor. Sie verwies auf die derzeit stark zerklüftete Dateninfrastrukturlandschaft in Deutschland, in der sich eine übergreifende Recherche von Daten häufig aufwendig und nutzerunfreundlich gestaltet. Jenseits von nationalen oder internationalen Forschungsprogrammen aus größeren sozial- oder wirtschaftswissenschaftlichen Einrichtungen, ist der Anteil von archivierten Forschungsdaten an der Menge der insgesamt produzierten Daten zudem noch sehr gering. Dies trifft insbesondere auf kleinere Forschungsprojekte zu, welchen keine eigenen Ressourcen für das Datenmanagement zur Verfügung stehen. Zielstellung des SowiDataNet-Projekts ist daher die Implementierung einer digitalen Infrastruktur für Forschungsdaten, welche eine eigenständige und niedrigschwellige Archivierung, Dokumentation und Distribution von Forschungsdaten ermöglicht und sich dabei am konkreten Bedarf der Scientific Community orientiert. Im Mittelpunkt stehen dabei quantitative Daten aus den Sozial- und Wirtschaftswissenschaften und damit aus zwei besonders empirisch orientierten Wissenschaftsdisziplinen. Hervorzuheben ist hierbei die Vernetzung von dezentral erzeugten Forschungsdaten aus unterschiedlichen wissenschaftlichen Einrichtungen oder auch von Einzelforschenden. Dies ist eine unbedingte Voraussetzung für eine nutzerfreundliche zentrale Recherche. Der Zugang zu den Daten erfolgt flexibel entweder direkt über SowiDataNet oder indirekt durch eine Weiterleitung an die datengebende Einrichtung.

Im weiteren Verlauf des Workshops stimmten zwei inhaltliche Input-Vorträge auf die Diskussion ein. Benedikt Fecher (HIIG und DIW Berlin) stellte in seinem Vortrag „Forschungsdaten teilen. Die Perspektive der Datennutzer/innen“ Ergebnisse aus seiner aktuellen Studie zum Data Sharing vor¹. Basierend auf einer umfangreichen Literaturanalyse wird darin ein konzeptioneller Analyserahmen entwickelt, der es erlaubt, die involvierten Akteure und ihre Interessen systematisch zueinander ins Verhältnis zu setzen. Das Konzept hilft, den Prozess des Data Sharing theoretisch zu erfassen und zugleich forschungspolitische Schlussfolgerungen zu ziehen. In Ihrem Vortrag „Zugang zu Forschungsdaten: Strategien des RatSWD“ gab Claudia Oellers, Leiterin der Geschäftsstelle des Rats für Sozial- und Wirtschaftsdaten, Einblicke in die Arbeit und Strategien dieser wichtigen Schnittstelle

¹ Fecher, Benedikt/Friesike, Sascha/Hebing, Marcel (2014): What Drives Academic Data Sharing? RatSWD Working Paper 236, http://www.ratswd.de/dl/RatSWD_WP_236.pdf [Letzter Zugriff: 15.07.2014]

zwischen Wissenschaft, Forschungsinfrastruktur und Politik. Sie skizzierte die Entwicklungen und Aktivitäten des Rats seit den 1990er Jahren, erörterte aber auch aktuelle Fragen nach den Möglichkeiten, um mehr Anerkennung für die Produktion und Bereitstellung von Forschungsdaten zu generieren. Des Weiteren wandte sie sich der zunehmenden internationalen und interdisziplinären Ausrichtung der empirischen Forschung sowie wichtigen Fragen des Datenschutzes und der Forschungsethik zu.

Zentraler Programmpunkt des Workshops war schließlich die Diskussion der Teilnehmer/innen an vier moderierten Thementischen. An Tisch 1 wurden „Erfahrungen in der Datenerzeugung und im Datenmanagement“ besprochen. Tisch 2 diskutierte „Standards der Beschreibung (Metadaten), Zugriffsrechte und Recherchemöglichkeiten“. Die „Anforderungen an eine zukünftige Forschungsdateninfrastruktur“ waren das Thema an Tisch 3, und Tisch 4 behandelte mögliche „Hemmnisse für die Nutzung seitens der Forscher/innen“. Die Themen stießen bei den Teilnehmenden auf großes Interesse, und an allen vier Tischen kam es zu einer angeregten Diskussion. Die Gespräche wurden durch Leitfragen strukturiert und zentrale Punkte durch die jeweiligen Tischmoderator/innen an den Pinnwänden festgehalten.

Im Folgenden sind die Ergebnisse der Diskussionen an den vier Thementischen zusammenfassend dokumentiert. Anschließend werden – auch unter Einbezug der Abschlussdiskussion – zentrale Fragen für die weitere Arbeit des Projekts formuliert.

2. Erfahrungen in der Datenerzeugung und im Datenmanagement

An Tisch 1 des als World-Café organisierten Projektworkshops sollten die Kenntnisse der Teilnehmenden im Umgang mit Forschungsdatenmanagement in Erfahrung gebracht werden. Ein Schwerpunkt sollte darauf liegen, den Status Quo in den einzelnen Institutionen zu erfassen, um daraus Rückschlüsse auf den Bedarf und die Anforderungen der aufzubauenden Infrastruktur im Rahmen des Projektes „SowiDataNet“ ziehen zu können.

Zur Kanalisation der Diskussion gab es drei Themenbereiche, die von der Moderation explizit angesprochen wurden. Zunächst wurde versucht, den Begriff „Forschungsdaten“ zu definieren und verschiedene Dimensionen zu identifizieren, anhand derer Forschungsdaten unterschieden werden können. Des Weiteren sollte geklärt werden, ob und in welcher Ausprägung an den Institutionen der Teilnehmenden bereits eine Forschungsdateninfrastruktur besteht und ob es strukturierte Vorgaben zum Umgang mit Forschungsdaten in Form von Data Policies und/oder Forschungsdatenmanagementplänen gibt. Schließlich sollten die bisherigen Erfahrungen bei der Nutzung externer Infrastrukturangebote betrachtet werden.

Begriffsklärung „Forschungsdaten“

Die Frage, welche Daten als „Forschungsdaten“ zu definieren sind, konnte nicht abschließend beantwortet werden. Es gab die Ansicht, dass es sich dabei um Daten handelt, die tatsächlich in der Forschung genutzt werden/wurden. Demgegenüber standen aber auch Auffassungen, den Begriff weiter zu fassen und auch anfallende Prozessdaten oder Ähnliches darunter zu subsumieren, auch wenn sie (noch) keine konkrete Verwendung in Forschungsprojekten finden.

In Bezug auf die Archivierung von Forschungsdaten hielten die Teilnehmenden überwiegend eine Relevanzprüfung für nötig. Kriterien, relevante Datensätze zu erkennen, auch im Hinblick auf langfristige Nutzbarkeit und schwer zu prognostizierende zukünftige Forschungsinteressen, müssten dafür erarbeitet werden. Die Gefahr, eine zukünftige Relevanz eines Datensatzes nicht zu erkennen und seine Archivierung vorschnell zu verwerfen, könnte umgangen werden, indem eine solche Relevanzprüfung nicht stattfindet. Ressourcen des Repositoriums müssten in diesem Fall aufgewendet werden, ohne dass dadurch ein Nutzen entstünde. Diese Aspekte gilt es gegeneinander abzuwägen. Ein Vorschlag aus der Runde war, die gespeicherten Datensätze nach bestimmten Zeiträumen einer Relevanzprüfung zu unterziehen und sie dann ggf. zu löschen, bspw. wenn ein Datensatz nach 20 Jahren nie nachgefragt wurde.

Auch die Frage nach den Datenformaten, in denen Forschungsdaten vorliegen können, konnte nicht abschließend geklärt werden. Neben klassischen Datensätzen im „rechteckigen“ Format (Zeilen und Spalten mit alphanumerischem Inhalt) und Tabellen kommen auch ganz andere Datenformate in Frage. So könnten zum Beispiel Bilder und Texte als Rohdaten für weitere Analysen dienen. Als Beispiel wurde ein Projekt am WZB erwähnt, in dem Parteiprogramme international vergleichend auf Basis einer quantitativen Inhaltsanalyse untersucht werden. Auch Syntax-Files könnten von Interesse sein, wenn sie bspw. dazu dienen, Daten aus verschiedenen Quellen zusammenzuführen, deren Zugriffsberechtigungen ein Speichern des tatsächlichen Datensatzes in einem externen Repository ausschließen. Die Frage, ob Syntax-Files als Forschungsdaten interpretiert werden können, wurde aktiv von der Moderation aufgeworfen und stieß bei den Teilnehmenden auf ein geteiltes Echo. Mehrere Teilnehmende hatten Erfahrung mit qualitativer Forschung. Abschriften und Audio- oder Videodateien von Interviews (standardisiert oder qualitativ) wären weitere Datenformate, die als Rohdaten berücksichtigt werden können.

Insgesamt ergab sich der Eindruck, dass es viele Formate geben kann, in denen Forschungsdaten vorliegen können, und dass im Projekt darüber entschieden werden muss, welche durch SowiData-Net abgedeckt werden sollen bzw. ob ein weitgehend formatoffener Ansatz erwünscht/machbar ist.

Neben den Datenformaten wurden weitere Dimensionen zur Unterscheidung von Datentypen diskutiert. So wurde auf den Unterschied zwischen Primär- und Sekundärdaten oder zwischen abgeschlossenen und fortlaufend aktualisierten Datensätzen verwiesen. Mehrere Teilnehmende gaben an, auch Daten externer kommerzieller Datenanbieter zu nutzen und diese mit Daten aus anderen Quellen zusammenzuspielen. Hier ist eindeutig, dass sie nicht die Rechte haben, solche Daten in einem Repository einzustellen. Auch wurde darauf verwiesen, dass der „Weg“ zwischen den Rohdaten und dem schließlich zur Analyse verwendeten Datensatz häufig eine Blackbox ist, die eine Replikation fremder Studien verhindert (z.B. Umgang mit Fehlern in den Rohdaten).

Mehrmals wurde betont, dass Nutzer/innen beim Einstellen von Forschungsdaten eine Versionierung der Daten vornehmen können müssen. Gleichzeitig sollte hinsichtlich Erweiterungen von Datensätzen und Fehlerkorrekturen geklärt werden, wer für eine Aktualisierung zuständig ist und ob diese zu leisten ist.

Stand der Erfassung von Forschungsdaten

In den meisten Instituten der Teilnehmenden gibt es derzeit weder eine Erfassung von Forschungsdaten, noch eine Data Policy, noch einen strukturierten Datenmanagementplan, an dem sich die Forschenden orientieren können/sollen. Bei einigen stehen diese Themen allerdings auf der Agenda.

Die am weitesten fortgeschrittene Ausnahme aus dem Kreis der Teilnehmenden bildet der SFB 882 an der Uni Bielefeld. Hier bestehen bereits Forschungsdateninfrastruktur, Datenmanagementplan etc. und diese werden auch genutzt. Die Forschenden sind von Seiten des SFB dazu verpflichtet, den Richtlinien zu folgen und ihre Daten in der „Virtuellen Forschungsumgebung“ abzulegen.²

Die ZBW hat mit dem EDaWaX-Projekt³ bereits eine Forschungsdateninfrastruktur im Haus, die allerdings nicht auf die eigene Institution beschränkt ist und der Speicherung von Datensätzen im Zusammenhang mit Publikationen zum Zwecke der Replikation sowie Nachnutzung dient.

Das DIPF erarbeitet derzeit einen Datenmanagementplan, und an der Universität Frankfurt/Main ist ein disziplinübergreifendes Forschungsdatenmanagement geplant, das an der Universitätsbibliothek angesiedelt sein wird. Dort besteht das Interesse, außerhalb der Universität Sichtbarkeit der Daten zu schaffen, indem die entsprechenden Metadaten zugänglich und durchsuchbar gemacht werden sollen. Eine Einbindung der Metadaten bei externen Repositorien scheint gewünscht zu sein.

Die Universität Bielefeld und GESIS geben Forschenden darüber hinaus auch Kurse zur Erstellung von Datenmanagementplänen bzw. zum Forschungsdatenmanagement.

Die Universität Bielefeld hat ein Webtool entwickelt, das durch konkrete Abfragen zu einem individuellen Forschungsprojekt die halbautomatisierte Generierung eines Datenmanagementplans ermöglicht, bestehend aus geeigneten, vordefinierten Textbausteinen.

Bisherige Nutzung externer Infrastrukturangebote

Keine/r der Teilnehmenden hat bisher einen eigenen Datensatz bei einer externen Infrastruktur hinterlegt. Die Diskussion wurde daher auf die Erfahrungen mit externer Infrastruktur aus der Perspektive von Datennutzer/innen sowie auf die Wünsche und Anforderungen an eine externe Dateninfrastruktur gelenkt.

Es entstand der Eindruck, dass die Mehrheit der Teilnehmenden ausschließlich über persönliche Hinweise/Mundpropaganda auf für ihre Forschungsinteressen geeignete Datensätze aufmerksam wird. Systematische Recherche nach Forschungsdaten findet nur vereinzelt statt. Dabei sind die Forschungsdatenzentren oder die Internetauftritte der Datenanbieter die wichtigsten Anlaufstellen bei der Suche.

Die zersplitterte Landschaft im Bereich der Forschungsinfrastruktur wurde mehrfach thematisiert. Die Existenz einer (oder einiger weniger) zentralen Anlaufstelle(n) ist wesentliche Bedingung für die Verwendung einer solchen Infrastruktur bei der Suche nach bestehenden Forschungsdaten. Damit sie genutzt wird, muss sich die Suche bei einer externen Infrastruktur lohnen in dem Sinne, dass sie in der Lage ist, per Online-Suche die relevanten Datensätze anzuzeigen. SowiDataNet muss hier ent-

² Überblick unter folgendem Link: <http://www.sfb882.uni-bielefeld.de/de/projects/inf/services>

³ <http://www.edawax.de/>

scheiden, ob es eine solche zentrale Anlaufstelle sein will und intensiv „Harvesting“ betreibt oder als Repository einen Schwerpunkt auf die Archivierung legt. Im letzteren Fall sollte stark darauf geachtet werden, das Angebot harvestingfähig zu gestalten, indem technische Schnittstellen und Metadatenschemata gegebenen Standards entsprechen.

Als weitere Anforderungen für eine effektive Online-Recherche wurden genannt:

- Einheitliche „Maßeinheiten“: Bspw. wurde in einer Quelle ein Panel als ein Datensatz aufgeführt, in einer anderen Quelle war jede Welle eines Panels ein eigener Datensatz.
- Hinweise auf die Qualität des Datensatzes.
- Hinweise auf die Relevanz des Datensatzes: Welche Publikationen basieren auf ihm?
- Dokumentation des Datensatzes: häufig auf Variablenebene; Zugriff auf Erhebungsinstrumente; Skalenhandbücher; Erhebungszeitraum; Stichprobengröße.
- Hinweise auf Nutzungsbedingungen.
- Verschlagwortung und Suchbarkeit der Inhalte.

Während der Diskussion wurde ein Aspekt angesprochen, der relevant scheint, obwohl er nicht direkt im Zusammenhang mit der Anforderungsanalyse für SowiDataNet steht: Die Zuständigkeiten innerhalb der Dateninfrastrukturlandschaft sollten zwischen den Akteuren und Akteurinnen (institutionell, disziplinspezifisch, interdisziplinär, international etc.) sinnvoll aufgeteilt werden, um einerseits doppelte Arbeit zu vermeiden und andererseits die gegenseitige Unterstützung und Zusammenarbeit zu sichern. Diese könnte gefährdet sein, wenn sich einzelne Akteure durch andere existenziell bedroht sehen.

Im Projekt muss ein sinnvoller Kompromiss aus hohen Anforderungen an die Dokumentation der einzelnen Datensätze zur sinnvollen Recherche mit dem Ziel der Nachnutzung vorhandener Forschungsdaten auf der einen Seite und der Anforderung einer möglichst einfachen, niedrigschwelligen und mit geringem Aufwand zu bewerkstelligenden Hinterlegung von Daten für die Produzent/innen auf der anderen Seite gefunden werden. Hierbei gilt es auch zu erörtern, inwieweit im Repository Personal dafür eingeplant werden kann, den Datenproduzent/innen Arbeit abzunehmen und die Anforderungen der Datennutzer/innen dennoch zu erfüllen.

Übersicht der Anforderungen an das Projekt – Tisch 1

Es ist zu klären,...

- ob in Bezug auf die Archivierung von Forschungsdaten eine Relevanzprüfung erfolgen soll.
 - welche Arten von Daten (Tabellen, Texte, Bilder, Audio, Video ...) abgedeckt werden sollen.
 - in welchen techn. Dateiformaten (z.B. XLS, SAV, JPG, AVI ...) Daten hinterlegt werden können (Ergänzung mit Blick auf Langzeitarchivierung – im Workshop noch nicht thematisiert)
 - ob der Weg von den Rohdaten zu den Use-Files vollständig dargestellt werden soll bzw. kann.
 - wo die Zuständigkeit für Aktualisierungen von Datensätzen und Fehlerkorrekturen liegt. Eine Versionierung der Daten muss in jedem Fall möglich sein.
 - ob SowiDataNet eher eine zentrale Anlaufstelle sein will und intensiv „Harvesting“ betreibt oder als Repository einen Schwerpunkt auf die Archivierung legt.
 - wie ein sinnvoller Kompromiss zwischen hohen Anforderungen an die Dokumentation seitens der Datensuchenden und möglichst niedrigen Anforderungen an die Datengeber/innen aussehen kann.
-

Formulierte Anforderungen an die Online-Recherche:

- Einheitliche Maßeinheiten verwenden (z.B. Welle vs. gesamtes Panel)
 - Hinweise auf Qualität des Datensatzes
 - Dokumentation: Stichprobengröße, Erhebungszeitraum, Skalenhandbücher, Erhebungsinstrumente, Variablenebene, Nutzungsbedingungen, Publikationen (Relevanz des Datensatzes)
 - Verschlagwortung und Suchbarkeit der Inhalte
-

3. Standards der Beschreibung (Metadaten), Zugriffsrechte und Recherchemöglichkeiten

Welche Metadaten(-schemata) sollen verwendet werden?

Die Nutzung von Metadatenstandards in den verschiedenen Instituten ist sehr heterogen. In den Instituten der meisten Teilnehmenden des Workshops findet kaum eine systematische Datenerfassung statt, und dementsprechend werden dort zurzeit Metadatenstandards – wenn überhaupt – nur in eingeschränkter Form verwendet. Beispielsweise hat das House of Finance Datenbestände, aber keine Tools zur standardisierten Daten-Dokumentation. Das DIW Berlin und die Uni Bielefeld nutzen DDI-Metadaten in begrenztem Umfang. GESIS verwendet DDI z.B. im Datenbestandskatalog (DBK) oder datorium und berücksichtigt diesen Metadatenstandard in weiteren Entwicklungen.

Die Anwendung von Metadatenstandards wird von allen Beteiligten angestrebt und als unerlässlich betrachtet. Da mittlerweile eine Vielzahl von Metadatenstandards existiert, muss ein gemeinsamer Nenner gefunden werden, mit dem die sozial- und wirtschaftswissenschaftlichen Daten verschiedener Institute einheitlich beschrieben und in SowiDataNet miteinander verknüpft werden können. Da die meisten Institute bereits mehr oder weniger mit dem disziplinspezifischen Metadaten-Standard DDI der Sozialwissenschaften vertraut sind, wird dieser als Basis verwendet und mit notwendigen Modifikationen bzw. Erweiterungen versehen. Die Kompatibilität z.B. mit DublinCore, DataCite oder da|ra ist dadurch gewährleistet, und der Datenaustausch mit anderen Systemen wird dadurch wesentlich erleichtert.

Wann immer möglich wird die Verwendung von kontrollierten Vokabularen und Klassifikationssystemen angestrebt. Dies sorgt für bessere Suchergebnisse und ermöglicht die Vergleichbarkeit von Datensätzen.

Wie könnten Zugriffsrechte für SowiDataNet aussehen?

Wünschenswert und notwendig sind gestaffelte Zugangsklassen auf die Forschungsdaten, die direkt über SowiDataNet herunterladbar sind. Für Daten, welche nur lokal über die Institutionen abrufbar sind, müssen die institutsspezifischen Zugriffsrechte gelten. Die dazugehörigen Metadaten werden allerdings in jedem Fall veröffentlicht.

Dem Vorschlag für folgende Zugangsklassen innerhalb von SowiDataNet wurde zugestimmt:

- a) Frei zugänglich
- b) Eingeschränkt zugänglich (nur nach Anfrage bei dem Institut oder Primärforscher)
- c) Zugang nach Embargofrist

d) Kein Zugang

Welche Recherchemöglichkeiten sollte SowiDataNet bieten?

Eine übergreifende Suche sollte einerseits in den Metadaten von SowiDataNet möglich sein. Andererseits müssten mehrere Filter implementiert werden, um die Suche nach bestimmten Kriterien eingrenzen zu können. So sollte es zum Beispiel möglich sein, nur nach Daten von bestimmten Institutionen zu recherchieren. Weitere wichtige Filter sind darüber hinaus:

- Primärforscher/in
- Jahr der Veröffentlichung
- Zugangsklasse
- Direkter Daten-Zugang über SowiDataNet/indirekter Zugang über Institution
- Fachgebiet
- Untersuchungsgebiet
- Grundgesamtheit
- Erhebungsverfahren

Außerdem sollte innerhalb der Dateien (z.B. SPSS) selbst nach Stichworten recherchiert werden können. Dies ist allerdings nur in Dateien möglich, die in SowiDataNet gespeichert sind und auf die direkt zugegriffen werden kann. Dateien, die in den Institutionen selbst archiviert werden, können nicht durchsucht werden. Es muss geprüft werden, welche Dateiformate durchsucht werden können, da nicht jedes Format dafür geeignet ist.

Wünschenswert sind auch statistische Angaben zu Publikationen, die mit Hilfe von Forschungsdaten aus SowiDataNet veröffentlicht wurden. Die Auflistung der entsprechenden Publikationstitel sollte möglich sein.

Die Forschungsdaten mit freier Zugänglichkeit sollten auch als Linked Open Data verfügbar sein. In diesem Fall müssen die Daten öffentlich zugänglich sein, die Ergebnisse über Google aufgelistet werden und auch über andere Repositorien recherchierbar sein.

Sonstiges

Um den Instituten bereits vor der Fertigstellung von SowiDataNet Hilfestellungen für die Forschungsdatendokumentation zu geben, wäre es sinnvoll, auf der SowiDataNet-Webseite entsprechende Informationen zu veröffentlichen. Mit diesen Arbeiten kann ab dem 01.08.2014 begonnen werden. Folgende Informationen sollten z.B. aufgenommen werden:

- Hinweise zu Metadatenstandards
- Modellhafte Workflows innerhalb des Forschungsdatenzklus
- Anleitungen zur Einrichtung von Repositorien
- Rechte/Lizenz-Management

Darüber hinaus ist eine Qualitätskontrolle der eingestellten Forschungsprojekte durch die Institutionen erwünscht, um eine höchstmögliche Qualität der Forschungsdaten-Dokumentationen zu erreichen und somit die Akzeptanz von SowiDataNet durch die Nutzer/innen und wissenschaftlichen Einrichtungen zu erreichen. Es bleibt zu diskutieren, ob und wie dies in der praktischen Umsetzung aus-

sehen kann und wie die Qualitätskontrolle der Forschungsprojekte von Forscher/innen ohne institutionelle Anbindung erfolgen könnte.

Übersicht der Anforderungen an das Projekt – Tisch 2

- Die Anwendung von Metadatenstandards wird als unerlässlich betrachtet. Da verschiedene existieren, ist ein gemeinsamer Nenner zu finden (DDI als Basis, auch um Kompatibilität z.B. mit DublinCore, DataCite oder da|ra zu gewährleisten)
 - Die Verwendung kontrollierter Vokabulare u. Klassifikationssysteme ist anzustreben.
 - Die Metadaten sind in jedem Fall zu veröffentlichen; für Datensätze sind hingegen gestaffelte Zugriffsrechte notwendig.
 - Die Recherche sollte übergreifend möglich und durch Filter eingrenzbar sein.
 - Auch eine Stichwortsuche innerhalb der Dateien, die auf SowiDataNet gespeichert sind, wäre wünschenswert.
 - Wünschenswert sind auch Verweise auf Publikationstitel, die unter Verwendung der Forschungsdaten aus SowiDataNet veröffentlicht wurden.
 - Forschungsdaten mit freier Zugänglichkeit sollten als Linked Open Data verfügbar sein.
 - Die Qualitätskontrolle der eingestellten Forschungsdaten sollte seitens der Institutionen erfolgen. Zu klären bleibt wie eine Qualitätskontrolle für Daten von Forschenden ohne Anbindung zu realisieren ist.
 - Bereits vor endgültiger Fertigstellung von SowiDataNet wäre es sinnvoll auf der Webseite wichtige Informationen zu veröffentlichen (Metadatenstandards, Workflows ...).
-

4. Anforderungen an eine zukünftige Forschungsdateninfrastruktur

Am Themen-Tisch 3 wurden Anforderungen an eine zukünftige Forschungsdateninfrastruktur diskutiert. Als Kernanforderung hat sich herauskristallisiert, dass SowiDataNet vor allem einfach bedienbar sein und die Sichtbarkeit der Forschungsdaten erhöhen soll. In ihrer Rolle als Datengeber/innen muss es den Forschenden so einfach wie möglich gemacht werden, ihre Forschungsdaten aus ihrer jeweiligen Arbeitsumgebung abzulegen, und zwar sowohl unter dem Aspekt der Bedienbarkeit (Usability) als auch unter dem Aspekt des zeitlichen Aufwands. Als Datensuchende erwarten die Forschenden eine Vielzahl an Informationen, die es ihnen erleichtert, den Nutzen eines Forschungsdatensatzes für ihre Arbeit abschätzen zu können. Damit ist das Spannungsfeld skizziert, dem sich SowiDataNet gegenüber sieht. Einerseits müsste eine Vielzahl an deskriptiven Metadaten erfasst werden, um den Anforderungen an die Suchbarkeit gerecht werden zu können. Andererseits sinkt die Akzeptanz der Forschenden als Datengeber/innen, wenn sie viele Metadaten (händisch) erfassen müssen. Im Folgenden werden diese und weitere Punkte aufgelistet, die diskutiert oder zumindest angesprochen wurden.

Provenance

Unter „Provenance“ wird allgemein die Dokumentation der Herkunft von Daten verstanden. Diese Informationen sollten mitgeführt werden. Darunter kann zum einen der Entstehungskontext verstanden werden (z.B. ein bestimmtes Drittmittelprojekt), aber auch die Information ob es sich um Rohdaten oder Scientific-Use-Files handelt (bzw. bei Nicht-Verfügbarkeit der Rohdaten zumindest der pauschale Hinweis, dass die vorliegenden Daten aus Rohdaten hervorgegangen sind).

Versionierung

Diskutiert wurden verschiedene Dimensionen der Versionierung von Forschungsdaten (zeitlich, nach Anonymisierungsgrad, nach Umfang, logisch: Optimierung/Korrektur von Forschungsdaten, Versionierung auf der Ebene ganzer Datensets oder auch von einzelnen Bestandteilen wie etwa Variablen). Unstrittig war, dass eine zeitliche Versionierung ermöglicht bzw. gefordert werden muss. Uneinigkeit herrschte darüber, ob „Vorschriften“ für die Versionierung von Forschungsdaten in ihren verschiedenen Ausprägungsformen entwickelt werden sollten. Zwar sei es sinnvoll, die aus Datenschutzgründen unterschiedlichen Ausprägungsformen von Forschungsdaten (Rohdaten, Scientific-Use-Files, Public-Use-Files etc.) bei der Versionierung zu berücksichtigen und dies bei SowiDataNet zu unterstützen, zugleich wurde jedoch angemerkt, dass insbesondere Einzelforschende und kleinere Projekte – mit hin die Kernzielgruppen von SowiDataNet – mit der Anonymisierung und Zugangsgeschichte (Versionierungen) überfordert seien und diese nicht liefern könnten. Der Prozess, wie die Forschenden von den Rohdaten zum End-File (Scientific-Use-File) gekommen sind, sollte dokumentiert werden und nachvollziehbar sein (siehe hierzu auch den Abschnitt zu „Provenance“), wobei dies als Möglichkeit angelegt, aber nicht erzwungen werden sollte.

Suchbarkeit/Nutzungsmöglichkeiten

Gewünscht wurden ausführliche Informationen zu den hinterlegten Datensätzen. Für die Suche und Beurteilung von Forschungsdatensätzen seien ausführliche Metadatenbeschreibungen nötig (Welche Variablen sind enthalten? Welche Skala wurde benutzt?). Zentral seien auch Aussagen über die Nutzungsmöglichkeiten (Kann ich auf die Daten wirklich zugreifen und wenn ja, wie?).

Ease of use

Sowohl bei der Suche als auch bei der Ablage (Archivierung) von Forschungsdaten ist „Einfachheit“ ein zentrales Kriterium für die Akzeptanz der Anwendung. Es muss den Forschenden als Datengeber/innen so einfach wie möglich gemacht werden ihre Daten zu archivieren. Eine einfache Übernahme der Daten aus internen Systemen und die einfache Weiterleitung in externe Systeme sollten gewährleistet werden. Normdaten sollten für eine möglichst präzise Beschreibung hinterlegt werden und aus z.B. einem Drop-down-Menu für Forschende auswählbar sein, wenn sie ihre Forschungsdaten ablegen. Bei dem Design der Anwendung sollten Usability-Aspekte berücksichtigt werden.

Rechtemanagement und Datenschutzaspekte

Was ist hinsichtlich des Datenschutzes, der Persönlichkeitsrechte oder in Bezug auf Lizenzfragen zu beachten? Angeregt wurde, Mustervorlagen für Datenschutzerklärungen für Datenerhebungen anzubieten.

Qualitätsprüfung

Die abgelegten Forschungsdaten sollten einer Qualitätsprüfung nach formalen – nicht nach inhaltlichen – Kriterien unterliegen, z.B. im Hinblick auf Vollständigkeit oder Konsistenz. Evtl. könnten die Datengeber/innen auch mittels einer Checkliste abgefragt werden, ob sie die Kriterien ihrerseits bereits abgeprüft haben. Eine Kennzeichnung, ob der hinterlegte Forschungsdatensatz valide sei, wäre

zumindest bei der Beurteilung der Daten hilfreich. Das Merkmal „valide“ als Suchkriterium bzw. als Filter aufzunehmen, wurde hingegen skeptisch beurteilt.

Wissenstransfer

Angeregt wurde, SowiDataNet auch als Wissenstransferplattform zu nutzen. So sei es hilfreich, wenn die Forschenden z.B. ihre Erkenntnisse zu methodischen Fragen der Nutzung von Datenbanken, wie z.B. Datastream oder Amadeus, teilen würden. Auch Informationen zur Verlinkung der Daten aus unterschiedlichen (kommerziellen) Datenbanken (Link-Tables) wären hilfreich. Eine weitere Option wäre, Datenpaper zur Beschreibung der hinterlegten Forschungsdaten mit anzubieten.

DOI

Die DOI-Vergabe für Forschungsdaten wurde quasi stillschweigend vorausgesetzt. Diskutiert wurde jedoch, auf welcher Granularitätsstufe DOIs vergeben werden sollten. Die DOIs sollten nicht zu granular vergeben werden, um eine vollständige Zuordnung der jeweiligen Forschungsdaten zu den Publikationen zu ermöglichen.

Sonstiges

Gefragt wurde, ob recherchierte bzw. gekaufte Daten bei SowiDataNet gespeichert werden könnten. Ggf. sollte eine Option für das Tagging durch Nutzer/innen angeboten und hierzu fachlich geeignete Klassifikationen zur Beschreibung der Forschungsdaten hinterlegt werden.

Übersicht der Anforderungen an das Projekt – Tisch 3

- SowiDataNet soll durch die Datengeber/innen sowohl hinsichtlich der Bedienbarkeit (Usability) als auch unter dem Aspekt des zeitlichen Aufwands möglichst einfach nutzbar sein. „Einfachheit“ ist ein zentrales Kriterium für die Akzeptanz der Anwendung!
- Gleichzeitig bestehen hohe Anforderungen hinsichtlich der Metadatenbeschreibung, Dokumentation zu den hinterlegten Datensätzen, respektive der Recherchemöglichkeiten → **Spannungsfeld – Kompromiss notwendig! (s. Abschnitt 2)**

Weitere Punkte:

- Provenance-Informationen sollten mitgeführt werden
 - Zeitliche Versionierung muss ermöglicht werden
 - Der Prozess von den Rohdaten zum End-File sollte dokumentiert und nachvollziehbar sein (jedoch kein Muss-Kriterium).
 - Angaben zur Nutzungsmöglichkeit (Zugriff auf Daten? Wenn ja, wie?) sind zentral!
 - Angeregt wurde, Mustervorlagen für Datenschutzerklärungen anzubieten
 - Die Qualitätsprüfung der Forschungsdaten sollte nach formalen – nicht nach inhaltlichen – Kriterien erfolgen.
 - Angeregt wurde SowiDataNet auch als Wissenstransferplattform zu nutzen.
 - Die DOI-Vergabe für Forschungsdaten wird vorausgesetzt.
 - Zu klären ist, ob recherchierte bzw. gekaufte Daten bei SowiDataNet gespeichert werden können.
-

5. Hemmnisse für die Nutzung seitens der Forscher/innen

In den Gesprächsrunden an Tisch 4 wurden sowohl Hemmnisse diskutiert, die der Nutzung bestehender und zukünftiger Forschungsdateninfrastrukturen seitens der Datenproduzent/innen entgegenstehen (können), als auch mögliche Perspektiven zum Abbau solcher Nutzungsbarrieren. Im Folgenden werden zunächst die Inhalte der beiden Diskussionsstränge zusammengefasst, anschließend werden Anforderungen für das Projekt dargestellt.

Nutzungsbarrieren seitens der Forschenden als Datenproduzent/innen

In der Diskussion über die Nutzungsbarrieren zeichneten sich zunächst zwei wesentliche Aspekte ab: Einerseits ging es um bestehende Bedenken bzw. Unsicherheiten der Forschenden, andererseits wurde auf mangelnde Ressourcen hingewiesen. In der ersten Gruppe wurde mehrfach und mit Nachdruck betont, dass seitens der Forschenden ein klarer und nachvollziehbarer Anspruch auf die Erstnutzung ihrer Daten besteht. Dies trifft insbesondere auf den Fall eigener Primärerhebungen zu. Eine zu frühzeitige Veröffentlichung der Daten stehe den eigenen Verwertungsansprüchen entgegen. Befürchtet wird, dass andere Forscher/innen „die Früchte der eigenen Arbeit ernten“. Hinzu kommt die Befürchtung, dass die Nachnutzung der Daten eine Aufdeckung von Fehlern in den eigenen Forschungsergebnissen und damit einen für das Wissenschaftssystem schwerwiegenden Reputationsverlust nach sich ziehen kann.

Zudem bestehen Unsicherheiten bezüglich des Datenschutzes und der Anonymisierung von Daten. So wird die Weitergabe der Daten an Dritte bspw. in Datenschutzerklärungen vieler Primärerhebungen ausgeschlossen. Selbst wenn sich diese Erklärungen rein datenschutzrechtlich nur auf personenbezogene Daten erstrecken, werden Auswirkungen auf die Validität der Untersuchungsergebnisse befürchtet (bspw. durch einen geringeren Rücklauf in Surveys), wenn Datenschutzerklärungen Hinweise auf die zukünftige Weitergabe der Daten enthielten. Die derzeit wachsende gesellschaftliche Sensibilität bzgl. der Nutzung und Weitergabe von Daten spielt hierbei sicher ebenfalls eine Rolle. Auch bei „heiklen Themen“ ist das Vertrauen der Befragten in die Forscher/innen eine wichtige Grundlage, um valide Angaben zu erhalten. Unsicherheiten bestehen zudem auch in Bezug auf rechtliche Aspekte, wie etwa hinsichtlich des Umgangs mit Nutzungslizenzen oder der Verwendung internationaler Daten.

In Bezug auf die Ressourcenfrage wurden v.a. die aktuellen Entwicklungstrends in den Rahmenbedingungen der Forschung angesprochen. Forschungsprojekte werden derzeit in immer kürzeren Projektlaufzeiten umgesetzt und mit knapperen finanziellen Mitteln ausgestattet. Die Folge ist ein Mangel an zeitlichen bzw. personellen Ressourcen. Immer häufiger sind es dabei sehr junge Forschende, die maßgeblich die Projektarbeit, und dabei vor allem die Datenerhebungen übernehmen. Letztlich wird somit die Befürchtung geäußert, dass neue Anforderungen des Datenmanagements mit einer weiteren Erhöhung der Arbeitsbelastung einhergehen und dass die zusätzliche Arbeit „am Ende doch bei den Wissenschaftler/innen hängen bleibt“.

In der Wahrnehmung der Forschenden reiht sich das Datenmanagement ein in eine Zunahme „administrativer“ Aufgaben: Projektanträge, Projektberichte und Mittelabrechnungen. Diese Tätigkeiten sind wenig beliebt: „Man möchte ja forschen und nicht Verwaltungskram erledigen.“ Ausdrücklich wurde erwähnt, dass die Eingabe umfangreicher Metadaten eher abschreckend wirke und als sehr

zeitaufwändig wahrgenommen wird. Schließlich wurde die Frage gestellt, ob sich der „ganze Aufwand überhaupt lohne“ bzw. ob sich durch Zahlen belegen ließe, dass die Daten tatsächlich in hohem Maße nachgenutzt werden. Vermutet wird, dass die Daten aufgrund ihrer sinkenden Halbwertszeit heutzutage sehr schnell an Bedeutung verlieren. Nicht vernachlässigt werden sollte schließlich die Frage der Ressourcenverteilung innerhalb von Instituten. Wenn (mehr) Geld für Infrastrukturprojekte ausgegeben wird, bspw. im Rahmen einer zukünftigen Dateninfrastruktur, fehlt das Geld an anderer Stelle, z.B. in der Forschung. Die Forschenden befürchten entsprechend, dass Mehrinvestitionen im Datenmanagement letztlich vom Forschungsbudget ihrer Institute abgehen.

Zwei weitere Punkte wurden darüber hinaus als wirkmächtige Hemmnisse bzw. Barrieren benannt. Erstens der mangelnde Stand des Wissens und der Aufklärung über die Möglichkeiten sowie den Nutzen des Datenmanagements. Dabei gehe es zunächst weniger um spezielle Aspekte wie die Vergabe von DOIs, sondern v.a. um ein fehlendes Basiswissen, das bereits zu Beginn von Projekten benötigt wird. Hierauf, so wurde in großer Übereinstimmung festgehalten, müsse insbesondere im Zuge der Einführung einer neuen Infrastruktur für Forschungsdaten ein gesondertes Augenmerk gelegt werden. Zweitens wurde betont, dass es schlichtweg an einem Belohnungssystem mangle, das heißt an individuellen Anreizen für die Forschenden: „Was hat man als Forschende/r davon?“ Maßgeblicher Grund hierfür ist die starke Publikationszentriertheit des Wissenschaftssystems. Veröffentlichungen neuer Ergebnisse bspw. in Form von Journal-Artikeln generieren die höchste Anerkennung. Die im Vorfeld geleisteten Arbeiten, gerade auch in Bezug auf die Datenerhebung, sind hingegen weniger sichtbar und die damit verbundenen Aufgaben wie etwa das Datenmanagement gelten daher nach wie vor eher als Pflicht denn als Kür.

Perspektiven zum Abbau der Nutzungsbarrieren

In der Diskussion über mögliche Perspektiven zum Abbau von Nutzungsbarrieren wurden zwei verschiedene Richtungen erörtert: erstens mögliche Anreize für die Forschenden und zweitens der externe Druck v.a. von Seiten der Fördereinrichtungen.⁴

Um Anreize zu setzen, wird in der aktuellen Debatte häufig auf die Zitation von Forschungsdaten verwiesen. Wenn diese auch auf absehbare Zeit nicht den Status wissenschaftlicher Fachpublikationen erlangen dürften, wäre auf diesem Wege zumindest etwas mehr Anerkennung für Datenerhebung und -management als eigenständige wissenschaftliche Leistungen zu ermöglichen. Denkbar wäre auch die Veröffentlichung als eine Art „Metadaten-Paper“, welches dann auch in den CV der Forschenden Aufnahme finden könnte. In Bezug auf die Veröffentlichung von Forschungsdaten wurden mehrere konkrete Anforderungen formuliert: Es muss möglich sein Embargofristen zu vergeben, Daten nur nach Einzelfallprüfung freizugeben und Informationen in unterschiedlichem Umfang zu veröffentlichen (bspw. zunächst nur Metadaten, später ausgewählte und letztlich vollständige Daten im Sinne einer Abstufung: Was muss? → Was sollte? → Was wäre optimal?).

Zur Zitation wurde allerdings einschränkend angemerkt, dass die Zitierfähigkeit von Forschungsdaten noch ein sehr neues, wenig verbreitetes und für Forschende auch noch recht abstraktes Verfahren

⁴ vgl. Andreoli-Versbach, Patrick/Mueller-Langer, Frank (2014): Open access to data: An ideal professed but not practiced. In: Research Policy, DOI: 10.1016/j.respol.2014.04.008
<http://www.sciencedirect.com/science/article/pii/S0048733314000699>.

darstellt. Der konkrete Nutzen wird sich daher wohl erst zukünftig zeigen. Entsprechend dürfe die momentane Wirksamkeit als Anreiz nicht überschätzt werden. Auch der Verweis auf forschungsethische Gesichtspunkte („gute wissenschaftliche Praxis“) mag seine Berechtigung haben, die Wirksamkeit als Anreiz wurde jedoch durch die Teilnehmenden ebenfalls zumindest z.T. bezweifelt. Wichtig wäre es daher vor allem, verstärkt zu vermitteln, dass die Arbeit der Forschenden selbst von der Nutzung einer Forschungsdateninfrastruktur profitieren kann. Ob in Bezug auf die längerfristige Arbeit mit den eigenen Forschungsdaten, die Arbeit in veränderlichen Projektteams oder die Nutzung in Folgeprojekten: Ein gutes FDM kann den Arbeitsaufwand verringern, die Arbeit mit den Daten erleichtern und helfen, die Qualität der Ergebnisse zu erhöhen bzw. Fehlerquellen zu reduzieren. Bessere Recherchemöglichkeiten sowie die potentielle Nutzung bisher kaum zugänglicher Datenbestände stellen weitere Anreize dar.

Neben der Perspektive individueller Anreize stand vor allem der externe Druck im Mittelpunkt der Tischgespräche. Dabei wurde der Fokus v.a. auf die Fördereinrichtungen gelegt (DFG, EU, Stiftungen). Regeln und Vorschriften, die mit der Vergabe finanzieller Mittel verbunden sind, wurden als wirkmächtiger Hebel eingeschätzt. Es wäre sogar vorstellbar, dass die letzte Rate einer Projektfinanzierung erst gezahlt wird, wenn die Forschungsdaten dokumentiert und veröffentlicht sind. An den jeweiligen Instituten könnten die Drittmittelstellen hierbei eine unterstützende Funktion übernehmen. Des Weiteren wurde eine Integration des Datenmanagements der Forschenden in das Evaluierungsverfahren der Leibniz-Gemeinschaft vorgeschlagen, was zusätzlich den Druck auf die Institute erhöhen würde. Die Position der Fördereinrichtungen ist bisher jedoch noch wenig eindeutig. Die DFG spricht bspw. bislang lediglich Empfehlungen aus, während in Projektanträgen bereits Gelder für das Datenmanagement einkalkuliert werden können.

Das notwendige Verhältnis zwischen Anreizen und externem Druck wurde am Thementisch 4 durchaus kontrovers diskutiert: einerseits gab es die Einschätzung, dass die Entwicklung derzeit bereits zu stark in Richtung einer Erhöhung des Drucks durch die Fördereinrichtungen gehe, es jedoch v.a. mehr Anreize als Motivation für die Forschenden brauche. Zu viel Druck könne, so die Befürchtung, auch Abwehrhaltungen verstärken! Andererseits wurde die Ansicht vertreten, dass es „ohne Druck nicht gehen wird“, dieser also weiter erhöht werden müsse. Einig waren sich die Teilnehmer/innen jedoch in der Einschätzung, dass nur ein Mix aus Anreizen und verpflichtenden Regeln („Zuckerbrot und Peitsche“) zu dauerhaften Veränderungen im Umgang mit Forschungsdaten führen könne.

Ferner wurde mehrfach die Rolle der Institute hervorgehoben. Diese könnten einerseits durch Serviceangebote die Arbeit der Forschenden unterstützen und damit zur Entlastung gerade im personell und finanziell knapp aufgestellten Projektkontext beitragen. Die „Betreuung vor Ort“ durch infrastrukturelle Serviceeinrichtungen müsse dabei administrativ angelegt und verstetigt sein, sodass die Forschenden hier Kontinuität und Verlässlichkeit erwarten können. Zudem sei eine lokale Schnittstelle ein wichtiges Feature einer künftigen Forschungsdateninfrastruktur. Wenn Daten im institutionellen Rahmen erfasst, dokumentiert und eingestellt werden – auch wenn sie dabei in eine institutsübergreifende Infrastruktur eingehen – könne dies als vertrauensbildende Maßnahme den Bedenken der Forschenden entgegenwirken.

Abschließend wurden mögliche Perspektiven zum Abbau des Wissens-/Aufklärungsmangels besprochen: Als besondere Herausforderung wird gesehen, dass mit der Einführung bzw. Veränderung von

Infrastrukturangeboten gleichzeitig auch das benötigte Wissen vermittelt werden muss: sowohl Wissen über die Existenz des neuen Angebots (Verbreitung) und die Möglichkeiten der Nutzung als auch generell über den Nutzen des FDM. Eine frühe Vermittlung der Inhalte bereits im Rahmen des Studiums wäre sicherlich wünschenswert. Später geht es v.a. um die Vermittlung von Basiswissen (einfache Grundlagen des Datenmanagements: Dateibenennung, Versionierung etc.). Solches Wissen sollte verstärkt in Form von praxisorientierten FAQs, Handreichungen, Leitlinien, einfachen Guidelines etc. bereitgestellt und auch online verfügbar gemacht werden. Online-Tools könnten sogar zur Erstellung von FDM-Plänen genutzt werden. Bei der Wissensvermittlung ist ein früher Kontakt zu den Forschenden wünschenswert, um bereits zu Beginn der Forschungsprojekte die „richtigen Weichen“ zu stellen, aber auch um die Forschenden mitzunehmen und nicht zum Projektende mit „überraschenden“ Arbeitsaufgaben zu überfordern.

Anforderungen für das Projekt

Die diskutierten Hemmnisse umfassen ein ausgedehntes Spektrum; genannt wurden u.a.: Bedenken und Unsicherheiten, mangelnde Ressourcen, mangelndes Wissen sowie fehlende Anreize. In Bezug auf die Bedenken/Unsicherheiten und das fehlende Wissen ist der allg. Auftrag an das Projekt v.a. darin zu sehen, dass die Einführung einer neuen Infrastruktur Hand in Hand gehen sollte mit der Verbreitung von Wissen auf zwei Ebenen: der Kenntnis darüber, dass es ein neues Angebot gibt und wie es genutzt werden kann, aber auch von Wissen, dass allg. über die Möglichkeiten und den Nutzen eines nachhaltigen FDM aufklärt. Dies könnte im Projekt durch die Bereitstellung von FAQs, Leitfäden, Checklisten etc. auf der Homepage umgesetzt werden. Den Ansprüchen auf Erstnutzung der Daten kann hingegen durch konkrete Embargofristen und durch die Möglichkeit, Metadaten und Datensätze in verschiedenen Stufen zu veröffentlichen, begegnet werden.

Betont wurde zudem die Notwendigkeit der lokalen Anbindung einer zukünftigen Infrastruktur. Einerseits hilft dies, Bedenken und Unsicherheiten abzubauen, andererseits kann auf der institutionellen Ebene auch ein Beitrag zur Entlastung der Forschenden geleistet werden (Ressourcenmangel). Die Diskussion über zukünftig zu verändernde Regularien und insbesondere die mögliche Erhöhung des Drucks seitens der Fördereinrichtungen betrifft schließlich die weitere Entwicklung des gesamten Forschungssystems und kann daher vom Projekt selbst sicher nur begrenzt aktiv mitgestaltet werden. Diese Perspektive sowie die derzeitigen Trends in den Forschungsbedingungen sind aber sicherlich als wichtige Rahmenbedingungen mitzudenken, gerade auch, wenn es um zusätzliche Arbeitsanforderungen an die Forschenden geht.

Hemmnisse, Nutzungsbarrieren	Anforderungen an das Projekt – Tisch 4
<i>Bedenken und Unsicherheiten</i>	
- Anspruch auf Erstnutzung	➤ Embargofristen, gestaffelte Zugangsklassen, gestaffelte Veröffentlichung von Informationen
- Möglichkeit der Aufdeckung von Fehlern	➤ ??? (...Forschungsethik)
- Datenschutzfragen	➤ ...sind zu klären, später in FAQs behandeln
- Lizenzfragen	➤ ...sind zu klären, später in FAQs behandeln
- Findet überhaupt eine Nachnutzung statt?	➤ langfristig: Nachweis über Nutzungsstatistik möglich? ➤ Allg.: lokale Anbindung der Infrastruktur als „vertrauensbildende Maßnahme“ beugt Unsicherheiten vor
<i>Ressourcenmangel</i>	
- Personell, finanziell (Mehrarbeit)	➤ Serviceangebote „vor Ort“ können Forschenden entlasten ➤ Geld für DM bereits in Projektanträgen einkalkulieren
- Institutsebene: mehr Geld für DM/Infrastruktur fehlt in der Forschung	➤ ???
<i>Fehlendes Wissen</i>	
	➤ (frühe Vermittlung bereits in Ausbildung) ➤ (später in Weiterbildungen v.a. Vermittlung praxisorientierten Basiswissens) ➤ Leitlinien, FAQs auf der Projekt-Homepage! ➤ „Zugpferde“, Multiplikatorinnen und Multiplikatoren, die vorneweg gehen (z.B. Institute?) ➤ Online-Tools für DM-Plan Erstellung (Uni Bielefeld)
<i>Fehlende Anreize</i>	
	➤ (Anerkennung durch Zitation von Forschungsdaten) ➤ „Metadaten-Paper“ für CV ➤ (Verweis auf Forschungsethik) ➤ Nutzung der Infrastruktur bereits als Teil des projektbezogenen Datenmanagements ➤ Neue Recherchemöglichkeiten
<i>Fehlende Regeln, Druck von „oben“</i>	
	➤ (klare Regeln statt Empfehlungen) ➤ (mehr Druck von Seiten der Fördereinrichtungen) ➤ (Integration in Evaluierungsverfahren)

6. Abschlussrunde und offene Fragen für das Projekt

Zum Abschluss des Workshops kamen alle Teilnehmenden erneut im Plenum zusammen. Die Moderator/innen der Thementische präsentierten die zentralen Ergebnisse der Tischgespräche, und in einer abschließenden Runde konnten weitere offene Fragen und Anregungen der Teilnehmenden aufgenommen werden. Folgende Themen wurden dabei angesprochen: Für die Zielstellung eines Workshops zur Anforderungsanalyse war die Gruppe der Forschenden als zentrale Zielgruppe leider unterrepräsentiert. Dies ist sicher bedauerlich, spiegelt aber auch die derzeitige Interessenslage der Forscher/innen an der Thematik wieder; so war es trotz entsprechender Bemühungen im Vorfeld des

Workshops auch nicht gelungen, mehr Forschende für eine Workshop-Teilnahme zu gewinnen. Daraus schloss sich der Hinweis an, dass es für das Projekt unerlässlich sei, über einen guten Argumentationsrahmen zu verfügen, der den Forschenden den Nutzen der künftigen Infrastruktur überzeugend darlegt. Weitere Anmerkungen bezogen sich auf die Nachnutzbarkeit von Forschungsdaten. Hierfür wurden einerseits umfangreiche Mindeststandards in der Datenbeschreibung gefordert, denn nur durch eine umfangreiche Dokumentation werde die Nutzung für die Forschenden attraktiv. Worauf jedoch der Einwand folgte, dass die Hürden für die Datengeber/innen eher gering zu halten seien (also gerade keine umfangreichen Pflichtangaben), um die Akzeptanz auf der Datengeberseite zu erhöhen. Es wurde zudem gefragt, wer die primäre Zielgruppe des Projekts sein soll, individuelle Forschenden oder Institutionen? Skepsis wurde dahingehend geäußert, dass ein Spagat zwischen beiden Ausrichtungen in einem Projekt zu leisten sei. Die Beteiligung von zwei Forschungsinstituten im Projektverbund spräche u.U. zunächst für einen institutionellen Zugang. Schließlich wurde gefragt, auf welchem Weg die Qualität der Daten gewährleistet werden soll. Das denkbare Spektrum reicht von einer Community-basierten Qualitätssicherung (z.B. durch Bewertung der hinterlegten Daten) über mittelbare (Vorgaben für Datengeber/innen) bis zu unmittelbaren (Review durch SowiDataNet) Verfahren der Qualitätssicherung durch die SowiDataNet-Betreibenden.

Vor dem Hintergrund der Tischgespräche und der Abschlussrunde lassen sich folgende zentrale offene Fragen für die weitere Projektarbeit festhalten:

Grundausrichtung:

- Wer soll die primäre Zielgruppe des Projekts sein: Einzelforschende vs. Institutionen? Ist beides zu leisten bzw. in einem Infrastrukturangebot integrierbar?
- Soll SowiDataNet eher eine zentrale Anlaufstelle für Recherchen nach Forschungsdaten sein und entsprechend intensiv „Harvesting“ betreiben oder primär als Repositorium den Schwerpunkt auf den Nachweis und, falls seitens der Datengeber/innen gewünscht, die Archivierung von Forschungsdaten legen (im letzten Falle jedoch zumindest harvestingfähig bleiben)?

Forschungsdaten:

- Welche Typen von Daten (Tabellen, Texte, Bilder usw.) sollen durch die Infrastruktur abgedeckt werden, bzw. ist ein formatoffener Ansatz erwünscht/machbar?
- Welchen Stellenwert haben „Verarbeitungsroutinen“ für Forschungsdaten (z.B. Syntax-Files)?
- Ausgehend von der Klärung der beiden vorherigen Fragen: In welchen Dateiformaten sollen die entsprechenden Informationsobjekte bereitgestellt werden können (falls nicht nur eine Archivierung durch Bitstream-Preservation angestrebt wird, sondern eine dauerhafte Nutzbarkeit der bereitgestellten Informationsobjekte sichergestellt werden soll)?
- Wie können Probleme der Daten-Versionierung mit Blick auf a) „lebende“ Datensätze (die fortgeschrieben werden) und b) Korrekturen vorhandener Daten gelöst werden? Wer ist für Aktualisierungen zuständig?
- Wie kann ein Kompromiss zwischen einerseits hohen Anforderungen an die Dokumentation zwecks guter Recherche-/Nachnutzungsmöglichkeiten und der Anforderung einer möglichst mit geringem Aufwand zu bewerkstellenden Hinterlegung von Daten durch die Produzent/innen andererseits (Akzeptanz) aussehen?

- Wie kann SowiDataNet die Qualität der über die Infrastruktur angebotenen Daten sichern (z.B. eigene Qualitätsprüfung, Prüfung durch Datengeber, Community-basierte Bewertung), und wie können Mechanismen der Qualitätssicherung nachvollziehbar für Nutzer/innen von SowiDataNet dargestellt werden?
- Metadatenschema DDI als Grundlage: Was sind Pflichtfelder, was sind empfohlene, was sind mögliche Felder?

Unterstützung u. Akzeptanz bei Datengeber/innen

- Inwiefern kann SowiDataNet dazu beitragen, Informationen und Kompetenzen in Sachen FDM in die Scientific Community zu vermitteln?
- Welche Argumente und Anreize überzeugen die Forschenden von Sinn und Nutzen der neuen Infrastruktur? Wie können diese in die Community vermittelt werden?
- Inwieweit kann den Datenproduzent/innen Arbeit abgenommen werden, um es ihnen so einfach wie möglich zu machen und eine hohe Akzeptanz zu erreichen, gleichzeitig aber die Anforderungen der Datennutzer/innen zu erfüllen? (Personal über Repositoryum? Lokale Service-Angebote der Institute?)