

DSZ-BO
Working Paper Series

Working Paper

Nr. 2

Datenmanagement
- Eine zentrale Voraussetzung für den Erfolg der
Data Sharing-Idee in den Sozialwissenschaften

Alexia Meyermann

Juli 2012

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

DSZ-BO

Working Paper Series

DSZ-BO Working Paper Series ist eine Schriftenreihe des Datenservicezentrum Betriebs- und Organisationsdaten (DSZ-BO) an der Universität Bielefeld. Darin erhalten Forscherinnen und Forscher die Möglichkeit ihre Arbeiten zu einem frühen Stadium einer breiten Öffentlichkeit vorzustellen. Veröffentlicht werden sollen Arbeiten im Bereich der empirischen, methodologischen oder theoretischen Organisationsforschung. Dies kann sowohl auf sozial- als auch wirtschaftswissenschaftliche oder psychologische Fragestellungen zutreffen beispielsweise aus den Bereichen der Arbeitsmarkt-, Gesundheits- oder Bildungsforschung und auch darüberhinausgehend.

Die DSZ-BO Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Sie können nicht über den Buchhandel, sondern nur online über das DSZ-BO bezogen werden.

Die Inhalte der DSZ-BO Working Papers stellen ausdrücklich die Meinung der jeweiligen Autorinnen und Autoren dar und nicht die des DSZ-BO.

DSZ-BO Working Paper Series
ISSN 12345678

Herausgeber:
Stefan Liebig, Leitung DSZ-BO
Martin Diewald, Leitung DSZ-BO

Kontakt:
Iris Braun
Tel.: 0521 106-6948 (Sekretariat: Regina Fischer)

Datenservicezentrum Betriebs- und Organisationsdaten
an der Universität Bielefeld
Postfach 10 01 31
33501 Bielefeld
dsz-bo@uni-bielefeld.de
www.uni-bielefeld.de/dsz-bo

Datenmanagement - Eine zentrale Voraussetzung für den Erfolg der Data Sharing-Idee in den Sozialwissenschaften

Alexia Meyermann

Abstract

Unter dem Begriff des *Datenmanagements* wird der systematische und planvolle Umgang mit Daten und deren Bearbeitung vor dem Hintergrund antizipierter Nutzungszwecke bezeichnet. Seit wenigen Jahren lässt sich in den Sozialwissenschaften international wie national ein Bedeutungsanstieg des Datenmanagements beobachten. So verlangen zunehmend mehr Forschungsförderer die Erstellung und Einreichung von Datenmanagementplänen und kontrollieren deren Umsetzung. Die aktuellen Diskussionen zum Thema Datenmanagement stehen in direktem Zusammenhang mit der steigenden Bedeutung, die Forschungsdaten im Allgemeinen – quasi als Output oder Produkt wissenschaftlicher Forschung – und Data Sharing im Besonderen, d.h. der Weitergabe und Weiternutzung von Forschungsdaten, zugewiesen wird. Der vorliegende Artikel vertritt die These, dass die erfolgreiche Realisierung von Data Sharing zentral von der Qualität des in den jeweiligen datenproduzierenden Forschungsprojekten angewandten Datenmanagements abhängt. Die Analyse der gegenwärtigen Situation zeigt jedoch, dass die gegenwärtigen Praktiken des Datenmanagements defizitär sind. Ungünstige Anreizstrukturen, fehlende Kenntnisse und eine unzureichende Ressourcenausstattung werden als Gründe hierfür diskutiert und auf notwendige Investitionen von Seiten aller Beteiligten, der Förderer, der Datenzentren sowie der Forscherinnen und Forscher hingewiesen.

Inhalt

1	EINLEITUNG	1
2	DATA SHARING IN DEN SOZIALWISSENSCHAFTEN	1
2.1	Data Sharing und die Vorteile der Sekundärforschung	1
2.2	Die nationalen und internationalen Rufe nach Data Sharing	3
2.3	Voraussetzungen erfolgreichen Data Sharing's: Das Verfügbarkeits- und das Nutzbarkeitsproblem und die Rolle der Datenzentren	4
3	DATENMANAGEMENT - ZUR HERSTELLUNG UND OPTIMIERUNG DER NUTZBARKEIT VON FORSCHUNGSDATEN	7
3.1	Datenmanagement	7
3.2	Teilbereiche des Datenmanagements und ihre Bedeutung für Data Sharing	9
3.2.1	Datendokumentation - Die Beschreibung der Daten mit Metadaten	9
3.2.2	Datensicherung	11
3.2.3	Rechtliche und Ethische Aspekte im Umgang mit Forschungsdaten	12
4	HERAUSFORDERUNGEN HINSICHTLICH EINER PROFESSIONALISIERUNG DES DATENMANAGEMENTS	14
5	AUSBLICK	18
6	LITERATUR	20
7	ANHANG	26

1 Einleitung

Seit Jahren steigen die Forderungen von Wissenschaftsorganisationen und Drittmittelgebern nach Data Sharing. Data Sharing - die Weitergabe und Weiternutzung von Daten durch Dritte - führt zu einer größtmöglichen Ausschöpfung bestehender Analysepotentiale und wird allgemein mit einer Reihe weiterer Vorteile verbunden. Zu nennen sind hier beispielsweise die Erhöhung der Objektivität und der Transparenz von Forschungsbefunden durch ein erhöhtes Maß an Replizierbarkeit. Nicht zuletzt kann Data Sharing zu einer stärkeren Legitimierung getätigter Investitionen in wissenschaftliche Forschung beitragen. Bei Betrachtung der Vorteile der Sekundärnutzung von Forschungsdaten ist es erstaunlich, wie wenig diese Form der empirischen Forschung derzeit noch praktiziert wird und welche geringe Bedeutung ihr in der heutigen Forschungspraxis zukommt. Diese als Ausgangspunkt nehmend zielt dieser Artikel darauf ab, einen möglicherweise zentralen erklärenden Faktor für die geringe Verbreitung von Data Sharing zu diskutieren: systematisches Datenmanagement.

Unter „systematischem Datenmanagement“ ist der gesteuerte und regulierte Umgang mit Forschungsdaten während des gesamten Forschungsprozesses und im Hinblick auf aktuelle und spätere Nutzungszwecke zu verstehen. Es wird die These vertreten, dass systematisches Datenmanagement für den Erfolg des Data Sharing zentral ist. Betrachtet man die bisherigen Aktivitäten seitens der Forschungsförderung zeigt sich aktuell eine interessante, den postulierten Zusammenhang nahelegende Entwicklung: Die bislang einseitige Fokussierung der Diskussion und Aktivitäten auf Data Sharing wird zunehmend ergänzt um solche das *Datenmanagement* selbst betreffende (vgl. van Eynden et al. 2011). Hinzu treten die Beobachtungen von Datenzentren, dass Data Sharing häufig weniger an der fehlenden Bereitschaft von Datenproduzenten zur Herausgabe ihrer Daten scheitert, als vielmehr an den fehlenden Kenntnissen und Ressourcen dieser im Bereich des Datenmanagements, denn Forscher sind oftmals „willing but unable“ (RIN 2008, 25). Datenmanagement spielt in der gegenwärtigen empirischen Sozialforschung keine große Rolle: Jenseits der Großprojekte, in welchen quantitative Massendatensätze produziert und angeboten werden wie dem SOEP oder NEPS¹, ist die Verbreitung systematischen Datenmanagements gering und dessen Qualität sehr unterschiedlich ausgeprägt.

Im Folgenden wird zunächst die Idee des Data Sharing vorgestellt und auf den wachsenden Druck seitens der Förderer hingewiesen. Als nächstes werden die Voraussetzungen für erfolgreiches Data Sharing diskutiert. Es wird argumentiert, dass Data Sharing gleichermaßen von der Verfügbarkeit von Daten als auch von deren Nutzbarkeit abhängt, und der Zusammenhang zwischen Data Sharing und Datenmanagement wird eingeführt. Im dritten Kapitel werden die zentralen Bestandteile von Datenmanagement (Datendokumentation, Datenaufbewahrung, rechtliche und ethische Aspekte) exemplarisch vorgestellt, und es wird jeweils auf bestehende, aktuelle Probleme hingewiesen.² Kapitel 4 fasst die Defizite im Hinblick einer Professionalisierung des Datenmanagements zusammen. Ausblickend wird auf die Bedeutung sämtlicher beteiligter Gruppen – den Forscherinnen und Forschern, den Förderern sowie den Datenzentren – hingewiesen.

2 Data Sharing in den Sozialwissenschaften

2.1 Data Sharing und die Vorteile der Sekundärforschung

Viele Jahre war es für empirische Sozialforscher üblich und die erste Wahl zur Beantwortung ihrer Forschungsfragen eigene Daten zu erheben, anstatt diese „nur“ sekundär zu nutzen. Dies wandelt sich seit einigen Jahrzehnten – in einigen Disziplinen stärker als in anderen.

1 SOEP: Sozio-ökonomisches Panel vgl. www.diw.de/soep ; NEPS: Nationales Bildungspanel, vgl. www.neps-data.de.

2 Es ist nicht Zweck dieses Artikels, eine Einführung in die Praxis des Datenmanagements zu geben. Auf entsprechende weiterführende Publikationen wird jedoch verwiesen.

Hierfür lassen sich mindestens zwei zentrale Ursachen anführen. *Zum einen* ist der Bestand an verfügbaren Daten gewachsen, und somit sind schlicht die Möglichkeiten gestiegen, Daten sekundär zu nutzen. Die Dateninfrastruktur innerhalb Deutschlands wird seit 2004 kontinuierlich unter dem Dach des Rat für Sozial- und Wirtschaftsdaten (RatSWD) ausgebaut (vgl. Habich et al. 2010). Es existieren die an Forschungsinstituten angesiedelten Datenzentren³, die hauseigene Daten zur Nutzung bereitstellen wie das Forschungsdatenzentrum (FDZ) des IAB, das des BIBB, die FDZ des Bundes und der Länder oder die FDZ einzelner Datensätze wie SOEP, SHARE oder PAIRFAM. Daneben gibt es Datenzentren, die Daten externer Forscherinnen und Forscher anbieten: das breit aufgestellte Datenzentrum German Microdata Lab der GESIS, das Datenzentrum PsychData für die Psychologie und stärker fachspezifisch ausgerichtete Datenzentren wie das FDZ-IQB für die Bildungsforschung, das IDSC am IZA für die internationale Arbeitsmarktforschung oder das DSZ-BO tätig für die empirische Organisationsforschung.⁴

Neben dem durch diese Entwicklungen stark gewachsenen, verfügbaren Datenbestand liegt der Anstieg der Sekundärforschung *zum anderen* an den Vorteilen, die damit selbst verbunden sind und für eine Nachrangigkeit von Primärforschung sprechen. Sekundäranalysen sind *erstens* ressourcenschonender als Primärforschungen, sie sind *zweitens* aufgrund forschungspraktischer Erwägungen häufig angebracht(er), *drittens* können sie einen Beitrag zur Qualität wissenschaftlicher Forschung leisten (vgl. zum Nutzen von Sekundäranalysen und Data Sharing z.B. King 1995, Dale 2004, Freese 2007, Seale 2009, Beagrie 2010: 26, van den Eynden et al. 2011: 3, Huschka et al. 2011, S.37ff.) und *viertens* erhöht sich die Zitationsrate des datenproduzierenden und Daten anbietenden Forschers. Die verschiedenen Aspekte werden im Folgenden näher erläutert.

(1) *Ressourcenschonende Forschung* bedeutet einen verantwortungsbewussten und möglichst effizienten Umgang mit Fördergeldern. Der „öffentliche-Gut“-Charakter von Forschungsdaten erhöht den Druck, diese der Allgemeinheit zur Re-Nutzung zur Verfügung zu stellen, um so den größtmöglichen Nutzen der Investitionen öffentlicher Mittel zu erzielen. Zudem können durch die nun mögliche, vollständigere Ausschöpfung vorhandener Analysepotentiale redundante Primärerhebungen vermieden, die Notwendigkeiten neuer Primärerhebungen begründet und neue Forschungsdesigns entsprechend der identifizierten Notwendigkeiten fokussiert werden, wenn Forschungslücken mittels Sekundäranalysen exakt(er) bestimmt wurden. Laut Weichselgartner (2011, 3) gehört zur Ausschöpfung vorhandener Analysepotentiale die Analyse von Daten unter neuen Fragestellungen oder mit anderen Analysemethoden. Neben den *finanziellen Erwägungen* hinsichtlich Amortisation und Effizienz eingesetzter Fördergelder gilt es, aufgrund *ethischer Erwägungen* redundante Primärerhebungen zu vermeiden. So sollten die Belastungen, denen Untersuchungsteilnehmer ausgesetzt werden, so gering wie möglich gehalten werden (vgl. Weichselgartner 2011, 3, vgl. Häder 2009, 8: „Datensparsamkeitsgebot“, §3a BDSG).⁵

(2) Zu den genannten Gründen für die steigende Bedeutung von Sekundäranalysen und damit des Data Sharing treten solche *forschungspraktischer Art*. Aufgrund aktueller

3 In Deutschland wird zwischen Forschungsdatenzentren und Datenservicezentren unterschieden (vgl. RatSWD, www.ratswd.de/dat/fdz.php). Da diese Unterscheidung hier inhaltlich nicht relevant ist, wird im Folgenden zusammenfassend von „Datenzentren“ gesprochen.

4 IAB steht für Institut für Arbeitsmarkt- und Berufsforschung, BIBB für Bundesinstitut für Berufsbildung, GESIS für Gesellschaft sozialwissenschaftlicher Infrastruktureinrichtungen (Leibniz-Institut für Sozialwissenschaften), IQB für Institut zur Qualitätsentwicklung im Bildungswesen, IDSC für International Data Service Center, IZA für Institut zur Zukunft der Arbeit, DSZ-BO für Datenservicezentrum Betriebs- und Organisationsdaten. PsychData – Forschungsdaten für die Psychologie ist am ZPID Leibniz-Zentrum für Psychologische Information und Dokumentation angesiedelt.

Für eine vollständige Übersicht der Datenzentren, siehe www.ratswd.de/dat/fdz.php

5 Weichselgartner nennt die Vorteile „Ökonomie“, „Exklusivität“, „Exhaustivität“, „Ethik“ und „Lehre, Kontrolle, Visibilität“ (2011, 3f.).

methodischer Herausforderungen und Entwicklungen (Mixed Methods, gesunkene Response Raten, unzureichende Sampling Frames) und eines gestiegenen methodischen Problembewusstseins⁶, sind die Anforderungen an die Qualität der Forschungsdaten auch in den Sozialwissenschaften gestiegen⁷. Daraus ergeben sich neue Anforderungen an die methodische Expertise und ein gesteigener Ressourcenbedarf in Primärforschungsprojekten. Diese können kaum im üblichen Forschungsalltag, von Einzelforschern erfüllt werden, für die die Datenerhebung nur einen Teilaspekt der gesamten Forschungstätigkeit ausmacht (neben Datenanalysen, Interpretationen, theoretischen Arbeiten u.v.m.).

(3) Nicht zuletzt bieten Sekundäranalysen neue Möglichkeiten der Validierung von Forschungsbefunden und erhöhen somit die Sicherheit wissenschaftlicher Erkenntnisse (*Wissenschaftlichkeit*). Dies erhöht oder festigt die Reputation des Wissenschaftssystems in der Gesellschaft (Freese 2007). Data Sharing kann insgesamt zum wissenschaftlichen Erkenntnisgewinn und Fortschritt beitragen (Key Perspectives Ltd. 2010, Haeussler 2010, NSF 2010, Huschka et al. 2011, 37f.). Es dient dazu, Vergleiche von Ergebnissen auf Basis verschiedener Datensätze durchzuführen, es eröffnet neue Möglichkeiten für Zeit- (präzisere Replizierbarkeit von Untersuchungen) oder Stichprobenvergleiche (vgl. auch Weichselgartner 2011, 3f.). Hinzu kommt, dass bestimmte Datenerhebungen schlicht nicht replizierbar sind, da sie einmalige, vergängliche Ereignisse messen, wie bspw. die deutsche Wiedervereinigung (Weichselgartner 2011). Unabhängig davon sind die Möglichkeiten zu Sekundäranalysen gerade für die vergleichsweise ressourcenschwachen Nachwuchswissenschaftler relevant und der Einsatz von Sekundäranalysen kann das Spektrum der üblichen universitären Lehre insgesamt erweitern.

(4) Ein direkter Nutzen für die datenproduzierende Forscherin selbst ergibt sich dann, wenn die Weiternutzung ihrer Daten mit einer erhöhten Zitationsrate einhergeht, wie bereits gezeigt werden konnte (Piwowar et al. 2007, Furman/Stern 2008; für eine kritische Diskussion siehe Abbott 2007) (*Reputationsgewinne des Einzelforschers*).⁸

2.2 Die nationalen und internationalen Rufe nach Data Sharing

Aufgrund der oben geschilderten und allgemein anerkannten Vorteile des Data Sharing gibt es mittlerweile eine Reihe institutioneller Initiativen, die die Förderung von Data Sharing in den Blick nehmen und Sekundärforschung forcieren. So schreibt die OECD 2007 in ihren Principles and Guidelines for Access to Research Data from Public Funding "Sharing and open access to publicly funded research data not only helps to maximise the research potential of new digital technologies and networks, but provides greater returns from the public investment in research."⁽¹⁰⁾ In Großbritannien ist der Wandel hin zu einer anderen Datenkultur bereits soweit vorangeschritten, dass einige Forschungsförderer eine Stellungnahme hinsichtlich Data Sharings zur vertraglichen Verpflichtung gemacht haben (so bspw. Economic and Social Research council (ESRC 2010), Natural Environment Research Council (NERC 2011)).

Die Entwicklungen auf der internationalen Ebene haben bereits begonnen auch die Situation in Deutschland zu verändern. So hat die DFG ihre Antragsmuster 2010 ergänzt um die Aufforderung: „... legen Sie bitte dar, welche Maßnahmen ergriffen wurden bzw. während der Laufzeit des Projektes getroffen werden, um die Daten nachhaltig zu sichern und ggf. für

6 Vergleichbar mit der Medizin sind in den vergangenen Jahren die Kenntnisse der Diagnostik nicht in gleichem Maße wie diejenigen der Therapie gestiegen. Dem gestiegenen Problembewusstsein und der gestiegenen Fehlerdiagnostik stehen noch keine hinreichend eindeutigen Lösungsvorschläge gegenüber.

7 Vergleiche für den Bereich der Surveyforschung die Diskussionen zum Total Survey Error und Data Quality (z.B. Biemer/Lyberg 2003); "as the statistical world moves towards the implementation of quality metrics, measures to improve quality and save costs"(Kreuter et al. 2010)

8 Bei Interesse an den individuellen und strukturellen Faktoren des Data Sharings sei auf Häusler 2009, 2010 verwiesen.

eine erneute Nutzung bereit zu stellen.“(DFG 2010, 32⁹). Für Langfristvorhaben gilt seit 2011: „Die Planung der Datendokumentation muss die langfristige Nutzung der Daten ermöglichen.“(DFG 2011, 3). Initiativen wie die Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen (2008) oder die GWK Kommission Zukunft der Informationsinfrastruktur (KII) der Leibniz-Gemeinschaft (2011) oder die aktuellen „Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften“ des Wissenschaftsrats (01/2011) stoßen in die gleiche Richtung.¹⁰

Betrachtet man die Tradition der Sekundärforschung Deutschlands im internationalen und interdisziplinären Vergleich (RIN 2008, Key Perspectives Ltd. 2010) besteht jedoch weiterer Handlungsbedarf. Die bisherigen Datenbestände beziehen sich vorwiegend auf Großprojekte, und eine „*Kultur des Data Sharing*“ (Huschka et al. 2011, 39) ist in den deutschen Sozialwissenschaften (noch) nicht verankert (Habich et al. 2010; GDF 2010a; GDF 2010b). Daraus folgt, dass eine Vielzahl von Daten(-schätzen) bis heute der Wissenschaft unwiderbringlich verloren gehen. „The cost of inaction is the loss of datasets which may or may not have had a role to play in the advancement of scholarship.“(RIN 2008, 4) Schätzungen über das die Sozialwissenschaften betreffende Ausmaß liegen unserer Kenntnis nach für die Sozialwissenschaften nicht vor¹¹, aber die Erfahrungen des DSZ-BO zeigen, dass Daten kleinerer Forschungsprojekte, wie sie u.a. in der deutschen, empirischen Organisationsforschung der Normalfall sind, regelmäßig *nicht* weitergenutzt werden. Aufgrund dieses Handlungsbedarfs sollen im Folgenden die Voraussetzungen für erfolgreiches Data Sharing diskutiert werden.

2.3 Voraussetzungen erfolgreichen Data Sharing's: Das Verfügbarkeits- und das Nutzbarkeitsproblem und die Rolle der Datenzentren

Um die potentiellen Vorteile von Data Sharing zu realisieren, bedarf es struktureller, organisatorischer und kultureller Veränderungen. Data Sharing verlangt in erster Linie die Bereitschaft von Forschern, die eigenen Daten weiterzugeben, aber ist auch – und das ist nicht minder wichtig – ein Ressourcenproblem. Es kostet Ressourcen, Daten (1) *verfügbar* zu machen, die (2) *nutzbar* sind.

(1) Verfügbar gemacht werden können Daten durch den Datenproduzenten selbst, über Repositorien wie sie Universitäten und einzelne Zeitschriften bieten oder über speziell dafür eingerichtete Datenzentren.¹² Die eigene Bereitstellung erfordert ein Ausmaß an Ressourcen, welches selten beim einzelnen Forscher vorhanden ist, dem in diesem Fall die gesamte Nutzeradministration obläge. Daher bietet sich die Übergabe an ein Datenzentrum, wie der

9 „... Wenn aus Projektmitteln systematisch (Mess-)Daten erhoben werden, die für die Nachnutzung geeignet sind, legen Sie bitte dar, welche Maßnahmen ergriffen wurden bzw. während der Laufzeit des Projektes getroffen werden, um die Daten nachhaltig zu sichern und ggf. für eine erneute Nutzung bereit zu stellen. Bitte berücksichtigen Sie dabei auch – sofern vorhanden – die in Ihrer Fachdisziplin existierenden Standards und die Angebote bestehender Datenrepositorien.“; ebenso enthalten im aktuellen Leitfadens, DFG 2012, 6).

10 Die gestiegene Bedeutung von Data Sharing spiegelt sich in der steigenden Zahl von Datenzentren national wie international. Zu den größten und bekanntesten Datenzentren gehören das US-amerikanische ICPSR (seit 1962), das britische UKDA (seit 1967, <http://ukda40.data-archive.ac.uk/>), in Deutschland die GESIS seit 1986). In den letzten Jahren ist in Deutschland ein starker Zuwachs von Datenzentren und eine Ausdifferenzierung der Dateninfrastruktur zu verzeichnen (vgl. RatSWD, www.ratswd.de/dat/fdz.php).

11 Schätzungen, die auf einer Fallstudie zur Archivierung von Film- und Fernsehmaterial beruhen, werden zitiert in RIN 2008, 4.

12 Innerhalb der Dateninfrastruktur werden heutzutage institutionelle Angebote der Universitäten und Universitätsbibliotheken von den fachbezogenen („domain-specific“) Angeboten von Datenzentren unterschieden (vgl. für weitergehende Informationen Dallmeier-Tiesen 2011, Pullinger/Wagner 2010, Siegert 2010, Green/Gutmann 2007). Hinzu kommen Angebote von Repositorien von Zeitschriften. In Deutschland im Bereich der Sozial- und Wirtschaftswissenschaften sind bislang vor allem die fachbezogenen Datenzentren aktiv.

GESIS oder je nach Disziplin z.B. dem IQB oder dem DSZ-BO an. Datenzentren zeichnen sich im Vergleich zu anderen Repositorien vor allem dadurch aus, dass sie neben der reinen Datenlagerung zusätzliche Leistungen anbieten und damit Datenproduzenten weitergehend entlasten (vgl. Huschka et al. 2011, 41ff.). So können Datenzentren über die Archiv- oder Lagerhaltungsfunktion¹³ hinaus, Informations-, Vermittlungs-, Bereitstellungs-, und Beratungsfunktionen übernehmen.¹⁴ Als im Vergleich zu den üblichen Lebensdauern von Forschungsprojekten langfristig existierende Institute lösen Datenzentren das Problem der Diskrepanz zwischen *short-term funding* und *long-term data sharing*.¹⁵ Als Vermittler zwischen Angebot und Nachfrage entlasten Datenzentren sowohl Datenproduzenten als auch Nutzer, da bilaterale Kooperationen für beide Seiten aufwendiger wären. Sie übernehmen das gesamte Nutzermanagement, d.h. die Administration, die Beratung und die Kontrolle der Einhaltung urheber- und datenschutzrechtlicher Bestimmungen. Dies beinhaltet die Vorhaltung entsprechender EDV, der nötigen personellen Ressourcen und der notwendigen Expertise, wie sie bspw. die Einhaltung des Datenschutzes bedarf.

(2) Die zweite Seite des Ressourcenproblems bezieht sich auf die Nutzbarkeit der Daten (i.S.v. usability). Während sich der Begriff der Verfügbarkeit auf das Datenangebot an sich und die reine Zugänglichkeit bezieht, bezeichnet der Nutzbarkeitsaspekt die Tatsache, ob die verfügbaren Daten *von Dritten* zu den gleichen *oder anderen* Forschungszwecken als den ursprünglich intendierten eingesetzt werden können und damit eher auf die Qualität als die Quantität des Datenangebots.

Die Nutzbarkeit von Forschungsdaten für Dritte hängt insbesondere von drei Dingen ab (ähnlich bei Büttner et al. 2011, S.14):

a) *Güte und Ausmaß der Dokumentation der Daten und der Prozesse der Datengenerierung*

Erst die hinreichende (lücken- und fehlerfreie) Dokumentation ermöglicht es Forscherinnen zu beurteilen, ob diese Daten, an deren Entstehungsprozess sie nicht beteiligt waren, zur Beantwortung der eigenen Forschungsfragen geeignet sind und die dafür notwendige Qualität aufweisen.

b) *Die rechtlichen und ethischen Bedingungen einer Weitergabe an Dritte müssen eingehalten werden.*

Die Weitergabe von Daten im Rahmen des Data Sharing darf nicht den Datenschutz beeinträchtigen, muss urheberrechtlich möglich und ethisch einwandfrei sein. (Artikel zu Recht)

c) *Die Art der Aufbewahrung oder Lagerung, d.h. wie die Daten und die zugehörigen Materialien vorliegen, sollte eine Weitergabe und -nutzung ermöglichen.*

So sollten die Forschungsdaten vollständig und in einem nutzbaren Format vorliegen. Handschriftliche Protokolle und Audiomitschnitte sollten möglichst digitalisiert worden sein, und die Formate sollten mit gängiger Soft- und Hardware verarbeitet werden können.

13 Mit dem Begriff der Archivierung wird üblicherweise die Langzeitarchivierung bezeichnet, welche mit besonderen Anforderungen einhergeht (vgl. z.B. nestor 2008, Klump 2011). Der Begriff der Lagerhaltung wurde gewählt, um von der Archivierung unterscheiden zu können. Nur wenige Datenzentren sind Datenarchive i.e.S. und bieten Langzeitarchivierung an (in D.: GESIS).

14 Hinzu treten Standardisierungsfunktionen, da die Transparenz von Studien zur Etablierung methodischer Standards beiträgt und die standardisiert gespeicherten Daten leichter vergleichbar sind.

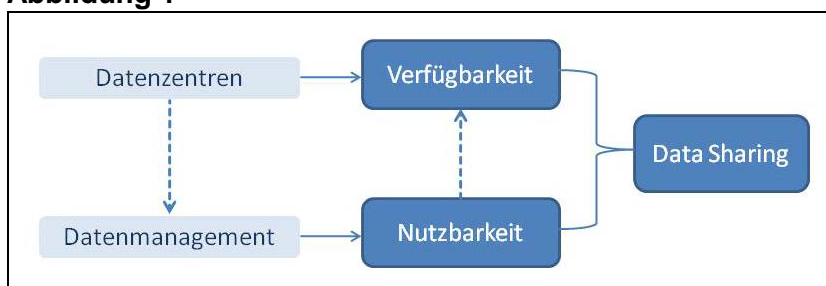
15 „Research funders and institutions should co-operate in seeking to ensure that long-term and sustainable arrangements are in place to preserve and make accessible the data that they deem to be of long-term value, and that such arrangements are not put at risk by short-term funding pressures.“(RIN 2008, S.9)

Im Gegensatz zum Verfügbarkeitsproblem, das durch die Existenz von Datenzentren und den Ausbau der Dateninfrastruktur adressiert wurde (vgl. Huschka et al. 2011, 36) und werden kann, ist eine Lösung des Nutzbarkeitsproblems allein durch Datenzentren und durch nachträgliche Maßnahmen nicht bzw. nur begrenzt möglich. Datensätze, die ohne die Angabe der Erhebungszeitpunkte übermittelt werden oder Variablen, die mit den mitgelieferten Fragebögen nicht übereinstimmen, zeigen beispielhaft, dass hier ohne das erneute Tätigwerden des Primärforschers die Nutzbarkeit der Daten durch Mitarbeiter von Datenzentren alleine nicht hergestellt werden kann. Auch die Suche nach Daten und Dokumenten auf dem heimischen PC oder in Büroschränken kann dem einzelnen Forscher nicht abgenommen werden. Ebenso wenig können einmal getätigte Zusicherungen gegenüber Probanden, die eine Nachnutzung ausschließen, im Nachhinein rückgängig gemacht werden. Für solche Arbeiten fehlt es dem Einzelforscher regelmäßig an den notwendigen Ressourcen, da mit Abschluss des Projekts auch die Förderung endet und die daran geknüpften befristeten Beschäftigungsverhältnisse auslaufen.

Festzuhalten bleibt: Zum einen ist die nachträgliche Nutzbarmachung von Forschungsdaten mit Aufwand verbunden, welcher dem Einzelforscher nur teilweise durch Datenzentren abgenommen werden kann. Zum zweiten ist die nachträgliche Herstellung der Nutzbarkeit auch unabhängig von Ressourcenbegrenzungen in vielen Fällen nur eingeschränkt oder gar nicht möglich ist. Darüber hinaus wirken Aspekte fehlender Nutzbarkeit auch direkt auf das Angebot an Daten selbst. So sind sowohl der schlechte Zustand von Daten als auch Bedenken hinsichtlich des Datenschutzes und des Urheberrechts häufig genannte Einwände von Forschern gegen eine Datenweitergabe (RIN 2008, 26f., Key Perspectives Ltd. 2010, 6, Huschka et al. 2011, 38f., Abbott 2007, 13ff.) An dieser Stelle verschärft sich das Nutzbarkeits- das Verfügbarkeitsproblem weiter. Wenn aber die nachträgliche Nutzbarmachung von Forschungsdaten kein gangbarer Weg ist – weder für die Forscher selbst noch für Datenzentren, dann werden *forschungsbegleitende* Maßnahmen notwendig. Solche forschungsbegleitenden, den Umgang mit Forschungsdaten betreffende Maßnahmen, die die Nutzbarkeit dieser gewährleisten können und nachträglichen Aufwand überflüssig machen, sind Maßnahmen des Datenmanagements. Jensen (2011, 72) schreibt: „Ein strukturiertes Konzept zum Datenmanagement ist für jedes Projekt essenziell, um die Arbeitsergebnisse während der Laufzeit zu sichern und zur weiteren Nutzung der Forschergemeinschaft bereitzustellen.“

Bevor im nächsten Abschnitt das Datenmanagement näher vorgestellt wird, veranschaulicht die nachstehende Abbildung das bisher Erläuterte. Erfolgreiches Data Sharing hängt sowohl von der Verfügbarkeit von Forschungsdaten als auch von der Nutzbarkeit dieser ab. Fehlende Nutzbarkeit beeinflusst das verfügbare Angebot an Daten negativ, wenn sie ein Grund ist, Daten nicht anzubieten. Während jedoch das Verfügbarkeitsproblem durch die Dienstleistungen von Datenzentren adressiert werden kann, kann die Lösung des Nutzbarkeitsproblems nur in einer Verbesserung des Datenmanagements durch die Datenproduzenten und durchgehend während der Forschungsprojekte liegen. Ein Prozess, den Datenzentren aber durchaus unterstützen können.

Abbildung 1



3 Datenmanagement - zur Herstellung und Optimierung der Nutzbarkeit von Forschungsdaten

3.1 Datenmanagement

Mit dem Begriff des Datenmanagements kann allgemein eine spezifische, auf die wissenschaftliche Branche und den Umgang mit Forschungsdaten ausgerichtete Form des Projektmanagements, der Arbeitsorganisation und -steuerung umschrieben werden. Es geht darum, die eigenen Arbeitsprozesse, die die Erzeugung von und den Umgang mit Forschungsdaten betreffen, möglichst *effizient* und *zielorientiert* zu organisieren und fortlaufend zu steuern. Wenn das Ziel der Datengenerierungsprozesse in Forschungsprojekten deren anschließende Nutzung ist, so ist das Ziel des Datenmanagements die *effiziente* Nutzung dieser Daten *in Abhängigkeit von den an sie gestellten Anforderungen*. Datenmanagement beinhaltet bereits im Rahmen der Projektplanung (vgl. Jensen 2011, 72ff.) die spezifischen Aufgaben und Probleme zu berücksichtigen, die im Zusammenhang mit Forschungsdaten entstehen, die nachfolgenden datenbezogenen Arbeitsprozesse dementsprechend zu planen und im Prozess zu steuern.

Datenmanagement adressiert genau die Aspekte, die die Nutzbarkeit von Forschungsdaten maßgeblich bestimmen. Insbesondere sind hier die Dokumentation, die Aufbewahrung oder Lagerung der Daten und Dokumente und die Einhaltung rechtlicher und ethischer Anforderungen zu erwähnen. Schließlich gilt es die Planung und Steuerung des Datenmanagements selbst nicht zu vergessen, worunter Punkte, wie der Ressourceneinsatz und die Zuweisung von Verantwortlichkeiten fallen. Wenngleich diese Aspekte auch heute schon vielfach bedacht werden, geht systematisches Datenmanagement darüber hinaus. Datenmanagement, das den Standards einer längerfristigen Nutzbarkeit durch Dritte genügt, ist (1) umfassend, (2) integrativ und (3) nutzerorientiert. Es ist umfassend, weil sämtliche der unterschiedlichen Teilaspekte im Umgang mit den Forschungsdaten berücksichtigt werden.¹⁶ Es ist integrativ, weil innerhalb einer Gesamtplanung diese Teilbereiche miteinander abgestimmt und die bestehenden Interdependenzen berücksichtigt werden. Schließlich ist Datenmanagement nutzerorientiert, da es stets an zuvor zu definierenden aktuellen und zukünftigen Nutzungszwecken auszurichten ist. Denn "the audience for the data may influence how the data are managed and shared" (vgl. ICPSR o.J., 9).

Zur Nutzerorientierung von Datenmanagement

Mit unterschiedlichen Nutzungszwecken (wer, wann, wie oft, wie lange, wo) gehen je unterschiedliche Anforderungen an Datenmanagement und die Ausgestaltung der Datenmanagementpraktiken einher.¹⁷ Allgemein gilt, je diverser die Nutzungszwecke, je unbekannter die potentiellen Nutzer und je entfernter im Hinblick auf Raum und Zeit diese sind, desto unspezifischer und universeller sind die Anforderungen an das Datenmanagement. Data Sharing stellt besonders unspezifische und universelle Anforderungen an Datenmanagement, da die Nutzung im Rahmen des Data Sharing relativ unbestimmbar und langfristig ist. Data Sharing in seiner allgemeinsten Form bezieht sich auf

¹⁶ Für eine vollständige und nicht zusammengefasste Auflistung der einzelnen Aspekte, siehe die im Anhang aufgeführte Vorlage zur Erstellung eines Datenmanagementplans des britischen Datenarchives UKDA.

¹⁷ Potentielle Nutzungen unterscheiden sich v.a. hinsichtlich der Aspekte, wer der oder die Nutzer sind, wann, wie oft und wo die Nutzung stattfinden soll (die Projektbeteiligten oder weitere Personen, in Kooperation oder selbständig, jetzt oder in Zukunft, einmalig oder häufiger, am Ort der Datenproduktion oder einem sonstigen).

Das schottische DCC unterscheidet zwischen den drei Anforderungsprofilen Funder Requirements, in-project data management planning und long-term preservation (http://dmponline.dcc.ac.uk/system/attachments/8/original/DCC_Checklist_DMP_v3_md_sj.pdf?1300724157). Je nachdem werden vom DCC unterschiedliche Empfehlungen hinsichtlich der Gestaltung (Inhalte und Umfang) von Datenmanagementplänen gegeben.

die Nutzung für einen unbegrenzten und unbekanntem Nutzerkreis zu unbekanntem und unbegrenzten Nutzungszwecken langfristig und dauerhaft. Folgende Beispiele können dies verdeutlichen:

- Um fremde Nutzer nicht von der Nutzung auszuschließen, bedarf es einer umfassenden Dokumentation, so dass die Datenerhebung nachvollziehbar ist und die Daten interpretierbar sind.
- Gegebenenfalls ist das Einverständnis von Auftraggebern oder Förderern einzuholen, um Daten an nicht am Projekt beteiligte Personen mit anderen Forschungsinteressen weitergeben zu können.
- Potentielle Nutzer könnten im In- oder Ausland sitzen, was englischsprachige Variablenbezeichnungen erforderlich machen könnte.
- Mit unterschiedlicher disziplinärer Herkunft einher gehen unterschiedliche Standards hinsichtlich der genutzten Software oder der methodischen Terminologie.
- Ist die Archivierung über ein bestimmtes Datenzentrum geplant, sind die dort vorhandenen Bestimmungen einzuhalten.¹⁸

Die Ausrichtung des Datenmanagements an Nutzungszwecken ist jedoch nicht nur im speziellen Fall Data Sharing relevant. Auch sonstige Nutzungszwecke zu antizipieren und die Inhalte von Datenmanagementpraktiken danach auszurichten, ist notwendig, um die jeweiligen Anforderungen dieser Nutzungen zu erfüllen und die Nutzungen realisieren zu können. Hierunter fällt der aus Sicht des Forschers zentrale Nutzungszweck: die Auswertung der Forschungsdaten zur Beantwortung der Forschungsfragen zu dessen Zweck die Datengenerierung durchgeführt wurde. Dies kann beinhalten, Kollaborationen zwischen den verschiedenen mit den Daten arbeitenden Forschern des Projektes optimal zu regeln oder Übergaben von Arbeiten nach dem Ausscheiden eines Mitarbeiters an die Kolleginnen reibungslos zu gestalten. Weitere Nutzungsformen des Primärforschers sind beispielsweise die gemeinsame Nutzung der Daten im Rahmen von Kooperationen mit Forschern außerhalb des Projektes oder die langfristige, über das Projektende hinausgehende Nutzung durch die diesen selbst. Letzere kann angestrebt werden, um zukünftige Datenanalysen mit neuen Fragestellungen oder Analyseverfahren zu ermöglichen, um das Forschungsdesign für Zeitvergleiche replizieren zu können oder um Rezensionen der eigenen Veröffentlichungen schnell und ausreichend begegnen zu können.¹⁹

Die Nutzerorientierung ist ein wesentlicher Bestandteil systematischen Datenmanagements und Datenmanagement damit unerlässlich, um die für Data Sharing hohen Bedingungen zu erfüllen. Datenmanagement ist aber auch erforderlich, um die Realisierung der *forschereigenen Nutzungswünsche* zu gewährleisten. Nur wenn die Nutzungsbedarfe frühzeitig vor oder mit Beginn eines Forschungsprojektes eruiert werden, können die nachfolgenden Arbeitsprozesse entsprechend gesteuert werden und die sich aus den jeweiligen Nutzungszwecken ergebenden Anforderungen soweit wie möglich gewährleistet werden. Nachfolgend werden einzelne Teilbereiche des Datenmanagements vorgestellt und beispielhaft auf die Notwendigkeit frühzeitiger Planungen ebenso wie auf Probleme, die sich beim Ausbleiben dieser ergeben können, hingewiesen.

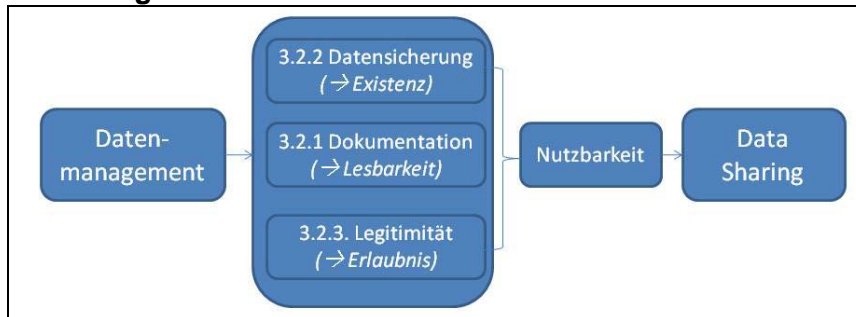
18 Für ein Beispiel siehe ICPSR 2011a, die Bestimmungen des US-amerikanischen Datenarchivs ICPSR.

19 Ergebnis einer Analyse der Nutzungszwecke kann es aber auch sein, dass Data Sharing nicht oder nur sehr eingeschränkt möglich gemacht werden kann; beispielsweise da die Kosten in dem spezifischen Fall zu hoch wären oder das Einverständnis der Partizipanden hierfür schlicht nicht eingeholt werden kann. Im Rahmen eines Datenmanagementplans können projekt- aber fachspezifische Besonderheiten berücksichtigt und gegenüber Förderern und der scientific community begründet werden. Möglich sind auch Einschränkungen des Zugangs bspw. durch Embargos oder Informationsrechte des Datenproduzenten zu Nutzungsanfragen.

3.2 Teilbereiche des Datenmanagements und ihre Bedeutung für Data Sharing

Um Data Sharing realisieren zu können, sind die verschiedenen Teilbereiche des Datenmanagements entsprechend zu gestalten. Zu den zentralen Teilbereichen des Datenmanagements gehören die Dokumentation der Daten, die diese erst lesbar machen (3.2.1), die sichere Aufbewahrung der Daten zum Schutz vor Verlust oder Manipulation (3.2.2) sowie die Prüfung rechtlicher und ethischer Erfordernisse im Hinblick auf die geplante Art der Datennutzung (3.2.3).²⁰

Abbildung 2



3.2.1 Datendokumentation - Die Beschreibung der Daten mit Metadaten

Zur Dokumentation von Forschungsdaten gehört die gezielte Bezeichnung oder Benennung der Daten selbst (Verschriftlichung, Digitalisierung) sowie die Ergänzung dieser um zusätzliche, die Studie und den Erhebungsprozess beschreibende Informationen, sog. Meta- und Paradata (vgl. Jensen 2011, 73f.). Dokumentieren bedeutet eine Beschreibung der Daten zu erstellen, die als eine Art Bedienungsanleitung für diese fungiert. Die Erfassung von Informationen über die Daten macht aus diesen erst ein eigenständiges und tauschbares Produkt.

Sollen die Daten im Rahmen des Data Sharing (d.h. für noch unbekannte Dritte für Sekundäranalysen) bereitgestellt werden, müssen sämtliche Informationen dokumentiert werden, die es Dritten ermöglichen, die Eignung der Daten für die eigenen Forschungszwecke zu beurteilen. Dritte benötigen Informationen sowohl über die *inhaltliche Relevanz* der Daten (relativ zur eigenen Forschungsfrage) als auch deren *methodische Qualität*.²¹ Hierzu dienen die Erfassung von Metadaten und das zur Verfügung stellen spezifischer Dokumente wie Fragebögen und Interviewleitfäden. Metadaten liefern zusätzliche Informationen über die Art und die Umstände der Datenerhebung und deren Zweck (z.B. Fragebogeninhalte, Projektlaufzeit, durchführende Organisationen, Feldphase usw.) (vgl. Gregory et al. 2009). „Metadata are often the only form of communication between the secondary analyst and the data producer, so they must be comprehensive and provide all of the needed information for accurate analysis.“ (ICSPR 2011b).

Zur Beurteilung der *methodischen Qualität* dienen spezifische Metadaten wie Responseraten, Feldzeiten oder Syntaxdateien. Neuerdings wird zur Beurteilung der Datenqualität verstärkt der Rückgriff auf Paradata empfohlen (vgl. Kreuter et al. 2010). Paradata stellen Messungen des Prozesses der Datenerhebung selbst dar (vgl. Kreuter et al. 2010).²² Hierzu gehören z.B. Kontaktprotokolle, Interviewerbeobachtungsprotokolle oder

²⁰ Eine vollständige Darstellung der Inhalte von Datenmanagement soll an dieser Stelle nicht gegeben werden. Interessierte können sich z.B. informieren bei Jensen 2011 und zukünftig auf der Homepage des DSZ-BO (www.uni-bielefeld.de/dsz-bo).

²¹ Ähnlich argumentiert das International Household Survey Network (IHSN 2009b).

²² Paradata sind Prozessdaten und damit eher eine spezifische Form von Mikrodaten. Im Unterschied zu den hier als Mikrodaten bezeichneten Daten sind sie nicht auf den eigentlichen

Audiomitschnitte der Interviews. Sind diese für Befragte und auch Nicht-Befragte verfügbar (wie bspw. Kontaktprotokolle oder Interviewerbeobachtungen), können sie für Nonresponse-Analysen oder zu Bias-Reduktion bspw. durch Gewichtung verwendet werden. Paradata, die die Interaktion zwischen Interviewer und Befragten messen (wie bspw. Audiomitschnitte von Telefoninterviews) können zur Analyse von Messfehlern genutzt werden. Nur wenn diese Art von Informationen erfasst und dokumentiert wurden, können die Analysepotentiale von Daten in Sekundäranalysen vollständig genutzt werden.

Dokumentationen dienen nicht nur dazu, Dritten die Beurteilung dieser zu ermöglichen. Daneben bilden Metadaten die grundlegenden Inhalte von Datenbestandskatalogen und Datenbanken, anhand welcher Datensätze recherchiert werden können. Ohne sie sind Daten schlicht nicht auffindbar und Data Sharing damit nicht erfolgreich. Auf die Bedeutung von „Informationsportalen“ für die „breite Nutzung in der Wissenschaft“ weisen auch Huschka et al. (2011) hin. Auch ergibt sich die Notwendigkeit zur Dokumentation von Forschungsdaten daher, dass die am Projekt beteiligten Personen häufig schon mit Ablauf des Forschungsprojekts für Rückfragen nicht mehr zur Verfügung stehen aufgrund der üblichen Fluktuation. Es treten praktische Probleme hinzu, wenn schriftliche Dokumentationen fehlen: Auf persönliche Erinnerungen der Beteiligten zurückgreifen zu müssen ist potentiell fehlerbehaftet und nicht alle Projektbeteiligten sind gleichermaßen informiert über sämtliche datenrelevante Prozesse, stattdessen liegt Wissen häufig diversifiziert vor. Dokumentationen liefern somit nicht nur die notwendigen Informationen für Sekundärnutzer, sondern ersparen auch den Aufwand für die Primärforscher, den das nachträgliche Rekonstruieren von Geschehnissen (z.B. über die Kontaktierung ehemaliger Projektbeteiligter) bedeuten würde. Zeitnahes Dokumentieren ist daher nicht nur sinnvoll für den Zweck der Sekundärnutzung, sondern auch mit Vorteilen für den Primärforscher selbst verbunden, der ja regelmäßig mit einer Vielzahl teilweise wechselnder Projektmitarbeiter über einen längeren Zeitraum hinweg Daten erhebt und bearbeitet. Und vom Verfassen eines Artikels, dessen Einreichung in einem Journal, möglicher Schleifen bei der Begutachtung bis hin zur Veröffentlichung können mehrere Jahre vergehen.

Bei der Verbreitung des Dokumentierens in der scientific community lassen sich große Defizite erkennen sowie Unterschiede in den dokumentierten Inhalten und dem Umfang der Dokumentationen. Dies zeigen Erfahrungen und Recherchen des DSZ-BO (vgl. Edler et al. 2012) und anderer (vgl. GDF 2010b, 19; Opitz/Mauer 2005; RIN 2009, Key Perspectives Ltd. 2010), wengleich systematische empirische Daten hierzu scheinbar fehlen. Die für die Sozialwissenschaften durchaus vorhandenen Standards der Dokumentation (wie DDI oder SDMX) sind noch ein ganzes Stück weit von einer einfachen Anwendbarkeit durch den Endnutzer (den Datenproduzenten) entfernt.²³ Datenzentren geben Inhalte und Umfang der Dokumentation vor, unterscheiden sich dahingehend aber hinsichtlich der verwendeten Begrifflichkeiten (vgl. bspw. WISDOM o.J.a, o.J.b und ICPSR 2011c).²⁴ Ein weiteres noch offenes Problem sind subdisziplinär unterschiedlich verwendete Terminologien, die einer Standardisierung der Datendokumentation, die mit den Vorteilen größerer Vergleichbarkeit, Transparenz und verbesserter Interpretierbarkeit der Daten einherginge, im Wege stehen.

Forschungszweck der Datenproduktion ausgerichtet, sondern haben den Erhebungsprozess selbst zum Gegenstand.

²³ Derzeit wird bspw. noch an DDI-Editoren zur leichteren Erfassung der Dokumentationsinhalte gearbeitet, die die Anwendbarkeit maßgeblich erhöhen würden (vgl. IASSIST 2011). Diese Arbeiten sind aber noch nicht abgeschlossen. Bereits erhältlich ist eine DDI-basierte Software zur Fragebogenentwicklung, QDDS, siehe <http://www.qdds.org/>.

²⁴ Einzelne Datenzentren bieten Online-Eingabemasken an, welche die Eingabe von Informationen erleichtern, z.B. PsychData am ZPID der Universität Trier <http://psychdata.zpid.de/index.php?main=give&sub=upload> oder im englischsprachigen Raum das Microdata Management toolkit von IHSN, IHSN 2009a)

Im Rahmen systematischen Datenmanagements müssten sowohl die Ressourcen (Hard- und Software, Personal) für solche Dokumentationsaufgaben bereitgestellt werden als auch die Kenntnisse darüber, was wie zu dokumentieren ist, vorhanden sein. Für die Notwendigkeit von Dokumentationen sollte sensibilisiert und ein Bewusstsein über dessen nutzenbringende Effekte geschaffen werden. Dokumentationen müssen frühzeitig bei der Projektplanung berücksichtigt werden. So sind Paradata nicht im Nachhinein erfassbar und Metadaten aufgrund der oben genannten Problematiken nur eingeschränkt und mit entsprechendem Mehraufwand. Da es nicht trivial ist zu entscheiden, welche Aspekte wie und wie ausführlich zu dokumentieren sind, sind Forscher gefragt sich die entsprechenden Kenntnisse anzueignen und/oder Beratungsleistungen von entsprechenden Stellen wie Datenservicezentren einzufordern.

3.2.2 Datensicherung

Neben der Beschreibung der Daten ist zur Herstellung einer längerfristigen Nutzbarkeit von Daten durch Dritte relevant, dass die Daten sicher verwahrt werden²⁵. Eine „sichere“ Verwahrung beinhaltet den Schutz der Daten vor unbefugten Zugriffen und Manipulationen sowie den Erhalt des Datenbestands selbst (Schutz vor Verlust) (vgl. Jensen 2011, 75f.). Der Erhalt des Datenbestands kann sowohl durch *technische Fehler* (Hardware- oder Software-Probleme, Computerviren) gefährdet sein, aufgrund derer Daten und Dateien unbrauchbar werden oder verloren gehen als auch durch *menschliche Fehler* wie versehentliches Löschen oder Verlegen von Dateien. Maßnahmen wie regelmäßige Backups, Firewalls, Virencans auf der einen, Versionierungen und Vergabe von Zugriffsrechten auf der anderen Seite sind zu installieren. Versäumnisse können auch rechtlich problematisch sein, wenn bspw. sensible Daten vor Zugriffen von außen nicht hinreichend geschützt werden.

Die zu ergreifenden Maßnahmen variieren in Abhängigkeit von Faktoren wie der Sensibilität der vorliegenden Daten, den Ansprüchen der Untersuchungsobjekte auf Vertraulichkeit, den Ansprüchen der Auftraggeber oder anderer Projektteilnehmer sowie der Anzahl dieser. Unabhängig von den Anforderungen der Sekundärnutzung variieren Maßnahmen der Datensicherung und -aufbewahrung von den Modalitäten der projektinternen Nutzung. Dies sind neben den geplanten Arbeiten im Rahmen der Datenerhebung und -analyse selbst Speicher- und Rechenkapazitäten, Mobilitätsanforderungen der Projektmitarbeiter und Vorhaben hinsichtlich kollaborativen Arbeitens. Insbesondere jedoch sind die zu installierenden Maßnahmen von der geplanten Dauer der Aufbewahrung der Forschungsdaten abhängig. Die langfristige Aufbewahrung von Daten (Archivierung) erfordert spezielle Kenntnisse und Ressourcen, die in den üblichen, befristeten Projektzusammenhängen nicht vorliegen. Aktuell mangelt es zudem an Standards bzgl. der zu nutzenden Dateiformate für Langzeitarchivierung (Long 2009, 336). Zum Zweck der Archivierung erscheint daher der Rückgriff auf externe Dienste unerlässlich, wie sie einzelne Universitäten (z.B. die Universitätsbibliothek Bielefeld) oder Datenarchive (z.B. die GESIS) anbieten.

Für erfolgreiches Data Sharing erforderlich ist nicht nur, dass alle relevanten Dateien noch vorhanden sind, sondern auch dass darauf zugegriffen werden kann. Bekannte Probleme aus Sicht von Datenzentren sind, dass Daten in den Forschungsprojekten nicht immer vollständig digitalisiert werden (z.B. wenn nicht sämtliche der geführten Interviews transkribiert werden), und dass Datenträger und genutzte Software veralten und späteres Zugreifen erschweren oder verhindern. Allgemein wird davon ausgegangen, dass der Wissenschaft hierdurch bis heute viele Datensätze und Datenschätze unwiederbringlich verloren gegangen sind (siehe die Beispiele bei Long 2009, 343ff. oder die Prognosen benannt in Key Perspectives Ltd. 2010, 4). Studien oder verlässliche Schätzungen zum Ausmaß dieses Verlusts liegen unseres Wissens jedoch bislang nicht vor.

25 Im Englischen auch mit dem Begriff data preservation beschrieben. Vgl. zum Beispiel ICPSR 2009.

Ebenso wie für die Arbeiten der Datendokumentation sind für diejenigen der Datenverwahrung spezifische Kenntnisse und Ressourcen notwendig. Im Zusammenhang mit der Datenverwahrung sind hier neben arbeitsorganisatorischen, auch technische und damit fachfremde Kenntnisse zu nennen. Versäumnisse können rechtlich problematisch sein (zum Beispiel keine Sicherung der Daten vor Zugriffen von außen) oder auch den Datenbestand derart beeinträchtigen, dass eine spätere Weitergabe und Weiternutzung ausgeschlossen ist. Die erforderlichen Bedarfe sind frühzeitig und umfassend zu berücksichtigen. So ist der Bedarf an technischen Ressourcen (Hard- und Software) bereits bei der Mittelbeantragung einzuplanen. Daten vor Verlust und Manipulation zu schützen, verlangt daher auch vor diesem Hintergrund nach systematischem Datenmanagement: „Digital data need to be actively managed over time to ensure that they will always be available and usable. This is important in order to preserve and protect our shared scientific heritage as technologies change.“(ICPSR o.J., 6) In der Regel bieten Universitäten und universitätsinterne Rechenzentren hierzu Leistungen an, die jedoch noch auf die Spezifika der jeweiligen Forschungsprojekte abzustimmen sind.²⁶

3.2.3 Rechtliche und Ethische Aspekte im Umgang mit Forschungsdaten

Um Data Sharing realisieren zu können, müssen bestimmte rechtliche und ethische Bedingungen während des Forschungsprojekts eingehalten werden. Die Bedeutung dieses Teilbereichs des Datenmanagements zeigt sich auch darin, dass Forschungsethik ebenso wie Datenschutz als zentrale Anliegen des RatSWD genannt werden (vgl. <http://idw-online.de/de/news421206>, Pressemitteilung vom 04.05.2011, GDF 2010b). Für Data Sharing muss es rechtlich möglich und ethisch vertretbar sein, Forschungsdaten an noch unbekannte Dritte außerhalb des eigentlichen Forschungszusammenhangs herauszugeben. Dem entgegenstehen können die Ansprüche der unterschiedlichen, beteiligten Parteien im Hinblick auf die Daten wie bspw. solche urheber- oder datenschutzrechtlicher Art oder die Rechte auf Information und Zustimmung bzgl. des Umgangs mit den Daten. Die Parteien, die mögliche Ansprüche besitzen, sind die an der Datenerhebung beteiligten Wissenschaftlerinnen (evtl. an verschiedenen Universitäten oder Instituten angesiedelt), die Untersuchungsobjekte (Personen, Haushalte oder Organisationen) und ggf. weitere wie Drittmittelgeber, Umfrageinstitute oder Auftraggeber.

Zentral sind die *datenschutzrechtlichen Ansprüche*, die den Untersuchungsobjekten Anonymität und Vertraulichkeit gewährleisten (vgl. Spindler/Hillegeist 2011, 67ff.). Anonymität muss auch bei einer Weitergabe der Daten an Dritte sichergestellt sein. Der Datenschutz ist in diesem Fall nicht nur von den Projektbeteiligten einzuhalten, sondern auch von den späteren Sekundärnutzern. Hierfür müssen die Daten in anonymisierter Form bereitgestellt werden, so dass eine Re-Identifikation ausgeschlossen werden kann.²⁷ Starke Anonymisierungen können jedoch die Analysepotentiale von Daten sehr einschränken und sollten daher sehr gezielt vorgenommen werden. Des Weiteren sind Datennutzungen durch Dritte üblicherweise (trotz Anonymisierung) auf die Einhaltung des Datenschutzes hin zu kontrollieren. Anonymisierung bei gleichzeitiger Bewahrung des Analysepotentials und die Kontrolle der Einhaltung des Datenschutzes sind Anforderungen der Sekundärnutzung durch Dritte zu leisten, um Data Sharing zu ermöglichen. Der einzelne Datenproduzent, der Forschungsdaten weitergeben möchte, kann hierzu auf eine Vielzahl von Datenzentren in Deutschland zurückgreifen, die in dieser Hinsicht eine langjährige Expertise aufweisen und diesen so von den Aufgaben der Anonymisierung und Datenschutzkontrolle entlasten.

26 (Informationen zum ICPSR disaster planning finden sich in ICPSR 2011d. hier: www.icpsr.umich.edu/icpsrweb/ICPSR/curation/disaster/index.jsp).

27 Daten sind mindestens so weit zu anonymisieren, dass ihre Re-Identifikation nur mit einem „unverhältnismäßigen Aufwand“ möglich wäre (vgl. BDSG §3, Abs. 7). Hierbei handelt es sich um die sogenannte „faktische Anonymität“. Die Bestimmung der faktischen Anonymität, ist nicht trivial (vgl. Wirth 1992, 2003; für qualitative Daten: Thomson et al. 2005), und wird im Zuge des technologischen Wandels scheinbar schwieriger (vgl. Schaar 2009).

Neben des Anspruchs auf Anonymität besitzen Untersuchungsobjekte einen *Anspruch auf Information*, den Oellers und Wegner ebenfalls als „Key right“ (Oellers/Wegner 2009, 2) bezeichnen. Aufgrund des Rechts auf informationelle Selbstbestimmung (vgl. Häder 2009, 5 und Schaar 2009, 3f.) aber auch aufgrund ethischer Erwägungen, sollten Individuen über die Inhalte, Zwecke und die Projektbeteiligten von Befragungen in Kenntnis gesetzt werden, so dass ihre Zustimmung zur Teilnahme als eine informierte Zustimmung („informed consent“) bezeichnet werden kann. Sollen die erhobenen Daten an Dritte weitergegeben werden, ist daher zu bedenken, ob die Befragten hierüber zu informieren sind und ob ihr Einverständnis eingeholt werden sollte. Solche Überlegungen müssen vor der Datenerhebung abgeschlossen sein. Denn die Probanden nach Abschluss der Erhebung, erneut zu kontaktieren, um deren Zustimmung einzuholen, ist mit erhöhten Kosten verbunden. Hinzu kommen die methodischen Aspekte im Hinblick auf das Forschungsdesign: Aus methodischer Sicht ist zu bedenken, in welcher Form (schriftlich/mündlich) und zu welchem Zeitpunkt der Befragung (währenddessen oder im Nachhinein, wann während des Interviews) dies zu geschehen hat, und welche Formulierungen zu wählen sind, so dass die Teilnahme und die Antworten der Untersuchungsobjekte am wenigsten beeinflusst werden.²⁸

In Forschungsprojekten werden unseren Erfahrungen nach häufig Formulierungen gegenüber Probanden verwendet, die eine Weitergabe an Dritte außerhalb des Projektes ausdrücklich ausschließen. Dies könnte vermieden werden, wenn im Rahmen systematischen Datenmanagements frühzeitig, die späteren Nutzungszwecke der Forschungsdaten bedacht und berücksichtigt werden.

In welchen Fällen und wie eine Informierung oder Zustimmung einzuholen ist, ist abhängig vom jeweiligen Projekt und nicht allgemein zu beantworten; auch weil hierzu bislang eindeutige Regelungen fehlen (vgl. Presser 1994, Oeller/Wegner 2009). Allgemeinhin werden Verknüpfungen mit weiteren Datenquellen (aus datenschutzrechtlichen Gründen) als zustimmungspflichtig angesehen und durchgeführt (vgl. z.B. Antoni et al. 2010, 19f. oder für Betriebe Spengler/Lorek 2010).

Neben den Ansprüchen der Untersuchungsobjekte bestehen Ansprüche der Projektbeteiligten im Sinne des *Urheberrechts* an den Datensätzen selbst oder an den Instrumenten der Datenerhebung (z.B. Fragebögen, Fragebatterien) (vgl. Spindler/Hillegeist 2011, 63ff.). Auch diese können die Möglichkeiten einer Weitergabe der Daten und Metadaten und dazugehöriger Materialien (wie Fragebögen) beeinträchtigen. Denn ohne die Zustimmung des Urhebers dürfen Daten nicht weitergegeben werden.²⁹ Gegebenenfalls ist das Einverständnis aller beteiligten Forscher, Universitäten und evtl. Verlagen einzuholen.

Ethische Erwägungen beziehen sich v.a. auf die Anonymität und die Informiertheit der Untersuchungsobjekte, d.h. auf den vertrauenswürdigen und transparenten Umgang mit diesen, aber auch auf Konflikte zwischen Forschern selbst (vgl. Oellers/Wegner 2009). Der Bereich der Forschungsethik ist in Deutschland durch Richtlinien und Empfehlungen der Deutschen Forschungsgemeinschaft (DFG), der Deutschen Gesellschaft für Soziologie (DGS) und dem Berufsverband Deutscher Soziologinnen und Soziologen (BDS) reguliert und

28 Die Forschungen zu informed consent sind zahlreich und befassen sich mit sehr unterschiedlichen Aspekten (Singer 1978, Sin 2005, zu Subpopulationen: Wiles et al. 2007), u.a. auch mit den Auswirkungen der Zustimmung zu Datenverknüpfungen (Record Linkage) auf die Selektivität von Antworten (vgl. Sala/Burton/Knies 2010). Forschungsbefunde zu informed consent explizit im Hinblick auf Data Sharing liegen bislang für die Sozialwissenschaften unseres Wissens nach nicht vor (Ergebnis einer Literaturrecherche in Web of Sciences vom August 2011).

29 Bei einer Weitergabe von Daten an ein Archiv zur Weiternutzung wird in der Regel nicht das Eigentum, das Urheberrecht, an den Daten übertragen. Dieses verbleibt beim Urheber. Stattdessen werden Nutzungsrechte eingeräumt²⁹. Für den Fall, dass das Urheberrecht bei mehreren Personen liegt (sog. joined ownership), ist das Einverständnis jeder dieser Personen einzuholen.

mit vergleichsweise geringen Sanktionen verbunden (vgl. Oellers/Wegner 2009, GDF 2010b). Die Tragweite ethischer Aspekte wird in den Sozialwissenschaften derzeit noch weithin unterschätzt, beobachten Oellers und Wegner: Forschungsethik werde innerhalb der deutschen Sozialwissenschaften bislang kaum wahrgenommen (vgl. Oellers/Wegner 2009). In einer der wenigen deutschen Studien zu diesem Thema in den Sozialwissenschaften konstatieren die Autoren enorme Defizite im internationalen und interdisziplinären Vergleich und fordern entsprechende Qualifizierungsmaßnahmen (Oellers/Wegner 2009 oder DFG 1998, 7).³⁰

Rechtliche und ethische Erwägungen in die Forschungsplanung und -praxis einzubeziehen ist nicht nur notwendig, um Data Sharing zu ermöglichen, sondern auch um die eigenen Forschungsvorhaben zu realisieren. Die verschiedenen Nutzungsarten sind potentiell gefährdet durch fehlende Einhaltung des Datenschutzes und Kontrollen dieser, Urheberrechtsstreitigkeiten oder fehlende Einwilligungen. Jenseits faktischer Rechtsbrüche und ethischer Probleme ist Vertrauen (i.S.v.Trust) zentral für den Erfolg von Datenerhebungen. Besonders relevant ist dies u.a. für Forschungen zu datenschutzrechtlich sensiblen Merkmalen, wie Regionalstudien, Studien zu speziellen Subpopulationen wie bspw. Kindern, Studien, die Verknüpfungen mehrerer Datenquellen planen und auch Panelstudien.

Angebracht scheint es daher, dass sämtliche Schritte des Forschungsprozesses unter rechtlichen und ethischen Gesichtspunkten frühzeitig geprüft werden (vgl. Häder 2009, 9ff.). Die Einhaltung der rechtlichen und ethischen Bedingungen kann zu einem erhöhten Bedarf an Ressourcen führen. Zur Einschätzung der rechtlichen Situation sind spezifische und Sozialforschern fachfremde Kenntnisse erforderlich, weshalb Forscher hier auf externe Beratungsleistungen angewiesen sind, die beispielsweise Hausjuristen oder Datenzentren erbringen können. Letztere können auch im Hinblick auf die methodische Umsetzung rechtlicher und ethischer Bedingungen beraten.

4 Herausforderungen hinsichtlich einer Professionalisierung des Datenmanagements

Die Standards des heutzutage in Forschungsprojekten üblicherweise angewandten Datenmanagements scheinen den aktuellen Anforderungen an die Nutzungsform Data Sharing, aber auch an andere, nicht zu genügen. Datenmanagement wird nicht standardmäßig und nicht im erforderlichen Ausmaß, d.h. umfassend, integrativ und nutzenorientiert, durchgeführt (RIN 2008, van den Eyden 2010). Für Deutschland lassen sich vor allem drei Defizite identifizieren, die die bisherige Situation kennzeichnen und einer Professionalisierung von Datenmanagement im Weg stehen: fehlende Anreize, Kenntnisse und Ressourcen.

Defizit Anreize

Das derzeitige wissenschaftliche Belohnungssystem ist neben der Höhe eingeworbener Drittmittel stark auf die Publikation von Forschungsartikeln ausgerichtet. Das Erstellen nutzbarer Datensätze, Datendokumentationen oder Methodenberichte hingegen bleibt weitestgehend unberücksichtigt (vgl. Büttner et al. 2011, 19). Forschungsdaten genießen nicht den Status eines eigenständigen, wertvollen, wissenschaftlichen „Gutes“, und es fehlt eine „Publikationskultur für Forschungsdaten“ (Büttner et al. 2011, 19). Dass die hiermit zusammenhängenden Tätigkeiten derzeit im Wissenschaftssystem nicht belohnt werden, könnte auch ein Grund dafür sein, dass Forscher die Aufgaben des Datenmanagements negativ wahrnehmen (Key Perspectives Ltd. 2010, RIN 2008: 26, GDF 2010b: 18f.). Falls aufgrund der eingangs beschriebenen hohen Attraktivität von Sekundäranalysen deren Anwendungen weiter steigen, besteht Anlaß zur Vermutung, dass die Bedeutung von

30 In Deutschland vorliegende Regelungen der Forschungsethik gehen bislang nicht über Empfehlungen und Richtlinien hinaus. Diese seien zudem kaum mit Konsequenzen verbunden und bezögen sich im Übrigen vor allem auf Konflikte zwischen Forschern und weniger auf die Ansprüche der Untersuchungsobjekte (vgl. Oellers/Wegner 2009).

Forschungsdaten und den diesen zugehörigen Materialien steigen wird. Gleichzeitig steigen mit dem Auf- und Ausbau der Dateninfrastruktur in Deutschland (Datenzentren) die Möglichkeiten von Forscherinnen und Forschern, Daten als eigenständiges und tauschbares Gut wahrzunehmen und zu behandeln.

Zunehmend entwickeln sich auch konkretere Anreize. So sind in den letzten Jahren die Möglichkeiten gestiegen, Artikel über Forschungsdaten in Datenjournalen wie European Data Watch (Schmollers Jahrbuch) zu publizieren (vgl. Dallmeier-Tiessen 2011) und die Daten in Datenbanken bzw. Informationsportalen zu listen. Möglichkeiten, die es weiter auszubauen gilt, betonen Huschka et al. (2011, 45). Und Sekundärnutzer von Forschungsdaten werden dazu angehalten in ihren Veröffentlichungen, die Daten zu zitieren, was die Reputation des Datenproduzenten erhöht. Zitationsstandards für Datensätze entwickeln sich.³¹

Anreize zum Datenmanagement ergeben sich weiter aus den Vorteilen, die damit für den Forscher selbst einhergehen: Für die einzelnen Forschungsprojekte kann professionelles Datenmanagement zu erheblichen Effizienzgewinnen und Qualitätsverbesserungen beitragen³², was insbesondere im modernen Wissenschaftssystem zunehmend eingefordert wird. Jedes Forschungsprojekt unterliegt dem Zwang zur Innovativität. Mit empirischen Erkenntnisgewinnen werden Forschungslücken kleiner und somit müssen die Vorgehensweisen diese zu schließen differenzierter werden. Mit den Erkenntnisgewinnen der Methodenforschung werden sie zudem quasi automatisch methodisch anspruchsvoller. Durch diese Prozesse entstehen ständig neue Anreize zur Anwendung neuer Methoden oder Methodenkombinationen ebenso wie zur Zusammenlegung von Ressourcen im Rahmen von Teamarbeit und bspw. interdisziplinären Kooperationen. Dies verlangt - zusammen mit dem Wettbewerb um Drittmittel und Publikationen sowie dem Legitimitätsdruck wissenschaftlicher Forschung vor dem Hintergrund knapper öffentlicher Kassen – nach einer Forschung die besonders effizient arbeitet, transparent und von hoher Qualität ist. Daraus folgt die Erfordernis eines professionellen Umgangs mit Forschungsdaten: Das heißt, das Planen und Steuern der gesamten Arbeitsprozesse im Hinblick auf die effiziente (und optimalerweise langfristige, auch durch Dritte mögliche) Nutzung der Daten. Hinzu kommt, dass durch die heutigen, technischen Ausstattungen von Projekten neue Möglichkeiten existieren, die Arbeit mit Forschungsdaten zu verbessern und effizient, kollaborativ und Synergieeffekte nutzend zu gestalten. Aufgrund der technischen Entwicklungen verbessern sich sowohl die Möglichkeiten für Data Sharing als auch für den Einsatz innovativer Methoden bspw. der Datenverknüpfung oder des Data Mining und neuer Datentypen (RIN 2009; Key Perspectives Ltd. 2010:3).

Jenseits von Belohnungen und positiven Anreizen beziehen sich die aktuellen Maßnahmen britischer und US-amerikanischer Förderinstitutionen v.a. darauf, Forscher zum Datenmanagement zu *verpflichten*. 2010 führte das britische Economic & Social Research Council die ESRC Research Data Policy (ESRC 2010) ein, die besagt, dass Förderanträge Aussagen zu Data Management and Sharing beinhalten müssen, welche begutachtet werden. Seit 01.01.2011 wird von der zentralen US-amerikanischen Forschungsförderung (der National Science Foundation, NSF) obligatorisch verlangt, dem üblichen Förderantrag

31 Für weitere Informationen zur Zitation von Forschungsdaten bspw. über sog. Digital Object Identifier, DOI, siehe vgl. Askita 2010, Brase 2010, Hausstein 2012.

32 Das zeigt auch die Existenz aktueller Projektvorhaben wie die Virtuelle Forschungsumgebung (VFU) des soeb am sofi Goettingen (www.soeb.de/soeb-3-virtuell) oder das INF-Projekt des SFB882 (www.sfb882.uni-bielefeld.de/projects/Teilprojekt_INF). Im ersten Projekt geht es darum, die Kooperation und Kollaboration zwischen den an verschiedenen Orten sitzenden und verschiedenen Instituten angehörenden Projektmitarbeitern über den Einsatz geeigneter IT-Lösungen zu ermöglichen und zu optimieren, was auch die gemeinsame Arbeit mit Daten beinhaltet. Im INF-Projekt geht es darum projektübergreifend Angebote zu schaffen, die den projektinternen Umgang mit Daten ebenso wie den projektübergreifenden Austausch gestalten. Durch INF können Ressourcen gemeinsam genutzt und Synergieeffekte erzielt werden.

einen Datenmanagementplan (DMP) hinzuzufügen (vgl. NSF 2010). Es wird davon ausgegangen, dass „Dissemination of data is necessary for the community to stimulate new advances as quickly as possible and to allow prompt evaluation of the results.“(NSF 2010, 1) Die Umsetzung des Datenmanagements wird fortlaufend überwacht und eine Evaluation dessen geht in die Begutachtung zukünftiger Förderanträge ein. Die Ausweitung der Foki der Förderinstitutionen, die sich zuvor nur auf Data Sharing bezogen, verweist auf die gestiegene Bedeutung, die diese dem Aspekt des Datenmanagements beimessen (vgl. auch Winkler-Nees 2011). Einzelne Journals verlangen schon länger die Übermittlung der Daten, auf denen die Publikationen beruhen. Dies spiegelt sich auch deutlich in den aktuellen DFG-Antragsbestimmungen wider, wonach Forscherinnen und Forscher entsprechende Maßnahmen zur nachhaltigen Sicherung und Bereitstellung der Daten für etwaige Sekundärnutzung zu ergreifen haben (vgl. DFG 2012).

Defizit Kenntnisse

Neben den fehlenden Anreizen sind es vor allem die fehlenden Kenntnisse im Zusammenhang mit Datenmanagement, die dessen breite und hochwertige Anwendung bislang verhindern. Das US-amerikanische Datenarchiv ICPSR hat nach der Einführung der Pflicht zur Erstellung eines Datenmanagementplans durch die NSF einen Ansturm auf ihre Services verzeichnet, und Universitätsbibliotheken sahen sich abrupt neuen Nutzerbedarfen, Beratungsbedarf zu DMPs, ausgesetzt (vgl. IASSIST Annual Conference 2011). Das deutet darauf hin, dass das Wissen unter US- Wissenschaftlern über einen professionellen Umgang mit Forschungsdaten nicht weit (genug) verbreitet gewesen zu sein scheint. Zu vermuten ist, dass diese Einschätzung auch für Deutschland zutrifft, da Forschungsdatenmanagement nicht zum Bestandteil der üblichen universitären (Graduierten- oder Postgraduierten-) Ausbildung gehört.

Die Situation scheint jedoch zweigeteilt zu sein (für UK: RIN 2008, 22, 37, 53). Auf der einen Seite stehen die Großprojekte, die quantitative Massendatensätze produzieren. In Deutschland zählen hierzu insbesondere das SOEP, der ALLBUS, der Mikrozensus und neuerdings SHARE, PAIRFAM, NEPS. Diese Daten sind in der Regel gut dokumentiert und aufbereitet, da ihre Sekundärnutzung von vorneherein angestrebt und berücksichtigt wurde (vgl. Habich et al. 2010, 7). Auf der anderen Seite stehen Daten aus kleineren Forschungsprojekten, qualitative Daten, Daten auf Basis komplexerer Forschungsdesigns oder Daten zu sog. speziellen Populationen, die in weit geringerem Maße zur Verfügung stehen.³³ In kleineren Forschungsprojekten wird bislang eine spätere Datenweitergabe zumeist nicht eingeplant – teilweise mit dem Argument, die eigenen Daten seien nicht relevant genug für eine Sekundärnutzung –, weshalb die hierfür notwendigen Vorkehrungen und Vorarbeiten nicht Teil der üblichen Arbeits- und Zeitplanungen sind. Hinzu kommt, dass Datenmanagement im Fall besonderer Forschungsdesigns vergleichsweise schwieriger ist, wenn anwenderfreundliche Standards fehlen. Forschungsdaten aus dieser Art von Projekten sind demzufolge oft nicht ausreichend dokumentiert und/oder aufbereitet, so dass nicht-beteiligte Dritte diese nicht ohne weiteres nutzen könnten (GDF 2010b, 19). Erfahrungen des DSZ-BO zeigen weiter, dass häufig Untersuchungsteilnehmerinnen Zusagen gemacht, die eine erneute Nutzung der Daten durch Dritte ausschließen. Die Anwendung eines Datenmanagements von geringer Qualität resultiert in Datensätzen, die vielleicht „gerade noch so“ für die eigenen Zwecke brauchbar sind - aber nicht jenseits davon, sowie innerhalb des Projektzeitraums - aber nicht darüber hinaus.

33 Letztere stehen im Fokus vom DSZ-BO, aber auch der GESIS, des IQB und des PsychData. Bei dieser Art von Daten sind sowohl das Verfügbarkeits- als auch das Nutzbarkeitsproblem virulenter. Die bisherigen Erfahrungen zeigen, dass es hier schwieriger ist, ein Angebot zu schaffen, weil zuvor aufwendige Recherche- und Akquisetätigkeiten nötig sind, weil Datenproduzenten der Weitergabe der eigenen Daten teilweise ablehnend gegenüber stehen, und weil die Daten häufig nicht die Mindestvoraussetzungen für eine Weitergabe erfüllen.

Insbesondere für solche Datenzentren, die das Spektrum jenseits der Massendatensätze abdecken, ist es sinnvoll mit ihren Dienstleistungen nicht nur auf die Akquise von Daten und deren Angebot zu zielen, sondern Datenproduzenten schon frühzeitig mit Blick auf späteres Data Sharing zu unterstützen und zu qualifizieren. Eine Reihe von Datenzentren haben die Notwendigkeit Datenmanagementkenntnisse zu verbreiten erkannt und ihre Dienstleistungen dementsprechend erweitert. Das UKDA hat dem Thema Datenmanagement in der neuesten Auflage seiner grundlegenden Veröffentlichung „Managing and Sharing Data“ (van den Eyden et al. 2011) im Vergleich zur vorhergehenden Ausgabe (2009) ein eigenes Kapitel gewidmet und diese um eine Checkliste im Anhang ergänzt. Daneben bietet das UKDA mehrmals im Jahr Workshops in und außerhalb der UK sowie ausführliche Informationen auf der Homepage (www.data-archive.ac.uk/create-manage). Auch das schottische Digital Curation Center (DCC) und das US-amerikanische Datenarchiv ICPSR bieten sehr umfassende Informationen zum Datenmanagement und zur Erstellung von Datenmanagementplänen (www.dcc.ac.uk/digital-curation). In Deutschland bietet die GESIS seit Herbst 2011 mit einem eigens eingerichteten Training Center Beratungen zu Datenmanagement und Datenarchivierung an (www.gesis.org/archive-and-data-management-training-and-information-centre) und im Jahr 2011 ist erstmals ein umfassendes Handbuch zum Thema Forschungsdatenmanagement erschienen (Büttner et al. 2011).

Auch in den Empfehlungen des RatSWD zeigt sich die Bedeutung und Notwendigkeit einer Professionalisierung des Datenmanagements (GDF 2011). Hierin werden sowohl die Verbesserung der Datendokumentation als auch die stärkere Diskussion und Berücksichtigung datenschutzrechtlicher Fragen und ethischer Aspekte genannt. Ausführliche Datendokumentationen werden dabei als Teil eines Qualitätsmanagements betrachtet. Auch wenn die verschiedenen Punkte einzeln und unabhängig voneinander diskutiert werden können, verlangt ihre Umsetzung im jeweiligen Forschungsprojekt eine *gemeinsame* Betrachtung, wie sie im Rahmen des Datenmanagements erfolgt.

Defizit Ressourcen

Gibt es den erläuterten Zusammenhang *professionelles Datenmanagement – Zunahme Data Sharing – Zunahme an Sekundärforschung*, kann Datenmanagement indirekt zu einer Einsparung von Ressourcen führen, die mit der Substitution von Primär- durch Sekundärforschung und gezielter einsetzbaren Primärerhebungen einhergeht. Jenseits dieses Vorteils für die Wissenschaft insgesamt, sollte Datenmanagement aber auch für den einzelnen Forscher ressourcensparend sein. Das gilt vor allem für den Vergleich der *nachträglichen* Nutzbarmachung von Daten für Data Sharing mit *forschungsbegleitendem* Datenmanagement. Forschungsbegleitendes Datenmanagement ist mit vergleichsweise geringerem Aufwand verbunden (z.B.: Long 2009: 345). Auch im Rahmen anderer Nutzungszwecke als derjenigen des Data Sharing sind Ressourceneinsparungen möglich. Bei Long finden sich die folgenden Beispiele aus der Praxis: „You have multiple versions of a file and do not know which is which. You cannot find a file and think you might have deleted it. You and a colleague are not sure which draft of your paper is the latest or find that there are two different "latest" drafts. You want the final version of the questionnaire and are not sure which file it is because two versions of the questionnaire include "final" in the name.“ (Long 2009: 19).

Systematisches Datenmanagement kann zeitaufwendige Suchen nach Dateien, die nicht nachvollziehbar benannt oder abgelegt wurden, ersparen. Auch kollaboratives Arbeiten wird erleichtert. Vor dem Hintergrund der möglichen Effizienzgewinne, könnten fehlende Ressourcen nur unter der Bedingung unzureichender Kenntnisse ein wirkliches Defizit darzustellen. Sobald die erforderlichen Kenntnisse vorliegen, könnte aus dem vermeintlichen Defizit ein Anreiz werden.

5 Ausblick

Gesellschaftliche Institutionen wie die *Forschung*, die öffentliche Gelder ausgeben, stehen immer auch vor einem Legitimierungsdruck gegenüber der Öffentlichkeit. Objektivität, Transparenz, Effizienz sowie die größtmögliche Amortisation der eingesetzten Mittel sind dabei die Bewertungskriterien der Wahl. Data Sharing ist von besonderer Bedeutung für den Erfolg und die Legitimität heutiger Wissenschaft und wie aufgezeigt wurde, ist der Erfolg des Data Sharing eng an die Professionalität und Verbreitung von Datenmanagement gekoppelt. Datenmanagement ist bislang in dem erforderlichen Ausmaß allerdings kaum verbreitet, und die fehlende Nutzbarkeit von Forschungsdaten ist daher ein großes Hemmnis bezüglich der Archivierung und Weitergabe von Daten für die Sekundärforschung.

Jedes Forschungsprojekt verwaltet (i.S.v. manage) zwar auch jetzt schon die eigenen Daten, dies geschieht jedoch nicht systematisch und umfassend genug: weder für die Zwecke des Data Sharing, noch für sonstige Zwecke gemessen an hohen Qualitätsstandards. Wenn aber Data Sharing die beschriebene Bedeutung zukommt und ein förderpolitisches Ziel ist, das mit steigender Tendenz eingefordert wird, dann sind Investitionen in die Förderung professionellen Datenmanagements von allen Beteiligten unerlässlich. Auch im internationalen Vergleich insbesondere mit dem angelsächsischen Raum, ist abzusehen, dass sich die Forschungs- und Förderlandschaft in Deutschland hinsichtlich des Umgangs mit Forschungsdaten zukünftig wandeln wird müssen.

Die genannten Defizite im Hinblick auf Anreize, Kenntnisse und Ressourcen sind dabei im besten Fall gleichermaßen zu adressieren. Förderer, Datenzentren und die Forschung selbst sind hierbei gefragt.

Förderer fällt die Aufgabe zu, notwendige finanzielle Mittel bereitzustellen und Lösungen zu entwickeln, den Widerspruch zwischen dem gewünschten und geforderten *long-term data sharing* und dem praktizierten *short-term funding* zu überbrücken: direkt durch die bessere Ausstattung von Forschungsprojekten und indirekt über die Finanzierung neuer Aufgabenbereiche von Datenzentren. Nachwuchswissenschaftlern muss die Gelegenheit gegeben werden, sich die neuen Kenntnisse anzueignen und die neuen Aufgaben wahrzunehmen. Qualifizierungsangebote von Datenzentren oder Universitäten müssen finanziert werden. Zudem können Förderer mit gezielten Maßnahmen auf die Anreizsituation einwirken, wobei hier neben Verpflichtungen und Sanktionen solche positiver Art nicht vernachlässigt werden sollten. Die Rolle von *Datenzentren* kann vor allem darin liegen durch ihre fachspezifischen Kenntnisse, den notwendigen Ausbau entsprechender Qualifikationen zu fördern und zu einer Sensibilisierung für die Thematik beizutragen, um den angestrebten Wandel der Anforderungen an Wissenschaftler und deren Aufgabenspektrum zu stützen (vgl. Huschka et al. 2011, 41 ff.). Eine Rolle, die Datenzentren außerhalb Deutschlands bereits verstärkt wahrnehmen. Das oben angesprochene Funktionsspektrum von Datenzentren (Informations-, Beratungs-, Bereitstellungs-, Archivierungsfunktion) um eine solche Qualifizierungsfunktion zu erweitern, liegt auch im Eigeninteresse von Datenzentren, da dies das potentielle Angebot an Daten erhöhen würde und die aufwendigeren, nachträglichen Arbeiten zur Herstellung von Nutzbarkeit verringerte. Insbesondere für kleinere, themenspezifische Datenzentren ist dies zentral. *Forscher* sind gefragt, nicht nur die Bereitschaft zu Data Sharing mitzubringen, sondern auch sich die fehlenden Kenntnisse des Datenmanagements anzueignen und Ressourcen dementsprechend einzusetzen. Ein hilfreiches Argument ist dabei sicherlich, dass Datenmanagement nicht nur eine zentrale Voraussetzung für erfolgreiches Data Sharing darstellt, sondern auch für die Forschungsprojekte selbst mit Vorteilen verbunden (und sogar zunehmend essentiell) ist. Diese Vorteile sind zu betonen und im besten Fall mit entsprechenden empirischen Studien zu belegen.

Mit der Entwicklung hin zum Data Sharing werden Forschungsdaten zunehmend zu einem eigenständigen und bedeutenden Gut wissenschaftlicher Forschung, sozusagen zu einem

Gut mit eigener „Produktqualität“. Sie etablieren sich damit neben dem bisher einzigen Outcome der Wissenschaft, der Publikation. Güte und Verbreitung professionellen Datenmanagements ist auch die Voraussetzung dafür, ob Daten der neuen Rolle gerecht werden können. Forschungsdaten als eigenständiger Outcome, die Einführung von Datenmanagement und Data Sharing können als Aspekte der Entwicklung hin zu einer *neuen Stufe der Professionalisierung* in der Forschung betrachtet werden. Für die Wissenschaft insgesamt sind damit möglicherweise Vorteile wie effizienter Einsatz von Ressourcen, erhöhte Chancen für Kooperationen und Kollaborationen auch über Disziplingrenzen hinweg, erhöhte Datenqualität und Transparenz wissenschaftlicher Ergebnisse verbunden. Wie weit diese Entwicklung geht, ob es gar zu einer Ausdifferenzierung wissenschaftlicher Arbeit (hin zu Datenspezialisten und Wissenschaftsdienstleistern, siehe bspw. Pampel et al. 2010) kommt, bleibt abzuwarten. Ob und wie dieser Prozess gelingt, wird aber in jedem Fall dazu beitragen, ob die bundesdeutsche, empirische Sozialforschung international wettbewerbsfähig bleibt und könnte nicht zuletzt auch Akzeptanz und Ansehen in der Bevölkerung beeinflussen.

6 Literatur

- Abbott, Andrew (2007): Notes on Replication. In: *Sociological Methods & Research*. 2007 (36), 210. DOI: 10.1177/0049124107306664
- Allianz-Initiative (2008): Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen. Online: <http://www.allianz-initiative.de> (Zugegriffen: 29. Mai 2012).
- Antoni, Manfred, Katrin Drasch, Corinna Kleinert, Britta Matthes, Michael Ruland, Annette Trahms (2010): *Arbeiten und Lernen im Wandel, Teil I: Überblick über die Studie*. FDZ Methodenreport 05/2010 (de), Nürnberg. (2., aktualisierte Fassung, März 2011)
- Askitas, Nikos (2010): What Makes Persistent Identifiers Persistent? In: *RatSWD Working Paper Series*. Juni 2010 (147). Online: http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_147.pdf (Zugegriffen: 29. Mai 2012).
- Beagrie, Charles (2010): *User Guide for Keeping Research Data Safe: Assessing Costs/Benefits of Research Data Management, Preservation and Re-use*. Version 1.0. Online: http://www.beagrie.com/KeepingResearchDataSafe_UserGuide_v1_Dec2010.pdf (Zugegriffen: 29. Mai 2012).
- Biemer, Paul und Lars E. Lyberg (2003): *Introduction to Survey Quality*. Hoboken, NJ: Wiley Interscience.
- Brase, Jan (2010): DataCite - A global registration agency for research data. In: *RatSWD Working Paper Series*. Juli 2010 (149). Online: www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_149.pdf (Zugegriffen: 29. Mai 2012).
- Büttner, Stephan, Hobohm, Hans-Christoph, Müller, Lars (Hg.) (2011): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock+Herchen.
- Büttner, Stephan, Hobohm, Hans-Christoph, Müller, Lars (2011): Research Data Management. In: Büttner, Stephan, Hobohm, Hans-Christoph, Müller, Lars (Hg.): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock+Herchen, S. 13-24.
- Corti, Louis and Bethany Brett (2011): Helping empower researchers and their institutions to manage and share research data. Unterlagen des Workshops WE1, erhalten im Rahmen des Workshops auf der Konferenz IASSIST 2011. www.rdl.sfu.ca/IASSIST/index.php/Program/category/tuesday_workshops/#we1
- Dale, Angela; Arber, Sara; Procter, Michael (2004): A Sociological Perspective on Secondary Analysis. In: Clive Seale (Hg.): *Social research methods. A reader*. London: Routledge (Routledge student readers), S. 136–140.
- Dallmeier-Tiessen, Sünje (2011): Strategien bei der Veröffentlichung von Forschungsdaten. In: *RatSWD Working Paper Series*. März 2011 (173). Online: http://www.ratswd.de/download/RatSWD_WP_2011/RatSWD_WP_173.pdf (Zugegriffen: 06. Juni 2012).
- DFG (Deutsche Forschungsgemeinschaft) (1998): *Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“*; Denkschrift. Weinheim: Wiley-VCH.
- DFG (Deutsche Forschungsgemeinschaft) (2010): *DFG-Vordruck 1.02-8/10, Merkblatt für Anträge auf Sachbeihilfen mit Leitfaden für die Antragstellung*. Online: www.dfg.de/download/programme/sachbeihilfe/antragstellung/1_02/1_02.pdf, (Zugegriffen 14.05.2011)

- DFG (Deutsche Forschungsgemeinschaft) (2011): DFG-Vordruck 1.01 – 10/11. Online: http://www.dfg.de/formulare/1_01/1_01.pdf, S.3 (Zugegriffen am 27.06.2012)
- DFG (Deutsche Forschungsgemeinschaft) (2012): Leitfaden für die Antragstellung – Projektanträge, DFG-Vordruck 54.01-1/12. Online: http://www.dfg.de/formulare/54_01/54_01_de.pdf, (Zugegriffen am 27.06.2012).
- Edler, Susanne, Alexia Meyermann, Jelena Hohlweg und Henrik Pruisken (2012): Hidden Data Treasures in German Organizational Research – Enhancing Visibility of Organizational Data. Vortrag, präsentiert auf der International Conference on Establishment Surveys, 11.-14.06.2012, Montreal, Kanada.
- ESRC (Economic and Social Research Council) (2010): ESRC Research Data Policy. Online: http://www.esrc.ac.uk/_images/Research_Data_Policy_2010_tcm8-4595.pdf (Zugegriffen: 05. Juni 2012).
- Freese, Jeremy (2007): Replication Standards for Quantitative Social Science. In: *Sociological Methods & Research* 36 (2), S. 153–172. Online: <http://smr.sagepub.com/content/36/2/153.full.pdf+html> (Zugegriffen: 05. Jun. 2012).
- Furman, Jeffrey L.; Stern, Scott (2008): Climbing atop the shoulders of giants: The impact of institutions on cumulative research. Paper presented at the CAMS Seminar. Online: http://www.en.cams.bwl.uni-muencheffurmann.de/files/furmanstern_climbing.pdf (Zugegriffen: 11. Juni 2012).
- GDF (German Data Forum) (2010a): Building on progress. Expanding the research infrastructure for the social, economic, and behavioral sciences. Opladen: Budrich UniPress.
- GDF (German Data Forum) (RatSWD) (2010b): Recommendations for Expanding the Research Infrastructure for the Social, Economic, and Behavioral Sciences. In: *RatSWD Working Paper Series* July (150). Online: www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_150.pdf (Zugegriffen: 05. Juni 2012).
- Green, Ann G.; Gutmann, Myron P. (2007): Building Partnerships Among Social Science Researchers, Institution-based Repositories and Domain Specific Data Archives. In: *OCLC Systems & Services* 23 (1), S. 35–53.
- Gregory, Arofan; Heus, Pascal; Ryssevik, Jostein (2009): Metadata. In: *RatSWD Working Paper Series*. March 2009 (57). Online: http://www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_57.pdf (Zugegriffen: 05. Juni 2012).
- Kommission Zukunft der Informationsinfrastruktur (KII) (2011): Gesamtkonzept für die Informationsinfrastruktur in Deutschland. Online: http://www.gwk-bonn.de/fileadmin/Papers/KII_Gesamtkonzept.pdf (Zugegriffen: 29.06.2011)
- Habich, Roland; Himmelreicher, Ralf K.; Huschka, Dennis (2010): Zur Entwicklung der Dateninfrastruktur in Deutschland. In: *RatSWD Working Paper Series* September 2010 (157). Online: http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_157.pdf (Zugegriffen: 05. Juni 2012).
- Häder, Michael (2009): Der Datenschutz in den Sozialwissenschaften. Anmerkungen zur Praxis sozialwissenschaftlicher Erhebungen und Datenverarbeitung in Deutschland. In: *RatSWD Working Paper Series*. Juni 2009 (90) Online: http://www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_90.pdf (Zugegriffen: 11. Juni 2012)
- Haeussler, Carolin (2010): Information-Sharing in Academia and the Industry: A Comparative Study. forthcoming Research Policy. In: *RatSWD Working Paper Series*. August 2010 (154). Online: www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_154.pdf (Zugegriffen: 05. Juni 2012).

- Hausstein, Brigitte (2012): Die Vergabe von DOI-Namen für Sozial- und Wirtschaftsdaten. Serviceleistungen der Registrierungsagentur da|ra. In: *RatSWD Working Paper Series*. Jan. 2012 (193). Online: www.ratswd.de/download/RatSWD_WP_2012/RatSWD_WP_193.pdf (Zugegriffen: 05. Juni 2012).
- Huschka, Denis, Oellers Claudia, Ott, Notburga, Wagner, G. Gert (2011): Datenmanagement und Data Sharing: Erfahrungen in den Sozial- und Wirtschaftswissenschaften. In: Büttner, Stephan, Hobohm, Hans-Christoph, Müller, Lars (Hg.): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock+Herchen, S. 35-48.
- IASSIST 37th Annual Conference (International Association for Social Science Information Services & Technology) (2011): Data Science Professionals: A Global Community of Sharing. 30.05.-03.06.2011, Vancouver, Kanada.
- ICPSR - Inter-University Consortium for Political and Social Research (2009): Principles and Good Practice for Preserving Data. In: *International Household Survey Network, IHSN Working Paper* December 2009 (003).
- ICPSR - Inter-University Consortium for Political and Social Research (o.J.): Guidelines for Effective Data Management Plans. Online: www.icpsr.umich.edu/files/ICPSR/dmp/DataManagementPlans-All.pdf (Zugegriffen: 05. Juni 2012).
- ICPSR - Inter-University Consortium for Political and Social Research (2011a): Appraisal Criteria in Detail. Online: www.icpsr.umich.edu/icpsrweb/ICPSR/curation/appraisal.jsp (Zugegriffen: 05. Juni 2012).
- ICPSR - Inter-University Consortium for Political and Social Research (2011b): Framework for Creating a Data Management Plan. Online: www.icpsr.umich.edu/icpsrweb/content/ICPSR/dmp/framework.html (Zugegriffen: 05. Juni 2012).
- ICPSR - Inter-University Consortium for Political and Social Research (2011c): Deposit Data and Findings. Online: www.icpsr.umich.edu/icpsrweb/ICPSR/access/deposit/index.jsp (Zugegriffen: 05. Juni 2012).
- ICPSR - Inter-University Consortium for Political and Social Research (2011d): Disaster Planning. Online: www.icpsr.umich.edu/icpsrweb/ICPSR/curation/disaster/___index.jsp (Zugegriffen: 05.06.2012).
- IHSN- International Household Survey Network (2009a): Microdata Management Toolkit. Online: www.surveynetwork.org/toolkit (Zugegriffen: 05. Juni 2012).
- IHSN- International Household Survey Network (2009b): Microdata documentation. Online: www.surveynetwork.org/home/index.php?q=tools/documentation (Zugegriffen: 05. Juni 2012).
- Jensen, Uwe (2011): Datenmanagementpläne. In: Büttner, Stephan, Hobohm, Hans-Christoph, Müller, Lars (Hg.): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock+Herchen, S. 71-82.
- Key perspectives Ltd (2010): Data dimensions: disciplinary differences in research in data sharing, reuse and long term viability. A comparative review based on sixteen case studies. DCC Scarp Synthesis Report, Version No. 1.1. Ed. by Digital Curation Centre. Online: www.dcc.ac.uk/sites/default/files/SCARP%20SYNTHESIS_FINAL.pdf (Zugegriffen: 05. Juni 2012).
- King, Gary (1995): Replication, Replication. In: *PS: Political Science & Politics* 28 (3), S. 444–452.
- Klump, Jens (2011): Langzeiterhaltung digitaler Forschungsdaten. In: Büttner, Stephan, Hobohm, Hans-Christoph, Müller, Lars (Hg.): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock+Herchen, S. 115-122.

- Kreuter, Frauke; Casas-Cordero, Carolina (2010): Paradata. In: *RatSWD Working Paper Series* April (136). Online: www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_136.pdf (Zugegriffen: 05. Juni 2012).
- Long, J. Scott (2009): *The workflow of data analysis using Stata*. College Station, Tex: Stata Press.
- NSF - National Science Foundation (2010): *Data Management for NSF SBE Directorate Proposals and Awards*. Online: www.nsf.gov/sbe/SBE_DataMgmtPlanPolicy.pdf (Zugegriffen: 05. Jun. 2012).
- NERC - Natural Environment Research Council (2011): *NERC data policy*. Online: www.nerc.ac.uk/research/sites/data/policy2011.asp (Zugegriffen: 05. Jun. 2012).
- nestor-Arbeitsgruppe „Vertrauenswürdige Archive – Zertifizierung“ (Hg.) (2008): *nestor-Kriterien - Kriterienkatalog vertrauenswürdige digitale Langzeitarchive*. Version 2. Nestor-Materialien 8. Frankfurt am Main: nestor c/o Deutsche Nationalbibliothek. Online: http://files.d-nb.de/nestor/materialien/nestor_mat_08.pdf (Zugegriffen: 27.06.2012)
- Oellers, Claudia; Wegner Eva (2009): Does Germany Need a (New) Research Ethics for the Social Sciences? In: *RatSWD Working Paper Series*. June 2009 (86). Online: www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_86.pdf (Zugegriffen: 06. Juni 2012).
- Organisation for Economic Co-operation and Development (2007): *OECD Principles and Guidelines for Access to Research Data from Public Funding*. OECD. Online: www.oecd.org/dataoecd/9/61/38500813.pdf (Zugegriffen: 06. Jun. 2012).
- Opitz, Diane und Reiner Mauer (2005): Erfahrungen mit der Sekundärnutzung von qualitativem Datenmaterial – Erste Ergebnisse einer schriftlichen Befragung im Rahmen der Machbarkeitsstudie zur Archivierung und Sekundärnutzung qualitativer Interviewdaten. In: *Forum Qualitative Sozialforschung* 6(1), Art. 43, www.qualitative-research.net/fqs-texte/1-05/05-1-43-d.htm.
- Pampel, Heinz, Roland Bertelmann und Hans-Christoph Hobohm (2010): „Data Librarianship“ – Rollen, Aufgaben, Kompetenzen. In: *RatSWD Working Paper Series*. Mai 2010 (144). Online: www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_144.pdf (Zugegriffen: 06. Juni 2012).
- Pflueger, Jessica; Pongratz, Hans J.; Trinczek, Rainer (2010): Fallstudien in der deutschen Arbeits- und Industriesoziologie. In: Hans J. Pongratz (Hg.): *Industriesoziologische Fallstudien. Entwicklungspotenziale einer Forschungsstrategie*. 1. Auflage. Berlin: Ed. Sigma, S. 23–67.
- Piwowar, Heather A.; Day, Roger S.; Fridsma, Douglas B. (2007): Sharing Detailed Research Data Is Associated with Increased Citation Rate. In: *PLoS ONE* 2 (3), Online: <http://dx.plos.org/10.1371/journal.pone.0000308> (Zugegriffen: 06. Juni 2012).
- Presser, Stanley (1994): Informed Consent and Confidentiality in Survey-Research. In: *Public Opinion Quarterly* 58 (3), S. 446–459.
- Pullinger, John; Wagner, Gert G. (2010): On the Respective Roles of National Libraries, National Archives and Research Data Centers in the Preservation of and Access to Research Data. In: *RatSWD Working Paper Series*. August 2010 (153). Online: www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_153.pdf (Zugegriffen: 05. Juni 2012).
- RIN (Report commissioned by the Research Information Network) (2008): *To Share or Not To Share: Publication and Quality Assurance of Research Data Outputs - Main Report*. Online: www.rin.ac.uk/system/files/attachments/To-share-data-outputs-report.pdf (Zugegriffen: 29.06.2012)
- Sala, Emanuela; Burton, Jonathan; Knies, Gundi (2010): Correlates of obtaining informed consent to data linkage: respondent, interview and interviewer characteristics. In: *ISER Working Paper Series*. Aug. 2010 (28). Online: www.iser.essex.ac.uk/publications/working-papers/iser/2010-28.pdf (Zugegriffen: 05. Juni 2012).

- Schaar, Peter (2009): Data protection and statistics – a dynamic and tension-filled relationship. In: *RatSWD Working Paper Series*. März 2009 (82). Online: www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_82.pdf (Zugegriffen: 05. Juni 2012).
- Seale, Clive (2009): Using Data Archives for Secondary Analysis. In: Clive Seale (Hg.): *Researching society and culture*. 2. ed., Repr. London: Sage, S. 355–365.
- Siegert, Olaf (2010): Speicherung und Publikation von Forschungsdaten. In: *RatSWD Working Paper Series*. Sep. 2010 (158). Online: www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_158.pdf (Zugegriffen: 05. Juni 2012).
- Sin, C. H. (2005): Seeking informed consent: Reflections on research practice. In: *Sociology- The Journal of the British Sociological Association* 39 (2), S. 277–294.
- Singer, E. (1978): Informed Consent-Consequences for Response Rate Response Quality in Social Survey. In: *American Social Review* 43 (2), S. 144–162.
- Spengler, Anja und Kerstin Lorek (2010). Verknüpfung und Abgleiche von Unternehmensregisterdaten des Statistischen Bundesamts mit Betriebsdaten des Instituts für Arbeitsmarkt- und Berufsforschung. FDZ Methodenreport 01/2010 (de), Nürnberg. (aktualisierte Fassung vom 13.08.2010)
- Spindler, Gerald und Tobias Hillegeist (2011). Rechtliche Probleme der elektronischen Langzeitarchivierung von Forschungsdaten. In: Büttner, Stephan, Hobohm, Hans-Christoph, Müller, Lars (Hg.): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock+Herchen, S. 63–69.
- Thomson, Denise; Bzdel, Lana; Golden-Biddle, Karen; Reay, Trish; Estabrooks, Carole A. (2005): Central Questions of Anonymization: A Case Study of Secondary Use of Qualitative Data. In: *Forum: Qualitative Social Research* 6 (1), S. 1–16. Online: <http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=16513431&site=ehost-live> (Zugegriffen: 05. Juni 2012).
- van den Eynden, Veerle; Bishop, Libby; Horton, Laurence; Corti, Louise (2010): Data Management Practices in the Social Sciences. Hg. v. UK Data Archive.
- van den Eynden, Veerle; Corti, Louise; Woollard, Matthew; Bishop, Libby; Horton, Laurence (2011): Managing and sharing data. A best practice guide for researchers. 3rd. Edition, fully revised. Colchester: UK Data Archive. Online: www.data-archive.ac.uk/media/2894/managingsharing.pdf (Zugegriffen: 05. Juni 2012).
- Weichselgartner, Erich (2011): Disziplinspezifische Aspekte des Archivierens von Forschungsdaten am Beispiel der Psychologie. In: *RatSWD Working Paper Series*. Dez. 2011 (179). Online: www.ratswd.de/download/RatSWD_WP_2011/RatSWD_WP_187.pdf (Zugegriffen: 05. Juni 2012).
- Wiles, Rose; Crow, Graham; Charles, Vikki; Heath, Sue (2007): Informed consent and the research process: Following rules or striking balances? In: *Sociological Research online* 12 (2). Online: www.socresonline.org.uk/12/2/wiles.html (Zugegriffen: 05. Jun. 2012).
- Winkler-Nees, Stefan (2011). Anforderungen an wissenschaftliche Informationsinfrastrukturen. In: *RatSWD Working Paper Series*. Juni 2011 (180). Online: www.ratswd.de/download/RatSWD_WP_2011/RatSWD_WP_180.pdf (Zugegriffen: 27.06.2012)
- WISDOM - Wiener Institut für sozialwissenschaftliche Dokumentation der Methodik (o.J.): Checklist. WISDOM - Wiener Institut für Sozialwissenschaftliche Dokumentation und Methodik. Online: www.wisdom.at/Datendokumentation/Da_Datenabgabe.aspx (Zugegriffen: 05. Jun. 2012).
- WISDOM - Wiener Institut für sozialwissenschaftliche Dokumentation der Methodik (o.J.a): Online: www.wisdom.at/Datendokumentation/PDF/FormularDatenbeschreibungQuantitativ.pdf (Zugegriffen: 05. Jun. 2012).

WISDOM - Wiener Institut für sozialwissenschaftliche Dokumentation der Methodik (o.J.b): Online:
www.wisdom.at/Datendokumentation/PDF/FormularDatenbeschreibungQualitativ.pdf
(Zugegriffen: 05. Jun. 2012).

Wissenschaftsrat (2011): Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und
Sozialwissenschaften des Wissenschaftsrats (01/2011) Online:
www.wissenschaftsrat.de/download/archiv/10465-11.pdf (Zugegriffen: 21.05.2012)

7 Anhang

Im Folgenden ist der Inhalt eines Datenmanagementplans nach den Vorgaben des britischen UKDA (Corti/Brett 2011) aufgeführt.

CONSENT
Will you be collecting personal, confidential or sensitive information?
Are you gaining informed consent for participation from research participants?
Are you discussing data sharing or archiving with the informants from whom you collect the data?
Are you gaining consent from informants to share data beyond your research?
Are you asking for written or verbal consent?
ANONYMISATION
Do you need to anonymise data e.g. to remove identifying information or personal data? You will need to think about when this will be done and how.
COPYRIGHT
If you are purchasing or re-using someone else's data sources have you considered how that data will be shareable?
Have you established who owns the copyright of your data? Might there be joint copyright?
DOCUMENTATION
How will you label and organize your data and files?
Are your data self-explanatory in terms of variable names, codes and abbreviations used? Would someone else be able to understand your data?
Do you have documentation which explains what your data means or gives context to your data? i.e. where an interview was carried out and with whom? Variable lists?
Have you considered how you will ensure the quality of your data? i.e. Data cleaning, checking transcriptions, data entry validation etc.
FORMATTING AND VERSION CONTROL
Which data formats will you use? i.e. Word, MaxQDA, SPSS etc Are these data formats likely to still be available to use in 10 years time?
If data are held in various places, how will you keep track of versions?
Will you know what the master version of your data files is?
TRANSCRIPTION
What conventions will you use for transcribing the data?
Who will do the transcription?
How will you ensure the confidentiality and security of your data during the transcription process?
SECURITY
If your data is held on a USB stick or other mobile storage device how will you ensure the safety and security of the data on it?
Are your files backed up regularly and are back-ups stored safely?
How will you securely store personal or sensitive data? (Digital and hard copy data) Do you need to keep secure all or just part of your data?
Do you have plans for how your data will be preserved once you have finished your research?
DATA ACCESS
Who has access to which data during and after research?
Are various access regulations needed?

Quelle: Corti/Brett 2011