

2

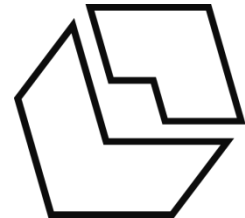
forschungsdaten
bildung **informiert**

Jessica Trixa und Thomas Ebel

jessica.trixa@gesis.org // thomas.ebel@gesis.org

**„Einführung in das Forschungs-
datenmanagement in der
empirischen Bildungsforschung“**

Zusammenfassung des Workshops im
Rahmen des CESSDA-
Trainingsprogramms



Version 1.0 // August 2015

Impressum

forschungsdaten bildung informiert // Nr. 2 (2015)

Herausgeber

Forschungsdatenzentrum (FDZ) Bildung am DIPF
Deutsches Institut für Internationale Pädagogische Forschung
Schloßstraße 29
60486 Frankfurt am Main

Redaktion und Layout

Alexander Schuster

Rechte

Nachdruck, auch auszugsweise, nur mit Genehmigung des FDZ Bildung gestattet.

www.forschungsdaten-bildung.de

Inhalt

Vorbemerkung	4
1 Modul 1: Einführung - Was ist Forschungsdatenmanagement (FDM)?.....	6
2 Modul 2: Metadaten und Dokumentation	10
3 Modul 3: Ethik, Datenschutz und Anonymisierung	12
4 Modul 4: Datenadministration und -speicherung.....	15
5 Modul 5: Urheberrecht und Lizenzierung	17
6 Modul 6: Datenweitergabe und Nachnutzung	18

Vorbemerkung

Diese Veröffentlichung fasst den Workshop mit dem Titel „Einführung in das Forschungsdatenmanagement in der empirischen Bildungsforschung“ zusammen, der vom Verbund Forschungsdaten Bildung¹ entwickelt und von CESSDA Training (GESIS) am 18./19. Mai 2015 am Kölner GESIS-Standort durchgeführt wurde.

Das Konzept des Workshops berücksichtigt die relevanten Themen des Forschungsdaten-managements und umfasst sechs Module:

- » Modul 1: Einführung - Was ist Forschungsdatenmanagement (FDM)?
- » Modul 2: Metadaten und Dokumentation
- » Modul 3: Ethik, Anonymisierung und Datenschutz
- » Modul 4: Datenadministration und -speicherung
- » Modul 5: Urheberrecht und Lizenzierung
- » Modul 6: Datenweitergabe und Nachnutzung

Forschungs- und die dazugehörigen Kontextdaten (Metadaten) sind zentrale Ressourcen für die Wissenschaft und der Umgang mit ihnen sollte dementsprechend verantwortungsvoll, systematisch und in wissenschaftlich adäquater Form erfolgen. Mit dem Begriff Forschungsdatenmanagement werden alle Schritte umfasst, die der Sicherung, Handhabung und Dokumentation der Daten dienen. Spätestens seit **Förderinstitutionen** auf nationaler und internationaler Ebene ihre Antragsbedingungen zunehmend um eine systematische Beschreibung davon, wie mit den generierten Daten während der Laufzeit und nach Abschluss des Projekts umgegangen werden soll, erweitern, sind die Themen Forschungsdatenmanagement und Datenmanagementpläne auch für die Forschenden von praktischer Relevanz (Jensen 2012).

Viele Abläufe und Maßnahmen eines systematischen Forschungsdatenmanagements sind unabhängig von der Forschungsdisziplin auf erhobene Forschungsdaten anwendbar. Die Daten im Feld der empirischen Bildungsforschung zeichnen sich aber durch einige Spezifika aus, auf die die Workshops eingehen werden. Unter diesen Besonderheiten sticht insbesondere die hohe Komplexität und Heterogenität der Forschungsdaten dieser Disziplin hinsichtlich ihrer methodischen, rechtlichen und organisatorischen Ausgangsbedingungen hervor (CESSDA Training, vgl. auch Stanat 2014). Die empirische Bildungsforschung zeichnet sich u. a. durch die Vielzahl der eingesetzten Erhebungsmethoden (quantitative und qualitative Methoden sowie diagnostische und Kompetenzmessungen) und den daraus resultierenden Datentypen (u. a. audiovisuelles Material, Transkripte von Interviews, quantitative Datensätze, Rohdaten medizinischer Diagnosewerkzeuge, bspw. EEG-Daten), der Schwierigkeit des Feldzuganges, insbesondere bei Studien mit Minderjährigen und Schülerinnen und Schülern und der Interdisziplinarität der Forschungsvorhaben aus. Diese Komplexität macht ein besonders vorausschauendes und strukturiertes Datenmanagement notwendig.

¹ Der Verbund Forschungsdaten Bildung besteht aus den drei Partnerinstituten DIPF, IQB und GESIS. Der Verbund wurde vom BMBF damit beauftragt, ein Trainingskonzept für eine Einführung in das Forschungsdatenmanagement zu erarbeiten und dieses in mehreren Workshops empirischen Forscherinnen und Forschern aus dem Bereich der Bildungsforschung vorzustellen. Weitere Informationen finden Sie unter www.forschungsdaten-bildung.de.

Die Verfasser danken herzlich Alexia Meyermann und Maike Porzelt (DIPF) sowie Claudia Neuendorf, Lisa Pegelow und Poldi Kuhl (IQB) für ihre Unterstützung.

Im Folgenden werden die Inhalte der Workshop-Module als Information für interessierte Forscherinnen und Forscher vorgestellt.

Literatur:

1. Stanat, P., 2014: Bereitstellung und Nutzung quantitativer Forschungsdaten in der Bildungsforschung: Memorandum des Fachkollegiums „Erziehungswissenschaft“ der DFG. Verfügbar unter:
www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/forschungsdaten_memorandum_fk_109.pdf, abgerufen am: 10.01.2015.
2. Jensen, U., 2012: Leitlinien zum Management von Forschungsdaten: Sozialwissenschaftliche Umfragedaten. GESIS Technical Reports, 2012/07. Verfügbar unter:
www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-07.pdf, abgerufen am: 25.09.2014.

1 Modul 1: Einführung - Was ist Forschungsdatenmanagement (FDM)?

Grundsätzlich bedeutet FDM die Pflege der eigenen Daten im Rahmen eines wissenschaftlichen, auf Forschungsdaten spezialisierten Projektmanagements. Dies beinhaltet die Entwicklung und Implementierung von Methoden und Richtlinien, welche dabei behilflich sein sollen, „die eigenen Arbeitsprozesse, die die Erzeugung von und den Umgang mit Forschungsdaten betreffen, möglichst effizient und zielorientiert zu organisieren und fortlaufend zu steuern“ (Meyermann 2012: 7).

Eine integrierte und aktive FDM-Strategie verfolgt das Ziel, alle datenrelevanten Fragen vom Studien- und Methodendesign über die Datenaufbereitung bis hin zur Datenarchivierung in die Planung eines Forschungsprojektes einzubeziehen und zu dokumentieren. **Alle Aspekte, die im Zusammenhang mit der Verwaltung, Organisation und Verarbeitung von Daten stehen, sind also Teil des Datenmanagements.** Darunter fallen nicht nur alle Aspekte des direkten Umgangs mit den Daten, bspw. die Aufbereitung und Bereinigung, sondern auch die Beschreibung und Kontextualisierung der Daten (*Metadaten*; Modul 2), bspw. die Beschreibung der Daten durch Begleitmaterialien, organisatorische und technische Maßnahmen (*Datenadministration*; Modul 4), bspw. Regeln zu Ordnerstrukturen, Datenformaten und Software, rechtliche Gesichtspunkte wie Datenschutz, Anonymisierung und ethische Fragen (Modul 3) und die Rechteverwaltung (*Urheberrecht und Lizenzierung und Datenweitergabe und Nachnutzung*, Module 5 und 6).

Unzureichendes Datenmanagement kann vielfältige Probleme hervorbringen. Diese reichen von datenschutzrechtlichen Problemen aufgrund unzulänglich anonymisierter personenbezogener² Daten über die unzureichende Dokumentation von Kontextinformationen bis hin zum Verlust der Ursprungsversionen von Dateien und der verstreuten Speicherung der Daten in verschiedenen Dateiversionen auf den Computern der Projektmitarbeiter/-innen.

Ein vorausschauend geplantes und entlang eines Datenmanagementplans (DMP) sorgfältig durchgeführtes FDM kann den wissenschaftlichen Forschungsprozess erleichtern und verbessern, denn es umfasst eine systematische Verwaltung der Daten, eine ständige Prüfung ihrer Vollständigkeit, Validität und Integrität und sorgt damit für wertvolle Forschungsdaten sowie transparente und empirisch nachvollziehbare Ergebnisse. Forschungsdatenmanagement sichert (1) die (langfristige) Interpretierbarkeit der Daten. Gut dokumentierte Daten können nicht nur von den Forschenden selbst zu einem späteren Zeitpunkt nachgenutzt werden, sondern (2) auch anderen Wissenschaftlern/-innen wird es ermöglicht, Forschungsdaten zu verstehen und nachzuvollziehen (Data Sharing). Außerdem erlauben dokumentierte Daten sowie abgesprochene und festgehaltene Abläufe (3) kollaboratives Arbeiten zwischen Wissenschaftlern. Ein sorgfältiger Umgang mit Forschungsdaten ist darüber hinaus (4) ein Qualitätssiegel für gute wissenschaftliche Praxis.

Forscher/-innen sollten bereits bei der Entwicklung ihrer Forschungsprojekte Aspekte des Datenmanagements definieren und einplanen. Sobald diese thematisiert und in einem DMP verschriftlicht wurden, können diese Prozesse noch vor Beginn der Datenerhebung strategisch umgesetzt werden. Vorausschauendes Handeln in der Vorphase des Forschungsprojektes kann den Archivierungsprozess wesentlich vereinfachen und dabei helfen, spätere Probleme in der Phase der Datenweitergabe bzw. -übergabe zu vermeiden. Umso früher Entscheidungen und Prozessabläufe im Verlauf des Forschungsdatenzyklus in einem Datenmanagement-

² Personenbezogene Daten sind definiert als Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbarer natürlicher Person (§3 Abs.1 Bundesdatenschutzgesetz, BDSG).

plan festgehalten werden, desto eher können diese Informationen zur Steuerung und Verwaltung der Studienprozesse und -ergebnisse genutzt und für die Nachnutzung der Daten im Anschluss an das Vorhaben in systematischer Form gebündelt zusammengefasst werden. Ein Beispiel hierfür sind die projektspezifischen Strategien zur Archivierung der Daten nach Projektende. Die hierbei zu beachtenden Aspekte und Fragen umfassen z. B. das Einholen von Einwilligungserklärungen der Studienteilnehmer, die Identifizierung und Anonymisierung personenbezogener und -beziehbarer Informationen, die langfristige Sicherung der Daten durch das eigene Institut oder Rechenzentrum oder die Abgabe der Daten an eine archivierende Organisation. Wird im Vorfeld bereits geklärt und festgehalten, wie mit der Archivierung umzugehen ist, ersparen sich die Projektmitarbeiter/-innen u. U. eine langwierige nachträgliche Implementierung einer Archivierungsstrategie, da die Archivierung bspw. durch das Einholen entsprechender Einwilligungen der Studienteilnehmer im Vorfeld rechtlich gesichert und vereinfacht werden kann. Ein frühzeitiges Datenmanagement bewahrt zudem vor Informationsverlusten, wenn zur Dokumentation erforderliche Angaben im Nachhinein nicht mehr konsistent und vollständig oder überhaupt nicht mehr zur Verfügung stehen, bspw. die Erinnerungen von Projektmitarbeiter/-innen fehlerbehaftet sind oder die Beteiligten die Stelle gewechselt haben (Meyermann 2012: 10).

Die Phasen des Forschungsdatenzyklus

Ein typischer Forschungsdatenzyklus eines empirischen Forschungsprojektes umfasst die Phasen Entwurf und Projektplanung, Planung des Untersuchungsdesigns, Datenerhebung und -aufbereitung, Datenauswertung, Publikation und Dokumentation sowie Archivierung und Bereitstellung der Daten (siehe Abb. 1). Wichtig ist dabei, dass FDM nicht erst einsetzt, wenn das Projekt abgeschlossen oder der Datensatz an ein Archiv übermittelt wird, sondern **alle Phasen des Forschungsdatenzyklus umfasst**, damit wichtige Abläufe und Ergebnisse zeitnah festgehalten werden und Richtlinien, bspw. Anonymisierungsstrategien, nicht nachträglich implementiert werden müssen. Aus diesem Grund sollte bereits in der Planungsphase vor der Datenerzeugung festgelegt werden, welche Arbeitsschritte und Aufgaben in den verschiedenen Phasen des Lebenszyklus bestehen. Abbildung 1 veranschaulicht den Forschungsdatenzyklus und erläutert, mit welchen Aspekten des FDM sich Forschende in jeder Phase konfrontiert sehen.

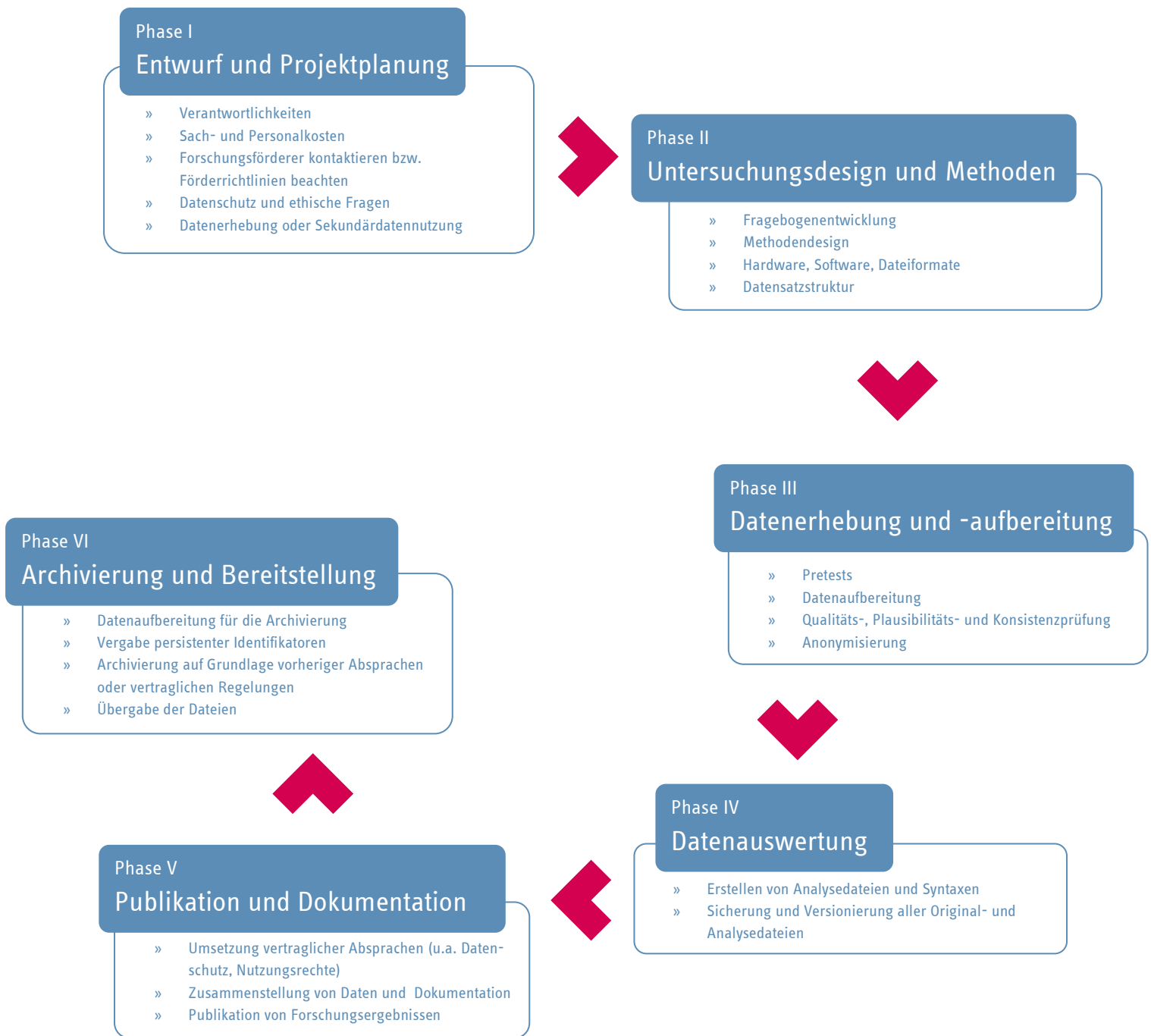


Abbildung 1: Phasen des Forschungszyklus und Themen für das Datenmanagement³

Viele Bestandteile des Datenmanagements können für jedes Forschungsprojekt entlang des in Abb. 1 veranschaulichten Forschungszyklus definiert werden. Allerdings ist die jeweilige Umsetzung der zu implementierenden Strategien und Richtlinien projekt- und disziplinabhängig, da sich Projekte in Design- und Methodenfragen sowie hinsichtlich der generierten Datentypen unterscheiden können. Beispielsweise entfallen für Studien mit rein sekundäranalytischen Auswertungen (Teile der) Untersuchungsdesign- und Datenerhebungsphase. Auch bei qualitativen Studien sind andere Schwerpunkte zu setzen. Aus diesem Grund muss für jedes Projekt individuell eine geeignete Datenmanagementstrategie erarbeitet werden.

³ Darstellung in Anlehnung an Jensen (2012: 12).

Datenmanagementplan

Ein Dokument, welches bei der Implementierung der vereinbarten Maßnahmen hilfreich sein kann, ist der **Datenmanagementplan (DMP)**. In einem DMP werden die im Projekt festgelegten Strategien und Richtlinien für Erhebung, Speicherung, Überprüfung, Sicherung, Archivierung und mögliche Weitergabe von Daten während des gesamten Forschungsdatenzyklus festgehalten. Er enthält außerdem Informationen über die Rechte an den Daten, den Zugang zu sowie die Verwaltung und Organisation der Daten und gibt Aufschluss darüber, welche Infrastruktur und welche Ressourcen benötigt werden und wie die Zuständigkeiten aufgeteilt sind.

Die Implementierung eines DMP ist die Grundlage für den verantwortungsvollen Umgang mit Forschungsdaten und damit Teil guter wissenschaftlicher Praxis (CESSDA Training 2014). Das Erstellen eines DMP wird im internationalen Kontext von Förderinstitutionen bereits eingefordert und die Mittel für eine Finanzierung des Datenmanagements werden zur Verfügung gestellt. Im deutschen Forschungskontext sei exemplarisch auf die Deutsche Forschungsgemeinschaft (DFG) verwiesen, die innerhalb ihres Leitfadens für die Antragstellung auf den Nutzen des Datenmanagements verweist und um Angaben zum Umgang mit den im Projekt erzielten Forschungsdaten bittet. Die DFG ist „[...] bestrebt, durch ihre Förderung auch zur Sicherung, Aufbewahrung und Nachnutzbarkeit von Forschungsdaten beizutragen“ (DFG, 2014, S. 5).⁴ Auch Publikationsorgane sind in verstärktem Maße an einer Veröffentlichung und Zugänglichmachung von Primärdaten interessiert, damit Beiträge repliziert und überprüft werden können. Damit wird FDM zu einem Anliegen für alle Forschenden. Explizit Forschende der empirischen Bildungsforschung werden durch das Memorandum des Fachkollegiums „Erziehungswissenschaft“ der DFG (Stanat 2014) angesprochen. Ihnen wird nahegelegt, zukünftig „bereits in der Planungs- und Antragsphase Überlegungen“ dazu anzustellen, „ob ein Nachnutzungspotenzial besteht und welche Konsequenzen dies später für die Form der Bereitstellung hat“ (Stanat 2014: 4).

Es sollte berücksichtigt werden, dass Datenmanagementpläne einen dynamischen Forschungsvorgang beschreiben und sich höchstwahrscheinlich während der Projektlaufzeit ändern werden. Sie sind „lebende Dokumente“, die im Laufe der Zeit um im Projektkontext entwickelte Aspekte wie Policies oder Guidelines erweitert werden (vgl. Jones 2011). Datenmanagementpläne sollten als integraler Teil von Forschungsanträgen und nicht als Nachsatz betrachtet werden. Bei der Erstellung eines Plans kann es von bedeutendem Vorteil sein, die Expertise Anderer zurate zu ziehen oder Tools zu verwenden, welche bereits in der Scientific Community entwickelt wurden. Eine Auflistung von Leitfäden, Guidelines und Tools zum Datenmanagement und zu Datenmanagementplänen können beispielsweise unter folgenden Links eingesehen werden:

1. Jones, Sarah, 2011: How to Develop a Data Management and Sharing Plan. DCC How-to Guides. Edinburgh: Digital Curation Centre. Verfügbar unter: www.dcc.ac.uk/resources/how-guides, abgerufen am 25.09.2014.
2. Guidelines für die Erstellung eines Datenmanagementplans des Inter-university Consortium for Political and Social Research (ICPSR). Verfügbar unter: www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/index.html, abgerufen am: 09.01.2015.
3. Datenmanagementplan Online-Tool des Digital Curation Centers (DCC): dmponline.dcc.ac.uk, abgerufen am: 09.01.2015.
4. Datenmanagementplan Online-Tool von DataOne: www.dataone.org/data-management-planning, abgerufen am: 09.01.2015.

⁴ DFG-Leitfaden für die Antragstellung, Vordruck 54.01 – 06/14, Verfügbar unter: www.dfg.de/formulare/54_01/54_01_de.pdf, abgerufen am 24.09.2014.

Literatur:

1. BDSG, Bundesdatenschutzgesetz, 1990. Stand: Neugefasst durch Bek. v. 14.1.2003 I 66; zuletzt geändert durch Art. 1 G v. 14.8.2009 I 2814: www.gesetze-im-internet.de/bdsg_1990/, abgerufen am 26.11.2014.
2. CESSDA Training-Webseite: www.gesis.org/en/archive-and-data-management-training-and-information-center/training-center-home, abgerufen am 08.05.2014.
3. DFG: Leitfaden für die Antragstellung, Vordruck 54.01 – 06/14. Verfügbar unter: www.dfg.de/formulare/54_01/54_01_de.pdf, abgerufen am 24.09.2014.
4. ICPSR: Elements of a Data Management Plan. Verfügbar unter: www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/elements.html, abgerufen am 25.09.2014.
5. Jensen, U., 2012: Leitlinien zum Management von Forschungsdaten: Sozialwissenschaftliche Umfragedaten. GESIS Technical Reports, 2012/07. Verfügbar unter: www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-07.pdf, abgerufen am: 25.09.2014.
6. Jones, S., 2011: How to Develop a Data Management and Sharing Plan. DCC How-to Guides. Edinburgh: Digital Curation Centre. Verfügbar unter: www.dcc.ac.uk/resources/how-guides, abgerufen am 25.09.2014.
7. Ludwig, J. & Enke, H. (Hrsg.), 2013: Leitfaden zum Forschungsdaten-Management. Verlag Werner Hülsbusch, Glückstadt, 2013. Verfügbar unter: www.wissgrid.de/publikationen/Leitfaden_Data-Management-WissGrid.pdf, abgerufen am: 25.09.2014.
8. Meyermann, A. & Porzelt, M., 2014: Hinweise zur Anonymisierung von qualitativen Daten. forschungsdaten bildung informiert, Nr. 1. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung. Verfügbar unter: www.forschungsdaten-bildung.de/forschungsdaten-bildung-informiert, abgerufen am 30.03.2015
9. Stanat, P., 2014: Bereitstellung und Nutzung quantitativer Forschungsdaten in der Bildungsforschung: Memorandum des Fachkollegiums „Erziehungswissenschaft“ der DFG. Verfügbar unter: www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/forschungsdaten_memorandum_fk_109.pdf. abgerufen am: 10.01.2015.

2 Modul 2: Metadaten und Dokumentation

Metadaten sind strukturierte oder unstrukturierte Informationen über Forschungsdaten und deren Management, die für jede Phase des Lebenszyklus empirischer Forschung zur Verfügung stehen. Von strukturierten Metadaten spricht man, wenn die Informationen auf standardisierten Vorgaben und Standards basieren (vgl. als weltweit genutzter Metadaten-Standard für die Sozialwissenschaften: www.ddialliance.org). Auf Variablen- und Fragenebene als Variablen- und Wertelabels, Fragebögen, Variablenlisten und Skalenhandbüchern dokumentieren Metadaten die inhaltliche Bedeutung der Daten und die Definition der Variablen (Jensen 2012: 62). Sie zeigen beispielsweise die Korrespondenz von Variablen und Fragen des Fragebogens, die numerische Kodierung der Fragekategorien, die inhaltliche Bedeutung und Herkunft einer Variable und ihrer Ausprägung. Metadaten auf Studienebene beschreiben Forschungsfragen, Projektbeteiligte, Studiendesigns so-

wie die eingesetzten Erhebungs- und Auswertungsinstrumente. Metadaten auf Studienebene sind insbesondere Methodenhandbücher und Projektberichte.

Metadaten haben drei zu unterscheidende Zwecke: Sie dienen der Interpretierbarkeit, der Wiederauffindbarkeit und der Verwaltung von Forschungsdaten. Nur in Verbindung mit diesen Metadaten sind Daten (1) inhaltlich verständlich und somit der Nachnutzung zugänglich. Metadaten bieten darüber hinaus die Möglichkeit, (2) Daten überhaupt erst zu finden. Durch die Angabe von Informationen wie Themen- und Fragestellung, Primärforscher/-innen und Erhebungszeitraum können Suchportale Wissenschaftler/-innen bei Interesse auf die entsprechenden Daten verweisen. Metadaten erleichtern überdies (3) die Sicherung, Verwaltung und Pflege von Forschungsdaten, indem sie eine einfachere und weniger fehleranfällige Verwaltung und Organisation ermöglichen. Sie sind damit unerlässlich für ein gutes Forschungsdatenmanagement.

Metadatenstandards sind dazu da, gemeinsam geteilte Strukturen und Spezifikationen zu definieren. Über diese lässt sich der Austausch von Metadaten über Institute und Forschungseinrichtungen hinweg einrichten. Der internationale Standard für die Beschreibung von Daten aus den Sozial-, Verhaltens- und Wirtschaftswissenschaften ist die Data Documentation Initiative (DDI), deren Ziel es ist, Metadatenstandards für alle Phasen des Forschungsdatenzyklus zu etablieren.

Idealerweise werden Entscheidungen zur Aufzeichnung von Metadaten sowie Dokumentation am Anfang der Forschung durch einen Datenmanagementplan festgehalten. Auf diese Weise sind viele der Metadaten bereits frühzeitig schriftlich fixiert. Viele Informationen, die für die Verwaltung und Nachnutzung der Forschungsergebnisse wichtig sind, entstehen jedoch erst später. Daher muss ebenfalls in allen weiteren Phasen der Forschung darauf geachtet werden, dass wichtige Informationen notiert und an zentraler Stelle als Metadaten festgehalten werden, damit diese nicht verloren gehen oder in Vergessenheit geraten. Dies kann in verschiedenen Formen geschehen, beispielsweise durch das Erstellen von Methodenberichten und Codebüchern, Scans der Fragebögen, Festhalten von Feldnotizen, dem Sichern und Kommentieren von Syntaxen und Entscheidungen zum Datenhandling.

Eine besondere Art von Metadaten sind **persistente Identifikatoren**. Was bei Büchern und Zeitschriften schon seit geraumer Zeit üblich ist, die eindeutige Kennzeichnung, ist auch in den Bereich der digitalen Objekte eingezogen. Wissenschaftler/-innen können eindeutige, persistente Identifikatoren für Ihre Daten, Datendokumentationen und andere digitale Objekte nutzen. Auf diese Weise vereinfachen Forscher/-innen die Referenzierbarkeit, Onlineverfügbarkeit und den Zugang zu ihren Forschungsdaten. Außerdem wird die Zitation von Forschungsdaten vereinheitlicht, sodass Datenerzeuger bei Nachnutzung ihrer Daten durch Dritte eine Anerkennung ihrer Arbeit erhalten.

Literatur:

1. Jensen, U., 2012: Leitlinien zum Management von Forschungsdaten von GESIS, Kapitel 4. Verfügbar unter: www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-07.pdf, abgerufen am: 09.01.2015
2. Gregory, A., Heus, P. & J. Ryssevick, 2009: Metadata. Working Paper No. 57. Berlin. Verfügbar unter: www.ratswd.de/download/workingpapers2009/57_09.pdf, abgerufen am: 31.11.2014
3. Day, M., 2005, DCC | Digital Curation Manual. Instalment on "Metadata". Verfügbar unter: www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/metadata/metadata.pdf, abgerufen am: 09.01.2015

4. International DOI Foundation (IDF): www.doi.org/index.html, abgerufen am: 09.01.2015
5. Data Documentation Initiative, DDI Alliance: www.ddialliance.org, abgerufen am: 15.01.2015

3 Modul 3: Ethik, Datenschutz und Anonymisierung

Empirische Wissenschaften untersuchen oft datenschutzrechtlich kritische Bereiche. Sie berühren dabei vertrauliche, persönliche und geschützte Informationen von Menschen. Insbesondere personenbezogene Daten unterliegen dem Datenschutz, da diese eine Identifizierung der Befragten ermöglichen. Auch Untersuchungen von Minderjährigen, Schülern und Schülerinnen, wie sie in der empirischen Bildungsforschung häufig vorkommen, sind mit besonderer Umsicht zu behandeln.

Für jede ethisch vertretbare Forschung ist es daher essentiell, eine Strategie zum Schutz der Identität der Studienteilnehmer/-innen zu entwerfen und umzusetzen. Um Ihren Mitgliedern Handlungsanleitungen und Empfehlungen für ethische Forschung anzubieten, haben verschiedene Fachgesellschaften wie bspw. die Deutsche Gesellschaft für Soziologie (DGS) und die Deutsche Gesellschaft für Erziehungswissenschaften (DGfE) Ethik-Kodizes erstellt.⁵ An diesen können sich Forschende orientieren.

Eine Strategie für ethisch vertretbare Forschung besteht grundsätzlich aus drei Säulen: (1) informierte Einwilligung, (2) Anonymisierung und (3) Nutzungs- und Zugangsbeschränkungen. Die informierte Einwilligung (informed consent) setzt voraus, dass Studienteilnehmer/-innen in die wesentlichen Aspekte des Forschungsvorhabens eingeweiht wurden, ihre Einwilligung zweckgebunden erfolgte und ihnen bekannt gemacht wurde, dass sie ihre – freiwillig erfolgte – Einwilligung jederzeit widerrufen können. Außerdem muss eine Einwilligung in der Regel schriftlich festgehalten werden (Jensen 2012: 14, vgl. www.forschungsdaten-bildung.de/einwilligung und §4 BDSG).

Ein weiterer Bestandteil einer solchen Strategie ist die Anonymisierung. „Die Anonymisierung von Forschungsdaten ist in all den Fällen notwendig, in denen keine Einwilligung der Probanden zur Verwendung ihrer persönlichen Daten vorliegt oder eine derartige Verwendung nicht durch eine entsprechende Rechtsvorschrift ermöglicht wird“.⁶ **Anonymisierung im Sinne des Bundesdatenschutzgesetzes bedeutet, dass Daten so verändert werden, dass ein Personenbezug nicht oder nur mit unverhältnismäßig hohem Aufwand (an Zeit, Kosten und Arbeitskraft) wiederhergestellt werden kann (vgl. §3 Abs.6 BDSG).** Anonymisierung ist aus einer Vielzahl ethischer Gründe erforderlich. Diese reichen dabei vom Schutz der Teilnehmer/-innen vor Identifizierung, insbesondere wenn sensible Themen erfragt werden, über die Verhinderung missbräuchlicher kommerzieller Verwendung bis hin zur Einhaltung des Datenschutzrechts. Durch (mindestens faktische) Anonymisierung entfallen strenge rechtliche Anforderungen zum Umgang mit personenbezogenen Daten, da diese nicht länger dem Anwendungsbereich der Datenschutzgesetze unterliegen (Kretzer 2003: 2; Arbeitsgruppe Datenschutz und qualitative Sozialforschung 2014: 7). Das heißt bspw., dass anonymisierte Forschungsdaten unbefristet und ohne gesonderte Einwilligung der Studienteilnehmer in die Archivierung archiviert und nachgenutzt werden dürfen (Jensen 2012: 67). Es können drei Formen der Anonymisierung unterschieden werden: formale, faktische und absolute Anonymisierung. Formale Anonymisierung bedeutet die getrennte Aufbewahrung direkter Identifizierungsmerkmale wie Name und Adresse von anderen Angaben. Eventuelle Zuordnungslisten müssen ebenfalls getrennt und sicher aufbewahrt werden und nur Berechtigten zur Verfü-

⁵ Ethik-Kodex der DGS: <http://www.soziologie.de/de/die-dgs/ethik-kommission/ethik-kodex.html> und der DGfE: http://www.dgfe.de/fileadmin/OrdnerRedakteure/Service/Satzung/Ethikkodex_2010.pdf, abgerufen am 09.04.2015.

⁶ <http://www.forschungsdaten-bildung.de/anonymisierung>, abgerufen am 09.04.2015.

gung stehen (Arbeitsgruppe Datenschutz und qualitative Sozialforschung 2014: 13). Die faktische Anonymisierung geht darüber hinaus. Hierzu werden schützenswerte Daten in ihrem Informationsgehalt reduziert und systematisch verändert, mit dem Ziel, Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit unverhältnismäßig hohem Aufwand einer Person zuordenbar zu machen. Naturgemäß bedeutet dies einen höheren Aufwand als bei der formalen Anonymisierung (Medjedovic, Witzel 2010: 75). Die absolute Anonymisierung schlussendlich erzwingt für besonders sensible Daten das Löschen von Informationen, wie beispielsweise kritischer Interviewpassagen, oder das Ersetzen sensibler Informationen durch Platzhalter. Faktische Anonymisierung ist aus Sicht des BDSG ausreichend.

Die Anonymisierung personenbezogener Daten sollte geschehen, sobald es der Forschungszweck zulässt. Daneben gibt es zusätzliche Maßnahmen, um den Schutz der Studienteilnehmer/-innen durch Einschränkungen des Zugangs und/oder der Nachnutzungsmöglichkeiten der Daten zu gewährleisten. Dazu zählen unter anderem die Empfehlung des Unterlassens einer Veröffentlichung kompletter qualitativer Interviews (Kretzer 2013: 7), ein kontrollierter Zugang zu anonymisierten Daten, vertragliche Regelungen bei Datenbereitstellung durch Wissenschaftler/-innen an Forschungsdatenzentren und vertragliche Einschränkungen der Nutzbarkeit der Daten für Nachnutzer/-innen, beispielsweise ein Zweckbindungsgebot, Reidentifizierungsverbot und Verbot der Weitergabe der Daten an Dritte (Kretzer 2013: 7). Ein solches rights management framework ist besonders für schwer zu anonymisierende Daten, beispielsweise Videodateien, die nicht oder nur unter hohen Kosten und Verlusten an Informationsgehalt anonymisiert werden können und qualitative Längsschnittstudien erforderlich (Bishop 2009: 262).

Der Anonymisierungsvorgang sollte bereits vor der Datenerhebung sorgfältig geplant sein und ist Teil des Datenmanagementplans. Auf diese Weise wird sichergestellt, dass im Idealfall informierte Einwilligungen der Studienteilnehmer/-innen vorliegen, die ebenfalls die Archivierung und Datenweitergabe ermöglichen. Außerdem ist der Anonymisierungsprozess auf diese Weise weniger ressourcenintensiv als bei späterer Anonymisierung.

Die Anonymisierungsstrategie hängt entscheidend von dem Datentyp ab.⁷ Qualitative Informationen, insbesondere audiovisuelles Material und offene Antwortmöglichkeiten in Befragungen, sind in der Regel wesentlich schwieriger und aufwändiger oder gar nicht zu anonymisieren. Unabhängig vom Datentyp sollten alle direkten Personenidentifikatoren durch Pseudonyme ersetzt werden. Darüber hinaus können Antwortkategorien vergrößert werden, insbesondere wenn die empirische Häufigkeit einzelner Ausprägungen sehr selten ist, oder sogar vereinzelt Variablen entfernt werden. Audiovisuelles Material kann bspw. durch digitale Manipulation wie Stimmverzerrung und Bildbearbeitung von kritischen Informationen bereinigt werden. Alle Änderungen sind in jedem Fall festzuhalten, beispielsweise als Syntax oder in separaten Änderungsprotokollen.

Die Einhaltung des Datenschutzes muss sowohl in der Erhebungs-, der Archivierungs- als auch der Datenweitergabephase eine wichtige Rolle spielen und betrifft viele Entscheidungen im Verlauf der Forschung, wie beispielsweise die Auswahl der Stichprobe und den Erstkontakt zu den Befragten, die Datenaufbereitung, die

⁷ Eine Übersicht über Anonymisierungsstrategien von v. a. quantitativen Daten bieten Kinder-Kurlanda, K. & Watteler, O., 2015: Hinweise zum Datenschutz. Rechtlicher Rahmen und Maßnahmen zur datenschutzgerechten Archivierung sozialwissenschaftlicher Forschungsdaten. Verfügbar unter: www.gesis.org/publikationen/gesis-papers, abgerufen am 30.03.2015. Für qualitative Daten finden sich Hinweise bei Meyermann, A. & Porzelt, M., 2014: Hinweise zur Anonymisierung von qualitativen Daten. *forschungsdaten bildung informiert*, Nr. 1. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung. Verfügbar unter: www.forschungsdaten-bildung.de/forschungsdaten-bildung-informiert, abgerufen am 30.03.2015.

Weitergabe der Daten an Archive und Forschungsdatenzentren zur Archivierung und Nachnutzung sowie die Auswahl geeigneter Daten für Sekundäranalysen und die Fixierung von Nachnutzungsbedingungen. Es ist daher empfehlenswert, Überlegungen zum Datenschutz zu einem festen Bestandteil des Datenmanagementplans zu machen.

Literatur:

1. Arbeitsgruppe Datenschutz und qualitative Sozialforschung, 2014: Datenschutzrechtliche Anforderungen bei der Generierung und Archivierung qualitativer Interviewdaten, RatSWD Working Paper Series. Verfügbar unter: www.ratswd.de/dl/RatSWD_WP_238.pdf, abgerufen am: 15.12.2014.
2. BDSG, Bundesdatenschutzgesetz, 1990. Stand: Neugefasst durch Bek. v. 14.1.2003 I 66; zuletzt geändert durch Art. 1 G v. 14.8.2009 I 2814: www.gesetze-im-internet.de/bdsg_1990/, abgerufen am 26.11.2014.
3. Bishop, L., 2009: Ethical Sharing and Reuse of Qualitative Data. Australian Journal of Social Issues 44 (3): 255 – 272.
4. Forschungsdaten-Bildung-Webseite: www.forschungsdaten-bildung.de/einwilligung, abgerufen am: 20.02.2015.
5. Häder, M., 2009: Der Datenschutz in den Sozialwissenschaften. Anmerkungen zur Praxis sozialwissenschaftlicher Erhebungen und Datenverarbeitung in Deutschland. RatSWD – Working Paper No. 90. Verfügbar unter: www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_90.pdf, abgerufen am: 19.01.2015.
6. Jensen, U., 2012. Leitlinien zum Management von Forschungsdaten. Sozialwissenschaftliche Umfragedaten. GESIS-Technical Reports 2012|07. Verfügbar unter: www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-07.pdf, abgerufen am 20.11.2014.
7. Kinder-Kurlanda, K. & Watteler, O., 2015: Hinweise zum Datenschutz. Rechtlicher Rahmen und Maßnahmen zur datenschutzgerechten Archivierung sozialwissenschaftlicher Forschungsdaten. Verfügbar unter: www.gesis.org/publikationen/gesis-papers, abgerufen am 30.03.2015.
8. Kretzer, S., 2013: Arbeitspapier zur Konzeptentwicklung der Anonymisierung/Pseudonymisierung in Qualiservice. Verfügbar unter: http://www.qualiservice.org/fileadmin/templates/qualiservice/Anonymisierungskonzept_Arbeitspapier.pdf, abgerufen am: 30.03.2015.
9. Medjedovic, I. & Witzel, A., 2010: Wiederverwendung qualitativer Daten – Archivierung und Sekundärnutzung qualitativer Interviewskripte. Wiesbaden: VS Verlag für Sozialwissenschaften.
10. Meyermann, A. & Porzelt, M., 2014: Hinweise zur Anonymisierung von qualitativen Daten. forschungsdaten bildung informiert, Nr. 1. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung. Verfügbar unter: www.forschungsdaten-bildung.de/forschungsdaten-bildung-informiert, abgerufen am 30.03.2015
11. Metschke, R. & Wellbrock, R., 2002. Datenschutz in Wissenschaft und Forschung. Materialien zum Datenschutz Nr. 28., 3. Aufl. Berlin, 2002: www.datenschutz-berlin.de/attachments/47/Materialien28.pdf?1166527077, abgerufen am 26.11.2014.

4 Modul 4: Datenadministration und -speicherung

Dateienorganisation

Eine systematische Dateienorganisation, die u.a. durch konsistente und transparente Ordnerstrukturen und Dateibenennungen implementiert wird, schafft eine Reihe von Vorteilen für Forschungseinrichtungen und -projekte. Durch das schnellere Auffinden von Ordnern und Dateien, vor allem für andere Nutzer/-innen in Forschungsk Kooperationen, ergeben sich Zeitersparnisse. Systematische Dateienorganisation trägt außerdem dazu bei, sicherzustellen, dass Dateien jederzeit eindeutig identifizierbar sind und verhindert damit gleichzeitig das Verwechseln und versehentliche Löschen.

Systematische Dateienorganisation erfolgt vor allem auf zwei Ebenen: durch *Ordnerstrukturen* und *Dateibenennungen*. Die Ordnerstruktur sollte konsistent, systematisch organisiert und unmittelbar verständlich sein. Um dies zu erreichen, empfiehlt sich

1. eine klare und selbsterklärende Verzeichnisstruktur zu nutzen,
2. in den hierarchisch höchsten Ordnername den Projekttitel, ggf. mit Datum (z. B. Projektlaufzeit), und einen eindeutigen Projektidentifikator zu integrieren,
3. untergeordnete Ordner dann beispielsweise nach Datum, Namen der beteiligten Wissenschaftler/-innen oder auch inhaltlich (z.B. entlang der Phasen im Forschungsdatenlebenszyklus) zu benennen und
4. innerhalb des Projekts Konsistenz zu wahren.

Die Dateibenennung sollte bereits Informationen zum Inhalt und Status einer Datei bieten, bspw.:

1. ID oder Studiennummer: zur Zuordnung zur Studie
2. Kürzel für Daten- bzw. Materialtyp: Interview, Video, Fragebogen etc.
3. Laufende Nummer von Datentypen: 001 etc.
4. Seriennummer: verschiedene Dateien je Datentyp (z. B. Ein Video besteht aus mehreren Dateien): a, b, c, etc.
5. Versionsnummer: falls Änderungen durchgeführt werden und dokumentiert werden sollen, zum Beispiel durch Anonymisierung

Daneben spielen außerdem Versionierungen, das Kennlichmachen von Änderungen und Entwicklungsstadien von Dateien sowie gegebenenfalls das Aufzeichnen der Änderungen, und die Verwendung empfohlener Dateiformate eine Rolle. Forschungsinstitute sollten sich um einheitliche, in der Fachdisziplin als Standard für den entsprechenden Datentyp geltende Formate bemühen. Diese sollten möglichst offen oder zumindest portierbar sein.

Datenspeicherung und Sicherheit

Forschungsdaten müssen sicher aufbewahrt und regelmäßig gesichert werden, um sie vor unautorisiertem Zugriff, Nutzung, Veränderung, Enthüllung und Zerstörung zu schützen. Vor allem sensible, personenbezogene oder -beziehbare, Daten bedürfen besonderer Schutzmaßnahmen. Beispielsweise sollten personenbezogene Daten frühestmöglich im Forschungsprozess von anderen Daten getrennt aufbewahrt werden, damit die Möglichkeit der Reidentifizierung von Befragten verringert wird. Der Workshop wird die Forscher/-innen

in die Lage versetzen, ihre Daten so zu schützen, dass dieses und andere Risiken möglichst ausgeschaltet oder zumindest reduziert werden.

Der Zugriff auf Forschungsdaten sollte in angemessener Weise beschränkt werden. Dafür steht eine Reihe technischer Maßnahmen zur Verfügung, die von Passwortschutz über Verschlüsselung bis hin zum Einsatz von Anti-Virus-Software und Firewalls reichen. Wichtig ist in diesem Bereich die Einhaltung von Standards, denn ein Passwortschutz ist nur so gut wie das gewählte Passwort und eine Verschlüsselung hilft nur, wenn der dazugehörige Schlüssel nicht als Post-it-Note am Monitor klebt oder per Email an die Kolleg/-innen verschickt wird.

Um ungewollten Datenverlust zu vermeiden, aber auch versehentliche Veränderungen der Daten rückgängig machen zu können, empfiehlt es sich, regelmäßig Backups zu erstellen. Denn: (1) alle digitalen Medien können jederzeit ausfallen, sei es durch Katastrophen wie Brände oder geborstene Wasserleitungen, oder durch begrenzte Lebensdauern der Medien und (2) digitale Formate ändern sich mit der Zeit und physikalische Speichermedien werden irgendwann obsolet. Im schlimmsten Fall sind sie dann nicht mehr les- und nutzbar. Dazu kommen (3) Anwenderfehler wie versehentliches Löschen der falschen Datei. Systematisches Forschungsdatenmanagement trägt dazu bei, schon im Vorfeld Forschungsdaten – dies betrifft auch Metadaten, Methodenberichte, Fragebögen etc. – so zu organisieren und zu sichern, dass sie nicht verloren gehen.

Die Datensicherung sollte in Form einer Backup-Strategie geregelt werden, die die folgenden Fragen klärt:

1. Was soll gesichert werden?
2. Wo können Daten gesichert werden?
3. Wie oft sollten Backups durchgeführt werden?
4. Welche Art von Backups wird durchgeführt (inkrementelle, vollständige)?
5. Wie lange werden die Backups aufbewahrt?

Unter Umständen sollen Daten aber explizit so gelöscht werden, dass sie nicht wiederherstellbar sind, beispielsweise bei der Entsorgung alter Hardware. Dies ist im Regelfall unter Windows-Systemen durch das einfache Löschen (und Entleeren des „Papierkorbes“) nicht geleistet. Hier muss gezielt spezielle Software eingesetzt werden, die Sektoren von Festplatten mehrfach überschreiben. Nur so oder durch die physikalische Vernichtung der Datenträger kann verhindert werden, dass Unbefugte den Zugriff auf Daten erhalten.

Literatur:

1. Forschungsdaten-Bildung-Webseite: www.forschungsdaten-bildung.de/tipps-material, abgerufen am: 10.01.2015.
2. Forschungsdaten-Bildung-Webseite: www.forschungsdaten-bildung.de/einwilligung, abgerufen am: 31.03.2015
3. Jensen, U., 2012: Leitlinien zum Management von Forschungsdaten von GESIS, Kapitel 3. Verfügbar unter: www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-07.pdf, abgerufen am: 09.01.2015
4. UK Data Archive-Webseite, Bereich Datenformate: www.data-archive.ac.uk/create-manage/format, abgerufen am: 10.01.2015

UK Data Archive-Webseite, Bereich Datenspeicherung: www.data-archive.ac.uk/create-manage/storage, abgerufen am: 10.01.2015

5 Modul 5: Urheberrecht und Lizenzierung

Werden Forschungsdaten veröffentlicht oder nachgenutzt, müssen Wissenschaftler/-innen sich mit den Themen geistiges Eigentum und Lizenzierung auseinandersetzen. In Deutschland wird der rechtliche Schutz des geistigen Eigentums aus den Bereichen der Wissenschaft durch das Urheberrecht gewährleistet. Gesetzliche Grundlage ist das „Gesetz über Urheberrecht und verwandte Schutzrechte“ vom 09.09.1965 (das sogenannte Urheberrechtsgesetz, kurz: UrhG) (Pierson et. al. 2007: 8). Schutz nach UrhG genießen Urheber von Werken der Literatur, Wissenschaft und Kunst (§1 UrhG). **Der rechtliche Status von Forschungsdaten ist häufig unklar**, unter bestimmten Umständen zählen aber auch Datenbankwerke als schützenswert.⁸ Oft ist schwer erkennbar, welche Werke geschützt sind und welche Werke frei verwendet werden dürfen. Als Beispiele strittiger Fälle seien hier Daten, die sich aus dem Internet herunterladen lassen, verwaiste Werke, das heißt Werke, deren Urheber unbekannt sind, sowie Datenbanken, die auf bekannten Fakten basieren, genannt. **Die Publikation von Forschungsdaten ohne Klarstellung der Bedingungen, unter denen diese genutzt werden können, ist daher in jedem Fall kontraproduktiv (Ball 2014: 2).**

Lizenzen dienen der Nachnutzung von Daten und der Absicherung der Datenerzeuger, indem sie Nachnutzer/-innen bestimmte Rechte einräumen. Lizenzierungen von Forschungsdaten sind Sekundärnutzer/-innen dadurch behilflich, dass frühzeitig im Forschungsprozess mögliche Restriktionen der Datennutzung erkannt werden, die unter Umständen eine Datennutzung oder eine Veröffentlichung der Forschungsergebnisse einschränken. Ein Beispiel für Lizenzen sind die Nachnutzungsordnungen von archivierenden Organisationen. Diese schränken bspw. den Zugriff auf Daten und deren Nachnutzungsmöglichkeiten durch die Nutzer ein. Ein weiteres Beispiel betrifft Test- und Kompetenzmessungsskalen und -items, die oftmals geschützt sind und deren Verwendung kostenpflichtig ist. Aber auch Datenerzeuger profitieren von der Lizenzierung. Lizenzen geben Datenerzeugern u. U. das Recht, Bedingungen zu definieren, unter denen die Daten genutzt werden können. Diese können dabei weniger restriktiv sein als die grundlegenden Bestimmungen des Urheberrechts.

Wissenschaftler/-innen stehen unter Umständen durch ihre Forschungsinstitute und sonstige Arbeitgeber vorgefasste Lizenzen zur Verfügung. Des Weiteren existiert eine Vielzahl an Standardlizenzen, auf die zurückgegriffen werden kann. Bekannte und bewährte Beispiele hierfür sind Creative Commons- und Open Data Commons-Lizenzen. Creative-Commons ist eine Non-Profit-Organisation, die verschiedene vorgefertigte Lizenzen zur Verfügung stellt, damit Urheber ihre rechtlich geschützten Werke zur Verwendung durch Andere freigeben können. Alle vorgefertigten Lizenzen sehen als Mindestvoraussetzung der Nutzung des geschützten Werks die Namensnennung des Urhebers vor. Des Weiteren können drei zusätzliche Bedingungen festgehalten werden: keine Zulassung kommerzieller Nutzung, keine Zulassung eigener Bearbeitung und die Einschränkung, dass eigene abgeleitete Werke unter denselben Bedingungen lizenziert werden müssen wie das Originalwerk. Aus der Kombination dieser Bedingungen ergeben sich sechs Standardlizenzen (vgl. creativecommons.org/licenses). Bei Open Data Commons handelt es sich um ein weiteres Projekt zur Erstellung und Pflege von Lizenzen. Die Lizenzen ähneln denen der Creative-Commons Organisation, sind aber speziell auf die Besonderheiten von Datenbankwerken ausgelegt (Ball 2014: 8).

Open Data bezeichnet den freien Zugriff sowie die freie Nutzung, Veränderung und Weitergabe von Daten ohne Einschränkungen (vgl. opendefinition.org). Open Data ermöglichen den Einsatz neuer Technologien im Umgang mit Daten (u.a. bezüglich Kombinierbarkeit, Analyse, Integration), wohingegen die Einsatzmöglichkeiten von Daten unter restriktiven Lizenzen eingeschränkt sind. Umso freier die Lizenzbestimmungen sind,

⁸ vgl. www.ip.mpg.de/files/pdf3/DSRI-Tagungsband-HA13-Hartmann.pdf, S. 510ff.

desto größer sind der Umfang von einsetzbaren Technologien und der Kreis potentieller Sekundärnutzer/-innen.

Literatur:

1. Ball, A., 2014: How to License Research Data. Digital Curation Centre. Verfügbar unter: www.dcc.ac.uk/resources/how-guides/license-research-data, abgerufen am: 13.01.2015.
2. Creative-Commons-Webseite: creativecommons.org, abgerufen am: 13.01.2015.
3. Open Data Commons-Webseite: opendatacommons.org, abgerufen am: 13.01.2015.
4. Opendefinition-Webseite: opendefinition.org, abgerufen am: 13.01.2015.
5. Pierson, M., Ahrens, T. & K. Fischer, 2007: Recht des geistigen Eigentums. Patente, Marken, Urheberrecht, Design. Verlag Franz Vahlen: München.
6. Taeger, J. (Hrsg.), 2013: Law as a Service (LaaS). Recht im Internet- und Cloud-Zeitalter. Oldenburger Verlag für Wirtschaft, Informatik und Recht. Verfügbar unter: www.ip.mpg.de/files/pdf3/DSRI-Tagungsband-HA13-Hartmann.pdf, abgerufen am: 13.04.2015.

6 Modul 6: Datenweitergabe und Nachnutzung

Auch nach der Laufzeit eines Forschungsprojekts, für das sie erhoben und verwendet wurden, bleiben Forschungsdaten eine wertvolle wissenschaftliche Ressource. Die Weitergabe der Daten zur Sekundärnutzung (Data Sharing) ist ein Nutzungszweck, der besonders unspezifische und universelle Anforderungen an das Datenmanagement stellt (Meyermann 2012: 7). Die Weitergabe der eigenen Forschungsdaten ermöglicht zukünftigen Projekten neue Untersuchungswege zu bestreiten, ohne die Arbeit und Mühe, welche mit einer Datenerhebung verbunden sind, wiederholen zu müssen. Es können sich zum Teil sogar neue Verwendungszwecke für die Daten ergeben, die vom/von der Primärforscher/-in ursprünglich nicht bedacht wurden. Hinzu kommt der Aspekt, dass Daten aus Forschungsprojekten, welche durch öffentliche Mittel finanziert wurden, im Anschluss zugänglich gemacht werden sollten. Gut dokumentierte Informationen ersparen Primärforscher/-innen und Nachnutzer/-innen zeitlichen Aufwand, da sie zum einen besser auffindbar sind und zum anderen auch im Hinblick auf mögliche Re-Analysen der eigenen Daten zu einem späteren Zeitpunkt nachvollziehbar bleiben. Sollte für das eigene Forschungsprojekt eine Datenarchivierung vorgesehen sein, so ist es von Vorteil, wenn diesbezügliche Pläne bereits in der Projektplanung und bei der Erstellung des Projektantrags entwickelt und in das weitere Vorgehen integriert werden.

Warum aber sollte man es in Betracht ziehen, die eigenen Daten, die aufwändig in einem Forschungsprojekt generiert wurden, Anderen zur Verfügung zu stellen? Gibt es legitime Gründe, die gegen die Weitergabe von Daten sprechen?

Gründe für die Datenweitergabe

Die Datenweitergabe bringt eine Vielzahl von Vorteilen mit sich, welche nicht nur die Nachnutzer/-innen der Daten und die Scientific Community, sondern auch die Datengeber selbst betreffen. Indem Forscher/-innen ihre Daten Anderen zur Verfügung stellen, können Ergebnisse repliziert, validiert und ggf. korrigiert werden. Dadurch unterstützen Datengeber den wissenschaftlichen Fortschritt und bürgen durch Transparenz für ihre wissenschaftliche Arbeit. Forschungsdatenzentren vergeben persistente Identifier für Datensätze. Dies ermöglicht die wissenschaftliche Würdigung der Datenerhebung durch deren systematische Zitation und steigert dadurch die eigene wissenschaftliche Reputation. Durch die Übergabe der Daten an Archive oder Repositori-

en wird zudem sichergestellt, dass Forschungsdaten langfristig gesichert und unter Einbehaltung hoher Sicherheits- und Qualitätsstandards zur Verfügung stehen. Die Daten werden in einer solchen Form dokumentiert, dass sie auch dann noch nachzuvollziehen und verständlich sind, wenn sie nicht mehr in regelmäßigem oder häufigem Gebrauch sind und sogar Jahre seit Projektende vergangen sind. Dies ist durchaus auch im Sinne der Primärforscher/-innen, die u. U. zukünftig wieder mit diesen Daten arbeiten wollen. Gerade auch Studierende und Promovierende profitieren darüber hinaus von der Verfügbarkeit „echter“ Daten, da sich hiermit wissenschaftliche Ergebnisse und Analysen unmittelbar replizieren und erlernen lassen. Nicht zuletzt wird auch in Deutschland die Forschungsförderung durch Geldgeber und Förderinstitutionen in zunehmendem Maße an Bedingungen geknüpft, welche eine Darlegung des Umgangs mit den Daten nach Beendigung des Projekts vorsehen.

Gründe gegen die Datenweitergabe

Wenn Forschungsdaten sensible und personenbezogene Informationen enthalten, kann die Veröffentlichung solcher Daten u. U. Datenschutzbestimmungen, Ethikkodizes und Vereinbarungen mit den Befragten oder mit Kooperationspartnern verletzen. Bei Forschung an Schulen und anderen Bildungseinrichtungen sind zusätzlich Länderbestimmungen sowie Absprachen mit Kultusministerien und weiteren Kooperationspartnern (Eltern, Lehrern, Schulen, etc.) zu beachten. Personenbezogene Daten müssen vor der Veröffentlichung anonymisiert werden. Datenarchive beraten Forschende zu Anonymisierungsleistungen und können diese u. U. selbst übernehmen. Eine Weitergabe von anonymisierten Forschungsdaten ist datenschutzrechtlich unbedenklich. Ist die Eigentümerschaft von Forschungsdaten (oder Instrumenten, Skalen, etc.) nicht eindeutig oder liegen die entsprechenden Rechte nicht in der Hand der Person, die diese bei einem Datenarchiv einreichen möchte, können die Daten nicht veröffentlicht werden. Wenn Daten einen finanziellen Wert darstellen, dann sollte die Datenweitergabe bzw. Archivierung mit den Kooperationspartnern und Dateneigentümern besprochen werden. Ggf. ist von einer Veröffentlichung der Daten abzusehen. Auch in diesem Fall kann ein Datenarchiv Beratungsleistungen anbieten und Forscher/-innen beispielsweise bei der Frage unterstützen, ob zumindest eine Teilveröffentlichung der Forschungsdaten möglich und sinnvoll erscheint. Außerdem können mit der archivierenden Organisation Zugriffs- und Nachnutzungsbeschränkungen vereinbart werden, bspw. können für Forschungsdaten, die die Grundlage für zukünftige Publikationen darstellen sollen, Embargofristen eingeräumt werden, die eine Datenweitergabe zu einem späteren Zeitpunkt ermöglichen.

Folgende Beispiele sollen die Herausforderungen, die Data Sharing für die Datenmanagementpraxis bedeutet, verdeutlichen (Meyermann 2012: 8):

1. Die Dokumentation der Daten und der Datenerhebung muss ausreichend sein, um Nachnutzern die Interpretation zu ermöglichen.
2. Gegebenenfalls wird das Einverständnis von Auftraggebern oder Förderern benötigt, um Daten an Dritte weitergeben zu können.
3. Das Einverständnis der Probanden ist einzuholen, wenn personenbezogene oder personenbeziehbare (nicht anonymisierte) Daten weitergegeben werden sollen.
4. Für nicht deutschsprachige Nutzungen sind englischsprachige Dokumentationen wie Variablenbezeichnungen erforderlich.
5. Innerhalb verschiedener Forschungsdisziplinen werden unterschiedliche Standards bzgl. der genutzten Software und der Terminologie genutzt.
6. Die Bestimmungen des archivierenden Datenzentrums sind einzuhalten.

Data Sharing ist nur ein Spezialfall der Ausrichtung des Datenmanagements an Nutzungszwecken (ebenda). Die Archivierung und Bereitstellung der Daten unterstützen auch Nutzungszwecke der Primärforscher/-innen und deren Kollegen/-innen. Beispiele hierfür sind die erneute Nutzung der Daten durch die Primärforscher/-innen nach Ablauf einiger Jahre, die Einarbeitung in Struktur und Kontext der Forschungsdaten durch neue Projektmitglieder bei Mitarbeiterfluktuationen und die Nachnutzung in einem Anschlussprojekt.

Literatur:

1. DataCite-Webseite. Eine Liste von Datenarchiven und Repositorien, inkl. deren Ingest-Richtlinien. Verfügbar unter: www.datacite.org/repolist, abgerufen am: 10.01.2015.
2. GESIS-Webseite. Beispiel für Zugangskategorien und -beschränkungen (GESIS): www.gesis.org/en/services/data-analysis/data-archive-service/usage-regulations/#3_Access_categories, abgerufen am: 10.01.2015.
3. Forschungsdaten-Bildung-Webseite: www.forschungsdaten-bildung.de/data-sharing, abgerufen am: 10.01.2015
4. Medjedovic, I. & Witzel, A., 2010: Wiederverwendung qualitativer Daten – Archivierung und Sekundärnutzung qualitativer Interviewskripte. Wiesbaden: VS Verlag für Sozialwissenschaften.
5. Meyermann, A. & Porzelt, M., 2014: Hinweise zur Anonymisierung von qualitativen Daten. forschungsdaten bildung informiert, Nr. 1. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung. Verfügbar unter: www.forschungsdaten-bildung.de/forschungsdaten-bildung-informiert, abgerufen am 30.03.2015
6. UK Data Archive-Webseite: www.data-archive.ac.uk/create-manage/planning-for-sharing, abgerufen am: 10.01.2015