

Auszug aus dem Buch:

Uwe Jensen
Sebastian Netscher
Katrin Weller (Hrsg.)

Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten

Grundlagen und praktische Lösungen
für den Umgang mit
quantitativen Forschungsdaten

Verlag Barbara Budrich
Opladen • Berlin • Toronto 2019

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;
detaillierte bibliografische Daten sind im Internet über
<http://dnb.d-nb.de> abrufbar.

© 2019 Dieses Werk ist beim Verlag Barbara Budrich erschienen und steht unter der Creative Commons Lizenz Attribution-ShareAlike 4.0 International (CC BY-SA 4.0):

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

Diese Lizenz erlaubt die Verbreitung, Speicherung, Vervielfältigung und Bearbeitung bei Verwendung der gleichen CC-BY-SA 4.0-Lizenz und unter Angabe der UrheberInnen, Rechte, Änderungen und verwendeten Lizenz.



Dieses Buch steht im Open-Access-Bereich der Verlagsseite zum kostenlosen Download bereit (<https://doi.org/10.3224/84742233>).

Eine kostenpflichtige Druckversion (Print on Demand) kann über den Verlag bezogen werden. Die Seitenzahlen in der Druck- und Onlineversion sind identisch.

ISBN 978-3-8474-2233-4 (Paperback)

eISBN 978-3-8474-1260-1 (eBook)

DOI 10.3224/84742233

Umschlaggestaltung: Bettina Lehfeldt, Kleinmachnow – www.lehfeldtgraphic.de

Lektorat: Nadine Jenke, Potsdam

Satz: Anja Borkam, Jena – kontakt@lektorat-borkam.de

Titelbildnachweis: Foto: Florian Losch

Druck: paper & tinta, Warschau

Printed in Europe

5. Dateiorganisation in empirischen Forschungsprojekten

Jonas Recker und Evelyn Brislinger

Forschungsdaten bilden die Grundlage empirischer Forschungsprojekte in den Sozialwissenschaften und stehen damit im Fokus der Aufmerksamkeit der Wissenschaftlerinnen und Wissenschaftler. Jedoch entstehen in Projekten neben erhobenen Forschungsdaten typischerweise große Mengen an weiteren Dokumenten, die im Rahmen des Forschungsdatenmanagements Beachtung finden sollten. Hierzu gehören einerseits Dokumente wie Besprechungsprotokolle oder E-Mail-Kommunikationen, die den Projektverlauf und die Projektorganisation dokumentieren und so z.B. getroffene Entscheidungen oder die Aufgabenverteilungen und Verantwortlichkeiten nachvollziehbar machen. Andererseits fallen aber auch Materialien an, die dazu geeignet sind, Auskunft über Entstehung, Aufbereitung, Inhalt und Qualität der erhobenen oder genutzten Daten zu geben – z.B. Informationen zur Entstehung des Erhebungsinstruments, Paradata über die Erhebungssituation, Syntaxdateien und Reports über die Datenaufbereitung. Diese Informationen tragen wesentlich dazu bei, den Forschungsprozess insgesamt nachvollziehbar zu machen. Es entsteht also im Projektverlauf eine Informationssammlung, die nicht nur im aktuellen Projektkontext, sondern auch darüber hinaus von Bedeutung ist. Zum Beispiel kann sie für die Primärforschenden eine wichtige Informationsquelle für nachfolgende Erhebungswellen darstellen. Aber auch für eine Nutzung der Daten durch Dritte zu Replikationszwecken oder zur Beantwortung neuer Forschungsfragen sind Teile dieser Informationssammlung unabdingbar, da sie Transparenz über den Forschungsprozess und insbesondere die Entstehung und Aufbereitung der Daten herstellen.

Der Großteil dieser bis zum Projektende entstehenden Informationen durchläuft eine Vielzahl von Bearbeitungsschritten, die z.T. durch unterschiedliche Personen vorgenommen werden. So liegen Daten und Dokumente am Ende eines Projekts oft in mehreren Aufbereitungsstufen und Versionen vor. Ein Projektteam steht damit vor der Herausforderung, diese heterogenen Informationen so zu organisieren und bereitzustellen, dass sie für ihre jeweiligen Adressatinnen und Adressaten leicht auffindbar und zugreifbar sind. Gerade in kollaborativen Forschungsprojekten kann es schwierig sein, sicherzustellen, dass Informationen einfach identifizierbar und auffindbar sind, dass sie nachfolgende Arbeitsschritte zeitnah erreichen, dass sie nicht irrtümlich gelöscht oder überschrieben werden, aus Back-up-Routinen herausfallen oder aus technischen Gründen verloren gehen. Damit steigen mit der Komplexität des Projektworkflows und der Anzahl der interagierenden Beteiligten auch die Anforderungen an die Strukturierung der Informationen und die Organisation des Informationsflusses (vgl. Macías/Petrakos 2006: 4). Folgen diese nicht verbindlichen, im Projekt abgestimmten Leitlinien, führt dies häufig dazu, dass im Laufe der Arbeit immer wieder ad hoc Entscheidungen über die Organisation von Dateien getroffen werden, die nicht für alle Beteiligten transparent und nachvollziehbar sind und daher nicht selten zu Informationsverlusten führen.

Dies zu vermeiden, ist Ziel eines systematischen Vorgehens bei der Organisation von Informationen (in Form von Dateien) und von Informationsflüssen in einem empirischen Forschungsprojekt. Ein solches Vorgehen trägt dazu bei, die Auffindbarkeit von Informationen zu verbessern und die Gefahr eines Verlustes von Daten oder relevanten Kontextinformationen zu minimieren. Darüber hinaus leisten transparente und gut strukturierte Arbeits- und Kommunikationsprozesse einen wichtigen Beitrag zur Erhöhung der Datenqualität: Sie verbessern die Zusammenarbeit der Beteiligten und helfen sicherzustellen, dass die erforderlichen Informationen zur richtigen Zeit und an der richtigen Stelle auffindbar und zugreifbar

sind. Am ehesten kann dies erreicht werden, wenn sich die Organisation von Dateien unmittelbar an den im Projekt definierten Arbeitsprozessen und Informationsflüssen orientiert und diese abbildet.

Die Grenzen zwischen genuinen Forschungsaktivitäten, Maßnahmen des Projektmanagements und des Datenmanagements verlaufen demzufolge teilweise fließend und eine eindeutige Zuordnung von Aktivitäten zu einem der drei Bereiche ist nicht immer möglich. Im Folgenden soll das Thema Dateiorganisation und Strukturierung von Informationen daher aus zwei Blickwinkeln betrachtet werden:

- Auf der *Prozessebene* stellt sich die Frage, wie Informationsflüsse über den gesamten Projektverlauf so ausgestaltet werden können, dass Informationen (z.B. Daten, Dokumentationen, Prozessinformationen) sicher dort verfügbar sind, wo sie benötigt werden. Dies erfordert zum einen das Wissen darüber, welche Daten, Dokumentationen sowie Prozessinformationen im Projektverlauf entstehen und von wem sie wann und in welcher Form genutzt bzw. weiter bearbeitet werden. Sollen die Projektergebnisse Dritten zur Nachnutzung zur Verfügung gestellt werden, erfordert es zum anderen die Verknüpfung der Forschungsdaten mit allen Kontextinformationen, die für das Verständnis und die Nutzbarkeit der Daten über die Projektlaufzeit hinaus von Bedeutung sind.
- Auf der *Dateiebene* gilt es im Rahmen des Forschungsdatenmanagements Organisationsprinzipien zu finden, mit Hilfe derer die im Projekt anfallenden Informationen sinnvoll geordnet werden können – sei es in einer lokalen oder webbasierten Verzeichnisstruktur oder auch in einer kollaborativen Arbeitsumgebung.

Vor allem in Forschungsprojekten mit regional verteilten Partnern werden zunehmend Onlineportale genutzt, die einen geschützten virtuellen Arbeitsraum für das Daten- und Projektmanagement sowie die Kommunikation und den Informationstransfer zwischen den beteiligten Forscherinnen und Forschern bieten. Die im Folgenden vorgestellten Empfehlungen zur Organisation von Dateien und Informationsflüssen lassen sich sowohl in verzeichnisbasierten Strukturen als auch in Online-Portalen umsetzen.

Im Abschnitt 5.1 soll zunächst aufgezeigt werden, wie Modelle, die den Lebenszyklus von Forschungsdaten beschreiben, genutzt werden können, um die im Projekt anfallenden Informationen auf der Datei- und der Prozessebene zu strukturieren. Diese theoretischen Überlegungen werden anschließend mit Hilfe eines Beispielszenarios konkretisiert (5.2). Erläuterungen und Hinweise zu allgemeinen Maßnahmen der Dateiorganisation – von der Dateibenennung über die Versionskontrolle bis hin zu Zugangsrechten und Back-up-Strategien – schließen das Kapitel ab (5.3).

5.1 Lebenszyklusmodelle für die Projekt- und Dateiorganisation

Wie oben dargestellt, erfordert das sinnvolle Strukturieren der im Projektverlauf entstehenden Informationen eine detaillierte Planung des Forschungsprozesses. Dies beinhaltet zum einen die Beschreibung der auszuführenden Aufgaben, die Festlegung der verantwortlichen Projektbeteiligten sowie die Definition der erwarteten Outputs. Zum anderen müssen die Wege der Informationen im Projektverlauf festgelegt und hierbei die Beziehungen zwischen den Informationen berücksichtigt werden. Die Aktualisierung oder Korrektur einer Information zieht sehr häufig Veränderungen in den Daten und Dokumentationen verknüpfter Projektphasen nach sich und erfordert einen guten Überblick über den Projektworkflow.

Modelle, die den Lebenszyklus von Forschungsdaten über den gesamten Forschungsprozess hinweg (und auch darüber hinaus) abbilden, sind hierbei ein hilfreiches Werkzeug. Sie lassen sich an die Bedürfnisse kleiner wie großer Projekte anpassen und ermöglichen eine detaillierte Strukturierung des Forschungsprozesses in einzelne Phasen, Workflows und

unmittelbare Arbeitsschritte. Mit Hilfe solcher Lebenszyklusmodelle lässt sich auch darstellen, welche Daten und Informationen in einer Projektphase erforderlich sind, neu produziert oder weiter bearbeitet werden bzw. an nachfolgende Phasen transferiert werden müssen. Sie bilden damit nicht nur einen geeigneten Rahmen für die Planung und Organisation der Kommunikation und Zusammenarbeit im Projektverlauf (Prozessebene), sondern hierauf aufbauend auch für die unmittelbare Organisation der Daten und Dokumentationen (Dateiebene). Projektergebnisse können so in strukturierter Weise abgelegt werden.

Es existiert eine Vielzahl verschiedener Lebenszyklusmodelle, die je nach Projekterfordernissen als Vorlage genutzt werden können. Sie beschreiben einfache bis sehr komplexe Projektworkflows und können unterschiedliche Ziele in den Fokus stellen – z.B. die Publikation von Projektergebnissen oder die Übergabe von Daten und Dokumentationen an ein Datenarchiv (vgl. Ball 2012) – oder sind auf bestimmte Studientypen abgestimmt, z.B. Längsschnittstudien oder Kulturvergleiche (vgl. Barkow et al. 2013; Survey Research Center 2016).¹ So bezieht das Lebenszyklusmodell der Data Documentation Initiative (DDI) die Archivierung, Dissemination und Nachnutzung von Forschungsdaten mit ein (vgl. Kapitel 2.2).

Das Survey Lifecycle Model (SLM) der Cross-Cultural Survey Guidelines (CCSG) definiert indessen sehr detailliert die Phasen, die empirische, international oder kulturell vergleichende Forschungsprojekte üblicherweise durchlaufen (vgl. Survey Research Center 2016; vgl. Schaukasten 5.1 für eine Übersicht der Phasen). Die Guidelines erläutern für die einzelnen Phasen die Arbeitsschritte einschließlich der Maßnahmen zur Sicherung der Datenqualität, benennen Verantwortlichkeiten und geben einen Überblick über die entstehenden Daten und Metadaten. Vergleichbar zum DDI-Modell schließt das Modell auch die Veröffentlichung und Nachnutzung der erhobenen Daten mit ein. Es erlaubt damit, die Nutzung der Daten und Metadaten sowie der Dokumentation für die Replikation und die sekundäranalytische Nutzung durch Dritte frühzeitig in der Projektplanung zu berücksichtigen.

Schaukasten 5.1: Phasen des CCSG Survey Lifecycle Modells	
1. Study Design and Organizational Structure	9. Pretesting
2. Study Management	10. Interviewer Recruitment, Selection, and Training
3. Tenders, Bids, & Contracts	11. Data Collection
4. Sample Design	12. Paradata and Other Auxiliary Data
5. Questionnaire Design	13. Data Harmonization
6. Instrument Technical Design	14. Data Processing & Statistical Adjustment
7. Translation	15. Data Dissemination
8. Adaptation	

Quelle: Survey Research Center (2016)

Bei der Organisation von Dateien und Dokumenten und der Strukturierung von Kommunikationsprozessen im Projekt anhand eines Lebenszyklusmodells kann wie folgt vorgegangen werden:

1. Ausgehend von der Art, Zielstellung und Komplexität des Forschungsprozesses werden die erforderlichen Phasen zu einem projektspezifischen Lebenszyklusmodell kombiniert. Zu berücksichtigende Faktoren können hierbei sein:

¹ Ein Überblick über verschiedene Lebenszyklusmodelle findet sich bei Ball (2012).

- a) *Art des Projektes*: Handelt es sich um ein Forschungsprojekt, das eine Datenerhebung plant (vgl. dazu Kapitel 3), oder werden bereits bestehende Daten nachgenutzt, etwa im Rahmen von Sekundäranalysen, wie in Kapitel 8 beschrieben?
 - b) *Zielstellung*: Werden die Daten ausschließlich für die eigene Verwendung oder auch für die Nachnutzung durch Dritte aufbereitet, wie im Zusammenhang mit dem Thema *Data Sharing* in Kapitel 7 diskutiert?
 - c) *Komplexität*: Handelt es sich beispielsweise um ein nationales Projekt einer/eines einzelnen Forschenden oder um eine multinationale Studie, in der regional verteilte Projektpartner zusammenarbeiten, wie im Kontext der Datenaufbereitung und Dokumentation in Kapitel 6 beleuchtet wird?
 - d) *Art der Datenquellen*: Geht es z.B. um Geodaten, die mit Umfragedaten verbunden werden sollen, bzw. Forschungsdaten, die aus Social Media gewonnenen werden, wie in den Kapiteln 11 und 12 thematisiert?
2. Für den entstehenden Projektworkflow werden die konkreten Arbeitsschritte und Prozesse beschrieben sowie die Beteiligten, ihre Rollen und ihre Verantwortlichkeiten benannt.
 3. Darauf aufbauend wird der erforderliche Informationstransfer zwischen den Phasen aufgeführt und die Interaktion der Beteiligten definiert: Welche Daten und Informationen werden aus vorausgehenden Phasen benötigt, welche werden neu produziert und welche müssen an nachfolgende Projektphasen oder Arbeitsschritte übergeben werden? Damit kann dieser Schritt auch die Selektion von Informationen und Materialien beinhalten, die am Ende des Projekts an ein Datenarchiv übergeben werden müssen.
 4. Schließlich werden die Erfordernisse in praktische Strategien und Maßnahmen zur Organisation und Strukturierung der Dateien im Projektverlauf umgesetzt, um Transparenz über die verfügbaren Daten und Informationen herzustellen, einen einfachen Zugang zu ihnen zu garantieren und den Informationsfluss im Projekt zu befördern.

Im nächsten Abschnitt soll anhand von Beispielen dargestellt werden, wie sich auf der Grundlage eines projektspezifischen Lebenszyklusmodells konkrete Workflows und Informationsflüsse definieren und die Organisation von Dateien praktisch umsetzen lässt.

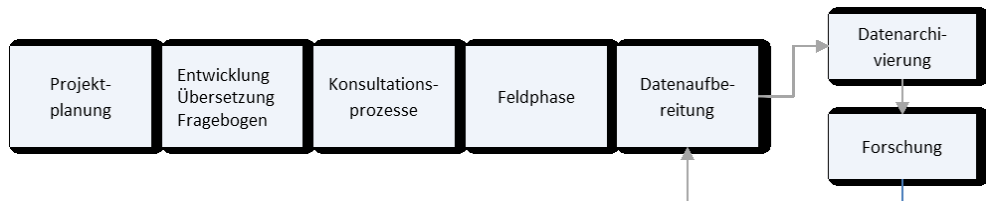
5.2 Beispielszenario: Dateiorganisation in einem multinationalen Primärforschungsprojekt

Betrachtet werden nachfolgend Beispiele für die Organisation von Dateien in einem multinationalen Primärforschungsprojekt, das Daten in mehreren Wellen selbst erhebt und diese nach Projektabschluss über ein Datenarchiv zur Nutzung bereitstellt. Als Beispiel verwenden wir die Europäische Wertestudie (European Values Study), da sie es ermöglicht, komplexe Ordnerstrukturen darzustellen, die sich dann an die Erfordernisse mittlerer oder kleinerer nationaler Projekte anpassen lassen. Unmittelbar kooperierende Projektbeteiligte sind in unserem Beispiel die zentralen Planungsgruppen, wie z.B. die Theoriegruppe und die Methoden-Gruppe, die für die Entwicklung des Erhebungsinstruments bzw. die Erhebung und Aufbereitung der Daten und ihre Qualität verantwortlich sind, sowie die regional verteilten nationalen Teams, die die Erhebungen in den jeweiligen Ländern umsetzen. Darüber hinaus erfolgt eine Abstimmung mit dem Datenarchiv über die zu übergebenden und zu publizierenden Daten und Materialien.

Zunächst wird ein Lebenszyklusmodell erarbeitet, in dem die erforderlichen Hauptschritte zur Realisierung der vereinbarten Projektziele definiert sind (vgl. Abbildung 5.1). Es schließt als Phasen die Projektplanung und Fragebogenentwicklung sowie die Datenerhebung und Datenaufbereitung ein und berücksichtigt eine sich anschließende Phase der Archivierung und Bereitstellung der Daten. Das multinationale Design der Europäischen Wertestudie erfordert als weitere Schritte die Übersetzung des Erhebungsinstruments in die

erforderlichen Sprachen und Konsultationsprozesse zwischen der Methodengruppe und den nationalen Teams, um die Vergleichbarkeit der Erhebungsinstrumente, Stichproben und Harmonisierungskonzepte zu überprüfen. Das Modell ist damit der Ausgangspunkt für eine detaillierte Projektplanung, in welcher für die einzelnen Phasen konkrete Aufgaben und Ergebnisse definiert, die Verantwortlichkeiten zwischen den Projektpartnern vereinbart und Schritte der Qualitätskontrolle implementiert werden (vgl. Macías/Petrakos 2006: 2).

Abbildung 5.1: Beispielhaftes Lebenszyklusmodell für ein multinationales Forschungsprojekt



Quelle: Eigene Darstellung

Betrachtet man den Bestand an Dateien, der in diesem Projekt organisiert werden muss, dann umfasst dieser zu Beginn des Projekts die Daten und Metadaten der vorausgehenden Erhebungswellen. Hierzu zählen z.B. das Erhebungsinstrument und die Frageübersetzungen und die Schritte der Datenharmonisierung, an die bei der Planung der neuen Erhebungswelle angeknüpft werden kann (vgl. zur Wiederverwendung bereits existierender Metadaten Kapitel 9.1.3). Im Projektverlauf werden neue Dokumente, Daten, Metadaten und Prozessinformationen für die aktuelle Erhebungswelle produziert, die als Originalinformationen, interne Arbeitsdateien und zu publizierende Dateien in verschiedenen Versionen abgelegt werden und z.T. besondere Maßnahmen der Datensicherheit erforderlich machen (vgl. Kapitel 4.3). Nach Projektabschluss werden ausgewählte Daten und Dokumentationen, die eine Nutzung durch Dritte unterstützen, in den vereinbarten Formaten an das Datenarchiv übergeben.

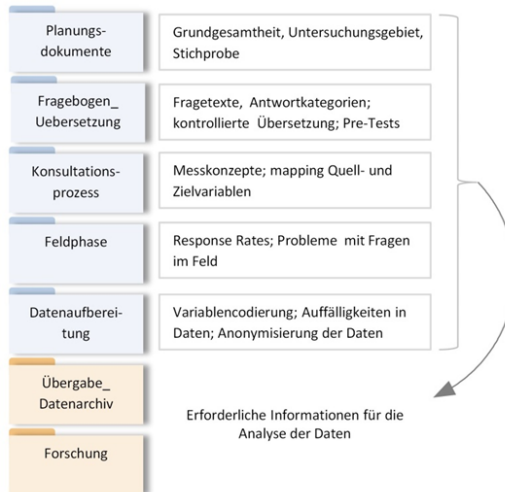
Für die Organisation dieser Dateien wird zu Beginn des Projektes eine Ordnerstruktur aufgebaut (vgl. Abbildung 5.2), die für Projektmitglieder intuitiv verständlich ist und von ihnen konsistent genutzt werden kann. Hierfür werden die Ordner der ersten Ebene aus den Phasen des Lebenszyklusmodells des Projekts – wie der Planungsphase, der Fragebogenübersetzung usw. – abgeleitet.

Die weitere Strukturierung in Unterordner folgt den unmittelbaren Arbeitsprozessen, die für die jeweiligen Projektphasen definiert sind. Somit wird es im Projektverlauf möglich, die Ordnerstrukturen auf der unmittelbaren Arbeitsebene zu erweitern, ohne dass sie für andere Projektmitglieder ihre Transparenz verlieren oder eine konsistente Nutzung erschwert wird. Hierfür werden im Projektkontext generelle und konkrete Regeln vereinbart und Vorgaben für eine standardisierte Beschreibung der Ordner entwickelt, z.B. in Form von Steckbriefen (s. die Schaukästen 5.2 und 5.3). Die Beschreibungen ermöglichen es den Projektbeteiligten, auf einen Blick zu erfassen, welche Arten von Informationen ein Ordner enthält, für wen diese zugänglich sind bzw. sein müssen und ob sie besondere Schutzanforderungen haben.

Generelle Regeln sollen helfen, den Informationsfluss im Projektverlauf zu garantieren. Das erfordert, die Inhalte für die Ordner zu vereinbaren und sicherzustellen, dass erforderliche Daten und Informationen mit den erforderlichen Zugriffsrechten versehen sind und für die jeweiligen Projektmitglieder zur Verfügung stehen. Konkrete Regeln beziehen sich unmittelbar auf die Organisation und Sicherheit der Dateien und sollen helfen, redundante Dateiablagen zu vermeiden, den Zugriff auf aktuelle Dateiversionen zu garantieren oder auch sensitive Informationen vor dem Zugriff sowie Originaldateien vor dem Überschreiben zu schützen. Sie legen darüber hinaus die Häufigkeit fest, in der Sicherungskopien für Dateien

und Back-ups für die Ordnerstruktur erstellt werden müssen und geben Anleitungen für die Benennung und Versionierung von Dateien (s. Abschnitt 5.3.4).

Abbildung 5.2: Hauptordner auf Grundlage des Lebenszyklusmodells



Quelle: Eigene Darstellung

Um den Übergang der Projektergebnisse in die Datenarchivierung und Nachnutzung möglichst einfach zu gestalten, sollte bereits zu Beginn des Projekts definiert werden, welche Daten und Dokumente für die Nachnutzung benötigt werden (s. Kapitel 3.2). Dazu gehören neben den Daten selbst auch ergänzende Dokumente, die es ermöglichen, die methodischen Grundlagen, den Inhalt und die Qualität der Daten sowie ihre Relevanz für das eigene Projekt zu beurteilen. Finale Versionen der entsprechenden Dateien werden dann nicht nur an eine nachfolgende Phase im Lebenszyklus weitergegeben, sondern finden in unserem Beispielszenario zusätzlich Eingang in den Ordner *Übergabe_ Datenarchiv*.

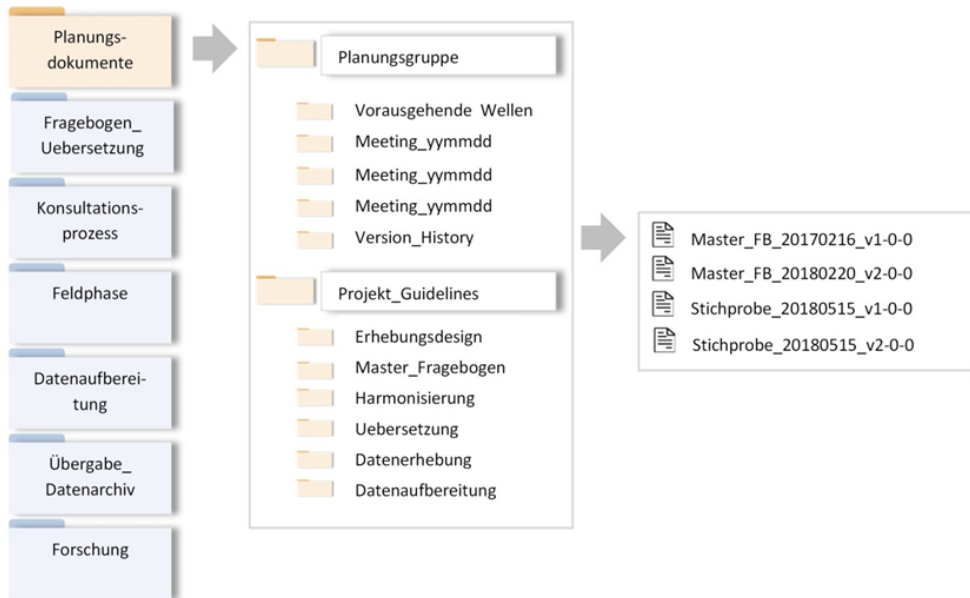
5.2.1 Beispiel 1: Ordner Planungsdokumente

Das erste Beispiel in Abbildung 5.3 beschreibt die Struktur des Hauptordners *Planungsdokumente*, der die Konzepte, Standards und Erhebungsinstrumente enthält, die zu Beginn eines Projektes entwickelt werden. Sie stellen für die Primärforschenden die Grundlage zur Durchführung des Projekts dar und sind für Sekundärforschende eine zentrale Informationsquelle bei der Nachnutzung der Daten. Die Aufgabe des Ordners ist es, für die zentralen Planungsgruppen die Entstehung und Veränderung der Dokumente nachvollziehbar zu machen und für alle Projektmitglieder jederzeit den Zugriff auf die aktuellen Versionen der Projekt-Guidelines zu garantieren.

Die Planungsdokumente werden vorrangig in Besprechungen der zentralen Planungsgruppen, wie der Methodengruppe und der Theoriegruppe bei der Europäischen Wertestudie, entwickelt und in den internen *Meetings-Ordern* abgelegt. An geeigneter Stelle (z.B. auf der übergeordneten Hierarchieebene oder im Ordner selbst) kann ein Ordnersteckbrief *Planungsdokumente* hinterlegt werden, wie in Schaukasten 5.2 dargestellt. Der Ordner *Version History* ermöglicht es, die im Projektverlauf vorgenommenen Veränderungen an den Planungsdokumenten nachzuvollziehen. Die verifizierten Versionen der Dokumente werden in einem Ordner *Projekt Guidelines* gespeichert, der bei komplexeren Projekten weiter in

Unterordner strukturiert werden kann, die thematisch den Lebenszyklusphasen folgen. Die Planungsdokumente stehen damit allen Akteuren zur Verfügung und können die Grundlage für nachfolgende Prozesse und Entscheidungen bilden. Nach Abschluss des Projekts sind sie Teil der Informationen, die zusammen mit den Daten an ein Datenarchiv übergeben werden.

Abbildung 5.3: Struktur des Ordners Planungsdokumente



Quelle: Eigene Darstellung

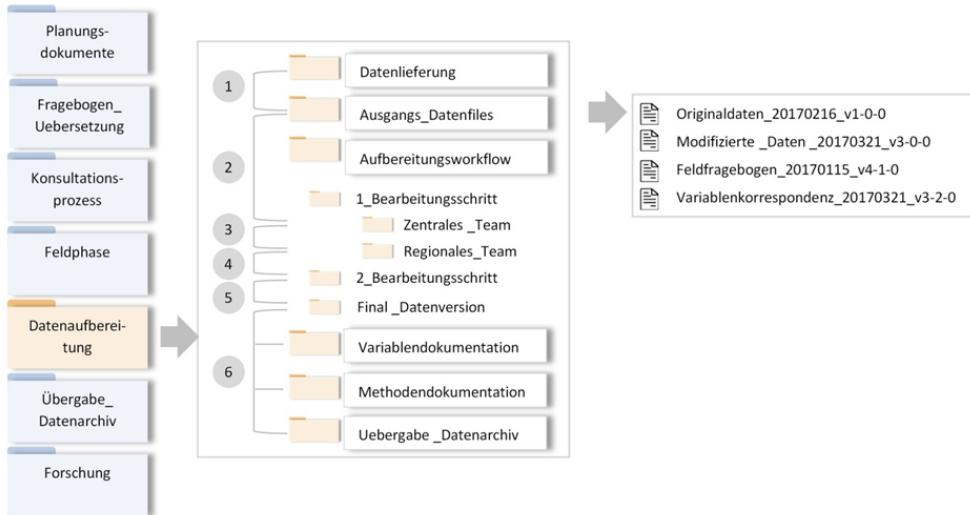
Schaukasten 5.2: Ordnersteckbrief <i>Planungsdokumente</i>	
<i>Kriterium</i>	<i>Beschreibung des Ordners</i>
Inhalt	Erhebungsinstrumente, methodische Konzepte, Standards
Verantwortlichkeit	Zentrale Planungsgruppe(n)
Quelle der Information	Meetings der zentralen Planungsgruppe(n)
weitere Quellen	Ergebnisse vorausgehender Erhebungswellen oder anderer Datenkollektionen
Informationstransfer	an alle folgenden Phasen
Zugang zu den Ergebnissen	alle Projektmitglieder
Zugang zu internen Dokumenten	zentrale Planungsgruppe(n)
besondere Anforderung	Version History für die offiziellen Planungsdokumente
Datensicherheit	--

Quelle: Eigene Darstellung

5.2.2 Beispiel 2: Ordner Datenaufbereitung

Der Hauptordner *Datenaufbereitung* enthält zum einen die Daten und Informationen, die aus der Feldphase an die Phase der Datenaufbereitung übergeben wurden. Er beinhaltet zum anderen die aufbereiteten Daten und Dokumentationen, die nach Abschluss der Projektarbeiten an ein Datenarchiv weitergegeben und für die Forschung bereitgestellt werden (siehe Abbildung 5.4). Seine Aufgabe hierbei ist es, den Dialog und die Zusammenarbeit zwischen den Projektpartnern zu unterstützen, Transparenz zu den entstehenden Daten und Informationen herzustellen und vorgenommene Modifikationen an den Daten nachvollziehbar zu machen (s. Kapitel 6).

Abbildung 5.4: Struktur des Ordners Datenaufbereitung



Quelle: Eigene Darstellung

Hierfür wird eine Ordnerstruktur (inklusive Steckbrief *Datenaufbereitung*, wie in Schaukasten 5.3 dargestellt) aufgebaut, die die Hauptschritte des Datenaufbereitungsworkflows nachbildet. Beschreibende Namen geben Auskunft über die Inhalte der Ordner und die verantwortlichen Akteure sowie den Status, in dem sich die jeweiligen Dateien im Workflow befinden. Die Akteure in diesem Prozess sind ein zentrales Team und regionale Teams, die zur Überprüfung und Verifizierung vorgenommener Bearbeitungsschritte Dateien untereinander transferieren.

- ① Die in der Feldphase erhobenen Daten und Informationen werden von dem regionalen Team in dem Ordner *Datenlieferung* abgelegt und vor dem Überschreiben geschützt. Sie werden vom zentralen Forschungsteam auf Vollständigkeit hin überprüft und in den Ordner *Ausgangs_Datenfiles* abgelegt. Dieser Ordner bildet den Ausgangspunkt für die weiteren Aufbereitungsschritte.
- ②③④ Der Aufbereitungsworkflow kann einen oder mehrere Bearbeitungsrunden umfassen. Hierbei werden die Daten von dem zentralen Team weiter überprüft und bearbeitet und in den Ordner *Zentrales_Team* abgelegt. Daran anschließend werden die Veränderungen von dem regionalen Team überprüft, bestätigt bzw. weitere Veränderungen vorgeschlagen und die Ergebnisse in den Ordner *Regionales_Team* abgelegt.
- ⑤ Ist die Bearbeitung und Überprüfung abgeschlossen, werden die verifizierten Files durch das zentrale Forschungsteam in den Ordner *Final_Datenversion* abgelegt und sind dem regionalen Team zugänglich.

- ⑥ Interne Prozessinformationen werden zusammen mit Informationen aus den vorausgehenden Projektphasen in Methodenberichten und Variablendokumentationen zusammengestellt. Sie werden zusammen mit den aufbereiteten Daten und ggf. Parادات² in dem Ordner *Uebergabe_Datenarchiv* gespeichert.

Schaukasten 5.3: Ordnersteckbrief <i>Datenaufbereitung</i>	
<i>Kriterium</i>	<i>Beschreibung des Ordners</i>
Inhalt	Daten, Metadaten, Dokumente, Prozessinformationen
Verantwortlichkeit	zentrales Aufbereitungsteam/nationales Team
Quelle der Information	Daten und Metadaten aus der Feldphase
weitere Quellen	Informationen aus den vorausgehenden Projektphasen
Weitergabe der Ergebnisse	für Analysen der Primärforscher und Übergabe an Datenarchiv
Zugang zu den Ergebnissen	Primärforschende/Sekundärforschende
besondere Anforderung	Sicherung der Originaldaten Version History für alle Datenmodifikationen
Datensicherheit	Sicherung von nicht faktisch anonymisierten Daten

Quelle: Eigene Darstellung

Die bisherigen Beispiele beschreiben die Dateiorganisation in einem größeren Forschungsprojekt. Ausgehend von dem Lebenszyklus des Projekts wurde eine Ordnerstruktur entwickelt, die es ermöglicht, die Daten und Informationen der Projektphasen strukturiert abzulegen und durch eine tiefere Gliederung in Unterordner die Zusammenarbeit der beteiligten Akteure zu unterstützen. Mittlere und kleine Projekte können dieses Modell einfach an ihre Erfordernisse anpassen. Das Lebenszyklus-Modell beinhaltet dann nur die für das Projekt erforderlichen Phasen und eine tiefere Strukturierung der Ordner wird nur vorgenommen, wenn z.B. die Versionierung von Files nicht ausreichend ist, um die einzelnen Bearbeitungsschritte der Projektmitglieder nachvollziehen zu können.

5.3 Allgemeine Maßnahmen und Hinweise zur Dateiorganisation

Zum Abschluss sollen noch allgemeine Handlungsempfehlungen zur Organisation, Benennung und Handhabung von Dateien im Projektkontext gegeben und so die Anpassung der in den vorangegangenen Beispielen geschilderten Maßnahmen an eigene Projekte erleichtert werden.

5.3.1 Konventionen für die Benennung von Dateien und Ordnern

Unabhängig vom Projektumfang verbessern durchdachte und konsistent angewandte Regeln für die Benennung von Dateien und Ordnern die Auffindbarkeit gesuchter Informationen. Sie verringern mittelbar auch die Gefahr, an veralteten oder falschen Versionen einer Datei zu

2 Bei Parادات handelt es sich um Daten, die im Prozess der Erhebung von Umfragedaten entstehen. Je nach Erhebungsmodus können das z.B. sein: die Anzahl und das Ergebnis der Kontaktaufnahmen mit dem Befragten, Zeitpunkt und Dauer der Befragung, die Beschreibung der Wohngegend bzw. des Haushalts oder auch das Antwortverhalten der befragten Personen (vgl. Felderer/Birg/Kreuter 2014: 357f).

arbeiten und dienen so der Qualitätssicherung. Projekte sollten daher Benennungskonventionen für Dateien und Ordner formulieren und schriftlich fixieren. Die Einhaltung der Konventionen sollte regelmäßig überprüft werden. Schaukasten 5.4 enthält Hinweise zu formalen Aspekten, die bei der Benennung von Ordnern und Dateien zu beachten sind.

Schaukasten 5.4: Namenskonventionen für Ordner und Dateien

Neben inhaltlichen Aspekten sind bei der Benennung von Dateien und Ordnern auch formale Regeln zu beachten. Dies hilft sicherzustellen, dass Datei- und Pfadnamen von verwendeter Software verarbeitet werden können. So sollten die Namen von Ordnern und Dateien:

- keine Leer- und Sonderzeichen enthalten (Ausnahme: Unterstriche und Bindestriche sind möglich),
- nicht zu lang sein, da unter Windows nur eine maximale Pfadlänge von 255 Zeichen möglich ist. Längere Pfade können bedeuten, dass Dateien nicht mehr geöffnet oder verschoben werden können oder dass automatisierte Back-ups nicht erfolgen.
- einheitlichen Regeln für Groß-/Kleinschreibung folgen, da manche Betriebssysteme oder Softwareprodukte diesbezüglich unterscheiden. D.h. *MeetingProtokoll.txt* und *Meetingprotokoll.txt* werden in manchen Umgebungen als identisch, in anderen als zwei unterschiedliche Dokumente interpretiert.

Quelle: Eigene Darstellung

Auf inhaltlicher Ebene können insbesondere Dateinamen Informationen transportieren, die es erlauben, den Inhalt der Datei zu erkennen – z.B. den Dokument- oder Dateityp, das Erstellungsdatum, die erstellende Person oder die Version. So enthält etwa der Dateiname *quest_che_ger_20170401_v1-0-1.docx* folgende Informationen:

- *quest*: Es handelt sich um einen Fragebogen;
- *che*: Der Fragebogen bezieht sich auf die Schweiz (Verwendung des Codes für Länder nach ISO 3166);
- *ger*: Die Sprache des Fragebogens ist Deutsch (Verwendung des Codes für Sprachen nach ISO 639);
- *20170401*: Die Datei wurde am 1. April 2017 erstellt (angegeben im Format YYYYMMDD);
- *v1-0-1*: Es handelt sich um Version 1.0.1.

Für die Benennung von Ordnern sind sprechende Namen zu empfehlen, wie in den Beispielen verwendet. Diese sorgen für eine möglichst intuitive Nutzbarkeit der Ordnerstruktur und erleichtern Projektmitarbeitenden so die Orientierung und das zügige Auffinden relevanter Informationen. Zu diesem Zweck ist es außerdem empfehlenswert, Ordner nicht zu tief zu strukturieren, sondern sich, wenn möglich, auf wenige Hierarchieebenen zu beschränken. Zentral abgelegte Kurzbeschreibungen der Ordner, z.B. in Form der beschriebenen Steckbriefe (vgl. Schaukästen 5.2 und 5.3), können die Orientierung für alle Projektmitarbeiterinnen und -mitarbeiter weiter erleichtern und damit einen effizienten Arbeitsprozess unterstützen.

5.3.2 Metadaten und Versionskontrolle

Um die Datei- und Ordnernamen nicht zu überfrachten, ist es empfehlenswert, weitere Informationen zum Inhalt oder Bearbeitungsstatus einer Datei auf leicht zugängliche Weise in anderer Form zu dokumentieren. So können Informationen zur Version und darüber, wer ein Dokument erstellt hat, wann es erstellt wurde und wann es durch wen zuletzt bearbeitet wurde im Kopfbereich des Dokuments oder als Variable im Datensatz mitgeführt werden. Wenn dies nicht möglich oder praktikabel ist, können diese Informationen auch in einer gesonderten Datei festgehalten werden.

Es kann zudem hilfreich sein, eingebettete Dateiiinformationen zu pflegen. Beispielsweise ermöglichen es alle gängigen Betriebssysteme, Schlagworte und Kategorien für Dateien zu vergeben. Diese können die gezielte Suche nach Informationen unterstützen, sollten aber

nicht als ausschließliches Mittel zur Speicherung dieser Informationen genutzt werden, da sie programm- und betriebssystemabhängig sind und bei Systemwechseln in der Regel nicht mittransportiert werden.

Schaukasten 5.5: Beispiel Versionshistorie und -informationen im Dokumentkopf				
Dokumenttitel	Länderfragebogen Schweiz (deutsch)			
Inhalt	Fragebogen für die Erhebungswelle 2018 (Schweiz, deutschsprachig)			
Dokument erstellt von	Susanne Dubois			
Dokumentstatus	Entwurf			
Verantwortlich	Susanne Dubois			
Beziehung zu anderen Dokumenten	Übersetzung des englischen Master-Questionnaire 2018, Version 1.0.0			
	Version	Geändert von	Änderungsprotokoll	Geändert am
	2.0.0	Mika Taylor	fehlende Fragen 43 und 44 eingefügt	2017-05-12
	1.1.2	Mika Taylor	Tippfehler im Erläuterungstext Frage 16 korrigiert	2017-05-08
	1.1.1	Alex Friedrich	Tippfehler in Fragen 2, 4, 11 korrigiert	2017-05-03
	1.1.0	Alex Friedrich	Überarbeitung der Intervieweranweisung in Fragen 10, 20, 25	2017-04-15
	1.0.0	Susanne Dubois	Fragebogen als Übersetzung erstellt	2017-04-01

Quelle: Eigene Darstellung

Die Versionierung und das Protokollieren von Änderungen an Dokumenten und Datensätzen dienen der Nachvollziehbarkeit und damit auch der Qualitätssicherung im Projekt. Es wird z.B. Transparenz darüber hergestellt, welche Veränderungen an Datensätzen und Dokumenten seit der ersten Erstellung vorgenommen wurden. Jedes Projekt sollte Regeln für die Versionierung festlegen und als Richtlinie schriftlich fixieren. Dies beinhaltet Regeln, wie Änderungen zu dokumentieren sind und welche Änderungen zu einer neuen Version führen. Bei der Erstellung von Versionierungsregeln ist auch darauf zu achten, dass Beziehungen zwischen unterschiedlichen Versionen zusammengehöriger Dateien ersichtlich bleiben.

Die Version einer Datei kann an ganz unterschiedlichen Stellen festgehalten werden, wie etwa im Dateinamen oder in der Datei selbst und verschiedene Formen annehmen. So kann die Version einer Datei etwa durch das Datum der letzten Bearbeitung oder eine fortlaufende Versionsnummer markiert werden. Verbreitete Verfahren sind hierbei das der *Major.Minor*-Versionsnummer (zweistellig) oder der *Major.Minor.Sub-minor*-Versionsnummer (dreistellig). Auch die Versionshistorie kann in einem separaten Dokument gepflegt oder direkt im Kopf des betreffenden Dokuments mitgeführt werden (s. Schaukasten 5.5).

Jensen (2012: 42) empfiehlt bei der Versionierung von Datensätzen im dreistufigen Verfahren folgende Regeln zum Hochzählen der Versionsnummer:

- „1. Position / Major-Nummer: Sie wird verändert, wenn
 - eine oder mehrerer Fälle [sic] in (aus) einem Datensatz eingefügt (gelöscht) werden;
 - eine oder mehrerer Variablen [sic] in (aus) einem Datensatz eingefügt (gelöscht) werden;
 - eine oder mehrere neue Wellen in einen integrierten Datensatz eingefügt werden;
 - ein o. mehrere Sample in den integrierten o. kumulierten Datensatz eingefügt werden.

- 2. Position / Minor-Nummer: Sie wird verändert, wenn Änderungen einer Variablen, d.h. bedeutungsrelevante Korrekturen oder Ergänzungen im Datensatz vorgenommen werden (Labels, Recodierungen, Datenformate, etc.).
- 3. Position / Revision-Nummer: Sie wird verändert, wenn etwa einfache Überarbeitungen von Labels ohne Bedeutungsrelevanz vorgenommen werden.“

Analog dazu können selbstverständlich auch Regeln für die Versionierung von Textdokumenten erstellt werden.

Versionskontrolle kann intellektuell erfolgen oder teilweise automatisiert mit Hilfe von Software wie z.B. *git* oder *Subversion* (vor allem für Quelltext), *Etherpad*, *Owncloud* (Textdokumente) oder dem *Open Science Framework*. Hierbei ist zu beachten, dass die Software zwar in der Regel die Änderungen an einem Dokument automatisch aufzeichnet, dass aber häufig eine Versionsnummer zur Bezeichnung eines bestimmten Stands manuell vergeben werden muss und so kommentiert werden sollte, dass deutlich wird, wie Versionen sich unterscheiden.

5.3.3 Zugangsrechte

Im Rahmen der Planung der Informationsflüsse und der Dateioorganisation im Projekt sollten insbesondere in kollaborativen Projekten auch differenzierte Zugangsrechte zu Ordnern sowie ggf. weitere Schutzmaßnahmen für sensitive Daten schriftlich festgelegt werden. Dies kann z.B. auf Netzlaufwerken über Berechtigungen erreicht werden. So empfiehlt es sich vor allem, Masterfiles durch Beschränkung der Schreibrechte gegen Überschreiben oder Löschen zu schützen und Änderungen nur an Arbeitsversionen der Daten und Dokumente vorzunehmen.

Wann immer im Projekt personenbezogene Daten anfallen, sind geeignete Schutzmaßnahmen zu ergreifen, damit diese Informationen nur von autorisierten Personen gelesen und bearbeitet werden können. Hierzu gehören restriktive Lese- und Schreibrechte, Passwortschutz und die Verwendung von Verschlüsselungsprogrammen.

Bei der technischen Verschlüsselung von Dateien oder Laufwerken werden diese von einem geeigneten Programm auf der Bit-Ebene so verändert, dass sie keine Informationen mehr transportieren, sondern nur noch aus bedeutungslosen Zeichenfolgen bestehen. Nur mit Hilfe des richtigen Schlüssels in Form eines Passworts können die Bitsequenzen der Originaldateien wiederhergestellt und lesbar gemacht werden. Selbst wenn nun unautorisierte Personen Zugriff auf Dateien oder Laufwerke erhalten – z.B. weil sie bei der Übermittlung über unsichere Kommunikationskanäle mitgelesen wurden (unverschlüsseltes WLAN, E-Mail), weil ein Datenträger in die falschen Hände geraten ist (Verlust, Diebstahl) oder weil jemand sich Zugang zu dem Server verschafft hat (Hacking) –, ist ein Auslesen der gespeicherten Informationen ohne Kenntnis des Schlüssels nicht möglich. Das bedeutet aber auch, dass bei Verlust des Schlüssels die Informationen für die Forschenden selbst nicht mehr zu rekonstruieren sind, weshalb dieser unbedingt sicher und vor Verlust geschützt aufbewahrt werden sollte.

Grundsätzlich ist von einer Speicherung und Übermittlung personenbezogener Daten ohne vorherige Verschlüsselung abzuraten. Einen Überblick über verfügbare Software bieten UK Data Service (2018) und der vierte Teil des CESSDA Expert Tour Guide on Data Management (2017).

Eine weitere Maßnahme zum Schutz von Personen, deren Daten in Forschungsprojekten erhoben und gespeichert werden, ist die getrennte Speicherung von direkten Identifizierungsmerkmalen und übrigen erhobenen Daten. Zu diesem Zweck würden z.B. Kontaktinformationen wie Namen und E-Mail-Adressen aus dem Datensatz entfernt und in einer separaten,

besonders geschützten, Datei gespeichert werden. Eine Zuordnung der Kontaktinformationen zum jeweiligen Fall im Datensatz erfolgt dann über eine ein-eindeutige ID (*Schlüssel*).

5.3.4 Back-up-Strategien

Unabhängig davon, wie Dateien im Projekt organisiert werden, sollte mit Hilfe einer geeigneten Back-up-Strategie dafür Sorge getragen werden, dass es nicht zu versehentlichen Verlusten von Dateien und Informationen kommt. Der Abschnitt „Backup“ im vierten Kapitel des CESSDA Expert Tour Guide on Data Management (CESSDA 2017³) empfiehlt hierzu die Orientierung an den folgenden Fragen:

1. Verfügt die eigene Institution über eine Back-up-Strategie? Wenn ja, sollten Projekte Details dieser Strategie in Erfahrung bringen und mit den eigenen Anforderungen abgleichen. Sofern aus Projekt-sicht Maßnahmen erforderlich sind, die über die institutionelle Strategie hinausgehen, sollten diese in der Projektplanung berücksichtigt und entsprechende Kosten eingeplant werden.
2. Was soll gesichert werden? Hier gilt es zu entscheiden, ob bei jedem Back-up der komplette Bestand an Dateien gesichert werden soll oder ob z.B. nur seit dem letzten Back-up aktualisierte Dateien in die Speicherung einfließen sollen.
3. Wie viele Kopien werden benötigt und wo sollen sie gesichert werden? In der Regel wird empfohlen, drei redundante Sicherungskopien der Projektdateien anzufertigen und möglichst an unterschiedlichen Orten aufzubewahren (z.B. auf Servern, die nicht im selben Gebäude stehen). Die Häufigkeit der Back-ups sollte sich daran orientieren, wie häufig (umfangreiche) Änderungen an den Dateien vorgenommen werden.
4. Wieviel Speicherkapazität wird benötigt? Wenn eine größere Kapazität benötigt wird, als z.B. die eigene Institution kostenlos zur Verfügung stellt, sollten entsprechende Kosten bei der Projektplanung berücksichtigt werden.
5. Gibt es Software, mit denen Back-ups gemäß den Projektanforderungen automatisiert durchgeführt werden können? Automatische Back-ups können helfen, Datenverlust durch menschliches Versagen (z.B. durch Vergessen oder versehentliches Überschreiben) zu verhindern. Wenn eine Software zur automatisierten Erstellung von Sicherungskopien verwendet wird, sollte dennoch regelmäßig überprüft werden, dass dies tatsächlich korrekt erfolgt ist.
6. Wie lange sollen Back-ups aufbewahrt und wie sollen sie vernichtet werden? Es ist häufig nicht notwendig, alle Sicherungskopien, die im Projektverlauf erstellt werden, über die gesamte Projektlaufzeit aufzubewahren. Insbesondere wenn Speicherkapazitäten knapp sind, kann zusätzlicher Speicherplatz durch die (geregelt) Löschung alter Sicherungskopien freigegeben werden. Im Fall von sensitiven, personenbezogenen Daten sollten zudem geeignete Maßnahmen zur sicheren Vernichtung der Informationen erfolgen (z.B. mehrfaches Überschreiben mit Hilfe geeigneter Software).
7. Wie werden personenbezogene Daten geschützt? Die Faustregel im Fall von personenbezogenen Daten lautet: Die Sicherungskopien müssen mit der gleichen Sorgfalt geschützt werden wie die zu sichernden Ausgangsdateien. D.h., auch hier sind, wie oben beschrieben, technische Schutzvorkehrungen wie Passwörter und Verschlüsselung einzusetzen. Zudem ist es bei sensitiven Daten u.U. empfehlenswert, weniger als drei Sicherungskopien anzufertigen.

Wenn unter Berücksichtigung dieser Fragen eine Back-up-Strategie erstellt wurde, sollten – wie in anderen Bereichen der Projekt- und Datenmanagementplanung auch – klare Verantwortlichkeiten benannt werden (vgl. CESSDA 2017). Diese sollten u.a. die (Kontrolle der) Durchführung der Back-ups und die regelmäßige Überprüfung der Lesbarkeit der gesicherten Dateien beinhalten. So ist sichergestellt, dass im Falle eines Datenverlusts keine schwerwiegenden Konsequenzen drohen, die den Projekterfolg gefährden könnten.

3 Die Absätze 1 bis 7 erstellten die Autor/-innen in Anlehnung an den englischsprachigen Guide.

5.4 Zusammenfassung

Im Rahmen des Forschungsdatenmanagements gilt es zu Beginn des Forschungsprojekts wichtige Maßnahmen zur Dateioorganisation und der Gestaltung von Informationsflüssen im Projekt abzustimmen und zu verschriftlichen. Auf der Prozessebene umfasst dies insbesondere die Beschäftigung mit der Frage, welche der im Projekt erhobenen und/oder verarbeiteten Informationen für welche Akteurinnen und Akteure zu welchem Zeitpunkt verfügbar sein müssen. Dies bezieht sich nicht nur auf die primären Forschungsobjekte, d.h. die Daten, sondern auch auf vielfältige Prozessinformationen. Auf der Dateiebene gilt es, das Bemühen um einen möglichst transparenten und effizienten Forschungsprozess mit Regeln zur Benennung und Versionierung von Dateien sowie differenzierten Zugangsrechten zu Informationen zu unterstützen. Übergeordnetes Ziel all dieser Maßnahmen ist die Vermeidung von Informationsverlusten sowie die Nachvollziehbarkeit des Forschungsprozesses mit allen getroffenen Entscheidungen, die Auswirkungen auf die Forschungsergebnisse haben. Dies unterstützt nicht nur den Forschungsprozess im Ausgangsprojekt, sondern erleichtert auch die Nachnutzung von Projektergebnissen durch Dritte.

Lebenszyklus-Modelle von Forschungsdaten können hilfreich dabei sein, Informationsflüsse im Projekt und die Strukturierung von Informationen in Dateien und Ordern zu planen und für alle Projektbeteiligten transparent umzusetzen. So kann anhand der Phasen des Modells veranschaulicht werden, welche Informationen wann und durch wen generiert werden und an wen sie im nächsten Schritt übergeben werden müssen. Sprechende Datei- und Ordernamen und die Ordnung von Dateien anhand der Phasen des Lebenszyklus unterstützen die Auffindbarkeit von Informationen. Die Authentizität und Integrität der gespeicherten Informationen lässt sich mit Hilfe von Versionierungsregeln und einem differenzierten Rechtemanagement für den Zugriff auf Dateien gewährleisten. Derartige Maßnahmen fördern einen reibungslosen Projektverlauf und die erfolgreiche Umsetzung der Forschungsziele.

Literaturverzeichnis

- Ball, Alex (2012): Review of Data Management Lifecycle Models. Version 1.0. REDm-MED Project Document. <http://opus.bath.ac.uk/28587/1/redm1rep120110ab10.pdf> [Zugriff: 02.05.2018].
- CESSDA (2017): Expert Tour Guide on Data Management. Chapter 4. Store. <https://www.CESSDA.eu/dmguide>.
- Barkow, Ingo/Block, William/Greenfield, Jay/Gregory, Arofan/Hebing, Marcel/Hoyle, Larry/Zenk-Möltgen, Wolfgang (2013): Generic Longitudinal Business Process Model. DDI Working Paper Series – Longitudinal Best Practices 5. <https://doi.org/10.3886/DDILongitudinal05>.
- Felderer, Barbara/Birg, Alexandra/Kreuter, Frauke (2014): Paradata. In: Baur, Nina /Blasius, Jörg (Hrsg.): Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer.
- International Standardisation Organisation (2007): ISO 639-3:2007 – Codes for the Representation of Names of Languages. Part 3: Alpha-3 Code for Comprehensive Coverage of Languages.
- International Standardisation Organisation (2013): ISO 3166-1:2013 – Codes for the Representation of Names of Countries and Their Subdivisions. Part 1: Country codes.
- Jensen, Uwe (2012): Leitlinien zum Management von Forschungsdaten. Sozialwissenschaftliche Umfragedaten. GESIS Technical Reports 2012/07. <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-320650>.
- Macías, Enrique Fernández/Petrakos, Michalis (2006): Quality assurance in the 4th European Working Conditions Survey. Paper, first EASR conference, Barcelona 18-22 July, 2005.
- Survey Research Center (2016): Guidelines for Best Practice in Cross-cultural Surveys. Full Guidelines. <http://ccsg.isr.umich.edu/> [Zugriff: 02.05.2018].

UK Data Service (2018): Data Encryption. <https://www.ukdataservice.ac.uk/manage-data/store/encryption> [Zugriff: 02.05.2018].

Linkverzeichnis

Etherpad: <http://etherpad.org/> [Zugriff: 02.05.2018].

European Values Study: <http://www.europeanvaluesstudy.eu/> [Zugriff: 02.05.2018].

Git: <https://git-scm.com/> [Zugriff: 02.05.2018].

Open Science Framework: <https://osf.io/> [Zugriff: 02.05.2018].

Owncloud: <https://owncloud.org/> [Zugriff: 02.05.2018].

Subversion: <https://subversion.apache.org/> [Zugriff: 02.05.2018].

