

Fachhochschule Potsdam  
Fachbereich Informationswissenschaften

# Workflow-Aspekte des Forschungsdatenmanagements in den Sozialwissenschaften

---

Eine qualitative Anforderungsanalyse am  
Wissenschaftszentrum Berlin für  
Sozialforschung

Arbeit zur Erlangung des  
akademischen Grades  
Master of Arts

**Christiane Behnert**

Matrikel: 7102

Bearbeitungszeitraum: 11. März - 13. August 2013  
Erstgutachter: Prof. Dr. Stephan Büttner  
Zweitgutachter: Sebastian Nix, M.A. LIS

## Inhaltsverzeichnis

Abbildungs- und Tabellenverzeichnis.....	4
Abkürzungsverzeichnis .....	5
Danksagung .....	6
1. Einleitung.....	7
1.1 Zielstellung .....	11
1.2 Aufbau und Vorgehensweise .....	12
2. Forschungsdaten in den Sozialwissenschaften .....	14
2.1 Datenlebenszyklus .....	20
2.2 Archivierung.....	24
2.3 Verfügbarkeit .....	27
3. Anforderungen an das Forschungsdatenmanagement.....	31
3.1 Allgemeine Anforderungen.....	33
3.1.1 Akteure im Forschungsdatenmanagement .....	36
3.1.2 Anforderungen der Forschenden .....	41
3.1.3 Forschungsdatenmanagement als Prozess.....	45
3.2 Besondere Anforderungen in den Sozialwissenschaften .....	48
3.3 Experteninterviews am WZB .....	56
3.3.1 Ausgangslage .....	58
3.3.2 Methodik.....	61
3.3.3 Leitfragen .....	64
3.3.4 Auswertung.....	66
3.4 Zusammenfassung und Auswirkungen auf den Workflow.....	77
4. Erstellung eines exemplarischen Workflows.....	82
4.1 Voraussetzungen .....	84

4.2 Workflow-Schritte.....	88
4.3 Erkenntnisse.....	94
5. Fazit.....	97
6. Quellenverzeichnis .....	99
Anhang.....	110
Anhang A) E-Mail .....	110
Anhang B) Transkriptionsregeln .....	111
Anhang C) Interview-Transkripte.....	113
C1) Interview 1 .....	114
C2) Interview 2 .....	131
C3) Interview 3 .....	139
Anhang D) Kategoriensystem .....	153
Anhang E) Legende der Notationselemente.....	189
Anhang F) Geschäftsprozessdiagramm .....	193

## Abbildungs- und Tabellenverzeichnis

Abbildung 1: Eigenschaften von Forschungsdaten im Vergleich zu Publikationen .....	15
Abbildung 2: DCC Curation Lifecycle Model .....	21
Abbildung 3: Domänenmodell.....	23
Abbildung 4: OAIS-Funktionseinheiten .....	25
Abbildung 5: Kernkompetenzen im Forschungsdatenmanagement .....	37
Abbildung 6: Akteure als Anbieter und Nutzer der Infrastruktur .....	39
Abbildung 7: Beziehungen zwischen Akteuren im Forschungsdatenmanagement.....	41
Abbildung 8: Ausgewählte Metadatenfelder in DDI-C (Version 2.1) .....	53
Abbildung 9: Ausgewählte Elemente in DDI-L (Version 3.1) .....	54
Abbildung 10: Datenlebenszyklus nach DDI-L.....	54
Abbildung 11: Ablaufmodell der induktiven Kategorienbildung.....	64
Abbildung 12: Pool und Nachrichtenfluss in BPMN 2.0 .....	85
Abbildung 13: Prozessbeispiel in BPMN 2.0.....	86
Abbildung 14: Gateways und Zwischenereignisse in BPMN 2.0 (Auswahl) .....	86
Abbildung 15: Datenobjekte und Artefakte in BPMN 2.0 (Auswahl) .....	87
Tabelle 1: Eckdaten der Experteninterviews.....	63
Tabelle 2: Leitfragen .....	66

## Abkürzungsverzeichnis

AIP	Archival Information Package
ALLBUS	Allgemeine Bevölkerungsumfrage der Sozialwissenschaften
BPMN	Business Process Model and Notation
CESSDA	Council of European Social Science Data Archives
DCC	Digital Curation Centre
DDI	Data Documentation Initiative
DFG	Deutsche Forschungsgemeinschaft
DINI	Deutsche Initiative für Netzwerkinformation
DIP	Dissemination Information Package
DIW	Deutsches Institut für Wirtschaftsforschung
DMP	Datenmanagementplan
DOI	Digital Object Identifier
ESS	European Social Survey
FDM	Forschungsdatenmanagement
FDZ	Forschungsdatenzentrum
GLES	German Longitudinal Election Survey
ICPSR	Inter-university Consortium for Political and Social Research
IDF	International DOI Foundation
JISC	Joint Information Systems Committee
NEPS	National Educational Panel Study
OAIS	Open Archival Information System
RatSWD	Rat für Sozial- und Wirtschaftsdaten
SIP	Submission Information Package
SOEP	Sozio-oekonomisches Panel
WZB	Wissenschaftszentrum Berlin für Sozialforschung
ZBW	Zentralbibliothek für Wirtschaftswissenschaften

## Danksagung

An dieser Stelle möchte ich mich bei all jenen bedanken, die zum Gelingen dieser Masterarbeit beigetragen haben. Ich danke vielmals den Experten für ihre Teilnahme an den Interviews sowie Roland Habich für die Kontaktaufnahme zu den Experten und die Informationen zu den Abläufen im Datenmanagement am WZB. Des Weiteren danke ich meinen beiden Gutachtern sehr herzlich für die hilfreichen Hinweise.

Ein großes Dankeschön geht auch an meine lieben Korrekturleserinnen in Berlin und Greifswald.

Schließlich gilt mein besonderer Dank meinem wunderbaren Lebensgefährten für die Unterstützung in jeglicher Hinsicht.

## 1. Einleitung

Die informationstechnologische Entwicklung hat zu einem Paradigmenwechsel in Wissenschaft und Forschung geführt: In den letzten Jahrzehnten ist eine *Datenflut* entstanden, hervorgerufen durch die Erforschung komplexer Phänomene mithilfe von Computersimulationen. Anhand bereits erhobener Daten werden Analysen unter neuen Gesichtspunkten durchgeführt; neue Forschungsprojekte basieren teilweise ausschließlich auf bestehenden Datensätzen. Die dafür eingesetzten Techniken unterscheiden sich maßgeblich von früheren Forschungsmethoden, so dass *data-intensive science* (auch *enhanced Science*, *eScience*)<sup>1</sup> als viertes Forschungsparadigma angesehen werden kann.<sup>2</sup>

Auch wenn die Datenflut zunächst hauptsächlich im Zusammenhang mit astronomischen Daten genannt wurde (automatisierte Messungen und Berechnungen), geht die Forschung in den Sozialwissenschaften ebenfalls mit einer stetig wachsenden Datenmenge einher: Automatisch erfasste Gesundheitsdaten (medizinische, personenbezogene Daten) sowie Daten in sozialen Netzwerken, Blog- und Foreneinträgen erfordern einen an die aktuellen Bedürfnisse der Forschenden angepassten Umgang sowohl im Hinblick an das Datenmanagement als auch bezüglich der Nachnutzung und der Weitergabe an Dritte.<sup>3</sup>

Zur Sicherung der *datenreichen Zukunft der Sozialwissenschaften*, fordert King unter anderem, das wachsende Interesse an der Nachnutzung der Primärdaten zu unterstützen, was sich sowohl im individuellen Forscherverhalten als auch auf Verlagsseite niederschlagen soll. Die Datenabgabe könne als Voraussetzung für die Publikation der daraus resultierenden Ergebnisse von wissenschaftlichen Zeitschriften eingefordert werden; zugleich müsse bereits in der Lehre die Wichtigkeit von Datenmanagement und Nachnutzbarkeit als Bestandteil guter wissenschaftlicher Praxis stärker betont werden.<sup>4</sup>

---

<sup>1</sup> Vgl. Bell u.a. (2009), S. XXX

<sup>2</sup> Vgl. Hey u.a. (2009), S. xviii ff.

<sup>3</sup> Vgl. King (2011), 719.

<sup>4</sup> Vgl. Ebenda, S. 720.

Um Forschung nicht nur innerhalb der einzelnen Fachdisziplinen, sondern auch interdisziplinär weiter voran zu treiben, müssen sich die wissenschaftspolitischen Rahmenbedingungen an den Wandel anpassen. So fordert bspw. die Deutsche Forschungsgemeinschaft (DFG) bei Projektanträgen die Beschreibung der Maßnahmen für die langfristige Sicherung der im Rahmen des betreffenden Projekts erhobenen Daten. Sie begründet dies mit dem Bestreben, den Umgang mit Forschungsdaten zu verbessern.<sup>5</sup> Anträge sozialwissenschaftlicher Langfristprojekte müssen Nachweise des *kontinuierlichen Datenmanagements* und Vorkehrungen für die praktische Umsetzung enthalten.<sup>6</sup>

Für den allgemeinen Umgang mit Forschungsdaten hat zudem die Allianz der deutschen Wissenschaftsorganisationen<sup>7</sup> im Juni 2010 sechs Grundsätze verabschiedet. Darin verankert ist unter anderem die *wissenschaftliche Anerkennung* der Datenveröffentlichung.<sup>8</sup> Dennoch bestehen weitere Hemmnisse bei Forschenden, ihre Daten zu teilen (*data sharing*). Dies liegt zum einen an dem Charakter sozialwissenschaftlicher Daten (sensible, personenbezogene Daten) und den hiesigen Datenschutzbestimmungen. Zum anderen fehlen Anreize für die Forschenden, ihre Daten anderen zur Verfügung zu stellen, insbesondere dann, wenn die Ergebnisse noch nicht veröffentlicht worden sind und somit ein Wettbewerbsvorteil verloren ginge.

Ein weiterer Hindernisfaktor ist mangelnde Aufklärung bezüglich des Datenmanagements sowie der technischen und organisationalen Infrastruktur. Viele Forschende vergeben weder Metadaten noch besitzen sie Kenntnis über ein geeignetes Repository, an das sie die Daten liefern könnten. Zudem sehen sie deren langfristige Archivierung nach Ende des Forschungsprojekts nicht in ihrer Verantwortung; gleichwohl möchten sie die Kontrolle über die Daten behalten, aus Angst vor Missbrauch oder Fehlinterpretation.<sup>9</sup>

---

<sup>5</sup> Vgl. DFG (Deutsche Forschungsgemeinschaft) (2012), S. 6.

<sup>6</sup> Vgl. DFG (Deutsche Forschungsgemeinschaft) (2011), S. 3.

<sup>7</sup> Zu ihr gehören: Alexander von Humboldt-Stiftung, Deutsche Akademie der Naturforscher Leopoldina, DFG, Deutscher Akademischer Austauschdienst, Fraunhofer-Gesellschaft, Helmholtz-Gemeinschaft, Hochschulrektorenkonferenz, Leibniz-Gemeinschaft, Max-Planck-Gesellschaft und Wissenschaftsrat.

<sup>8</sup> Vgl. Allianz der deutschen Wissenschaftsorganisationen (2010), S. [1].

<sup>9</sup> Vgl. Verhaar u.a. (2010), S. 20 ff.



Studien<sup>10</sup> zum Thema Forschungsverhalten im Hinblick auf Datennutzung, -speicherung und -zugriff zeigen auf, dass die Einstellung zu *data sharing* in den jeweiligen Fachdisziplinen unterschiedlich ist: Forschende im naturwissenschaftlichen Bereich scheinen eher dazu geneigt, ihre Daten zu veröffentlichen, als deren Kolleginnen und Kollegen anderer Fachdisziplinen.<sup>11</sup>

Dieser Feststellung steht jedoch die Tatsache gegenüber, dass *data sharing* in den Sozialwissenschaften seit längerem praktiziert wird und Deutschland sogar eine Vorreiterrolle einnimmt bezüglich des Aufbaus einer entsprechenden Forschungsinfrastruktur. Die fach- und auch themenspezifischen Forschungsdatenzentren (FDZ) und Datenarchive<sup>12</sup> stellen Daten aus großen, meist national durchgeführten Studien zur Verfügung, während die Archivierung und Weitergabe von Daten aus kleinen Projekten die Ausnahme bilden.<sup>13</sup>

Die Veröffentlichung der Forschungsdaten ohne entsprechende Metadaten oder Einhaltung von Standards, auch bei der Archivierung, stellt jedoch keine Lösung des Problems dar.<sup>14</sup> Um die langfristige Sicherung und Nachnutzung von Forschungsdaten zu gewährleisten, müssen neben der geeigneten Infrastruktur auch informationswissenschaftliche „Werkzeuge“ zum Einsatz kommen.

Mit dem Thema Forschungsdatenmanagement (FDM) befasst sich das Bibliothekswesen seit einigen Jahren. Dies zeigen diverse Publikationen<sup>15</sup> und die Präsentation zahlreicher Projekte und Erfahrungsberichte auf den Bibliothekartagen der letzten Jahre.<sup>16</sup> Dabei kann die Frage nach der informationswissenschaftlichen bzw. bibliothekarischen Verantwortung nicht nur mit den praktischen Hilfsmitteln beantwortet werden: Nach der *DIKW-Hierarchie (Data - Information - Knowledge -*

---

<sup>10</sup> Die Ergebnisse einer vergleichenden Literaturstudie von 15 Quellen aus Australien, Europa und den USA liefert Feijen (2011).

<sup>11</sup> Vgl. Kuipers & Van der Hoeven (2009), S. 20.

<sup>12</sup> Vgl. Habich u.a. (2010), S. 4; Huschka u.a. (2011), S. 35 ff.; Dallmeier-Tiessen u.a. (2009), S. 6.

<sup>13</sup> Vgl. Weichselgartner u.a. (2011), S. 193.

<sup>14</sup> Vgl. King (2011), S. 720.

<sup>15</sup> Siehe z.B. Abschlussarbeiten von Rümpel (2010); Simukovic (2012); sowie die Handbücher von: Büttner u.a. (2011a); Neuroth u.a. (2012).

<sup>16</sup> Siehe Vorträge auf dem OPUS-Server des Berufsverbands Bibliothek und Information e.V. zum Thema „Elektronische Publikationen und Forschungsdaten, elektronisches Publizieren“, URL: <http://www.bib-info.de/verband/publikationen/opus/opus-themen/elektronische-publikationen-elektronisches-publizieren.html> [Zugriff am: 22.04.2013].

*Wisdom*) beruht das Informationsmanagement auf den *Daten*, welche summiert mit der *Bedeutung* auf der nächst höheren Ebene zu *Informationen* werden. Demnach liegt es nahe, auch das Management von Forschungsdaten im informationswissenschaftlichen Bereich anzusiedeln.<sup>17</sup>

---

<sup>17</sup> Vgl. Hobohm (2010), S. 223.

## 1.1 Zielstellung

Da Forschungsdaten je nach Fachgebiet unterschiedlicher Ausprägung sind (z.B. astronomische im Gegensatz zu medizinischen Daten), müssen sie ihren fachspezifischen Anforderungen entsprechend verwaltet werden. Die Masterarbeit konzentriert sich auf den sozialwissenschaftlichen Bereich und bezieht sich auf quantitative Daten, da qualitative Daten andere Eigenschaften besitzen und dadurch andere Anforderungen an den Umgang mit ihnen bestehen.

Das Ziel der vorliegenden Arbeit besteht darin, einen exemplarischen Workflow für das Forschungsdatenmanagement im Wissenschaftszentrum Berlin für Sozialforschung (WZB) zu erstellen und die daraus resultierenden Erkenntnisse für die praktische Umsetzung aufzuzeigen.

## 1.2 Aufbau und Vorgehensweise

Für die Einführung in das Thema Forschungsdaten in den Sozialwissenschaften werden zunächst fachspezifische Begrifflichkeiten und der Forschungsablauf kurz dargestellt. In diesem Zusammenhang spielt die Gestaltung des Datenlebenszyklus eine wichtige Rolle, daher werden zwei ausgewählte Modelle näher vorgestellt. Die langfristige Verfügbarkeit von Forschungsdaten setzt deren vertrauenswürdige Archivierung voraus. Aus diesem Grund werden wichtige Aspekte bezüglich der digitalen Langzeitarchivierung in einem eigenen Unterkapitel vorgestellt. Anschließend wird ein Überblick über Möglichkeiten des Zugriffs sozialwissenschaftlicher Forschungsdaten gegeben (Kapitel 2).

Hieran schließt sich der Hauptteil der Arbeit: Um einen Workflow erstellen zu können, müssen zunächst die Anforderungen an das FDM im sozialwissenschaftlichen Bereich analysiert werden. Dazu werden allgemeine Anforderungen und die fachspezifischen Besonderheiten anhand aktueller deutschsprachiger, sowie Literatur aus dem anglo-amerikanischen Raum und Europa untersucht (Kapitel 3).

Für die Herausstellung sozialwissenschaftlicher Anforderungen an das FDM werden zudem Experteninterviews am Wissenschaftszentrum Berlin für Sozialforschung durchgeführt, die sich an vier Leitfragen orientieren:

1. Welche Arten von quantitativen Daten werden als Forschungsdaten wahrgenommen?
2. Wie gestaltet sich der Datenlebenszyklus?
3. Wie werden die Daten dokumentiert?
4. Wie sieht die Nachnutzung von Daten aus?

Die Auswahl der Forschenden für die Experteninterviews basiert auf drei Forschungsprojekten, die derzeit am WZB durchgeführt werden, wobei deren Forschungsdaten sich jeweils in unterschiedlichen Stadien ihres Lebenszyklus befinden. So soll sich ein umfassendes Bild der Anforderungen an das FDM über den gesamten Datenlebenszyklus ergeben.

Die Experteninterviews werden aufgezeichnet und transkribiert. Die Transkripte werden anschließend anonymisiert (Anhang C) und einer qualitativen Inhaltsanalyse unterzogen. Das Ergebnis ist ein System induktiv gebildeter Kategorien mit den jeweiligen Textpassagen aus den Transkripten (Anhang D). Anhand der literaturbasierten Anforderungsanalyse und den Ergebnissen aus den Experteninterviews werden die Anforderungen zusammengefasst und deren Auswirkungen auf die Gestaltung des Workflows erläutert.

Die Ergebnisse der Anforderungsanalyse bilden die Basis für die Erstellung des exemplarischen Workflows (Kapitel 4). Für die Darstellung der Aufgabenabfolge und der jeweiligen Zuständigkeiten wird die Form eines Geschäftsprozessdiagramms gewählt. Hierfür wird der web-basierte *Signavio Process Editor* unter Verwendung der standardisierten Modellierungssprache BPMN 2.0 (*Business Process Model and Notation*) genutzt, in die vorab eine kurze Einführung gegeben wird. Des Weiteren ist eine Legende über die verwendeten Elemente in Anhang E beigefügt. Die Workflow-Schritte werden ausführlich erläutert und die daraus resultierenden Erkenntnisse, auch im Hinblick auf das WZB und die praktische Umsetzbarkeit, aufgezeigt.

Für das abschließende Fazit werden die wichtigsten Erkenntnisse der Arbeit nochmals zusammengefasst (Kapitel 5).

## 2. Forschungsdaten in den Sozialwissenschaften

Es gibt keine einheitliche Definition von Forschungsdaten, die im Detail für jede Fachdisziplin gilt. Die DFG empfiehlt, die Definition von Forschungsprimärdaten innerhalb der jeweiligen wissenschaftlichen Gemeinschaft selbst zu entscheiden. Forschungsprimärdaten sind allgemein solche „Daten, die im Verlauf von Quellenforschungen, Experimenten, Messungen, Erhebungen oder Umfragen entstanden sind. Sie stellen die Grundlage für die wissenschaftlichen Publikationen dar.“<sup>18</sup> Primärdaten werden je nach Fachdisziplin unterschiedlich interpretiert: Messdaten, bereits aufbereitete Daten oder Daten als Grundlage wissenschaftlicher Publikationen können zum allgemeinen Verständnis als *Forschungsdaten* bezeichnet werden.<sup>19</sup>

Inwieweit sich Forschungsdaten von wissenschaftlichen Publikationen unterscheiden, zeigt Abbildung 1 auf Seite 15 in einer Gegenüberstellung von Merkmalen – insbesondere im Hinblick auf Aspekte der (Nach-)Nutzung. Diese verdeutlicht, dass sich im Umgang mit Forschungsdaten zwangsläufig neue Anforderungen ergeben; in Kapitel 3 werden diese näher untersucht.

Um eine Vorstellung von Forschungsdaten in den Sozialwissenschaften zu bekommen, folgt ein kurzer Überblick über Forschungsinhalte und den Forschungsablauf.<sup>20</sup> Die Sozialwissenschaften sind keine eigenständige Wissenschaftsdisziplin. Der Begriff ist vielmehr ein Sammelbegriff für unterschiedliche Fachdisziplinen, welche „Fragen des gesellschaftlichen Zusammenlebens behandeln“.<sup>21</sup> Als Untermenge der Sozialwissenschaften stellt die empirische Sozialforschung quantitative und qualitative Methoden bereit, mit denen soziale Tatbestände systematisch erfasst und gedeutet werden.<sup>22</sup> *Empirisch* bedeutet dabei, dass Aussagen direkt oder indirekt auf Erfahrungen beruhen, die an ihnen überprüft werden können.<sup>23</sup>

<sup>18</sup> DFG (Deutsche Forschungsgemeinschaft) (2009), S. 2.

<sup>19</sup> Vgl. Klump (2009), S. 104/105.

<sup>20</sup> Das Ziel dieser Arbeit ist es nicht, eine detaillierte Beschreibung über sozialwissenschaftliche Forschung zu liefern; die nachfolgenden Erläuterungen sollen dem Verständnis sozialwissenschaftlicher Begrifflichkeiten in den nachfolgenden Kapiteln dienen.

<sup>21</sup> Quandt & Mauer (2012), S. [61].

<sup>22</sup> Vgl. Bortz & Döring (2006), S. V; Atteslander (2010), S. 3.

<sup>23</sup> Vgl. Fuchs-Heinritz u.a. (2011), S. 165.

Research publication	Research data
Information transformed into results	Information not transformed into results
Use requires basic software and instruments and their command	Use often requires special software and instruments and their command
Self-explanatory	Requires additional information and documentation if not archived
Should not include sensitive information	May include sensitive and confidential information
Use does not require permission	Use often requires permission
Ownership and copyright often clear	Ownership and copyright often unclear
Openly accessed by the scientific community for a fee or for free	Several degrees of openness (from completely open to closed)
Understood as scientific output (mentioned in the CV)	At the moment not understood as scientific output/merit even if the data were published (usually not mentioned in the CV)
Ready to be used by others as such	Use requires processing

Abbildung 1: Eigenschaften von Forschungsdaten im Vergleich zu Publikationen<sup>24</sup>

Der Forschungsablauf wird von Atteslander in fünf Phasen unterteilt:<sup>25</sup>

1. Problembenennung (Hypothesenformulierung)
2. Gegenstandsbenennung (Definition des Forschungsgegenstandes)
3. Durchführung (Anwendung von Forschungsmethoden)
4. Analyse (Auswertungsverfahren)
5. Verwendung (von Ergebnissen)

Im Hinblick auf das FDM sind die Phasen der Durchführung, Analyse und Verwendung von Bedeutung. Zur Durchführung gehören nach Festlegung des Forschungsdesigns die Datenerhebung und Datenerfassung. Das Forschungsdesign stellt die Untersuchungsform dar, mit der die Daten erhoben werden sollen, z.B. ob Beobachtungen oder Befragungen durchgeführt, ob Daten einmalig oder zu verschiedenen Zeitpunkten an

<sup>24</sup> Quelle: Kuula & Borg (2008), S. 28.

<sup>25</sup> Vgl. Atteslander (2010), S. [21] ff.

denselben Personen erhoben werden sollen (*Panel*<sup>26</sup>).<sup>27</sup> Für die eigentliche Datenerhebung müssen organisatorische Vorbereitungen getroffen werden, z.B. für Befragungen die Schulung der Interviewer, der Fragebogenentwurf oder die Durchführung von Pre-tests. Sind die Daten erhoben, müssen sie erfasst (Rohdatensatz) und in einer bestimmten Art strukturiert und aufbereitet werden (Datensatz), damit die Auswertung erfolgen kann. Mit der Aufbereitung findet zudem eine Datenbereinigung statt, bei der der Datensatz auf Fehler (z.B. Schreibfehler, Übertragungsfehler) geprüft wird. Etwaige fehlende Werte (*Missing Values*) können ersetzt werden, da diese bei der statistischen Auswertung unter Umständen zu Problemen führen. Dieser Vorgang heißt Imputation.<sup>28</sup>

In der empirischen Sozialforschung werden Untersuchungsobjekte mit für die Fragestellung relevanten Merkmalen beschrieben. Diese Merkmale können unterschiedliche Ausprägungen besitzen. Um die Menge aller Merkmalsausprägungen definieren zu können, wurde der Begriff *Variable* eingeführt. Die Variable „Geschlecht“ besitzt z.B. die Ausprägungen „männlich“ und „weiblich“; eine Variable ist auch eine Eigenschaft, die vorhanden oder nicht vorhanden ist, z.B. „Raucher“, „Nichtraucher“.<sup>29</sup>

Die Art der Daten lässt sich der Methodik entsprechend in quantitative und qualitative unterteilen. Quantitative Daten bezeichnen „die Menge aller Merkmalsmessungen“,<sup>30</sup> sie entstehen durch Zählen, Urteilen, Testen, Befragen, Beobachten und physiologische Messungen. Die quantitative Analyse untersucht Messwerte mithilfe statistischer Verfahren.<sup>31</sup> Hierfür werden die Statistikprogramme SPSS und STATA am häufigsten genutzt. Zu den qualitativen Methoden zählen ebenfalls Befragen und Beobachten.<sup>32</sup> Diese zeichnen sich jedoch im Gegensatz zu quantitativer Forschung durch die Verwendung nicht-standardisierter Instrumente aus. Hier werden verbale bzw. nicht-numerische Daten analysiert. Es werden allerdings auch *mixed-method*-Ansätze verfolgt, bei denen bspw. verbale Daten (Interviewtranskripte) zusätzlich einer

---

<sup>26</sup> Vgl. Schnell u.a. (2011), S. 230.

<sup>27</sup> Vgl. Ebenda, S. 7.

<sup>28</sup> Vgl. Fuchs-Heinritz u.a. (2011), S. 297.

<sup>29</sup> Vgl. Bortz & Döring (2006), S. 2.

<sup>30</sup> Ebenda, S. 3.

<sup>31</sup> Vgl. Ebenda, S. 298.

<sup>32</sup> Vgl. Ebenda, S. X/XI.



quantitativen Analyse unterzogen werden (wie z.B. Messen von Worthäufungen).<sup>33</sup>

Soziale Daten lassen sich außerdem nach Art der Untersuchungseinheit auf drei Ebenen unterscheiden:

1. Zur *Makroebene* gehören internationale Organisationen, die Daten produzieren, z.B. UNO, WHO, OECD.
2. Auf der *Mesoebene* sind Regierungen und Verbände sowie Kirchen zu nennen, deren Umfragen sich auf ausgewählte Gebiete beschränken.
3. Zur *Mikroebene* zählen Umfrageinstitute und spezifische Forschungsinstitute, die Auftragserhebungen durchführen, wie z.B. Meinungsbarometer oder auch Radio- und Fernsehquotenmessungen, d.h. hier steht das individuelle Handeln im Vordergrund.

Alle Ebenen sind insofern miteinander verbunden, als dass Probleme auf globaler Ebene (Makro) auch das Individuum beeinflussen (Mikro), z.B. Auswirkungen der Finanzkrise. Umgekehrt wirkt sich beispielsweise der demografische Wandel wiederum auf die Makro- und Mesoebene aus, z.B. hinsichtlich (regionaler) Aus- und Abwanderungen.<sup>34</sup>

Die Auswertung der Daten, die im Rahmen einer Studie neu erhoben wurden, ist die *Primäranalyse* (Primärdaten). Werden bereits vorhandene Daten nachgenutzt, handelt es sich um eine *Sekundäranalyse*. Hierbei werden die (Roh)-Daten mit anderen Methoden oder im Hinblick einer neuen Forschungsfrage analysiert.<sup>35</sup> Dabei ist darauf zu achten, dass „*das Auswahlverfahren, welches für die Sekundäranalyse benutzten Daten angewendet wurde, auch mit dem nunmehr angestrebten Aussagenbereich kompatibel [ist]*“.<sup>36</sup> Wenn bspw. Aussagen über Erwachsene getroffen werden sollen, die Daten jedoch nur von Jugendlichen vorliegen, wäre eine Sekundäranalyse nicht durchführbar. Da sozialwissenschaftliche Forschungsdaten oftmals den Charakter einer Momentaufnahme über einen bestimmten gesellschaftlichen Zustand besitzen, können sie für zu-

---

<sup>33</sup> Vgl. Quandt & Mauer (2012), S. 71/72.

<sup>34</sup> Vgl. Atteslander (2010), S. 354/355.

<sup>35</sup> Vgl. Bortz & Döring (2006), S. 370.

<sup>36</sup> Schnell u.a. (2011), S. 244.

künftige Sekundäranalysen äußerst wertvoll sein, auch wenn dies zum Zeitpunkt der Primärauswertung noch nicht vorherzusehen ist.<sup>37</sup>

Datenquellen für Sekundäranalysen sind in einer Vielzahl vorhanden. So gibt es die Daten der amtlichen Statistik (z.B. Statistisches Bundesamt, Landesämter, kommunale Ämter), staatlicher Behörden (z.B. Bundesversicherungsanstalt), internationaler Organisationen (EU, OECD u.a.) sowie kommerzieller Institutionen (z.B. Banken, Versicherungen). Bei diesen entstehen zwei Arten von Daten: *Aggregatdaten* und *prozessproduzierte Daten*. Aggregatdaten sind Daten, die aus Merkmalen einzelner Untersuchungseinheiten abgeleitet bzw. zusammengefasst sind (Anteile, Mittelwerte), z.B. Wahlbeteiligung pro Stadtbezirk.<sup>38</sup> Prozessproduzierte Daten entstehen automatisch im Rahmen von Verwaltungsprozessen, z.B. bei Kranken- und Rentenversicherungen.<sup>39</sup>

Die Verwendung der Forschungsergebnisse erfolgt durch einen Abschlussbericht an die Förderorganisation und durch Buch- bzw. Zeitschriftenveröffentlichungen.<sup>40</sup> Gegebenenfalls müssen Daten vorab anonymisiert werden, um datenschutzrechtliche Vorschriften einzuhalten oder aus Gründen der Forschungsethik (sensible Daten). Dabei werden die Daten gelöscht, die Lesern Rückschlüsse auf die Identifikation der individuellen Untersuchungsobjekte (Personen, Organisationen) ermöglichen.

Bei qualitativen Daten ist eine Anonymisierung problematisch, da das Löschen aller personenbezogenen Daten den Informationsgehalt mindert oder sogar verfälscht. Aus diesem Grund werden die betreffenden Daten durch erfundene ersetzt, deren Informationsgehalt möglichst nah an den Original-Daten sein sollte (Pseudonymisierung).<sup>41</sup> Die Form der Datenweitergabe entscheidet über den Anonymisierungsgrad: faktisch<sup>42</sup> anonymisierte Daten werden als *scientific use files* Forschenden mit nachweislicher

---

<sup>37</sup> Vgl. Quandt & Mauer (2012), S. 62.

<sup>38</sup> Vgl. Schnell u.a. (2011), S. 244.

<sup>39</sup> Vgl. Quandt & Mauer (2012), S. 67.

<sup>40</sup> Vgl. Schnell u.a. (2011), S. 10.

<sup>41</sup> Vgl. Fuchs-Heinritz u.a. (2011), S. 37.

<sup>42</sup> Eine faktische Anonymisierung liegt vor, wenn bei hohem Zeit- und Arbeitsaufwand der Personenbezug theoretisch rekonstruiert werden kann. Bei absoluter Anonymisierung darf die Möglichkeit nicht bestehen. Vgl. <http://www.forschungsdatenzentrum.de/anonymisierung.asp> [Zugriff: 11.8.2013].

Institutszugehörigkeit zur Verfügung gestellt. Für die Weitergabe in Form von *public use files* müssen die Daten absolut anonymisiert werden.<sup>43</sup>

Nähere Erläuterungen zur Verfügbarkeit von sozialen Forschungsdaten sind in Kapitel 2.3 enthalten. Als Voraussetzung für die langfristige Verfügbarkeit müssen die Daten archiviert werden, dazu erläutert Kapitel 2.2 die notwendigen Voraussetzungen. Zunächst wird jedoch der Lebenszyklus von Forschungsdaten vorgestellt.

---

<sup>43</sup> Vgl. Quandt & Mauer (2012), S. 68.

## 2.1 Datenlebenszyklus

Für die Darstellung des Lebenszyklus von Forschungsdaten existieren mehrere Modelle. Sie alle stellen die einzelnen Phasen heraus, teilweise jedoch in unterschiedlichem Detaillierungsgrad.<sup>44</sup> So wird beispielsweise die Phase der Datenerstellung oft als „Create“ bezeichnet und dabei sind alle Aktivitäten, die zur Erstellung der Daten führen, impliziert. Ist dies der Fall, stellt die langfristige Archivierung einen großen Teil des Lebenszyklus dar, wie das *Curation Lifecycle Model* des Digital Curation Centre (DCC) zeigt, während der *DDI-Lifecycle* (siehe Abbildung 10 auf Seite 54) die Datensammlung und -auswertung explizit beinhaltet. Dabei gestaltet sich der Datenlebenszyklus in den jeweiligen Fachdisziplinen unterschiedlich, wodurch fachspezifische Anforderungen an das FDM bestehen.<sup>45</sup>

Im Folgenden werden zwei ausgewählte Modelle zum Datenlebenszyklus kurz vorgestellt: das DCC Curation Lifecycle Model und das Domänenmodell von Treloar & Harboe-Ree.

Das *DCC Curation Lifecycle Model* (siehe Abbildung 2 auf Seite 21) zeigt einerseits die allgemeinen Tätigkeiten über den gesamten Lebenszyklus (vom Kern ausgehend): die Datenbeschreibung und Bereitstellung der Erhaltungsinformationen (administrative, technische, strukturelle Metadaten), die Erhaltungsplanung, die Beobachtung der wissenschaftlichen Community bezüglich sich ändernder Arbeitsweisen und Standards sowie die Pflege (*Curation*) und Erhaltung (*Preservation*) der Daten.

Des Weiteren werden die sieben Phasen im Einzelnen aufgeführt und sind analog zu *Curation* und *Preservation* als fortwährender Prozess dargestellt (vgl. die Pfeildarstellung), in dem die Archivierung der Daten für die langfristige Verfügbarkeit vorausgesetzt wird. Mit der abgeschlossenen Konzeptentwicklung wird der Kreislauf mit der Datenerstellung bzw. Nachnutzung bereits vorhandener Daten gestartet (*Create or Receive*). Nachdem die gewonnenen Forschungsdaten bewertet wurden, können die für archivierungswürdig befundenen Daten zur Archivierung übergeben und die anderen gelöscht werden (*Appraise & Select; Dispose*). Mit der Übergabe (*Ingest*)

<sup>44</sup> Vgl. Ball (2012).

<sup>45</sup> Vgl. Aschenbrenner & Neuroth (2011), S. [101].

startet der Erhaltungsprozess, der die Durchführung von Archivierungsmaßnahmen (*Preservation Action*) und die Speicherung (*Store*) nach sich zieht und mit dem Zugriff sowie der Nachnutzung erneut in den Teilprozess der Datenpflege übergeht (*Access, Use & Reuse*). Nun können aus den Daten neue Forschungsdaten gewonnen werden, für diesen Zweck werden sie z.B. in ein neues Format übertragen (*Migrate*) oder es wird ein Teil eines Datensatzes ausgewählt und nachgenutzt.<sup>46</sup>

Die Komplexität des gesamten Forschungsdatenmanagements wird anhand des Modells bereits deutlich.<sup>47</sup> Der Prozess der Archivierung ist in diesem Modell übersichtlich dargestellt, er wird im nachfolgenden Kapitel 2.2 unter Berücksichtigung des OAIS-Referenzmodells noch ausführlicher beschrieben.



Abbildung 2: DCC Curation Lifecycle Model<sup>48</sup>

<sup>46</sup> Vgl. DCC (Digital Curation Centre) (2008).

<sup>47</sup> Vgl. Rümpel (2011), S. 29.

<sup>48</sup> Quelle: <http://www.dcc.ac.uk/resources/curation-lifecycle-model> [Zugriff: 11.8.2013].

Das Domänenmodell von Treloar & Harboe-Ree untergliedert den Datenlebenszyklus in drei Domänen und stellt die unterschiedlichen Anforderungen an das FDM hinsichtlich der domänenspezifischen Zugriffs- und Nutzungsmöglichkeiten in den Vordergrund (siehe Abbildung 3 auf Seite 23).<sup>49</sup>

Die *Private Research Domain* beinhaltet die Entstehung und Analyse von Forschungsdaten, auf die in der Regel nur die primär Forschenden oder ausgewählte Beteiligte Zugriff haben. Die Forschenden verwalten ihre Daten mehr oder weniger selbständig und vergeben Metadaten nur in geringem Maße, da in diesem Bereich noch keine Archivierungsmaßnahmen (im Hinblick auf die langfristige Archivierung der Daten) vorgenommen werden. Für die Speicherung kommt ein *Research Data Store* zum Einsatz.

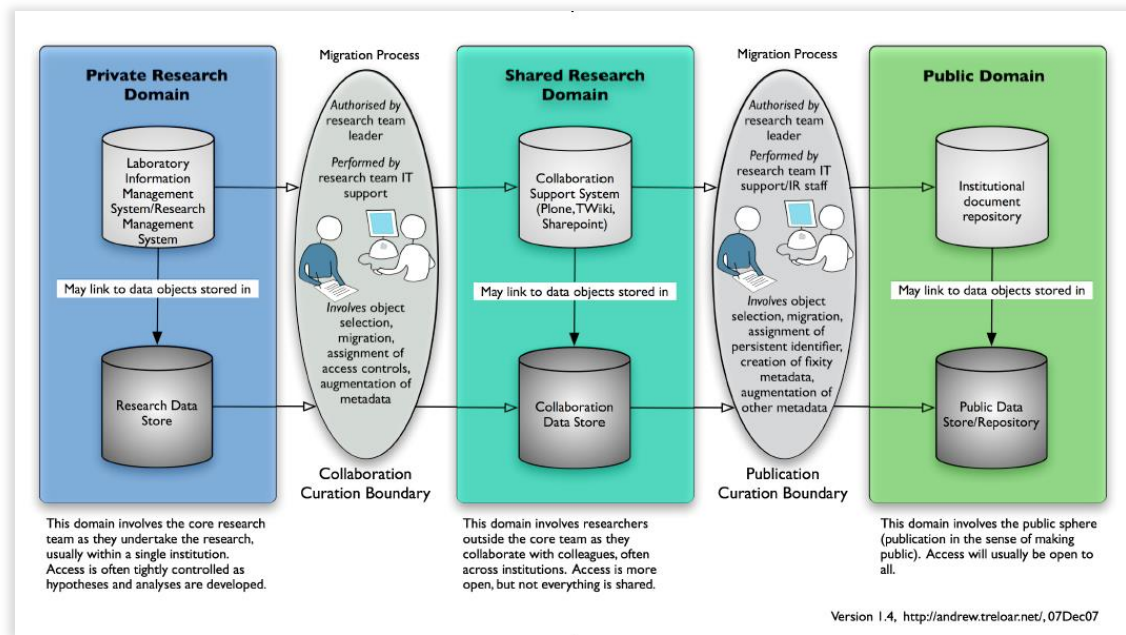
Für die *Shared Research Domain* werden Kopien der Forschungsdaten bereits (in Auswahl) mit anderen Forschenden geteilt, auch außerhalb der eigenen Institution. Hierfür müssen die Daten in einem Repository gehalten werden, um die Strukturierung der Daten besser darstellen zu können (*Collaboration Data Store*), d.h. es gibt einen höheren Bedarf an Metadaten im Vergleich zur ersten Domäne. Außerdem müssen Zugriffsrechte für die kollaborative Nutzung der Daten entsprechend erweitert werden.

Nach Abschluss der Forschung werden ausgewählte Daten innerhalb der *Public Domain* publiziert. Die Nutzung erfolgt in der Regel uneingeschränkt, wodurch ein höherer Bedarf an Metadaten und Archivierungsmaßnahmen besteht. Die Datenverwaltung findet in diesem Bereich vollständig getrennt von den Forschenden statt. Die Daten können über ein institutionelles Repository (*Public Data Store/Repository*) öffentlich zur Verfügung gestellt werden.<sup>50</sup>

---

<sup>49</sup> Vgl. Rümpel (2011), S. 27.

<sup>50</sup> Vgl. Treloar & Harboe-Ree (2008), S. 5-7.

Abbildung 3: Domänenmodell<sup>51</sup>

Bei Betrachtung der beiden Modelle fällt auf, dass sie unterschiedliche Sichtweisen auf den Datenlebenszyklus zeigen.<sup>52</sup>

Das DCC-Modell identifiziert die einzelnen Phasen von der Datenerstellung bis zur Nachnutzung: Der Datenlebenszyklus ist ein fortlaufender Prozess, in dem neu erstellte Forschungsdaten wiederum nachgenutzt werden können und der Kreislauf von Pflege und Erhaltung von neuem beginnt. Auf diese Weise wird deutlich, dass sich die Datenpflege (*Curation*) und Datenerhaltung (*Preservation*) gegenseitig bedingen.

Das Domänenmodell legt den Schwerpunkt auf die unterschiedlichen (speicher)-technischen Anforderungen an die Daten hinsichtlich der Zugriffsmöglichkeiten und beleuchtet weniger die fortlaufenden Aktivitäten innerhalb einzelner Phasen. So wird unter anderem auf Details der Datenerstellung verzichtet und der Charakter eines Kreislaufs nicht in dem Maße herausgestellt, wie es im DCC-Modell der Fall ist.

Für die Beschreibung der Anforderungen an das FDM und die Erstellung des exemplarischen Workflows ist der Prozess-Charakter des Datenlebenszyklus von besonderer Bedeutung, wie im Verlauf der Arbeit aufgezeigt wird.

<sup>51</sup> Quelle: Ebenda, S. 6.

<sup>52</sup> Vgl. Rümpel (2011), S. 27 ff.

## 2.2 Archivierung

Forschungsdaten werden mit dem Ziel der langfristigen Verfügbarkeit und Nachnutzbarkeit archiviert. Das bedeutet, dass die reine Speicherung der Daten auf Bit-Ebene nicht ausreicht, es müssen Vorkehrungen getroffen und Maßnahmen ergriffen werden, damit die Daten auch nach Jahren (und Jahrzehnten) ihrer Entstehung richtig gelesen und vor allem interpretiert werden können. Um dieses Ziel zu erreichen, ist die Wahrung der Authentizität und Integrität der Daten unabdingbar. Diese werden aus rechtlicher Sicht zwar nur bei medizinischen Daten verlangt, bilden jedoch die Basis für die Vertrauenswürdigkeit der Daten und in die jeweilige archivierende Einrichtung. Aus diesem Grund sollte auch bei der Archivierung von sozialwissenschaftlichen Daten darauf geachtet werden.<sup>53</sup> Damit Daten integer sind, müssen deren Unversehrtheit, Lesbarkeit und Interpretierbarkeit gewahrt bleiben.

Die Sicherstellung der Unversehrtheit bedeutet „*die Bit-genaue Erhaltung der archivierten Daten ohne Ansehen ihres Inhalts*“.<sup>54</sup> Dies kann durch Prüfsummen (*checksums*) bei Kopiervorgängen (zwecks Sicherheitskopien oder Erneuerung des Speichermediums) erreicht werden. Dabei wird die Prüfsumme vor und nach dem Kopiervorgang kontrolliert – ist sie identisch, sind es auch die kopierten Daten. Zur Sicherstellung der Lesbarkeit werden Daten in neue Formate konvertiert oder die Software, mit der die Daten gelesen werden, migriert. Dass die Daten interpretierbar bleiben, stellt sich im disziplinübergreifenden Kontext schwierig dar. Hier kommt es insbesondere auf Dokumentationen und Metadaten an, die einerseits den Inhalt betreffen, aber auch technische Metadaten und Informationen zur Herkunft der Daten (Provenienz) beinhalten.<sup>55</sup> Um die langfristige Interpretierbarkeit der Daten zu ermöglichen, sind demzufolge auch disziplinspezifische Anforderungen an Dokumentations- und Metadatenstandards vorhanden.<sup>56</sup> Diese werden unter anderem in Kapitel 3.2 behandelt.

---

<sup>53</sup> Vgl. Spindler & Hillegeist (2010), S. 19.

<sup>54</sup> Lautenschlager (2011), S. [149].

<sup>55</sup> Speziell für die Langzeitarchivierung gibt es den Metadatenstandard PREMIS, der digitale Objekte als intellektuelle Einheiten (*Intellectual Entities*) begreift, die im Zusammenhang mit Ereignissen (*Events*), beteiligten Akteuren (*Agents*) und rechtlichen Regelungen (*Rights*) stehen. Vgl. Hein (2012), S. 98 ff. Siehe auch <http://www.loc.gov/standards/premis/> [Zugriff: 5.8.2013].

<sup>56</sup> Vgl. Lautenschlager (2011), S. [149] ff.



Für die Umsetzung der oben genannten sowie technischen Anforderungen an die Langzeitarchivierung spielt das *Open Archival Information System* (OAIS) eine signifikante Rolle. Dieses standardisierte<sup>57</sup> Referenzmodell sieht die Langzeitarchivierung als Organisation, in der Mensch und System miteinander agieren. Es beinhaltet Standards und Verantwortlichkeiten sowie ein Informations- und Funktionsmodell. Darüber hinaus dient es der Kommunikation über den Aufbau von Langzeitarchiven.<sup>58</sup>

Abbildung 4 zeigt das Funktionsmodell, welches die Archivierung ab der Übergabe der Daten durch den Produzenten bis zur Nachfrage durch den Endnutzer darstellt. Die Funktionseinheiten sind die Übernahme (*Ingest*), der Archivspeicher (*Archival Storage*), die Datenverwaltung (*Data Management*), die Erhaltungsplanung (*Preservation Planning*), der Zugriff (*Access*) und die Administration (*Administration*).<sup>59</sup>

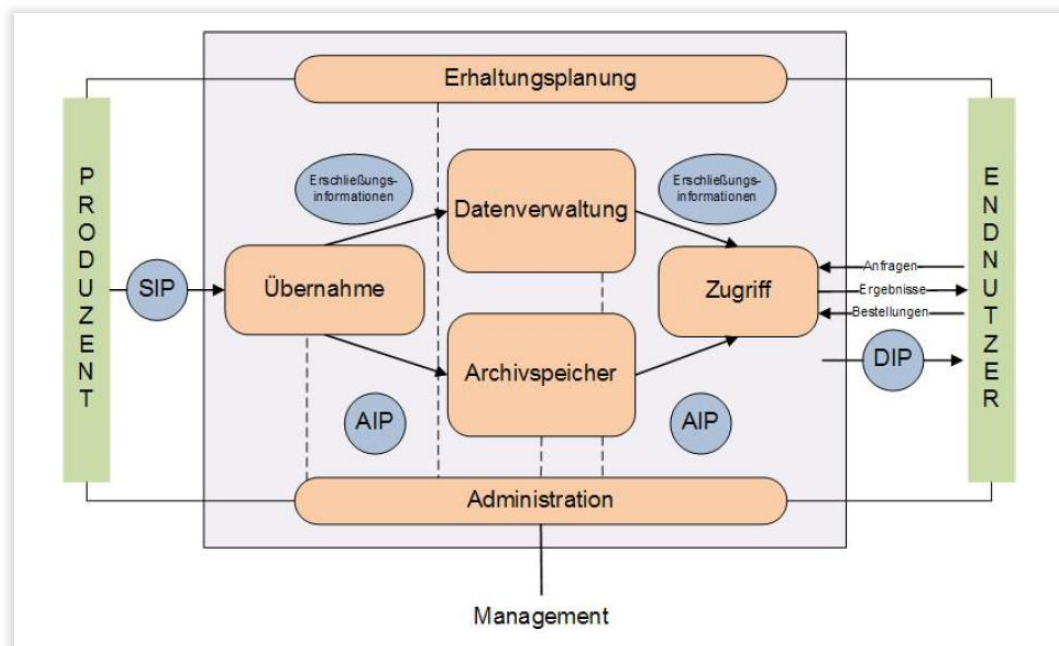


Abbildung 4: OAIS-Funktionseinheiten<sup>60</sup>

Nach der Übernahme des SIP (*Submission Information Package*) wird das AIP (*Archival Information Package*) erzeugt, welches getrennt von den dazugehörigen Metadaten im Archivspeicher abgelegt wird. Bei einer Anfrage werden die gewünschten Daten als DIP (*Dissemination Information Package*) herausgegeben, niemals die originalen AIP.

<sup>57</sup> ISO-Standard 14721:2003.

<sup>58</sup> Vgl. Schimpf (2012), S. 51.

<sup>59</sup> Vgl. nestor-Arbeitsgruppe OAIS-Übersetzung / Terminologie (2012), S. 32 ff.

<sup>60</sup> Quelle: Ebenda, S. 32.

Eine weitere Voraussetzung für die vertrauenswürdige langfristige Archivierung von Forschungsdaten ist neben der OAIS-Konformität des Archivsystems die Bewertung und Selektion der Daten, denn nicht alle können bzw. müssen aufbewahrt werden. Daher ist – wie bei Archivierungsvorgängen üblich – vorab eine Auswahl der archivierungswürdigen Daten zu treffen. Das *DCC Curation Lifecycle Model* stellt diese als eigenständige Phase dar.

Die Auswahl erfolgt anhand unterschiedlicher Kriterien: Neben der Datenmenge (die vor allem in den als datenintensiv geltenden Naturwissenschaften einige Terabyte einnehmen kann) ist der Aufwand der Datenaufbereitung ein wichtiges Auswahlkriterium. Die Qualität der Daten geht einher mit ihrer Beschreibung, denn ohne entsprechende Dokumentationen sind die Forschungsdaten nicht interpretierbar, wodurch die Datenintegrität nicht vollständig gewährleistet wäre (s.o.). Warum quantitative sozialwissenschaftliche Daten ausführlicher Dokumentationen und Metadaten bedürfen, wird in Kapitel 3.2 näher erläutert.

Weitere Auswahlkriterien können geografische Beschränkungen sein, im Hinblick auf nationale oder internationale Datenarchive wie das *UK Data Archive*<sup>61</sup> oder das Datenarchiv des *Inter-university Consortium for Political and Social Research (ICPSR)*<sup>62</sup>. Die Signifikanz oder Einmaligkeit der Daten sollte ausschlaggebend für die Auswahl sein. Durch nicht reproduzierbare Rahmenbedingungen oder Datenerhebungen mit Bezug zu historischen oder politischen Ereignissen sind solche Daten besonders wertvoll und somit unbedingt archivierungswürdig.<sup>63</sup>

Spätestens vor der Übergabe der Daten ist zu prüfen, ob rechtliche Einschränkungen bedacht wurden.<sup>64</sup> In den Sozialwissenschaften unterliegen erhobene Daten oft dem Datenschutz, aus diesem Grund müssen Anonymisierungsverfahren angewendet werden. Bezüglich der rechtlichen Rahmenbedingungen sei an dieser Stelle auf Kapitel 3.1 verwiesen.

---

<sup>61</sup> Vgl. Eynden u.a. (2011), S. 1.

<sup>62</sup> Vgl. ICPSR (Inter-university Consortium for Political and Social Research) (2012), S. 4.

<sup>63</sup> Vgl. Weichselgartner u.a. (2011), S. 195 ff.

<sup>64</sup> Vgl. Spindler & Hillegeist (2011); Spindler & Hillegeist (2010).

## 2.3 Verfügbarkeit

In Kapitel 2.2 wurde erklärt, dass die Archivierung der Forschungsdaten die Voraussetzung für die langfristige Verfügbarkeit dieser darstellt. Damit die Daten nachnutzbar sind, müssen sie nicht nur unter vertrauenswürdigen Bedingungen archiviert werden, sondern auch auffindbar und zitierfähig sein. Dies wird zum einen durch die eindeutige, permanente Identifizierbarkeit der Daten erreicht, zum anderen durch entsprechende Zugangshilfen, damit die Daten recherchiert werden können.

Für die Identifizierbarkeit und Referenzierbarkeit können Forschungsdaten mit einem *Digital Object Identifier (DOI)* versehen werden. Dieser ermöglicht die Identifizierung des digitalen Objekts unabhängig von seinem Speicherort (im Gegensatz zur URL). Viele Fachverlage verwenden DOI bereits seit einigen Jahren zur Identifikation ihrer E-Books und E-Journals.

Für die Vergabe von DOI bei Forschungsdaten spricht die breite Verwendung und die Überwachung durch die *International DOI Foundation (IDF)*<sup>65</sup> im Hinblick auf die notwendige Persistenz. DOI werden bereits von einigen forschungsdatenhaltenden Einrichtungen genutzt. Ein DOI-Name (z.B. „10.4232/1.10028“) kann dabei jederzeit von der jeweiligen Registrierungsagentur selbständig aufgelöst werden, sodass die im DOI-Namen enthaltenen Adressinformationen nicht verloren gehen können.<sup>66</sup>

Eine bei der IDF akkreditierte DOI-Registrierungsagentur ist *DataCite*, die als 2009 gegründetes, internationales Konsortium das Ziel verfolgt, Forschenden Methoden zum zuverlässigen Lokalisieren, Identifizieren und Zitieren von Forschungsdaten anzubieten.<sup>67</sup> In Deutschland hat das GESIS – Leibniz-Institut für Sozialwissenschaften in Kooperation mit der ZBW (beide sind Vollmitglieder in *DataCite*<sup>68</sup>) den Service *da|ra* – Registrierungsagentur für Sozial- und Wirtschaftsdaten<sup>69</sup> aufgebaut. Im Jahr 2010 wurde zu Testzwecken mit der Registrierung von Forschungsdaten aus dem GESIS-

---

<sup>65</sup> Siehe <http://www.doi.org/> [Zugriff: 7.8.2013].

<sup>66</sup> Vgl. Hausstein (2012), S. 138 ff.

<sup>67</sup> Vgl. DataCite (2009), S. [1].

<sup>68</sup> Vgl. Hausstein (2012), S. 140.

<sup>69</sup> Auf der Startseite <http://www.da-ra.de/> [Zugriff: 7.8.2013] können DOI aufgelöst werden.

Datenarchiv begonnen.<sup>70</sup> Zurzeit befindet sich das von der DFG geförderte Projekt in der Etablierungsphase.<sup>71</sup>

Das GESIS-Datenarchiv stellt sozialwissenschaftliche Forschungsdaten aus nationalen und international vergleichbaren Umfragen bereit.<sup>72</sup> Über den Datenbestandskatalog<sup>73</sup> können bspw. Daten der *Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften* (ALLBUS), der *German Longitudinal Election Study* (GLES) oder des *European Social Survey* (ESS) gefunden werden.<sup>74</sup> Neben der (am OAIS-Referenzmodell orientierten<sup>75</sup>) Langzeitarchivierung der Daten offeriert es die „*Bereitstellung und Unterhaltung notwendiger Dienste, um Sekundäranalysen, Replikationen sowie vergleichende Untersuchungen über Raum und Zeit zu ermöglichen und den Zugang zu internationalen Forschungsdaten zu gewährleisten*“.<sup>76</sup> Die Serviceangebote werden vielfach nachgefragt. So gab es bspw. mehr als 30.000 Datenweitergaben im Jahr 2011, pro Jahr wurden bereits mehr als tausend Anfragen gezählt.<sup>77</sup>

Datenarchive stellen die langfristige Verfügbarkeit der Daten sicher, indem sie die Daten und dazugehörigen Dokumentationen von dem Datenproduzenten übernehmen. Für die Forschenden besteht der Vorteil darin, dass ihnen die Verantwortung für die dauerhafte Aufbewahrung abgenommen wird. Das Modell bietet sich daher an, wenn Forschende zeitlich begrenzt als Datenproduzenten agieren.<sup>78</sup> Ein Nachteil ist dabei, dass das Archiv im Allgemeinen nicht alle inhaltlichen Fragen detailliert beantworten kann, weil die für die Datenerstellung verantwortlichen Forschenden nach der Datenübergabe nicht mehr involviert sind. Analog zum GESIS-Datenarchiv für Deutschland sind entsprechende Infrastrukturen in anderen Ländern für die Verfügbarkeit von sozialwissenschaftlichen Forschungsdaten zuständig. So gibt es in Österreich das *Wiener Institut für sozialwissenschaftliche Dokumentation und Methodik* (WISDOM)<sup>79</sup>;

---

<sup>70</sup> Vgl. Hausstein (2012), S. 141.

<sup>71</sup> Vgl. <http://www.da-ra.de/de/ueber-uns/> [Zugriff: 11.8.2013].

<sup>72</sup> Vgl. Mauer (2012), S. 197.

<sup>73</sup> Siehe <http://www.gesis.org/unser-angebot/recherchieren/datenbestandskatalog/> [Zugriff: 11.8.2013].

<sup>74</sup> Vgl. <http://www.gesis.org/unser-angebot/recherchieren/> [Zugriff: 11.8.2013].

<sup>75</sup> Vgl. Mauer (2012), S. 201 ff.

<sup>76</sup> Jensen (2012), S. 71.

<sup>77</sup> Vgl. Mauer (2012), S. 207.

<sup>78</sup> Vgl. Huschka u.a. (2011), S. 42.

<sup>79</sup> Siehe <http://www.wisdom.at/datenarchiv/> [Zugriff: 7.8.2013].

in der Schweiz besteht mit *DARIS - Daten- und Forschungsinformationsservice*<sup>80</sup> die Möglichkeit zur Archivierung und Nachnutzbarkeit sozialwissenschaftlicher Forschungsdaten; das *UK Data Archive*<sup>81</sup> stellt die größte Sammlung an Forschungsdaten der Sozial- und Geisteswissenschaften im Vereinigten Königreich dar. Für den internationalen Zugang stellt das bereits erwähnte ICPSR in den USA mit seinem Datenarchiv<sup>82</sup> eine entsprechende Infrastruktur bereit.

Im Gegensatz zu Datenarchiven bieten Forschungsdatenzentren einen „wissenschaftlich unterstützenden inhaltlichen Service um die Daten herum, der nur erbringbar ist, weil die das FDZ betreibenden Datenproduzenten in der Regel die besten Experten im Umgang mit den eigenen Daten sind“.<sup>83</sup> Das Datenzugangsmodell eines Forschungsdatenzentrums bietet sich an, wenn aufgrund besonderer inhaltlicher Komplexität oder anspruchsvoller Datensatzstrukturen die Kommunikation zwischen Datenproduzent bzw. wissenschaftlichem Fachpersonal und Nutzer unerlässlich ist. Bei hoher Datenschutzsensibilität und dadurch bedingter Zugriffsbeschränkungen können über ein FDZ zugängliche Forschungsdaten unter der Kontrolle des Datenproduzenten verbleiben. Der Nachteil liegt jedoch in einem relativ hohen finanziellen Aufwand für die Datenproduzenten, der hauptsächlich in der Einführungsphase anfällt. Außerdem fehlt eine zentrale Anlaufstelle, welche die Zugangswege und Dokumentationen der Daten übersichtlich verwaltet.<sup>84</sup> Eine Übersicht des Rats für Sozial- und Wirtschaftsdaten (RatSWD) listet bspw. folgende akkreditierte Forschungsdatenzentren auf:

- Survey of Health, Ageing and Retirement in Europe (SHARE),
- Forschungsdatenzentrum des Statistischen Bundesamtes (FDZ-Bund),
- Forschungsdatenzentrum des Deutschen Zentrums für Altersfragen (FDZ-DZA),
- National Educational Panel Study (NEPS) und
- das am Deutschen Institut für Wirtschaftsforschung Berlin (DIW) produzierte Sozio-oekonomische Panel (FDZ-SOEP).<sup>85</sup>

<sup>80</sup> Siehe <http://www2.unil.ch/daris/?lang=de> [Zugriff: 7.8.2013].

<sup>81</sup> Siehe <http://www.data-archive.ac.uk/> [Zugriff: 7.8.2013].

<sup>82</sup> Siehe <http://www.icpsr.umich.edu> [Zugriff: 7.8.2013].

<sup>83</sup> Huschka u.a. (2011), S. 44.

<sup>84</sup> Vgl. Ebenda, S. 41 ff.

<sup>85</sup> Vgl. <http://ratswd.de/forschungsdaten/fdz> [Zugriff: 11.8.2013].

Der RatSWD<sup>86</sup> ist ein vom Bundesministerium für Bildung und Forschung 2004 eingerichtetes unabhängiges Gremium von empirisch arbeitenden Wissenschaftlerinnen und Wissenschaftlern. Er verfolgt das Ziel, die Forschungsdateninfrastruktur für die Sozial- und Wirtschaftswissenschaften zu verbessern, zu diesem Zweck fungiert er als Interessensvermittler zwischen den Datenproduzenten und Nutzern.<sup>87</sup> Mit diversen Veranstaltungen und zahlreichen Publikationen, wie z.B. die mehrmals jährlich erscheinenden RatSWD Working Papers, wird zur Information und zum Austausch beigetragen.

Um den Zugang zu Forschungsdaten auf globaler Ebene zu erleichtern, wurde der Aufbau eines weltweiten, interdisziplinären, Online-Verzeichnisses von Forschungsdaten-Repositorien im Rahmen des Projekts *re3data.org – Registry of Research Data Repositories*<sup>88</sup> begonnen. Mit diesem sollen Forschungsdaten-Repositorien vernetzt und standardisiert werden, auch um *data sharing* zu fördern. Die Eingabe der Daten erfolgt mithilfe eines eigens entwickelten XML-Schemas, das die Erschließung in über 50 Haupt- und Unterkriterien ermöglicht und dabei auch kontrollierte Vokabulare nutzt.<sup>89</sup>

Um die Anforderungen an das Forschungsdatenmanagement in den Sozialwissenschaften analysieren zu können, wurden in diesem Kapitel grundlegende Begrifflichkeiten erläutert und ein kurzer Überblick über den Ablauf empirischer Forschung gegeben. Die Darstellung des Datenlebenszyklus spielt für das FDM eine signifikante Rolle. Es wurde aufgezeigt, dass die langfristige und vertrauenswürdige Archivierung als Bestandteil des Datenlebenszyklus die Voraussetzung für die Nutzbarkeit von Forschungsdaten darstellt. Neben der Datenarchivierung sind jedoch geeignete Instrumente zur Auffindbarkeit von Forschungsdaten notwendig. Welche Tätigkeiten im FDM zur nachhaltigen Nutzbarkeit beitragen und welche Anforderungen vor allem im sozialwissenschaftlichen Bereich bestehen, wird im folgenden Kapitel untersucht.

---

<sup>86</sup> Siehe auch <http://ratswd.de/> [Zugriff: 7.8.2013].

<sup>87</sup> Vgl. RatSWD (Rat für Sozial- und Wirtschaftsdaten) (2011), S. 1.

<sup>88</sup> Siehe <http://www.re3data.org/> [Zugriff: 7.8.2013].

<sup>89</sup> Vgl. Scholze u.a. (2013), S. [2] ff.

### 3. Anforderungen an das Forschungsdatenmanagement

Um Forschungsdaten sowohl langfristig verfügbar als auch nutzbar machen zu können, bedarf es ihrer systematischen Erfassung, Aufbereitung, Dokumentation und Sicherung unter Beachtung rechtlicher und ethischer Rahmenbedingungen.<sup>90</sup> Dieser Prozess umfasst den gesamten Datenlebenszyklus und stellt die Wahrung der Authentizität und Integrität der Forschungsdaten sicher. Dabei spielen Transparenz und Nachvollziehbarkeit eine signifikante Rolle, denn der Zweck erfolgreichen Datenmanagements ist der, dass *„Daten ihre wissenschaftliche Aussagekraft behalten und [auch] für die [erneute] Auswertung zugänglich bleiben“*.<sup>91</sup>

Zusammengefasst lässt sich das Ziel von FDM nach Meyermann wie folgt beschreiben:

*„Es geht darum, die eigenen Arbeitsprozesse, die die Erzeugung von und den Umgang mit Forschungsdaten betreffen, möglichst effizient und zielorientiert zu organisieren und fortlaufend zu steuern.“*<sup>92</sup>

Im Hinblick auf die digitale Langzeitarchivierung bedeutet effizientes und zielorientiertes FDM eine Erleichterung bei der Übernahme (*Ingest*) der Daten in das Archiv: Wenn bereits alle notwendigen Metadaten vergeben sind und die erforderlichen Standards bei der Erfassung berücksichtigt wurden, resultiert dies mitunter aus einem erfolgreichen FDM. So spiegelt sich die Qualität des Datenmanagements in dem Aufwand bei der Übernahme zu Beginn des eigentlichen Archivierungsprozesses wider. An dieser Stelle müssen neben der Qualitätskontrolle je nach deren Ergebnis gegebenenfalls zusätzliche Ressourcen aufgewendet werden, um die Forschungsdaten beispielsweise in das richtige Format zu überführen oder fehlende Metadaten nachträglich zu ergänzen. Meist kann dies nur in Rücksprache mit dem Datenproduzenten – der Forscherin oder dem Forscher – erfolgen, was zusätzlich zeitlichen und personellen Aufwand hinsichtlich der (erneuten) Kontaktaufnahme bedeutet.

In englischsprachigen Quellen wird für ‚Forschungsdatenmanagement‘ neben (*research*) *data management* auch häufig der Begriff *data curation* verwendet. Die

---

<sup>90</sup> Vgl. Meyermann (2012), S. 4 ff.

<sup>91</sup> Büttner u.a. (2011b), S. 14.

<sup>92</sup> Meyermann (2012), S. 7.

Bedeutung ist jedoch dieselbe – das *Verwalten* bzw. die *Pflege* von Forschungsdaten. *Data curation* ist ein Kernprozess der *digital curation*, ebenso wie die langfristige Datenerhaltung (*data preservation*), die bereits im Zusammenhang mit dem in Kapitel 2.1 vorgestellten *DCC Curation Lifecycle Model* angesprochen wurde.

Nicht nur dieses Modell stellt die Tätigkeiten innerhalb des Datenlebenszyklus als fortwährenden Prozess dar.<sup>93</sup> Die Orientierung am Datenlebenszyklus schlägt sich einhellig in der für diese Arbeit recherchierten Literatur<sup>94</sup> nieder und nimmt den größten Stellenwert bezüglich der Anforderungen an das FDM ein – sowohl fachübergreifend als auch spezifisch für den sozialwissenschaftlichen Bereich.

In den folgenden Unterkapiteln werden Anforderungen an das erfolgreiche Management von Forschungsdaten näher erläutert. Damit die Anforderungsanalyse nicht eine rein literaturbasierte ist, werden anhand von drei Experteninterviews am WZB Informationen und Erfahrungswerte aus der Praxis einbezogen, sodass am Ende des Kapitels eine praxisnahe Zusammenfassung der Anforderungen an das Forschungsdatenmanagement im sozialwissenschaftlichen Bereich erfolgen kann.

---

<sup>93</sup> Vgl. Rümpel (2011), S. 27/28; DCC (Digital Curation Centre) (2013).

<sup>94</sup> Büttner u.a. (2011b); ICPSR (Inter-university Consortium for Political and Social Research) (2012); Jensen (2012); Jensen (2011); Jones (2011); Klump (2009); Ludwig & Enke (2013); Mauer (2012); Meyermann (2012); Quandt & Mauer (2012); Rümpel (2011); Treloar & Harboe-Ree (2008); Verhaar u.a. (2010); Weichselgartner (2011); Wiggins u.a. (2013).



### 3.1 Allgemeine Anforderungen

Die Archivierung und Nachnutzung von Forschungsdaten wird nicht nur von der DFG und der Allianz der deutschen Wissenschaftsorganisationen für Deutschland gefordert,<sup>95</sup> sondern es gibt auch internationale *Data Policies*<sup>96</sup> und Empfehlungen. So hat beispielsweise die OECD im Jahr 2007 Richtlinien herausgegeben, die „eine Steigerung des gesellschaftlichen Nutzens durch frei zugängliche Forschungsdaten [fordern]“.<sup>97</sup> Die dreizehn Grundsätze beinhalten u.a. *transparency, interoperability, quality, efficiency* und *sustainability*.<sup>98</sup>

Auf europäischer Ebene wurde durch die Europäische Kommission die Förderung des *freien Zugangs* zu Forschungsdaten betont. Dafür soll eine E-Infrastruktur zur *gemeinsamen Nutzung wissenschaftlicher Informationen* (als Ergebnis öffentlich finanzierter Forschungsprojekte), gefördert im Rahmen von „Horizont 2020“, einem Forschungs- und Innovationsförderprogramm der EU, geschaffen werden. Ziel ist, die Rentabilität der jährlichen Investitionen in dem Bereich Forschung und Entwicklung zu steigern und die *Innovationskapazität der EU* zu stärken.<sup>99</sup>

Dass im Rahmen des Forschungsdatenmanagements nicht nur der freie Zugang zu Forschungsdaten und deren Qualität, sondern auch Nachhaltigkeit und Transparenz einen großen Stellenwert einnehmen (müssen), ist bekannt. Im Gegensatz zu der Vielzahl an Richtlinien von der OECD wurden die Anforderungen an das FDM von der Arbeitsgruppe „Elektronisches Publizieren“ der Deutschen Initiative für Netzwerkinformation (DINI) in fünf Positionen thematisiert, in der ebenfalls die Qualität als eigenständiges Kriterium auftaucht:

- *Disziplinarität*: Fachspezifika sind zu berücksichtigen;
- *Organisation*: die Gestaltung eines Workflows *auf institutioneller Ebene* wird empfohlen;

---

<sup>95</sup> Siehe auch Kapitel 1.

<sup>96</sup> Vgl. Pampel & Bertelmann (2011), S. [49] - 61.

<sup>97</sup> Ebenda, S. 51.

<sup>98</sup> Vgl. OECD (Organisation for Economic Co-operation and development) (2007), S. 5.

<sup>99</sup> Vgl. Europäische Kommission (2012), S. 1 ff.

- *Zugang*: der Zugang muss transparent und effektiv für die Forschung gestaltet sein und unter rechtssicheren Bedingungen erfolgen, zudem müssen Standards Anwendung finden und Forschungsdaten korrekt nachgewiesen sein;
- *Qualität*: die Forschenden sind für die inhaltliche Qualität der von ihnen erstellten Daten verantwortlich, für die Erfüllung der formalen Qualitätskriterien sind die infrastrukturellen Serviceeinrichtungen zuständig;
- *Technik*: die Sicherung und langfristige Archivierung muss vertrauenswürdig sein, die Wahrung der Integrität und Authentizität der Daten hat oberste Priorität.<sup>100</sup>

Den unter den Punkten *Disziplinarität* und *Organisation* genannten Anforderungen wird in der vorliegenden Arbeit mit dem Fokus auf den Sozialwissenschaften und der Erstellung eines exemplarischen Workflows auf institutioneller Ebene (für das WZB) Rechnung getragen.

*Transparenz* und *Effektivität* werden im Zusammenhang mit dem Datenzugang genannt, wobei die *Qualität* auch hier indirekt in Form des korrekten Nachweises auftaucht. Die formale Qualität bezüglich Einheitlichkeit (Standards) und Vollständigkeit der Metadaten(felder) wird gewährleistet durch die infrastrukturellen Serviceanbieter. Die Qualität des Inhalts liegt in der Verantwortung der Datenersteller (Wissenschaftlerinnen und Wissenschaftler).

Hinsichtlich des *Zugangs* zu Forschungsdaten müssen ebenso rechtssichere Bedingungen vorliegen (z.B. unterschiedliche Zugriffsrechte je nach Nutzungsgruppe). Zudem beginnt die Klärung rechtlicher Belange bereits vor der digitalen Langzeitarchivierung, da datenschutzrechtliche Anforderungen Beachtung finden und Nutzungsrechte geklärt sein müssen. Eine Zusammenfassung<sup>101</sup> der rechtlichen Anforderungen an die Langzeitarchivierung von Forschungsdaten ist im *Handbuch Forschungsdatenmanagement* zu finden. Im Folgenden wird der Inhalt dieser Zusammenfassung kurz wiedergegeben.

---

<sup>100</sup> Vgl. Dallmeier-Tiessen u.a. (2009), S. 7 ff.

<sup>101</sup> Vgl. Spindler & Hillegeist (2011), S. [63] - 69.

Rechtliche Hindernisse bei der Datenweitergabe an Dritte sind in den meisten Fachdisziplinen eher weniger zu finden. Aufgrund der zumeist personenbezogenen Daten im Rahmen sozialwissenschaftlicher Erhebungen kommen datenschutzrechtliche Belange im Hinblick auf Langzeitarchivierung, und darauf beruhend das Ziel der Nutzbarkeit, hier besonders zum Tragen. Probleme für die Langzeitarchivierung lassen sich demzufolge vermeiden, wenn auf die Übergabe personenbezogener Daten verzichtet wird bzw. die Daten anonymisiert an das Archiv übertragen werden.<sup>102</sup>

Urheberrechtliche Streitfragen bei der Langzeitarchivierung können in der Regel ausgeschlossen werden, da wissenschaftliche Primärdaten nicht urheberrechtlich geschützt sind. Der Grund besteht darin, dass die Daten nicht aus einer für den Urheberschutz notwendigen *persönlichen geistigen Schöpfung* heraus entstanden sind.

Zu achten ist jedoch auf vertraglich geregelte Nutzungsrechte: Wenn in Arbeitsverträgen, z.B. mit der Hochschule, oder in der vertraglichen Vereinbarung mit dem Förderer (bei Drittmittelprojekten), zumindest einfache Nutzungsrechte eingeräumt sind, die eine Archivierung und Veröffentlichung der Daten erlauben, steht diesen aus rechtlicher Sicht nichts im Weg.<sup>103</sup>

Die Ansprüche an die digitale Langzeitarchivierung sind aus Sicht der OECD-Richtlinien unter dem Punkt *Nachhaltigkeit* fest verankert. Nach dem Positionspapier der DINI kann der Anspruch der Nachhaltigkeit durch die Wahrung der Datenintegrität und -authentizität als erfüllt betrachtet werden. Die Archivierung und Weitergabe von Forschungsdaten muss auf dem Einsatz vertrauenswürdiger Technik beruhen.

Policies und Richtlinien nationaler und internationaler Organisationen und wissenschaftlicher Vereinigungen beinhalten grundsätzliche Ansprüche an den Umgang mit Forschungsdaten im Allgemeinen. Auf diese Weise wurden zumindest Rahmenbedingungen auf wissenschaftspolitischer Ebene geschaffen.

Die spezifischeren Anforderungen an das Forschungsdatenmanagement definieren sich jedoch durch die Bedürfnisse der Forschenden und die erforderlichen Kompetenzen und Tätigkeiten der beteiligten Akteure. Diese gilt es zunächst zu identifizieren.

---

<sup>102</sup> Vgl. Ebenda, S. 69.

<sup>103</sup> Vgl. Ebenda, S. [63] ff.

### 3.1.1 Akteure im Forschungsdatenmanagement

Auf der einen Seite gibt es die Anbieter (Betreiber) der Infrastruktur, die sich unter anderem informationswissenschaftlicher, bibliothekarischer Methoden bedienen. Diese werden in englischsprachigen Quellen als *data manager* und *data librarian* bezeichnet. Zudem wird der *data scientist* genannt, der sowohl informationswissenschaftliche oder Informatik-Kenntnisse besitzt als auch mit fachspezifischem Wissen selbst aktiv an der Datenerstellung beteiligt ist.

Die andere Seite ist die der Nutzerinnen und Nutzer der Infrastruktur, also der Datenproduzenten (*data creators*) und derjenigen, die die Daten für ihre Forschung (sekundär) analysieren – die Datenerstellung und Datenanalyse kann durch dieselben Personen erfolgen.<sup>104</sup> Für die Datenanalyse und -aufbereitung wird hier – wie oben erwähnt – die Rolle des *data scientist* aufgeführt. Eine genauere Vorstellung der Begriffsbedeutungen ermöglichen die folgenden Definitionen von Swan & Brown:

*“Data creators or data authors [are] researchers with domain expertise who produce data. These people may have a high level of expertise in handling, manipulating and using data, gained through experience [...].*

*Data managers [are] people who are computer scientists, information technologists or information scientists and who take responsibility for computing facilities, storage, continuing access and preservation of data. [...]*

*Data librarians [are] originating from the library community, trained and specialising in the curation, preservation and archiving of data. Originally, the term data librarian seemed to be confined to librarians dealing with social science data, but the title now encompasses people with data skills in all disciplines. [...]*

*Data scientists [are] people who work where the research is carried out – or, in the case of data centre personnel, in close collaboration with the creators of the data [...]. In origin and training they may be domain experts, computer scientists or information technologists and their career development may have required them to assimilate skills*

---

<sup>104</sup> Vgl. Swan & Brown (2008), S. 8.

from a discipline from which they did not originate. So, a data scientist in systems biology may be a biologist by origin who has acquired very considerable computing skills and a data scientist whose background is in software engineering may have acquired a considerable degree of biological knowledge.”<sup>105</sup>

Für den englischsprachigen Raum wurde der Versuch einer visuellen Darstellung der Rollen und erforderlichen Kompetenzen von Donnelly (siehe Abbildung 5) unternommen, sie basiert auf den oben genannten Rollenbeschreibungen von Swan & Brown und wurde bereits mehrmals im Zusammenhang mit „Data Librarianship“ zitiert.<sup>106</sup>

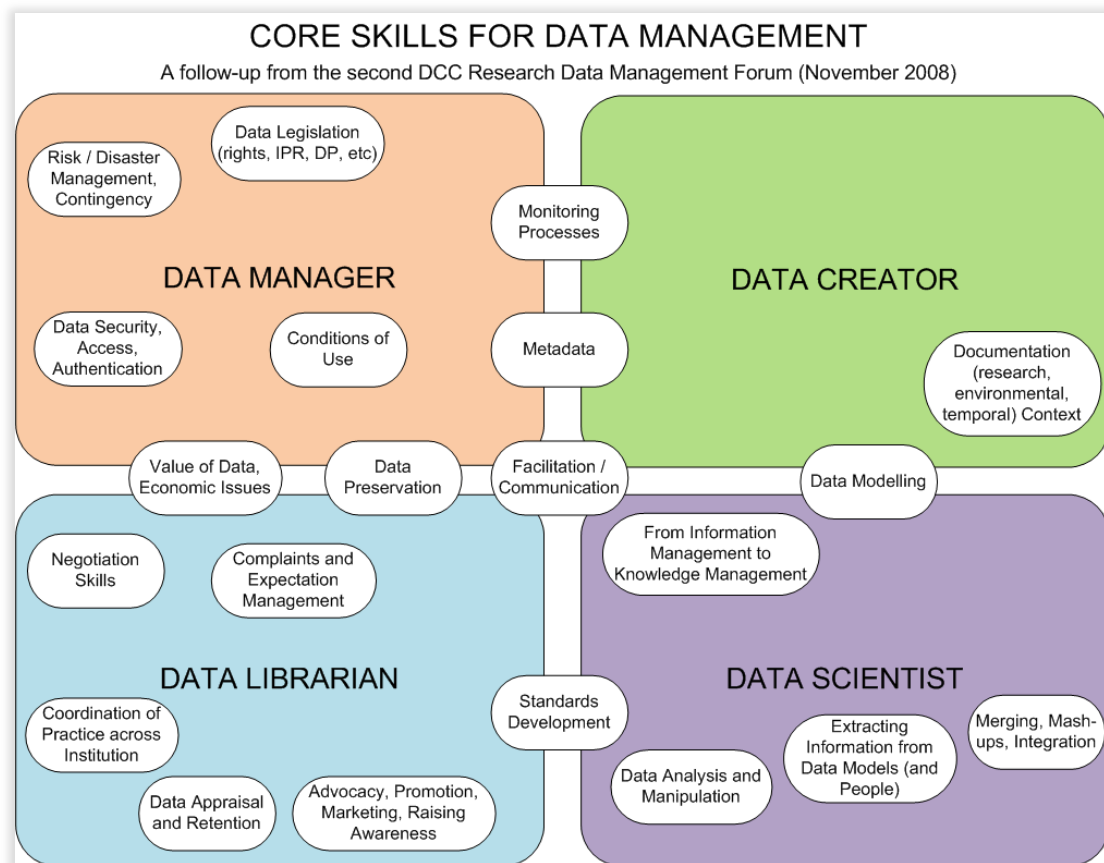


Abbildung 5: Kernkompetenzen im Forschungsdatenmanagement<sup>107</sup>

<sup>105</sup> Ebenda, S. 8. Reihenfolge der Auflistung abweichend vom Original.

<sup>106</sup> Vgl. Pampel u.a. (2009), Folie 20; Rümpel (2010), S. 38.

<sup>107</sup> Quelle: Donnelly (2008).

Eine exakte Übertragung der Rollen in den deutschen Kontext ist allein aufgrund der unterschiedlichen Ausprägungen der Berufsbezeichnungen nicht möglich. So tauchen im anglo-amerikanischen Raum diverse unterschiedliche Bezeichnungen für *data librarian* auf, z.B. *e-science librarian*, *data service librarian*, *research data management librarian*.<sup>108</sup> Auch wenn die Aufgaben des *data librarian* mit Tätigkeiten von Bibliothekarinnen und Bibliothekaren in Deutschland übereinstimmen, so gibt es keine offizielle berufliche Bezeichnung,<sup>109</sup> zumal die Aufgabengebiete der einzelnen Akteure oft nicht klar voneinander zu trennen sind, wie beispielsweise im Zusammenhang mit Metadaten. Zudem stellt sich bezüglich der Unterscheidung der Akteure die Frage, „*ob die Rolle des Data Scientist wirklich notwendig ist oder dies nicht von [data librarians] übernommen werden könnte*“.<sup>110</sup>

Im Verständnis der Autorin sollen *data scientists* - anders als nach obiger Definition von Swan & Brown - als diejenigen Forschenden angesehen werden, die bereits vorhandene Daten nachnutzen, indem sie sie analysieren und aufbereiten. Zudem sind sie *data creators*, also zugleich Urheber, Produzenten dieser Daten, da als Ergebnis von Analyse und Aufbereitung wiederum neue Forschungsdaten entstehen.<sup>111</sup> Die lt. Abbildung 5 auf Seite 37 dieser Rolle zugewiesenen Tätigkeiten und Kompetenzen stützen diese Sichtweise, bis auf einen Tätigkeitsbereich – *From Information Management to Knowledge Management* – dieser ist der einzige, der einen direkten informationswissenschaftlichen Bezug herstellt; die anderen Tätigkeiten, wie z.B. *Data Analysis and Manipulation*, spiegeln die bekannten Aufgaben von Forschenden innerhalb der Analyse-Phase des Forschungsablaufs wider.

*Data scientists* können demnach zu den (zukünftig) Nutzenden gezählt werden, wobei sie ihre Rolle mit der der *data creators* tauschen. Dies veranschaulicht Abbildung 6 auf Seite 39 mithilfe der beiden Pfeile auf der Nutzerseite. Sie folgt einer vereinfachten

---

<sup>108</sup> Vgl. Rümpel (2010), S. 42.

<sup>109</sup> Rümpel (2010) führt den deutschen Begriff ‚datenorientierter Bibliothekar‘ ein.

<sup>110</sup> Rümpel (2010), S. 70.

<sup>111</sup> Eine alternative Sichtweise wäre eine Rollenüberschneidung des *data scientist* und *data librarian*, sodass deren Aufgaben von ein und derselben Person wahrgenommen werden könnten. Voraussetzung dafür wäre eine entsprechende Ausbildung nach US-Amerikanischem Vorbild (Fachwissenschaftlicher Abschluss und zusätzlich Master in LIS). Ob sich in Deutschland ein eigenes Berufsfeld für den *data scientist* entwickelt und auch langfristig in der Praxis durchsetzen kann, nicht zuletzt im Hinblick auf personelle Ressourcen, bleibt abzuwarten.

Grafik Donnellys und verzichtet auf die Zuweisung von Kompetenzen, um zunächst die Akteure lediglich zu benennen und in den Kontext des FDM einzuordnen. Die Darstellung des Kreises soll in Anlehnung an das *DCC Curation Lifecycle Model* symbolisieren, dass sich die Akteure in einem fortwährenden Prozess befinden und jede Rolle ihren Anteil zum gesamten und erfolgreichen FDM beiträgt:

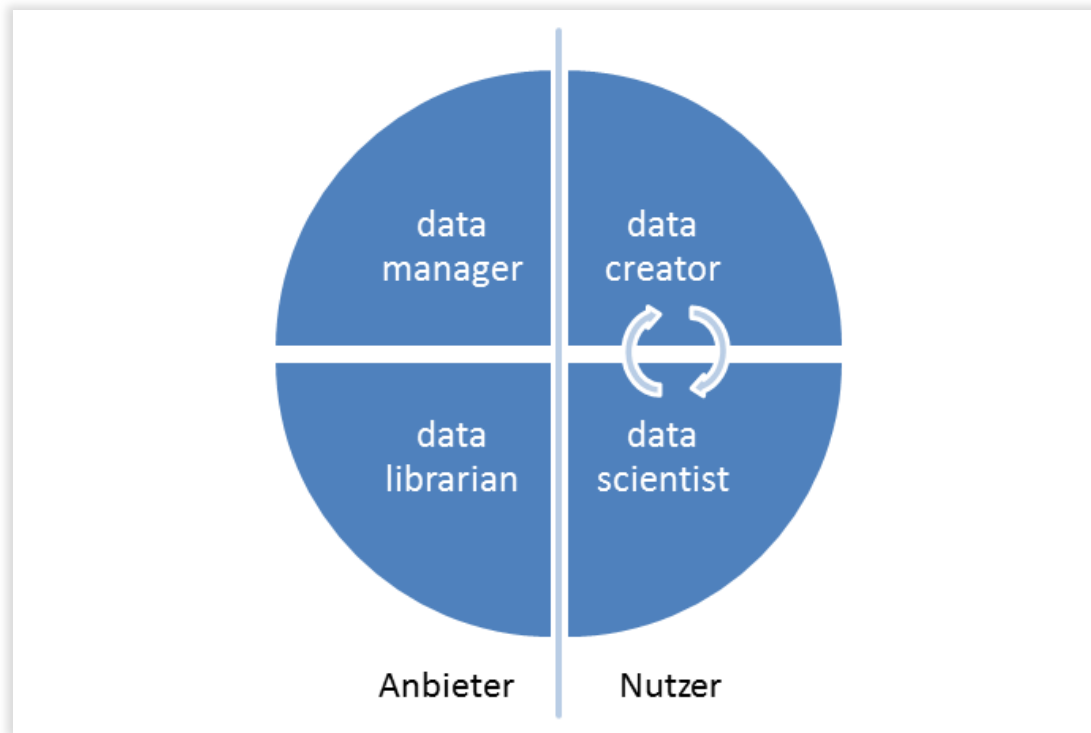


Abbildung 6: Akteure als Anbieter und Nutzer der Infrastruktur<sup>112</sup>

Nach Benennung und Einordnung der Akteure in den Kontext des FDM kann der Blick auf deren Kompetenzen und Tätigkeiten gerichtet werden.

Die Aufgaben der Infrastrukturanbieter, also die des *data manager* und *data librarian*, zusammengefasst unter dem Begriff *data curation*, zeichnen sich neben der Langzeitarchivierung insbesondere durch ihren Dienstleistungscharakter aus; hervorzuheben sei hier neben der Verwaltung die Unterstützungs- und Beratungstätigkeit.<sup>113</sup> Von ihren Kompetenzen profitieren einerseits die Datenproduzenten, die ihre Forschungsdaten vollständig dokumentieren, mit Metadaten anreichern und dabei die notwendigen Standards berücksichtigen müssen. Voraussetzung hierfür ist eine erfolgreiche

<sup>112</sup> Quelle: Eigene Darstellung nach Donnelly (2008).

<sup>113</sup> Vgl. Winkler-Nees (2011), S. 12.

Kommunikation zwischen den Akteuren und die kompetente fachliche Beratung von Seiten der Infrastrukturanbieter.

Andererseits nehmen die (zukünftigen) Nutzerinnen und Nutzer der Infrastruktur für die Datenanalyse – und gegebenenfalls für die Recherche nach den gewünschten Daten – vorab den Datenzugang und die Beratung der Anbieter in Anspruch. Hierbei nutzen sie neben den eigentlichen Forschungsdaten die von den Datenproduzenten erfassten und mitgelieferten Interpretationshilfen (Beschreibungen der Forschungsprojekte, Dokumentationen, Metadaten bspw. auf Studien- und Variablenebene).

*„Allgemein gilt, je diverser die Nutzungszwecke, je unbekannter die potentiellen Nutzer und je entfernter im Hinblick auf Raum und Zeit diese sind, desto unspezifischer und universeller sind die Anforderungen an das Datenmanagement.“<sup>114</sup>*

Abbildung 7 auf Seite 41 stellt die Beziehungen zwischen Anbieter- und Nutzerseite aufgrund ihrer Aufgaben dar: Die Anbieter gewährleisten den Datenzugang (Langzeitarchivierung) und vermitteln Methoden (Informationskompetenz); die Akteure auf Nutzerseite sind die (zukünftigen) Datennutzer und Datenproduzenten, welche das Infrastrukturangebot wahrnehmen. Auf Nutzerseite besteht eine indirekte Beziehung (gekennzeichnet durch die gestrichelte Linie des Pfeils) zwischen den beiden Akteuren, da es sich um dieselbe Person handeln kann. Wenn dies nicht der Fall ist und es sich um einen, zum Zeitpunkt der Erhebung unbekanntem, zukünftigen Datennutzer handelt, kann die Beziehung ebenso als indirekte angesehen werden. Der Grund hierfür ist, dass im Idealfall die beiden Akteure nicht miteinander in direktem Kontakt stehen. Eine direkte Kommunikation erfolgt lediglich mit dem Infrastrukturanbieter, auch ohne den persönlichen Kontakt, wenn sich der Bedarf auf den reinen Datenzugang beschränkt.

---

<sup>114</sup> Meyermann (2012), S. 7.



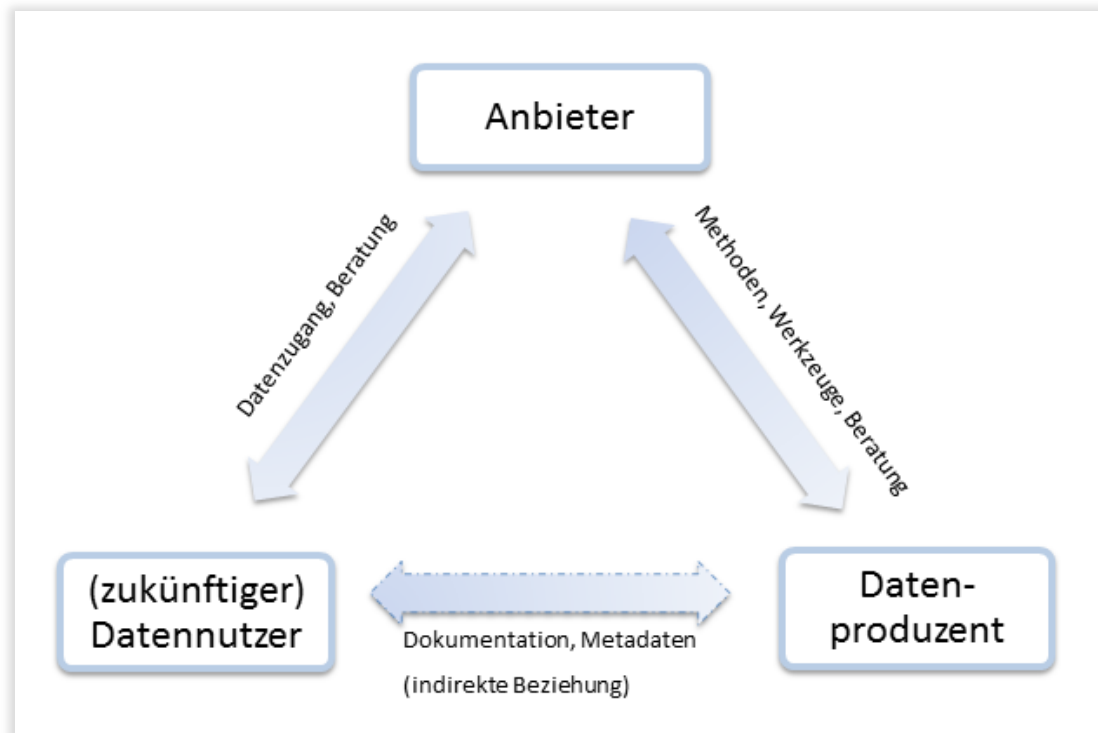


Abbildung 7: Beziehungen zwischen Akteuren im Forschungsdatenmanagement<sup>115</sup>

Die Aufgaben der Akteure zeigen, in welcher Beziehung diese zueinander stehen. Anhand dieser lassen sich allgemeine, übergeordnete Anforderungen formulieren, für die als erstes die Anforderungen der Forschenden untersucht werden.

#### 3.1.2 Anforderungen der Forschenden

Die in der Einleitung angesprochene vergleichende Literaturstudie von Feijen mit dem Titel *What researchers want*<sup>116</sup> liefert Erkenntnisse zu den Bedürfnissen von Wissenschaftlerinnen und Wissenschaftlern in Bezug auf *data sharing* und leitet aus diesen die entsprechenden Anforderungen an die infrastrukturellen Serviceeinrichtungen ab. Auf die Ergebnisse der Studie greift unter anderem Winkler-Nees zurück<sup>117</sup> und beschreibt *Anforderungen an wissenschaftliche Informationsinfrastrukturen*, die sich dem Dienstleistungscharakter der Infrastruktur-anbieter entsprechend als stark kundenorientiert darstellen.

<sup>115</sup> Quelle: Eigene Darstellung.

<sup>116</sup> Vgl. Feijen (2011).

<sup>117</sup> Vgl. Winkler-Nees (2011), S. 6 ff.

Aus Sicht der Datenproduzenten und -nutzer, die während des Forschungsprozesses ihre Daten erheben und für die spätere Weitergabe (z.B. an ein Datenarchiv) erfassen und aufbereiten, müssen die Arbeitsabläufe mit einem möglichst geringen Zeitaufwand verbunden sein. Die zu verwendenden Werkzeuge (z.B. Eingabemasken) müssen sich in den wissenschaftlichen Arbeitsalltag leicht integrieren und benutzen lassen: *„Researchers expect tools and services to support their day-to-day work within the research project, and long-term/public requirements must be subordinate to that interest.“*<sup>118</sup> Hier besteht die Anforderung an die Forschenden (*data creators* und *data scientists*) in der Akzeptanz neuer Methoden und Werkzeuge sowohl bei der Erstellung als auch bei der Aufbereitung und Analyse der Daten. Die Arbeitsweise der Forschenden bildet demnach den wichtigsten Bezugspunkt: *„Tools and services must be in tune with researchers’ workflows, which are often disciplinespecific (and sometimes even project-specific).“*<sup>119</sup>

Ein unkomplizierter Zugang zu den Daten und Hilfsmitteln ist gefragt; für die Nutzung der Werkzeuge müssen Unterstützungsangebote durch Personal, auch mittels persönlichen Kontakts, bereitgestellt werden<sup>120</sup>: *“Support must be local, hands-on, and available when needed.“*<sup>121</sup>

Die Vermittlung von Informationskompetenz ist eine Aufgabe wissenschaftlicher Bibliotheken. Im Bereich des FDM kann die Vermittlung von Datenkompetenz (*data literacy*)<sup>122</sup> ebenfalls durch diese erfolgen bzw. als integraler Bestandteil informationswissenschaftlicher Services angesehen werden.<sup>123</sup>

Das Fachwissen des Personals in den infrastrukturellen Serviceeinrichtungen nimmt dabei einen wichtigen Stellenwert ein. Jedoch geht es nicht darum, Einzelheiten und detailliertes informationswissenschaftliches Fachwissen den Forschenden „aufzudrängen“, sondern vielmehr, für das Thema zu sensibilisieren und für die Inanspruchnahme der Unterstützungsangebote zu werben.

---

<sup>118</sup> Feijen (2011), S. 4.

<sup>119</sup> Ebenda, S. 4.

<sup>120</sup> Vgl. Winkler-Nees (2011), S. 7 ff.

<sup>121</sup> Feijen (2011), S. 4.

<sup>122</sup> Vgl. Rümpel (2010), S. 45, 55, 69.

<sup>123</sup> Vgl. Winkler-Nees (2011), S. 16/17.

*„So kann von der wissenschaftlichen Seite nicht erwartet werden, dass eine allumfassende Kenntnis aller verfügbaren Standards vorhanden ist. Es ist jedoch erforderlich, dass nicht nur die Bereitschaft, hier entsprechende Expertise hinzuzuziehen vorhanden ist, sondern das Bewusstsein und die Verantwortung vorherrscht, dass nur unter deren Berücksichtigung ein professionelles Datenmanagement möglich ist.“<sup>124</sup>*

Allgemein gilt wie bei jedem Projekt: Akteure müssen eng kooperieren und miteinander kommunizieren. Sie müssen sich vor allem international ausrichten, um interdisziplinären Forschungsansprüchen Rechnung zu tragen.<sup>125</sup>

Die einfache Bedienung – *„[t]ools and services must be easy to use...“<sup>126</sup>* – ist für die Forschenden in ihrem Arbeitsalltag ein besonders wichtiges Kriterium. Für die langfristige Archivierung hingegen sehen sich die Forschenden nicht in der Verantwortung; hier sind die Kompetenzen der Infrastrukturanbieter gefragt. Außerdem gilt es im Hinblick auf Re- und Sekundäranalysen von vorneherein alle rechtlichen Unsicherheiten aus dem Weg zu räumen und ebenso den Schutz vor Datenmissbrauch zu gewährleisten. Auch die Angst vor einer zukünftigen Kommerzialisierung der Daten trägt nicht zur Bereitschaft der Datenweitergabe bei. An dieser Stelle sind Förderorganisationen und wissenschaftspolitische Entscheidungsträger in der Verantwortung, da es um die Schaffung entsprechender, einheitlicher und rechtlich bindender Rahmenbedingungen geht.<sup>127</sup>

Neben den angesprochenen Bedürfnissen von Wissenschaftlerinnen und Wissenschaftlern sind weitere Anforderungen in der nachfolgenden Auflistung aus der vergleichenden Literaturstudie von Feijen zusammengefasst:

- *“Tools and services must be in tune with researchers’ workflows, which are often disciplinespecific (and sometimes even project-specific).*
- *Researchers resist top-down and/or mandatory schemes.*

---

<sup>124</sup> Ebenda, S. 13.

<sup>125</sup> Vgl. Ebenda, S. 16/17.

<sup>126</sup> Feijen (2011), S. 4.

<sup>127</sup> Vgl. Winkler-Nees (2011), S. 16.

- *Researchers favour a “cafeteria” model in which they can pick and choose from a set of services.*
- *Tools and services must be easy to use.*
- *Researchers must be in control of what happens to their data, who has access to it, and under which conditions. Consequently, they want to be sure that whoever is dealing with their data (data centre, library, etc.) will respect their interests.*
- *Researchers expect tools and services to support their day-to-day work within the research project, and long-term/public requirements must be subordinate to that interest.*
- *The benefits of the support must be clearly visible – not in three years’ time, but now.*
- *Support must be local, hands-on, and available when needed.”<sup>128</sup>*

Obwohl nicht alle genannten Ergebnisse an dieser Stelle diskutiert werden, wird auf diese im weiteren Verlauf der Arbeit unter Berücksichtigung fachspezifischer Anforderungen und praktischer Erfahrungen näher eingegangen.

Die Anforderungen an das Forschungsdatenmanagement ergeben sich also in hohem Maße aus den Bedürfnissen der Forschenden: Werkzeuge und Methoden müssen sich leicht in den Forschungsalltag integrieren lassen und bestehende Arbeitsweisen der Forschenden dürfen nicht durch einen enorm hohen Zeitaufwand negativ beeinflusst werden. Die Anforderungen an die Anbieter bzw. Betreiber der Informationsinfrastruktur bestehen in der praktischen Umsetzung dieser Ansprüche, wobei die zwischenmenschliche Kommunikation auf Anbieter- und Nutzerseite unabdingbar ist, was die Sozialkompetenz als nicht zu vernachlässigenden Faktor kennzeichnet. Des Weiteren sind aus Anbietersicht für die Bewältigung der Aufgaben anerkannte Standards und einheitliche Vorgehensweisen unter Einbezug von Best Practices für das erfolgreiche und effiziente FDM maßgeblich.

---

<sup>128</sup> Feijen (2011), S. 4.

### 3.1.3 Forschungsdatenmanagement als Prozess

Wie zu Beginn des Kapitels 3 bereits betont, orientieren sich die Aufgaben im Datenmanagement an dem gesamten Datenlebenszyklus, wie in diversen Publikationen verdeutlicht wird. Im Folgenden werden ausgewählte Veröffentlichungen kurz vorgestellt.

Der im Rahmen des WissGrid-Projekts<sup>129</sup> entstandene *Leitfaden zum Forschungsdatenmanagement* von Ludwig & Enke stellt die einzelnen *Aufgaben im Lebenszyklus von Forschungsdaten* heraus, ohne dabei fachspezifische Besonderheiten im Forschungsprozess inhaltlich zu erläutern. Innerhalb des Datenlebenszyklus werden phasenübergreifende Aufgaben gesondert aufgeführt, wie beispielsweise organisatorische Aufgaben bezüglich der Datenarchivierung und Kostenplanung sowie die Bereitstellung geeigneter Identifikatoren.<sup>130</sup> Eine zur Nachnutzung erstellte Checkliste orientiert sich an den genannten Aufgaben, ist jedoch *als generisches Instrument* zu verstehen, d.h. sie sollte je nach Zielgruppe (z.B. Nutzungsgruppe der *Privaten Domäne* oder *Gruppen-domäne* nach dem Domänenmodell<sup>131</sup>) erweitert oder eingeschränkt werden.<sup>132</sup>

Analog zu diesem disziplinübergreifenden Leitfaden wurden von Jensen *Leitlinien zum Management von Forschungsdaten* publiziert, die sich ausschließlich auf sozialwissenschaftliche Umfragedaten, quantitative Daten, beziehen.<sup>133</sup>

Beide Publikationen richten sich nach den durch den Datenlebenszyklus vorgegebenen Phasen und Aufgaben, wonach bereits vor Projektbeginn mit der Planungsphase die Ansprüche an die langfristige Archivierung mit bedacht werden müssen. Dabei ist es essentiell, alle Überlegungen und Entscheidungen schriftlich festzuhalten. Hierfür gibt es ein geeignetes Instrument: den Datenmanagementplan (DMP).

Datenmanagementpläne beinhalten die systematische Beschreibung des Umgangs mit Forschungsdaten während und nach der Projektlaufzeit in Form eines strukturierten Konzepts. Signifikante Kernelemente eines DMP sind beispielsweise die Beschreibung

<sup>129</sup> Siehe [www.wissgrid.de/](http://www.wissgrid.de/) [Zugriff: 9.6.2013].

<sup>130</sup> Vgl. Ludwig & Enke (2013), S. 14 ff.

<sup>131</sup> Vgl. Schäfer u.a. (2013), S. 48.

<sup>132</sup> Vgl. Ludwig & Enke (2013), S. 17/18.

<sup>133</sup> Vgl. Jensen (2012), S. 9.

zu den im Projekt genutzten, bearbeiteten oder neu erzeugten Forschungsdaten; Metadaten zum Studien- und Methodendesign und natürlich der Daten an sich; Informationen zur Datensicherheit, Vertraulichkeit, Qualitätskontrolle, zu einmaligen und laufenden Kosten und zu dem Nutzungs- und Rechtemanagement.<sup>134</sup>

Das Erstellen eines DMP erfolgt idealerweise durch die Forschenden in Kooperation und Rücksprache mit den Anbietern bzw. Betreibern der Infrastruktur (*data manager*, *data librarian*, Personal des entsprechenden Datenarchivs oder Forschungsdatenzentrums). Das Wissen, z.B. über geeignete Metadatenstandards und Datenformate bis zu technischen Lösungen von (kurzfristiger) Sicherung und langfristiger Archivierung, ist bei diesen Akteuren vorhanden. Auch hier besteht die Anforderung darin, den Arbeits- und Zeitaufwand der Forschenden bezüglich solcher Fragen so gering wie möglich zu halten und dabei bestmögliche Unterstützung zu geben.

Hilfen zum Erstellen eines Datenmanagementplans sind bereits in Form von diversen Vorlagen (*templates*)<sup>135</sup> und Informationsangeboten vorhanden, wie Beispiele aus dem englischsprachigen Raum zeigen: Das *Massachusetts Institute of Technology* (MIT) stellt online umfangreiche Informationen zum Thema Datenmanagement zur Verfügung, die von Gründen für das Datenmanagement über Checklisten und DMP bis hin zu Zitation von Forschungsdaten und rechtlichen Fragestellungen reichen. Zusätzlich werden spezielle Informationen für den Umgang mit sozialwissenschaftlichen Daten geboten.<sup>136</sup>

Die *Columbia University* bietet ebenfalls neben Informationen zum Datenmanagement disziplinspezifische Vorlagen im doc-Format an, darunter auch zu dem *Social, Behavioral and Economic Sciences Directorate* (SBE).<sup>137</sup>

---

<sup>134</sup> Vgl. Jensen (2011), S. [71] ff.

<sup>135</sup> Eine ausführliche Darstellung der einzelnen Vorlageninhalte soll an dieser Stelle nicht erfolgen, der praktische Bezug zu den Aufgaben und Verantwortlichkeiten im Rahmen des Datenmanagements wird in Kapitel 4 behandelt.

<sup>136</sup> Vgl. <http://libraries.mit.edu/guides/subjects/data-management/index.html> [Zugriff: 11.8.2013].

<sup>137</sup> Vgl. <http://scholcomm.columbia.edu/data-management/data-management-plan-templates/> [Zugriff: 11.8.2013].

Das DCC und Joint Information Systems Committee (JISC) haben eine Anleitung für das Erstellen eines DMP herausgegeben, die viele Fragen aus Sicht der Förderer beleuchtet und betont, dass Datenmanagement ein integraler Bestandteil des gesamten Forschungsprozesses ist: „*Reviewers will look for evidence that data management is embedded throughout your proposal and forms an integral part of your research process.*“<sup>138</sup>

Ebenfalls vom DCC wurde eine Checkliste veröffentlicht, die Forschenden bei der Antragserstellung helfen soll, alle erforderlichen Inhalte zu bedenken und Verantwortlichkeiten zu benennen, wobei auch hier auf das *DCC Curation Lifecycle Model* verwiesen wird. Vorab wird betont, dass die Checkliste jedoch keinen Schritt-für-Schritt-Workflow zum Datenmanagement darstellt, die Beschäftigung mit der Checkliste aber inhaltlich mit der Erstellung eines solchen einhergeht.<sup>139</sup>

Auch wenn die genannten Vorlagen teilweise für den universitären oder auch fachübergreifenden Kontext entwickelt worden sind, lassen sie sich in ihren Grundzügen auf fachspezifische, institutionelle Bedürfnisse übertragen und können an die jeweiligen Projektbedingungen angepasst werden. So gibt es für das FDM im sozialwissenschaftlichen Bereich Publikationen, die ebenfalls als Rahmenkonzept angesehen und angepasst werden können.<sup>140</sup>

Die Anforderungen an das FDM im sozialwissenschaftlichen Kontext werden im folgenden Kapitel näher erläutert.

---

<sup>138</sup> Jones (2011), S. 8.

<sup>139</sup> Vgl. Donnelly & Jones (2009), S. 1/2.

<sup>140</sup> Vgl. Jensen (2011), S. [71].

## 3.2 Besondere Anforderungen in den Sozialwissenschaften

Die in Kapitel 3.1 genannten Rahmenbedingungen und Anforderungen an das FDM gelten im Allgemeinen für alle Fachdisziplinen, die Bedürfnisse der Wissenschaftlerinnen und Wissenschaftler sind im Großen und Ganzen dieselben. Dennoch gibt es einige fachspezifische Besonderheiten, die es zu berücksichtigen gilt.<sup>141</sup> Die Beachtung disziplinärer Ansprüche wird im DINI-Positionspapier ebenfalls betont.<sup>142</sup>

Zunächst tragen neben den disziplinübergreifenden Empfehlungen und Richtlinien, die bereits in Kapitel 3.1 erwähnt wurden, sowohl interdisziplinäre als auch institutionelle Policies zur Schaffung entsprechender Rahmenbedingungen bei. Als Dachorganisation der sozialwissenschaftlichen Datenarchive in Europa hat das CESSDA (Council of European Social Science Data Archives)<sup>143</sup> ein Übereinkommen zur Zusammenarbeit der Datenarchive geschlossen. Darin heißt es: *“All CESSDA archives are committed to open access to data and operate similar dissemination policies.”*<sup>144</sup> Dies zeigt, dass *“Policies nicht nur auf Wissenschaftler [und Wissenschaftlerinnen], sondern auch auf disziplinäre Infrastruktureinrichtungen wirken.”*<sup>145</sup> Somit wird der Forderung nach disziplinspezifischen Anpassungen nachgekommen.<sup>146</sup>

Auf institutioneller Ebene orientieren sich die jeweiligen Einrichtungen beispielsweise an den Vorgaben der Förderorganisation. Es werden aber auch Impulse aus anderer Richtung geliefert: Die Leibniz-Gemeinschaft, zu der auch das WZB und GESIS gehören, ist zwar keine Förderorganisation, hat jedoch die DFG-Empfehlung zur zehnjährigen Datensicherung für ihre *Policy* übernommen; diese Regeln *„sollen den Instituten als Rahmen für die Ausgestaltung und Umsetzung eigener Überlegungen gemäß der Strukturen, der Rechtsform und der Aufgaben in den Instituten dienen“*.<sup>147</sup> Auch die Max-Planck-Gesellschaft orientiert sich in ihren *Regeln zur Sicherung guter*

---

<sup>141</sup> Vgl. Feijen (2011), S. 4.

<sup>142</sup> Vgl. Dallmeier-Tiessen u.a. (2009), S. 7.

<sup>143</sup> Siehe auch Kapitel 2.3.

<sup>144</sup> Vgl. <http://www.cessda.org/sharing/dissemination/> [Zugriff: 11.8.2013].

<sup>145</sup> Pampel & Bertelmann (2011), S. 53.

<sup>146</sup> Vgl. Ebenda, S. 53.

<sup>147</sup> Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (1999), S. [1] ff.



wissenschaftlicher Praxis an den Formulierungen der DFG und fordert die zentrale oder institutionelle Sicherung von Primärdaten für zehn Jahre.<sup>148</sup>

Obwohl diese allgemeinen und auch disziplinspezifischen Richtlinien den notwendigen wissenschaftspolitischen Rahmen schaffen, fehlen jedoch weitere, konkrete Vorgaben, die sich beispielsweise auf den Einsatz personeller Ressourcen beziehen. Außerdem sollten die Positionen im Hinblick auf Fehlverhalten im Umgang mit Forschungsdaten und *data sharing* in den jeweiligen Forschungsdisziplinen klar definiert sein, auch auf institutioneller Ebene.

Für das FDM im sozialwissenschaftlichen Bereich spielen rechtliche Belange vor allem bezüglich des Datenschutzes aufgrund der Natur quantitativer sozialwissenschaftlicher Daten eine besondere Rolle. Personenbezogene Daten, sensible Daten sind in der empirischen Sozialforschung keine Seltenheit und Forschende sind sich der Brisanz solcher Daten bewusst. Datenschutzrechtliche Vorgaben werden oft als Hindernis für *data sharing* gesehen, sollten dennoch kein Grund sein, auf die Weitergabe von Forschungsdaten zu verzichten.

*“Generell gilt, dass der Daten- und Persönlichkeitsschutz durch die Anwendung entsprechender Vorkehrungen strikt und umfassend entlang der Gesetze eingehalten werden muss, dies jedoch niemals ein Argument dafür sein kann Daten nicht zugänglich zu machen.”<sup>149</sup>*

Anonymisierungsmethoden bei der Datenaufbereitung und vor der Übergabe an ein Datenarchiv sowie unterschiedliche Nutzungsmodelle mit entsprechender Rechteverwaltung gewährleisten die weitere Datennutzung unter Beachtung gesetzlicher Regelungen.

Da bereits zu Beginn des Forschungsprojekts (mit der Konzeption) und des Datenlebenszyklus die spätere Archivierung und Weitergabe der Daten mitbedacht werden müssen, sollten datenschutzrechtliche Belange und zukünftige Nutzungsbedingungen ebenso frühzeitig im Projekt geklärt werden. Solche Überlegungen gehen mit der Erstellung eines Datenmanagementplans einher, der unter anderem die von der

---

<sup>148</sup> Vgl. MPG (Max-Planck-Gesellschaft) (2009), S. 4.

<sup>149</sup> Huschka u.a. (2011), S. 40.

Förderorganisation geforderten nachgewiesenen Maßnahmen zum Datenmanagement und zur Datensicherung enthält. Wie diesbezüglich in Kapitel 3.1 beschrieben, müssen die infrastrukturellen Serviceeinrichtungen an dieser Stelle entsprechende Dienstleistungen anbieten.

Um die Wissenschaftlerinnen und Wissenschaftlern bei der Erstellung eines Datenmanagementplans zu unterstützen und kompetente Beratung zu geben, ist ein Aspekt von entscheidender Bedeutung: Kommunikation. Überall dort, wo Projekte (ob wissenschaftlich oder industriell) geplant und durchgeführt werden, ist eine erfolgreiche zwischenmenschliche Kommunikation essentiell für den Erfolg des gesamten Projekts.

Auch im FDM trägt Kommunikation zwischen den beteiligten Akteuren maßgeblich zur Effizienz und Effektivität bei. Kompetente Beratung gelingt jedoch nur dann, wenn alle Akteure dasselbe Verständnis über fachspezifische Begrifflichkeiten besitzen und sich alle sicher sind, auch wirklich dasselbe zu meinen. Damit die Anbieter der Infrastrukturen die Bedürfnisse der Forschenden verstehen und nachvollziehen können, müssen ein gewisses Maß an Grundkenntnissen zu Forschungsabläufen und Begrifflichkeiten vorhanden sein. Diese Voraussetzung spiegelt sich auch im Aufbau der vorliegenden Arbeit wider: Bevor die Anforderungen an das FDM im Hinblick auf entsprechende Workflow-Aspekte in den Sozialwissenschaften analysiert werden können, gibt das zweite Kapitel zu Beginn einen kurzen Überblick über den Forschungsablauf und liefert Begriffsdefinitionen zu quantitativen sozialwissenschaftlichen Daten. Des Weiteren sind Kenntnisse über die vorhandenen Infrastrukturen unabdingbar, wie sie für die Sozialwissenschaften in den Kapiteln 2.2 und 2.3 aufgezeigt werden.

Das Wissen über sozialwissenschaftliche Standards bei der Datenerhebung, -analyse und -aufbereitung stellt neben informationswissenschaftlichen Kompetenzen eine wichtige Anforderung an die infrastrukturellen Serviceeinrichtungen dar. Es dient sowohl der Erstellung von Datenmanagementplänen als auch den Tätigkeiten in den weiteren Phasen des Datenlebenszyklus.<sup>150</sup> Um auf die Informationsbedürfnisse der Forschenden eingehen zu können, müssen Studienabläufe und formale Standards

---

<sup>150</sup> Da dieses Kapitel die Anforderungen an das FDM analysiert, wird auf die ausführliche Beschreibung aller Aufgaben in den einzelnen Phasen des Datenlebenszyklus verzichtet; in Kapitel 4 sind die wichtigsten Tätigkeiten anhand der Workflow-Schritte zusammenfassend dargestellt.

bekannt sein. Die Anforderungen an das FDM in den Sozialwissenschaften richten sich demzufolge auch nach den Inhalten und Fragestellungen der jeweiligen Forschungsprojekte. Im Folgenden werden exemplarisch einige Standards für das Management quantitativer sozialwissenschaftlicher Daten aufgezeigt.

Die in Kapitel 3.1.3 angesprochenen *Leitlinien zum Management von Forschungsdaten* beinhalten neben einer Checkliste zum Forschungsdatenmanagement<sup>151</sup> alle notwendigen Aufgaben und Tätigkeiten im gesamten Datenlebenszyklus. Neben der Beschreibung der Datenaufbereitung und Metadatenvergabe der Forschungsdaten werden ebenfalls sozialwissenschaftliche Standards erläutert: Internationale Standards sind beispielsweise die Klassifikationssysteme ISCED für Schulsysteme und Schultypen, ISCO von der International Labour Organisation als Klassifikation für Berufsgruppen sowie die Standards ISO 639 zur Kodierung von Sprachen und ISO 3166 für bestehende Staaten. Diese Klassifikationen und Normen dienen der Vergleichbarkeit von Forschungsdaten, vor allem im Hinblick auf internationale Anforderungen.<sup>152</sup>

Neben der Verwendung der oben exemplarisch genannten sozialwissenschaftlichen Standards müssen ebenso Metadaten über den gesamten Forschungsprozess vergeben werden: Das Studiendesign muss dokumentiert werden, dazu gehören Informationen zur Fragebogenentwicklung (Metadaten zu den Messinstrumenten), zur Datenerhebung (Informationen zu Pretests, Übersetzungsprozesse bei der Fragebogenerstellung, Kommentare der Forscher als Metainformationen) und auch sogenannte Paradata, die den Datenerhebungsprozess an sich beschreiben und bei computergestützten Befragungen automatisch erzeugt werden.<sup>153</sup>

Nach der Datenerhebung erfolgt die -auswertung, die für eine spätere Nachnutzung ebenfalls beschrieben werden muss, da „[d]ie Art und Weise, wie Daten analysiert werden und welche Analysemethoden dabei eingesetzt werden, [...] notwendige Informationen über den Analyseprozess dar[stellen]“.<sup>154</sup> Hierbei ist die Harmonisierung von Daten, d.h. „die Transformation von Daten aus unterschiedlichen Quellen in

---

<sup>151</sup> Vgl. Jensen (2012), S. 19/20.

<sup>152</sup> Vgl. Ebenda, S. 45 ff.

<sup>153</sup> Vgl. Jensen u.a. (2011), S. [83] ff.

<sup>154</sup> Ebenda, S. 85.

*standardisierte Maße...*<sup>155</sup> zu nennen, beispielsweise werden Daten bezüglich der Höhe von Einkommen harmonisiert, um eine vergleichende Analyse überhaupt zu ermöglichen, da Einkommen in verschiedenen Ländern unterschiedlich hohe Ausprägungen besitzt.<sup>156</sup>

Für die Dokumentation der Forschungsdaten ist es notwendig, Metadaten auf unterschiedlichen Ebenen zu vergeben – zum einen auf Studienebene, zum anderen auf Variablenebene. Als sozialwissenschaftlicher Dokumentationsstandard hat sich der Standard der 2003 gegründeten DDI-Alliance (*Data Documentation Initiative*) etabliert.

Der DDI-Standard besteht einerseits aus dem *DDI-Codebook (DDI-C)*, das eine Metadatenpezifikation für Mikrodaten, Aggregatdaten und geografische Informationen darstellt; zum anderen aus dem modular aufgebauten *DDI-Lifecycle (DDI-L)*, mit dem Ziel, den gesamten Datenlebenszyklus vollständig mit Metadaten (im XML-Format) zu beschreiben. Mit dem *Codebook* und dem *Lifecycle* stehen „zwei unterschiedliche Entwicklungslinien von Metadatenstrukturen“<sup>157</sup> zur Verfügung.<sup>158</sup>

Die Spezifikation DDI-C ist der ursprünglich entwickelte Standard (bis zur aktuellen Version 2.5) zur Beschreibung eines Datensatzes auf Studien- und Variablenebene.<sup>159</sup> Abbildung 8 auf Seite 53 zeigt eine Auswahl der Metadatenelemente der Version 2.1. Die DDI-Alliance stellt diverse Software-Tools<sup>160</sup> zum Datenmanagement bereit, wie den *DDI Validator* zur Validierung von DDI-C und DDI-L XML-Code, aber auch Beispiel-Codebooks in XML, PDF und HTML.<sup>161</sup> Auffallend ist der große Umfang der Codebooks: Die Variablenbeschreibung stellt jeweils mit mehreren hundert PDF-Seiten den Hauptanteil dar.<sup>162</sup>

---

<sup>155</sup> Ebenda, S. 85.

<sup>156</sup> Vgl. Ebenda, S. 84/85.

<sup>157</sup> Jensen (2012), S. 48.

<sup>158</sup> Ziel der vorliegenden Arbeit ist es nicht, eine vergleichende Analyse bestehender Metadatenstandards zu präsentieren. Weitere Informationen zur DDI-Spezifikation sind in den nachfolgend zitierten Quellen zu finden sowie allgemein über die Webseiten der DDI-Alliance: <http://www.ddialliance.org/> [Zugriff: 11.8.2013].

<sup>159</sup> Vgl. <http://www.ddialliance.org/Specification/> [Zugriff: 11.8.2013].

<sup>160</sup> Vgl. <http://www.ddialliance.org/resources/tools> [Zugriff: 11.8.2013].

<sup>161</sup> Vgl. <http://www.ddialliance.org/resources/tools/codebooks> [Zugriff: 11.8.2013].

<sup>162</sup> Vgl. Beispiele des ICPSR zu *Retirement History Longitudinal Survey Series*: <http://www.ddialliance.org/resources/tools/codebooks/retirement> [Zugriff: 11.8.2013].

Gruppe	Elemente (Auszug)
Document description	<ul style="list-style-type: none"> <li>▪ Titel, Autoren, Publikation, Zitation des DDI Dokuments</li> </ul>
Study description	<ul style="list-style-type: none"> <li>▪ Primärforscher, Institutionen, Projektinformationen</li> <li>▪ Typ der Studie, Art der Daten und Abstract ihrer Inhalte</li> <li>▪ Zeitliche und geographische Angaben zur Erhebung</li> <li>▪ Methoden, Grundgesamtheit, Analyseeinheit, Stichprobe</li> <li>▪ Literatur des Projektes zur Studie (z. B. Forschungsbericht; Analysen)</li> <li>▪ Versionierung, bibliographische Zitation der Datendatei</li> <li>▪ z. B. Verweise auf digitalisierte Fragebogen, Methodenberichte o. ä.</li> </ul>
Variable description	<ul style="list-style-type: none"> <li>▪ Variablenamen, Typ und Labels</li> <li>▪ Code, Typ und Häufigkeiten</li> <li>▪ Hinweise zur Codierung oder Berechnung von Werten (Notes)</li> <li>▪ Fragetexte und Antwortkategorien</li> <li>▪ Intervieweranweisungen, Filter</li> </ul>
Data files description	<ul style="list-style-type: none"> <li>▪ Anzahl der Variablen, Fälle</li> <li>▪ Namen, Formate und Versionen</li> </ul>

Abbildung 8: Ausgewählte Metadatenfelder in DDI-C (Version 2.1)<sup>163</sup>

Im Gegensatz zu *DDI-C* gewährleistet die *Lifecycle*-Spezifikation die Beschreibung der gesamten Datenlebenszyklusstruktur. Wenn die Dokumentation die verschiedenen Phasen des Forschungsprozesses und somit des Datenlebenszyklus berücksichtigt, entspricht dies folgerichtig dem Anspruch eines am Lebenszyklus orientierten FDM.

Der praktische Vorteil des modularen Aufbaus von *DDI-Lifecycle* besteht in dem Austausch der Metadaten zwischen Kooperationspartnern. Insbesondere im Hinblick auf die spätere Langzeitarchivierung und Nachnutzbarkeit ist dies sinnvoll, da bereits vorhandene Metadaten in einem standardisierten Format die Datenübergabe an ein Archiv wesentlich erleichtert.

Der modulare Aufbau spiegelt sich in der Dokumentationsstruktur wider (siehe Abbildung 9 auf Seite 54): So ist z.B. das Modul „Data Collection“ der zweiten Phase des Lebenszyklus gleichgesetzt (siehe Abbildung 10 auf Seite 54). Zudem ist der Lebenszyklus als fortwährender Prozess dargestellt, bei dem die Archivierung für die langfristige Nutzbarkeit der Daten notwendig ist.

<sup>163</sup> Quelle: Jensen (2012), S. 49.

Modul	Dokumentationselemente (Metadaten)
Study Unit	<ul style="list-style-type: none"> <li>▪ Umfangreiche Metadaten zur Identifikation einer Studie (Titel, Forscher), zur bibliographischen Zitation und zur Nutzung dauerhafter Identifikatoren z. B. in Form einer DOI (Digital Object Identifier)</li> <li>▪ Abstract, Zweck, Antrag und Förderung einer Einzelstudie</li> <li>▪ Thematische, zeitliche und räumliche Abdeckung der Studie</li> <li>▪ Untersuchungseinheit(en), Messkonzepte der Daten</li> <li>▪ Publikationen und Verweise auf digitalisierte Materialien: Fragebogen, Methodenberichte o. ä.</li> </ul>
Data Collection	<ul style="list-style-type: none"> <li>▪ Methodologie; Erhebungszeiträume und -ereignisse</li> <li>▪ Mehrsprachige Fragetexte und Antwortdomänen (Text, Numerisch, Codeschema etc.)</li> <li>▪ Messinstrument, Ablauf- und Kontrollstrukturen, Intervieweranweisungen, Codierungsanweisungen</li> </ul>
Logical Product	<ul style="list-style-type: none"> <li>▪ Logische Beschreibung der Forschungsdaten einer Studie.</li> <li>▪ Variablenschema, Variablendefinition, Kategorien-, Codelisten</li> <li>▪ NCubes; Gruppierung von NCubes und Variablen zur Beschreibung aggregierter Daten</li> <li>▪ Logische Beziehungen der Daten untereinander, z. B. durch Kennzeichnung von mehreren Datensätzen (Records) je Befragtem</li> </ul>

Abbildung 9: Ausgewählte Elemente in DDI-L (Version 3.1)<sup>164</sup>

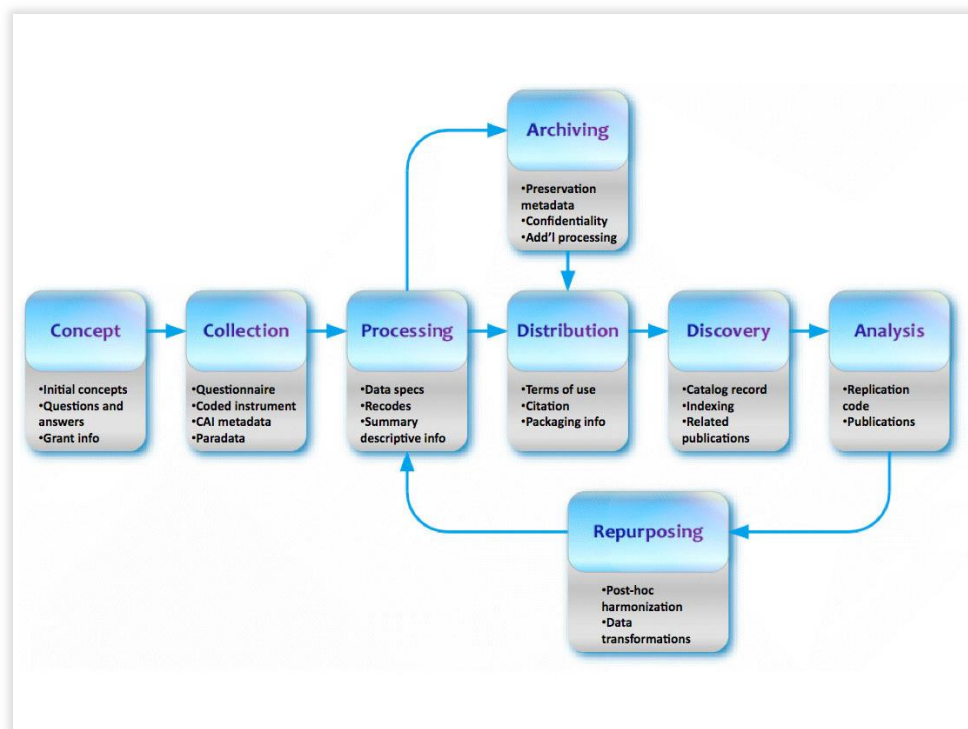


Abbildung 10: Datenlebenszyklus nach DDI-L<sup>165</sup>

<sup>164</sup> Quelle: Ebenda, S. 50.

<sup>165</sup> Quelle: Ebenda, S. 48, nach DDI-Alliance.

Laut DDI-L (Phase „Data Collection“) fließen ebenfalls Informationen über die Instrumente der Datenerhebung in die Dokumentation mit ein; in der empirischen Sozialforschung sind dies in der Regel Befragungen und Beobachtungen, wobei der Beobachtung nicht nur quantitative Analysen folgen. Innerhalb des Forschungsprozesses entstehen unterschiedliche Arten von Daten: Rohdaten, edierte Daten in Vorbereitung für die Veröffentlichung, anonymisierte Daten. Welche Daten wie und in welcher Phase entstehen bzw. verarbeitet werden, muss ebenfalls Bestandteil der Metadaten sein. Im Hinblick auf personenbezogene Daten gelten die besonderen Anforderungen an die Anonymisierung, um datenschutzrechtliche Bestimmungen nicht zu verletzen.<sup>166</sup>

Quantitative sozialwissenschaftliche Daten werden in der Regel nicht maschinell produziert (mit Ausnahme von prozessproduzierten Daten). Im Gegensatz zu Daten aus geologischen Bohrungen oder meteorologischen Messungen, werden sie intellektuell produziert und erfordern somit eine nähere, ausführlichere Beschreibung, um eine zukünftige Nachvollziehbarkeit zu ermöglichen.

Nicht zuletzt aufgrund der Natur der Daten und der vorhandenen Datenschutzbestimmungen bestehen Hemmnisse bei sozialwissenschaftlich Forschenden bezüglich *data sharing*. Hinzu kommt die Angst vor Verlust oder Einschränkung der Wettbewerbsfähigkeit bei einer frühzeitigen Datenweitergabe und vor Missbrauch der Daten durch Dritte. Solche emotionalen Hürden beeinflussen den Willen zur Datenweitergabe in weit höherem Maße als technische Barrieren.<sup>167</sup>

*„The traditional view that the primary outputs of a project are journal articles prevails, with the fate of the underlying data coming further down the list of priorities.“<sup>168</sup>*

Ob die besonderen Anforderungen an das FDM in den Sozialwissenschaften laut Literaturstudie den praktischen Gegebenheiten entsprechen, soll mithilfe der Experteninterviews verifiziert werden.

---

<sup>166</sup> Vgl. Spindler & Hillegeist (2011), S. 69.

<sup>167</sup> Vgl. Feijen (2011), S. 4.

<sup>168</sup> Key Perspectives (2010), S. 9.

### 3.3 Experteninterviews am WZB

Experteninterviews bieten die Möglichkeit einer vergleichsweise leichten und kurzfristig realisierbaren Daten- oder vielmehr Wissenserhebung. Sie können im Gegensatz zu (Massen-)Befragungen oder Beobachtungen in relativ kurzer Zeit durchgeführt und je nach Umfang der Gespräche und Art der inhaltlichen Analysetechnik zeitnah ausgewertet werden. Die Gewinnung der zu Interviewenden stellt sich ebenfalls einfacher dar, denn Professionalität, Erfahrung und das eigene Bewusstsein ihrer wissenschaftlichen Bedeutung sind mitunter Faktoren, welche die Expertinnen und Experten zur Teilnahme am Interview motivieren.<sup>169</sup>

Kennzeichnend für Experteninterviews ist der Expertenstatus der Interviewten. Nicht die Gesamtperson als Individuum wird analysiert, denn „[d]er Kontext [...] ist ein organisatorischer oder institutioneller Zusammenhang, der mit dem Lebenszusammenhang der darin agierenden Personen gerade nicht identisch ist...“.<sup>170</sup> Als Experten gelten Personen, die Verantwortung für eine Problemlösung tragen oder über spezielle Informationen verfügen. Dabei sind es nicht zwingend die offiziellen Entscheidungsträgerinnen und -träger, denen der Expertenstatus verliehen wird, sondern diejenigen, die über das detaillierte und anwendungsbezogene Wissen über interne Abläufe und Probleme verfügen.

Es sind zwei Arten von Expertenwissen unterscheidbar: Geben Expertinnen und Experten als Zielgruppe der Untersuchung Auskunft über ihr eigenes Handlungsfeld, definiert sich ihr Erfahrungswissen als *Betriebswissen*. Repräsentieren sie hingegen einen zur Zielgruppe ergänzenden Expertenkreis und sollen *Informationen über die Kontextbedingungen des Handelns der Zielgruppe* geliefert werden, handelt es sich um *Kontextwissen*.

Die Perspektive auf eine Form des Expertenwissens beeinflusst die spätere Auswertung. Für die vorliegende Arbeit soll das *Betriebswissen* der Expertinnen und

---

<sup>169</sup> Vgl. Bogner & Menz (2005), S. 7 ff.

<sup>170</sup> Meuser & Nagel (2005), S. 72/73.



Experten erforscht werden; für die Analyse „*wird ein kategoriales Gerüst als Bezugsrahmen [...] vorausgesetzt*“.<sup>171</sup>

*Dort, wo sich das Forschungsinteresse auf die ExpertInnen als Zielgruppe und nicht als Kontextgröße richtet, wird es in der Auswertung darum gehen, die entsprechenden Wissens- und Handlungsstrukturen, Einstellungen und Prinzipien theoretisch zu generalisieren, Aussagen über Eigenschaften, Konzepte und Kategorien zu treffen...*<sup>172</sup>

Eine Vergleichbarkeit der Interviews wird einerseits durch die Zugehörigkeit zu derselben Institution und andererseits mit der leitfadensorientierten Interviewführung erreicht.<sup>173</sup> Mithilfe der zuvor entwickelten Leitfragen ist zum einen eine thematische Annäherung an bestimmte Sachverhalte gegeben, der Interviewverlauf folgt jedoch keinem strikten Ablaufmodell, wodurch eine gewisse Offenheit des Interviewverlaufs vorhanden ist. Außerdem soll mit der Formulierung der Leitfragen verhindert werden, dass sich die befragende Person als *inkompetenter Gesprächspartner* herausstellt und die befragte Person ihre Entscheidung zur Teilnahme an dem Interview bereut.<sup>174</sup>

Für das im Rahmen dieser Arbeit zu erhebende Expertenwissen bietet das außeruniversitäre Wissenschaftszentrum Berlin für Sozialforschung beste Voraussetzungen. Durch die Vielzahl praxisbezogener, sozialwissenschaftlicher Forschungsprojekte<sup>175</sup> können exemplarisch Erkenntnisse über den Umgang der Forschenden mit den Forschungsdaten in unterschiedlichen Phasen des Forschungsprozesses bzw. Datenlebenszyklus gewonnen werden.

Im nachfolgenden Kapitel wird zunächst die Ausgangslage am WZB näher beschrieben, um einen kurzen Überblick über das Datenmanagement sowie Hintergrundinformationen zu geben. Anschließend folgt eine Erläuterung der Vorgehensweise zur Interviewdurchführung und der inhaltlichen Analyse. Danach werden die Interviewleitfragen vorgestellt und letztlich die Ergebnisse ausgewertet.

---

<sup>171</sup> Ebenda, S. 76.

<sup>172</sup> Ebenda, S. 77.

<sup>173</sup> Vgl. Ebenda, S. 75-81.

<sup>174</sup> Vgl. Ebenda, S. 77/78.

<sup>175</sup> Vgl. <http://www.wzb.eu/de/ueber-das-wzb> [Zugriff: 11.8.2013].

### 3.3.1 Ausgangslage

Das WZB ist eine gemeinnützige GmbH, die zu 75% vom Bund und zu 25% vom Land Berlin finanziert wird, zudem ist sie Mitglied der Leibniz-Gemeinschaft. Hier forschen über 160 Wissenschaftlerinnen und Wissenschaftler zu unterschiedlichen sozialwissenschaftlichen Themen, u.a. Bildung, Wandel politischer Systeme und Migration.<sup>176</sup> Die Forschungseinheiten sind in sechs Schwerpunkte und die Projektgruppe der Präsidentin strukturiert und beinhalten weitere Abteilungen, Forschungs-, Nachwuchs- und Projektgruppen.<sup>177</sup>

Die seit 2005 bestehende Abteilung „Zentrales Datenmanagement“ bietet primär für die Wissenschaftlerinnen und Wissenschaftler des WZB Serviceleistungen rund um die Erfassung, Beschaffung und Nutzung von Forschungsdaten an und verwaltet alle datenschutzrelevanten Maßnahmen, insbesondere Zugriffsrechte.<sup>178</sup> Dies ist vor allem hinsichtlich der Datenbeschaffung aus externen Quellen wichtig, wie bspw. bei den Daten der Amtlichen Statistik<sup>179</sup> oder des SOEP.

Für die systematische Erfassung der am WZB vorhandenen Daten kommt ein Meta-informationssystem zum Einsatz, welches als Forschungsdokumentations- und Berichtssystem („WZBaktiv“)<sup>180</sup> unter anderem kurze Beschreibungen der am WZB erstellten Datensätze erfasst. An dieser Stelle sei erwähnt, dass die Wissenschaftlerinnen und Wissenschaftler erfahrungsgemäß den für die Erfassung benötigten Zeitaufwand trotz des stark reduzierten Metadatenschemas als zu hoch und den Eingabeprozess der Metadaten als eher unerfreulich empfinden.<sup>181</sup>

Die Forschungsdaten werden seit 2012 über einen Testzugang und ab 2013 offiziell bei da|ra<sup>182</sup> registriert. Über die Webseiten des GESIS sind einige Datensätze aus WZB-Projekten zugänglich, wie z.B. die Daten der Deutschen Wahlstudie 2009-2017 (GLES),

---

<sup>176</sup> Vgl. <http://www.wzb.eu/de/ueber-das-wzb> [Zugriff: 11.8.2013].

<sup>177</sup> Vgl. <http://www.wzb.eu/de/forschung> [Zugriff: 11.8.2013].

<sup>178</sup> Vgl. <http://www.wzb.eu/de/ueber-das-wzb/organisation/service-und-verwaltung/datenmanagement> [Zugriff: 11.8.2013].

<sup>179</sup> Hierzu zählen die Statistiken aller offiziellen Ämter, z.B. des Statistischen Bundesamts: <https://www.destatis.de/DE/Startseite.html> [Zugriff: 23.5.2013].

<sup>180</sup> Vgl. Allmendinger (2013), S. 99.

<sup>181</sup> Quelle: Persönliches Gespräch mit R. Habich, 28.05.2013.

<sup>182</sup> Siehe auch Kapitel 2.3.

an deren Erstellung das WZB beteiligt ist.<sup>183</sup> Öffentlich einsehbar sind auf den Webseiten des WZB Informationen zu ausgewählten Datenbanken,<sup>184</sup> die hier (auch teilweise) produziert wurden.

Der empfohlene DDI-Standard<sup>185</sup> wird bei der Metadatenvergabe jedoch (noch) nicht genutzt, da mit diesem Daten aus Panel-Studien nicht komplett abgebildet werden können.<sup>186</sup> Die aktuelle Version 3.1 erlaubt zwar eine Beschreibung von Panel-Daten, jedoch nicht die vollständige Darstellung der konkreten Panel-Struktur.<sup>187</sup>

Die Bibliothek bietet neben zahlreichen gedruckten Statistiken ebenfalls Online-Zugriff auf Datensätze: Unter anderem werden über das Portal *OECD iLibrary* 23 statistische Datenbanken zur Verfügung gestellt, weitere Beispiele sind die *International financial statistics*, *World development indicators online* (Weltbank-Statistiken) und das Statistik-Portal *Statista*.<sup>188</sup>

Das Thema Forschungsdatenmanagement wird von der Bibliothek und dem Zentralen Datenmanagement kooperativ bearbeitet. So wurde beispielsweise ein Vortragsblock im Rahmen des Soziologie-Kongresses 2012 unter dem Titel „Forschungsdatenmanagement in den Sozialwissenschaften: Möglichkeiten, Grenzen, Perspektiven“ gemeinsam organisiert<sup>189</sup> und an der Publikation des RatSWD-Working Papers zum Thema Forschungsinfrastrukturen der Sozial- und Geisteswissenschaften in Deutschland<sup>190</sup> aktiv mitgewirkt.

Eine eigene Policy zur guten wissenschaftlichen Praxis und zum Umgang mit Forschungsdaten ist am WZB nicht vorhanden. Jedoch gibt es den „Code of Conduct“, der zwar hauptsächlich Maßnahmen zur Karriereförderung beinhaltet, aber auf den *Regeln zur Sicherung guter wissenschaftlicher Praxis* der Leibniz-Gemeinschaft beruht.

---

<sup>183</sup> Vgl. <http://www.gesis.org/wahlen/gles/> [Zugriff: 11.8.2013].

<sup>184</sup> Vgl. <http://www.wzb.eu/de/publikationen/datenbanken> [Zugriff: 11.8.2013].

<sup>185</sup> Siehe auch Kapitel 3.2.

<sup>186</sup> Quelle: Persönliches Gespräch mit R. Habich, 28.05.2013.

<sup>187</sup> Quelle: E-Mail von W. Zenk-Möltgen, 24.06.2013.

<sup>188</sup> Vgl. <http://www.wzb.eu/de/bibliothek/bestand-recherche/datenbanken> [Zugriff: 11.8.2013].

<sup>189</sup> Vgl. DGS (Deutsche Gesellschaft für Soziologie) (2012), S. 135.

<sup>190</sup> Vgl. Bove u.a. (2011).

Um den Empfehlungen hinsichtlich guter wissenschaftlicher Praxis auf europäischer Ebene zu begegnen, hat das WZB im April dieses Jahres als eine der ersten Einrichtungen Deutschlands die „Europäische Charta für ForscherInnen“ unterzeichnet, die jedoch viele Maßnahmen vorsieht, die am WZB bereits praktisch umgesetzt werden.<sup>191</sup> Die Charta sieht im Umgang mit Forschungs- bzw. Primärdaten vor, dass Daten korrekt zu zitieren sind – andernfalls läge ein Plagiat vor; die Datenerhebung und -auswertung muss entsprechend datenschutzrechtlicher Reglementierungen erfolgen. Re- und Sekundäranalysen sind mit Zustimmung der Verantwortlichen zu ermöglichen:

*“Methods of collection and analysis, the outputs and, where applicable, details of the data should be open to internal and external scrutiny, whenever necessary and as requested by the appropriate authorities.”<sup>192</sup>*

Trotz der Tätigkeiten der Abteilungen „Zentrales Datenmanagement“ und „Bibliothek und wissenschaftliche Information“ zur Unterstützung bei der Recherche, Nutzung und Erfassung von Forschungsdaten ist der Umgang mit ihnen aus infrastruktureller Sicht nicht einheitlich: Es existieren je nach Forschungsprojekt Unterschiede hinsichtlich der Speicherung und des Zugriffs – sogenannte „Insellösungen“.<sup>193</sup> Die wissenschaftspolitischen Rahmenbedingungen sind mit den Richtlinien der Leibniz-Gemeinschaft und der EU-Charta vorhanden. Es fehlt jedoch eine einheitliche, institutionelle Struktur für das gesamte Forschungsdatenmanagement, die von den Wissenschaftlerinnen und Wissenschaftlern in Gänze angenommen wird.

Der Aufbau einer solchen Infrastruktur ist allerdings in Planung: Mit dem Projekt *SowiDataNet - Forschungsdatenverbund für die Sozial- und Wirtschaftswissenschaften* soll in Kooperation mit GESIS, ZBW (Zentralbibliothek für Wirtschaftswissenschaften) und DIW eine Infrastruktur zur Archivierung quantitativer sozial- und wirtschaftswissenschaftlicher Daten entwickelt werden. Der Projektbeginn ist für Anfang 2014 geplant.<sup>194</sup>

---

<sup>191</sup> Vgl. <http://www.wzb.eu/de/news/wzb-unterzeichnet-eu-charta> [Zugriff: 11.8.2013].

<sup>192</sup> <http://ec.europa.eu/euraxess/index.cfm/rights/europeanCharter> [Zugriff: 11.8.2013].

<sup>193</sup> Quelle: Persönliches Gespräch mit R. Habich, 28.05.2013.

<sup>194</sup> Quelle: Persönliches Gespräch mit R. Habich, 28.05.2013.

Nach Vorstellung der Ausgangssituation werden in den nachfolgenden Kapiteln die qualitative Analyse der Interviews sowie die zuvor entwickelten Leitfragen beschrieben und schließlich die Ergebnisse der Analyse ausgewertet.

#### 3.3.2 Methodik

Die Auswahl der Interviewpartner<sup>195</sup> erfolgte im Hinblick auf die derzeit am WZB laufenden Forschungsprojekte. Um einen umfassenden Überblick über die Anforderungen an das Forschungsdatenmanagement während des gesamten Datenlebenszyklus zu erhalten, wurden Experten ausgewählt, deren Projekte sich in unterschiedlichen Phasen befinden. Die Interviewpartner sind alle aktiv an der Projektdurchführung und der Datenauswertung beteiligt, wodurch sie über spezielle Informationen zu internen Abläufen verfügen. Somit kann den Interviewpartnern für die im Rahmen der vorliegenden Masterarbeit zu bearbeitende Fragestellung der Expertenstatus zugeschrieben werden.

Auf Wunsch der Interviewpartner erfolgt die Auswertung ohne namentliche Nennung ihrer Person oder des jeweiligen Forschungsprojektes. Es kann lediglich die Phase, in der sich das Forschungsprojekt zum Zeitpunkt der Interviewdurchführung befindet, benannt werden.

Für die Verschriftlichung der Interviews wurde ein Transkriptionssystem gewählt, dessen Basisregeln den Ansprüchen einer rein *inhaltlichen* Analyse genügen. Das Gesprächsanalytische Transkriptionssystem (GAT) wurde bewusst nicht verwendet, da der zeitliche und finanzielle Aufwand für das Transkribieren in keinem Verhältnis zum erwartbaren Nutzen stünde.<sup>196</sup> Für die inhaltliche Analyse ist die Kenntnis darüber, *was* ausgesagt wird, vollkommen ausreichend; im Gegensatz dazu zielt die Transkription nach GAT zusätzlich darauf ab, *wie* etwas gesagt wird. Die Regeln des verwendeten

---

<sup>195</sup> Um dem Wunsch nach Anonymität der Interviewpartner nachzukommen und eventuellen Rückschlüssen auf das Geschlecht vorzubeugen, wird im weiteren Textverlauf bezüglich der Interviewpartner ausschließlich das generische Maskulinum verwendet.

<sup>196</sup> Die Anfertigung der Transkripte erfolgte durch einen Dienstleister (<http://www.schreibservice-dienstleistung.at/>, Zugriff: 24.6.2013); fachspezifische Begriffe wurden von der Autorin nachträglich ergänzt.

Transkriptionssystem von Dresing & Pehl sind in Anhang B zu finden.<sup>197</sup> Die vollständigen Transkripte sind in anonymisierter Form in Anhang C beigefügt.<sup>198</sup>

Ziel der Interviews war es, das Betriebswissen der Experten zu ermitteln. Der Analyse soll dementsprechend ein *kategoriales Gerüst* zugrunde liegen. Allgemein stellt die Entwicklung eines Kategoriensystems den zentralen Punkt der qualitativen Analyse dar.<sup>199</sup>

Die qualitative Inhaltsanalyse hat zum Ziel, Text- oder auch andere Kommunikationsmaterialien systematisch zu bearbeiten. Es gibt dabei vier Verfahren, die als Analysetechnik infrage kommen:

Eine Technik ist die *Zusammenfassung*, mit dem Ziel, das Material unter anderem durch das Auslassen und Generalisieren ausgewählter Textstellen zu reduzieren, sodass ein *überschaubarer Kurztext* entsteht. Diese Technik bietet sich an, wenn sich das Forschungsinteresse auf die inhaltliche Ebene des Materials beschränkt. Das Vorgehen bei der zusammenfassenden Inhaltsanalyse wird auch für die *induktive Kategorienbildung* genutzt.<sup>200</sup> Hierbei wird nach Festlegung eines Definitionskriteriums schrittweise das Material durchgearbeitet, um Kategorien zu entwickeln, die wiederum auch zu Hauptkategorien zusammengefasst werden können.<sup>201</sup> Die Technik der *Explikation* hat das Gegenteil der Zusammenfassung zum Ziel: Es wird zusätzliches Material herangezogen, um unklare Textstellen zu verstehen. Bei der *strukturierenden Inhaltsanalyse* wird das Material z.B. nach bestimmten Aspekten gefiltert oder nach vorab festgelegten Kriterien eingeschätzt.<sup>202</sup>

*„In qualitativ orientierter Forschung wird aber auf eine systematische Ableitung von Auswertungsgesichtspunkten aus dem Material, also eine induktive Kategorienbildung, großer Wert gelegt.“<sup>203</sup>*

---

<sup>197</sup> Vgl. Dresing & Pehl (2012), S. 25 ff.

<sup>198</sup> Die originalen Audiodateien mit den Experteninterviews sind aufgrund des Wunsches nach Anonymität nicht auf beiliegender DVD enthalten und werden lediglich für interne Zwecke aufbewahrt.

<sup>199</sup> Vgl. Mayring (2008), S. 43, 53.

<sup>200</sup> Vgl. Mayring (2012), S. 472.

<sup>201</sup> Vgl. Mayring (2000), S. [4].

<sup>202</sup> Vgl. Mayring (2012), S. 473; Mayring (2008), S. 58.

<sup>203</sup> Mayring (2002), S. 115.

Um dem Anspruch qualitativ orientierter Forschung gerecht zu werden und weil die *inhaltliche* Ebene betrachtet werden soll, wird für die Inhaltsanalyse der Experteninterviews die Technik der induktiven Kategorienbildung verwendet. Durch die schrittweise, systematische Bearbeitung der Interviewtranskripte und Bildung der Kategorien wird die Interpretation der Ergebnisse nachvollziehbar begründet.

Das Ausgangsmaterial für die Inhaltsanalyse bilden die Transkripte der Interviews.

Interview / Studie	Länge	Phase	Anhang
1 (1 Experte)	17 S. (44,10 min)	Datenerhebung	C1
2 (1 Experte)	8 S. (26,45 min)	Datenauswertung	C2
3 (2 Experten)	14 S. (36,54 min)	Abschlussphase (Erste Ergebnisse sind publiziert)	C3

Tabelle 1: Eckdaten der Experteninterviews

Die zu beantwortende Fragestellung lautet: Wie gestaltet sich das FDM und welche Erfahrungen haben die Experten mit FDM? Die Leitfragen werden in Kapitel 3.3.3 gesondert aufgeführt. Sie bilden die Grundlage für das Selektionskriterium, nach dem die Kategoriendefinition erfolgen soll, d.h. wird in der Textpassage eine Aussage zu einer der Leitfragen getroffen und kann mit ihr eine der dazugehörigen Fragen beantwortet werden, soll eine Kategorie erzeugt werden.

Danach wird der Text *Zeile für Zeile* durchgearbeitet und geprüft, ob das Kriterium erfüllt ist. Die jeweilige Textstelle wird entweder einer bereits verwendeten Kategorie zugeordnet (*Subsumtion*), oder es wird eine neue Kategorie gebildet. Nachdem die Hälfte des Textes auf diese Weise durchgearbeitet ist, erfolgt eine Revision der Kategorien, um zu prüfen, ob das Selektionskriterium erfüllt ist und die gebildeten Kategorien zur Beantwortung der Fragestellung zielführend sind (*formative Reliabilitätsprüfung*). Nach vollständiger Durcharbeitung werden die Kategorien einer abschließenden Überprüfung unterzogen (*summative Reliabilitätsprüfung*). Das Ergebnis der Analyse ist ein Kategoriensystem, welches schließlich im Sinne der

Fragestellung interpretiert wird.<sup>204</sup> Eine ausführliche Beschreibung des Kategoriensystems erfolgt mit der Auswertung in Kapitel 3.3.4. Die Vorgehensweise der induktiven Kategorienbildung veranschaulicht das Ablaufmodell in der nachfolgenden Abbildung 11.

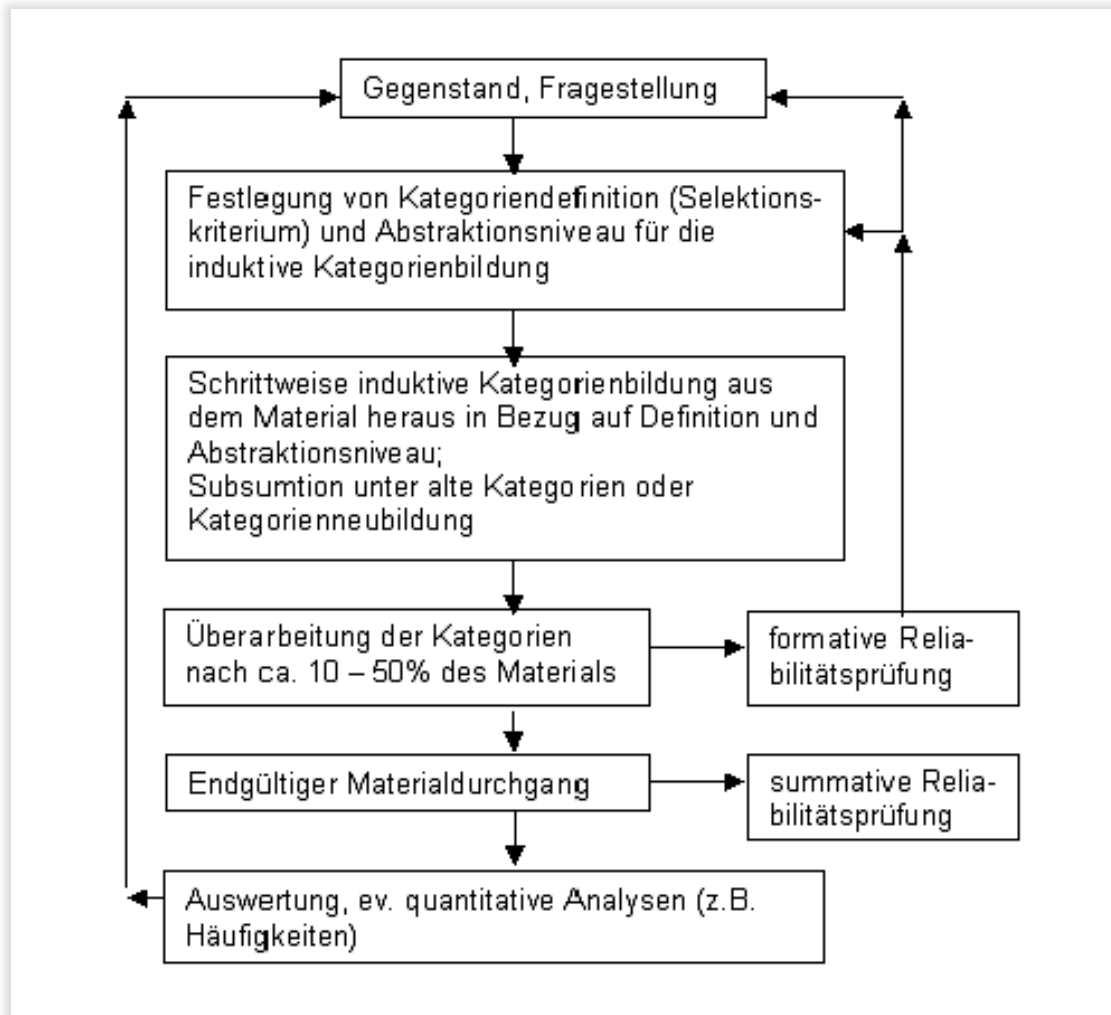


Abbildung 11: Ablaufmodell der induktiven Kategorienbildung<sup>205</sup>

### 3.3.3 Leitfragen

In Vorbereitung auf die Experteninterviews wurden Leitfragen entwickelt, die insgesamt vier Bereiche abdecken: Definition von Forschungsdaten, Phasen des Datenlebenszyklus, Umgang mit der Erstellung von Studiendokumentationen bzw. Metadatenvergabe und Nachnutzung von Forschungsdaten.

<sup>204</sup> Vgl. Mayring (2008), S. 74 ff.

<sup>205</sup> Quelle: Mayring (2000), S. [4].



Tabelle 2 zeigt die Leitfragen mit den dazugehörigen Einzelaspekten in Form von Unterfragen, die während der Interviews bei Bedarf gestellt wurden.

Leitfrage	Einzelaspekte/Unterfragen
1. Welche Arten von quantitativen Daten werden als Forschungsdaten wahrgenommen?	<ul style="list-style-type: none"> <li>➤ Welche Arten von Daten entstehen?</li> <li>➤ Welche Erhebungsart(en) / Formate fallen an?</li> <li>➤ Welche Software wird zur Erhebung / Auswertung verwendet?</li> <li>➤ Welche Merkmale / Eigenschaften weisen Forschungsdaten auf?</li> </ul>
2. Wie gestaltet sich der Datenlebenszyklus?	<ul style="list-style-type: none"> <li>➤ Wie sieht der Projektablauf aus, wann entstehen (die meisten) Daten?</li> </ul>
3. Wie werden die Daten dokumentiert?	<ul style="list-style-type: none"> <li>➤ Werden Standards genutzt (z.B. nur sozialwiss. Klassifikationssysteme oder auch andere)?</li> <li>➤ Werden Dokumentations- oder Beschreibungsvorlagen verwendet?</li> <li>➤ Wie werden wann von wem für welchen Zeitraum welche Metadaten vergeben?</li> <li>➤ Wer ist verantwortlich? (Qualitätssicherung)</li> <li>➤ Wie/wo werden die Daten gesichert?</li> <li>➤ Wie stellen sich Verbleib / Sicherung / Verantwortlichkeit für die Metadaten nach Ende der Projektlaufzeit dar?</li> <li>➤ Sind Defizite bewusst bzw. Wünsche vorhanden?</li> </ul>
4. Wie sieht die Nachnutzung von Daten aus?	<ul style="list-style-type: none"> <li>➤ Wann und unter welchen Bedingungen werden Daten (faktisch oder absolut anonymisiert) weitergegeben? (z.B. Einverständniserklärungen, Embargo-Fristen)</li> </ul>

	<ul style="list-style-type: none"> <li>➤ Werden Daten auch unabhängig vom eigenen Forschungsinteresse an andere Konsumenten weitergegeben? (Interdisziplinarität)</li> <li>➤ Besteht Interesse an einer langfristigen Archivierung?</li> <li>➤ Wurden vorhandene Daten für das eigene Forschungsprojekt nachgenutzt?</li> <li>➤ Inwieweit wird die Bereitschaft zur Datenweitergabe von anderen (Kollegen, Förderern, der eigenen Institution) beeinflusst?</li> <li>➤ Wie stellen sich Verbleib / Sicherung / Verantwortlichkeit für die Daten nach Ende der Projektlaufzeit dar?</li> <li>➤ Sind Defizite bewusst bzw. Wünsche vorhanden?</li> </ul>
--	--

Tabelle 2: Leitfragen

### 3.3.4 Auswertung

Die im Rahmen der qualitativen Inhaltsanalyse gebildeten Kategorien stellen die Basis für die Interpretation der Ergebnisse dar. Es wurden insgesamt 25 Kategorien (K1 bis K25) entwickelt, die in Abgrenzung zu den vier Leitfragen im Kategoriensystem<sup>206</sup> tabellarisch aufgezeigt werden. Die dazugehörigen Textpassagen wurden ebenfalls aufgelistet und die jeweiligen Textzeilen der Transkripte in Klammern eingefügt.<sup>207</sup> Die Inhalte der Textpassagen wurden stichpunktartig zusammengefasst. Die letzte Spalte der Tabelle enthält die Ergebnisse dieser Zusammenfassungen zu den einzelnen Kategorien. Da das Textmaterial als Ganzes analysiert wurde, stellen die einzelnen Aussagen allgemeine Gedanken dar und beziehen sich nicht auf eine konkrete Studie.

Bei Betrachtung des Kategoriensystems fällt auf, dass manche Kategorien im Vergleich zu anderen sehr schwach besetzt sind, d.h. ihnen konnte lediglich eine Textpassage

<sup>206</sup> Siehe Anhang D.

<sup>207</sup> Personen- oder projektbezogene Aussagen wurden von der Autorin entfernt und die Textstellen durch [...] gekennzeichnet.

zugeordnet werden (siehe K3, K8, K9). Auch wurden manche Kategorien auf Grundlage der Aussagen eines einzelnen Experten gebildet (siehe K3, K8, K9, K18, K25). Dies kann mit der offenen Interviewführung begründet werden: Nicht allen Einzelaspekten wurde sich in jedem Interview auf die gleiche Weise genähert, sodass die Unterfragen in der Leitfragentabelle teilweise nicht oder nur geringfügig beantwortet werden können, wie beispielsweise die Frage zu den Datenformaten (siehe K3). Aufgrund der inhaltlichen Bedeutung dieser Aussagen wurde auf die Bildung einer eigenständigen Kategorie jedoch nicht verzichtet.

Im Folgenden werden die Ergebnisse des Kategoriensystems näher erläutert.

Für den Begriff *Forschungsdaten* kann keine eindeutige Definition gegeben werden. Es ist dennoch ein breites Verständnis von Forschungsdaten vorhanden. Die Unterscheidung von Forschungsdaten erfolgt primär in Fremddaten, die nachgenutzt werden und Daten, die selbst erhoben bzw. weiter verarbeitet werden. Die nachgenutzten Fremddaten sind solche, die von Dienstleistern speziell zu diesem Zweck produziert werden, wie bspw. das SOEP oder NEPS. Auch automatisch generierte Prozessdaten, wie sie z.B. in Amtsgerichten entstehen, können Bestandteil einer Analyse sein.

Forschungsdaten werden als Informationsquellen angesehen, wobei sie nicht mit Daten im Allgemeinen gleichgesetzt werden, sondern eine Untermenge bilden. Forschungsdaten sind Daten, die zum Zweck der Analyse in einer aufbereiteten Form bestehen. Dabei ist die Unterscheidung in qualitative und quantitative Daten ebenfalls abhängig von der Art der Auswertung. Eine Klassifikation von Forschungsdaten kann auch in Abhängigkeit von der Art der Untersuchungseinheit erfolgen: Daten werden auf der Makro-, Meso- und Mikroebene gesammelt und ausgewertet. Bei den Primärdaten handelt es sich zumeist um statistische Daten, die mit den gängigen Softwaretools SPSS oder STATA ausgewertet werden. Einzelne Dateiformate wurden allerdings nicht benannt.

Der Begriff *Datenlebenszyklus* ist den Experten nicht geläufig, dennoch ist ein allgemeines Verständnis über die einzelnen Phasen im Zusammenhang mit der Datenerhebung und -auswertung vorhanden. Dabei wird deutlich, dass bei der Erhebung die meisten Daten entstehen und wiederum neue Forschungsdaten das Resultat der Aus-

wertung sind. Die Löschung von Daten nach Abschluss des Forschungsprozesses gehört ebenfalls zum Datenlebenszyklus. Diese Tatsache ist stärker im Bewusstsein verankert, als dass die langfristige Archivierung Teil des Datenlebenszyklus ist. Dessen Verlauf wird allerdings durch äußere Rahmenbedingungen beeinflusst, dies betrifft insbesondere die Weitergabe: Fehlende Ressourcen und die Ungewissheit über die Zukunft nach Ende des Forschungsprojekts wirken sich nachteilig auf diese Phase aus, da für die Weitergabe personenbezogene Daten anonymisiert werden müssen und Anonymisierung eine eigenständige, zusätzliche Aufgabe darstellt.

An der Erstellung von *Dokumentationen* und der Vergabe von *Metadaten* sind unterschiedliche Personen beteiligt. Der für die inhaltliche Bearbeitung verantwortliche Akteur ist im Idealfall zugleich für die entsprechende Beschreibung zuständig, wie z.B. für Kodierungen; hauptverantwortlich ist letztlich die Projektleitung. Die Vergabe von Metadaten geschieht schließlich im Hinblick auf die Datenweitergabe (z.B. für die Registrierung bei da|ra), zum Zweck der Nachweisbarkeit und um zukünftig Interessierten das Auffinden der Daten zu ermöglichen.

Auch Dokumentationen werden als hilfreich angesehen. Der Mehrwert ergibt sich hierbei nicht nur für die zukünftige Nutzung: Ohne das Wissen über verwendete Kodierregeln werden weitere Analysen erschwert oder sind überhaupt nicht durchführbar. Für die Arbeit innerhalb des eigenen Projektes dienen Dokumentationen als Erinnerungsstütze über getroffene Entscheidungen und als Mittel zur Überprüfung über bestimmte Abläufe, z.B. Interviewverläufe. Protokolliert werden daher neben Befragungen und Fragebögen hauptsächlich die Methodik und Kodierregeln.

*„[A]ber es ist auch wichtig für einen selber, dass man weiß, okay wie hat man entschieden unter Umständen, das merkt man erst später. Das war nicht die gute Entscheidung, das muss man auch dokumentiert haben, beispielsweise in welchen Fällen hat man was entschieden?“<sup>208</sup>*

---

<sup>208</sup> Interview 1, Zeilen 190-193.

Die Vorgehensweise beinhaltet weniger den Einsatz von Standards, es werden jedoch Vorlagen genutzt, die an die Bedürfnisse des jeweiligen Projekts entsprechend angepasst werden.

Die Experten sind sich des Wertes von Beschreibungen in Form von Metadaten oder Dokumentationen bewusst, systematisches Vorgehen wird als wünschenswert gesehen. Dennoch wird in der Praxis zumeist auf das Erstellen ausführlicher Dokumentationen verzichtet. Das Fehlen solcher ist eindeutig ressourcenbedingt, da diesbezügliche personelle Kapazitäten in der Projektplanung nicht enthalten sind und der hohe Aufwand den Nutzen kaum rechtfertigt. Mangelndes Interesse an den Daten außerhalb der Institution sowie die fehlende Anerkennung motivieren nicht sonderlich zur Erstellung ausführlicher Dokumentationen.

*„Also man erwartet denke ich, dass man das automatisch macht, aber es ist mit Aufwand verbunden und der wird, glaube ich nirgends genügend reflektiert und [...] da kriegt man auch nichts dafür also. Der Output für uns sind Artikel oder Studien und nicht ein gutes Datenmanagement zu machen oder eine gute Datendokumentation.“<sup>209</sup>*

Zugleich gibt es widersprüchliche Einstellungen: Einerseits nützen Dokumentationen der eigenen Arbeit, andererseits wird keine Notwendigkeit dafür gesehen, auch aufgrund der unbekanntenen Zielgruppe.

*„Und ja die Frage für wen macht man es dann, weil selber, selber weiß ich, wie ich es gemacht habe, wenn ich die Daten auswerte und ich weiß nicht, wer jetzt vielleicht mit den Daten noch mal arbeitet.“<sup>210</sup>*

Die langfristige Archivierung von Dokumentationen ist nicht eindeutig klar. Aus Datenschutzgründen müssen personenbezogene Daten gelöscht werden, z.B. auch in Interviewprotokollen. Das bedeutet, dass auch Dokumentationen, die für die eigene Arbeit (als Mittel zur Überprüfung) erstellt wurden, nicht automatisch mit den Daten weitergegeben werden dürfen. Hier ist zu klären, ob aus datenschutzrechtlicher Sicht Einwände bestehen.

---

<sup>209</sup> Interview 3, Zeilen 401-407.

<sup>210</sup> Interview 3, Zeilen 364-366.

Das Thema *Nachnutzung* stellt einen gravierenden Aspekt im FDM dar.<sup>211</sup> Die Experten sind einer Datenweitergabe generell nicht abgeneigt. Es müssen jedoch bestimmte Bedingungen vorliegen, damit Daten weiter gegeben werden. Diese Bedingungen sind zum einen, dass die Weitergabe der Daten nur in anonymisierter Form erfolgt, wie es die Datenschutzbestimmungen vorgeben. Mit der Anonymisierung soll verhindert werden, dass ein Personenbezug aus den Daten hergestellt werden kann, dieser Vorgang der Anonymisierung erfordert einen hohen Zeit- bzw. Arbeitsaufwand.

*„[S]o sind die Arbeitsbedingungen in der Wissenschaft, also in meiner Freizeit mache ich keine Datenanonymisierung.“<sup>212</sup>*

Forschungsdaten werden in der Regel mit zeitlicher Verzögerung, also nach Ablauf einer Embargo-Frist, zur *Nachnutzung* bereitgestellt. Mittels eines entsprechenden Nutzungsvertrages kann die Weitergabe kontrolliert erfolgen. Für diese Bedingungen gibt es verschiedene Gründe: Zunächst ist zu berücksichtigen, dass eine Datenerhebung sehr aufwendig ist.

*„Und der Grund ist einfach, ich meine Daten erheben ist ein Haufen, Haufen, Haufen Arbeit.“<sup>213</sup>*

Hinzu kommt, dass unter den gegebenen Arbeitsbedingungen (und allgemein den wissenschaftspolitischen Rahmenbedingungen) die Publikation der Ergebnisse als wissenschaftlicher Output einen höheren Stellenwert einnimmt als die Veröffentlichung der zugrunde liegenden Daten. Dies wurde bereits im Zusammenhang mit Dokumentationen und Metadatenvergabe deutlich.

*„Aber keiner von uns ist unbefristet und es wird einfach geguckt, was man veröffentlicht hat.“<sup>214</sup>*

*„[S]trategisch soll man immer sagen, nein ich erhebe keine Daten. Das ist meine zweite Datenerhebung, ich werde in meinen Lebenslauf*

---

<sup>211</sup> Die Interviews ließen sich zum Teil auch quantitativ auswerten: Beispielsweise nimmt allein das Thema *Nachnutzung* im Kategoriensystem 88 Textpassagen ein, dies sind 52 % der Textpassagen insgesamt (die Summe der Textpassagen zu den anderen drei Leitfragen beträgt 80).

<sup>212</sup> Interview 1, Zeilen 314-315.

<sup>213</sup> Interview 1, Zeilen 374-375.

<sup>214</sup> Interview 1, Zeilen 379-380.

*ziemliche Löcher haben, weil ich eben Daten erhoben habe und am Endeffekt vollzeitbeschäftigt war mit Datenerhebung. Und in der Zwischenzeit nicht, also veröffentlicht habe...<sup>215</sup>*

Die Experten erheben daher für sich den Anspruch, ihre erhobenen Daten auch als erste auszuwerten und zu publizieren, um sich auf dem Markt profilieren zu können.

*„Ja einfach Forschung oder Wissenschaft funktioniert ja daher, dass man Neues hervorbringt ja und durch die (...) Neuigkeit der (...) Erkenntnisse und natürlich Daten, sich profilieren kann und veröffentlicht wird. Und dadurch ist es halt ein bisschen systematisch halt angelegt, dass Forscher und Forscherinnen teilweise nicht gerne ihre Daten, zumindest nicht zum aktuellen Zeitpunkt dann herausgeben einfach. Weil sie die Ersten sein wollen, die damit veröffentlichen möchten.“<sup>216</sup>*

*„Also es hat weniger mit, ein Jahr nach Veröffentlichung zu tun, sondern dass man selber auch in der Lage ist zu veröffentlichen.“<sup>217</sup>*

Die Weitergabe von Forschungsdaten erfolgt auch in Abhängigkeit von der betreffenden Person bzw. Zielgruppe. Wird mit der Weitergabe zur Nachwuchsförderung beigetragen, sind die Experten eher bereit, ihre Daten zur Verfügung zu stellen, auch ohne zeitliche Verzögerung.

*„Also wir können also dazu sagen, dass unsere Daten so ziemlich einmalig sind und wir uns natürlich nicht dort gegen generell gegen eine Nutzung verschließen. Also beispielsweise haben wir eine Doktorandin aus [...] die gerne mit unseren Daten arbeitet oder so verschiedene Leute, da passiert schon ein bisschen was. Aber wir sehen es gegenwärtig nicht als sinnvoll an, wenn wir beispielsweise die Daten frei zur Verfügung stellen würden.“<sup>218</sup>*

---

<sup>215</sup> Interview 1, Zeilen 384-388.

<sup>216</sup> Interview 2, Zeilen 331-336.

<sup>217</sup> Interview 1, Zeilen 366-367.

<sup>218</sup> Interview 3, Zeilen 246-250.

Wenn die Auswertung der eigenen Daten durch andere Forschende dem eigenen Forschungsinteresse nützt, bestehen ebenfalls geringere Widerstände, die Daten weiterzugeben.

*„Also wir, in der Hinsicht ist es vielleicht noch mal wichtig, dass wir versuchen auch durch solche externe Nutzung bestimmte Themen, die wir nicht bearbeiten oder zu denen wir auch nicht mehr kommen werden, durchaus dort, ja noch auswerten zu lassen so.“<sup>219</sup>*

Einer Weitergabe an Externe oder Fachfremde stehen die Experten dennoch teilweise skeptisch gegenüber, aus Angst vor Datenmissbrauch, Fehlinterpretationen oder Rufschädigung.

*„Na ja so ganz frei würde ich das nicht weg, rausgeben glaube ich. (lacht) Wenn jemand sagt, hey uns interessiert das nur mal und, und (...). Ich habe da auch in mancher Hinsicht Bedenken, ich meine man kann ja auch viel Unsinn machen, mit so einem Datensatz. Man kann ja auch den hernehmen, um zu zeigen, na ja Ihr habt nicht sauber gearbeitet oder, oder. Oder in der und der Hinsicht (...) sind Defizite vorhanden, also.“<sup>220</sup>*

*„Also vielleicht noch mal zusammenfassend, wir sind generell bereit zur Weitergabe, aber würden uns das genau angucken, was damit gemacht werden soll. Also das würde ich eher als unsere Verantwortung jetzt sehen. [...] Dass da nicht irgendwas mit passiert, was nicht im Sinne (...) der Daten auch liegt.“<sup>221</sup>*

Abgesehen von den Befürchtungen, die Daten könnten missverstanden werden oder die Ergebnisse aus Sekundär- bzw. Re-Analysen zu unerwünschten Interpretationen führen, bestehen generell Zweifel an einem künftigen Nutzungsinteresse. Begründet wird dies mit dem vermeintlichen Aktualitätsverlust der Daten, zudem beschränkt sich

---

<sup>219</sup> Interview 3, Zeilen 460-462.

<sup>220</sup> Interview 3, Zeilen 474-478.

<sup>221</sup> Interview 3, Zeilen 590-596.



das Interesse anderer auf die Ergebnisse der Auswertung, z.B. in Form von Publikationen.

*„Also ob sie jemand benutzt in zehn Jahren mag ich bezweifeln...“<sup>222</sup>*

*„Und ich bin mir allerdings damit im Klaren, dass solche Daten Verfallsdatum auch haben. Und nach ein paar Jahren damit keiner mehr arbeiten will oder arbeiten möchte. Das heißt, dass das übliche Problem der Datenfriedhöfe hier auch bei uns besteht.“<sup>223</sup>*

*„Wir haben es eher mit einem Feld zu tun, die praktisch erst mal die Auswertung haben wollen und nicht die Daten direkt.“<sup>224</sup>*

Wie die Bereitschaft zur Datenweitergabe von Kolleginnen und Kollegen (auch im WZB) wahrgenommen wird, ist nicht hinreichend bekannt. Innerhalb der eigenen Institution wird die Bereitschaft zur Weitergabe höher gesehen, als bei Nicht-Angehörigen oder sogar Fachfremden. Es ist jedoch zu unterscheiden, ob es sich um quantitative oder qualitative Daten handelt.

*„[J]a ich würde das so einschätzen, dass es IM Haus so relativ unkompliziert ist, von daher ist schon Bereitschaft ist klar.“<sup>225</sup>*

*„[A]nsonsten muss ich sagen, kenne ich nicht sehr viele, also direkt mittelbar Kollegen und Kolleginnen die Daten erheben und also natürlich kann man sich jetzt fragen, also welche Daten? Also ich kenne natürlich Kollegen und Kolleginnen, die qualitative Daten erheben. Und die geben definitiv ihre Daten nicht weiter.“<sup>226</sup>*

Die Bereitschaft zur Weitergabe ist höher, wenn die Ergebnisse bereits publiziert sind. Hier kommt erneut der Anspruch auf Erstveröffentlichung zum Ausdruck.

*„Ich habe eigentlich kaum Erfahrung. Aber doch, also sozusagen ein, zwei wenige, wo ich schon sagen würde, dass die Leute auch, also*

---

<sup>222</sup> Interview 1, Zeile 460.

<sup>223</sup> Interview 3, Zeilen 250-253.

<sup>224</sup> Interview 3, Zeilen 334-335.

<sup>225</sup> Interview 2, Zeilen 308-309.

<sup>226</sup> Interview 1, Zeilen 539-542.

*mein Eindruck wäre schon, die Leute sitzen schon auch auf Ihren Daten. Weil es generell, also ich hatte wenig mit anderen Projekten zu tun direkt.[...] Aber mein Eindruck wäre schon, solange die Leute auf ihren, an ihren Projekten sitzen und daran arbeiten, dann an Publikation arbeiten und da Leute direkt dran sind, ist es natürlich was, wo man erst mal auf seinen Sachen auch sitzt und guckt, was man damit macht.“<sup>227</sup>*

Eine andere Sichtweise beinhaltet, dass unterschiedliches Forschungsinteresse und andere Auswertungsmethoden den Konkurrenzdruck mindern und somit einer Weitergabe an externe oder fachfremde Forschende offener begegnet wird. Inwieweit diese Ansicht in der Praxis geteilt wird, kann im Rahmen dieser Arbeit allerdings nicht näher untersucht werden.

*„Und wenn jemand forscht zu was ganz anderem oder würde ganz oft eine ganz andere Perspektive an, auswerten, dann sind die Widerstände wahrscheinlich auch geringer...“<sup>228</sup>*

Neben der Datenweitergabe gehören Auffindbarkeit und Verfügbarkeit zu Aspekten der Nachnutzung. Für die Verfügbarkeit wird der Datenmanager zugleich als Verantwortlicher und Ansprechperson gesehen. Positiv hervorzuheben ist an dieser Stelle, dass die Notwendigkeit von Migrationen bekannt ist, auch wenn der Begriff nicht explizit genannt wurde:

*„Na Pflege in dem Fall gibt es ja nicht mehr, aber nee wobei die immer auf dem technischen Stand sind. Meine Vermutung wäre, dass - dadurch, dass es ein Kooperationsprojekt ist, müsste man entscheiden, ob wir oder das Partnerinstitut und ob beide das machen. Und dann kann ich mir vorstellen, dass wenn das hier wäre, die Abteilung hat einen Datenmanager, dass er dann dafür verantwortlich wäre, dass jemand weiß, wo die Daten sind...“<sup>229</sup>*

---

<sup>227</sup> Interview 3, Zeilen 539-548.

<sup>228</sup> Interview 2, Zeilen 354-356.

<sup>229</sup> Interview 1, Zeilen 664-668.

Für die Nachweisbarkeit spielen Metadaten und die Registrierung der Daten eine wichtige Rolle. Zudem müssen Zugriffsrechte vorab geregelt sein, um den datenschutzrechtlichen Anforderungen nachzukommen.

Wie im Zusammenhang mit den Phasen des Datenlebenszyklus bereits ersichtlich wurde, wird die langfristige Archivierung nicht als expliziter Bestandteil des FDM angesehen. Dementsprechend sind Informationen über die Archivierung der erhobenen Daten nur oberflächlich vorhanden. Dieser Umstand ist durchaus nachvollziehbar, da es nicht die Aufgabe der Forschenden ist, sich um die Archivierung zu kümmern, sondern die Anbieter der Infrastrukturen diesbezüglich die Verantwortung tragen. Die geringe Kenntnis über die Zukunft der erhobenen Daten kann jedoch zum Anlass genommen werden, das Bewusstsein der Forschenden dahingehend zu schärfen, dass bereits Entscheidungen in einer frühen Phase des Datenlebenszyklus Auswirkungen auf die spätere Archivierung haben können. Diesbezüglich ist die Rücksprache mit den Infrastrukturanbietern von wesentlicher Bedeutung.

Die Archivierung der Daten wird allgemein als sinnvoll angesehen. Wenn Daten gelöscht werden, dann geschieht dies ausschließlich aus Datenschutzgründen. Doch handelt es sich in dem Fall nicht um Primärdaten, sondern um Beschreibungen für interne Zwecke (z.B. Protokolldaten).

Ein weiterer Aspekt der Nachnutzung liegt in der Fremddatennutzung für das eigene Forschungsprojekt. Im Zusammenhang mit den Merkmalen bzw. Eigenschaften von Forschungsdaten wurde bereits auf von Dienstleistern produzierte Datenbestände (z.B. SOEP) hingewiesen. Wenn Fremddaten genutzt werden, betrifft dies überwiegend jene großen Datensätze oder Daten der Amtlichen Statistik, auch werden Strukturdaten, also soziodemografische Daten (z.B. Angaben zu Geschlecht, Alter, Bildungsabschluss)<sup>230</sup> in die Analyse miteinbezogen. Bereits erhobene Daten werden zum Zweck der Vergleichbarkeit nachgenutzt, wenn z.B. eine Studie als eine Art „Neuaufgabe“ einer älteren wiederholt wird.

Die Betrachtung der Fremddatennutzung liefert neben den bestehenden Arbeitsbedingungen und der mangelnden Anerkennung für die Datenerhebung eine weitere

---

<sup>230</sup> Quelle: Persönliches Gespräch mit R. Habich, 06.08.2013.

Begründung für die bedingte Bereitschaft zur Datenweitergabe: Wenn die Experten selbst keine Forschungsdaten aus kleinen oder mittelgroßen Projekten für die eigene Forschung nachnutzen, sehen sie dementsprechend auch keine zwingende Notwendigkeit zur Weitergabe der eigenen Daten.

Insgesamt zeigen die Ergebnisse der Experteninterviews die Diskrepanz zwischen Theorie und Praxis. Die Bereitschaft, die eigenen Daten mit anderen Forschenden oder Interessierten zu teilen, wurde ausführlich erörtert. Das Bewusstsein für Datensatzbeschreibungen in Form von Dokumentationen und Metadaten sowie die Aufbewahrung von Forschungsdaten ist definitiv vorhanden, auch wenn die Vorteile von den Experten in unterschiedlichem Maße herausgestellt wurden. Die praktische Umsetzung der Anforderungen kann jedoch nur unter Berücksichtigung der gegebenen Arbeitsbedingungen und letztlich im Verständnis wissenschaftspolitischer Rahmenbedingungen erfolgen. Das Defizit an Ressourcen für die Datendokumentationen und die Vorbereitung für die Datenweitergabe muss bei der Entwicklung eines Workflows einkalkuliert werden, andernfalls wäre die Planung unrealistisch. Insofern bestehen neben den nicht-technischen Barrieren zum *data sharing*, wie bereits in Kapitel 3.2 beschrieben, durchaus ressourcenbedingte Hindernisse.

*„[A]lso mir ist auf jeden Fall bewusst, was alles gemacht wird und was eigentlich auch so theoretisch gefordert wird, was man hinsichtlich Daten machen sollte und so weiter. Aber die Praxis sieht glaube ich einfach anders aus.“<sup>231</sup>*

---

<sup>231</sup> Interview 3, Zeilen 662-664.

### 3.4 Zusammenfassung und Auswirkungen auf den Workflow

Die Notwendigkeit zur Aufbewahrung und Weitergabe von Forschungsdaten wird sowohl international als auch national auf wissenschaftspolitischer Ebene in diversen Policies und *Regeln zur guten wissenschaftlichen Praxis* belegt. Auch wenn die Formulierung konkreter Vorgaben zur praktischen Umsetzung zumeist den Forschungseinrichtungen überlassen bleibt, stellen die Rahmenbedingungen die Basis der allgemeinen Anforderungen an das FDM dar. Dabei sind nicht nur Nachhaltigkeit und Transparenz wichtige Aspekte, sondern auch die Berücksichtigung von disziplinären Besonderheiten. Aus diesem Grund lassen sich die Anforderungen in fachübergreifende und disziplinspezifische unterteilen.

Es wurden zunächst die beteiligten Akteure und deren Rollen identifiziert. Dabei zeigte sich, dass deren Tätigkeiten sich nicht eindeutig voneinander trennen lassen und insbesondere ein gemeinsames Verständnis über die Rolle des *data scientist* im deutschsprachigen Raum (noch) nicht vorhanden ist. Im Rahmen dieser Arbeit wurden die Akteure in die Anbieter (*data manager, data librarian*) und Nutzer (*data creator, data scientist*) der Infrastrukturen eingeteilt und die Beziehungen zwischen ihnen dargestellt. Es wurde ersichtlich, dass die Anforderungen sich nicht allein nach den Tätigkeiten der Akteure richten, sondern insbesondere nach den Bedürfnissen und Arbeitsweisen der Forschenden. Hierzu beinhalten die Ergebnisse der Literaturstudie von Feijen unter anderem, dass die Methoden und Werkzeuge des FDM sich leicht in den Arbeitsalltag integrieren lassen müssen, damit sie die für eine längerfristige Nutzung notwendige Akzeptanz auf Seiten der Wissenschaftlerinnen und Wissenschaftler erfahren. Daraus resultiert die Erkenntnis, dass die Anbieter der infrastrukturellen Serviceeinrichtungen, hauptsächlich die Bibliotheken, die Vermittlung von Datenkompetenz analog zur Informationskompetenz in ihr Dienstleistungsangebot integrieren und für die Inanspruchnahme ihrer Angebote werben müssen, mit dem Ziel, den Zeitaufwand der Forschenden so gering wie möglich zu halten.

Ein effizientes und erfolgreiches Forschungsdatenmanagement orientiert sich am gesamten Datenlebenszyklus und stellt somit einen fortlaufenden Prozess dar. Das Erstellen eines Datenmanagementplans erfolgt im Idealfall durch die Forschenden in

Zusammenarbeit mit den Infrastrukturanbietern und unter Nutzung von Vorlagen, die an die jeweiligen Projektbedingungen und Forschungsinhalte angepasst werden.

Derartige Vorlagen gibt es unter anderem in Form von Checklisten, die nicht nur im Rahmen universitärer Forschung eingesetzt werden und zudem disziplinspezifische Aspekte hervorheben.

Um die besonderen Anforderungen in den Sozialwissenschaften bei dem FDM zu berücksichtigen, ist die Kenntnis über Fachbegriffe und Forschungsabläufe erforderlich – nicht nur als Voraussetzung für die erfolgreiche Kommunikation zwischen allen Beteiligten. Es wurde erläutert, dass es sich bei quantitativen sozialwissenschaftlichen Daten häufig um personenbezogene, unter Umständen auch um sensible Daten handelt, die aus datenschutzrechtlichen Gründen nur in anonymisierter Form an Dritte weitergegeben werden dürfen. Das bedeutet, dass die verwendeten Anonymisierungsmethoden ebenfalls dokumentiert und aufbewahrt werden müssen, um den Ansprüchen an Transparenz und Nachvollziehbarkeit nachzukommen. Der Dokumentationsaufwand ist dadurch höher, auch weil Forschungsprimärdaten zumeist intellektuell produziert sind, im Gegensatz zu Daten manch anderer Fachdisziplinen (z.B. Daten als Ergebnis maschineller Messungen in der Geologie oder Meteorologie). Dadurch bedarf es sehr ausführlicher Beschreibungen zu Methodik, Forschungsdesign und Messinstrumenten.

Des Weiteren sollte neben informationswissenschaftlicher auch das Vorhandensein sozialwissenschaftlicher Standards und Methoden bekannt sein, wie bspw. Klassifikationsstandards zu Berufs- oder Länderkodierungen. Als Dokumentationsstandard wurden die zwei Spezifikationen *Codebook* und *Lifecycle* des DDI-Standards überblicksartig vorgestellt. Obwohl den Experteninterviews zufolge für die Erstellung von Dokumentationen eher Vorlagen als Klassifikationsstandards genutzt werden, bedeutet dies nicht, dass Standards keine oder nur eine untergeordnete Rolle für die Erstellung eines Workflows spielen sollten, sondern diesbezüglich Aufklärungsbedarf besteht. An dieser Stelle zeigt sich erneut die Diskrepanz zwischen den theoretischen Anforderungen und den begrenzten praktischen Umsetzungsmöglichkeiten.

Die Ergebnisse der Experteninterviews am WZB untermauern die der literaturbasierten Anforderungsanalyse im Allgemeinen. Die wichtigste Erkenntnis ist, dass die

emotionalen Hemmnisse (Angst vor Datenmissbrauch, Verlust der Wettbewerbsfähigkeit, Rufschädigung) die Bereitschaft zur Datenweitergabe tatsächlich in höherem Maße zu beeinflussen scheinen als technische Barrieren, welche in Form mangelnder Ressourcen vorhanden sind. Die bestehenden Rahmenbedingungen erzielen nicht die gewünschte Wirkung auf die Bereitschaft zur Datenweitergabe. Innerhalb der wissenschaftlichen Gemeinschaft und bei wissenschaftspolitischen Entscheidungsträgern muss ein Umdenken erfolgen, damit die Forderungen nach *data sharing* und langfristiger Verfügbarkeit von Forschungsdaten auch in der Praxis umsetzbar sind: Die erhobenen Daten müssen ebenso wie die Publikation der Ergebnisse als wissenschaftliche Leistung und Forschungsergebnis anerkannt werden und dementsprechend müssen die Arbeitsbedingungen gestaltet sein.

Da die Projektplanungen in der Regel keine zusätzlichen Ressourcen für das Datenmanagement bzw. explizite Tätigkeiten in Vorbereitung auf die Datenweitergabe (Metadatenvergabe, Anonymisierung personenbezogener Daten) enthalten, muss der Zeitaufwand der Forschenden so gering wie möglich gestaltet werden.

Die Kommunikation und Kooperation zwischen den beteiligten Akteuren kann in jeder Phase des Datenlebenszyklus erfolgen und ist im Hinblick auf die spätere Archivierung von wesentlicher Bedeutung, da sich Entscheidungen in einer frühen Phase unter Umständen negativ auf die spätere Archivierung auswirken. Im Vorfeld ist es wichtig, über die Vorteile des FDM und der Inanspruchnahme entsprechender Serviceangebote zu informieren und positive Beispiele aufzuzeigen (z.B. verbesserte Rechercheergebnisse durch Metadatenvergabe, internationale Vergleichbarkeit durch Nutzung internationaler Standards): „*The benefits of the support must be clearly visible – not in three years’ time, but now.*“<sup>232</sup>

Welche konkreten Auswirkungen haben die Ergebnisse der Anforderungsanalyse schließlich auf die Erstellung des Workflows?

Zu Beginn des Workflows ist zu klären, um welche Arten von Daten es sich handelt. Für Fremddaten, die nachgenutzt werden, müssen entsprechende Schnittstellen sowohl zu Recherchezwecken als auch für den Zugriff vorhanden sein. Daten, die selbst erhoben

---

<sup>232</sup> Feijen (2011), S. 4.

oder weiter bearbeitet werden, sind in qualitative und quantitative Daten zu unterscheiden. Dies ist ebenfalls abhängig von der Art der Auswertung, d.h. an dieser Stelle erfolgt die Nachfrage bei der forschenden Person. Da sich der Workflow auf das FDM quantitativer Daten beschränkt, wirkt sich die Unterscheidung unmittelbar auf die restlichen Aktivitäten aus: Für qualitative Daten muss eine alternative Verfahrensweise ermittelt werden. Zu beachten ist außerdem, dass mit der Auswertung wiederum neue Forschungsdaten entstehen. An dieser Stelle ist zu überlegen, wie mit diesen Daten umgegangen wird.

An der Erstellung von Dokumentationen und der Vergabe von Metadaten sind unterschiedliche Personen beteiligt, hauptverantwortlich ist letztlich die Projektleitung. Hier ist zu überlegen, welchem Akteur diese Tätigkeiten zugeschrieben werden und wie die Qualitätskontrolle erfolgen kann. Auch der Zeitpunkt oder Zeitrahmen muss klar definiert sein, da spätestens für die Weitergabe die Metadaten vollständig vorhanden sein müssen, Dokumentationen jedoch auch als Hilfsmittel *während* der Analyse- oder Auswertungsphase genutzt werden. Für die Erstellung von Dokumentationen können zudem Vorlagen nachgenutzt werden, die flexibel an die jeweiligen Bedingungen anpassbar sein müssen. An dieser Stelle muss die *Möglichkeit* zur Vorlagennutzung in den Workflow integriert werden, sie sollte aber nicht zwingender Bestandteil sein, denn: „*Researchers resist top-down and/or mandatory schemes.*“<sup>233</sup>

Im Hinblick auf die Weitergabe ist zu klären, ob aus Datenschutzgründen die Daten in Dokumentationen oder die Primärdaten anonymisiert werden müssen, d.h. die Metadaten dürfen nicht automatisch mit den Primärdaten zusammen weitergegeben werden. Falls eine Anonymisierung notwendig ist, stellt diese eine eigenständige, zusätzliche Aufgabe dar, die dementsprechend als eigener Workflow-Schritt definiert wird. Dies hat zwei Gründe: Zum einen wird der zusätzliche Arbeitsaufwand der Forschenden als solcher anerkannt (emotionales Hindernis), zum anderen kann dieser Schritt übersprungen werden, falls eine Weitergabe nicht erfolgt, wodurch der weitere Prozessablauf nicht beeinträchtigt würde.

---

<sup>233</sup> Ebenda, S. 4.



Die Datenweitergabe wird in der Regel mit zeitlicher Verzögerung erfolgen, aus diesem Grund muss der Workflow eine Embargo-Frist als Schranke beinhalten. In diesem Zusammenhang müssen die Daten registriert werden und Zugriffsrechte müssen geregelt sein. Für diese Aufgaben sehen die Experten den Datenmanager in der Verantwortung, der zugleich die Ansprechperson für die Forschenden bezüglich Rückfragen und zukünftiger Verfügbarkeit darstellt. Es ist zu prüfen, inwieweit dieses Rollenverständnis für die Erstellung des exemplarischen Workflows weiterhin bestehen kann.

Um der Forderung nach einer kontrollierten Weitergabe nachzukommen und somit Datenmissbrauch vorzubeugen, sollte die Weitergabe in jedem Fall in Rücksprache mit den Forschenden erfolgen, denn: *„Researchers must be in control of what happens to their data, who has access to it, and under which conditions. Consequently, they want to be sure that whoever is dealing with their data (data centre, library, etc.) will respect their interests.“*<sup>234</sup>

Neben der Datenweitergabe sollte auch die Datenlöschung als Workflow-Schritt mitberücksichtigt werden, falls Daten aus datenschutzrechtlicher Sicht nach Abschluss des Forschungsprojekts zu löschen sind und auch nicht in anonymisierter Form weitergegeben werden, wie beispielsweise Adressdaten von Interviewteilnehmern.

Die Ergebnisse der Anforderungsanalyse sollen die Grundlage für die Erstellung des exemplarischen Workflows bilden. Es konnten wichtige Erkenntnisse gewonnen werden, die für die praktische Umsetzung des Workflows maßgeblich sind.

Das folgende Kapitel 4 erläutert die Erstellung des exemplarischen Workflows und zeigt dafür notwendige Voraussetzungen auf. Die einzelnen Schritte werden aufgelistet und anhand des Prozessmodells visualisiert. Daraus gewonnene Erkenntnisse werden abschließend erläutert.

---

<sup>234</sup> Ebenda, S. 4.

## 4. Erstellung eines exemplarischen Workflows

Ziel dieser Arbeit ist die Erstellung eines exemplarischen Workflows für das FDM am WZB unter Berücksichtigung der Ergebnisse aus der Anforderungsanalyse. Mithilfe eines Workflows soll ein Prozess, der auf der fachlich-konzeptionellen Ebene innerhalb einer Einrichtung erarbeitet wurde, auf der operativen Ebene umgesetzt werden.<sup>235</sup> Dabei steht die Optimierung der Arbeitsabläufe unter organisatorischen Gesichtspunkten im Vordergrund. Im Allgemeinen stellt ein Workflow vorab definierte, gut strukturierte Abläufe dar, die weitestgehend automatisiert erfolgen bzw. durch Informationsverarbeitungssysteme unterstützt und mithilfe eines Workflow-Management-Systems kontrolliert werden. Zudem zeichnet er sich durch seinen repetitiven Charakter aus.<sup>236</sup>

Prozess- und Workflow-Management findet insbesondere in Unternehmen statt, da die Umsetzung ihrer strategischen Ziele ohne effiziente Abläufe und den effektiven Einsatz entsprechender Ressourcen nicht zu erreichen ist.<sup>237</sup> Doch auch wissenschaftliche Bibliotheken widmen sich vermehrt der Integration von Workflow-Systemen, wie bspw. bei der Durchführung von Digitalisierungsprojekten, in denen es um das Management von Digitalisaten ab dem Zeitpunkt ihrer Entstehung (Scannen) und Aufbereitung (Nachbearbeitung der Bilddateien) über die Erschließung (Vergabe von Struktur- und Metadaten) bis hin zu ihrer Sicherung und Online-Präsentation (Verlinkung) geht.<sup>238</sup>

Da auch das FDM allgemein die Verwaltung digitaler Daten (Objekte) beinhaltet, lassen sich die Vorteile eines Workflows dementsprechend übertragen:

*„Workflows are ideal for systematically, accurately, and repeatedly running routine procedures: managing data capture from sensors or*

---

<sup>235</sup> Vgl. Gadatsch (2013), S. 2/3.

<sup>236</sup> Vgl. Ebenda, S. 42 ff.

<sup>237</sup> Vgl. Ebenda, S. 1.

<sup>238</sup> Eine Digitalisierungssoftware mit integriertem Workflow-Management, die von zahlreichen Bibliotheken, Archiven und Museen nicht nur in Deutschland genutzt wird, ist *Goobi*. Nähere Informationen sind zu finden unter: [www.goobi.org](http://www.goobi.org) [Zugriff: 27.7.2013].

*instruments; cleaning, normalizing, and validating data; securely and efficiently moving and archiving data...<sup>239</sup>*

Systematisches und strukturiertes Vorgehen sind im FDM von großer Bedeutung. So weist die bereits in Kapitel 3 erwähnte DINI auf die Gestaltung eines Workflows hin:

*„Es empfiehlt sich, den Umgang mit Forschungsdaten auf institutioneller Ebene im Rahmen eines definierten Workflows zu gestalten. Dabei sollten die jeweiligen Akteure und ihre Rollen beschrieben werden. [...] Wissenschaftlichen Infrastruktureinrichtungen wie Bibliotheken, Daten- und Rechenzentren kommt dabei eine entscheidene Bedeutung zu: Sie unterstützen Forschende beim Umgang mit wissenschaftlichen Daten und etablieren lokale Dienste.<sup>240</sup>*

Die Etablierung lokaler Dienste folgt unter anderem der Forderung, Forschungsdaten für zehn Jahre an der Institution, an der sie entstanden sind, zu sichern. Eine langfristige, vertrauenswürdige Archivierung ist jedoch nicht von jeder Institution, an der Forschung betrieben wird, praktisch umsetzbar, noch wäre dies erstrebenswert.

*„[FDM] findet in dem Bewusstsein statt, dass lokale Lösungen Bestandteil einer übergreifenden Forschungsdateninfrastruktur sein müssen.<sup>241</sup>*

Voraussetzung für eine erfolgreiche übergreifende Forschungsdateninfrastruktur ist zunächst das erfolgreiche FDM auf institutioneller Ebene: Damit Forschungsdaten systematisch aufbereitet und erschlossen werden, um schließlich langfristig nutzbar sein zu können, bedarf es eines strukturierten Workflows. Der exemplarische Workflow für das FDM wird auf abstrakter Ebene anhand eines Geschäftsprozessdiagramms<sup>242</sup> modelliert und die einzelnen Schritte in Kapitel 4.2 erläutert. Die daraus resultierenden Erkenntnisse werden mit Blick auf das WZB in Kapitel 4.3 aufgezeigt. Zunächst werden die notwendigen Voraussetzungen genannt.

---

<sup>239</sup> Goble & De Roure (2009), S. 141.

<sup>240</sup> Dallmeier-Tiessen u.a. (2009), S. 7.

<sup>241</sup> Büttner u.a. (2011b), S. 14.

<sup>242</sup> Aus technischer Sicht handelt es sich nicht um einen *Workflow*, da im Rahmen dieser Arbeit operative Aspekte nicht in dem erforderlichen Umfang berücksichtigt werden können. Die Begriffe *Workflow* und *Prozess* werden jedoch erfahrungsgemäß in der Praxis oft synonym verwendet.

## 4.1 Voraussetzungen

Für die Entwicklung des exemplarischen Workflows sind sowohl formale als auch inhaltliche Voraussetzungen erforderlich.

Hinsichtlich der Darstellung des Prozesses ist zu entscheiden, welche Modellierungssprache verwendet werden soll. Für die Erstellung des exemplarischen Workflows wird in dieser Arbeit die standardisierte Sprache BPMN 2.0 (Business Process Model and Notation)<sup>243</sup> genutzt. Mit dieser können Geschäftsprozessdiagramme, Konversationsdiagramme, Kollaborationsdiagramme und weitere geschäftsrelevante Abläufe abgebildet werden. Die Prozessdiagramme dienen einerseits der internen Kommunikation und dem gemeinsamen Verständnis innerhalb einer Organisation, können aber auch als Programmiergrundlage dienen. Im Rahmen dieser Arbeit dient das Prozessdiagramm in erster Hinsicht der Veranschaulichung und mit Blick auf eine eventuelle Nachnutzung<sup>244</sup> zur Kommunikation.

Die BPMN wird nicht nur zur Prozessmodellierung in Unternehmen genutzt, es gibt bereits Praxisbeispiele aus dem informationswissenschaftlichen Bereich: So gab es bei der Deutschen Nationalbibliothek ein Projekt zur Einführung von BPMN<sup>245</sup> und im Rahmen einer Diplomarbeit wurde ein Konzept für die Optimierung der Fotoauftragsverwaltung an der Fotothek der Stiftung Preußische Schlösser und Gärten Berlin-Brandenburg unter Einbezug von BPMN-Modellen erstellt.<sup>246</sup> Zudem werden die Themen Prozessanalyse und Prozessmodellierung an der Fachhochschule Potsdam im Zusammenhang mit Records Management und digitaler Archivierung (Masterstudien-gang Informationswissenschaften) behandelt.

Für die Modellierung des Prozesses mit BPMN 2.0 wird der browserbasierte *Signavio Process Editor* in der Version 7.2.0<sup>247</sup> verwendet. Neben den standardisierten Notationsregeln<sup>248</sup> wurden die Modellierungsrichtlinien von Signavio beachtet,<sup>249</sup> welche

---

<sup>243</sup> Vgl. Freund; Rücker (2010), S. 9.

<sup>244</sup> Die beiliegende DVD enthält ebenfalls die im XML-Format exportierten Diagramme (\*.bpmn).

<sup>245</sup> Vgl. Diebel (2011).

<sup>246</sup> Vgl. Kastner (2010).

<sup>247</sup> Weitere Informationen siehe unter <http://www.signavio.com/de/> [Zugriff: 29.7.2013].

<sup>248</sup> Vgl. Allweyer (2009); Freund & Rücker (2010); Götz (2011).

durch die Einheitlichkeit des Modellierungsstils zur Verbesserung der Gesamtqualität des Modells beitragen. Die Richtlinien enthalten unter anderem Vorgaben zum Layout und zur Prozessstruktur.<sup>250</sup>

Im Folgenden wird für das Verständnis der Prozessdiagramme ein kurzer Überblick über die wichtigsten verwendeten Notationselemente gegeben.<sup>251</sup> Eine entsprechende Legende ist in Anhang E beigefügt.

Der Prozess findet innerhalb eines Pools statt, der in mehrere Lanes (Schwimmbahnen) unterteilt ist. Diese repräsentieren die beteiligten Rollen und sind in der Regel horizontal dargestellt. Zwischen Pools können Nachrichten ausgetauscht werden, jedoch nicht innerhalb eines Pools (siehe Abbildung 12).

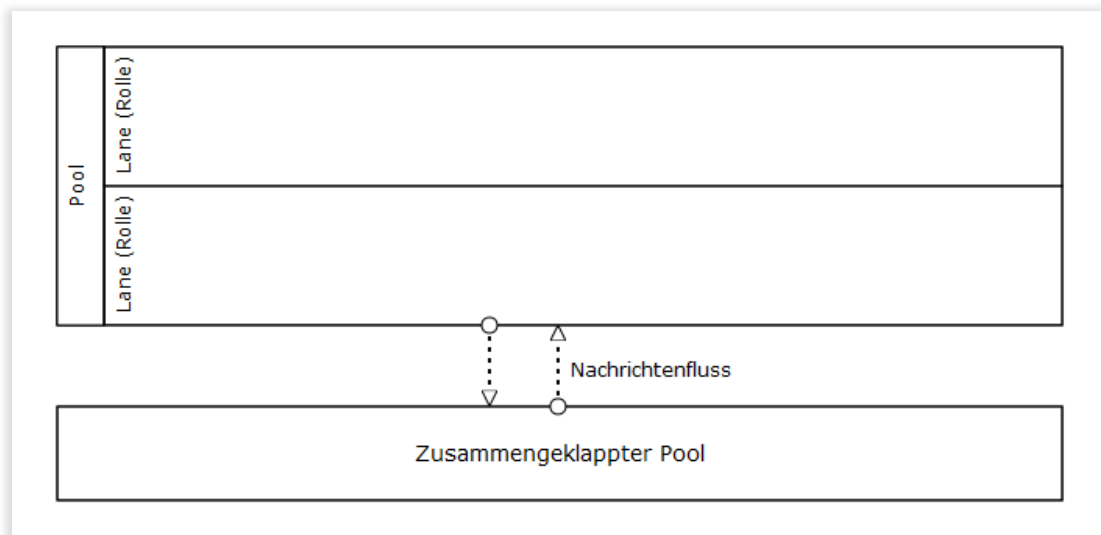


Abbildung 12: Pool und Nachrichtenfluss in BPMN 2.0<sup>252</sup>

Der Prozess besteht im Allgemeinen aus Aufgabenabfolgen, die mittels Sequenzflüssen verbunden sind und ist gekennzeichnet durch ein Start- und ein Endereignis (siehe Abbildung 13 auf Seite 86). Die Lane, in der die Aufgabe abgebildet ist, stellt zugleich

<sup>249</sup> Anmerkung der Autorin: Mithilfe der Syntax-Kontrolle und Signavio-Standardregeln konnten Fehler in den Prozessdiagrammen gefunden und vollständig korrigiert werden.

<sup>250</sup> Vgl. <http://www.modeling-guidelines.org/de/> [Zugriff: 11.8.2013].

<sup>251</sup> Für die Workflow-Modellierung wird sich auf ausgewählte Elemente beschränkt, um die Verständlichkeit nicht unnötig zu erschweren; dabei ist die Verwendung der Elemente mitunter abhängig vom individuellen Modellierungsstil. Eine gute Übersicht über alle Elemente ist zu finden unter <http://www.bpmb.de/index.php/BPMNPoster> [Zugriff: 30.7.2013], jedoch können nicht alle mit dem *Signavio Process Editor* modelliert werden, wie z.B. die Aktivitätsmarkierung mit einer Schleife für die Darstellung der wiederholten Ausführung.

<sup>252</sup> Quelle: Eigene Darstellung unter Nutzung des *Signavio Process Editors*.

die ausführende Rolle (verantwortliche Person) dar. Eine Aufgabe kann wiederum einen Unterprozess darstellen, zu dem ein eigenes Diagramm verlinkt werden kann.

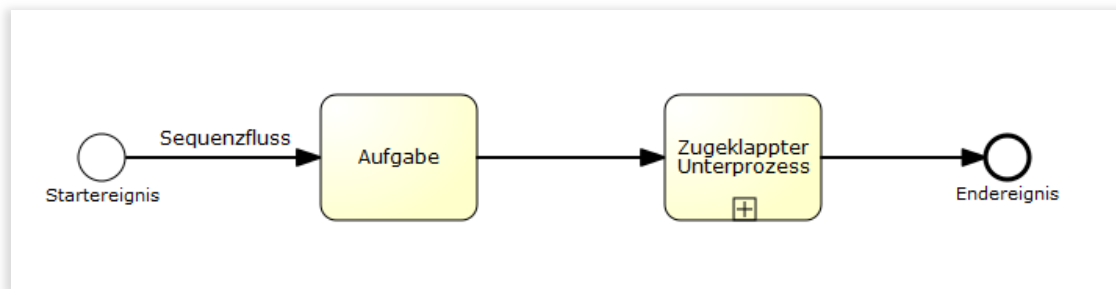


Abbildung 13: Prozessbeispiel in BPMN 2.0<sup>253</sup>

Innerhalb des Prozesses können unterschiedliche Ereignisse und Verzweigungen eingefügt werden, die den weiteren Verlauf beeinflussen, d.h. die nachfolgende Aufgabe wird nur bei Eintritt des Ereignisses oder Vorliegen der Bedingung ausgeführt. Es gibt eine Vielzahl an Ereignissen sowie unterschiedliche Verzweigungselemente (Gateways) (siehe Abbildung 14). Das datenbasierte exklusive Gateway (ODER) führt den Prozess an *einem* von mehreren möglichen Sequenzflüssen weiter, wohingegen das inklusive Gateway (nicht ausschließendes ODER) *einen oder mehrere* Ausgänge ermöglicht. Einem parallelen Gateway (UND) folgt der gleichzeitige Ausgang mehrerer Sequenzflüsse. Zwischenereignisse sind größtenteils entweder eintretend oder auslösend, wie z.B. das Erhalten und Versenden von Nachrichten.



Abbildung 14: Gateways und Zwischenereignisse in BPMN 2.0 (Auswahl)<sup>254</sup>

<sup>253</sup> Quelle: Eigene Darstellung unter Nutzung des *Signavio Process Editors*.

<sup>254</sup> Quelle: Eigene Darstellung unter Nutzung des *Signavio Process Editors*.

Des Weiteren können Datenobjekte (z.B. Datenspeicher und Datenobjekt) und Artefakte (z.B. IT-System, Textanmerkung) modelliert werden (siehe Abbildung 15). Mit ihnen lassen sich insbesondere Datenflüsse und Speichervorgänge übersichtlicher darstellen. Zum besseren Verständnis können neben den allgemeinen Beschriftungen der Elemente zusätzliche Kommentare eingefügt werden.

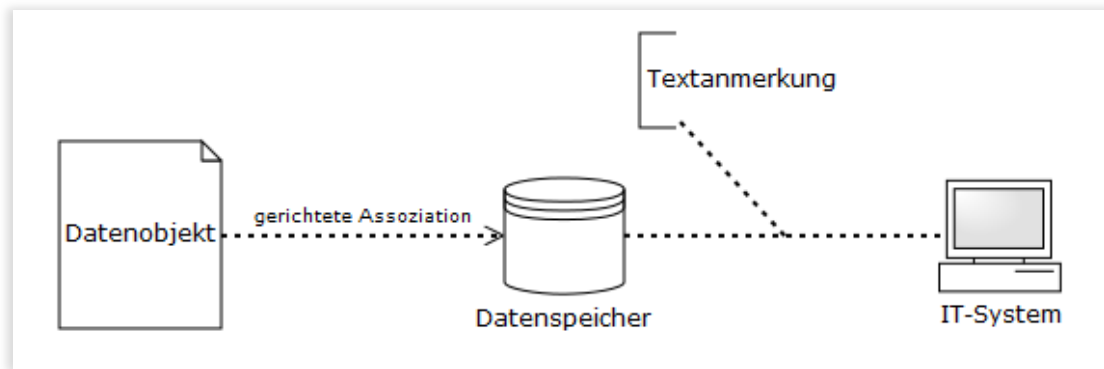


Abbildung 15: Datenobjekte und Artefakte in BPMN 2.0 (Auswahl)<sup>255</sup>

Eine inhaltliche Voraussetzung zur Erstellung des Workflows ist die Identifikation der beteiligten Akteure und Festlegung ihrer Tätigkeitsbereiche im Vorhinein. Auch für die formale Modellierung in BPMN 2.0 ist diese Festlegung notwendig, da ohne die Benennung der Rollen keine Zuordnung der Tätigkeiten erfolgen kann und der Prozess inhaltlich unvollständig wäre. Wie in Kapitel 3.1.1 aufgezeigt, erfolgt in dieser Arbeit die Unterteilung der Rollen in Anbieter und Nutzer der Infrastruktur. Dementsprechend wurden zwei Pools modelliert: Der obere beinhaltet den *data librarian* und *data manager* in jeweils einer Lane, der untere den *data scientist* und *data creator* zusammen in einer Lane, da diese Tätigkeiten von derselben Person übernommen werden kann.

Die Verantwortlichkeiten und Tätigkeiten im Forschungsdatenmanagement wurden im Rahmen der Anforderungsanalyse und insbesondere im Hinblick auf das WZB mithilfe der Experteninterviews herausgestellt. Es wurden bereits in Kapitel 3.4 daraus resultierende Auswirkungen auf den Workflow aufgezeigt. Diese Anforderungen gilt es entsprechend umzusetzen.

<sup>255</sup> Quelle: Eigene Darstellung unter Nutzung des *Signavio Process Editors*.

## 4.2 Workflow-Schritte

In dem Prozessdiagramm<sup>256</sup> laufen zwei Prozesse ab, die miteinander in Verbindung stehen: Der untere Pool visualisiert den Prozess bei den Forschenden (Nutzer der Infrastruktur), der obere zeigt den Ablauf auf Anbieterseite (Anbieter der Infrastruktur). Die Prozesse sind jeweils gekennzeichnet durch das Start- und Endereignis.

Der Prozess bei den Forschenden startet mit dem Beginn der Projektplanung, woran sich die erste Aufgabe **Konzept entwickeln** anschließt, die zwar als eigener Unterprozess angesehen werden kann, für die Zielstellung des Workflows jedoch nicht notwendigerweise modelliert werden muss. Ein Ergebnis dieser Aufgabe ist der Datenmanagementplan, dessen Inhalte teilweise in den Projektantrag fließen, wie es zum Teil von Förderorganisationen gefordert wird. Dies ist dargestellt durch ein Datenobjekt mit einer gerichteten Assoziation zur Aufgabe **Projektantrag stellen**.

Bei den Anbietern der Infrastruktur startet der Prozess damit, dass eine Information über den Beginn der Projektplanung eingegangen ist. Darauf folgt die Aufgabe **Konzeptentwicklung unterstützen**, welche parallel von dem *data manager* und dem *data librarian* ausgeführt wird (paralleles Gateway) und zugleich einen Unterprozess<sup>257</sup> bildet.

In diesem Unterprozess erfolgt die erste Anfrage, die von dem *data librarian* an den Nutzer geht, bezüglich der Art der Daten (auslösendes Nachrichtenzwischenereignis).<sup>258</sup> Ist die Rückmeldung eingetroffen (eintretendes Nachrichtenzwischenereignis), wird der Prozess fortgesetzt. Es folgt ein datenbasiertes exklusives Gateway (ODER-Gateway), an dem die Entscheidung getroffen wird, ob es sich um quantitative oder qualitative Daten handelt. Letzteres führt zum Abbruch des Prozesses, da für das Management qualitativer Daten ein alternativer Workflow genutzt werden muss. (Dies darf jedoch nicht zu der Schlussfolgerung führen, dass eine Unterstützung bei der Konzepterstellung nur in Bezug auf quantitative Daten erfolgt.) Müssen quantitative Daten verwaltet werden, folgt eine Anfrage zur Klärung einer zukünftigen Nachnutzung

<sup>256</sup> Siehe Anhang F und auf beiliegender DVD.

<sup>257</sup> Siehe Anhang F und auf beiliegender DVD.

<sup>258</sup> Die Nachrichtenflüsse sind für eine bessere Übersichtlichkeit in unterschiedlichen Farben dargestellt – ausgehend vom Anbieter sind sie grün, ausgehend vom Nutzer sind sie violett.



Dritter und einer Embargo-Frist (auslösendes Nachrichtenzwischenereignis beim *data librarian*).

Die Rückmeldung wird in der Aufgabe **DMP-Fragen klären** aufgefangen, welche von beiden Akteuren in dem Anbieter-Pool kooperativ ausgeführt wird. Hier werden bspw. Fragen zu Methodendesign (Instrumente der Datenerhebung), Verantwortlichkeiten, Qualitätssicherung, die Einhaltung datenschutzrechtlicher Bestimmungen (z.B. Anonymisierungsmethoden, Rechtemanagement) beantwortet sowie technische und finanzielle Anforderungen behandelt. Im Detail sind Fragen zur Dokumentation und Metadatenvergabe auf Studien- und Variablenebene (Formate, Standards) zu klären.<sup>259</sup> Das bedeutet, dass an dieser Stelle bereits besprochen wird, ob die Forschungsdaten ausschließlich archiviert oder auch zur Nachnutzung zur Verfügung gestellt werden dürfen, da sich diese Entscheidung zum gegebenen Zeitpunkt auf den Dokumentationsaufwand auswirkt. Allerdings läuft der Prozess auch ohne endgültige Entscheidung bezüglich einer Embargo-Frist weiter und endet schließlich mit einer Rückmeldung über den DMP beziehungsweise über das Konzept an die Forschenden (Nachrichtenendereignis).

Nachdem auf Nutzerseite der Projektantrag gestellt wurde, wird solange gewartet, bis das Projekt genehmigt ist (Signal-Zwischenereignis). Das bedeutet, der Prozess wird nicht weiter ausgeführt, wenn das Signal ausbleibt, weil der Antrag abgelehnt wurde.

Je nachdem, ob ein Bedarf an Fremddaten zur Nachnutzung für das eigene Forschungsprojekt bei den Nutzern vorhanden ist, wird ein Rechercheauftrag an den *data librarian* versendet, bei dem die Aufgabe **Recherche nach Fremddaten durchführen** erfolgt. Nach Abschluss der Aufgabe wird das Recherche-Ergebnis an den Nutzer gesendet. Ist nach der Aufgabe **Recherche-Ergebnis sichten** das Recherche-Ergebnis positiv, erfolgt eine Rückmeldung an den *data manager*, welcher die Tätigkeit **Zugriff auf Fremddaten ermöglichen** ausführt. Ist das Recherche-Ergebnis negativ und der Fremddatenbedarf auch weiter vorhanden, wird die Recherche erneut ausgeführt.

Ist die Rückmeldung über den Fremddatenzugriff erfolgt, wird als nächstes die Tätigkeit **Daten erheben** ausgeführt, die analog zur Konzeptentwicklung ebenfalls als eigener Unterprozess dargestellt werden könnte. Die zuvor angeforderten Fremddaten

---

<sup>259</sup> Vgl. Jensen (2011), S. [71] ff.

(Datenobjekt) gelangen über eine Schnittstelle aus dem externen Datenarchiv (Datenspeicher) zu den Forschenden, welche die Daten im Rahmen der Datenerhebung verwenden. Das Ergebnis der Datenerhebung sind die Rohdaten, die in der Aufgabe **Daten analysieren und auswerten** genutzt werden, ebenso wie die Fremddaten (gerichtete Assoziation).

Die Information über die Datenerhebung wird an den *data librarian* gesendet, was zur Aufgabe **Daten sichern (laufend)** beim *data manager* führt. Die Sicherung wird laufend vorgenommen, weil sie nicht innerhalb einer Aufgabe für den gesamten Prozess abgeschlossen werden kann. Bei der Analyse und Auswertung entstehen wiederum neue Forschungsdaten (Datenobjekt). Parallel dazu erfolgt die **Metadatenvergabe**, da die Ergebnisse der Experteninterviews unter anderem zeigten, dass Dokumentationen auch für die eigene Arbeit genutzt werden.

Die Metadatenvergabe ist als eigener Unterprozess<sup>260</sup> modelliert, das Starterereignis ist hier der Beginn der Datenanalyse und Auswertung. Zunächst werden die für die Archivierung erforderlichen Metadaten unter Nutzung von Vorlagen (für ausführliche Dokumentationen) und Standards (z.B. DDI) vergeben (Aufgabe **Pflichtfelder für Archivierung belegen**). Diese Pflichtfelder (z.B. Bezeichnung der Studie, Ersteller, Erhebungszeitraum, Kodierregeln, Methodenbericht, Originalfeldfragebogen bzw. das Messinstrument)<sup>261</sup> decken die Kernanforderungen für die Archivierung ab, sind aber für das Verständnis zukünftiger Nutzer in der Regel nicht ausreichend. Die Informationen zu den Pflichtfeldern erhalten die Nutzer von der Anbieterseite, nach Abschluss der Aufgabe wird eine Rückmeldung gegeben.

Es schließt sich eine Entscheidung an: Wenn zusätzlich optionale Metadatenfelder zu belegen sind (wie im Fall von *data sharing*) folgt die Aufgabe **ausführliche Metadaten vergeben**. Auch hier erhalten die Forschenden die notwendigen Informationen zu den Datenfeldern aus dem Anbieter-Pool (z.B. zusätzliche Berichte und Dokumentationen insbesondere über die Auswertung zur Erleichterung von Sekundär- und Re-Analysen).<sup>262</sup> Eventuell könnte die Metadatenvergabe in der Praxis auch teilweise vom

---

<sup>260</sup> Siehe Anhang F und auf beiliegender DVD.

<sup>261</sup> Vgl. Jensen (2012), S. 69.

<sup>262</sup> Vgl. Ebenda, S. 70.

*data librarian* in Absprache mit dem *data manager* übernommen werden, z.B. indem konkrete Vorschläge zu inhaltlichen Beschreibungen unterbreitet werden. Ohne endgültige Zustimmung der Forschenden sollte die Metadatenvergabe auf Studien- und Variablenebene jedoch nicht erfolgen.

Die Vergabe von Metadaten ist mitunter nicht nach einmaliger Ausführung abgeschlossen, es können auch im Zuge der weiteren Datenauswertung Metadaten bearbeitet oder hinzugefügt werden. Dies ist zwar nicht im Modell dargestellt, ist aber als impliziert aufzufassen. Dass die Metadatenvergabe von der Anbieterseite begleitet wird, ist im Gesamtprozessdiagramm durch die parallel auszuführende Aufgabe **Metadatenvergabe unterstützen** hervorgehoben.

Die Qualitätskontrolle wird sowohl von den Forschenden (inhaltlich) als auch auf Anbieterseite (formal) vorgenommen. Wenn die Aufgabe **Qualitätskontrolle (inhaltlich) durchführen** erfolgreich abgeschlossen und die Daten u.a. auf Plausibilität und logische Konsistenz geprüft wurden,<sup>263</sup> wird eine Nachricht an den *data librarian* gesendet. Andernfalls werden die Aufgaben der Datenanalyse/-auswertung und Metadatenvergabe solange wiederholt, bis die Kontrolle erfolgreich ist (ODER-Gateways). Der *data librarian* kommuniziert an dieser Stelle die Bitte um die Datenauswahl hinsichtlich der Weitergabe sowie um die Entscheidung bezüglich einer Embargo-Frist. Während auf Nutzerseite die Aufgabe **Daten bewerten / Auswahl treffen** ausgeführt wird, findet die formale Qualitätskontrolle durch den *data manager* und *data librarian* statt.

Der Unterprozess **Qualitätskontrolle (formal) durchführen**<sup>264</sup> beinhaltet die formale Prüfung der Metadaten, sowohl bezüglich technischer (z.B. Erhaltungsmetadaten, Formate, Gewährleistung von Authentizität und Integrität der Daten durch Prüfsummen und Versionsverwaltung) als auch nicht-technischer Fehler (z.B. Vollständigkeit der Metadaten und Dokumentationen, Überprüfen von Variablenwerten auf Zulässigkeit und formaler Übereinstimmung mit Fragen und Antwortkategorien des Fragebogens, formale Prüfung der Identifikationsnummern).<sup>265</sup> Beide parallel ausgeführten Tätigkeiten erfolgen unter der Nutzung von Standards bzw. Vorlagen (Datenobjekt).

---

<sup>263</sup> Vgl. Ebenda, S. 32.

<sup>264</sup> Siehe Anhang F und auf beiliegender DVD.

<sup>265</sup> Vgl. Jensen (2012), S. 35 ff.

Ist ein Fehler gefunden, muss dieser korrigiert werden. Dabei ist zu beachten, ob es sich um einen nicht-technischen (Aufgabe des *data librarian*) oder technischen (Aufgabe des *data manager*) handelt. Sind beide Fehlerarten aufgetreten, können die Aufgaben **Fehler korrigieren** auch parallel erledigt werden (inklusive Gateway bzw. nicht-ausschließende ODER-Verzweigung). Falls an dieser Stelle inhaltliche Fehler erkannt werden, die im Rahmen der inhaltlichen Qualitätskontrolle durch die Forschenden übersehen wurden, könnten eventuell auch diese in der Praxis nach entsprechender Rücksprache von dem *data librarian* behoben werden.

Die Aufgabe impliziert, dass die Fehlerkorrektur solange ausgeführt wird, bis alle Fehler beseitigt sind; auf die Modellierung einer weiteren Verzweigung wurde daher verzichtet. Nach Abschluss wird eine Rückmeldung an die Forschenden (Nachrichtenebene) gesendet, die jedoch nicht mit einer neuen Aufgabe oder einem nachfolgenden Ereignis verbunden ist (Nachrichtenfluss endet an der Pool-Grenze), da diese Rückmeldung den weiteren Prozessverlauf nicht beeinflusst.

Die Forschungsdaten als Produkt der Datenauswertung sowie die Metadaten werden vom *data manager* im lokalen Datenspeicher abgelegt, wie die Aufgabe **Forschungs- und Metadaten sichern (laufend)** beschreibt.

Der Prozess läuft weiter, wenn die Entscheidung über die Datenweitergabe getroffen wurde und die Mitteilung den *data librarian* erreicht hat. Sollen die Daten archiviert werden (nicht zwingend im Sinne einer Langzeitarchivierung), übernimmt dieser die Aufgabe **Daten registrieren (DOI-Vergabe)**, der *data manager* führt die Aufgabe **Daten löschen** durch, wobei an dieser Stelle bspw. mit der informationstechnologischen Abteilung kommuniziert wird (zusätzlicher Beteiligter), um die Daten (ggf. Rohdaten, Meta- oder auch Paradata) im lokalen Speicher und Backup-System auf Bit-Ebene vollständig zu löschen. Die Ausführung dieser beiden Tätigkeiten schließt sich nicht gegenseitig aus (inklusive Gateway), da nicht zwingend alle Daten an das Datenarchiv weitergegeben werden, selbst wenn kein *data sharing* vorgesehen ist.

Die Frage nach *data sharing* ist auf Nutzerseite sehr wichtig, da die Antwort mitunter entscheidet, ob die Daten bei Bedarf anonymisiert werden müssen. Sollen Daten nach Projektende geteilt werden und ist Anonymisierungsbedarf (z.B. aufgrund personen-

bezogener Daten) vorhanden, folgen die Aufgaben **Daten anonymisieren** und **Anonymisieren dokumentieren**. Die Dokumentation über die verwendeten Anonymisierungsmethoden (Datenobjekt) wird an den *data manager* weitergeleitet, der die Aufgabe **Dokumentation über Anonymisierung sichern** durchführt.

Nachfolgend kann die Datenübergabe an die Archivierungseinrichtung (rot markierter zusätzlicher Benutzer) eingeleitet werden (**Ingest vorbereiten**), bei der unter anderem auf Vollständigkeit, Lesbarkeit und Virenfreiheit geprüft wird, und ggf. Datenformate in das gewünschte Archivierungsformat konvertiert werden.<sup>266</sup> Im Rahmen dieser Tätigkeit wird ein letztes Mal mit den Forschenden bezüglich der Embargo-Frist kommuniziert, unter der Bedingung, dass *data sharing* erfolgen soll und die Entscheidung über die Frist nicht bereits zu einem früheren Zeitpunkt getroffen wurde. Ist dies nicht der Fall, wird die letzte Aufgabe im Prozess auf Nutzerseite durchgeführt: **Entscheidung über Embargo-Frist treffen**.

Ist diese Aufgabe abgeschlossen, wird die Entscheidung dem *data manager* mitgeteilt, damit die Information über die Embargo-Frist mit der darauffolgenden Aufgabe **Ingest durchführen** an das Datenarchiv mit allen anderen Forschungs- und Metadaten (SIP) übergeben und im Archivsystem entsprechend hinterlegt werden kann. Der Prozess endet mit einer Benachrichtigung zur Datenübergabe an die Nutzer der Infrastruktur.

Für die Archivierungseinrichtung entstehen neue Aufgaben bezüglich der (OAIS-konformen) Archivierung, die in einem weiteren Prozess modelliert werden können.<sup>267</sup> Mithilfe des Prozessmodells sollten insbesondere die erforderlichen Kommunikationsflüsse und Tätigkeiten im Rahmen des Forschungsdatenmanagements von der Entstehung der Daten bis zur Archivierung visualisiert werden. Die langfristige Verfügbarkeit digitaler Forschungsdaten und die Wahrung ihrer Authentizität und Integrität ist nur durch eine für diese Aufgaben spezialisierte Archivierungseinrichtung gewährleistet.<sup>268</sup>

---

<sup>266</sup> Vgl. Ebenda, S. 71 ff.

<sup>267</sup> Dies ist jedoch nicht Bestandteil der vorliegenden Arbeit.

<sup>268</sup> Im Hinblick auf das WZB käme das Modell des Datenarchivs infrage, da dieses die Verantwortung für die langfristige Verfügbarkeit der Daten aus zumeist zeitlich begrenzten Projekten übernehmen könnte. (Siehe auch Kapitel 2.3 für die Vor- und Nachteile von Datenarchiven vs. Forschungsdatenzentren.)

### 4.3 Erkenntnisse

Nachdem das Prozessdiagramm auf einer eher abstrakten Ebene erstellt wurde, werden in diesem Kapitel die sich daraus ableitenden Erkenntnisse für den FDM-Workflow am WZB beschrieben.

Ein Testdurchlauf des Workflows anhand eines Beispieldatensatzes kann am WZB im Rahmen dieser Arbeit aus Zeitgründen nicht vorgenommen werden. Jedoch ergab ein Gespräch mit dem Leiter der Abteilung Datenmanagement, dass die einzelnen Workflow-Schritte plausibel dargestellt wurden.<sup>269</sup>

Bei Betrachtung des Workflows fällt auf, dass die Kommunikation fast ausschließlich zwischen den Nutzern und dem *data librarian* verläuft. Der *data librarian* stellt die untere Lane im Anbieter-Pool dar, in diesen gehen die Nachrichtenflüsse. Auf diese Weise zeigt sich die Schnittstellenfunktion der Informationswissenschaften zwischen den Forschenden und der technischen Anbieterseite (DV-Abteilung). Allgemein zeigt das Prozessdiagramm den *notwendigen* Informationsaustausch, alle *möglichen* Nachrichtenflüsse wurden der Übersichtlichkeit wegen nicht modelliert.

Hinsichtlich der Embargo-Frist sollte visualisiert werden, wann mit den Forschenden Rücksprache zu halten ist bzw. zu welchem Zeitpunkt spätestens die Frist festgelegt sein sollte. Modelliert wurden diesbezüglich Nachrichtenflüsse in drei Phasen des Datenlebenszyklus: während der Planungsphase (Konzepterstellung), nach Abschluss der Datenauswertung und unmittelbar vor der Weitergabe der Daten an die Archivierungseinrichtung (Ingest-Vorbereitung). Für die Hinterlegung der Embargo-Frist im System sollte die Möglichkeit bestehen, Datensätze partiell für die Nachnutzung bereit zu stellen, sodass z.B. nicht alle Variablen ab demselben Zeitpunkt verfügbar sind.<sup>270</sup> Da diese Unterteilung den Prozessablauf auf Anbieterseite nicht beeinflusst, wurde auf deren Modellierung ebenfalls verzichtet.

Zudem werden die Aufgaben vom *data librarian* und dem *data manager* größtenteils kooperativ ausgeführt (siehe parallele Gateways im Prozessdiagramm). Dies entspricht

---

<sup>269</sup> Quelle: Persönliches Gespräch mit R. Habich, 26.07.2013.

<sup>270</sup> Quelle: Persönliches Gespräch mit R. Habich, 26.07.2013.

dennoch der internen Aufgabenteilung: Der *data manager* ist im Hinblick auf technische Fragen (Datensicherung, Archivierung) involviert, die Verantwortlichkeiten des *data librarian* konzentrieren sich auf nicht-technische Aspekte (Erschließungsfragen, Recherche).

Der Zeitaufwand bei den Forschenden ist im Rahmen des FDM so gering wie möglich zu halten. Mithilfe eines strukturierten Workflows soll dieses Ziel erreicht werden. Die Tätigkeiten bei den Nutzern der Infrastruktur sind zwar zahlreich, stellen jedoch keinen extra Zeitaufwand für das FDM dar, denn die Aufgaben entsprechen im Allgemeinen denen innerhalb des Forschungsablaufs bzw. sind nicht neu hinzugekommen.

Neben der nicht vorhandenen technischen Infrastruktur, welche im Rahmen des geplanten Kooperationsprojekts *SowiDataNet* entwickelt werden soll, fehlen weitere Voraussetzungen zur praktischen Umsetzung des Prozessmodells am WZB:

Die Aufgabenteilung kann nicht in der Form erfolgen, da es am WZB keinen *data librarian* gibt. Zudem ist der *data manager* lt. Prozessmodell mit eher technischen Aufgaben betraut sowie für die Datensicherung zuständig (in Kooperation mit der IT-Abteilung) und führt vorbereitende Maßnahmen für die Archivierung durch. Zurzeit findet jedoch keine einheitlich geregelte Übergabe der am WZB entstehenden Forschungsdaten, inklusive der erforderlichen Metadaten, an eine Archivierungseinrichtung statt. Daher müsste auch die Rolle des *data managers* überdacht und an die Anforderungen des FDM lt. Workflow angepasst werden.

Nicht nur technisch-organisatorische Voraussetzungen sind zu erfüllen, auch aus methodischer Sicht existieren Lücken: Für die Metadatenvergabe wird kein einheitlicher Standard verwendet, Vorlagen werden von Forschenden abteilungsintern erstellt und nachgenutzt, die formale Qualitätskontrolle kann demzufolge nicht in der gewünschten (wie im Prozessdiagramm dargestellten) Form durchgeführt werden. Bereits bei der Konzeptentwicklung sind diese Defizite erkennbar. Da die Erstellung eines Datenmanagementplans für Projektanträge nicht zwingend gefordert wird, werden nicht alle notwendigen Überlegungen bereits in der Planungsphase angestellt. Dass sich Entscheidungen in einer frühen Phase des Datenlebenszyklus auch auf den späteren Ab-

lauf auswirken, wurde beispielsweise anhand der Modellierung des Anonymisierungsbedarfs bei geplantem *data sharing* verdeutlicht.

Mit der technischen Infrastruktur muss zugleich ein geeignetes Workflow-Management-System integriert werden, damit der Prozessablauf auf operativer Ebene kontrolliert und für alle Beteiligten transparent gestaltet werden kann (z.B. bezüglich des Zeitpunkts zur erforderlichen Kommunikation). Der Nachrichtenaustausch könnte dann über das System stattfinden, d.h. die Forschenden müssen nicht zwingend eine E-Mail versenden, wenn es bspw. um die Information über die abgeschlossene Metadatenvergabe geht.

Wie die Anforderungsanalyse ergeben hat, müssen sich die Werkzeuge des FDM leicht in den Arbeitsalltag der Wissenschaftlerinnen und Wissenschaftler integrieren lassen. Dahingehend ist zu prüfen, welches System auf dem Markt die erforderlichen und gewünschten Funktionalitäten besitzt und für das WZB infrage kommt. Dies setzt die Erstellung eines Anforderungskatalogs voraus, für die das Prozessmodell dienlich sein kann.

Sind die Vorkehrungen für die Nutzung der technischen Infrastruktur getroffen, sollten Methoden des Veränderungsmanagements (Akzeptanzmanagements) zum Einsatz kommen. Wenn strukturelle Veränderungen in einer Einrichtung vorgenommen werden, müssen alle Beteiligten nicht nur informiert sondern auch zu deren Akzeptanz motiviert werden. Anhand von Informationsmaterial und Schulungsangeboten (Kennenlernen der Ansprechpartner und Aufgaben, Umgang mit neuen Werkzeugen) wird auf die bevorstehende Einführung eines neuen Systems aufmerksam gemacht, wobei der Nutzen für die Forschenden dabei jederzeit im Vordergrund stehen sollte.



## 5. Fazit

Die Aufbewahrung von Forschungsdaten für Sekundär- und Re-Analysen ist Bestandteil guter wissenschaftlicher Praxis und wird von wissenschaftspolitischen Entscheidungsträgern und Förderorganisationen sowohl auf nationaler als auch internationaler Ebene gefordert. Um diese Forderung erfüllen zu können, müssen entsprechende Infrastrukturen für das Management von Forschungsdaten ab ihrer Entstehung bis zu ihrer (langfristigen) Archivierung geschaffen bzw. bestehende Informationsinfrastrukturen ausgebaut werden.

Im Rahmen des Forschungsdatenmanagements gilt es, diesbezüglich fachspezifische Anforderungen zu berücksichtigen, da Forschungsdaten je nach Fachdisziplin unterschiedlich wahrgenommen werden. Auch der Datenlebenszyklus stellt sich, bedingt durch den Forschungsablauf, mitunter etwas anders dar.

Für die empirische Sozialforschung stellen Datenarchive und Forschungsdatenzentren sozialwissenschaftliche Forschungsdaten zur Nachnutzung bereit. Auf diese Weise ist der Zugang zu umfangreichen Datenquellen, wie bspw. Daten der Amtlichen Statistik oder aus dem ALLBUS, gewährleistet. Im Gegensatz dazu werden Forschungsdaten aus kleineren Projekten nicht in dem gewünschten Maße zur Verfügung gestellt.

Forschungseinrichtungen, an denen Daten erhoben und ausgewertet werden, sollen diese mindestens zehn Jahre aufbewahren. Das bedeutet, dass ein geeigneter Workflow für das Forschungsdatenmanagement erforderlich ist, damit Daten bereits auf institutioneller Ebene systematisch aufbereitet und erschlossen werden können, um schließlich zur Nachnutzung zur Verfügung zu stehen.

Um einen institutionellen Workflow erstellen zu können, müssen vorab die spezifischen Anforderungen analysiert werden. Da bei Forschenden oftmals Hemmnisse bestehen, ihre Daten mit anderen zu teilen, ist es unerlässlich, deren Bedürfnisse zu kennen um die Anforderungen an das FDM an ihnen auszurichten.

Die im Rahmen dieser Arbeit durchgeführte Anforderungsanalyse zeigte sowohl allgemeine als auch besondere Anforderungen in den Sozialwissenschaften auf. Die Ergebnisse aus den Experteninterviews sind diesbezüglich von signifikanter Bedeutung,

da sie die Gestaltung des Workflows maßgeblich beeinflussen. So wurden unter anderem die Nutzung von Vorlagen und Dokumentationen für die Arbeit der Forschenden berücksichtigt und die Anonymisierung im Falle von *data sharing* als eigenständiger Arbeitsschritt visualisiert.

Das Ergebnis dieser Arbeit ist ein institutioneller Workflow, der exemplarisch für das WZB erstellt wurde und projektübergreifend angewendet werden kann. Im Rahmen des Workflows wurden Akteure und Verantwortlichkeiten festgelegt, um Forschungsdaten ab dem Zeitpunkt ihrer Entstehung bis zur Archivierung systematisch erfassen und aufbereiten zu können. Dies setzt die Kommunikation aller Beteiligten voraus, die auch innerhalb des Workflows laufend stattfinden muss. Aufgrund mangelnder Ressourcen auf Seite der Forschenden sind der Zeit- und Arbeitsaufwand bei der (inhaltlichen) Erschließung und Aufbereitung ihrer Daten, vor allem im Hinblick auf die datenschutzbedingte Anonymisierung, so gering wie möglich zu halten.

Für die praktische Umsetzung des Workflows müssen die erforderlichen Ressourcen zur Verfügung stehen, um nicht nur die für ein erfolgreiches Forschungsdatenmanagement notwendige technische Infrastruktur aufzubauen – Forschungsdaten bedürfen zudem einer laufenden Pflege durch kompetentes Personal, welches sich durch ein Grundverständnis für sozialwissenschaftliche Begriffe und Forschungsabläufe auszeichnet und zugleich informationswissenschaftliche Werkzeuge und Methoden einsetzt.

Abschließend kann folgende Empfehlung ausgesprochen werden: Der Umgang mit Forschungsdaten sollte in einer institutionellen Policy festgelegt werden. Auch die interne Kommunikation bezüglich der Anforderungen und Verantwortlichkeiten im Rahmen des Forschungsdatenmanagements kann zum Abbau bestehender Hemmnisse für das *data sharing* beitragen. Langfristig kann dies nur gelingen, wenn die gewonnenen Forschungsdaten als wissenschaftlicher Output anerkannt werden und in ihrer Wertschätzung mit der Publikation der Ergebnisse gleichgestellt sind.

## 6. Quellenverzeichnis

### Literaturquellen

**Allianz der deutschen Wissenschaftsorganisationen (2010):** Grundsätze zum Umgang mit Forschungsdaten.

URL: [http://www.allianzinitiative.de/fileadmin/user\\_upload/Home/Video/Grunds%C3%A4tze%20Umgang%20mit%20Forschungsdaten.pdf](http://www.allianzinitiative.de/fileadmin/user_upload/Home/Video/Grunds%C3%A4tze%20Umgang%20mit%20Forschungsdaten.pdf) [Zugriff: 17.4.2013]

**Allmendinger, J. (Hrsg.) (2013):** Bericht 2012.

Berlin: WZB (Wissenschaftszentrum Berlin für Sozialforschung).

URL: [http://www.wzb.eu/sites/default/files/publikationen/wzb\\_berichte/wzb-bericht-2012.pdf](http://www.wzb.eu/sites/default/files/publikationen/wzb_berichte/wzb-bericht-2012.pdf) [Zugriff: 12.6.2013]

**Allweyer, T. (2009):** BPMN 2.0 Business Process Model and Notation : Einführung in den Standard für die Geschäftsprozessmodellierung. Norderstedt: Books on Demand.

**Aschenbrenner, A.; Neuroth, H. (2011):** Forschungsdaten-Repositoryen.

In: Büttner, S.; Hobohm, H.-C.; Müller, L. (Hrsg.): *Handbuch*

*Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen, S. [101] - 114.

URL: [http://opus.kobv.de/fhpotsdam/volltexte/2011/232/pdf/2.5\\_Forschungsdaten\\_Repositoryen.pdf](http://opus.kobv.de/fhpotsdam/volltexte/2011/232/pdf/2.5_Forschungsdaten_Repositoryen.pdf) [Zugriff: 4.4.2013]

**Atteslander, Peter (2010):** Methoden der empirischen Sozialforschung. Berlin: Erich Schmidt.

**Ball, A. (2012):** Review of data management lifecycle models.

Bath, UK: University of Bath.

URL: <http://opus.bath.ac.uk/28587/1/redm1rep120110ab10.pdf> [Zugriff: 11.4.2013]

**Bell, G.; Hey, T.; Szalay, A. (2009):** Computer science : beyond the data deluge.

In: *Science*, Band 323, Ausgabe 5919, 2009, S. 1297 – 1298.

**Bogner, M.; Menz, W. (2005):** Das theoriegenerierende Experteninterview :

Erkenntnisinteresse, Wissensformen, Interaktion ; zur Einführung in ein unübersichtliches Problemfeld.

In: Bogner, A.; Littig, B.; Menz, W. (Hrsg.): *Das Experteninterview : Theorie, Methode, Anwendung*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 7 – 70.

**Bortz, J.; Döring, N. (2006):** Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Berlin: Springer.

**Bove, H.; Habich, R.; Nix, S.; Quast, A. (2011):** Überlegungen zu den Forschungsinfrastrukturen in den Sozial- und Geisteswissenschaften : die Sicht von Serviceeinrichtungen aus dem Bereich „Bibliotheken und wissenschaftliche Information“. Berlin: RatSWD.

URL: [http://www.ratswd.de/download/RatSWD\\_WP\\_2011/RatSWD\\_WP\\_183.pdf](http://www.ratswd.de/download/RatSWD_WP_2011/RatSWD_WP_183.pdf) [Zugriff: 19.3.2013]

**Büttner, S.; Hobohm, H.-C.; Müller, L. (Hrsg.) (2011a):** Handbuch Forschungsdatenmanagement. Bad Honnef: Bock + Herchen.  
URL: <http://opus.kobv.de/fhpotsdam/volltexte/2011/241/pdf/HandbuchForschungsdatenmanagement.pdf> [Zugriff: 7.3.2013]

**Büttner, S.; Hobohm, H.-C.; Müller, L. (2011b):** Research Data Management.  
In: Büttner, S.; Hobohm, H.-C.; Müller, L. (Hrsg.): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen, S. [13] – 24.  
URL: [http://opus.kobv.de/fhpotsdam/volltexte/2011/225/pdf/1.1\\_Research\\_Data\\_Management.pdf](http://opus.kobv.de/fhpotsdam/volltexte/2011/225/pdf/1.1_Research_Data_Management.pdf) [Zugriff: 4.4.2013]

**Dallmeier-Tiessen, S.; Dobratz, S.; Gradmann, S.; Horstmann, W.; Kleiner, E.; Pampel, H.; Schirmbacher, P.; Schmidt, B.; Scholze, F.; Schulze, M. (2009):** Positionspapier Forschungsdaten. DINI - Deutsche Initiative für Netzwerkinformation e.V., Arbeitsgruppe „Elektronisches Publizieren“.  
URL: <http://pub.uni-bielefeld.de/download/2522749/2522765> [Zugriff: 8.5.2013]

**DataCite (2009):** DataCite statutes : final version.  
URL: <http://www.datacite.org/docs/datacite-statutes-final.pdf> [Zugriff: 7.8.2013]

**DCC (Digital Curation Centre) (2008):** The DCC Curation Lifecycle Model.  
URL: <http://www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf> [Zugriff: 4.8.2013]

**DFG (Deutsche Forschungsgemeinschaft) (2009):** Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten.  
URL: [http://www.dfg.de/download/pdf/foerderung/programme/lis/ua\\_inf\\_empfehlungen\\_200901.pdf](http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf) [Zugriff: 17.4.2013]

**DFG (Deutsche Forschungsgemeinschaft) (2011):** Ergänzender Leitfaden für die Beantragung von Langfristvorhaben. DFG-Vordruck 1.01 – 10/11.  
URL: [http://www.dfg.de/formulare/1\\_01/1\\_01.pdf](http://www.dfg.de/formulare/1_01/1_01.pdf) [Zugriff: 18.4.2013]

**DFG (Deutsche Forschungsgemeinschaft) (2012):** Leitfaden für die Antragstellung : Projektanträge. DFG-Vordruck 54.01 – 10/12.  
URL: [http://www.dfg.de/formulare/54\\_01/54\\_01\\_de.pdf](http://www.dfg.de/formulare/54_01/54_01_de.pdf) [Zugriff: 18.4.2013]

**DGS (Deutsche Gesellschaft für Soziologie) (2012):** [Hauptprogramm] 36. Kongress der deutschen Gesellschaft für Soziologie ; 1. - 5. Oktober. Bochum ; Dortmund: DGS.  
URL: [http://www.soziologie.de/uploads/media/Hauptprogramm\\_DGS2012\\_02.pdf](http://www.soziologie.de/uploads/media/Hauptprogramm_DGS2012_02.pdf) [Zugriff: 23.5.2013]

**Diebel, C. (2011):** Einführung von Geschäftsprozessmethoden bei der Deutschen Nationalbibliothek : Entwicklung einer Arbeitshilfe und eines Style-Guides für den Einsatz von BPMN.  
In: Oßwald, A.; Meinhardt, H.; Rösch, H.; Tappenbeck, I. (Hrsg.): *MALIS-Praxisprojekte 2011 : Projektberichte aus dem berufsbegleitenden Masterstudiengang Bibliotheks- und Informationswissenschaft der Fachhochschule Köln*. Wiesbaden: Dinges & Frick, S. S. [71] – 90.

- Donnelly, M.; Jones, S. (2009):** DCC data management plan content checklist : draft template for consultation. Digital Curation Centre.  
URL: <http://www.era.lib.ed.ac.uk/handle/1842/3343> [Zugriff: 9.5.2013]
- Dresing, T.; Pehl, T. (2012):** Praxisbuch Transkription : Regelsysteme, Software und praktische Anleitungen für qualitative ForscherInnen. Marburg: Eigenverlag.  
URL: <http://www.audiotranskription.de/Praxisbuch-Transkription.pdf>  
[Zugriff: 6.6.2013]
- Europäische Kommission (Hrsg.) (2012):** Wissenschaftliche Daten: freier Zugang zu Forschungsergebnissen wird Innovationskapazität der EU stärken.  
[Pressemitteilung, 17. Juli].  
URL: [http://europa.eu/rapid/press-release\\_IP-12-790\\_de.pdf](http://europa.eu/rapid/press-release_IP-12-790_de.pdf) [Zugriff: 26.6.2013]
- Eynden, V. v. d.; Corti, L.; Woollard, M.; Bishop, L.; Horton, L. (2011):** Managing and sharing data: best practice for researchers.  
Colchester: UK Data Archive, University of Essex.  
URL: <http://data-archive.ac.uk/media/2894/managingsharing.pdf> [Zugriff: 12.4.2013]
- Feijen, M. (2011):** What researchers want : a literature study of researchers' requirements with respect to storage and access to research data.  
Utrecht: Surf Foundation.  
URL: <http://www.surf.nl/en/publicaties/Pages/Whatresearcherswant.aspx>  
[Zugriff: 19.3.2013]
- Freund, J.; Rücker, B. (2010):** Praxishandbuch BPMN 2.0. München ; Wien: Hanser.
- Fuchs-Heinritz, W.; Klimke, D.; Lautmann, R.; Rammstedt, O.; Stäheli, U.; Weischer, C.; Wienold, H. (Hrsg.) (2011):** Lexikon zur Soziologie.  
Wiesbaden: Springer Verlag für Sozialwissenschaften.
- Gadatsch, A. (2013):** Grundkurs Geschäftsprozess-Management : Methoden und Werkzeuge für die IT-Praxis ; eine Einführung für Studenten und Praktiker. 7. Aufl.  
Wiesbaden: Vieweg + Teubner.
- Goble, C.; De Roure, D. (2009):** The impact of workflow tools on data-centric research.  
In: Hey, T.; Tansley, S.; Tolle, K. (Hrsg.): *The fourth paradigm : data-intensive scientific discovery*. Redmond, Wash.: Microsoft Research, S. 137 – 145.  
URL: [http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th\\_paradigm\\_book\\_part3\\_goble\\_deroure.pdf](http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_part3_goble_deroure.pdf)  
[Zugriff: 21.6.2013]
- Götz, M. (2011):** BPMN 2.0 Tutorial : kompakte Einführung in die BPMN 2.0.  
iTransparent.  
URL: [http://www.itransparent.de/sites/default/files/BPMN\\_2\\_0\\_Tutorial\\_Business\\_Process\\_Modeling\\_Notation\\_Deutsch.pdf](http://www.itransparent.de/sites/default/files/BPMN_2_0_Tutorial_Business_Process_Modeling_Notation_Deutsch.pdf) [Zugriff: 21.6.2013]

- Habich, R.; Himmelreicher, R.; Huschka, D. (2010):** Zur Entwicklung der Dateninfrastruktur in Deutschland. Berlin: RatSWD.  
URL: [http://www.ratswd.de/download/RatSWD\\_WP\\_2010/RatSWD\\_WP\\_157.pdf](http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_157.pdf)  
[Zugriff: 24.3.2013]
- Hausstein, B. (2012):** Die Vergabe von DOI-Namen für Sozial- und Wirtschaftsdaten : Serviceleistungen der Registrierungsagentur da|ra.  
In: Altenhöner, R.; Oellers, C. (Hrsg.): *Langzeitarchivierung von Forschungsdaten : Standards und disziplinspezifische Lösungen*. Berlin: Scivero, S. 137–145.
- Hein, S. (2012):** Metadaten für die Langzeitarchivierung.  
In: Altenhöner, R.; Oellers, C. (Hrsg.): *Langzeitarchivierung von Forschungsdaten : Standards und disziplinspezifische Lösungen*. Berlin: Scivero, S. 87–109.
- Hey, T.; Tansley, S.; Tolle, K. (Hrsg.) (2009):** Jim Gray on eScience : a transformed scientific method ; based on the transcript of a talk given by Jim Gray to the NRC-CSTB in Mountain View, CA, on January 11, 2007.  
In: *The fourth paradigm : data-intensive scientific discovery*. Redmond, Wash.: Microsoft Research, S. xvii – xxxi.  
URL: [http://research.microsoft.com/en-us/UM/redmond/about/collaboration/fourthparadigm/4th PARADIGM BOOK complete HR.pdf](http://research.microsoft.com/en-us/UM/redmond/about/collaboration/fourthparadigm/4th_PARADIGM_BOOK_complete_HR.pdf) [Zugriff: 22.4.2013]
- Hobohm, H.-C. (2010):** DIKW-Hierarchie. In: Umlauf, K.; Gradmann, S. (Hrsg.): *Lexikon der Bibliotheks- und Informationswissenschaft : LB. 1, Lfg. 3*. Stuttgart: Hiersemann, S. 222 – 223.
- Huschka, D.; Oellers, C.; Ott, N.; Wagner, G. G. (2011):** Datenmanagement und Data Sharing: Erfahrungen in den Sozial- und Wirtschaftswissenschaften.  
In: Büttner, S.; Hobohm, H.-C.; Müller, L. (Hrsg.): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen, S. [35] – 48.  
URL: [http://opus.kobv.de/fhpotsdam/volltexte/2011/227/pdf/1.3 Datenmanagement und Data Sharing.pdf](http://opus.kobv.de/fhpotsdam/volltexte/2011/227/pdf/1.3_Datenmanagement_und_Data_Sharing.pdf) [Zugriff: 19.3.2013]
- ICPSR (Inter-university Consortium for Political and Social Research) (2012):** Guide to social science data preparation and archiving : best practice through the data life cycle. Ann Arbor, Mich.: ICPSR, Institute for Social Research, University of Michigan.  
URL: <http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf> [Zugriff: 9.5.2013]
- Jensen, U. (2011):** Datenmanagementpläne.  
In: Büttner, S.; Hobohm, H.-C.; Müller, L. (Hrsg.): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen, S. [71] – 82.  
URL: [http://opus.kobv.de/fhpotsdam/volltexte/2011/230/pdf/2.3 Datenmanagement plAne.pdf](http://opus.kobv.de/fhpotsdam/volltexte/2011/230/pdf/2.3_DatenmanagementplAne.pdf) [Zugriff: 4.4.2013]

- Jensen, U. (2012):** Leitlinien zum Management von Forschungsdaten : sozialwissenschaftliche Umfragedaten.  
Köln: GESIS - Leibniz-Institut für Sozialwissenschaften.  
URL: [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/gesis\\_methodenberichte/2012/TechnicalReport\\_2012-07.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-07.pdf) [Zugriff: 11.4.2013]
- Jensen, U.; Katsanidou, A.; Zenk-Möltgen, W. (2011):** Metadaten und Standards.  
In: Büttner, S.; Hobohm, H.-C.; Müller, L. (Hrsg.): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen, S. [83] – 100.  
URL: [http://opus.kobv.de/fhpotsdam/volltexte/2011/231/pdf/2.4\\_Metadaten\\_und\\_Standards.pdf](http://opus.kobv.de/fhpotsdam/volltexte/2011/231/pdf/2.4_Metadaten_und_Standards.pdf) [Zugriff: 4.4.2013]
- Jones, S. (2011):** How to develop a data management and sharing plan.  
Digital Curation Centre.  
URL: [http://www.dcc.ac.uk/webfm\\_send/486](http://www.dcc.ac.uk/webfm_send/486) [Zugriff: 12.4.2013]
- Kastner, C. (2010):** Geschäftsprozessmanagement am Beispiel der Fotothek der Stiftung Preußische Schlösser und Gärten Berlin-Brandenburg.  
Fachhochschule Potsdam, Diplomarbeit.
- Key Perspectives (2010):** Data dimensions : disciplinary differences in research data sharing, reuse and long term viability ; a comparative review based on sixteen case studies.  
Edinburgh: DCC (Digital Curation Center).  
URL: <http://www.dcc.ac.uk/sites/default/files/documents/publications/case-studies/SCARP%20SYNTHESIS.pdf> [Zugriff: 24.3.2013]
- King, G. (2011):** Ensuring the data-rich future of the social sciences.  
In: *Science*, Band 331, Ausgabe 6018, 10.2.2011, S. 719–721.  
URL: <http://gking.harvard.edu/files/datarich.pdf> [Zugriff: 7.5.2013]
- Klump, J. (2009):** Digitale Forschungsdaten.  
In: Neuroth, H.; Oßwald, A.; Scheffel, R.; Strathmann, S.; Jehn, M. (Hrsg.): *nestor Handbuch : eine kleine Enzyklopädie der digitalen Langzeitarchivierung*.  
Online-Version 2.0, Kap. 17.10.  
URL: [http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor\\_handbuch\\_artikel\\_292.pdf](http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_292.pdf) [Zugriff: 22.4.2013]
- Kuipers, T.; Van der Hoeven, J. (2009):** PARSE.Insight: Insight into digital preservation of research output in Europe ; survey report.  
URL: [http://www.parse-insight.eu/downloads/PARSE-Insight\\_D3-4\\_SurveyReport\\_final\\_hq.pdf](http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf) [Zugriff: 5.4.2013]
- Kuula, A.; Borg, S. (2008):** Open access to and reuse of research data : the state of the art in Finland. Tampere: Finnish Social Science Data Archive.  
URL: [http://www.fsd.uta.fi/fi/julkaisut/julkaisusarja/FSDjs07\\_OECD\\_en.pdf](http://www.fsd.uta.fi/fi/julkaisut/julkaisusarja/FSDjs07_OECD_en.pdf) [Zugriff: 5.4.2013]

- Lautenschlager, M. (2011):** Institutionalisierte „Data Curation Services“.  
In: Büttner, S.; Hobohm, H.-C.; Müller, L. (Hrsg.): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen, S. [149] - 156.  
URL: [http://opus.kobv.de/fhpotsdam/volltexte/2011/236/pdf/3.1\\_Institutionalisierte\\_Data\\_Curation\\_Services.pdf](http://opus.kobv.de/fhpotsdam/volltexte/2011/236/pdf/3.1_Institutionalisierte_Data_Curation_Services.pdf) [Zugriff: 11.8.2013]
- Ludwig, J.; Enke, H. (Hrsg.) (2013):** Leitfaden zum Forschungsdaten-Management : Handreichungen aus dem WissGrid-Projekt. Glückstadt: Hülsbusch.  
URL: [http://www.wissgrid.de/publikationen/Leitfaden\\_Data-Management-WissGrid.pdf](http://www.wissgrid.de/publikationen/Leitfaden_Data-Management-WissGrid.pdf) [Zugriff: 19.3.2013]
- Mauer, R. (2012):** Das GESIS Datenarchiv für Sozialwissenschaften.  
In: Altenhöner, R.; Oellers, C. (Hrsg.): *Langzeitarchivierung von Forschungsdaten : Standards und disziplinspezifische Lösungen*. Berlin: Scivero, S. 197 – 215.
- Mayring, P. (2000):** Qualitative Inhaltsanalyse.  
In: *Forum Qualitative Sozialforschung*, Band 1, Ausgabe 2.  
URL: <http://ojs-test.cedis.fu-berlin.de/fqs-test/index.php/fqs/article/download/1089/2384> [Zugriff: 23.6.2013]
- Mayring, P. (2002):** Einführung in die qualitative Sozialforschung. Eine Anleitung zu qualitativem Denken. Weinheim: Beltz.
- Mayring, P. (2008):** Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim: Beltz.
- Mayring, P. (2012):** Qualitative Inhaltsanalyse.  
In: Flick, U.; Kardorff, E. v.; Steinke, I. (Hrsg.): *Qualitative Forschung : ein Handbuch*. Reinbek bei Hamburg: Rowohlt Taschenbuch-Verl., S. 468 – 475.
- Meuser, M.; Nagel, U. (2005):** Experteninterviews - vielfach erprobt, wenig bedacht. : ein Beitrag zur qualitativen Methodendiskussion.  
In: Bogner, A.; Littig, B.; Menz, W. (Hrsg.): *Das Experteninterview : Theorie, Methode, Anwendung*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 71 – 93.
- Meyermann, A. (2012):** Datenmanagement : eine zentrale Voraussetzung für den Erfolg der Data Sharing-Idee in den Sozialwissenschaften. Bielefeld: Datenservicezentrum Betriebs- und Organisationsdaten an der Universität Bielefeld.  
URL: [http://www.uni-bielefeld.de/dsz-bo/pdf/2012-07-23\\_WP2\\_Datenmanagement.pdf](http://www.uni-bielefeld.de/dsz-bo/pdf/2012-07-23_WP2_Datenmanagement.pdf) [Zugriff: 25.4.2013]
- MPG (Max-Planck-Gesellschaft) (2009):** Regeln zur Sicherung guter wissenschaftlicher Praxis. URL: <http://www.mpg.de/229457> [Zugriff: 19.6.2013]
- nestor-Arbeitsgruppe OAIS-Übersetzung / Terminologie (Hrsg.) (2012):** Referenzmodell für ein Offenes Archiv-Informationssystem : Deutsche Übersetzung. nestor-materialien, 16.  
URL: [http://files.d-nb.de/nestor/materialien/nestor\\_mat\\_16.pdf](http://files.d-nb.de/nestor/materialien/nestor_mat_16.pdf) [Zugriff: 4.8.2013]



**Neuroth, H.; Strathmann, S.; Oßwald, A.; Scheffel, R.; Klump, J.; Ludwig, J. (Hrsg.) (2012):** Langzeitarchivierung von Forschungsdaten: Eine Bestandsaufnahme. Boizenburg: Hülsbusch.  
URL: <http://nbn-resolving.de/urn:nbn:de:0008-2012031401> [Zugriff: 15.4.2013]

**OECD (Organisation for Economic Co-operation and development) (2007):** OECD principles and guidelines for access to research data from public funding. Paris: OECD Publications.  
URL: <http://www.oecd.org/science/sci-tech/38500813.pdf> [Zugriff: 11.6.2013]

**Pampel, H.; Bertelmann, R. (2011):** „Data Policies“ im Spannungsfeld zwischen Empfehlungen und Verpflichtung.  
In: Büttner, S.; Hobohm, H.-C.; Müller, L. (Hrsg.): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen, S. [49] – 61.  
URL: [http://opus.kobv.de/fhpotsdam/volltexte/2011/228/pdf/2.1\\_Data\\_Policies\\_im\\_Spannungsfeld\\_zwischen\\_Empfehlung\\_und\\_Verpflichtung.pdf](http://opus.kobv.de/fhpotsdam/volltexte/2011/228/pdf/2.1_Data_Policies_im_Spannungsfeld_zwischen_Empfehlung_und_Verpflichtung.pdf) [Zugriff: 4.4.2013]

**Pampel, H.; Hobohm, H.-C.; Bertelmann, R. (2009):** „Data Librarianship“ - Rollen, Aufgaben, Kompetenzen. Erfurt: 98. Deutscher Bibliothekartag.  
URL: <http://nbn-resolving.de/urn:nbn:de:0290-opus-6985> [Zugriff: 8.5.2013]

**Quandt, M.; Mauer, R. (2012):** Sozialwissenschaften.  
In: Neuroth, H.; Strathmann, S.; Oßwald, A.; Scheffel, R.; Klump, J.; Ludwig, J. (Hrsg.): *Langzeitarchivierung von Forschungsdaten : eine Bestandsaufnahme*. Boizenburg: Hülsbusch, S. [61] – 81.  
URL: <http://www.nestor.sub.uni-goettingen.de/bestandsaufnahme/index.php>  
[Zugriff: 19.3.2013]

**RatSWD (Rat für Sozial- und Wirtschaftsdaten) (2011):** Aufgaben und Struktur des Rates für Sozial- und Wirtschaftsdaten (RatSWD).  
URL: [http://ratswd.de/dl/doc/RatSWD\\_Aufgaben\\_Struktur\\_Apr11.pdf](http://ratswd.de/dl/doc/RatSWD_Aufgaben_Struktur_Apr11.pdf)  
[Zugriff: 7.8.2013]

**Rümpel, S. (2010):** Data Librarianship : Anforderungen an Bibliothekare im Forschungsdatenmanagement. Fachhochschule Potsdam, Diplomarbeit.  
URL: <http://opus4.kobv.de/opus4-fhpotsdam/files/134/10600.pdf> [Zugriff: 13.3.2013]

**Rümpel, S. (2011):** Der Lebenszyklus von Forschungsdaten.  
In: Büttner, S.; Hobohm, H.-C.; Müller, L. (Hrsg.): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen, S. [25] – 34.  
URL: [http://opus.kobv.de/fhpotsdam/volltexte/2011/226/pdf/1.2\\_Der\\_Lebenszyklus\\_von\\_Forschungsdaten.pdf](http://opus.kobv.de/fhpotsdam/volltexte/2011/226/pdf/1.2_Der_Lebenszyklus_von_Forschungsdaten.pdf) [Zugriff: 7.3.2013]

**Schäfer, L.; Pampel, H.; Klump, J.; Häbler, T. (2013):** Bericht Symposium „Forschungsdaten-Infrastrukturen (FDI 2013)“. GFZ (GeoForschungsZentrum Potsdam).  
URL: [http://www.forschungsdaten.org/wp-content/uploads/FDI2013\\_Symposium\\_Materialien.pdf](http://www.forschungsdaten.org/wp-content/uploads/FDI2013_Symposium_Materialien.pdf) [Zugriff: 14.3.2013]

**Schnell, R.; Hill, P. B.; Esser, E. (2011):** Methoden der empirischen Sozialforschung. München: Oldenbourg.

**Scholze, F.; Goebelbecker, H.-J.; Pampel, H. (2013):** re3data.org - Registry of Research Data Repositories. Leipzig: 102. Deutscher Bibliothekartag.  
URL: <http://nbn-resolving.de/urn:nbn:de:0290-opus-13826> [Zugriff: 10.4.2013]

**Schrimpf, S. (2012):** Herausforderungen der Archivierung sozial-, verhaltens- und wirtschaftswissenschaftlicher Datenbestände.  
In: Altenhöner, R.; Oellers, C. (Hrsg.): *Langzeitarchivierung von Forschungsdaten : Standards und disziplinspezifische Lösungen*. Berlin: Scivero, S. 51 – 65.

**Simukovic, E. (2012):** Enhanced publications - Integration von Forschungsdaten beim wissenschaftlichen Publizieren. Humboldt-Universität zu Berlin, Masterarbeit.  
URL: <http://nbn-resolving.de/urn:nbn:de:kobv:11-100206672> [Zugriff: 19.3.2013]

**Spindler, G.; Hillegeist, T. (2010):** Langzeitarchivierung wissenschaftlicher Primärdaten.  
In: Neuroth, H.; Oßwald, A.; Scheffel, R.; Strathmann, S.; Huth, K. (Hrsg.): *nestor Handbuch : eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Online-Version 2.3, Kap. 16.3.  
URL: <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-20100617312>  
[Zugriff: 25.5.2013]

**Spindler, G.; Hillegeist, T. (2011):** Rechtliche Probleme der elektronischen Langzeitarchivierung von Forschungsdaten.  
In: Büttner, S.; Hobohm, H.-C.; Müller, L. (Hrsg.): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen, S. [63] – 69.  
URL: [http://opus.kobv.de/fhpotsdam/volltexte/2011/229/pdf/2.2\\_Rechtliche\\_Probleme\\_der\\_elektronischen\\_Langzeitarchivierung\\_von\\_Forschungsdaten.pdf](http://opus.kobv.de/fhpotsdam/volltexte/2011/229/pdf/2.2_Rechtliche_Probleme_der_elektronischen_Langzeitarchivierung_von_Forschungsdaten.pdf)  
[Zugriff: 6.5.2013]

**Swan, A.; Brown, S. (2008):** The skills, role and career structure of data scientists and curators : an assessment of current practice and future needs ; report to the JISC. Truro: [Key Perspectives].  
URL: <http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dataski/lscareersfinalreport.pdf> [Zugriff: 30.5.2013]

**Treloar, A.; Harboe-Ree, C. (2008):** Data management and the curation continuum: how the Monash experience is informing repository relationships. Melbourne: VALA2008 Conference.  
URL: [http://www.valaconf.org.au/vala2008/papers2008/111\\_Treloar\\_Final.pdf](http://www.valaconf.org.au/vala2008/papers2008/111_Treloar_Final.pdf)  
[Zugriff: 29.4.2013]

**Verhaar, P.; Mitova, M.; Rutten, P.; Van der Weel, A.; Birnie, F.; Wagenaar, A.; Gloerich, J. (2010):** Data curation in arts and media research. Utrecht: Surf Foundation.  
URL: [http://www.surf.nl/nl/themas/openonderzoek/cris/Documents/SURFshare\\_Collectionen\\_Data\\_Curation\\_in\\_Arts\\_and%20Media\\_Research\\_DEF.pdf](http://www.surf.nl/nl/themas/openonderzoek/cris/Documents/SURFshare_Collectionen_Data_Curation_in_Arts_and%20Media_Research_DEF.pdf)  
[Zugriff: 9.4.2013]

**Weichselgartner, E. (2011):** Forschungsdaten in der Psychologie: Disziplinspezifische und disziplinübergreifende Bedürfnisse. Zusammenfassung des Forums (2) der 5. Konferenz für Sozial- und Wirtschaftsdaten. Berlin: RatSWD.

URL: [http://www.ratswd.de/download/RatSWD\\_WP\\_2011/RatSWD\\_WP\\_187.pdf](http://www.ratswd.de/download/RatSWD_WP_2011/RatSWD_WP_187.pdf)

[Zugriff: 28.3.2013]

**Weichselgartner, E.; Günther, A.; Denhard, I. (2011):** Archivierung von Forschungsdaten.

In: Büttner, S.; Hobohm, H.-C.; Müller, L. (Hrsg.): *Handbuch*

*Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen, S. [191] - 202.

URL: [http://opus.kobv.de/fhpotsdam/volltexte/2011/239/pdf/3.4\\_Archivierung\\_von\\_Forschungsdaten.pdf](http://opus.kobv.de/fhpotsdam/volltexte/2011/239/pdf/3.4_Archivierung_von_Forschungsdaten.pdf) [Zugriff: 4.4.2013]

**Wiggins, A.; Bonney, R.; Graham, E.; Henderson, S.; Kelling, S.; LeBuhn, G.; Littauer, R.; Lotts, K.; Michener, W.; Newman, G.; Russell, E.; Stevenson, R.; Weltzin, J. (2013):** Data management guide for public participation in scientific research. Albuquerque, NM: DataOne.

Albuquerque, NM: DataOne.

URL: <http://www.dataone.org/sites/all/documents/DataONE-PPSR-DataManagementGuide.pdf> [Zugriff: 3.4.2013]

**Winkler-Nees, S. (2011):** Anforderungen an wissenschaftliche Informationsinfrastrukturen. Berlin: RatSWD.

URL: [http://www.ratswd.de/download/RatSWD\\_WP\\_2011/RatSWD\\_WP\\_180.pdf](http://www.ratswd.de/download/RatSWD_WP_2011/RatSWD_WP_180.pdf)

[Zugriff: 28.3.2013]

**Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (Hrsg.) (1999):** Regeln zur Sicherung guter wissenschaftlicher Praxis.

URL: [http://www.leibniz-](http://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Forschung/Regeln_zur_Sicherung_guter_wissenschaftlicher_Praxis.pdf)

[gemeinschaft.de/fileadmin/user\\_upload/downloads/Forschung/Regeln\\_zur\\_Sicherung\\_guter\\_wissenschaftlicher\\_Praxis.pdf](http://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Forschung/Regeln_zur_Sicherung_guter_wissenschaftlicher_Praxis.pdf) [Zugriff: 12.6.2013]

*Sonstige Quellen*

**CESSDA (Council of European Social Science Data Archives) (2012):** Dissemination.  
URL: <http://www.CESSDA.org/sharing/dissemination/> [Zugriff: 11.8.2013].

**DCC (Digital Curation Centre) (2013):** DCC Curation Lifecycle Model.  
URL: <http://www.dcc.ac.uk/resources/curation-lifecycle-model> [Zugriff: 3.8.2013]

**Donnelly, M. (2008):** Research Data Management Forum : Core Skills Diagram.  
URL: <http://data-forum.blogspot.de/2008/12/rdmf2-core-skills-diagram.html>  
[Zugriff: 30.5.2013]

**Habich, R. (2013):** Persönliches Gespräch mit Christiane Behnert, 28. Mai.

**Habich, R. (2013):** Persönliches Gespräch mit Christiane Behnert, 26. Juli.

**Habich, R. (2013):** Persönliches Gespräch mit Christiane Behnert, 6. August.

**o. A. (2013):** Berufsverband Bibliothek Information e.V. : OPUS-Themen ;  
Elektronisches Publizieren.  
URL: <http://www.bib-info.de/verband/publikationen/opus/opus-themen/elektronische-publikationen-elektronisches-publizieren.html>  
[Zugriff: 11.8.2013].

**o. A. (2013):** da|ra – Registrierungsagentur für Sozial- und Wirtschaftsdaten : Über  
uns.  
URL: <http://www.da-ra.de/de/ueber-uns/> [Zugriff: 11.8.2013].

**o. A. (2013):** GESIS : Wahlstudien.  
URL: <http://www.gesis.org/wahlen/gles/> [Zugriff: 11.8.2013].

**o. A. (2013):** The European Charter for Researchers.  
URL: <http://ec.europa.eu/euraxess/index.cfm/rights/europeanCharter>  
[Zugriff: 11.8.2013].

**o. A. [2013]:** WZB: WZB unterzeichnet EU-Charta.  
URL: <http://www.wzb.eu/de/news/wzb-unterzeichnet-eu-charta> [Zugriff: 11.8.2013].

**o. A. (2012):** DDI Alliance : DDI Specification.  
URL: <http://www.ddialliance.org/Specification/> [Zugriff: 11.8.2013].

**o. A. (2012):** DDI Alliance : Marked-up Codebooks (Under Construction).  
URL: <http://www.ddialliance.org/resources/tools/codebooks> [Zugriff: 11.8.2013].

**o. A. (2012):** GESIS : Unser Angebot ; Phase 1 Recherchieren.  
URL: <http://www.gesis.org/unser-angebot/recherchieren/> [Zugriff: 11.8.2013].

**o. A. (2011):** DDI Alliance : Retirement History Longitudinal Survey Series.  
URL: <http://www.ddialliance.org/resources/tools/codebooks/retirement>  
[Zugriff: 11.8.2013].

- o. A. (2011):** Statistische Ämter des Bundes und der Länder : Forschungsdatenzentren ; Datenzugang | Anonymität von Mikrodaten.  
URL: <http://www.forschungsdatenzentrum.de/anonymisierung.asp>  
[Zugriff: 11.8.2013].
- o. A. (2009):** DDI Alliance : DDI Tools.  
URL: <http://www.ddialliance.org/resources/tools> [Zugriff: 11.8.2013].
- o. A. (o. J.):** BPMN Modelling Guidelines : die 5 Kategorien der Modellierungsrichtlinien.  
URL: <http://www.modeling-guidelines.org/de/> [Zugriff: 11.8.2013].
- o. A. (o. J.):** Columbia University Libraries / Information Services : Data Management Plan Templates.  
URL: <http://scholcomm.columbia.edu/data-management/data-management-plan-templates/> [Zugriff: 11.8.2013].
- o. A. (o. J.):** MIT Libraries : Data Management and Publishing ; Manage Your Data.  
URL: <http://libraries.mit.edu/guides/subjects/data-management/index.html>  
[Zugriff: 11.8.2013].
- o. A. (o. J.):** RatSWD – Rat für Sozial- und Wirtschaftsdaten : Akkreditierte Datenzentren - Ein vielfältiges Datenangebot.  
URL: <http://ratswd.de/forschungsdaten/fdz> [Zugriff: 11.8.2013].
- o. A. (o. J.):** WZB: Bibliothek ; Datenbanken.  
URL: <http://www.wzb.eu/de/bibliothek/bestand-recherche/datenbanken>  
[Zugriff: 11.8.2013].
- o. A. (o. J.):** WZB : Forschung.  
URL: <http://www.wzb.eu/de/forschung> [Zugriff: 11.8.2013].
- o. A. (o. J.):** WZB : Publikationen ; Datenbanken.  
URL: <http://www.wzb.eu/de/publikationen/datenbanken> [Zugriff: 11.8.2013]
- o. A. (o. J.):** WZB : über das WZB.  
URL: <http://www.wzb.eu/de/ueber-das-wzb> [Zugriff: 11.8.2013]
- o. A. (o. J.):** WZB: Zentrales Datenmanagement.  
URL: <http://www.wzb.eu/de/ueber-das-wzb/organisation/service-und-verwaltung/datenmanagement> [Zugriff: 11.8.2013].
- Zenk-Möltgen, W. (2013):** E-Mail an Christiane Behnert, 24. Juni.

## Anhang

### Anhang A) E-Mail

#### nestor-Praktikertag / Frage zum DDI-Standard

---

Zenk-Möltgen, Wolfgang <Wolfgang.Zenk-Moeltgen@gesis.org>  
An: "christiane.behnert@gmail.com" <christiane.behnert@gmail.com>

24. Juni 2013 11:33

Sehr geehrte Frau Behnert,

Frau Schuman hat mir Ihre Anfrage weitergeleitet und ich will versuchen, Ihnen ein paar Informationen dazu zusammenzustellen. Zusammenfassend kann man sagen, dass auch Panel-Daten mit DDI beschreibbar sind, dass aber ein paar Informationen zur Panel-Struktur nicht besonders strukturiert abgelegt werden. Es hängt dann z.T. am Aufbau des konkreten Datensatzes, wie die Struktur der Paneldaten für den Benutzer transparent dokumentiert ist. Das gilt für DDI-Lifecycle (DDI ab Version 3.0), denn in DDI-Codebook (bis Version 2.5) kann zwar jeder Datensatz, also auch ein Panel-Datensatz, dokumentiert werden, jedoch nicht die konkrete Panel-Struktur.

Zur Dokumentation von Längsschnittdaten in DDI gibt es eine Reihe von "Best Practise" papers, an denen ich z.T. mitgearbeitet habe:

<http://www.ddialliance.org/resources/publications/working/BestPractices/LongitudinalData>

Vielleicht ist das Paper "Presenting Longitudinal Studies to End Users Effectively Using DDI Metadata" interessant für Sie, denn darin gibt es auch im Anhang ein DDI Example mit einer (fiktiven) Panelstudie. Ein konkretes echtes Beispiel, das in DDI3 dokumentiert ist, ist das LISS Panel: <http://www.lissdata.nl/lissdata/> benutzt die Software Questasy, die auf DDI3 basiert. Auch das SOEP entwickelt gerade eine Nachfolgeversion von SOEPinfo auf der Basis von DDI3 (s. im Anhang eine Präsentation der Kollegen vom DIW dazu).

Vielleicht hilft Ihnen das schon weiter? Ansonsten können Sie mir gerne noch Fragen dazu schicken.

Mit freundlichen Grüßen,

Wolfgang Zenk-Möltgen

## Anhang B) Transkriptionsregeln

### Einfaches Transkriptionssystem

1. Es wird wörtlich transkribiert, also nicht lautsprachlich oder zusammenfassend. Vorhandene Dialekte werden möglichst wortgenau ins Hochdeutsche übersetzt. Wenn keine eindeutige Übersetzung möglich ist, wird der Dialekt beibehalten, zum Beispiel: Ich gehe heuer auf das Oktoberfest.
2. Wortverschleifungen werden nicht transkribiert, sondern an das Schriftdeutsch angenähert. Beispielsweise wird aus „Er hatte noch so’n Buch genannt“ wird zu „Er hatte noch so ein Buch genannt“ und „hamma“ wird zu „haben wir“. Die Satzform wird beibehalten, auch wenn sie syntaktische Fehler beinhaltet, beispielsweise: „bin ich nach Kaufhaus gegangen.“
3. Wort- und Satzabbrüche sowie Stottern werden geglättet bzw. ausgelassen, Wortdoppelungen nur erfasst, wenn sie als Stilmittel zur Betonung genutzt werden: „Das ist mir sehr, sehr wichtig.“. „Ganze“ Halbsätze, denen nur die Vollendung fehlt, werden jedoch erfasst und mit dem Abbruchzeichen / gekennzeichnet.
4. Interpunktion wird zu Gunsten der Lesbarkeit geglättet, das heißt bei kurzem Senken der Stimme oder uneindeutiger Betonung, wird eher ein Punkt als ein Komma gesetzt. Dabei sollen Sinneinheiten beibehalten werden.
5. Pausen werden durch drei Auslassungspunkte in Klammern (...) markiert.
6. Verständnissignale des gerade nicht Sprechenden wie „mhm, aha, ja, genau, ähm“ etc. werden nicht transkribiert. AUSNAHME: Eine Antwort besteht NUR aus „mhm“ ohne jegliche weitere Ausführung. Dies wird als „mhm (bejahend)“, oder „mhm (verneinend)“ erfasst, je nach Interpretation.
7. Besonders betonte Wörter oder Äußerungen werden durch GROSSSCHREIBUNG gekennzeichnet.

8. Jeder Sprecherbeitrag erhält eigene Absätze. Zwischen den Sprechern gibt es eine freie, leere Zeile. Auch kurze Einwürfe werden in einem separaten Absatz transkribiert. Mindestens am Ende eines Absatzes werden Zeitmarken eingefügt.

9. Emotionale nonverbale Äußerungen der befragten Person und des Interviewers, die die Aussage unterstützen oder verdeutlichen (etwa wie lachen oder seufzen), werden beim Einsatz in Klammern notiert.

10. Unverständliche Wörter werden mit (unv.) gekennzeichnet. Längere unverständliche Passagen sollen möglichst mit der Ursache versehen werden (unv., Handystörgeräusch) oder (unv., Mikrofon rauscht). Vermutet man einen Wortlaut, ist sich aber nicht sicher, wird das Wort bzw. der Satzteil mit einem Fragezeichen in Klammern gesetzt. Zum Beispiel: (Xylomethanolin?) Generell werden alle unverständlichen Stellen mit einer Zeitmarke versehen, wenn innerhalb von einer Minute keine Zeitmarke gesetzt ist.

11. Die interviewende Person wird durch ein „I:“, die befragte Person durch ein „B:“ gekennzeichnet. Bei mehreren Interviewpartnern (z.B. Gruppendiskussion) wird dem Kürzel „B“ eine entsprechende Kennnummer oder Name zugeordnet (z.B. „B1:“, „Peter:“).

12. Die Transkription wird als Rich Text Format (.rtf Datei) gespeichert. Benennung der Datei entsprechend des Audiodateinamens (ohne Endung wav, mp3).  
Beispielsweise: Interview\_04022011.rtf oder interview\_schmitt.rtf

Quelle:

Dresing, T.; Pehl, T. (2012): Praxisbuch Transkription : Regelsysteme, Software und praktische Anleitungen für qualitative ForscherInnen. Marburg: Eigenverlag, S. 26-28.

URL: <http://www.audiotranskription.de/Praxisbuch-Transkription.pdf>

[Zugriff: 6.6.2013]



## **Anhang C) Interview-Transkripte**

Die Transkripte beginnen formatierungsbedingt (Zeilennummern) jeweils auf einer neuen Seite.

## 1 C1) Interview 1

### 2 Teil a)

3

4

5 I: Okay, also es geht um das Thema Forschungsdaten und ich habe ein paar Leitfragen notiert  
6 und ich würde sagen, wir fangen einfach mal an und Sie erzählen einfach dazu, wie Sie  
7 möchten und ich werde dann einfach Rückfragen stellen, je nachdem. Zunächst würde mich  
8 interessieren, welche Arten von quantitativen Daten Sie als Forschungsdaten überhaupt  
9 wahrnehmen. Also was sind für Sie Forschungsdaten? #00:00:25-9#

10

11 B: Quantitative Daten? #00:00:27-1#

12

13 I: Ja. #00:00:28-4#

14

15 B: Also ich würde sagen natürlich alle Erhebungen, sozusagen Individualdaten und die können  
16 irgendwie schon, also bekannten SOEP, NEPS, die Daten die wir selber hier erheben mit  
17 Mikrozensus ist jetzt sozusagen nicht unbedingt eine freiwillige Erhebung, aber gehören zum,  
18 für mich auch zu Forschungsdaten. Dann eher Aggregatdaten, die beispielsweise jetzt, wie  
19 heißen diese, jetzt fällt mir der Name nicht mehr ein, diese die auch beispielsweise auf  
20 europäischer Ebene, also wo Staaten beispielsweise eher dargestellt sind? Mit wie viel Prozent  
21 der Leute dort einen akademischen Abschluss haben. Oder wie viel sind arbeitslos, also das  
22 eher Aggregatdaten sind. Es ist jetzt, also ich empfinde sie auch als Forschungsdaten, aber  
23 damit habe ich nie gearbeitet, sind beispielsweise auch Dokumenten. Also ist es je nachdem,  
24 wie viele Dokumente man hat, ist auch eine Art von quantitativen, ja Forschungsdaten. Also da  
25 sind weniger Individuen, also zu sagen direkt befragt, aber natürlich ist auch ein Produkt (unv.)  
26 #00:01:48-0#, ja. #00:01:49-3#

27

28 I: Also verstehe ich das richtig, dass Forschungsdaten für Sie alles das ist, was ein Produkt von  
29 Forschung ist? Und was weiter verwendet wird im Gegensatz zu Publikation? #00:02:01-3#

30

31 B: Na Publikationen könnten natürlich auch Forschungsdaten sein, also Daten sind einfach  
32 Informationsquellen für mich. Und je nach Fragestellung, wenn ich beispielsweise eine  
33 Fragestellung habe, wo ich Publikationen analysiere oder Anzahl der Publikationen, Typ der  
34 Publikationen, also das wären dann auch Forschungsdaten. Die müsste ich mir irgendwo eher  
35 zusammen vermutlich basteln, da gibt es ja keine Quellen und dann wären auch  
36 Forschungsdaten da. #00:02:33-3#

37

38 I: Okay und mit welchen quantitativen Daten haben Sie jetzt in der aktuellen Studie zu tun?  
39 Also wie sehen die aus, welche Formate sind das so? #00:02:42-0#

40

41 B: Also aktuell, was jetzt gerade läuft? #00:02:44-8#

42

43 I: Ja. #00:02:45-0#

44

45 B: Wir befragen zum einen, also ich weiß jetzt nicht wie konkret, aber können Sie mir sagen, ob  
46 das die Art von Antwort ist, jetzt [...]\*. Also es handelt sich um Individualdaten [...]\*. #00:03:21-  
47 2#

48

- 49 I: Okay. Ja zum Datenlebenszyklus, also zum ganzen Projektablauf, wie sieht das genau aus und  
50 wie entstehen oder wann entstehen da die meisten Daten? #00:03:34-6#  
51
- 52 B: Ich glaube ich habe die Frage nicht verstanden. (lacht) #00:03:37-1#  
53
- 54 I: (lacht) Also es gibt, also die Daten durchleben ja einen bestimmten Zyklus, also sie werden  
55 erfasst, erhoben und bearbeitet und so weiter. Und bei dem aktuellen Projekt was Sie  
56 bearbeiten, wie sieht das überhaupt aus da bei den Daten, also wo entstehen zum Beispiel die  
57 meisten Daten in welcher Phase? #00:03:57-6#  
58
- 59 B: Die meisten im Sinne von Quantität? #00:03:59-8#  
60
- 61 I: Ja. #00:04:00-7#  
62
- 63 B: Na ja [es]\* werden die meisten Daten jetzt entstehen, also wir befinden uns jetzt gerade in  
64 der Feldphase. Und jetzt gerade führen wir Interviews durch, also praktisch mit [...]\*. Und da es  
65 ja auch relativ kleines Projekt ist, ist sie anders als beim SOEP [...]\*. Und eher jetzt werden wir  
66 die meisten Daten erheben, nur beispielsweise wie gesagt [...]\*. #00:05:39-5#  
67
- 68 I: Mhm okay. Und wie werden die Daten bei Ihnen dokumentiert, also beschrieben? Also  
69 sowohl auf Variablenebene als auch auf Studienebene insgesamt, also zum Stichwort  
70 Metadaten zum Beispiel? #00:05:55-1#  
71
- 72 B: Also Informationen beispielsweise über die [...]\*. wurde, wie die Befragung gelaufen ist.  
73 #00:06:01-6#  
74
- 75 I: Ja genau so was zum Beispiel genau. Wie das alles zustande ist? #00:06:03-6#  
76
- 77 B: Na ja, also wir haben als Allererstes natürlich ganze Prozedurenbeschreibungen, bevor wir  
78 überhaupt ins Feld gegangen sind, wo wir irgendwie das Design und Vorgehen beschrieben  
79 haben, das mussten wir auch machen. Weil wir zwar eine individuelle Befragung machen um,  
80 aber um die Leute zu erreichen mussten wir [...]\*. und sprechen mussten wir auch offiziell über  
81 den [...]\*. Datenschutz gehen. Und sie wollten schon im vorab eine Prozedurenbeschreibung  
82 haben, in anderen Projekten auch in meinem früheren Projekt, [...]\*, hat man das eher im  
83 Nachhinein dokumentiert, wie man das gemacht hat (lacht). Also man hat sich schon alles  
84 vorgestellt und dann mit der Befragung (Institut aus?) #00:06:47-7# war es schon beredet,  
85 aber nicht so formell und so ordentlich, wie wir das jetzt gemacht haben. #00:06:51-8#  
86
- 87 I: Mhm. #00:06:52-2#  
88
- 89 B: Und jetzt wo die Befragungen laufen haben wir zum einen Interviewer Protokolle, also zum  
90 einen haben wir Informationen über [...]\*, zum Teil, ich weiß gar nicht mehr, bei welchem Amt  
91 man einsehen kann, beispielsweise welche [...]\*. sie sind und welche [...]\*. sie da haben und [so  
92 weiter]\*. Und dann haben wir, da haben wir so eine Art Stammblatt für jede [...]\*. und dann bei  
93 den Interviews selber, muss ein Interviewer und Interviewerinnen ein Protokoll führen. In dem  
94 dann Besonderheiten vermerkt werden, beispielsweise ob alles so gelaufen ist und alles so  
95 war, wie eben in diesem Stammblatt beschrieben oder ob bestimmte Vorkommnisse waren?  
96 Wir hatten auch eine Menge, wir hatten auch beispielsweise einen Feueralarm, also zwei  
97 Feueralarme, also heute hatten wir auch schon einen zweiten Feueralarm. Na also und oder ob  
98 (...) entgegen den Absprachen, keine [...]\*.stattfanden und bestimmten [...]\*. Also das  
99 versuchen wir jetzt im Verlauf sofort zu dokumentieren, es schriftlich von den Interviewern  
100 und Interviewerinnen und jetzt ist auch eine Sache, die ich dann tatsächlich direkt auch immer

101 mache, ist, das ich versuche auch immer hier zu sein, wenn die Interviewer stattfinden,  
102 Interviews stattfinden. Und dann auch kurz danach mit dem Interviewer zu sprechen und zu  
103 fragen, weil manchmal denkt man bestimmte Sachen sind gar nicht wichtig und schreibt man  
104 sie nicht auf. Und dann protokolliere ich das sozusagen, also schaue ich mir immer die  
105 Protokolle noch mal an und schaue (...) und wenn bestimmte Sachen, die vielleicht eher so  
106 (unv.) #00:08:36-2# nicht so wichtig waren, aber wo ich denke, das ist ganz gut das zu wissen,  
107 dann protokolliere ich das auch. Das Zweite, was wir machen, wobei wir haben jetzt gerade  
108 angefangen, ist, dass wir dokumentieren, also es sind jetzt PAPI - Fragebogen, also Pen and  
109 Pencil nee, Paper and Pencil heißen das, das ist Papierfragebogen. Die Daten müssen natürlich  
110 eingelesen und eingegeben werden, ist, dass wir auch alles dokumentieren, also  
111 beispielsweise, wenn jemand eine Antwort korrigiert hat, weil das müssen wir sowieso per  
112 Hand nachträglich also korrigieren. Oder wenn beispielsweise eine Person nicht verstanden  
113 hat, dass nur eine Antwort möglich war und dann hat zwei Antworten angekreuzt. Also ist  
114 natürlich immer der Mittel der Befragung, also bei bestimmten Befragungen wie online, also  
115 wenn man das so programmiert, dass es nicht möglich ist, zweifache also Mehrfachnennung zu  
116 machen, kann es ganz nicht passieren, der Befragte selber merkt es sofort. Bei PAPI ist  
117 natürlich, hat man sich was Schönes vorgestellt, aber natürlich dann sind die Leute auch  
118 INDIVIDUELL (lacht) und das dokumentieren wir auch. Das müssen wir auch dokumentieren,  
119 weil wir danach natürlich auch entscheiden müssen, wie wir diese Daten edieren. Also und das  
120 ist und das läuft JETZT, also sobald wir die Daten angeben, alle Arten von Problemen, auch  
121 Probleme, die eben nicht unbedingt Probleme sind, aber wo wir sagen, okay, wenn wir das  
122 machen, machen wir das jetzt auch einfach ordentlich. Dokumentieren wir das alles, wenn es  
123 fertig ist, okay. #00:10:04-0#

124  
125 I: Okay und diese Dokumentationen, wie lange werden die aufbewahrt für den gesamten  
126 Verlauf der Studie oder noch länger, dass man im Nachhinein noch mal nachschauen kann?  
127 #00:10:14-6#

128  
129 B: Ach das ist eine gute Frage, so genau weiß ich das nicht. Ich glaube, auf jeden Fall bis (...)   
130 Verlauf der Studie, ich glaube nicht, dass es darüber hinaus sein wird. Weil (...) also gerade,  
131 also (...) also ich hätte ein bisschen Angst, dass es sensible Informationen da drinnen halt,  
132 allgemein sind zwar meistens keine Namen oder so was, also wenn Namen sind, dann wird das  
133 jetzt nicht protokolliert, sondern sofort gelöscht. #00:10:46-7#

134  
135 I: Mhm. #00:10:47-5#

136  
137 B: Aber beispielsweise, ich meine, wenn man zwei Sachen, zwei Kreuze macht, ist eine Sache,  
138 aber wenn man beispielsweise bei offener Angabe, zum Beispiel nur sagen sollte, welchen [...] \*  
139 man machen will [...] \* und jemand zwei [...] \* eingibt, ich weiß es nicht. Ich glaube es wäre  
140 einfach für mich besser, also aber darüber haben wir noch gar nicht diskutiert/ #00:11:07-4#

141  
142 I: Aha. #00:11:07-6#

143  
144 B: /.. wenn wir dann nach Ablauf der Studie sie löschen, weil nach Ablauf der Studie ist es  
145 sowieso auch, also wegen der Datenweitergaben gerade, muss man sich sowieso viele  
146 Gedanken machen, weil da müssen sie anonymisiert sein und [und so weiter] \*. Zusätzlich auch  
147 denke, dass gar keine offenen Angaben sein werden, sondern nur codierte Angaben und dann  
148 nützt auch nichts, also das muss ich dann dokumentieren, welche Regeln ich ja getroffen habe.  
149 Ob ich sage, okay bei unzulässigen Mehrfachangaben würde Code-Missings oder eben eine  
150 Editions- Missing so vergeben oder bei bestimmten Sachen, ich denke jetzt bei [...] \*, wenn man  
151 sagt, okay, was versucht das sozusagen in eine Kategorie zu binden, wenn das möglich ist. Also  
152 man muss das dokumentieren und sagen, wie man das gehabt hat, aber man muss jetzt nicht

153 wissen, Fall B - hat das und das angegeben. #00:11:59-4#  
154  
155 I: Ja. Und diese normalen Dokumentationen, warum das und das so gemacht wurde. Wird das  
156 dann noch weiterhin aufbewahrt oder? #00:12:07-6#  
157  
158 B: Ja, ja das muss man machen, weil man das dann nachher auch nicht mehr weiß/ #00:12:10-  
159 5#  
160  
161 I: Mhm. #00:12:11-2#  
162  
163 B: / ...unter Umständen. Also nicht nur bei der Datenweitergabe, sondern unter Umständen  
164 weiß ich auch nicht mehr in drei Jahren, wie haben wir damals entschieden? Also ich glaube, es  
165 ist auch ganz sinnvoll für einen selber. #00:12:23-1#  
166  
167 I: Ja. #00:12:23-2#  
168  
169 B: Also ich merke das jetzt auch, was Sie vorhin gefragt hatten, nämlich bei diesen, also was  
170 eher bei [...]\* angeht, deshalb mache ich das jetzt auch sofort. Weil irgendwie, die erste Woche  
171 wusste ich noch, [...]\* „A“ war so und [...]\* „B“ war so und [...]\* „C“ war so. Schon ab  
172 [...]\* wusste ich nicht mehr, welche war [...]\*, bei [...]\*/ #00:12:41-5#  
173  
174 I: (lacht). #00:12:41-7#  
175  
176 B: /...also, wenn man das nicht sofort dokumentiert, ist dann weg. #00:12:44-4#  
177  
178 I: Ja. #00:12:45-3#  
179  
180 B: Und was man schriftlich hat, ist erst mal da/ #00:12:47-8#  
181  
182 I: Mhm. #00:12:48-2#  
183  
184 B: / .. also ob man das danach noch versteht, weil manchmal macht man einen ganz kurzen  
185 Stichpunkt, ist ja nur was anderes ohnehin. #00:12:53-3#  
186  
187 I: Mhm. #00:12:54-3#  
188  
189 B: Deshalb glaube ich, ist einfach wichtig, das wichtig natürlich für außerhalb, dass jemand  
190 fragen kann, aber es ist auch wichtig für einen selber, dass man weiß, okay wie hat man  
191 entschieden unter Umständen, das merkt man erst später. Das war nicht die gute  
192 Entscheidung, das muss man auch dokumentiert haben, beispielsweise in welchen Fällen hat  
193 man was entschieden? #00:13:11-8#  
194  
195 I: Ja. #00:13:12-6#  
196  
197 B: Und deshalb ist einfach wichtig und in dem Fall bei uns [...]\*, wäre auch schön, wenn man  
198 beispielsweise, wenn sich als gute Lösung bestimmte Regeln, die wir auch noch, also wir sind  
199 gerade wie gesagt, noch bei den Interviews. #00:13:25-0#  
200  
201 I: Mhm. #00:13:25-7#  
202  
203 B: Mit bestimmten Regeln auch beibehält und nicht jedes Mal ad hoc entscheidet, also deshalb  
204 macht man auch so eine Beziehung. #00:13:32-6#

205  
206 I: Und sind Sie [...] \* dafür verantwortlich ist dann, weil Sie meinten, Sie notieren sich das dann  
207 gleich, also sind Sie [die Person, die] \* diese Sachen dokumentiert? #00:13:41-5#  
208  
209 B: Also jetzt bei den Daten selber nicht. Das machen die studentischen Hilfskräfte. #00:13:46-  
210 4#  
211  
212 I: Mhm. #00:13:46-8#  
213  
214 B: Bei den Protokollen machen das die Interviewer, aber ich schaue mir immer die Protokolle  
215 an, ohne einen Mitarbeiter, ich habe mir erst, auch die ersten Protokolle von den Daten  
216 angeguckt und jetzt wird das [...] \* machen. Also ich werde schon am Ende, aber dann nicht  
217 jetzt bei [...] \*, sondern ich werde jetzt, also ich erstmal einfach bei [...] \*, weil ich auch die  
218 meisten Erfahrungen habe, irgendwie so ein bisschen geguckt, was lohnt sich. Ist natürlich  
219 auch viel Arbeit diese ganze Dokumentation, nicht, also ist immer auch eine Abwägungssache.  
220 #00:14:17-0#  
221  
222 I: Aha. #00:14:17-2#  
223  
224 B: Und dann, ja ist viel einfacher irgendwas zu korrigieren ohne es zu dokumentieren oder  
225 nicht? Aber eben danach ist es weg. Danach weiß man das nicht mehr, vor allen Dingen in  
226 unserem Fall, weil wir haben ja Papierfragebogen, wir haben nicht unterschiedliche  
227 Datensätze, wie in anderen, nicht. Da hat man einen unedierten Datensatz und dann hat man  
228 einen edierten Datensatz, aber unser Datensatz ist erstmal Papier, also wenn ich das direkt in  
229 den Daten ändere, dann muss ich das noch mal in Papier nachschauen. #00:14:45-0#  
230  
231 I: Mhm. #00:14:46-0#  
232  
233 B: Ja. Nee aber es werden die studentischen Hilfskräfte das machen und dann, die  
234 wissenschaftlichen Mitarbeiter sozusagen ja. Anschauen und dann eben, wenn die dann eine  
235 Masse auch an [...] \* haben, dann kann man auch schauen, okay welche Probleme gibt es eben,  
236 die regelmäßig auftauchen und da kann man sich auch nach regelmäßigen Lösungen suchen.  
237 Also jetzt, was jetzt für (unv.). #00:15:08-7#.  
238  
239 I: Ja okay. Jetzt muss ich mal kurz schauen. (lacht) #00:15:15-3#  
240  
241 B: Ja. #00:15:16-4#  
242  
243 I: (...) Zur Datenweitergabe, wann und unter welchen Bedingungen werden die Daten dann  
244 weitergegeben? Also wenn das Projekt auch zu Ende ist? #00:15:30-0#  
245  
246 B: Da muss ich auch unser Protokoll, nee nicht Protokoll, wie heißt das?  
247 Prozedurenbeschreibung nachschauen, nicht dass ich jetzt was Falsches sage. Also so lange das  
248 Projekt läuft, werden sowieso keine Daten weitergegeben. #00:15:48-4#  
249  
250 I: Mhm. #00:15:49-1#  
251  
252 B: Zurzeit können sowieso nur Projektmitarbeiter und Mitarbeiterinnen Zugang zu den Daten  
253 haben, also zurzeit haben wir gar keine Daten. (lacht) Wir erheben sie gerade. (lacht)  
254 #00:16:01-4#  
255  
256 I: (lacht). #00:16:01-8#

257  
258 B: Und am Ende werden Daten weitergegeben in anonymisierter Form, was es genau heißt,  
259 also müssen wir uns eben im Laufes des Projektes Gedanken machen. #00:16:17-9#  
260  
261 I: Mhm. #00:16:18-9#  
262  
263 B: Also wobei klar ist, keine personenbezogenen Daten, also Namen oder so was, also ich  
264 meine die fragen das gar nicht, aber es könnte natürlich sein, dass irgendjemand einen Namen  
265 drin schreibt oder irgendwie. #00:16:31-1#  
266  
267 I: Ja. #00:16:31-6#  
268  
269 B: Aber jetzt natürlich könnte auch sein, dass eine Kombination von Merkmalen einfach macht  
270 die Leute zu rekonstruieren, wer die Leute sind/ #00:16:40-2#  
271  
272 I: Mhm. #00:16:40-4#  
273  
274 B: / ..und darüber muss man sich erst mal Gedanken machen, auch Gedanken machen in Bezug  
275 auf Wellen, weil jetzt bei einer Datenerhebung glaube ich, ist jetzt nicht so problematisch, ich  
276 meine, die sind alle [...]\*. Also dass man sich dort auch Gedanken macht, welche Kombination  
277 von Merkmalen man unter Umstände gröber darstellt. Also für die Datenweitergabe darstellen  
278 muss. #00:17:22-9#  
279  
280 I: Mhm. #00:17:23-4#  
281  
282 B: Und dann, also das ist auf jedem Fall eine Sache, die wir machen werden, dann werden wir  
283 glaube ich, (...) ich bin mir nicht sicher, ob wir das jetzt machen oder machen, nicht machen  
284 werden. Aber auf jeden Fall war ein Gedanke, dass wir auch noch mal neue Nummer vergeben.  
285 #00:17:36-3#  
286  
287 I: Mhm. #00:17:37-7#  
288  
289 B: Also dass es noch schwieriger ist zu rekonstruieren unter Umständen, welche [...]\*.?  
290 #00:17:45-3#  
291  
292 I: Mhm, also Identifikationsnummer? #00:17:47-3#  
293  
294 B: Neue Identifikationsnummer vergeben. Und genau. Und dann gibt es irgendwie die  
295 institutionelle Anbindung, also sprich, da werden Daten weitergegeben, nur an Leute die eine  
296 institutionelle Anbindung vorweisen können, weil meistens, also muss man hier auch am WZB  
297 versprechen, dass man mit, ja also wissenschaftlichen Arbeiten und [so weiter]\*. Und  
298 Datenschutz und so und ich glaube genau, wenn wir irgendwie Scientific User File glaube ich  
299 machen. Also eben, dass also, nachdem wir eben überlegt haben, was dürften wir weitergeben  
300 oder nicht? Aber wenn Ihnen das ganz konkret interessiert, wäre am besten wenn ich diesen  
301 Teil der Pozedurenbeschreibung weiterleite und ich weiß nicht oder hatte ich es/ #00:18:41-1#  
302  
303 I: Ich hatte von, ich hatte es von Herrn Habich schon bekommen. #00:18:42-9#  
304  
305 B: Genau, weil ehrlich gesagt, das haben wir [...]\*. gemacht und auch mehrmals diskutiert,  
306 sowohl im Projekt, was können wir leisten, was können wir nicht leisten? #00:18:50-7#  
307  
308 I: Mhm. #00:18:51-9#

309  
310 B: Weil das ist auch immer ein Problem. Anonymisierung ist ein extra Arbeitsschritt, und  
311 Projekte sind aber nie so (...) weder ausgestattet, noch auch von der Laufzeit, nicht, also ich  
312 meine, man hat irgendwie gerade irgendwie was erhoben, man muss eigentlich  
313 veröffentlichen. Muss das und das und das und da bleibt man Daten, das Projekt ist zu Ende,  
314 die Stelle ist zu Ende, so sind die Arbeitsbedingungen in der Wissenschaft, also in meiner  
315 Freizeit mache ich keine Datenanonymisierung. #00:19:18-7#  
316  
317 I: Mhm. #00:19:19-1#  
318  
319 B: Also wenn an sich früh genug Gedanken macht, dann kann man das, (lacht) das ist jetzt ein  
320 bisschen eher unser Gedanke, dass wir jetzt einfach überlegen, weil vor allen bei der nächsten  
321 Welle, also ab Anfang schwierig ist. #00:19:29-3#  
322  
323 I: Mhm. #00:19:30-6#  
324  
325 B: Aber wie gesagt, um das genauer nach wie dann am Endeffekt, wir das dann entschieden  
326 haben, da würde ich noch mal auf die Prozedur schauen/ #00:19:38-0#  
327  
328 I: Mhm. #00:19:38-4#  
329  
330 B: Ich weiß sie nicht mehr auswendig. #00:19:40-3#  
331  
332 I: Okay. (lacht) Wie sieht das zum Beispiel aus mit Embargo-Fristen, also würden Sie die Sachen  
333 oder die Daten weitergeben nach einer bestimmten Zeit, nachdem Sie die Ergebnisse selbst  
334 veröffentlicht haben erst oder würden Sie die sofort weitergeben nach Studienende?  
335 #00:19:57-6#  
336  
337 B: Ja ich, ich glaube das ist auch eher eine Sache, die man nach Studienende entscheidet, so  
338 dann auch mal entscheidet, wie viel man veröffentlicht hat oder nicht. #00:20:09-9#  
339  
340 I: Mhm. #00:20:10-7#  
341  
342 B: Beziehungsweise, wie weit ist man mit bestimmten Veröffentlichungen oder nicht?  
343 #00:20:15-1#  
344  
345 I: Mhm. #00:20:16-7#  
346  
347 B: Also ich würde das jetzt nicht ausschließen wollen, dass wir das unter Umständen auch  
348 machen. #00:20:22-3#  
349  
350 I: Mhm. Eine Embargo-Frist einhalten? #00:20:24-6#  
351  
352 B: Genau, dass man sagt, ja man gibt die Daten, aber man gibt sie erst nach einem Jahr oder  
353 zwei Jahren raus. #00:20:30-1#  
354  
355 I: Mhm mhm. Und wenn Sie das so entscheiden würden, was wäre der Grund dahinter?  
356 #00:20:33-5#  
357  
358 B: Also ich wäre sowieso nicht [die Person, die]\* das entscheidet, sondern das wären die [...] \*  
359 Projektleiter, die [...] \*, die das Projekt auch angeworben haben, aber noch mal die Frage?  
360 #00:20:47-2#



361  
362 I: Wenn Sie sich für eine Embargo-Frist entscheiden würden, also dass sie zum Beispiel erst ein  
363 Jahr nach Veröffentlichung die Daten weitergeben, aus welchem Grund wäre das dann?  
364 #00:20:56-9#  
365  
366 B: Also es hat weniger mit, ein Jahr nach Veröffentlichung zu tun, sondern dass man selber  
367 auch in der Lage ist zu veröffentlichen. #00:21:06-4#  
368  
369 I: Mhm. #00:21:06-9#  
370  
371 B: Also es wäre ja jetzt nicht so, dass ich sage und das darf erst ein Jahr, nachdem ich  
372 veröffentlicht habe, ausgegeben werden, sondern ich will zumindest ein oder / also ich kann  
373 jetzt nicht dazu oder nicht so, wie so, so viel, wie ich wollte, um Muster zu veröffentlichen und  
374 erst nachdem, also wenn ich jetzt sozusagen ich oder [...] die Chancen haben. Und der Grund  
375 ist einfach, ich meine Daten erheben ist ein Haufen, Haufen, Haufen Arbeit. #00:21:37-0#  
376  
377 I: Mhm. #00:21:38-5#  
378  
379 B: Ich weiß jetzt nicht, inwiefern Sie sich mit dem wissenschaftlichen System auskennen? Aber  
380 keiner von uns ist unbefristet und es wird einfach geguckt, was man veröffentlicht hat.  
381  
382 I: Mhm. #00:21:51-2#  
383  
384 B: Und wie viel Zeit. Und eigentlich strategisch hat das, strategisch soll man immer sagen, nein  
385 ich erhebe keine Daten. Das ist meine zweite Datenerhebung, ich werde in meinen Lebenslauf  
386 ziemliche Löcher haben, weil ich eben Daten erhoben habe und am Endeffekt  
387 vollzeitbeschäftigt war mit Datenerhebung. Und in der Zwischenzeit nicht, also veröffentlicht  
388 habe und bevor man die Daten, nachdem man die Daten erhoben hat, muss man sie eben  
389 auch eingeben, je nachdem oder edieren. #00:22:19-2#  
390  
391 I: Mhm. #00:22:19-5#  
392  
393 B: Das muss man immer und erst dann kann man veröffentlichen, sehr oft, eine  
394 Veröffentlichung dauert genau ein (unv.) #00:22:26-0# dann hat man den Datensatz, also ich  
395 glaube quantitativ, aber qualitativ. Dann muss man erst mal aufbereiten und dann macht man  
396 die Analysen und dann schreibt man und dann macht man noch mal die Analysen. Und das  
397 reicht man irgendwo ein und dann muss man das noch mal überarbeiten. Und das fände ich  
398 ehrlich gesagt auch ein bisschen unfair, also ich sage das ehrlich jetzt auch über mein altes  
399 Projekt, aber auch in diesen, dann ist man irgendwie in der Laufzeit des Projekts ziemlich gut  
400 beschäftigt Daten zu erheben. #00:22:51-5#  
401  
402 I: Mhm. #00:22:52-9#  
403  
404 B: Und dann ist das Projekt zu Ende, am besten Fall im Laufzeit dieses Projekt hat man auch  
405 mal die Daten anonymisiert, sonst kann man sie auch gar nicht weitergeben. #00:23:00-3#  
406  
407 I: Mhm. #00:23:00-6#  
408  
409 B: Also sprich, man macht irgendwie nur eine reinste Dienstleistung, also es wäre eine reinste  
410 Dienstleistung und dann gibt man die Daten weiter. Und mache ich ein Beispiel und Sie  
411 kommen und benutzen Sie meine Daten und ich unter Umständen habe ich nicht mal eine,  
412 eine Veröffentlichung daraus, aber die ganze Arbeit reingesteckt. #00:23:14-9#

413  
414 I: Mhm. #00:23:15-7#  
415  
416 B: Und unter Umständen habe ich nicht mal mehr eine Stelle, weil das Projekt abgelaufen ist  
417 dann. (lacht) #00:23:21-5#  
418  
419 I: (lacht) Ja. #00:23:21-9#  
420  
421 B: Also von der Perspektive, wie gesagt ist jetzt weniger, wie ich veröffentlichen in 2012 und  
422 die Daten dürfen erst in 2014, sondern man sagt, okay das Projekt lief bis 2016 und man hat in  
423 diesen drei Jahren sehr, sehr viel Arbeit reingesteckt. #00:23:37-8#  
424  
425 I: Mhm. #00:23:38-5#  
426  
427 B: Dass man sagt, zumindest die Leute, die eben an der Datenerhebung und Konzeption und  
428 [so weiter]\*, sollen sie zumindest die Chancen haben. Ob sie eine Stelle haben, steht das auch  
429 noch in einem anderen Stern, aber zumindest die Chancen haben, irgendwas mit diesen Daten  
430 zu machen, bevor eben jemand anderer kommt, der eben mit diesen Daten noch gar nichts zu  
431 tun hatte. Und der eben in der Zeit, wo man in der Datenerhebung war, schön mit anderen  
432 Daten, die von jemand anderen erhoben wurden, schön publiziert hat. Weil ich meine sonst,  
433 wie gesagt kann man gleich packen und gehen und sagen okay, wo ist das nächste Arbeitsamt?  
434 #00:24:10-9#  
435  
436 I: (lacht) #00:24:11-7#  
437  
438 B: Also das finde ich, gegeben den Arbeitsbedingungen ist (...) fänd ich fair. #00:24:20-3#  
439  
440 I: Mhm okay. #00:24:21-3#  
441  
442 B: Aber wie gesagt, ist auch nicht meine Entscheidung, aber? (lacht) #00:24:24-1#  
443  
444 I: Okay. (lacht) Wie sieht das eigentlich aus mit der Archivierung der Daten? Also die werden  
445 noch eine bestimmte Zeit aufgehoben, aufbewahrt nach Studienende? #00:24:36-3#  
446  
447 B: Also welche Daten meinen Sie? #00:24:38-2#  
448  
449 I: Also die Daten, die Sie erhoben haben? #00:24:40-1#  
450  
451 B: Personenbezogene Daten oder die Erhebungsdaten sozusagen? #00:24:44-5#  
452  
453 I: Die - alle oder gibt es nur einen Teil davon - also? (lacht) #00:24:48-9#  
454  
455 B: Also die, sagen wir so die Daten die aus dem Fragebogen, die man aus den Fragbogen  
456 gewonnen hat, na die bleiben ja ewig bestehen. #00:25:01-8#  
457  
458 I: Mhm. #00:25:02-6#  
459  
460 B: Also ob sie jemand benutzt in zehn Jahren mag ich bezweifeln, aber die bleiben, also das  
461 wäre auch eine Verschwendung, warum sollte man die wegschmeißen? #00:25:10-3#  
462  
463 I: Mhm. #00:25:10-8#  
464

465 B: Personenbezogene Daten, jetzt in dem Fall hier, [...] die werden nach Abschluss der Studie  
466 gelöscht, beide, das sind ja sensible Daten und ich meine, was macht man mit den Adressen,  
467 abgesehen davon, die sind vermutlich nicht mal aktuell, aber selbst wenn? #00:25:30-2#  
468

469 I: Ja okay. (lacht) #00:25:30-4#  
470

471 B: Also wäre ja noch schlimmer, also ich meine, dass sind echt personenbezogene Daten, das  
472 steht auch im Datenschutz, die löschen wir. Also nachdem wir, wenn die Studie abgeschlossen  
473 ist und dementsprechend sie nicht mehr gebraucht werden, um sie wieder zu kontaktieren, die  
474 werden gelöscht. Aber die Informationen, die also die Erhebungsinformation, die sie uns  
475 gegeben haben, die bleibt. #00:25:50-8#  
476

477 I: Mhm. #00:25:51-3#  
478

479 B: Ja. Aber das ist ja immer so, also/ #00:25:54-0#  
480

481 I: Okay. #00:25:55-1#  
482

483 B: Die Lebensverlaufs-Studie, die in den achtziger Jahren gelaufen sind vom Max Planck  
484 Institut, jetzt mal nur ein Beispiel, das ich kenne. #00:26:01-2#  
485

486 I: Mhm. #00:26:01-7#  
487

488 B: Die gibt es noch, weil wie gesagt, es kann jemand heute kommen und sagen, ich möchte  
489 gerne den Aspekt, diesmal ist es natürlich mehr was Historisches, ist nicht mehr ganz aktuell,  
490 aber mit diesem Aspekt von - keine Ahnung, die Kriegsgeneration in Deutschland kommt  
491 demnächst. Oder jetzt mit NEPS hat man Leute interviewt, nicht, die heute 50 sind und wollen  
492 wir vergleichen bestimmte Sachen von denen heute 50- und 60-jährigen zu den Leuten, die in  
493 den 80-iger Jahren 50 / 60, warum denn nicht? #00:26:34-0#  
494

495 I: Mhm. #00:26:34-7#  
496

497 B: Warum soll man sie wegschmeißen, also ich meine, man hat die Leute stundenlang gequält.  
498 (lacht) #00:26:37-8#  
499

500 I: (lacht). Also Sie würden die dann im Prinzip selbst eventuell auch noch mal irgendwann  
501 nutzen oder auch für andere Leute, dass die anderen nachnutzen? #00:26:48-3#  
502

503 B: Nee das wäre auch was anderes, das wäre auch (unv.), #00:26:51-0# ja. (Es klopft an der  
504 Tür) JA? #00:26:54-0#  
505

506 Ende des 1. Teils  
507  
508

509 **Teil b)**  
510

511 I: Okay jetzt kann es weitergehen. Genau, meine Frage war, ob Sie auch für die aktuelle Studie  
512 schon Daten, die Sie nicht selbst erhoben haben nutzen, also quasi selbst Fremddaten  
513 nachnutzen? #00:00:16-3#  
514

515 B: Nee, also für die aktuelle Studie habe ich das, also haben wir das nicht gemacht. #00:00:21-  
516 8#

517  
518 I: Mhm. #00:00:22-1#  
519  
520 B: Ich glaube, wir werden das auch nicht machen, weil wenn wir das machen könnten,  
521 bräuchten wir die Studie ja gar nicht. Also natürlich haben wir bestimmte Daten wie, (...)   
522 welche waren das denn, also wie, (...) wie viele (...) [...] es gibt oder wie die [...] in den  
523 unterschiedlichen [...] sich verteilen. [...] Also das schon, aber in dem Sinne anderen Daten,  
524 im Sinne von [...] nicht. #00:01:08-3#  
525  
526 I: Aha okay. #00:01:09-4#  
527  
528 B: Mhm. #00:01:10-2#  
529  
530 I: Ja jetzt möchte ich Sie gerne von, bezüglich Ihrer eignen Erfahrung allgemein gerne fragen  
531 zur Datenweitergabe. Wissen Sie wie das allgemein von Kollegen auch hier im Haus so  
532 wahrgenommen wird oder praktiziert wird? Also wie, denken alle so wie Sie mit, dadurch, dass  
533 es ja ein sehr hoher Zeitaufwand ist, die Daten zu erheben, möchte man erst mal selbst die  
534 publizieren, bevor die Daten jemand anders dann bekommt, also wie haben Sie, welches  
535 Gefühl haben Sie? #00:01:42-3#  
536  
537 B: Na ja jetzt vom Kollegen und Kolleginnen, also ich meine, jetzt ist natürlich NEPS da und sie  
538 machen massenhaft Datenerhebung, aber das ist ja ganz anders geregelt. NEPS ist eben ein  
539 Datendienstleister. Das ist jetzt kein Projekt oder ansonsten muss ich sagen, kenne ich nicht  
540 sehr viele, also direkt mittelbar Kollegen und Kolleginnen die Daten erheben und also natürlich  
541 kann man sich jetzt fragen, also welche Daten? Also ich kenne natürlich Kollegen und  
542 Kolleginnen, die qualitative Daten erheben. Und die geben definitiv ihre Daten nicht weiter.  
543 Aus vielen Gründen, viele kann man auch nachvollziehen, viele Dinge finde ich allerdings bei  
544 der qualitativen eben Datenerhebung, qualitative Forschung ist schon ein bisschen  
545 problematisch. Weil man quantitativ eben (...) viel, viel mehr dokumentieren muss und eben  
546 zugänglich auch machen muss und transparent machen muss. Und qualitativ wie gesagt, ich  
547 habe Verständnis, dass es ein bisschen schwieriger ist, aber zum Teil dann ist es auch ein  
548 bisschen, (...) wie sagt man, eben nicht immer ganz transparent und nachvollziehbar, was man  
549 macht oder nicht macht. Aber auf jeden Fall, die geben definitiv ihre Daten nicht weiter. Ich  
550 weiß auch von anderen Projekten, die eher beispielsweise Dokumente und  
551 Gesetzsdatenbanken aufgebaut haben, da kann ich ehrlich gesagt nicht sagen, inwiefern, so das  
552 habe ich weder mitbekommen, dass sie die weitergegeben haben, aber auch nicht  
553 mitbekommen. Also „A“ kann ich, würde ich eben nicht meine Hand ins Feuer legen, das wir  
554 nicht gemacht haben, aber auch weiß ich gar nicht, ob sie das nicht gemacht haben, weil sie  
555 gar nicht gefragt wurden, kann auch sein. #00:03:28-5#  
556  
557 I: Okay. #00:03:29-0#  
558  
559 B: Also bei mir war jetzt der Fall, also jetzt noch nicht [...]\*, wobei witzigerweise kam auch  
560 schon (lacht) eine Anfrage, von jemand von außerhalb. Aber ich meine zurzeit habe ich da echt  
561 ein einfaches Spiel, wo ich auch nur denke, sorry, wir haben erst mal gar keine Daten also. Da  
562 brauche ich auch gar nichts zu sagen, und bevor du oder sie die Daten benutzen, möchten wir  
563 gerne hier die benutzen. Also mein altes Projekt war tatsächlich so und das war, also Anfragen  
564 kamen, da habe ich auch nicht Nein gesagt, das kam immer ein bisschen darauf an, davon ab,  
565 wer gefragt hat, also beispielsweise die erste Veröffentlichung hat [eine Studentische  
566 Hilfskraft]\* gemacht, [...] weiß ich mehr. Mit den Daten nicht nur und das war BEVOR  
567 ÜBERHAUPT ich oder meine anderen Kollegen irgendwas mit den Daten gemacht haben, aber  
568 das finde ich jetzt was anderes. Wäre jemand komplett Fremdes gewesen, hätte ich vermutlich

569 Nein gesagt, aber da auch wiederum aus zwei Gründen zu der Zeit oder eigentlich bis jetzt  
570 noch, sind diese Daten noch nicht anonymisiert. Also von meinem altes Projekt und das sind  
571 auf jedem Fall Daten, da muss man viel Arbeit und Energie reinstecken, sie zu anonymisieren.  
572 Das Projekt war sowieso super knapp geplant, ich mache das jetzt nicht nebenbei, also auch  
573 noch dazu. Und von daher wäre eigentlich ein Datenschutzproblem gewesen, also wo sowieso  
574 klar ist, da kann ich nirgend, nirgend jemand, also da kann ich selber nicht mit den Daten nach  
575 Hause gehen sozusagen. #00:05:01-6#

576

577 I: Mhm. #00:05:01-8#

578

579 B: Weil eine mögliche Kooperation wäre jetzt was anderes und jetzt würde ich auch ein  
580 bisschen anders darüber nachdenken, aber wie gesagt, also das ist jetzt nicht per se, dass ich  
581 immer sage, nee bevor ich oder nicht. Also ist schon anders gegangen und ich kann mir aber  
582 schon vorstellen, dass es, es ist eine interessante Frage, werde ich mal nachfragen, dass ich  
583 nicht [die einzige Person]\* bin, [die]\* so denkt. Ich weiß selbst, wenn man eben Dokumente an  
584 der Dokumentendatenbank aufbaut, das ist einfach viel, viel Arbeit, und dass man erst mal  
585 sagt, ja gerne steht das irgendwann zur Verfügung, aber erstmal möchte ich sie benutzen.

586 #00:05:37-2#

587

588 I: Mhm. #00:05:38-5#

589

590 B: Und das, ja wie gesagt, das ist was anderes beim SOEP oder auch bei NEPS, weil das ist eben  
591 anders gedacht, das ist eben nur eine Produktion da gedacht. Aber NEPS kann ich auch nicht  
592 mich melden und sagen, schiebe mir die Daten rüber, man muss auch einen Antrag stellen  
593 oder man muss sich dabei, ich weiß gar nicht, weil ich jetzt mit NEPS noch nicht Interesse  
594 hatte. Aber vielleicht irgendwann werde ich auch Interesse haben, also inwiefern, also beim  
595 SOEP muss man einfach zahlen und irgendwie, nicht. Ich weiß jetzt nicht wie es bei NEPS ist, ob  
596 man beim DFG auch einen kleinen Antrag stellen muss und diese Art von Sachen Arbeit meint,  
597 da sind, wie gesagt die Bedingungen ein bisschen anders. Da ist die Ausstattung allerdings auch  
598 ein bisschen eine andere. Also ich meine, [...]\*. (lacht) #00:06:22-4#

599

600 I: Okay. #00:06:24-1#

601

602 B: Ja ihre Arbeitsverträge sind zum Teil auch ein bisschen anders, also ja. #00:06:29-0#

603

604 I: Ach so okay. (lacht) Mhm. Ja, jetzt muss ich noch gucken, ob ich noch irgendwas, (...) jetzt  
605 muss ich doch noch mal nachfragen bei der, zur Datenweitergabe/ #00:06:43-2#

606

607 B: Ja. #00:06:44-1#

608

609 I: /..also Sie würden, Sie sagten eine studentische Hilfskraft hat einmal Daten in ihrer  
610 Abschlussarbeit verwendet, die ansonsten noch von keinem weiter/ #00:06:49-7#

611

612 B: Ja. #00:06:50-7#

613

614 I: / ..publiziert worden waren? Das konnte sie machen, weil sie im Rahmen hier die, also weil  
615 sie an der Erhebung beteiligt war, also? #00:06:59-2#

616

617 B: Ja. Also sie war jetzt nicht an der Erhebung beteiligt, sie war an der Datenedition beteiligt.

618 #00:07:03-0#

619

620 I: Ah okay. #00:07:03-6#

621  
622 B: Genau. #00:07:03-7#  
623  
624 I: Und wenn jetzt zum Beispiel jemand von einem ganz anderen Fachgebiet kommen würde  
625 und fragen würde, ob er die Daten bekäme? #00:07:12-3#  
626  
627 B: Also wie gesagt, jetzt hat jetzt weniger mit dem Fachgebiet zu tun, also ich meine, in einem  
628 ganz anderen Fachgebiet, da muss ich mich fragen, was willst du mit diesen Daten (lacht)  
629 machen. So und es hat einfach damit zu tun, es ist und das würde ich jetzt auch bei der  
630 aktuellen Studie machen. Also wenn jetzt eine studentische Mitarbeiterin kommen würde,  
631 entweder von uns oder ich meine, es ist ein Kooperationsprojekt mit einem anderen Institut,  
632 das wäre das genau das Gleiche und sagen würde, er möchte gerne meinen Rahmen, also im  
633 Rahmen des Projekts, meine BA oder MA- Arbeit schreiben. Also sprich, er würde einfach so  
634 ein Thema machen, der irgendwie passt, hätte ich auch ÜBERHAUPT KEINE PROBLEME zu  
635 sagen, ja. #00:07:50-9#  
636  
637 I: Mhm. #00:07:51-3#  
638  
639 B: Überhaupt keine Probleme, also wir sind jetzt hier ein relativ großes Projekt, da müsste man  
640 sich absprechen, das nicht zwei gleichzeitig ohne das voneinander zu wissen, das machen.  
641 Aber das sehe ich was anders, weil sie sind eben auch mit involviert in dem Projekt und auch  
642 zum Tragen des Projekts. Plus eben, da ist auch was anderes wegen Datenschutz. Weil ich  
643 meine, wir haben zurzeit selber anonymisierten Daten, wir haben noch nicht anonymisierten  
644 Daten. Würde jemand eben komplett anders, von irgendwo anders herkommen, wie gesagt,  
645 zum einen fände ich das so ein bisschen unfair, gegenüber die die hier arbeiten, nur ich denke  
646 ja, ich möchte auch gerne mal mit den Daten machen, aber leider hatte ich bis jetzt (lacht)  
647 keine Zeit. Oder meine studentischen Hilfskräfte oder [meine Kolleginnen und Kollegen]\*  
648 hatten gar nicht Zeit und das Zweite wäre, (...) ich kann diese Daten gar nicht geben. Ich habe  
649 versprochen, dass ich diese Daten gar nicht gebe, weil die Leute haben mir persönlich es eben  
650 vorgegeben, die eben nicht rechts und links geben sollte. Und das wie gesagt, es wäre dann  
651 zusätzliche Arbeit und ich kann dann nicht kontrollieren/ #00:08:56-9#  
652  
653 I: Mhm. #00:08:57-5#  
654  
655 B: /...das wäre dann, ja. Also das, daran würde ich das festmachen, also. #00:09:02-2#  
656  
657 I: Ja okay. #00:09:02-5#  
658  
659 B: Ja. #00:09:03-2#  
660  
661 I: Wer ist dann eigentlich nach Projektende verantwortlich dafür, für die Datenweitergabe oder  
662 Pflege der Daten? Also wer ist dann/ #00:09:11-6#  
663  
664 B: Na Pflege in dem Fall gibt es ja nicht mehr, aber nee wobei die immer auf dem technischen  
665 Stand sind. Meine Vermutung wäre, dass - dadurch, dass es ein Kooperationsprojekt ist,  
666 müsste man entscheiden, ob wir oder das Partnerinstitut und ob beide das machen. Und dann  
667 kann ich mir vorstellen, dass wenn das hier wäre, die Abteilung hat einen Datenmanager, dass  
668 er dann dafür verantwortlich wäre, dass jemand weiß, wo die Daten sind und das keine  
669 Ahnung, manchmal, also ich habe das jetzt noch nie so erlebt bei Daten, aber trotzdem, ich  
670 meine kennt man manchmal bei alten Word – Programmen, also bestimmte Sachen kann ich /  
671 #00:09:53-4#  
672

673 I: Mhm. #00:09:53-9#  
674  
675 B: /...nicht mehr benutzen, dass sie immer auf dem neuesten technischen Stand sind.  
676 (Telefonklingeln) #00:09:57-7#  
677  
678 Ende Teil 2  
679  
680  
681 **Teil c)**  
682  
683  
684 B: Entschuldigung wo waren wir? #00:00:03-3#  
685  
686 I: (lacht) Wir waren bei Datenweitergabe, wo es nach dem, also wie die, wer verantwortlich ist  
687 dafür nach dem Projektende. #00:00:11-2#  
688  
689 B: Genau, also wo ich einfach denke, wäre eben der Datenmanager, dafür hat man eben einen  
690 Datenmanager, also er macht natürlich viel bei Datenerhebung/ #00:00:19-7#  
691  
692 I: Mhm. #00:00:20-7#  
693  
694 B: /..aber könnte ich denken, dass wäre auch seine Zuständigkeit, wie gesagt, Pflege wäre was  
695 anderes, aber auf jeden Fall ein Ansprechpartner und dass er eben weiß, wer / wo die Daten  
696 sind und unter Umständen, eben wo sie ist, die anonymisierte Version des Datensatzes. Was  
697 muss man machen, weil es Scientific User File, das wär ja wirklich ein Fall auch in enger  
698 Koordination oder Kooperation mit Roland Habich oder wer immer, wenn wir so weit sind,  
699 auch zuständig wäre. Weil eben, also müsste man eben auch mit ihm absprechen, also auch  
700 mit dem Datenschützer absprechen, also wir haben überlegt, das wären die Schritte und  
701 danach wäre das anonym und was heißt dann eine institutionelle Anbindung? Und was  
702 müssen die Leute irgendwie beweisen und, nicht. Wie macht man das per Mittel, also macht  
703 das in, schickt man eine CD, müssen sie hier sein, weil das ja auch noch eine Frage ist. Und bei  
704 bestimmten Datensätzen kannst du, kann man gar nicht benutzen, sondern, also nicht wie den  
705 SOEP oder Mikrozensus bekommt man eine CD, sondern man muss vor Ort sie benutzen.  
706 #00:01:20-5#  
707  
708 I: Mhm. #00:01:21-1#  
709  
710 B: Das wäre auch eine Möglichkeit, ja. #00:01:23-5#  
711  
712 I: Ja okay. #00:01:25-1#  
713  
714 B: Ja und da, ja. #00:01:24-5#  
715  
716 I: Ja natürlich. (lacht) #00:01:26-5#  
717  
718 B: Dafür würde ich mich jetzt auf jeden Fall langfristig nicht zuständig fühlen. (lacht) Nein.  
719 #00:01:30-0#  
720  
721 I: Okay. (lacht) Ich hatte vorhin bei der Datendokumentation noch eine Frage vergessen, die  
722 ich stellen wollte. #00:01:37-1#  
723  
724 B: Ja. #00:01:37-2#

- 725  
726 I: Nutzen Sie denn bestimmte Standards oder Vorlagen, also nach welchem Schema gehen Sie  
727 vor, wenn Sie das / die Studie dokumentieren (...) oder beziehungsweise beschreiben?  
728 #00:01:48-2#  
729  
730 B: Ja. Also bei dieser Prozedurbeschreibung habe ich sehr von [...] \* profitiert. (räuspert sich) Da  
731 ist natürlich diese institutionelle Anbindung sehr von Vorteil. Wir haben, ich habe einfach  
732 gefragt und netterweise haben sie uns ein Beispiel zur Verfügung gestellt und daran haben wir  
733 uns orientiert. Natürlich stark angepasst, weil unser Design ist ein anderes und [so weiter] \*.  
734 #00:02:13-9#  
735  
736 I: Mhm. #00:02:14-2#  
737  
738 B: Und eben wir deswegen keine Datenproduzenten sind, sondern es ein Projekt und so.  
739 #00:02:19-1#  
740  
741 I: Ja. #00:02:19-8#  
742  
743 B: Genau. Und (...) das kann ich sagen war als Vorlage. Aber es war jetzt weniger als Vorlage,  
744 sondern / doch es war eine Vorlage. Aber ich glaube, man muss es jetzt nicht so machen, nur  
745 es hat extrem meine Arbeit erleichtert, dass ich wissen konnte, wie sieht so was aus, was eben  
746 schon über einen [...] \* Datenschützer oder überhaupt Datenschutz in seiner Art  
747 durchgegangen ist. Und muss ich nicht das Rad neu erfinden. #00:02:45-6#  
748  
749 I: Ja. #00:02:45-6#  
750  
751 B: Man bestimmte auch Formulierungen hat, also ist manchmal echt eine Rechtssprache. Und  
752 dann habe ich da natürlich auch noch mit Roland Habich mehrmals bei bestimmten Sachen  
753 gesprochen, dann hat er es auch noch mal geguckt, weil wie gesagt, es gibt bestimmte, für  
754 mich Synonyme und dagegen sind nicht Synonyme. Weil rechtlich heißt das was anderes oder  
755 in dem Fach heißt das was anders. So für die Datenedition, ehrlich gesagt habe ich mich  
756 orientiert an meine letzte Studie und ich versuche irgendwie mich zu erinnern, an wen ich  
757 mich orientiert habe. Ich glaube, ich habe vermutlich so eine Vorlage von Daten, also von so  
758 einer Art Maske, die bei der [...] \* studie, benutzt worden ist. Natürlich angepasst dann an den  
759 jeweiligen Feldern. #00:03:36-2#  
760  
761 I: Mhm. #00:03:36-5#  
762  
763 B: Genau, weil der Datenmanager hat jahrelang, also bei der [...] \* studie, wo auch dann eben  
764 Daten erhoben wurden, dort auch ediert wurden und von daher hat man eben ein bisschen  
765 Erfahrung. Also wie gesagt, man muss es nicht unbedingt machen, ich habe keine Ahnung wie  
766 SOEP ihre Daten dokumentiert, ob sie die überhaupt dokumentieren, also gehe ich davon aus,  
767 sie machen ja kaum Edition. Es macht einfach nur einfacher, wenn man weiß, okay jetzt muss  
768 ich mir das nicht von selber überlegen. Dafür sagt man okay, das kann ich diese Vorlage und  
769 dann passe ich meine Vorlage, also diese Vorlage an meine Bedürfnisse, so machen wir das.  
770 #00:04:14-5#  
771  
772 I: Können Sie mir kurz erklären, wie die Vorlage aussah, also welche Felder oder welche Sachen  
773 standen da drin, also wie muss ich mir das vorstellen? #00:04:22-2#  
774  
775 B: Also jetzt bei den Daten beispielsweise? #00:04:23-7#  
776



777 I: Ja. Sie meinten, Sie haben so eine Maske benutzt? #00:04:26-9#  
778  
779 B: Nee eine Maske ist eine, nee Maske war das falsche Wort, ich sage mal eine Excel Tabelle  
780 und da ist irgendwie eine Spalte, ganz klar muss die Case-ID, also die (unv.) -Nummer sein.  
781 Sonst (lacht) braucht man die gar nicht zu führen. #00:04:40-2#  
782  
783 I: Ja. (lacht) #00:04:40-1#  
784  
785 B: Und dann steht, beides benutzen wir kaum, aber das haben wir sehr viel für meine alten  
786 benutzt, den Bereich, also wo das Problem auftauchte und wie gesagt jetzt bei uns war, ich  
787 habe das übernommen. Aber im Endeffekt, ich meine, der Fragebogen ist auch kürzer und  
788 einfach, ich meine die sind eben [...]\*. Ob jetzt so wichtig ist zu wissen, die Frage ist eher eine  
789 Frage, [...]\* ich meine 99 Prozent der Fragebögen sind [...]\*, also nee. Benutzen wir kaum.  
790 Dann natürlich eine Fragenummer, dann die Beschreibung des Problems. Und dann haben  
791 wir eine Spalte - Lösung Fragzeichen, also Editions-vorschlag, also jeder soll eben, bin ein großer  
792 Fan von Teams und jeder, also mehr Leute denken mehr als eins. #00:05:36-2#  
793  
794 I: Mhm. (lacht) #00:05:37-7#  
795  
796 B: Und dann kommt eine Spalte und dann eben Datum und Kürzel und keine Ahnung was und  
797 dann kommt, die haben wir noch nicht benutzt, aber die werden wir dann benutzen –  
798 Editionswege. Und die Idee ist dann einfach, warum man das macht, es natürlich interessiert  
799 mich zu wissen ob bestimmten Personen, deshalb die (...) viele Probleme hatten dann und  
800 dann unter Umstände weiß ich, der oder sie vielleicht hat den Fragebogen nicht so ganz  
801 verstanden oder da lief irgendwas offensichtlich schief. Aber gerade für die Edition interessiert  
802 mich beispielsweise erst zu sortieren nach bestimmten gleichen Problemen. #00:06:10-6#  
803  
804 I: Mhm. #00:06:11-8#  
805  
806 B: Also, nicht, also das, also entweder als Rückmeldung, ha die Frage haben wir uns nett  
807 überlegt, aber offensichtlich, da hinkt etwas oder eben okay, da das sind ähnlichen Problemen,  
808 dann können wir ähnliche Lösungen suchen. Und dann, wenn wir eine Lösung gefunden haben  
809 für dieses Problem, dann ist es für alle gleich. Und nicht für Person „A“ entscheide ich diese  
810 Lösung und dann kommt da irgendwie zwei Wochen später Person „B“, ich hab vergessen,  
811 dass ich das Problem schon hatte und entscheide ich anders. Also damit hat es eher zu tun.  
812 #00:06:44-6#  
813  
814 I: Mhm. #00:06:45-0#  
815  
816 B: Und dann kommt eben danach, wenn man das eben umsetzt, aber so weit sind wir ja noch  
817 gar nicht, dass man sagt - umgesetzt. Also man hat ja okay, jetzt haben wir die Regel oder die  
818 Entscheidung getroffen und dann haben wir tatsächlich die Daten geändert. Aber soweit sind  
819 wir noch nicht, weil ja erst mal die Daten noch eingeben. #00:07:00-0#  
820  
821 I: Mhm okay.(lacht) Ja ich glaube, wir haben eigentlich alles soweit besprochen, was ich mir so  
822 vorgestellt habe. #00:07:09-5#  
823  
824 B: Ja schön. #00:07:10-8#  
825  
826 I: Genau. Genau. #00:07:13-9#  
827  
828 B: Okay. #00:07:14-2#

829  
830 I: Dann bedanke ich mich sehr herzlich. (lacht) #00:07:15-6#  
831  
832 B: Nichts zu danken. #00:07:18-3#  
833  
834  
835 Ende des Interviews  
836  
837  
838 Anmerkung der Autorin: [...] \* = anonymisiert

## 1 C2) Interview 2

2 I: Okay, also es geht um das Thema Forschungsdaten und ich habe vier Leitfragen vorbereitet.  
3 Aber Sie können auch einfach erzählen, wie es Ihnen einfällt und gucken wir mal wie wir dann  
4 zurecht kommen und ich stelle auch gegebenenfalls Rückfragen. 00:00:17-8

5  
6 B: Mhm 00:00:18-4

7  
8 I: Zunächst würde mich interessieren welche Arten von quantitativen Daten Sie als  
9 Forschungsdaten wahrnehmen? Also es geht nur um quantitative Daten. Also was sind für Sie  
10 Forschungsdaten, welche Merkmale oder Eigenschaften müssen Sie besitzen, damit sie  
11 Forschungsdaten sind? 00:00:36-8

12  
13 B: Mhm. Gut die Frage ist halt, ob was erst zu Forschungsdaten wird, wenn ich die Rohdaten  
14 quasi in eine beforschbare (...) Form bringe. (...) Ehrlich gesagt, (lacht) ich kann mit der Frage  
15 nicht so viel anfangen? 00:01:13-1

16  
17 I: Okay. 00:01:11-8

18  
19 B: Also. 00:01:14-8

20  
21 I: Ja ist einfach nur welche Daten zum Beispiel sind in diesem Projekt entstanden, also in der  
22 Studie die noch läuft. Also was sind das für Daten? 00:01:22-8

23  
24 B: Okay. Also das sind quasi Interviewdaten aus größtenteils standardisierten, Fragebögen,  
25 also das heißt, in dem Sinne ist es auch eine Mischung aus, also ja gut quantitativen Daten  
26 oder primär quantitativen Daten einfach durch die Standardisierung. Und die offenen Fragen,  
27 es gibt natürlich auch offene Fragen und dann kann okay, was sind das jetzt für Daten? Weil sie  
28 halt noch nicht in der quantifizierten Form vorliegen, sondern also qualitativ ausgewertet  
29 werden können. Und das ist wahrscheinlich auch häufig der Fall. Dass die Daten ja per se nicht  
30 rein quantitativ oder qualitativ sind, sondern es ja auf die Auswertung (lacht) auch ankommt.  
31 Aber da ist natürlich Frage, okay, was wo (haben sie?) 00:02:13-1 die Erhebung, was für  
32 Instrumente benutze ich dann, dann ist es meist ja relativ klar. Dahin gehend also, wenn dann  
33 die Frage ist, eben, was sind Forschungsdaten? Würde ich sagen, das ist ja schon eben ein  
34 zweiter Schritt, wenn Daten aufbereitet sind oder zumindest aus einer bestimmten  
35 Perspektive gesehen werden. Hinsichtlich schon der Auswertung, also da sehe ich das dann als  
36 quantitative Daten, oder wenn man zum Beispiel Prozessdaten hat. Ich habe auch mit Daten  
37 der Rentenversicherung gearbeitet zum Beispiel aus dem Forschungsdatensatz. Dann kann ich  
38 es natürlich auf alle möglichen Wege auswerten, aber es ist halt erst mal ein Test und dann (...)  
39 steckt dahinter immer schon eine Auswertungsintensivierung, aber wenn ich da jetzt Kriterien  
40 angeben müsste. Es gibt natürlich Daten, über die sich dann der Forscher mehr freut, ja, was  
41 irgendwie Fallzahlen angeht oder die Selektion oder Stichprobenqualität, ja. Oder aber  
42 prinzipiell kann ich keine Definition für Forschungsdaten geben. 00:03:35-1

43  
44 I: Okay. Gut dann gehen, machen wir mit der zweiten Frage weiter. Da geht es jetzt um den  
45 Datenlebenszyklus. Sagt Ihnen der Begriff etwas oder soll ich dazu was sagen? 00:03:44-9

46  
47 B: Ich kann mir ein bisschen was darunter vorstellen, aber also, man muss halt dazu sagen,  
48 [...] also? 00:03:55-0

49

- 50 I: Das macht nichts. (lacht) 00:03:56-5  
51
- 52 B: Okay, ja dann, vielleicht können kurz was, was sagen dann? 00:03:59-3  
53
- 54 I: Ja. Dann würde ich erst mal gerne wissen, was stellen Sie sich ja drunter vor, dann vielleicht  
55 ist es das ja. (lacht) 00:04:03-9  
56
- 57 B: Ja gut Datenlebenszyklus, jetzt weiß ich nicht, dazu gehört wahrscheinlich irgendwie eine  
58 erste Phase der Erhebung, also wenn Daten erhoben werden, also wenn es eine  
59 Primärerhebung gibt, dann die Primärerhebung, dann die Datenaufbereitung, Bereinigung und  
60 wie, was auch immer einen auswertbaren Zustand in den Datensatz zu bringen. Dann hängen  
61 an die Primärauswertung der, also meist haben dann die Forscher die die Daten erheben auch  
62 noch Auswertungsinteresse. Bei uns ist es jetzt auch ein bisschen zu kurz gekommen, also es  
63 ist im Moment mit [...] Dann Sekundärauswertung beziehungsweise noch die Daten, (...) ja  
64 gut Daten aufgehoben, registriert werden, zugänglich sind, dann eine Sekundärauswertung.  
65 Mehr weiß (lacht) ich, ob zu einem Zyklus auch die Löschung von Daten gehört theoretisch,  
66 was ja immer schade ist, ja aber wahrscheinlich auch vorkommt. (lacht) 00:05:13-0  
67
- 68 I: Okay. 00:05:12-8  
69
- 70 B: Ja. Das stelle ich mir darunter vor. 00:05:17-8  
71
- 72 I: Ja. Das sind auch im Prinzip die Phasen, aber nicht alle, also es gehört zum Beispiel auch als  
73 Abschlussphase der Löschvorgang oder eben die Archivierung der Daten dazu, wenn sie nicht  
74 gelöscht werden. Aber das sind im Prinzip die Phasen, die Sie auch genannt haben. 00:05:30-1  
75
- 76 B: Beziehungsweise, gut was da nicht, eine Menge interne Daten ist halt (unv.) 00:05:32-4 und  
77 eben dann Anonymisierung, also wie dann meinetwegen Adressdaten und so dann trennen na  
78 ja, ja. Sie kennen (lacht) sich ja wahrscheinlich besser aus, als ich. Ja. 00:05:44-7  
79
- 80 I: (lacht) Okay das war eigentlich auch schon fast die Frage. Mich würde noch interessieren, in  
81 welcher Phase entstehen denn die meisten Daten? Also wie sehen Sie das? 00:05:56-2  
82
- 83 B: Die meisten Daten entstehen (...) ja eigentlich während der Erhebungsphase. Also (...) die  
84 Frage ist, ob die Forschungsdaten (lacht) oder Daten generell, was ja alles sein kann. 00:06:25-  
85 0  
86
- 87 I: Sagen wir mal alles, ja. Alle möglichen Daten, die anfallen. 00:06:27-8  
88
- 89 B: Ja also wahrscheinlich während der, ja das ist schwierig zu quantifizieren, ich meine  
90 während der Auswertung, also sei es jetzt primär, Primärauswertung oder  
91 Sekundärauswertung produziert man natürlich eine ganze Menge an neuen Variablen, Zahlen,  
92 auch Daten. Also das kann natürlich so ein Umfang mehr, für mehr Datensatz übersteigen. Das  
93 ist natürlich da, der (...) auf Ihre vorige Frage können das natürlich dann auch Forschungsdaten  
94 sein, aber (lacht) die dann natürlich wiederum auch ausgewertet werden können, nicht. (...)   
95 Also im Forschungsprozess selber verfolgen wir oder mit denen dann auch gearbeitet wird  
96 und. Also Sie wollen jetzt wahrscheinlich keine Ja-Nein- Antwort haben, also /00:07:35-2  
97
- 98 I: Sie können einfach antworten, wie Sie meinen. (lacht) 00:07:38-5  
99
- 100 B: Ja ich würde trotzdem dabei bleiben, eigentlich bei der Erhebung, so im engeren Sinne.  
101 00:07:47-6

102  
103 I: Okay. 00:07:47-9  
104  
105 B: So Daten ja. 00:07:49-7  
106  
107 I: Okay. Das ist eigentlich auch eine ganz nette Überleitung, Sie sagten gerade, dass bei der  
108 Auswertung ja auch jede Menge Daten entstehen, zum Beispiel neue Variablen. Aber die  
109 nächste Frage ist jetzt, wie werden die Daten dann dokumentiert oder zum Beispiel auch mit  
110 Metadaten versehen. Oder Beschreibungen werden verfasst. Also wie sieht es da aus?  
111 00:08:10-0  
112  
113 B: Na gut wenn, der Primärdatensatz sollte durch den vom Befragungsinstitut oder das Institut,  
114 was die Codierung vornimmt, beschrieben werden. Bei den quasi abgeleiteten Variablen hängt  
115 es dann noch mal davon ab, was also in welchen Arbeitsschritt das ist irgendwie und wie es.  
116 Also bei uns war es jetzt auch der Fall, wir haben halt den codierten Datensatz bekommen von  
117 dem Befragungsinstitut und haben natürlich an dem Datensatz dann noch irgendwie  
118 weitergearbeitet und ihn auch noch mal in einem anderen Schritt, einen Datensatz zur  
119 Veröffentlichung gebastelt. Und Dokumentation da oder sagen wir Metadaten passiert halt  
120 dann für die Registrierung. Währenddessen hat man verschiedene Datensätze, die dann mehr  
121 oder weniger gut, entweder durch die (...) also zumindest wie wir gearbeitet haben, haben wir  
122 die Datensätze selber kaum mit Metadaten oder Datensatzbeschreibung versehen, weil es halt  
123 Work in Progress einfach war. 00:09:35-8  
124  
125 I: Und was genau meinen Sie damit? Also ist / kann man ad hoc immer nach Bedarf oder wie,  
126 was muss ich darunter verstehen? 00:09:46-3  
127  
128 B: Die / was Work in Progress war? 00:09:48-0  
129  
130 I: Ja. Also wie haben die Metadaten nicht, also Sie haben Work in Progress also ab und zu,  
131 nicht in regelmäßigen Abständen oder? 00:09:58-5  
132  
133 B: Ja eigentlich also/ (...) Also für die Weitergabe ja, aber sonst, wenn wir damit gearbeitet  
134 haben, aber (...) (unv.) 00:10:31-4 (...) ja also eigentlich kaum mit Datensatzbeschreibungen  
135 gearbeitet oder ist es irgendwie, ja. (lacht) 00:10:45-0  
136  
137 I: Okay. 00:10:45-1  
138  
139 B: Also es waren es hauptsächlich Arbeitsdatensätze, ja. 00:10:51-7  
140  
141 I: Ich hatte von Herrn Habich diesen Projektbericht, den es schon gibt mal bekommen und  
142 habe den auch durchgelesen und da stand, dass die Interviewprotokolle in einem Datensatz  
143 quasi ausgewertet wurden und der auch für weitere Analysen zur Verfügung steht. Wurden  
144 dazu irgendwie Dokumentationen angefertigt? Also diese Interviewprotokolle sind / 00:11:14-  
145 2  
146  
147 B: Wie die auch entstanden sind oder? 00:11:15-1  
148  
149 I: ... die auch eine Art von Beschreibungen? Oder also das sind ja im Prinzip Dokumentationen?  
150 00:11:21-5  
151  
152 B: Mhm. 00:11:23-7  
153

- 154 I: Wer war denn dafür zum Beispiel verantwortlich für diese Protokollerstellungen? 00:11:29-8  
155
- 156 B: Also für die Protokollerstellung und die Erstellung, also und auch die quasi Codierung dieser  
157 Protokolldaten war halt das Befragungsinstitut verantwortlich. Und diese Daten sind halt jetzt  
158 auch in den quasi Rohdatensatz oder den Daten aus dem wir von ihnen bekommen haben  
159 auch drin. Der wird dann bei der Weitergabe sind zumindest diese Protokolldaten, die Daten  
160 da drinnen veröffentlicht wird, sind die dann nicht mehr drin. Aber (...) jetzt (...) ist also da [...]\*
- 161 und ich ja nicht mehr am WZB sind, dann ist ja [...]\*, quasi hat so ein wenig mehr oder weniger  
162 die Verantwortung für das Projekt. Das heißt Anfragen für eine Auswertung von diesen Daten  
163 und es würden dann an [...] gehen. 00:12:33-1  
164
- 165 I: Mhm. 00:12:33-6  
166
- 167 B: Und ich meine wir haben, während der Erhebung natürlich mit den Protokolldaten auch  
168 gearbeitet einfach zum Monitoring und zur Überprüfung so die Interviews laufen. Ja, (...) ja, ja.  
169 00:13:01-5  
170
- 171 I: Okay. Gibt es da eine Person, die da speziell verantwortlich ist für, für die Aufbewahrung  
172 dieser Protokolle oder Beschreibungen? Ist das [die Person, die]\* Sie gerade genannt haben,  
173 [die]\* man auch ansprechen muss, wenn es um die Weitergabe geht? 00:13:13-7  
174
- 175 B: Also [...] ist generell, also weil das ja nicht getrennt ist, jetzt die quasi Interviewdaten und  
176 diese Protokolldaten, also nicht zwei verschiedene Positionen. Also für die Daten jetzt ist eine,  
177 also für das Projekt an sich oder für die Weiterführung von Veröffentlichungen und Anfragen,  
178 es geht theoretisch erst mal an [...]\*, ja. Genau. 00:13:40-7  
179
- 180 I: Okay. 00:13:39-5  
181
- 182 B: Und dann haben wir ja noch, also Sie meinen jetzt am WZB ist [...] hat sich ja viel mit den  
183 Daten beschäftigt. [...] Auszubildend[...] war oder halt mitgearbeitet hat. Aber sage ich mal  
184 für die Verantwortung ist, das geht erstmal an [...]\*, wenn da (...) Anfragen kämen oder so.  
185 00:14:00-9  
186
- 187 I: Mhm. 00:14:01-2  
188
- 189 B: Und ja (...) genau und eben ja mit dem Befragungsinstitut, wo die Daten ja auch liegen, ist es  
190 eben ausgemacht, dass die bis zum Ende [des Projekts]\*, das dann alles löschen, ja. Genau.  
191 00:14:19-2  
192
- 193 I: Mhm. Also alle Protokolldaten oder? 00:14:24-4  
194
- 195 B: Die Protokolldaten und die Adressdaten und ja. 00:14:29-5  
196
- 197 I: Okay. Die werden dann gelöscht, weil das Projekt dann zu Ende ist oder? 00:14:32-9  
198
- 199 B: Genau. 00:14:33-2  
200
- 201 I: Aha okay. 00:14:34-2  
202
- 203 B: Also das Projekt wurde jetzt im, ist im Prinzip [...] zumindest Erhebungen und  
204 Datenbindungen et cetera abgeschlossen und auch, also es gab ja verschiedene  
205 Projektbeteiligte in den verschiedenen Forschungsinteressen. Und also soweit ist das

- 206 abgeschlossen, wobei natürlich jetzt die Auswertungen und Publikationen und so noch laufen,  
207 aber genau. Also da ist die Befragung et cetera abgeschlossen ist und das ist dann (unv.).  
208 00:15:05-9  
209
- 210 I: Okay ja dann sind wir eigentlich auch schon mitten im Thema Nachnutzung der Daten, (lacht)  
211 da haben Sie jetzt ja schon einige Sachen genannt. Wissen Sie, unter welchen Bedingungen die  
212 Daten dann weitergegeben werden, also wer kann da Anfragen stellen, wissen Sie das oder?  
213 00:15:22-0  
214
- 215 B: Also soweit wie wir verblieben sind, wird das, werden die Daten registriert und sollen dann,  
216 ich glaube da war noch nicht ganz klar, wir hatten keine Zeit mehr, aber entweder hätte da die  
217 GESIS oder am WZB direkt verfügbar sein. Und die Idee war (...) da, also das ist schon  
218 zumindest in dieser Weitergabeform, da wir eben manche Sachen aufgrund von  
219 Anonymisierung da rausnehmen mussten, dass es frei zugänglich ist, aber die Idee war so eine  
220 Sperre von zwei oder drei Jahren oder so. 00:16:03-2  
221
- 222 I: Mhm. 00:16:05-0  
223
- 224 B: ...zu irgendeiner Art Auswertung, aber da ist halt die Erstauswertung halt am WZB  
225 vorgenommen werden, ja. 00:16:10-3  
226
- 227 I: Also das heißt, Entschuldigung ich wollte Sie nicht unterbrechen, dass die Publikation der  
228 Ergebnisse dann erfolgt und dann zwei / drei Jahre später die Daten erst weitergegeben werden?  
229 Oder auch wenn die Auswertung schon erfolgt ist, aber noch nicht die Ergebnisse publiziert  
230 sind? 00:16:24-3  
231
- 232 B: Also ich weiß, also einmal weiß ich nicht wie das jetzt weil wir eben das noch nicht dann  
233 registriert hatten, weil dann die Zeit gefehlt ist, das war eine rein zeitliche Begrenzung. Also  
234 sonst müsste man es ja auch irgendwie also koppeln, an welche Art Auswertung. Und dann  
235 wäre es natürlich okay, wer schreibt dann was und ab wie vielen Artikeln, die jetzt vom WZB  
236 publiziert werden, also deswegen war da glaub ich auch die Idee, man macht, man sagt einfach  
237 zwei Jahre oder so und geht dann davon aus, dass sage ich mal, bis dahin einige Ergebnisse  
238 publiziert wurden und dann danach, also es ist halt, also dann verpufft also der  
239 Neuigkeitseffekt sowieso ein bisschen, ja. 17:06-3  
240
- 241 I: Okay. 00:17:06-0  
242
- 243 B: ....ja aber wie gesagt, ich bin da jetzt auch gerade nicht im Bilde, wie ob das jetzt so gemacht  
244 wird, wie da der Stand ist? 00:17:15-8  
245
- 246 I: Mhm okay. 00:17:16-8  
247
- 248 B: Ja. 00:17:18-9  
249
- 250 I: Okay. Wissen Sie zufällig, wie das aussieht, wenn jetzt die Daten weitergegeben werden, ob  
251 das auch interdisziplinär möglich wäre oder also könnte da jeder, von jedem möglichen  
252 Fachgebiet die Daten nutzen oder? 00:17:32-7  
253
- 254 B: B: Ich denke schon, ja also zumindest von unserer Seite war da keine Beschränkung geplant.  
255 00:17:37-5  
256
- 257 I: Mhm. 00:17:38-8

258  
259 B: Und dann, ja wäre dann, vielleicht wer auch immer bei der GESIS dann anfragt oder halt den  
260 Datensatz findet, ist ja auch über die WZB Seite dann, also sollte der Datensatz ja auch online  
261 sein, oder zumindest die Beschreibung, wer das dann findet, das ist ja immer die Frage, (lacht)  
262 das ist im Prinzip für alle Fächer nutzbar ja, ja. 00:18:06-9  
263  
264 I: Okay. Wie sehen Sie das denn mit der langfristigen Archivierung, also die Daten werden ja  
265 normalerweise zehn Jahre dann an der Institution, an der sie erhoben wurden, aufbewahrt.  
266 Also haben Sie davon schon gehört oder? Okay (lacht) oder wissen Sie, wie das aussieht mit  
267 der langfristigen Archivierung? Also jetzt abgesehen von den Daten, die gelöscht werden nach  
268 Projektende. 00:18:29-7  
269  
270 B: Mhm. 00:18:30-8  
271  
272 I: Aber die dann weitergegeben werden sollen? Können? Okay. 00:18:34-8  
273  
274 B: So. (lacht) 00:18:35-9  
275  
276 I: Kein Problem. 00:18:37-0  
277  
278 B: (unv.) Mhm. 00:18:37-6  
279  
280 I: Wurden eigentlich für das Projekt bereits vorhandene Daten nachgenutzt, also Fremddaten,  
281 die Sie von irgendwoher bekommen haben? 00:18:46-4  
282  
283 B: Also indirekt, also quasi Vorläufer der Studie oder die Studie ist ja entstanden nach so, [...] \*  
284 waren und dann kam die Idee halt eine größere quantitative Studie zu machen. 00:19:13-4  
285  
286 I: Mhm. 00:19:14-0  
287  
288 B: Und also, ich meine diese Daten wurden glaube ich eher so, dann als (...) quasi explorativ für  
289 Fragebogenentwicklung et cetera benutzt, ja. 00:19:28-4  
290  
291 I: Mhm. 00:19:28-8  
292  
293 B: Ja. Aber sonst andere Daten benutzt haben? (...) Also was wir für die Auswertung genutzt  
294 haben, weil ja einige Fragen (...) aus allgemeinen Bevölkerungsumfragen genommen wurden,  
295 dann einfach zum Vergleich, also das war bei uns jetzt nicht sehr, ja tief gehend die anderen,  
296 bei den anderen Projektpartnern, die machen wirklich systematische Vergleiche auch. Also da  
297 wären dann, (...) ich weiß nicht genau was sie für Daten, es waren halt größere allgemeine  
298 Bevölkerungsumfragen, so SOEP oder und da ja. 00:20:14-6  
299  
300 I: Mhm. 00:20:14-8  
301  
302 B: European Social Survey, ja. 00:20:16-9  
303  
304 I: Okay. Wie sieht das aus mit der Daten, also in Bezug auf die Datenweitergabe, wissen Sie das  
305 auch von anderen Kollegen vielleicht aus dem Haus oder ehemaligen Kollegen? Wie die  
306 Bereitschaft da aussieht, also? 00:20:36-7  
307  
308 B: [...] \* also ich hatte das (...) (...) ja ich würde das so einschätzen, dass es IM Haus so relativ  
309 unkompliziert ist, von daher ist schon Bereitschaft ist klar. 00:20:56-1



310  
311 I: Also innerhalb des Hauses? 00:20:56-6  
312  
313 B: Ja. 00:20:57-4  
314  
315 I: Mhm. 00:20:58-7  
316  
317 B: Aber ich denke, das kommt auch immer ganz darauf an. Also bei uns war es jetzt ja ebenso  
318 weil eher relativ viele, eben Partner involviert waren, hier war es klar, dass es halt (...) also ein  
319 Datensatz ist der in dem Rahmen irgendwie frei nutzbar sein soll, vom WZB zumindest  
320 erstmal/00:21:23-4  
321  
322 I: Mhm. 00:21:23-8  
323  
324 B: Und (...) ja ich denke, das ist dann auch sehr unterschiedlich, also wenn es jetzt so einen  
325 Datensatz gäbe, wo darin jetzt für eine Forscherin oder ein Forscher sehr exklusive Daten sind.  
326 Dann kann ich mir vorstellen, dass genau ein bisschen mehr wie auf Widerstände trifft, nicht.  
327 Und / aber wie gesagt, in dem Fall jetzt war es eher da breit sage ich mal angelegt. 00:21:51-7  
328  
329 I: Und warum würde man da vielleicht auf Widerstände stoßen? 00:21:55-1  
330  
331 B: Ja einfach Forschung oder Wissenschaft funktioniert ja daher, dass man Neues hervorbringt  
332 ja und durch die (...) Neuigkeit der (...) Erkenntnisse und natürlich Daten, sich profilieren kann  
333 und veröffentlicht wird. Und dadurch ist es halt ein bisschen systematisch halt angelegt, dass  
334 Forscher und Forscherinnen teilweise nicht gerne ihre Daten, zumindest nicht zum aktuellen  
335 Zeitpunkt dann herausgeben einfach. Weil sie die Ersten sein wollen, die damit veröffentlichen  
336 möchten. 00:22:34-4  
337  
338 I: Mhm. 00:22:34-8  
339  
340 B: (unv.) 00:22:36-4 Es kann natürlich, sieht natürlich anders aus, wenn man dann zusammen  
341 eine Veröffentlichung plant oder so ja. 00:22:44-3  
342  
343 I: Mhm. 00:22:45-9  
344  
345 B: Ja. Beziehungsweise kommt es auch darauf an, waren wir auch ein bisschen versucht zu  
346 koordinieren eben, wenn man jetzt diese neuen Daten hat und es ist ja schon, also gerade eine  
347 [...] (\* (...) schon eben die letzte größere [...]) (\*, also schon wieder was Neues. Und dass wir da so  
348 ein bisschen versucht haben zu koordinieren, eben wer veröffentlicht, wer arbeitet zu was,  
349 nicht, von den Projektpartnern irgendwie das oder/ 00:23:24-6  
350  
351 I: Mhm. 00:23:24-9  
352  
353 B: ... wenn man sich das da eben ein bisschen aufteilen kann und ich meine, es kann in einem  
354 anderen Kontext auch möglich sein, man hat zwar neue Daten, aber kriegt eine Anfrage. Und  
355 wenn jemand forscht zu was ganz anderem oder würde ganz oft eine ganz andere Perspektive  
356 an, auswerten, dann sind die Widerstände wahrscheinlich auch geringer, nicht. 00:23:44-2  
357  
358 I: Mhm. 00:23:44-6  
359  
360 B: ...wie gesagt, die Veröffentlichung geht und (...) ja und der Erste zu sein, ja. (lacht) 00:23:51-  
361 7

362  
363 I: Mhm okay. 00:23:53-5  
364  
365 B: Ja. 00:23:53-9  
366  
367 I: Ich würde gerne noch wissen, wenn jetzt die Daten über das WZB oder GESIS verfügbar sind  
368 und weiter genutzt werden können, wo sehen Sie da die Verantwortlichkeit? Also, wer wäre  
369 dann dafür verantwortlich, zum Beispiel wenn es hier am WZB wäre? 00:24:11-6  
370  
371 B: (...). 00:24:14-6  
372  
373 I: Also wer wäre eigentlich beteiligt daran, dass die Daten auch längerfristig verfügbar wären?  
374 00:24:21-2  
375  
376 B: Also für die Datenaufbewahrung, et cetera haben wir immer mit Herrn Habich Rücksprache  
377 getroffen, meistens wie das so ist. (...) So die explizite Verantwortungsklä rung (lacht) wie  
378 immer, also für uns war es wichtig zu wissen, wer ist eben für dieses Projekt weiter  
379 verantwortlich, da ist eben [...] \* jetzt (...) und ich denke, dass [...] \* dann mit Rücksprachen,  
380 Rücksprache mit Herrn Habich da quasi dann den Hut auf hat. 00:24:51-3  
381  
382 I: Mhm. 0 00:24:51-3  
383  
384 B: Ja und so ein bisschen haben wir das dann auch an seine, also zumindest, was die  
385 Aufbewahrung angeht, (...) bei ihm gelassen und [...] \* eben, ja. 00:25:03-8  
386  
387 I: Okay, können Sie sich vorstellen was alles dazugehört, dass die Daten auch ganz lange, also  
388 langfristig archiviert werden? Also haben Sie, welche Vorstellung haben Sie davon? 00:25:17-8  
389  
390 B: Na ja zumindest erst mal, dass, also wir hatten, ja dieses tolles Metadatenschema und so,  
391 dass das also verfügbar ist und die Daten registriert sind und dann diesen, wie hieß es denn?  
392 Impermanent? 00:25:45-1  
393  
394 I: Identifier? 00:25:44-8  
395  
396 B: Diesen De. 00:25:46-0  
397  
398 I: DOI? 00:25:46-5  
399  
400 B: Diese DOI-Nummer einfach bekommen. Genau, permanent identifier. Und (...) ja ansonsten  
401 (...) wäre es natürlich gut da Ansprechpersonen zu haben, aber na ja oder zumindest die  
402 Institution, über die Institution, dass es wieder auffindbar ist und ja (...) ja, ja. 00:26:31-9  
403  
404 I: Okay. (lacht) Ja. Das war es dann, also ich, wir haben eigentlich alle Fragen besprochen, die  
405 ich mir so vorgestellt hatte. Und dann bedanke ich mich ganz herzlich. 00:26:42-7  
406  
407 B: Gern. 00:26:43-4  
408  
409 I: Ich schalte dann jetzt das Gerät auch aus. (lacht) 00:26:45-0  
410  
411 Ende des Interviews  
412  
413 Anmerkung der Autorin: [...] \* = anonymisiert

### 1 C3) Interview 3

- 2 I: Fangen wir mal an. 00:00:02-4  
3  
4 B1: Geht das alles so oder soll ich noch was unterstellen? 00:00:05-1  
5  
6 I: Ich glaube wenn ich das so, das sollte eigentlich funktionieren. 00:00:08-7  
7  
8 B1: Wir haben da zwei Mikrofone, das würde glaube ich / 00:00:11-5  
9  
10 I: Genau, genau, okay, ja. Also zunächst würde mich interessieren, welche Arten von Daten  
11 beziehungsweise quantitativen Daten speziell Sie als Forschungsdaten wahrnehmen? Also was  
12 sind für Sie Forschungsdaten? 00:00:26-6  
13  
14 B1: Willst du anfangen, soll ich anfangen? 00:00:30-8  
15  
16 B2: Ich kann kurz anfangen, na ja für uns, da wir hauptsächlich auch quantitativ arbeiten und  
17 damit natürlich sozusagen wenig auf der qualitativen Mikroebene, sind für uns hauptsächlich  
18 die großen Datensätze interessant. Die irgendwie zur Verfügung stehen, also so was wie das  
19 SOEP zum Beispiel oder auch, es gibt sozusagen in unseren Bereich eine spezielle, spezielle  
20 Erhebung der „Freiwilligen Survey“ heißt das im vier Jahresabstand, nicht? 00:00:58-5  
21  
22 B1: Mhm. 00:00:58-4  
23  
24 B2: Jetzt schon zum dritten Mal gelaufen ist und wo zum Thema [...] sehr viel bundesweit  
25 abgefragt wird und das natürlich große Datensätze, wo dann auch die Daten generell zur  
26 Verfügung stehen. Das würde ich vielleicht als einen Bereich sehen und ein zweiter Bereich ist  
27 natürlich Daten, die wir selber in Projekten nah erhoben haben. Was wir sozusagen natürlich  
28 dann mit dem viel größeren Einfluss da auch verbunden haben, was wir da erheben und was  
29 wir dann auch bekommen. Und wir haben auch gerade eine größere [...]befragung gemacht  
30 2011/2012. 00:01:27-3  
31  
32 I: Mhm. 00:01:27-2  
33  
34 B2: Und arbeiten da auch möglich da jetzt auch relativ viel mit den eigenen Daten. 00:01:32-9  
35  
36 I: Was sind das so für Daten? Also welche Arten in welcher Form? 00:01:36-8  
37  
38 B2: Na gut was wir da gemacht haben, ist ich weiß nicht, ob [...] Ihnen da jetzt schon mal auch  
39 so ein, können wir Ihnen auch noch mal geben so einen Fragebogen. Also wir haben im Prinzip  
40 eine [...]befragung gemacht [...] 00:01:54-6  
41  
42 I: Mhm. 00:01:54-5  
43  
44 B2: [...] Und da sind relativ viel Strukturdaten dabei. Wie die [...] aufgestellt sind, wie viel  
45 Mitarbeiter [...] die haben, hat sich da was verändert. Und aber auch so ein paar, Fragen, was  
46 Entwicklung angeht, retrospektiv aber auch in die Zukunft zum Beispiel, wie sich auch  
47 Finanzeinnahmen verändert haben. Ja das vielleicht im Groben. Und sozusagen natürlich,  
48 waren für uns jetzt hier, wir haben auch ein paar dabei, aber relativ wenig so

49 Einstellungsfragen, möglich weil wir eben auf dieser [...]\*-ebene angesetzt haben. Und dann im  
50 Zweifelsfall, wir immer nur sozusagen, ja einzelne Interview - Personen fragen konnten, ob sie  
51 irgendeine Einschätzung treffen können für die [...]\*, aber nicht ganz auf der persönlichen  
52 Ebene. 00:02:52-1  
53  
54 I: Okay. 00:02:51-9  
55  
56 B2: Genau. Na ja und das sind sozusagen in den meisten Fällen bei uns auf jeden Fall Daten, die  
57 so im Querschnitt, also sozusagen vorliegen, wenn wir so einzelne Projekte machen. Da haben  
58 wir jetzt also einfach nur einmal im Querschnitt eine Messung. Der Vorteil von anderen  
59 größeren Märkten, das ist natürlich häufiger, dass dann eben auch einfach ein Längsschnitt die  
60 Möglichkeiten gibt, was auszuwerten, wie SOEP oder auch im „Freiwilligen Survey“ ist jetzt die  
61 dritte Welle. 00:03:18-9  
62  
63 I: Mhm. 00:03:19-4  
64  
65 B2: Das hat natürlich nochmal ganz andere, hat natürlich noch mal einen ganz anderen Zugang  
66 und hat ganz andere Möglichkeiten. (...) Ja. 00:03:26-9  
67  
68 B1: Okay wenn ich ergänzen kann, im Prinzip kann man, wenn man systematisiert sagen, wir  
69 arbeiten mit Daten auf der Makro-, Meso- und Mikroebene, auf der Makroebene viel auch mit  
70 Daten der Amtlichen Statistik. Also die beispielsweise Gesamtbeschäftigung in Deutschland  
71 oder [...]\* aus der Einkommensteuerstatistik, die verwenden wir also Amtliche Statistik vor  
72 allen Dingen auf der Makro-Ebene. Auf der Meso-Ebene hat [...]\* das schon richtig  
73 beschrieben, mit der [...]\*-erhebung. Viele [...]\*-bezogene Daten für uns sind beispielsweise aber  
74 auch neben den selbst erhobenen Daten, Daten der Statistiken der, dieser [...]\* von  
75 Bedeutung, also der [...]\* die alle vier Jahre eine Statistik herausgeben oder auch was die  
76 eigentlich im Internet für Daten publizieren. Auf der Mikro-Ebene hat [...]\* auch schon gesagt,  
77 auf der Meso-Ebene sind eigentlich auch unsere selbst erhobenen Daten spielen da auch eine  
78 Rolle. Also das wäre auf der Meso-Ebene, sind die einzelnen [...]\* Mikroebene, „Freiwilligen  
79 Survey“ wurde schon genannt. Also wir selbst haben auch eine Einschaltung in eine  
80 Untersuchung gemacht in den letzten Jahren. Die GB - Untersuchung „Gute Arbeit“ und  
81 konnten uns dann mit einigen Fragen im Prinzip einkaufen. Das sind so die Daten auf der  
82 Mikro-Ebene, wie es darum gibt die Angaben einzelner Personen zu sammeln oder  
83 auszuwerten. Ja angefangen von sozio-demografischen Merkmalen bis Einschätzung,  
84 Zufriedenheiten. So würde ich das vielleicht grob klassifizieren, wenn man das klassifiziert,  
85 würde ich sagen Makro, Meso, Mikro und nachher natürlich die unterschiedlichsten  
86 Datenquellen. Amtliche Statistik wäre so eine, dass das runtergeht bis zur selbst erhobenen  
87 Dokumentenanalyse, Internetrecherchen. Also ein relativ breites Spektrum was wir abdecken,  
88 nur was für meine Begriffe damit etwas zusammen mit unserer Forschungsthematik  
89 zusammenhängt. [...]\* ist eben ein breites Feld, ist ja eine Kategorie, die mehr  
90 wissenschaftstheoretisch entstanden ist und ja dadurch tun wir natürlich sehr breit an die  
91 Sachen rangehen. Und da kommt glaube ich dieses breite Verständnis von Forschungsdaten  
92 auch zustande. 00:06:07-0  
93  
94 B2: Das glaube ich auch, also was ich ziemlich spannend fand, was ich vorher auch nicht  
95 kannte, das weiß ich nicht auf welcher Ebene, im Prinzip sind es auch Makro, Makrodaten  
96 oder. (lacht) Was wir gemacht haben auch im Rahmen des Projekts, des letzten ist  
97 Registerdaten uns mal anzuschauen. Also das ist das, was so glaube ich, klassischer Weise unter  
98 prozessproduzierten Daten oder so, die jetzt nicht sozusagen irgendwie in Verwaltungsakten  
99 oder eben durch Registrierung bei [...]\* entstehen. Und die sind mittlerweile auch online über  
100 ein Registerportal verfügbar und dadurch relativ einfach im Verhältnis zu noch, weiß ich nicht

101 vor also in den 90igern oder so. 00:06:41-2  
102  
103 B1: Ja. 00:06:42-0  
104  
105 B2: Wenn du schon mal dran warst, einsehbar. 00:06:43-9  
106  
107 B1: Karteikarten ja. 00:06:44-9  
108  
109 B2: Und da konnte man mittlerweile zum Beispiel jetzt auch regionale Verteilungen von [...] \*  
110 so teilweise sich anschauen und das ist natürlich schon auch relativ spannend. 00:07:01-3  
111  
112 B1: Also [...] \*register, [...] \*register / 00:07:04-6  
113  
114 B2: Ja, ja. 00:07:04-8  
115  
116 B1: ... na ja am ehesten [...] \*register, das sind so ein paar Sachen mit denen wir auch in der  
117 letzten Zeit gearbeitet haben. 00:07:10-8  
118  
119 I: Mhm. 00:07:11-2  
120  
121 B2: Also speziell jetzt für unseren Forschungsbereich ist es auf jeden Fall, weil ich einmal  
122 ziemlich gut mitarbeiten kann. 00:07:19-0  
123  
124 B1: Ja. 00:07:19-9  
125  
126 I. Mhm. Okay. Die nächste Frage, da geht es um den Datenlebenszyklus. Ist Ihnen das ein  
127 Begriff? Also kennen Sie den Begriff Datenlebenszyklus? 00:07:30-5  
128  
129 B1:Nö, habe ich noch nicht gehört. 00:07:31-6  
130  
131 B2: Ich habe ihn nicht gehört. (lacht) 00:07:32-6  
132  
133 I: Okay kann ich dazu erstmal was sagen. (lacht) Also Daten durchlaufen einen bestimmten  
134 Zyklus von der Entstehung, also Erhebung über die Auswertung, bis hin zur Archivierung und  
135 gegebenenfalls auch weitere Nachnutzung und/ 00:07:44-9  
136  
137 B1: Und dann Papierkorb. 00:07:45-3  
138  
139 I: Genau. (lacht)  
140  
141 B1: Sie haben, Sie haben erstens den Bedarf vergessen und dann, dass sie weggeschmissen  
142 werden. 00:07:52-0  
143  
144 I: (lacht) Genau also auch entweder Archivierung oder auch Löschung und ja da würde ich  
145 gerne wissen, wie aus Ihrer Sicht sich dieser Zyklus gestaltet? Also dieser Datenlebenszyklus?  
146 Wie der aus Ihrer Sicht aussieht? 00:08:08-3  
147  
148 B2: Mhm. 00:08:08-6  
149  
150 B1: Da müssen wir aber jetzt wahrscheinlich einengen, dass was wir erzählt haben mit Register  
151 und so weiter, spielt ja nichts dabei so die Rolle, Sie meinen sicherlich selbst erhobene Daten  
152 dann? 00:08:21-8

153  
154 I: Das kann alles sein, aber ja, selbst Erhobene. 00:08:24-6  
155  
156 B2: Also dafür könnten wir es am ehesten sagen, weil natürlich sozusagen, wenn es wir erst  
157 mal, einmal zugreifen auf irgendeinen Datensatz, der auch irgendwie hier frei verfügbar  
158 irgendwie - im Haus/00:08:32-5  
159  
160 I: Genau. 00:08:33-2  
161  
162 B2: ...wir müssen das natürlich haben, wie mit den/00:08:34-3  
163  
164 I: Dann haben sie den nachgenutzt ja. 00:08:35-5  
165  
166 B2:... Zyklus relativ wenig zu tun, wir benutzen ihn dann punktuell und dann, na ja. Ja was kann  
167 man zum Zyklus sagen? Also es gibt natürlich das sind solche Standardschritte, dass davon  
168 angefangen, dass man am Anfang relativ Arbeit damit hat, die Daten irgendwie zu erheben und  
169 die in Form zu bringen, dass man damit arbeiten kann, mit so gängiger Statistiksoftware? Also  
170 wir arbeiten meistens mit SPSS, wobei auch teilweise mit STATA. Und da muss es eigentlich  
171 alles entsprechend erstmal aufbereitend werden und dann teilweise auch irgendwie mit  
172 Syntax vorbearbeitet, bis man dann diese ganzen Auswertungen machen kann. Und ab dem  
173 Zeitpunkt würde ich sagen, also wäre sozusagen immer meine Vorstellung, man kann erstmal  
174 sozusagen irgendwann so einen Grunddatensatz der wirklich sozusagen mit (unv.) 00:09:21-3  
175 wo die Basics fertig sind, mit denen dann oft halt verschiedene Leute arbeiten können. Und  
176 das ist ja meistens das Problem bei Daten. Wenn die Daten dann ausgewertet und verarbeitet  
177 werden, dass man mit Syntax arbeitet, bestimmte Sachen immer wieder neu macht oder Leute  
178 machen verschiedene Sachen und wie koordiniert man das miteinander. Und das ist zumindest  
179 bei uns so das Vorgehen, das wir da so einen fertigen Datensatz im Prinzip dann irgendwann  
180 hatten. 00:09:45-7  
181  
182 I: Mhm. 00:09:46-3  
183  
184 B2: Der einen bestimmten Stand hatte natürlich nur und dann individuell die Leute versucht  
185 haben, das eben sozusagen mit Syntax-Dateien weiter zu bearbeiten. Und sich darum dann an  
186 den Daten, so lange daran gearbeitet wird an diesen Grunddaten erstmal sich natürlich nicht  
187 mehr so viel ändert. Und die Frage ist natürlich, was sozusagen jetzt weiter passiert, derzeit  
188 werten wir sie noch aus. (lacht) Was damit weiter passiert ist schwer zu sagen, das hängt  
189 natürlich auch von ziemlich viel Rahmenbedingen glaube ich ab. Also weiß ich nicht, wir hatten  
190 jetzt ein Projekt, das hat eine Laufzeit von [...] Jahren. Da kann man natürlich ein bisschen was  
191 auswerten. Aber die Frage ist natürlich was ist danach, also wenn sozusagen wir haben dann  
192 Projektstellen hier auf diese [...] Jahre laufen, dann sind die weg. Und eben so relativ häufig  
193 ist heutzutage, weil dann gehen die Leute wo anders hin, arbeiten im nächsten Projekt. Dann  
194 sind die Daten im Prinzip erst mal da, aber die Frage wäre natürlich, wird da weiter irgendwie  
195 mit gearbeitet, die Leute die da irgendwie die als erhoben haben oder daran mitgearbeitet  
196 haben, arbeiten die da weiter? Oder wird das irgendwie sozusagen weiter irgendwie zur  
197 Verfügung gestellt, da gibt es natürlich auch immer verschiedene, bei so kleineren Projekten  
198 glaube ich verschiedene Probleme, die damit verbunden sind, die einfach jetzt komplett  
199 öffentlich zu machen. 00:10:59-9  
200  
201 I: Mhm. Welche wären das zum Beispiel? 00:11:01-5  
202  
203 B2: Na gut, eine Grundfrage ist natürlich wahrscheinlich erst mal, wenn es sozusagen  
204 Forschung ist, die von einer bestimmten Seite finanziert wird, denke ich geht es erst mal

205 darum, dass dann auch bestimmte Publikationen, erstmal bestimmte Veröffentlichungen da  
206 sein müssen. Die dann auch von der Seite, die dort gearbeitet hat und in Abstimmung mit der,  
207 mit denen die es finanziert hat also wir hatten jetzt eine Finanzierung von [...] \*erstmal  
208 vorlegen müssen. Und man nicht sozusagen die Daten, sobald man sie erhoben hat ins Internet  
209 stellt, dass jeder damit arbeiten kann und unter Umständen dann die Sachen veröffentlicht  
210 hat, bevor man selber (lacht) irgendwas ausgewertet hat. Denn das ist gerade bei Daten, weil  
211 die wahrscheinlich bei diesen Verläufen von Daten heutzutage auch ein relativ wichtiger Punkt.  
212 Dass es natürlich irgendwie Leute die Daten erheben, auch damit natürlich den Anspruch  
213 verbinden, damit jetzt zum Beispiel was zu publizieren oder sei es, dass Leute, also ich fange  
214 jetzt gerade an irgendeine Doktorarbeit zu schreiben. Und das ist dabei, (...) finde ich  
215 zumindest so ein bisschen eigentlich nicht 100-prozentig glaube ich geklärt Bereich. Also wie  
216 sozusagen, die Ansprüche darauf dann sind, sozusagen würden. Aber das ist glaube ich eine  
217 zentrale Frage, weil da geht ja auch darum ob, wie der sozusagen weiter veröffentlicht, frei  
218 verfügbar gemacht wird oder nicht. Aber das ist ja in den seltensten Projekten, also ich kenne  
219 eigentlich keine, zumindest sozusagen von diesen kleineren oder medium-großen Studien.  
220 Wenn man dann irgendwie, einfach mal sagt, der möchte mal in die Daten reingucken. Man  
221 hat eine Studie bekommen, da ist da so was ausgewertet worden, das geht eigentlich  
222 prinzipiell ja nicht. Ich kann jetzt nicht zu irgendeinem Institut, irgendeiner anderen  
223 Forschungseinrichtung, kann man anfragen, aber ich würde vermuten, dass die Chancen relativ  
224 gering sind da einfach mal die Daten zu bekommen, nur wo was ausgewertet wurde. 00:12:43-  
225 4

226

227 I: Mhm. 00:12:43-8

228

229 B1: Ja wir sind eigentlich eine relativ kleine Einheit, die keine Mittel hat oder keine, nicht gut  
230 ausgestattet ist vom Haus aus, sondern nur über Drittmittel lebt. Und wir haben keine  
231 Perspektive. Im Prinzip läuft die Gruppe aus im [...] \* Und insofern stellen sich die Fragen,  
232 Datenlebenszyklus sicherlich anders als beispielsweise hier beim Bildungspanel oder ähnlichen  
233 Geschichten, die mit sehr vielen Stellen ausgestattet sind, wo die Forschung auf 10, 15 Jahre  
234 angelegt ist. Oder was weiß ich noch länger wie die Professoren oder die Autoren da sind und  
235 insofern gehen wir da vielleicht mit einer anderen Perspektive ran. Also wir fragen natürlich  
236 was bringt uns das, erstens und jetzt, da denke ich, da sind, was schon genannt wurde, die  
237 Sachen beschränkt, wir haben eigentlich gar keine Zeit um uns, um unsere Daten so  
238 systematisch zu kümmern. Also wir müssen produzieren, Ausstoß haben, ja und unter dem  
239 Gesichtspunkt, dass die Perspektive unklar ist, wer wird nach [...] \* überhaupt da sein, um  
240 mit den Daten zu arbeiten? Ja, ist die Motivation jetzt dort etwas systematisch anzulegen oder  
241 nicht sehr hoch? In Zukunft wird das, hat [...] \* auch schon gesagt, dass die Frage der  
242 Konkurrenz natürlich steht. 00:14:24-4

243

244 B2: Mhm. 00:14:24-6

245

246 B1: Also wir können also dazu sagen, dass unsere Daten so ziemlich einmalig sind und wir uns  
247 natürlich nicht dort gegen generell gegen eine Nutzung verschließen. Also beispielsweise  
248 haben wir eine [Person, die] \* gerne mit unseren Daten arbeitet oder so verschiedene Leute, da  
249 passiert schon ein bisschen was. Aber wir sehen es gegenwärtig nicht als sinnvoll an, wenn wir  
250 beispielsweise die Daten frei zur Verfügung stellen würden. Und ich bin mir allerdings damit im  
251 Klaren, dass solche Daten Verfallsdatum auch haben. Und nach ein paar Jahren damit keiner  
252 mehr arbeiten will oder arbeiten möchte. Das heißt, dass das übliche Problem der  
253 Datenfriedhöfe hier auch bei uns besteht. 00:15:12-9

254

255 B2: Mhm. 00:15:13-6

256

257 B1: Also das muss man einfach so klar sehen, aber wie gesagt also ich denke, das hängt zwar  
258 ein bisschen mit Ressourcen und Perspektiven zusammen, die diesen Datenlebenszyklus  
259 beeinflussen und wo wir denke ich auch, in gewisser Weise bremsen und. Ich kenne das aus  
260 anderen Projekten schon, wo das so ähnlich gelaufen ist, nicht? Ich meine, man selber sieht,  
261 dass die Daten einmalig sind, dass man da was Besonderes hat. Aber letztendlich kommt man  
262 aus unterschiedlichen Erwägungen, weil meistens sind es Finanzierungsfragen, nicht dazu alles  
263 auszuwerten und irgendwann ist die Sache nicht mehr aktuell und spielt keine Rolle mehr. Und  
264 wie gesagt landet nichts im Papierkorb, aber es wird abgelegt und ist nicht mehr aktuell, nicht  
265 und für solche Forschungsrichtungen, wo eben die klare Perspektive fehlt, ist das ja auch  
266 vielleicht das Ideale, warum sollen wir (unv.) 00:16:12 aufbauen und es geht nicht weiter  
267 damit? 00:16:15-0

268

269 I: Mhm. Und wenn Sie die Daten quasi vernichten oder löschen wie auch immer, wie sieht das  
270 denn aus, also wie gehen Sie da vor? 00:16:24-8

271

272 B1: Na wir löschen sie nicht, wir (lacht) die bleiben liegen in so einem Ding auf den Rechnern  
273 oder/00:16:29-7

274

275 B2: Das gibt es ja im Prinzip heutzutage eigentlich kaum noch, also sozusagen man müsste  
276 sämtliche Rechner auf dem das irgendwie hier liegt, ja und dann ist das ja hier über das  
277 Netzwerk im WZB auch irgendwo im zentralen Back-up. Also das wäre auch schon glaube ich  
278 ein relativer Aufwand, alles, was mit den Daten zu tun hat, (lacht) irgendwie löschen zu wollen,  
279 das/ 00:16:46-8

280

281 I: Mhm. 00:16:47-6

282

283 B1: Alles stellen sie eben nicht in die Datenbanken in die großen rein oder haben da nicht  
284 unbedingt die große Motivation dafür. 00:16:56-5

285

286 I: Okay. 00:16:56-7

287

288 B1: Denn wir bekommen ja auch nichts da dafür. 00:16:59-4

289

290 I: Okay. Sie sagten gerade, dass Sie, dass Ihnen Ressourcen fehlen, sich um die Daten so richtig  
291 zu kümmern, wenn ich das richtig verstanden habe, dokumentieren Sie denn die Daten die Sie  
292 erheben und auswerten? Oder die Frage ist eigentlich, also wie werden die Daten bei Ihnen  
293 dokumentiert, also nutzen Sie Klassifikationssysteme, Metadaten, Beschreibungen? 00:17:24-3

294

295 B1: Na ja erstmal muss man sagen, bevor du dann die praktischen Sachen sagst, wir sind  
296 natürlich eine übersichtliche Gruppe, wir sind jetzt zurzeit [...] Mitarbeiter, [...] Hilfskräfte, da  
297 läuft natürlich vieles auf Zuruf. Und nicht zu systematisch und das engt jetzt natürlich ein  
298 bisschen ein einfach oder. Viele Fragen, die man sonst stellen würde oder stellen müsste und  
299 wo es ja auch entsprechende Regelungen oder wollen wir mal so sagen Empfehlungen oder  
300 Erfahrungen gibt, (...) erübrigen sich da vielleicht, aber du kannst ja noch mal ein bisschen was  
301 praktisches sagen. 00:18:00-2

302

303 B2: So viel kann ich auch gar nicht zu sagen, weil sozusagen tatsächlich also, ich kann einfach  
304 nur sagen, also wir haben, keine glaube ich eigenständige, irgendwie  
305 Datenklassifikationssystematik. Wo wir jetzt quasi so verschiedene Datensätze oder  
306 verschiedene Sachen irgendwie wirklich, wirklich inhaltlich irgendwie beschrieben hätten. Gut  
307 bei einigen die wir nutzen, gibt es sozusagen so natürlich Beschreibungen und wenn wir sie  
308 von woanders nutzen und eigene Produzierte haben, wenn ich in so (unv.) 00:18:27-7 haben



309 wir jetzt einen größeren Datensatz. 00:18:30-2  
310  
311 I: Mhm. 00:18:29-5  
312  
313 B1: Na die zwei mit dem / 00:18:32-2  
314  
315 B2: Im Prinzip sozusagen diese/00:18:33-4  
316  
317 B1: Mit der guten Arbeit und da mit zwei/00:18:33-6  
318  
319 B2: ...diese gute Arbeit, das sind zwei Dateien genau. Und darum würden wir unsere  
320 Beschreibung, die wir in eigenen Publikationen dafür benutzen, aber na ja. 00:18:40-6  
321  
322 B1: ... und oft funktioniert es eben auf Zuruf. Ich weiß, die Kollegin die die gute Arbeit macht,  
323 wenn ich da was brauche, gehe ich da hin und erkundige mich und hole halt die [...] \*erhebung,  
324 [...] \*, das läuft eben noch überschaubar eigentlich. 00:18:57-7  
325  
326 I: Mhm. 00:18:58-1  
327  
328 B1: Aber wie gesagt, vieles hängt eben mit der Konstruktion zusammen, also dass wir dort,  
329 sicherlich könnte man das stärker systematisieren und dann ein bestimmtes Schema dort  
330 reinbringen. 00:19:11-9  
331  
332 I: Mhm. 00:19:12-2  
333  
334 B1: Aber, Interesse ist beschränkt von außen, das zu nutzen. Wir haben es eher mit einem Feld  
335 zu tun, die praktisch erst mal die Auswertung haben wollen und nicht die Daten direkt. Und  
336 zweitens ist so unser Kreis, der sich damit beschäftigt, klein, sodass vieles nicht in dieser  
337 verriegelten, strukturierten Formen unbedingt laufen muss. Erzielt auch mehr auf Aufwand das  
338 und wie gesagt, wir haben jetzt Laufzeiten von Projekten von [...] \* Monaten und (...) [...] \* und  
339 so weiter. Das (...) da wird, da ist ja kein Geld zur Verfügung eigentlich. Also wenn über alles  
340 Drittmittel funktioniert, wer soll das, da müsste man ja eigene Drittmittel einwerben also.  
341 Entweder würde bestimmt im Haus dann was zur Verfügung gestellt werden oder ansonsten  
342 können wir uns das einfach nicht leisten, glaube ich. (lacht) 00:20:10-3  
343  
344 I: Also Dokumentationen mit der Datenvergabe sind gar nicht vorgesehen dann und aus  
345 Zeitgründen und Geldgründen einfach gar nicht machbar? Okay. 00:20:17-7  
346  
347 B2: Also da, also mir schwebt so was teilweise sozusagen zumindest so eine ausführliche  
348 Dokumentation von so Daten ja auch immer vor. Aber es ist also wirklich mit, wenn man erst  
349 mal an bestimmten Stellen mit anfängt, auch ein ziemlicher Aufwand, selbst wenn man einen  
350 Datensatz hätte und selbst sozusagen nur mal auch so ein bisschen so eine, (...) na ja man  
351 hätte ja verschiedenste Punkte. Man hat sozusagen jetzt irgendwelche Fragen, in einer  
352 bestimmten Art gestellt, hat dann aber irgendwie rausgefunden, da muss man noch danach ein  
353 bisschen was bearbeiten, um sie auswerten zu können. Das Wissen, was sozusagen die Schritte  
354 sind, die man mit den ursprünglichen Daten gemacht hat. Die ja teilweise schon sinnvoll  
355 wären, wenn man hier jetzt vielleicht mal so ein bisschen verschriftlicht dokumentieren würde,  
356 damit eine andere Person hinterher noch mal später arbeitet, dann nicht von vorne anfängt.  
357 Oder sich da umfänglich auch, wenn es einarbeiten müsste. Aber es ist natürlich, müsste man  
358 eigentlich tatsächlich sozusagen, das ist dann ein Nebenjob. Den man sich damit schafft, um so  
359 was sozusagen in so einer Form zu dokumentieren, um es für andere leichter zu machen, sich  
360 da einzuarbeiten und da fehlt tatsächlich, also würde ich sagen einfach die Zeit/ 00:21:20-2

361  
362 B1: Die Zeit und die Perspektive, nicht? 00:21:22-6  
363  
364 B2: Und ja die Frage für wen macht man es dann, weil selber, selber weiß ich, wie ich es  
365 gemacht habe, wenn ich die Daten auswerte und ich weiß nicht, wer jetzt vielleicht mit den  
366 Daten noch mal arbeitet. Und dann in meiner Freizeit mich hinzusetzen und dann da eine  
367 Dokumentation zu machen, dass vielleicht jemand damit noch mal einfacher arbeiten kann, ist  
368 natürlich die Motivation relativ gering, würde ich sagen. Aber ich fände so was immer, wenn es  
369 so was gibt, finde ich gut also. Zum Beispiel gut ist auch eine andere Frage, ich habe halt früher  
370 [...] gearbeitet, da gibt es jetzt zum Beispiel sehr, doch recht ausführliche Dokumentationen.  
371 Wobei es da natürlich auch noch mal ganz andere, über den Längsschnitt und dann verändern  
372 sich irgendwelche Kategorien. Und weiß ich, alles was natürlich auch einfach viel dringender  
373 notwendig, weil irgendein Nutzer, der dort einfach mal so rein guckt, überhaupt nicht über die  
374 ganzen Jahre den Überblick mehr haben kann in kürzester Zeit. Sondern sich da ein Jahr  
375 einarbeiten müsste. 00:22:10-0  
376  
377 B1: Großer Nutzerkreis und so weiter, na das ist ja eine Infrastrukturerhebung, also es ja,  
378 gehört zu der Infrastruktur der Sozialwissenschaften dazu. 00:22:16-5  
379  
380 B2: Ja. 00:22:17-3  
381  
382 B1: Während wir ja da wirklich so im 17. Ring um den Saturn kreisen vielleicht. (lacht)  
383 00:22:23-5  
384  
385 B2: Na ja und vor allen bei dem, also als ich da gearbeitet habe, also es gibt natürlich, es gibt  
386 auch glaube ich zumindest eine, ich glaube sogar zwei/ 00:22:29-1  
387  
388 B1: Stellen. 00:22:30-5  
389  
390 B2: Stellen, wo Leute sich tatsächlich nur mit dem reinen Datenmanagement beschäftigen  
391 [...]\*. 00:22:34-3  
392  
393 I: Mhm. 00:22:35-1  
394  
395 B2: Die sind dafür einfach da, die haben da Vollzeitstellen für und werden dafür bezahlt. Das ist  
396 natürlich dann auch noch mal eine andere, ein anderes Level, aber gut das sind aber wirklich  
397 auch, kann man nicht ganz vergleichen mit dem, was wir selbst erheben. 00:22:48-1  
398  
399 B1: Ich glaube, dass auch Geldgeber nicht, nicht (lacht) unbedingt da erfreut sein würden,  
400 wenn man Datenmanagement jetzt noch mit rein nimmt und dann noch 20.000 Euro mit  
401 draufschlägt. (lacht) Das wäre glaube ich eher ein Grund zu sagen, na ja. (...) Also man erwartet  
402 denke ich, dass man das automatisch macht, aber es ist mit Aufwand verbunden und der wird,  
403 glaube ich nirgends genügend reflektiert und/00:23:14-9  
404  
405 B2. Mhm. 00:23:15-2  
406  
407 B1: ... und (...) da kriegt man auch nichts dafür also. Der Output für uns sind Artikel oder  
408 Studien und nicht ein gutes Datenmanagement zu machen oder eine gute  
409 Datendokumentation. Ich will Sie nicht entmutigen hier, aber/ 00:23:38-3  
410  
411 B2: (lacht) 00:23:37-3  
412

413 I: (lacht). 00:23:37-6  
414  
415 B2: Ja was im Prinzip natürlich auch, also man kann das ja durchaus auch kritisch sehen, ich  
416 finde das schon oft, es ist auch schon ein Problem. Also wenn du in der Wissenschaft natürlich  
417 so, auf so ein Output in so einer Art und Weise immer mehr geachtet wird. Also das ist sicher  
418 auch eine Frage, was sozusagen an bestimmten Enden dann vielleicht auch (...) hinten drüber  
419 fällt, was vielleicht auch aber sinnvoll wäre. 00:24:00-6  
420  
421 I: Mhm. 00:24:02-5  
422  
423 B2: Ja. 00:24:04-9  
424  
425 I: Okay. Dann habe ich noch eine letzte Frage zur Nachnutzung der Daten, Sie haben ja schon  
426 angedeutet, dass Sie eigentlich nicht wissen, wer das dann noch weiter nutzt, oder das auch  
427 nicht langfristig im Blick haben. Aber angenommen, es würde jetzt jemand Interesse an den  
428 Daten bekunden. Und das Projekt ist aber schon abgeschlossen ein Jahr oder zwei Jahre.  
429 Würde es da irgendjemanden geben an den er sich wenden könnte? Oder, also wie sähe da  
430 überhaupt die Verantwortung aus danach? 00:24:30-8  
431  
432 B2: Na ja [...] \* geht noch. 00:24:31-9  
433  
434 B1: Ja, ja. (lacht) Also solange noch jemand hier ist, geht das schon, aber/ 00:24:36-0  
435  
436 B2: Also hier war eine spezielle Situation, glaube ich jetzt ist/ 00:24:38-5  
437  
438 B1: ...also wir haben, wir haben das ja auch schon eine bisschen angesprochen, ich meine es ist  
439 natürlich für die kurzen Projektlaufzeiten. Und in der Regel macht man ja seine Projekte doch  
440 mit jüngeren Leuten, die dort ihre Qualifikation auch machen wollen, ihre Promotion, sodass  
441 immer dieses Spannungsfeld einfach vorhanden ist. Gebe ich etwas weg, was eigentlich der die  
442 Auswertung hier noch den Leuten vorbehalten bleibt, die hier das Ding durchgeführt haben.  
443 Also diese Konkurrenzsituation ist schon vorhanden denke ich mir. Aber das ist, wobei wir  
444 generell ja recht offen sind, glaube ich, wenn jemand kommen würde, und würde da seine  
445 spezielle Frage stellen und die Daten, da ist nicht unbedingt, dass wir das blockieren würden,  
446 ja? 00:25:31-7  
447  
448 B2: Hast du auch schon gesagt, also es gibt auf jeden Fall eine Person, die gerade [eine  
449 wissenschaftliche Arbeit] \* schreibt und die Datenarbeit, dann auch noch eine [weitere  
450 wissenschaftliche Arbeit] \*. 00:25:39-9  
451  
452 B1: Masterarbeiten mache ich ja genau, wir haben eine ehemalige Hilfskraft von uns, die in  
453 [...] \* jetzt. 00:25:45-3  
454  
455 I: Mhm. 00:25:46-0  
456  
457 B2: Auch damit mitarbeitet, also das ist so jetzt, aus dem Kreis natürlich auch welche.  
458 00:25:50-0  
459  
460 B1: Also wir, in der Hinsicht ist es vielleicht noch mal wichtig, dass wir versuchen auch durch  
461 solche externe Nutzung bestimmte Themen, die wir nicht bearbeiten oder zu denen wir auch  
462 nicht mehr kommen werden, durchaus dort, ja noch auswerten zu lassen so. 00:26:04-3  
463  
464 I: Mhm. 00:26:05-4

465  
466 B1: Also das wäre durchaus möglich und ich denke, das funktioniert bei uns auch  
467 unkompliziert. 00:26:12-1  
468  
469 I: Und angenommen es würde jemand kommen und an einer Fragestellung arbeiten wollen,  
470 die aber gar nichts mit Ihrem Thema zu tun hätte, aber trotzdem würde er sich für Ihre Daten  
471 interessieren? Wären Sie dann auch offen gegenüber einer Weitergabe oder? Hätten Sie dann  
472 Zweifel? 00:26:26-0  
473  
474 B1: Na ja so ganz frei würde ich das nicht weg, rausgeben glaube ich. (lacht) Wenn jemand  
475 sagt, hey uns interessiert das nur mal und, und (...). Ich habe da auch in mancher Hinsicht  
476 Bedenken, ich meine man kann ja auch viel Unsinn machen, mit so einem Datensatz. Man kann  
477 ja auch den hernehmen, um zu zeigen, na ja Ihr habt nicht sauber gearbeitet oder, oder. Oder  
478 in der und der Hinsicht (...) sind Defizite vorhanden, also. Also/  
479 00:26:58-2  
480  
481 I: Mhm. 00:26:57-3  
482  
483 B1: ...ich denke schon, dass wir uns diesen, solchen Sachen auch stellen würden, aber wenn  
484 generell jemand Unbekanntes kommt und nur sagt, ich interessiere mich mal für euren  
485 Datensatz. Und da steckt kein thematisches Interesse dahinter, da würde ich auch eher  
486 denken, dass da Unfug mit passieren kann. Weil man bestimmte Sachen nicht richtig  
487 einordnen kann, sicherlich auch auf Fehlen solcher Dokumentationen oder ausreichenden  
488 Dokumentationen, also insofern, wir haben ja was, wir haben ja einen Methodenbericht, auch  
489 von dem Umfrageinstitut zu dieser Untersuchung. Also so ist es nicht, dass wir gar nichts  
490 haben, ja. Aber ich denke schon, dass es ein gewisses fachliches Interesse und fachliche  
491 Kenntnisse auch bei solchen Datensätzen erforderlich ist. 00:27:42-1  
492  
493 I: Mhm. 00:27:42-3  
494  
495 B1: Denn so und so kann man, hätte auch, wir haben ja selber dann oft Skrupel, dass wir sagen,  
496 wenn wir zu tief in eine Unterersetzung in Gruppen reingehen, sind dort nur noch wenige Fälle  
497 vorhanden, also das werten wir nicht aus. Wenn ich das jemand aber so ohne weiteres  
498 weitergebe, dann habe ich ja aber überhaupt keine Gewähr, dass der da nicht irgendwas mit  
499 machen würde. Und wir dann aber als die Daten liefernde Organisation dafür stehen würden.  
500 Und das wäre dann sehr komisch dann, das was wir nicht uns getraut haben auszuwerten, weil  
501 nur fünf Einheiten da noch vertreten sind. Wenn das jemand dann, dann hochschießt und zu  
502 den Ergebnissen kommt, würde es auf uns ja zurückfallen letztendlich. So würde ich das  
503 jedenfalls sehen. 00:28:27-9  
504  
505 B2: Also ich denke das auch, ich weiß nicht (unv.) 00:28:26 weiß nicht wie berechtigt das  
506 wirklich ist diese Bedenken, aber ich hätte das auch so oft tatsächlich, was Du meintest, dass  
507 bestimmte Auswertungen die man macht, da trifft man bestimmte Entscheidungen. Soll es  
508 also in Kategorien zusammengefasst werden im einfachsten Fall, oder bei komplexeren  
509 Rechenmodellen, dass auch oft eine ziemliche so ein bisschen eine Spielerei ist. Und jetzt  
510 sozusagen einfach jemand, (...) also keine Ahnung ich kann es nicht einschätzen. Aber  
511 sozusagen es gibt auf jeden Fall bei vielen Sachen in der Statistik glaube ich die Möglichkeit,  
512 die Daten einfach mal sich so anzugucken und was anderes rauszukriegen. (...) Und dann, das  
513 ist natürlich auch sozusagen, damit hat man natürlich eine gewisse, eine gewisse  
514 Interpretation so halt darüber über die eigenen Daten. Das ist schon, glaube ich, kann ich mir  
515 auch vorstellen, durchaus eine weitverbreitete Einstellung (lacht) in dem Bereich. 00:29:18-5  
516

- 517 I: Mhm. 00:29:19-1  
518  
519 B2: Ja. 00:29:19-6  
520  
521 B1:Denn wir wissen im Prinzip, was bei bestimmten Sachen dahinter steht. Warum das so ist  
522 und das kann man glaube ich selbst mit einer perfekten Dokumentation nicht alles vermitteln.  
523 Ich glaube auch, dass da man unterscheiden muss, wie groß die Datensätze sind, also unser  
524 [...] \*datensatz sind [...] \*, die wir dort mit dem, die geantwortet haben auf den Fragebogen.  
525 Wenn ich andere Untersuchungen sehe, wie Freiwilligen Survey - 20.000 und so weiter. Da tut  
526 sich natürlich auch vieles wieder nivellieren im Prinzip. Und das habe ich natürlich mit solchen  
527 mittelgrößeren Datensätzen eher weniger. Und wenn ich dann, was man bei uns ja machen  
528 kann, runtergehe auf Länder und so halt auch einzelne Länder, dann wird es natürlich sehr  
529 kleinteilig. Und ja einfach, ich denke auch, dass man dort, dass da etwas rauskommen könnte,  
530 was nicht in unseren, unbedingt in unserem Interesse ist. (lacht) 00:30:12-0  
531  
532 B2: Mhm. 00:30:12-6  
533  
534 I: Wie denken Sie ist die Einstellung hier im Haus dazu, zur Datenweitergabe, Nachnutzung?  
535 Also denken viele so wie Sie oder? Unterhalten Sie sich darüber mit Kollegen? 00:30:24-1  
536  
537 B1: Wenig. 00:30:24-0  
538  
539 B2: Ich habe eigentlich kaum Erfahrung. Aber doch, also sozusagen ein, zwei wenige, wo ich  
540 schon sagen würde, dass die Leute auch, also mein Eindruck wäre schon, die Leute sitzen  
541 schon auch auf Ihren Daten. Weil es generell, also ich hatte wenig mit anderen Projekten zu  
542 tun direkt. 00:30:42-1  
543  
544 I: Mhm. 00:30:42-6  
545  
546 B2: Aber mein Eindruck wäre schon, solange die Leute auf ihren, an ihren Projekten sitzen und  
547 daran arbeiten, dann an Publikation arbeiten und da Leute direkt dran sind, ist es natürlich  
548 was, wo man erst mal auf seinen Sachen auch sitzt und guckt, was man damit macht. 00:30:57-  
549 6  
550  
551 B1: Ja und die Dissertationen die verzögern das natürlich, nicht. 00:31:00-8  
552  
553 I: Wie meinen Sie das? 00:31:01-8  
554  
555 B1: Na ja die Leute sind zwar in dem Projekt eingebunden, aber die Dissertationen, die werden  
556 ja in längeren Zeiträumen in der Regel realisiert, sodass das über dem Projektzeitraum  
557 hinausgeht. 00:31:13-6  
558  
559 I: Ah okay. 00:31:14-1  
560  
561 B1: Und ich meine dadurch gibt es natürlich ein gewisses Interesse oder eine Befürchtung, dass  
562 da was vorweggenommen werden könnte. Das kann man ja nicht einschätzen mit welcher  
563 Kapazität dann beispielsweise eine andere Institution die Sachen dann realisiert. Also diesen,  
564 dieser gewisse zeitliche Abstand, bevor man so was freigibt, ist glaube ich nicht zu  
565 vernachlässigen. Ich weiß es auch gar nicht, wie es jetzt Zentralarchiv Köln heißt es ja nicht  
566 mehr, nicht, wie heißt das denn jetzt? 00:31:46-4  
567  
568 I: Sie meinen das GESIS Datenarchiv? 00:31:48-2

569  
570 B1: GESIS Datenarchiv heißt es jetzt. Mit welcher Verzögerung die Daten dort eingehen. Aber  
571 da gibt es eben wie gesagt auch unterschiedliche Kategorien, wenn ich so den ALLBUS sehe  
572 und so weiter. Die ja teilweise als Infrastruktur gedacht sind, die, erwartet man, dass die  
573 schnell ihre Daten weitergeben, während - ich glaube von uns erwartet man das nicht  
574 unbedingt. 00:32:12-1  
575  
576 B2: Ich glaube, das muss man wirklich unterscheiden, also es gibt einfach so eine Erhebung, wo  
577 sehr viel öffentliche Gelder reingehen/ 00:32:16-9  
578  
579 B1: Ja. 00:32:17-2  
580  
581 B2: ..wo das von Anfang an als genaue Infrastruktur-Daten, wie SOEP, ALLBUS angelegt ist, die  
582 sozusagen dann auch für die breitere Forschungscommunity zur Verfügung stehen sollen. Und  
583 wo das wahrscheinlich gar nicht, wo da niemand sagen könnte, wir geben jetzt mal dieses Jahr  
584 den ALLBUS nicht raus. Denn das gibt es glaube ich, geht einfach nicht oder das SOEP, das  
585 würde nicht gehen. Das kann dort niemand, kann dort selbst im SOEP-Team könnte das  
586 niemand entscheiden. Mal eine Welle SOEP nicht rauszugeben, so darum und da sind einfach  
587 darum, ist es schon noch mal eine andere Ebene, als wenn so ein projektbezogen so  
588 Datensätze entstehen. 00:32:52-5  
589  
590 B1: Also vielleicht noch mal zusammenfassend, wir sind generell bereit zur Weitergabe, aber  
591 würden uns das genau angucken, was damit gemacht werden soll. Also das würde ich eher als  
592 unsere Verantwortung jetzt sehen. 00:33:03-5  
593  
594 I: Mhm. 00:33:03-8  
595  
596 B1: Dass da nicht irgendwas mit passiert, was nicht im Sinne (...) der Daten auch liegt.  
597 00:33:11-4  
598  
599 I: Und wie würden Sie sich da absichern? Sie würden zum einen, zunächst einmal wissen  
600 wollen, wer das ist und was sie mit den Daten vorhat? 00:33:16-4  
601  
602 B2: Genau. 00:33:16-8  
603  
604 B1: Mhm. 00:33:17-7  
605  
606 I: Und das würden Sie sich wahrscheinlich auch schriftlich geben lassen? (lacht) 00:33:19-2  
607  
608 B1: Und dann gibt es ja über Roland Habich glaube ich so ein Datennutzungsvertrag oder was.  
609 00:33:24-8  
610  
611 I: Mhm. 00:33:25-6  
612  
613 B1: Da würden wir schon gucken, dass der zustande kommt dann. 00:33:29-6  
614  
615 I: Mhm. 00:33:30-3  
616  
617 B1: Aber wie gesagt, so ein großer Run ist auf unsere Daten auch nicht, nicht? Man möchte  
618 eher unsere Ergebnisse haben, also. 00:33:39-7  
619  
620 B2: Ja das halt um welche Daten, nicht? 00:33:43-1

- 621  
622 B1: Ja. 00:33:43-1  
623  
624 I: Mhm. 00:33:44-3  
625  
626 B2: Aber das ist vielleicht bei so, so Themen, sehr themenfokussierten Projekterhebungen  
627 auch eher der Fall, könnte ich mir vorstellen, dass dann eher Interesse dann dran besteht, aus  
628 möglichen, aus Fachöffentlichkeit. Aber eben bei uns gerade auch aus dem Bereich selbst, wo  
629 wir geforscht haben, da was zu erfahren dann, was da rausgekommen ist, stärker ist. Das jetzt  
630 von irgendeiner anderen Einrichtung irgendwelche Anfragen kommen, jetzt mit den Daten zu  
631 arbeiten. Selbst wenn, wenn selbst wenn es Leute von Unis waren, wo eigentlich die meisten  
632 Anfragen gingen so in die Richtung, ob man dann die Power Point Präsentation kriegen kann  
633 oder dann (...) eben die Publikationen, ja. 00:34:19-5  
634  
635 B1: Oder so Qualifizierungsarbeiten, Masterarbeiten und Ähnliches, nicht? Mal so. Aber das  
636 hängt ein bisschen glaube ich mit unserem Forschungsfeld zusammen. Das ist eben eine Nische  
637 in Deutschland, da gibt es nicht so viele, die auf DEM Feld forschen und die empirisch forschen,  
638 die sind (lacht) sind noch geringer der Anteil. Da sind, viele sind [...] \*wissenschaftler, die da mit  
639 den Daten auch nicht unbedingt was am Hut haben und dadurch entsteht das eigentlich, nicht.  
640 Dadurch haben wir auch ein bisschen Alleinstellungsmerkmal in der Hinsicht, dass wir (...) diesen  
641 empirischen Teil eigentlich in diese Forschung reinbringen. Und das ist sicherlich  
642 anders, wenn wir Bildungsforschung oder/ 00:35:03-6  
643  
644 B2: Arbeitsmarkt. 00:35:03-7  
645  
646 B1: oder Sozialstrukturforschung oder was Ähnliches machen würden, also da gibt es ja dann  
647 doch Hunderte oder Tausende, die auf diesem Feld sitzen und, und insofern ja (...) wird  
648 vielleicht auch eine Lücke bleiben, wenn wir das Empirische nicht mehr machen. Aber so sieht  
649 es aus auch, also, ich denke schon, dass strukturell das nicht so durchdacht ist, als wenn solche  
650 Nischenbereiche. Wo wir gar nicht so, wo der Bereich, den wir beforschen nicht so klein ist.  
651 Aber von der Forschung sind wir eben ein Nischenbereich. Und in der Hinsicht stellen sich viele  
652 Fragen vielleicht etwas anders. 00:35:52-1  
653  
654 I: Okay. Ja also ich habe keine weiteren Fragen mehr. 00:35:55-8  
655  
656 B1: Okay, hoffentlich können Sie da was anfangen? (lacht) 00:35:58-8  
657  
658 B2: (lacht) 00:35:59-1  
659  
660 I: Ach doch, ich denke auf jedem Fall. (lacht) 00:36:00-9  
661  
662 B1: Ja da, also mir ist auf jedem Fall bewusst, was alles gemacht wird und was eigentlich auch  
663 so theoretisch gefordert wird, was man hinsichtlich Daten machen sollte und so weiter. Aber  
664 die Praxis sieht glaube ich einfach anders aus. 00:36:18-0  
665  
666 B2: Mhm. 00:36:17-6  
667  
668 B1: Na das ist eben der konkrete Fall entscheidend und man muss wahrscheinlich, wenn Sie  
669 auch Ihre Arbeit schreiben, unterschiedliche Ebenen einziehen, welche Art von Forschung ist  
670 es. 00:36:28-1  
671  
672 I: Mhm. 00:36:28-7

673

674 B1: Sind es große Forschungslinien, die eine Perspektive und Struktur und Potenzial haben,  
675 oder sind es solche kleinen Orchideen (...) Forschungsfelder? (lacht) 00:36:49-4

676

677 I: Okay. Ja. 00:36:50-6

678

679 B1: Ja dann sind wir gespannt, was Sie rauskriegen? 00:36:52-4

680

681 I: (lacht) Ja. 00:36:54-3

682

683 Ende des Interviews

684

685

686 Anmerkung der Autorin: [...] \* = anonymisiert



## Anhang D) Kategoriensystem

Kategorie	Textpassage (Interview, Zeile)	Zusammenfassung	Interpretation
<b>Leitfrage Forschungsdaten</b>			
K1: Keine oder unklare Definition	<p>„Gut die Frage ist halt, ob was erst zu Forschungsdaten wird, wenn ich die Rohdaten quasi in eine beforschbare (...) Form bringe. (...) Ehrlich gesagt, (lacht) ich kann mit der Frage nicht so viel anfangen?“(2, 13-15)</p> <p>„Oder aber prinzipiell kann ich keine Definition für Forschungsdaten geben.“(2, 41-42)</p>	<p>keine genaue Definition bekannt</p> <p>keine genaue Definition bekannt</p>	Eine eindeutige Definition kann nicht gegeben werden.
K2: Merkmale / Eigenschaften von Forschungsdaten	<p>„Also ich würde sagen natürlich alle Erhebungen, sozusagen Individualdaten und die können irgendwie schon, also bekannten SOEP, NEPS, die Daten die wir selber hier erheben mit Mikrozensus ist jetzt sozusagen nicht unbedingt eine freiwillige Erhebung, aber gehören zum, für mich auch zu Forschungsdaten. Dann eher Aggregatdaten [...] also ich empfinde sie auch als Forschungsdaten, aber damit habe ich nie gearbeitet, sind beispielsweise auch Dokumenten. Also ist es je nachdem, wie viele Dokumente man hat, ist auch eine Art von quantitativen, ja Forschungsdaten.“(1, 15-24)</p>	<p>Fremddaten (z.B. Quelle SOEP, NEPS), selbst erhobene Daten, bearbeitete Daten (Aggregatdaten), Analyse von Dokumenten</p>	Eine Unterscheidung von Forschungsdaten erfolgt primär in Fremddaten, die nachgenutzt werden und Daten, die selbst erhoben bzw. weiter verarbeitet werden.

<p>„Publikationen könnten natürlich auch Forschungsdaten sein, also Daten sind einfach Informationsquellen für mich. Und je nach Fragestellung, wenn ich beispielsweise eine Fragestellung habe, wo ich Publikationen analysiere oder Anzahl der Publikationen, Typ der Publikationen, also das wären dann auch Forschungsdaten.“ (1, 31-34)</p>	<p>Daten sind Informationsquellen, Analyse von Publikationen</p>
<p>„Dahin gehend also, wenn dann die Frage ist, eben, was sind Forschungsdaten? Würde ich sagen, das ist ja schon eben ein zweiter Schritt, wenn Daten aufbereitet sind oder zumindest aus einer bestimmten Perspektive gesehen werden. Hinsichtlich schon der Auswertung, also da sehe ich das dann als quantitative Daten, oder wenn man zum Beispiel Prozessdaten hat.“(2, 32-36)</p>	<p>Daten zum Zweck der Auswertung / Analyse, Fremddaten (prozessproduzierte Daten)</p>
<p>„[D]a wir hauptsächlich auch quantitativ arbeiten und damit natürlich sozusagen wenig auf der qualitativen Mikroebene, sind für uns hauptsächlich die großen Datensätze interessant. Die irgendwie zur Verfügung stehen, also so was wie das SOEP zum Beispiel...“ (3, 16-19)</p>	<p>Fremddaten (SOEP)</p>
<p>„[N]atürlich große Datensätze, wo dann auch die Daten generell zur Verfügung stehen. Das würde ich vielleicht als einen Bereich sehen und ein zweiter Bereich ist natürlich Daten, die wir selber in Projekten nah erhoben haben.“(3, 25-27)</p>	<p>Fremddaten, selbst erhobene Daten</p>
<p>„Und da sind relativ viel Strukturdaten dabei. Wie die [...] aufgestellt sind, wie viel Mitarbeiter, Engagierte die haben, hat sich da was verändert.“(3, 44-45)</p>	<p>Fremddaten (Strukturdaten)</p>

<p>„Was wir gemacht haben auch im Rahmen des Projekts, des letzten ist Registerdaten uns mal anzuschauen. Also das ist das, was so glaube ich, klassischer weise unter prozessproduzierten Daten oder so, die jetzt nicht sozusagen irgendwie in Verwaltungsakten oder eben durch Registrierung bei [...] entstehen. Und die sind mittlerweile auch online über ein Registerportal verfügbar...“ (3, 96-100)</p>	<p>Fremddaten (prozessproduzierte Daten)</p>
<p>„[!]m Prinzip kann man, wenn man systematisiert sagen, wir arbeiten mit Daten auf der Makro-, Meso- und Mikroebene, auf der Makroebene viel auch mit Daten der Amtlichen Statistik.“(3, 68-70)</p>	<p>Unterteilung nach Untersuchungseinheit (Makro-, Meso- und Mikro- Ebene)</p>
<p>„Das sind so die Daten auf der Mikro-Ebene, wie es darum gibt die Angaben einzelner Personen zu sammeln oder auszuwerten. Ja angefangen von sozio-demografischen Merkmalen bis Einschätzung, Zufriedenheiten. So würde ich das vielleicht grob klassifizieren, wenn man das klassifiziert, würde ich sagen Makro, Meso, Mikro und nachher natürlich die unterschiedlichsten Datenquellen.“(3, 81-86)</p>	<p>Unterteilung nach Untersuchungseinheit (Makro-, Meso- und Mikro- Ebene)</p>
<p>„Auf der Meso-Ebene hat [...] das schon richtig beschrieben, mit der [...]erhebung. Viele [...]bezogene Daten für uns sind beispielsweise aber auch neben den selbst erhobenen Daten, Daten der Statistiken der, dieser [...] von Bedeutung, also der [...] die alle vier Jahre eine Statistik herausgeben oder auch was die eigentlich im Internet für Daten publizieren.“(3, 72-76)</p>	<p>selbst erhobene Daten, Fremddaten</p>

	<p>„Und die offenen Fragen, es gibt natürlich auch offene Fragen und dann kann okay, was sind das jetzt für Daten? Weil sie halt noch nicht in der quantifizierten Form vorliegen, sondern also qualitativ ausgewertet werden können. Und das ist wahrscheinlich auch häufig der Fall. Dass die Daten ja per se nicht rein quantitativ oder qualitativ sind, sondern es ja auf die Auswertung (lacht) auch ankommt. Aber da ist natürlich Frage, okay, was wo (haben sie?) 00:02:13-1 die Erhebung, was für Instrumente benutze ich dann, dann ist es meist ja relativ klar.“(2, 26-32)</p> <p>„Also (...) die Frage ist, ob die Forschungsdaten (lacht) oder Daten generell, was ja alles sein kann.“(2, 83-84)</p>	<p>Unterscheidung von quantitativen und qualitativen Daten auch abhängig von Art der Auswertung</p> <p>Forschungsdaten als Untermenge von Daten</p>	
K3: Datenformate und Software	<p>„Also wir arbeiten meistens mit SPSS, wobei auch teilweise mit STATA. Und da muss es eigentlich alles entsprechend erstmal aufbereitend werden und dann teilweise auch irgendwie mit Syntax vorbearbeitet, bis man dann diese ganzen Auswertungen machen kann.“(3, 169-172)</p>	<p>Dateiformate zur statistischen Auswertung (mit Software SPSS, STATA)</p>	
<b>Leitfrage Datenlebenszyklus</b>			
K4: Begriff ist nicht oder eher nicht bekannt	<p>„Ich glaube ich habe die Frage nicht verstanden.“(1, 52)</p> <p>„Ich kann mir ein bisschen was darunter vorstellen, aber also, man muss halt dazu sagen, [...] * also?“(2, 47-48)</p> <p>„Nö, habe ich noch nicht gehört.“(3, 129)</p> <p>„Ich habe ihn nicht gehört.“(3, 131)</p>	<p>Begriffsvorstellung vorhanden</p> <p>Begriff unbekannt</p> <p>Begriff unbekannt</p>	<p>Der Begriff ‚Datenlebenszyklus‘ ist nicht geläufig.</p>

<p>K5: Phasen des Datenlebenszyklus</p>	<p>„Ja gut Datenlebenszyklus, jetzt weiß ich nicht, dazu gehört wahrscheinlich irgendwie eine erste Phase der Erhebung, also wenn Daten erhoben werden, also wenn es eine Primärerhebung gibt, dann die Primärerhebung, dann die Datenaufbereitung, Bereinigung und wie, was auch immer einen auswertbaren Zustand in den Datensatz zu bringen. Dann hängen an die Primärauswertung der, also meist haben dann die Forscher die die Daten erheben auch noch Auswertungsinteresse. [...] Dann Sekundärauswertung beziehungsweise noch die Daten, (...) ja gut Daten aufgehoben, registriert werden, zugänglich sind, dann eine Sekundärauswertung. Mehr weiß (lacht) ich, ob zu einem Zyklus auch die Löschung von Daten gehört theoretisch, was ja immer schade ist, ja aber wahrscheinlich auch vorkommt.“(2, 57-66)</p> <p>„Sie haben erstens den Bedarf vergessen und dann, dass sie weggeschmissen werden.“(3, 141-142)</p> <p>„Ja was kann man zum Zyklus sagen? Also es gibt natürlich das sind solche Standardschritte, dass davon angefangen, dass man am Anfang relativ Arbeit damit hat, die Daten irgendwie zu erheben und die in Form zu bringen, dass man damit arbeiten kann, mit so gängiger Statistiksoftware?“(3, 166-169)</p>	<p>Datenerhebung, Datenaufbereitung, Datenbereinigung, Datenauswertung (Primär- und Sekundärauswertung), Datenspeicherung, Datenregistrierung (für Zugänglichkeit), Sekundärauswertung, Löschung</p> <p>Bedarf (Konzeptualisierung), Löschung</p> <p>Datenerhebung, - aufbereitung, Analyse / Auswertung</p>	<p>Ein allgemeines Verständnis über die Phasen ist vorhanden, die langfristige Archivierung wird nicht explizit genannt.</p>
---	---	--	--

	<p>„Und ab dem Zeitpunkt würde ich sagen, also wäre sozusagen immer meine Vorstellung, man kann erstmal sozusagen irgendwann so einen Grunddatensatz der wirklich sozusagen mit (unv.) 00:09:21-3 wo die Basics fertig sind, mit denen dann oft halt verschiedene Leute arbeiten können. Und das ist ja meistens das Problem bei Daten. Wenn die Daten dann ausgewertet und verarbeitet werden, dass man mit Syntax arbeitet, bestimmte Sachen immer wieder neu macht oder Leute machen verschiedene Sachen und wie koordiniert man das miteinander. Und das ist zumindest bei uns so das Vorgehen, das wir da so einen fertigen Datensatz im Prinzip dann irgendwann hatten.“(3, 172-180)</p>	Grunddatensatz als Ergebnis der Aufbereitung und Basis für weitere Analysen	
K6: Unklarheit durch Projektbedingungen	<p>„Und die Frage ist natürlich, was sozusagen jetzt weiter passiert, derzeit werten wir sie noch aus. (lacht) Was damit weiter passiert ist schwer zu sagen, das hängt natürlich auch von ziemlich viel Rahmenbedingen glaube ich ab.“(3, 187-189)</p> <p>„Dann sind die Daten im Prinzip erst mal da, aber die Frage wäre natürlich, wird da weiter irgendwie mit gearbeitet, die Leute die da irgendwie die als erhoben haben oder daran mitgearbeitet haben, arbeiten die da weiter? Oder wird das irgendwie sozusagen weiter irgendwie zur Verfügung gestellt, da gibt es natürlich auch immer verschiedene, bei so kleineren Projekten glaube ich verschiedene Probleme, die damit verbunden sind, die einfach jetzt komplett öffentlich zu machen.“(3, 193-199)</p>	<p>Zukunft ungewiss</p> <p>Zukunft ungewiss, Weitergabe ungewiss</p>	Die Phasen des Datenlebenszyklus (vor allem Auswertung, Weitergabe) werden durch Rahmenbedingungen beeinflusst, wobei mangelnde Ressourcen den wichtigsten Aspekt darstellen.

	<p>„Ja wir sind eigentlich eine relativ kleine Einheit, die keine Mittel hat oder keine, nicht gut ausgestattet ist vom Haus aus, sondern nur über Drittmittel lebt. Und wir haben keine Perspektive. Im Prinzip läuft die Gruppe aus im [...]“* Und insofern stellen sich die Fragen, Datenlebenszyklus sicherlich anders als beispielsweise hier beim Bildungspanel oder ähnlichen Geschichten, die mit sehr vielen Stellen ausgestattet sind, wo die Forschung auf 10, 15 Jahre angelegt ist.“(3, 229-234)</p> <p>„Also das muss man einfach so klar sehen, aber wie gesagt also ich denke, das hängt zwar ein bisschen mit Ressourcen und Perspektiven zusammen, die diesen Datenlebenszyklus beeinflussen und wo wir denke ich auch, in gewisser Weise bremsen...“(3, 257-259)</p>	<p>fehlende Ressourcen, Zukunft ungewiss</p> <p>fehlende Ressourcen, Zukunft ungewiss</p>	
K7: Datenintensive Phase	<p>„Na ja [es]* werden die meisten Daten jetzt entstehen, also wir befinden uns jetzt gerade in der Feldphase.“(1, 63-74)</p> <p>„Die meisten Daten entstehen (...) ja eigentlich während der Erhebungsphase.“ (2, 83)</p>	<p>Datenerhebung</p> <p>Datenerhebung</p>	Bei der Datenerhebung entstehen die meisten Daten.
K8: Auswertung	<p>„Ja also wahrscheinlich während der, ja das ist schwierig zu quantifizieren, ich meine während der Auswertung, also sei es jetzt primär, Primärauswertung oder Sekundärauswertung produziert man natürlich eine ganze Menge an neuen Variablen, Zahlen, auch Daten. Also das kann natürlich so ein Umfang mehr, für mehr Datensatz übersteigen. Das ist natürlich da, der (...) auf Ihre vorige Frage können das natürlich dann auch Forschungsdaten sein, aber (lacht) die dann natürlich wiederum auch ausgewertet werden können.“(2, 89-94)</p>	<p>neue Daten entstehen, die wiederum Forschungsdaten sind</p>	Die Auswertung erzeugt neue Forschungsdaten.

K9: Anonymisierung	„...und eben dann Anonymisierung, also wie dann meinetwegen Adressdaten und so dann trennen na ja, ja.“(2, 76-78)	Anonymisierung für Weitergabe	Die Anonymisierung als Vorbereitung der Weitergabe ist eine eigenständige, zusätzliche Aufgabe.
<b>Leitfrage Dokumentation und Metadaten</b>			
K10: Involvierte Personen	<p>„Also [...] ist generell, also weil das ja nicht getrennt ist, jetzt die quasi Interviewdaten und diese Protokolldaten, also nicht zwei verschiedene Positionen. Also für die Daten jetzt ist eine, also für das Projekt an sich oder für die Weiterführung von Veröffentlichungen und Anfragen, es geht theoretisch erst mal an [...]*, ja.“(2, 175-178)</p> <p>„Also jetzt bei den Daten selber nicht. Das machen die studentischen Hilfskräfte.“(1, 209)</p> <p>„Bei den Protokollen machen das die Interviewer, aber ich schaue mir immer die Protokolle an, ohne einen Mitarbeiter, ich habe mir erst, auch die ersten Protokolle von den Daten angeguckt und jetzt wird das [...] machen.“(1, 214-216)</p> <p>„Na gut wenn, der Primärdatensatz sollte durch den vom Befragungsinstitut oder das Institut, was die Codierung vornimmt, beschrieben werden.“(2, 113-114)</p> <p>„Also für die Protokollerstellung und die Erstellung, also und auch die quasi Codierung dieser Protokolldaten war halt das Befragungsinstitut verantwortlich.“(2, 156-157)</p>	<p>Gemeinsame Verantwortlichkeit für Projekt und Metadaten</p> <p>Eingabe durch Stud. Hilfskräfte</p> <p>Protokolle von Interviewenden erstellt</p> <p>Kodierung und Beschreibung aus einer Hand</p> <p>Kodierung und Beschreibung aus einer Hand</p>	An der Erstellung von Dokumentationen und Vergabe von Metadaten sind unterschiedliche Personen aktiv beteiligt; wer verantwortlich für den Inhalt ist, ist auch für dessen Beschreibung zuständig; die gemeinsame Verantwortung liegt bei der Projektleitung.



K11: Gründe für Dokumentation / Metadaten	<p>„Und Dokumentation da oder sagen wir Metadaten passiert halt dann für die Registrierung. Währenddessen hat man verschiedene Datensätze, die dann mehr oder weniger gut, entweder durch die (...) also zumindest wie wir gearbeitet haben, haben wir die Datensätze selber kaum mit Metadaten oder Datensatzbeschreibung versehen, weil es halt Work in Progress einfach war.“(2, 119-123)</p>	Dokumentation / Metadaten für Weitergabe	Dokumentationen nützen zum einen der Arbeit innerhalb des eigenen Projekts und zum anderen den zukünftig Nutzenden.
	<p>„Also für die Weitergabe ja, aber sonst, wenn wir damit gearbeitet haben [...] ja also eigentlich kaum mit Datensatzbeschreibungen gearbeitet....“ (2, 133-135)</p>	Dokumentation / Metadaten für Weitergabe	
	<p>„Das müssen wir auch dokumentieren, weil wir danach natürlich auch entscheiden müssen, wie wir diese Daten edieren. Also und das ist und das läuft JETZT, also sobald wir die Daten angeben, alle Arten von Problemen, auch Probleme, die eben nicht unbedingt Probleme sind, aber wo wir sagen, okay, wenn wir das machen, machen wir das jetzt auch einfach ordentlich. Dokumentieren wir das alles, wenn es fertig ist, okay.“(1, 118-123)</p>	Mittel zur Überprüfung, Entscheidungsfindung	
	<p>„Also nicht nur bei der Datenweitergabe, sondern unter Umständen weiß ich auch nicht mehr in drei Jahren, wie haben wir damals entschieden? Also ich glaube, es ist auch ganz sinnvoll für einen selber.“(1, 163-165)</p>	Erinnerungsstütze	

<p>„Also ich merke das jetzt auch, was Sie vorhin gefragt hatten, nämlich bei diesen, also was eher bei [...]* angeht, deshalb mache ich das jetzt auch sofort. Weil irgendwie, die erste Woche wusste ich noch, [...]* „A“ war so und [...]* „B“ war so und [...]* „C“ war so. Schon ab [...]* wusste ich nicht mehr, welche war [...]*, bei [...]*/also, wenn man das nicht sofort dokumentiert, ist dann weg.“(1, 169-176)</p>	Erinnerungsstütze	
<p>„[A]ber es ist auch wichtig für einen selber, dass man weiß, okay wie hat man entschieden unter Umständen, das merkt man erst später. Das war nicht die gute Entscheidung, das muss man auch dokumentiert haben, beispielsweise in welchen Fällen hat man was entschieden?“(1, 190-193)</p>	Erinnerungsstütze	
<p>„Und dann, ja ist viel einfacher irgendwas zu korrigieren ohne es zu dokumentieren oder nicht? Aber eben danach ist es weg. Danach weiß man das nicht mehr, vor allen Dingen in unserem Fall, weil wir haben ja Papierfragebogen, wir haben nicht unterschiedliche Datensätze, wie in anderen, nicht. Da hat man einen unedierten Datensatz und dann hat man einen edierten Datensatz, aber unser Datensatz ist erstmal Papier, also wenn ich das direkt in den Daten ändere, dann muss ich das noch mal in Papier nachschauen.“ (1, 224-229)</p>	Erinnerungsstütze	
<p>„Und ich meine wir haben, während der Erhebung natürlich mit den Protokolldaten auch gearbeitet einfach zum Monitoring und zur Überprüfung so die Interviews laufen.“(2, 167-168)</p>	Mittel zur Überprüfung, Entscheidungsfindung	

	<p>„Zusätzlich auch denke, dass gar keine offenen Angaben sein werden, sondern nur codierte Angaben und dann nützt auch nichts, also das muss ich dann dokumentieren, welche Regeln ich ja getroffen habe.“(1, 146-148)</p> <p>„Deshalb glaube ich, ist einfach wichtig [...] für außerhalb, dass jemand fragen kann...“(1, 189-190)</p> <p>„Das Wissen, was sozusagen die Schritte sind, die man mit den ursprünglichen Daten gemacht hat. Die ja teilweise schon sinnvoll wären, wenn man hier jetzt vielleicht mal so ein bisschen verschriftlicht dokumentieren würde, damit eine andere Person hinterher noch mal später arbeitet, dann nicht von vorne anfängt. Oder sich da umfänglich auch, wenn es einarbeiten müsste.“ (3, 353-357)</p> <p>„Aber ich fände so was immer, wenn es so was gibt, finde ich gut also. Zum Beispiel gut ist auch eine andere Frage, ich habe halt früher [...] gearbeitet, da gibt es jetzt zum Beispiel sehr, doch recht ausführliche Dokumentationen.“(3, 368-370)</p>	<p>Hilfsmittel für zukünftig Nutzende</p> <p>Nachweisbarkeit ermöglichen</p> <p>Hilfsmittel für zukünftig Nutzende</p> <p>hilfreich bei Fremddaten-nutzung</p>	
K12: Was wird dokumentiert?	<p>„Weil wir zwar eine individuelle Befragung machen um, aber um die Leute zu erreichen mussten wir in [...] gehen und sprechen mussten wir auch offiziell über den [...] Datenschutz gehen. Und sie wollten schon im vorab eine Prozedurenbeschreibung haben, in anderen Projekten auch in meinem früheren Projekt, [...] hat man das eher im Nachhinein dokumentiert, wie man das gemacht hat (lacht). [...] aber nicht so formell und so ordentlich, wie wir das jetzt gemacht haben.“ (1, 79-85)</p>	<p>Prozedurenbeschreibung (Methodik, Informationen zur Weitergabe)</p>	<p>Es werden hauptsächlich die Methodik und Kodierregeln beschrieben.</p>

<p>„Und jetzt wo die Befragungen laufen haben wir zum einen Interviewer Protokolle, also zum einen haben wir Informationen über [...]*, zum Teil, ich weiß gar nicht mehr, bei welchem Amt man einsehen kann, beispielsweise welche [...]* sie sind und welche [...]* sie da haben und [so weiter]*. Und dann haben wir, da haben wir so eine Art Stammbblatt für jede [...]* und dann bei den Interviews selber, muss ein Interviewer und Interviewerinnen ein Protokoll führen. In dem dann Besonderheiten vermerkt werden, beispielsweise ob alles so gelaufen ist und alles so war, wie eben in diesem Stammbblatt beschrieben oder ob bestimmte Vorkommnisse waren?“(1, 89-95)</p>	Protokolle der Befragungen
<p>„Die Daten müssen natürlich eingelesen und eingegeben werden, ist, dass wir auch alles dokumentieren, also beispielsweise, wenn jemand eine Antwort korrigiert hat, weil das müssen wir sowieso per Hand nachträglich also korrigieren. Oder wenn beispielsweise eine Person nicht verstanden hat, dass nur eine Antwort möglich war und dann hat zwei Antworten angekreuzt.“(1, 109-113)</p>	Fragebögen
<p>„Zusätzlich auch denke, dass gar keine offenen Angaben sein werden, sondern nur codierte Angaben und dann nützt auch nichts, also das muss ich dann dokumentieren, welche Regeln ich ja getroffen habe.“(1, 146-148)</p>	Kodierregeln
<p>„Na gut wenn, der Primärdatensatz sollte durch den vom Befragungsinstitut oder das Institut, was die Codierung vornimmt, beschrieben werden.“(2, 113-114)</p>	erhobene Daten, Kodierungen

	"[W]ir haben ja was, wir haben ja einen Methodenbericht, auch von dem Umfrageinstitut zu dieser Untersuchung. Also so ist es nicht, dass wir gar nichts haben, ja. "(3, 488-490)	Methodik	
K13: Vorgehensweise	<p>„Ja. Also bei dieser Prozedurbeschreibung habe ich sehr von [...] profitiert. (räuspert sich) Da ist natürlich diese institutionelle Anbindung sehr von Vorteil. Wir haben, ich habe einfach gefragt und netterweise haben sie uns ein Beispiel zur Verfügung gestellt und daran haben wir uns orientiert. Natürlich stark angepasst, weil unser Design ist ein anderes und [so weiter]*.“(1, 730-733)</p> <p>„Aber es war jetzt weniger als Vorlage, sondern / doch es war eine Vorlage. Aber ich glaube, man muss es jetzt nicht so machen, nur es hat extrem meine Arbeit erleichtert, dass ich wissen konnte, wie sieht so was aus, was eben schon über einen [...] Datenschützer oder überhaupt Datenschutz in seiner Art durchgegangen ist. Und muss ich nicht das Rad neu erfinden.“(1, 743-747)</p> <p>„So für die Datenedition, ehrlich gesagt habe ich mich orientiert an meine letzte Studie und ich versuche irgendwie mich zu erinnern, an wen ich mich orientiert habe. Ich glaube, ich habe vermutlich so eine Vorlage von Daten, also von so einer Art Maske, die bei der [...] studie, benutzt worden ist. Natürlich angepasst dann an den jeweiligen Feldern.“(1, 755-759)</p> <p>„Es macht einfach nur einfacher, wenn man weiß, okay jetzt muss ich mir das nicht von selber überlegen. Dafür sagt man okay, [...] dann passe ich meine Vorlage, also diese Vorlage an meine Bedürfnisse, so machen wir das.“(1, 767-769)</p>	<p>Vorlagen vorteilhaft, an eigene Bedürfnisse angepasst</p> <p>Vorlagen vorteilhaft</p> <p>Vorlagen an eigene Bedürfnisse angepasst</p> <p>Vorlagen vorteilhaft, an eigene Bedürfnisse angepasst</p>	Für die Erstellung von Dokumentationen kommen keine Standards aber Vorlagen zum Einsatz.

	<p>„So viel kann ich auch gar nicht zu sagen, weil sozusagen tatsächlich also, ich kann einfach nur sagen, also wir haben, keine glaube ich eigenständige, irgendwie Datenklassifikationssystematik. Wo wir jetzt quasi so verschiedene Datensätze oder verschiedene Sachen irgendwie wirklich, wirklich inhaltlich irgendwie beschrieben hätten.“(3, 303-306)</p>	keine Nutzung von Standards / Klassifikationssystemen	
K14: Gründe für fehlende ausführliche Dokumentationen / Metadaten	<p>„Ist natürlich auch viel Arbeit diese ganze Dokumentation, nicht, also ist immer auch eine Abwägungssache.“(1, 218-219)</p> <p>„Aber es ist also wirklich mit, wenn man erst mal an bestimmten Stellen mit anfängt, auch ein ziemlicher Aufwand...“(3, 348-349)</p> <p>„Aber es ist natürlich, müsste man eigentlich tatsächlich sozusagen, das ist dann ein Nebenjob. Den man sich damit schafft...“(3, 357-358)</p> <p>„Na ja erstmal muss man sagen, bevor du dann die praktischen Sachen sagst, wir sind natürlich eine übersichtliche Gruppe, wir sind jetzt zurzeit [...] * Mitarbeiter, [...] * Hilfskräfte, da läuft natürlich vieles auf Zuruf. Und nicht zu systematisch und das engt jetzt natürlich ein bisschen ein einfach oder. Viele Fragen, die man sonst stellen würde oder stellen müsste und wo es ja auch entsprechende Regelungen oder wollen wir mal so sagen Empfehlungen oder Erfahrungen gibt, (...) erübrigen sich da vielleicht...“(3, 295-300)</p> <p>„Und ja die Frage für wen macht man es dann, weil selber, selber weiß ich, wie ich es gemacht habe, wenn ich die Daten auswerte und ich weiß nicht, wer jetzt vielleicht mit den Daten noch mal arbeitet.“(3, 364-366)</p>	<p>Dokumentation ist sehr aufwendig</p> <p>Dokumentation ist sehr aufwendig</p> <p>Dokumentation ist sehr aufwendig</p> <p>Keine Notwendigkeit für die alltägliche Arbeit</p> <p>Keine Notwendigkeit für die alltägliche Arbeit, Zielgruppe unbekannt</p>	<p>Das Fehlen ausführlicher Dokumentationen ist hauptsächlich ressourcenbedingt, da entsprechende personelle Kapazitäten (Geld) nicht in der Projektplanung enthalten sind und der hohe Aufwand den Nutzen kaum rechtfertigt.</p>

<p>„Interesse ist beschränkt von außen, das zu nutzen. Wir haben es eher mit einem Feld zu tun, die praktisch erst mal die Auswertung haben wollen und nicht die Daten direkt. Und zweitens ist so unser Kreis, der sich damit beschäftigt, klein, sodass vieles nicht in dieser verriegelten, strukturierten Formen unbedingt laufen muss.“ (3, 334-337)</p>	<p>Mangelndes Interesse von außen</p>
<p>„Denn wir wissen im Prinzip, was bei bestimmten Sachen dahinter steht. Warum das so ist und das kann man glaube ich selbst mit einer perfekten Dokumentation nicht alles vermitteln.“(3, 521-522)</p>	<p>Zweifel an Mehrwert</p>
<p>„[W]ir haben eigentlich gar keine Zeit um uns, um unsere Daten so systematisch zu kümmern.“(3, 237-238)</p>	<p>Keine Zeit</p>
<p>„Das (...) da wird, da ist ja kein Geld zur Verfügung eigentlich. Also wenn über alles Drittmittel funktioniert, wer soll das, da müsste man ja eigene Drittmittel einwerben also. Entweder würde bestimmt im Haus dann was zur Verfügung gestellt werden oder ansonsten können wir uns das einfach nicht leisten, glaube ich.“ (3, 339-342)</p>	<p>kein Geld</p>
<p>„[U]m so was sozusagen in so einer Form zu dokumentieren, um es für andere leichter zu machen, sich da einzuarbeiten und da fehlt tatsächlich, also würde ich sagen einfach die Zeit.“(3, 358-360)</p>	<p>keine Zeit</p>
<p>„Und dann in meiner Freizeit mich hinzusetzen und dann da eine Dokumentation zu machen, dass vielleicht jemand damit noch mal einfacher arbeiten kann, ist natürlich die Motivation relativ gering, würde ich sagen.“(3, 336-368)</p>	<p>keine Zeit</p>

	<p>„Die sind dafür einfach da, die haben da Vollzeitstellen für und werden dafür bezahlt. Das ist natürlich dann auch noch mal eine andere, ein anderes Level, aber gut das sind aber wirklich auch, kann man nicht ganz vergleichen mit dem, was wir selbst erheben.“(3, 395-397)</p> <p>„Ich glaube, dass auch Geldgeber nicht, nicht (lacht) unbedingt da erfreut sein würden, wenn man Datenmanagement jetzt noch mit rein nimmt und dann noch 20.000 Euro mit draufschlägt. (lacht) Das wäre glaube ich eher ein Grund zu sagen, na ja.“(3, 399-401)</p> <p>„Also man erwartet denke ich, dass man das automatisch macht, aber es ist mit Aufwand verbunden und der wird, glaube ich nirgends genügend reflektiert und [...] da kriegt man auch nichts dafür also.“(3, 401-407)</p> <p>„Der Output für uns sind Artikel oder Studien und nicht ein gutes Datenmanagement zu machen oder eine gute Datendokumentation.“(3, 407-409)</p>	<p>keine Zeit, kein Geld</p> <p>kein Geld</p> <p>hoher Aufwand, fehlende Anerkennung</p> <p>Dokumentation nicht Teil des wiss. Outputs</p>	
K15: Verbesserungsbedarf	<p>„[S]icherlich könnte man das stärker systematisieren und dann ein bestimmtes Schema dort reinbringen.“(3, 329-330)</p> <p>„Also da, also mir schwebt so was teilweise sozusagen zumindest so eine ausführliche Dokumentation von so Daten ja auch immer vor.“(3, 347-348)</p>	<p>Systematisches Vorgehen</p> <p>Verbesserungsbedarf bewusst</p>	Ein systematisches Dokumentieren ist wünschenswert.



	<p>„[A]lso man kann das ja durchaus auch kritisch sehen, ich finde das schon oft, es ist auch schon ein Problem. Also wenn du in der Wissenschaft natürlich so, auf so ein Output in so einer Art und Weise immer mehr geachtet wird. Also das ist sicher auch eine Frage, was sozusagen an bestimmten Enden dann vielleicht auch (...) hinten drüber fällt, was vielleicht auch aber sinnvoll wäre.“ (3, 415-419)</p>	Dokumentation sinnvoll	
K16: Langfristige Archivierung der Dokumentationen	<p>„Ach das ist eine gute Frage, so genau weiß ich das nicht. Ich glaube, auf jeden Fall bis (...) Verlauf der Studie, ich glaube nicht, dass es darüber hinaus sein wird.“(1, 129-130)</p> <p>„Und ja (...) genau und eben ja mit dem Befragungsinstitut, wo die Daten ja auch liegen, ist es eben ausgemacht, dass die bis zum Ende [des Projekts]*, das dann alles löschen, ja. [...] Die Protokolldaten und die Adressdaten und ja.“(2, 189-195)</p>	<p>Archivierung unklar</p> <p>Löschen aus Datenschutzgründen</p>	Dokumentationen sind nicht für die Langzeitarchivierung gedacht.
<b>Leitfrage Nachnutzung</b>			
K17: Bedingungen für Weitergabe	<p>„Und am Ende werden Daten weitergegeben in anonymisierter Form, was es genau heißt, also müssen wir uns eben im Laufes des Projektes Gedanken machen.“(1, 258-259)</p> <p>„Also wobei klar ist, keine personenbezogenen Daten, also Namen oder so was, also ich meine die fragen das gar nicht, aber es könnte natürlich sein, dass irgendjemand einen Namen drin schreibt oder irgendwie.“(1, 263-265)</p>	<p>Weitergabe nur in anonymisierter Form</p> <p>keine Weitergabe personenbezogener Daten</p>	Eine Weitergabe kann nur unter Einhaltung der Datenschutzbestimmungen erfolgen, d.h. die Daten müssen vorher anonymisiert werden. Eine Möglichkeit ist die (kontrollierte) Weitergabe nach Ablauf einer Embargo-Frist, auch im Rahmen eines Nutzungsvertrags.

<p>„[I]ch kann diese Daten gar nicht geben. Ich habe versprochen, dass ich diese Daten gar nicht gebe, weil die Leute haben mir persönlich es eben vorgegeben...“(1, 648-650)</p>	<p>keine Weitergabe personenbezogener Daten</p>
<p>„...da wir eben manche Sachen aufgrund von Anonymisierung da rausnehmen mussten...“(2, 218-219)</p>	<p>keine Weitergabe personenbezogener Daten</p>
<p>„Genau, dass man sagt, ja man gibt die Daten, aber man gibt sie erst nach einem Jahr oder zwei Jahren raus.“(1, 352-353)</p>	<p>Weitergabe mit zeitlicher Verzögerung</p>
<p>„[D]ie Idee war so eine Sperre von zwei oder drei Jahren oder so.“(2, 219-220)</p>	<p>Weitergabe mit zeitlicher Verzögerung</p>
<p>„[D]as war eine rein zeitliche Begrenzung. Also sonst müsste man es ja auch irgendwie also koppeln, an welche Art Auswertung.“(2, 233-234)</p>	<p>Weitergabe mit zeitlicher Verzögerung</p>
<p>„Ja und die Dissertationen die verzögern das natürlich [...] Na ja die Leute sind zwar in dem Projekt eingebunden, aber die Dissertationen, die werden ja in längeren Zeiträumen in der Regel realisiert, sodass das über dem Projektzeitraum hinausgeht.“(3, 551-557)</p>	<p>Weitergabe mit zeitlicher Verzögerung</p>
<p>„Und ich meine dadurch gibt es natürlich ein gewisses Interesse oder eine Befürchtung, dass da was vorweggenommen werden könnte. Das kann man ja nicht einschätzen mit welcher Kapazität dann beispielsweise eine andere Institution die Sachen dann realisiert. Also diesen, dieser gewisse zeitliche Abstand, bevor man so was freigibt, ist glaube ich nicht zu vernachlässigen.“(3, 561-565)</p>	<p>Weitergabe mit zeitlicher Verzögerung</p>

	<p>„Also so lange das Projekt läuft, werden sowieso keine Daten weitergegeben.“(1, 247-248)</p> <p>„Und dann gibt es ja über Roland Habich glaube ich so ein Datennutzungsvertrag oder was. [...] Da würden wir schon gucken, dass der zustande kommt dann.“(3, 608-613)</p> <p>„Aber das ist, wobei wir generell ja recht offen sind, glaube ich, wenn jemand kommen würde, und würde da seine spezielle Frage stellen und die Daten, da ist nicht unbedingt, dass wir das blockieren würden, ja?“(3, 443-446)</p> <p>„Also ich wäre sowieso nicht [die Person, die]* das entscheidet, sondern das wären die [...] * Projektleiter, die [...] *, die das Projekt auch angeworben haben...“(1, 358-359)</p>	<p>Weitergabe nicht vor Projektende</p> <p>Nutzungsvertrag</p> <p>Datenweitergabe eher befürwortet</p> <p>Entscheidung zur Weitergabe liegt bei Projektleitung</p>	
K18: Anforderungen an Anonymisierung	<p>„Aber jetzt natürlich könnte auch sein, dass eine Kombination von Merkmalen einfach macht die Leute zu rekonstruieren, wer die Leute sind [...]und darüber muss man sich erst mal Gedanken machen...“(1, 269-274)</p> <p>„Also dass man sich dort auch Gedanken macht, welche Kombination von Merkmalen man unter Umstände größer darstellt. Also für die Datenweitergabe darstellen muss.“(1, 276-278)</p> <p>„Aber auf jeden Fall war ein Gedanke, dass wir auch noch mal neue Nummer vergeben. [...]Neue Identifikationsnummer vergeben.“ (1, 284-294)</p>	<p>Daten dürfen keine Rückschlüsse auf Personenbezug zulassen</p> <p>Daten dürfen keine Rückschlüsse auf Personenbezug zulassen</p> <p>Identifikationsnummer erneuern</p>	<p>Mit der Anonymisierung soll verhindert werden, dass ein Personenbezug aus den Daten hergestellt werden kann. Anonymisierung ist sehr zeit- bzw. arbeitsaufwendig.</p>

	<p>„Weil das ist auch immer ein Problem. Anonymisierung ist ein extra Arbeitsschritt, und Projekte sind aber nie so (...) weder ausgestattet, noch auch von der Laufzeit, nicht, also ich meine, man hat irgendwie gerade irgendwie was erhoben, man muss eigentlich veröffentlichen. Muss das und das und das und da bleibt man Daten, das Projekt ist zu Ende, die Stelle ist zu Ende, so sind die Arbeitsbedingungen in der Wissenschaft, also in meiner Freizeit mache ich keine Datenanonymisierung.“(1, 310-315)</p>	Anonymisierung hoher Zeit-/Arbeitsaufwand	
K19: Gründe für Weitergabe unter bestimmten Bedingungen	<p>„Dann muss man erst mal aufbereiten und dann macht man die Analysen und dann schreibt man und dann macht man noch mal die Analysen. Und das reicht man irgendwo ein und dann muss man das noch mal überarbeiten.“(1, 395-397)</p> <p>„Und der Grund ist einfach, ich meine Daten erheben ist ein Haufen, Haufen, Haufen Arbeit.“ (1, 374-375)</p> <p>„Aber keiner von uns ist unbefristet und es wird einfach geguckt, was man veröffentlicht hat.“(1, 379-380)</p> <p>„Und eigentlich strategisch hat das, strategisch soll man immer sagen, nein ich erhebe keine Daten. Das ist meine zweite Datenerhebung, ich werde in meinen Lebenslauf ziemliche Löcher haben, weil ich eben Daten erhoben habe und am Endeffekt vollzeitbeschäftigt war mit Datenerhebung. Und in der Zwischenzeit nicht, also veröffentlicht habe...“(1, 384-388)</p> <p>„Da ist die Ausstattung allerdings auch ein bisschen eine andere. Also ich meine, [...]* [...] Ja ihre Arbeitsverträge sind zum Teil auch ein bisschen anders, also ja.“(1, 597-602)</p>	<p>Datenerhebung ist hoher Aufwand</p> <p>Datenerhebung ist hoher Aufwand</p> <p>Arbeitsbedingungen, nur Publikationen zählen</p> <p>Arbeitsbedingungen, nur Publikationen zählen</p> <p>Arbeitsbedingungen</p>	<p>Die Weitergabe erfolgt nur unter bestimmten Bedingungen, da Datenerhebung sehr aufwendig ist und unter den gegebenen Arbeitsbedingungen nur die Publikation der Ergebnisse als wiss. Output zählt. Daher haben die Forschenden den Anspruch, ihre Daten als erste auszuwerten um publizieren zu können. Die Datenweitergabe erfolgt hauptsächlich aus Gründen der Nachwuchsförderung und aus eigenem fachlichen Interesse. Einer Weitergabe an Externe oder Fachfremde wird eher skeptisch gegenübergestanden, aus Angst vor Datenmissbrauch, Fehlinterpretationen oder Rufschädigung.</p>

<p>„Ja wir sind eigentlich eine relativ kleine Einheit, die keine Mittel hat oder keine, nicht gut ausgestattet ist vom Haus aus, sondern nur über Drittmittel lebt. Und wir haben keine Perspektive.“ (3, 229-231)</p>	<p>Arbeitsbedingungen, finanzielle Rahmenbedingungen</p>
<p>„Ja ich, ich glaube das ist auch eher eine Sache, die man nach Studienende entscheidet, so dann auch mal entscheidet, wie viel man veröffentlicht hat oder nicht.“(1, 337-338)</p>	<p>Anrecht auf Erstauswertung/ veröffentlichung</p>
<p>„Also es hat weniger mit, ein Jahr nach Veröffentlichung zu tun, sondern dass man selber auch in der Lage ist zu veröffentlichen.“ (1, 366-367)</p>	<p>Anrecht auf Erstauswertung/ veröffentlichung</p>
<p>„[B]evor du oder sie die Daten benutzen, möchten wir gerne hier die benutzen.“(1, 562-563)</p>	<p>Anrecht auf Erstauswertung/ veröffentlichung</p>
<p>„Ich weiß selbst, wenn man eben Dokumente an der Dokumentendatenbank aufbaut, das ist einfach viel, viel Arbeit, und dass man erst mal sagt, ja gerne steht das irgendwann zur Verfügung, aber erstmal möchte ich sie benutzen.“(1, 583-585)</p>	<p>hoher Aufwand, Anrecht auf Erstauswertung/ veröffentlichung</p>
<p>„[N]ur ich denke ja, ich möchte auch gerne mal mit den Daten machen, aber leider hatte ich bis jetzt (lacht) keine Zeit. Oder meine studentischen Hilfskräfte oder meine Kolleginnen hatten gar nicht Zeit...“(1, 645-648)</p>	<p>Anrecht auf Erstauswertung/ veröffentlichung, keine Zeit</p>

„Und dann wäre es natürlich okay, wer schreibt dann was und ab wie vielen Artikeln, die jetzt vom WZB publiziert werden, also deswegen war da glaub ich auch die Idee, man macht, man sagt einfach zwei Jahre oder so und geht dann davon aus, dass sage ich mal, bis dahin einige Ergebnisse publiziert wurden und dann danach, also es ist halt, also dann verpufft also der Neuigkeitseffekt sowieso ein bisschen, ja.“(2, 234-239)

„In Zukunft wird das, hat [...] \* auch schon gesagt, dass die Frage der Konkurrenz natürlich steht.“(3, 241-242)

Anrecht auf Erstauswertung/-  
veröffentlichung

Konkurrenzdruck

„Und das fände ich ehrlich gesagt auch ein bisschen unfair, also ich sage das ehrlich jetzt auch über mein altes Projekt, aber auch in diesen, dann ist man irgendwie in der Laufzeit des Projekts ziemlich gut beschäftigt Daten zu erheben. [...] Und dann ist das Projekt zu Ende, am besten Fall im Laufzeit dieses Projekt hat man auch mal die Daten anonymisiert, sonst kann man sie auch gar nicht weitergeben. [...]Also sprich, man macht irgendwie nur eine reinste Dienstleistung, also es wäre eine reinste Dienstleistung und dann gibt man die Daten weiter. Und mache ich ein Beispiel und Sie kommen und benutzen Sie meine Daten und ich unter Umständen habe ich nicht mal eine, eine Veröffentlichung daraus, aber die ganze Arbeit reingesteckt. [...]Und unter Umständen habe ich nicht mal mehr eine Stelle, weil das Projekt abgelaufen ist dann.[...] Also von der Perspektive, wie gesagt ist jetzt weniger, wie ich veröffentlichen in 2012 und die Daten dürfen erst in 2014, sondern man sagt, okay das Projekt lief bis 2016 und man hat in diesen drei Jahren sehr, sehr viel Arbeit reingesteckt. [...]Dass man sagt, zumindest die Leute, die eben an der Datenerhebung und Konzeption und [so weiter]\*, sollen sie zumindest die Chancen haben. Ob sie eine Stelle haben, steht das auch noch in einem anderen Stern, aber zumindest die Chancen haben, irgendwas mit diesen Daten zu machen, bevor eben jemand anderer kommt, der eben mit diesen Daten noch gar nichts zu tun hatte. Und der eben in der Zeit, wo man in der Datenerhebung war, schön mit anderen Daten, die von jemand anderen erhoben wurden, schön publiziert hat. Weil ich meine sonst, wie gesagt kann man gleich packen und gehen und sagen okay, wo ist das nächste Arbeitsamt?“  
(1, 397- 433)

Arbeitsbedingungen,  
Zeitmangel, Datenerhebung  
ist hoher Aufwand, Anrecht  
auf Erstauswertung/  
veröffentlichung,  
Konkurrenzdruck

<p>„Ja einfach Forschung oder Wissenschaft funktioniert ja daher, dass man Neues hervorbringt ja und durch die (...) Neuigkeit der (...) Erkenntnisse und natürlich Daten, sich profilieren kann und veröffentlicht wird. Und dadurch ist es halt ein bisschen systematisch halt angelegt, dass Forscher und Forscherinnen teilweise nicht gerne ihre Daten, zumindest nicht zum aktuellen Zeitpunkt dann herausgeben einfach. Weil sie die Ersten sein wollen, die damit veröffentlichen möchten.“(2, 331-336)</p>	<p>Nur Publikationen zählen als wiss. Output, Konkurrenzdruck</p>
<p>„Und man nicht sozusagen die Daten, sobald man sie erhoben hat ins Internet stellt, dass jeder damit arbeiten kann und unter Umständen dann die Sachen veröffentlicht hat, bevor man selber (lacht) irgendwas ausgewertet hat. Denn das ist gerade bei Daten, weil die wahrscheinlich bei diesen Verläufen von Daten heutzutage auch ein relativ wichtiger Punkt. Dass es natürlich irgendwie Leute die Daten erheben, auch damit natürlich den Anspruch verbinden, damit jetzt zum Beispiel was zu publizieren...“(3, 208-213)</p>	<p>Anrecht auf Erstauswertung/-veröffentlichung</p>
<p>„Gebe ich etwas weg, was eigentlich der die Auswertung hier noch den Leuten vorbehalten bleibt, die hier das Ding durchgeführt haben. Also diese Konkurrenzsituation ist schon vorhanden denke ich mir.“(3, 441-443)</p>	<p>Anrecht auf Erstauswertung/-veröffentlichung, Konkurrenzdruck</p>
<p>„Na gut, eine Grundfrage ist natürlich wahrscheinlich erst mal, wenn es sozusagen Forschung ist, die von einer bestimmten Seite finanziert wird, denke ich geht es erst mal darum, dass dann auch bestimmte Publikationen, erstmal bestimmte Veröffentlichungen da sein müssen. Die dann auch von der Seite, die dort gearbeitet hat und in Abstimmung mit der, mit denen die es finanziert hat also wir hatten jetzt eine Finanzierung von [...] *erstmal vorlegen müssen.“(3, 203-208)</p>	<p>Verpflichtung zur Erstveröffentlichung von Drittmittelgeber</p>



„Also mein altes Projekt war tatsächlich so und das war, also Anfragen kamen, da habe ich auch nicht Nein gesagt, das kam immer ein bisschen darauf an, davon ab, wer gefragt hat, also beispielsweise die erste Veröffentlichung hat [eine Studentische Hilfskraft]\* gemacht, [...]\*, weiß ich mehr. Mit den Daten nicht nur und das war BEVOR ÜBERHAUPT ich oder meine anderen Kollegen irgendwas mit den Daten gemacht haben, aber das finde ich jetzt was anderes. Wäre jemand komplett Fremdes gewesen, hätte ich vermutlich Nein gesagt, aber da auch wiederum aus zwei Gründen zu der Zeit oder eigentlich bis jetzt noch, sind diese Daten noch nicht anonymisiert.“(1, 563-570)

„Weil eine mögliche Kooperation wäre jetzt was anderes und jetzt würde ich auch ein bisschen anders darüber nachdenken, aber wie gesagt, also das ist jetzt nicht per se, dass ich immer sage, nee bevor ich oder nicht.“(1, 579-581)

„Also wenn jetzt eine studentische Mitarbeiterin kommen würde, entweder von uns oder ich meine, es ist ein Kooperationsprojekt mit einem anderen Institut, das wäre das genau das Gleiche und sagen würde, er möchte gerne meinen Rahmen, also im Rahmen des Projekts, meine BA oder MA- Arbeit schreiben. Also sprich, er würde einfach so ein Thema machen, der irgendwie passt, hätte ich auch ÜBERHAUPT KEINE PROBLEME zu sagen, ja.“(1, 630-635)

„Es kann natürlich, sieht natürlich anders aus, wenn man dann zusammen eine Veröffentlichung plant oder so ja.“(2, 340-341)

Weitergabe an ausgewählte Personen,  
Nachwuchsförderung

Weitergabe aus eigenem fachlichen Interesse

Weitergabe aus eigenem fachlichen Interesse

Weitergabe aus eigenem fachlichen Interesse

<p>„Also wir, in der Hinsicht ist es vielleicht noch mal wichtig, dass wir versuchen auch durch solche externe Nutzung bestimmte Themen, die wir nicht bearbeiten oder zu denen wir auch nicht mehr kommen werden, durchaus dort, ja noch auswerten zu lassen so.“(3, 460-462)</p>	<p>Weitergabe aus eigenem fachlichen Interesse</p>
<p>„Also wir können also dazu sagen, dass unsere Daten so ziemlich einmalig sind und wir uns natürlich nicht dort gegen generell gegen eine Nutzung verschließen. Also beispielsweise haben wir [eine Person, die]* gerne mit unseren Daten arbeitet oder so verschiedene Leute, da passiert schon ein bisschen was. Aber wir sehen es gegenwärtig nicht als sinnvoll an, wenn wir beispielsweise die Daten frei zur Verfügung stellen würden.“(3, 246-250)</p>	<p>Weitergabe an ausgewählte Personen, Nachwuchsförderung</p>
<p>„...also es gibt auf jeden Fall eine Person, die gerade [eine wissenschaftliche Arbeit]* schreibt...“(3, 448-449)</p>	<p>Weitergabe an ausgewählte Personen, Nachwuchsförderung</p>
<p>„Masterarbeiten mache ich ja genau, wir haben eine ehemalige Hilfskraft von uns, die in [...] jetzt.“(3, 452-453)</p>	<p>Weitergabe an ausgewählte Personen, Nachwuchsförderung</p>
<p>„Na ja so ganz frei würde ich das nicht weg, rausgeben glaube ich. (lacht) Wenn jemand sagt, hey uns interessiert das nur mal und, und (...). Ich habe da auch in mancher Hinsicht Bedenken, ich meine man kann ja auch viel Unsinn machen, mit so einem Datensatz. Man kann ja auch den hernehmen, um zu zeigen, na ja Ihr habt nicht sauber gearbeitet oder, oder. Oder in der und der Hinsicht (...) sind Defizite vorhanden, also.“(3, 474-478)</p>	<p>Angst vor Datenmissbrauch oder Rufschädigung</p>

„...ich denke schon, dass wir uns diesen, solchen Sachen auch stellen würden, aber wenn generell jemand Unbekanntes kommt und nur sagt, ich interessiere mich mal für euren Datensatz. Und da steckt kein thematisches Interesse dahinter, da würde ich auch eher denken, dass da Unfug mit passieren kann. Weil man bestimmte Sachen nicht richtig einordnen kann, sicherlich auch auf Fehlen solcher Dokumentationen oder ausreichenden Dokumentationen, also insofern, wir haben ja was, wir haben ja einen Methodenbericht, auch von dem Umfrageinstitut zu dieser Untersuchung. Also so ist es nicht, dass wir gar nichts haben, ja. Aber ich denke schon, dass es ein gewisses fachliches Interesse und fachliche Kenntnisse auch bei solchen Datensätzen erforderlich ist.“(3, 483-491)

Angst vor Datenmissbrauch oder Fehlinterpretationen aufgrund von fehlenden Informationen / Fachkenntnissen

„Wenn ich das jemand aber so ohne weiteres weitergebe, dann habe ich ja aber überhaupt keine Gewähr, dass der da nicht irgendwas mit machen würde. Und wir dann aber als die Daten liefernde Organisation dafür stehen würden. Und das wäre dann sehr komisch dann, das was wir nicht uns getraut haben auszuwerten, weil nur fünf Einheiten da noch vertreten sind. Wenn das jemand dann, dann hochschießt und zu den Ergebnissen kommt, würde es auf uns ja zurückfallen letztendlich. So würde ich das jedenfalls sehen.“(3, 497-503)

Angst vor Rufschädigung

„Und ja einfach, ich denke auch, dass man dort, dass da etwas rauskommen könnte, was nicht in unseren, unbedingt in unserem Interesse ist.“(3, 529-530)

Angst vor Datenmissbrauch

	<p>„Also vielleicht noch mal zusammenfassend, wir sind generell bereit zur Weitergabe, aber würden uns das genau angucken, was damit gemacht werden soll. Also das würde ich eher als unsere Verantwortung jetzt sehen. [...] Dass da nicht irgendwas mit passiert, was nicht im Sinne (...) der Daten auch liegt.“ (3, 590-596)</p> <p>„[Q]ualitative Forschung ist schon ein bisschen problematisch. Weil man quantitativ eben (...) viel, viel mehr dokumentieren muss und eben zugänglich auch machen muss und transparent machen muss. Und qualitativ wie gesagt, ich habe Verständnis, dass es ein bisschen schwieriger ist, aber zum Teil dann ist es auch ein bisschen, (...) wie sagt man, eben nicht immer ganz transparent und nachvollziehbar, was man macht oder nicht macht.“ (1, 544-549)</p>	<p>Angst vor Datenmissbrauch</p> <p>Transparenz und Nachvollziehbarkeit durch Weitergabe</p>	
K20: Zweifel an Interesse von außen	<p>„Also ob sie jemand benutzt in zehn Jahren mag ich bezweifeln...“ (1, 460)</p> <p>„Also wie gesagt, jetzt hat jetzt weniger mit dem Fachgebiet zu tun, also ich meine, in einem ganz anderen Fachgebiet, da muss ich mich fragen, was willst du mit diesen Daten (lacht) machen.“ (1, 627-629)</p> <p>„Und ich bin mir allerdings damit im Klaren, dass solche Daten Verfallsdatum auch haben. Und nach ein paar Jahren damit keiner mehr arbeiten will oder arbeiten möchte. Das heißt, dass das übliche Problem der Datenfriedhöfe hier auch bei uns besteht.“ (3, 250-253)</p>	<p>Zweifel an zukünftigem Interesse</p> <p>Zweifel an Interesse fachfremder Personen</p> <p>Daten verlieren an Aktualität, mangelndes Interesse</p>	<p>Es bestehen Zweifel an einer zukünftigen Nutzung der Daten, da diese an Aktualität verlieren und das Interesse sich eher auf die Ergebnisse (z.B. Publikationen) beschränkt.</p>

	<p>„Aber letztendlich kommt man aus unterschiedlichen Erwägungen, weil meistens sind es Finanzierungsfragen, nicht dazu alles auszuwerten und irgendwann ist die Sache nicht mehr aktuell und spielt keine Rolle mehr. Und wie gesagt landet nichts im Papierkorb, aber es wird abgelegt und ist nicht mehr aktuell...“(3, 261-264)</p> <p>„Alles stellen sie eben nicht in die Datenbanken in die großen rein oder haben da nicht unbedingt die große Motivation dafür.“ (3, 283-284)</p> <p>„Wir haben es eher mit einem Feld zu tun, die praktisch erst mal die Auswertung haben wollen und nicht die Daten direkt.“(3, 334-335)</p> <p>„Aber da gibt es eben wie gesagt auch unterschiedliche Kategorien, wenn ich so den ALLBUS sehe und so weiter. Die ja teilweise als Infrastruktur gedacht sind, die, erwartet man, dass die schnell ihre Daten weitergeben, während - ich glaube von uns erwartet man das nicht unbedingt.“(3, 570-574)</p> <p>„Aber wie gesagt, so ein großer Run ist auf unsere Daten auch nicht, nicht? Man möchte eher unsere Ergebnisse haben, also.“ (3, 617-618)</p>	<p>Daten verlieren an Aktualität, mangelndes Interesse</p> <p>mangelndes Interesse</p> <p>mangelndes Interesse</p> <p>mangelndes Interesse</p> <p>Ergebnisse wichtiger als Daten</p>	
K21: Bereitschaft zur Datenweitergabe im Umfeld der Institution	<p>„[A]nsonsten muss ich sagen, kenne ich nicht sehr viele, also direkt mittelbar Kollegen und Kolleginnen die Daten erheben und also natürlich kann man sich jetzt fragen, also welche Daten? Also ich kenne natürlich Kollegen und Kolleginnen, die qualitative Daten erheben. Und die geben definitiv ihre Daten nicht weiter.“ (1, 539-542)</p>	<p>Bereitschaft nicht bekannt, qualitative Daten werden nicht weitergegeben</p>	<p>Die Bereitschaft wird wenig diskutiert. Innerhalb der eigenen Institution wird die Bereitschaft zur Weitergabe höher gesehen, als bei Nicht-Angehörigen oder sogar Fachfremden, bei qualitativen Daten ist sie eher</p>

<p>„[E]s ist eine interessante Frage, werde ich mal nachfragen, dass ich nicht [die einzige Person]* bin, [die]* so denkt.“(1, 582-583)</p>	Bereitschaft nicht bekannt	<p>gering. Die Bereitschaft zur Weitergabe nach Veröffentlichung ist größer, weil der Konkurrenzdruck abnimmt. Unterschiedliches Forschungsinteresse und andere Auswertungsmethoden sprechen für die Weitergabe an Externe.</p>
<p>„Und (...) ja ich denke, das ist dann auch sehr unterschiedlich, also wenn es jetzt so einen Datensatz gäbe, wo darin jetzt für eine Forscherin oder ein Forscher sehr exklusive Daten sind. Dann kann ich mir vorstellen, dass genau ein bisschen mehr wie auf Widerstände trifft...“(2, 324-326)</p>	Bereitschaft beschränkt	
<p>„[D]a gibt es natürlich auch immer verschiedene, bei so kleineren Projekten glaube ich verschiedene Probleme, die damit verbunden sind, die einfach jetzt komplett öffentlich zu machen.“(3, 197-199)</p>	Bereitschaft beschränkt	
<p>„Ich habe eigentlich kaum Erfahrung. Aber doch, also sozusagen ein, zwei wenige, wo ich schon sagen würde, dass die Leute auch, also mein Eindruck wäre schon, die Leute sitzen schon auch auf Ihren Daten. Weil es generell, also ich hatte wenig mit anderen Projekten zu tun direkt.[...] Aber mein Eindruck wäre schon, solange die Leute auf ihren, an ihren Projekten sitzen und daran arbeiten, dann an Publikation arbeiten und da Leute direkt dran sind, ist es natürlich was, wo man erst mal auf seinen Sachen auch sitzt und guckt, was man damit macht.“(3, 539-548)</p>	Bereitschaft beschränkt, Weitergabe mit zeitlicher Verzögerung, Anspruch auf Erstausswertung/-veröffentlichung	

	<p>„Und das ist dabei, (...) finde ich zumindest so ein bisschen eigentlich nicht 100-prozentig glaube ich geklärt Bereich. Also wie sozusagen, die Ansprüche darauf dann sind, sozusagen würden. Aber das ist glaube ich eine zentrale Frage, weil da geht ja auch darum ob, wie der sozusagen weiter veröffentlicht, frei verfügbar gemacht wird oder nicht. Aber das ist ja in den seltensten Projekten, also ich kenne eigentlich keine, zumindest sozusagen von diesen kleineren oder medium-großen Studien. Wenn man dann irgendwie, einfach mal sagt, der möchte mal in die Daten reingucken. Man hat eine Studie bekommen, da ist da so was ausgewertet worden, das geht eigentlich prinzipiell ja nicht. Ich kann jetzt nicht zu irgendeinem Institut, irgendeiner anderen Forschungseinrichtung, kann man anfragen, aber ich würde vermuten, dass die Chancen relativ gering sind da einfach mal die Daten zu bekommen, nur wo was ausgewertet wurde.“(3, 214-224)</p> <p>„[...]“ also ich hatte das (...) (...) ja ich würde das so einschätzen, dass es IM Haus so relativ unkompliziert ist, von daher ist schon Bereitschaft ist klar.“(2, 308-309)</p> <p>„[J]a also zumindest von unserer Seite war da keine Beschränkung geplant.“(2, 254)</p> <p>„Und wenn jemand forscht zu was ganz anderem oder würde ganz oft eine ganz andere Perspektive an, auswerten, dann sind die Widerstände wahrscheinlich auch geringer...“(2, 354-356)</p>	<p>Weitergabe nach Veröffentlichung</p> <p>Bereitschaft für Weitergabe innerhalb des Hauses vorhanden</p> <p>keine Beschränkung</p> <p>Konkurrenzdruck geringer bei unterschiedlichem Forschungsinteresse</p>	
K22: Verfügbarkeit, Auffindbarkeit	<p>„Na Pflege in dem Fall gibt es ja nicht mehr, aber nee wobei die immer auf dem technischen Stand sind. Meine Vermutung wäre, dass - dadurch, dass es ein Kooperationsprojekt ist, müsste man entscheiden, ob wir oder das Partnerinstitut und ob beide das</p>	<p>Migrationsbedarf vorhanden, Datenmanager als verantwortliche Person</p>	<p>Für die Verfügbarkeit wird der Datenmanager als Verantwortlicher und Ansprechperson gesehen. Als</p>

<p>machen. Und dann kann ich mir vorstellen, dass wenn das hier wäre, die Abteilung hat einen Datenmanager, dass er dann dafür verantwortlich wäre, dass jemand weiß, wo die Daten sind...“ (1, 664-668)</p>		<p>Datennachweis sind Metadaten und die Registrierung der Daten wichtig. Zugriffsrechte müssen geregelt sein.</p>
<p>„Genau, also wo ich einfach denke, wäre eben der Datenmanager, dafür hat man eben einen Datenmanager, also er macht natürlich viel bei Datenerhebung [...] aber könnte ich denken, dass wäre auch seine Zuständigkeit, wie gesagt, Pflege wäre was anderes, aber auf jeden Fall ein Ansprechpartner und dass er eben weiß, wer / wo die Daten sind und unter Umständen, eben wo sie ist, die anonymisierte Version des Datensatzes. Was muss man machen, weil es Scientific User File, das wär ja wirklich ein Fall auch in enger Koordination oder Kooperation mit Roland Habich oder wer immer, wenn wir so weit sind, auch zuständig wäre.“(1, 689-699)</p>	<p>Datenmanager als Ansprechperson zum Verbleib der Daten</p>	
<p>„Dafür würde ich mich jetzt auf jeden Fall langfristig nicht zuständig fühlen.“(1, 718)</p>	<p>Verantwortung nicht bei Forschenden</p>	
<p>„Also für die Datenaufbewahrung, et cetera haben wir immer mit Herrn Habich Rücksprache getroffen, meistens wie das so ist. (...) So die explizite Verantwortungs-klärung (lacht) wie immer, also für uns war es wichtig zu wissen, wer ist eben für dieses Projekt weiter verantwortlich, da ist eben [...] * jetzt (...) und ich denke, dass [...] * dann mit Rücksprachen, Rücksprache mit Herrn Habich da quasi dann den Hut auf hat.“(2, 367-380)</p>	<p>Datenmanager in Kooperation mit Projektleitung</p>	
<p>„Weil eben, also müsste man eben auch mit ihm absprechen, also auch mit dem Datenschützer absprechen, also wir haben überlegt, das wären die Schritte und danach wäre das anonym und was heißt dann eine institutionelle Anbindung? Und was müssen die Leute</p>	<p>Nutzung unter Beachtung des Datenschutzes, Zugriffsmöglichkeiten entsprechend wählen</p>	



irgendwie beweisen und, nicht. Wie macht man das per Mittel, also macht das in, schickt man eine CD, müssen sie hier sein, weil das ja auch noch eine Frage ist. Und bei bestimmten Datensätzen kannst du, kann man gar nicht benutzen, sondern, also nicht wie den SOEP oder Mikrozensus bekommt man eine CD, sondern man muss vor Ort sie benutzen.“(1, 699-705)

„Also soweit wie wir verblieben sind, wird das, werden die Daten registriert und sollen dann, ich glaube da war noch nicht ganz klar, wir hatten keine Zeit mehr, aber entweder hätte da die GESIS oder am WZB direkt verfügbar sein.“(2, 215-217)

„Und dann, ja wäre dann, vielleicht wer auch immer bei der GESIS dann anfragt oder halt den Datensatz findet, ist ja auch über die WZB Seite dann, also sollte der Datensatz ja auch online sein, oder zumindest die Beschreibung, wer das dann findet, das ist ja immer die Frage, (lacht) das ist im Prinzip für alle Fächer nutzbar ja, ja.“ (2, 259-262)

„Na ja zumindest erst mal, dass, also wir hatten, ja dieses tolles Metadatenschema und so, dass das also verfügbar ist und die Daten registriert sind und dann diesen, wie hieß es denn? Impermanent? [...]Diese DOI-Nummer einfach bekommen. Genau, permanent identifiziert. Und (...) ja ansonsten (...) wäre es natürlich gut da Ansprechpersonen zu haben, aber na ja oder zumindest die Institution, über die Institution, dass es wieder auffindbar ist und ja (...) ja, ja.“(2, 390-402)

Registrierung der Daten für  
Nachweisbarkeit

Nachweisbarkeit bzw. Zugang  
über Institution, keine  
Beschränkung auf  
Institutszugehörigkeit

Metadaten, DOI,  
Ansprechperson, Zugang über  
Institution

	<p>"Und dann gibt es irgendwie die institutionelle Anbindung, also sprich, da werden Daten weitergegeben, nur an Leute die eine institutionelle Anbindung vorweisen können, weil meistens, also muss man hier auch am WZB versprechen, dass man mit, ja also wissenschaftlichen Arbeiten und [so weiter]*. Und Datenschutz und so und ich glaube genau, wenn wir irgendwie Scientific User File glaube ich machen. Also eben, dass also, nachdem wir eben überlegt haben, was dürften wir weitergeben oder nicht?" (1, 294-300)</p>	Zugriff für Forschende mit Institutsangehörigkeit	
K23: Archivierung / Löschung	<p>„Personenbezogene Daten, jetzt in dem Fall hier, [...] die werden nach Abschluss der Studie gelöscht, beide, das sind ja sensible Daten...“(1, 465-466)</p> <p>„Also wäre ja noch schlimmer, also ich meine, dass sind echt personenbezogene Daten, das steht auch im Datenschutz, die löschen wir. Also nachdem wir, wenn die Studie abgeschlossen ist und dementsprechend sie nicht mehr gebraucht werden, um sie wieder zu kontaktieren, die werden gelöscht.“(1, 471-474)</p> <p>„Das gibt es ja im Prinzip heutzutage eigentlich kaum noch, also sozusagen man müsste sämtliche Rechner auf dem das irgendwie hier liegt, ja und dann ist das ja hier über das Netzwerk im WZB auch irgendwo im zentralen Back-up. Also das wäre auch schon glaube ich ein relativer Aufwand, alles, was mit den Daten zu tun hat, (lacht) irgendwie löschen zu wollen...“(3, 275-278)</p> <p>„Also die, sagen wir so die Daten die aus dem Fragebogen, die man aus den Fragbogen gewonnen hat, na die bleiben ja ewig bestehen.“(1, 455-456)</p>	<p>Löschung aus Datenschutzgründen</p> <p>Löschung aus Datenschutzgründen</p> <p>Vollständige Löschung nicht möglich</p> <p>Archivierung sinnvoll</p>	Die Archivierung der Daten wird als sinnvoll angesehen. Wenn Daten gelöscht werden, dann aus Datenschutzgründen. Informationen über Archivierung sind nicht immer vorhanden.

	<p>„[A]ber die bleiben, also das wäre auch eine Verschwendung, warum sollte man die wegschmeißen?“(1, 460-461)</p> <p>„Aber die Informationen, die also die Erhebungsinformation, die sie uns gegeben haben, die bleibt.“(1, 474-475)</p> <p>„Warum soll man sie wegschmeißen, also ich meine, man hat die Leute stundenlang gequält.“(1, 497)</p> <p>[keine Kenntnis] (2, 270-278)</p>	<p>Archivierung sinnvoll</p> <p>Archivierung sinnvoll</p> <p>Archivierung sinnvoll</p> <p>Archivierung unklar / unbekannt</p>	
K24: Fremddaten-nutzung	<p>„Nee, also für die aktuelle Studie habe ich das, also haben wir das nicht gemacht. [...] Ich glaube, wir werden das auch nicht machen, weil wenn wir das machen könnten, bräuchten wir die Studie ja gar nicht. Also natürlich haben wir bestimmte Daten wie, (...) welche waren das denn, also wie, (...) wie viele (...) [...] es gibt oder wie die [...] in den unterschiedlichen [...] sich verteilen.[...] Also das schon, aber in dem Sinne anderen Daten, im Sinne von [...] nicht.“(1, 515-524)</p> <p>„Also indirekt, also quasi Vorläufer der Studie oder die Studie ist ja entstanden nach so, [...] waren und dann kam die Idee halt eine größere quantitative Studie zu machen.“(2, 283-284)</p>	<p>keine Daten aus anderen Projekten genutzt, außer offizielle Strukturdaten</p> <p>Inhaltlich orientiert an älterer Studie ("Neuaufgabe")</p>	<p>Bereits erhobene Daten werden für das eigene Forschungsprojekt eher aus Infrastruktureinrichtungen oder der Amtlichen Statistik nachgenutzt, außer bei einer Wiederholung einer Studie.</p>

	<p>„Aber sonst andere Daten benutzt haben? (...) Also was wir für die Auswertung genutzt haben, weil ja einige Fragen (...) aus allgemeinen Bevölkerungsumfragen genommen wurden, dann einfach zum Vergleich, also das war bei uns jetzt nicht sehr, ja tief gehend die anderen, bei den anderen Projektpartnern, die machen wirklich systematische Vergleiche auch. Also da wären dann, (...) ich weiß nicht genau was sie für Daten, es waren halt größere allgemeine Bevölkerungsumfragen, so SOEP oder und da ja.“ (2, 293-298)</p>	Daten in geringem Umfang für Vergleichsanalysen nachgenutzt	
K25: Diskrepanz zwischen Theorie und Praxis	<p>„[A]lso mir ist auf jeden Fall bewusst, was alles gemacht wird und was eigentlich auch so theoretisch gefordert wird, was man hinsichtlich Daten machen sollte und so weiter. Aber die Praxis sieht glaube ich einfach anders aus. [...] Na das ist eben der konkrete Fall entscheidend und man muss wahrscheinlich, wenn Sie auch Ihre Arbeit schreiben, unterschiedliche Ebenen einziehen, welche Art von Forschung ist es.“(3, 662-670)</p> <p>„Aber von der Forschung sind wir eben ein Nischenbereich. Und in der Hinsicht stellen sich viele Fragen vielleicht etwas anders.“ (3, 651-652)</p>	nicht alle Anforderungen praktisch umsetzbar	Die praktische Umsetzung der Anforderungen ist abhängig von der Art des Forschungsprojekts und den gegebenen Arbeitsbedingungen.
		Rahmenbedingungen beeinflussen FDM	

Anmerkung der Autorin: [...] \* = anonymisiert

## Anhang E) Legende der Notationselemente<sup>271</sup>

### **Task**

Eine Task ist eine atomare Arbeitseinheit. Sie repräsentiert die Aufgabe, die zu tun ist.

### **Zugeklappter Unterprozess**

Ein zugeklappter Unterprozess ist eine verfeinerbare Aktivität. Er kann mit einem anderen Prozessdiagramm verlinkt werden.

### **Datenbasiertes exklusives Gateway**

Es handelt sich um eine Fallunterscheidung. Abhängig von Verzweigungsbedingungen geht der Prozessfluss an genau einem ausgehenden Zweig weiter. Bei einer Zusammenführung wird auf eine der eingehenden Kanten gewartet, um den Prozess anschließend fortzuführen.

### **Paralleles Gateway**

Bei einer verzweigenden Verwendung werden sämtliche nachfolgenden Zweige gleichzeitig ausgeführt. Bei der Zusammenführung wird auf alle eingehenden Kanten gewartet, bevor der Prozessfluss fortgesetzt wird.

### **Inklusives Gateway**

Bei einer Verzweigung werden eine oder mehrere Wege abhängig von den Verzweigungsbedingungen eingeschlagen. Bei einer Zusammenführung wird auf alle eingehenden aktiven Kanten gewartet.

### **Pool**

Pools repräsentieren Organisationseinheiten und können durch Lanes weiter aufgefächert bzw. unterstrukturiert werden.

---

<sup>271</sup> Quelle: Alle Abbildungen der Elemente und Erläuterungen sind ohne Änderungen übernommen aus einer automatisch generierten Legende des *Signavio Process Editor*, Version 7.2.0.

## Lane

Pools und Lanes repräsentieren Rollen. Lanes repräsentieren Verantwortlichkeiten, wie etwa Organisationseinheiten, Stellen oder IT-Systeme. Lanes können hierarchisch in weitere Unter-Lanes gegliedert sein.

## **Zusätzlicher Beteiligter**

Ein zusätzlicher Beteiligter kann an einen Task oder Unterprozess angeheftet werden. Mit ihm können zusätzliche Beziehungen einer Rolle oder Person zu der jeweiligen Aufgabe abgebildet werden (z.B. fachliche Verantwortung, Benachrichtigte Personen, etc.).

## **Textanmerkung**

Jedes Objekt kann durch eine Text-Anmerkung kommentiert werden.

## **Datenobjekt**

Ein Datenobjekt repräsentiert Informationen, die durch den Prozess fließen, wie z.B. Dokumente, E-Mails oder Briefe.

## **IT-System**

Ein IT-System kann mittels Assoziation mit einer Aktivität verbunden werden. Es handelt sich um ein System oder eine Software-Applikation, die im Laufe der Aktivität verwendet wird.

## **Datenspeicher**

Ein Datenspeicher ist ein Ort, wo der Prozess Daten lesen oder schreiben kann, z.B. eine Datenbank oder ein Aktenschrank. Er existiert unabhängig von der Lebensdauer der Prozessinstanz.

## **Startereignis**

Hier beginnt der normale Prozessfluss durch eine nicht näher beschriebene Auslösebedingung.

 **Nachrichten-Zwischenereignis**

Das Nachrichten-Zwischenereignis wartet auf den Empfang einer Nachricht und setzt daraufhin die Prozessausführung fort.

 **Signal-Zwischenereignis**

Im Laufe der Prozessausführung wird auf das Auffangen eines Signals gewartet, bis die Ausführung fortgesetzt wird. Die Signalgebung kann über verschiedene Prozesse hinweg stattfinden.

 **Nachrichten-Zwischenereignis**

Das auslösende Nachrichten-Zwischenereignis versendet eine Nachricht und setzt daraufhin die Prozessausführung unmittelbar fort.

 **Endereignis**

Untypisiertes Endereignis, welches ein normales Ende eines Prozesses markiert.

 **Nachrichten-Endereignis**

Zum Abschluss des Prozesses wird noch eine Nachricht versendet.

 **Sequenzfluss**

Ein Sequenzfluss definiert die Ausführungsreihenfolge von Aktivitäten.

 **Ungerichtete Assoziation**

Die Zuordnung eines Datenobjekts zu einem Sequenzfluss über eine ungerichtete Assoziation weist auf eine Informationsübergabe zwischen den Aktivitäten hin.

 **Gerichtete Assoziation**

Eine gerichtete Assoziation zeigt den Informationsfluss. Ausgehende Kanten zeigen Schreiben, eingehende zeigen Lesen an.

## **Nachrichtenfluss**

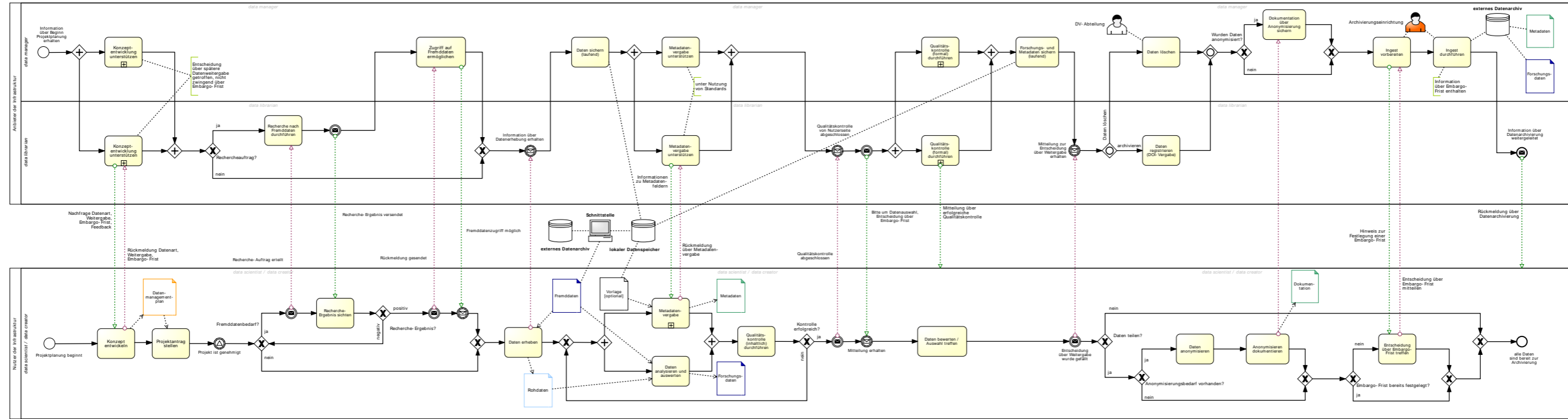
Nachrichtenflüsse symbolisieren Informationen, die über organisatorische Grenzen hinweg fließen. Nachrichtenflüsse können an Pools, Aktivitäten oder Nachrichten-Ereignissen angeheftet sein. Die Abfolge des Nachrichtenaustauschs kann durch das Kombinieren von Nachrichtenfluss und Sequenzfluss spezifiziert werden.



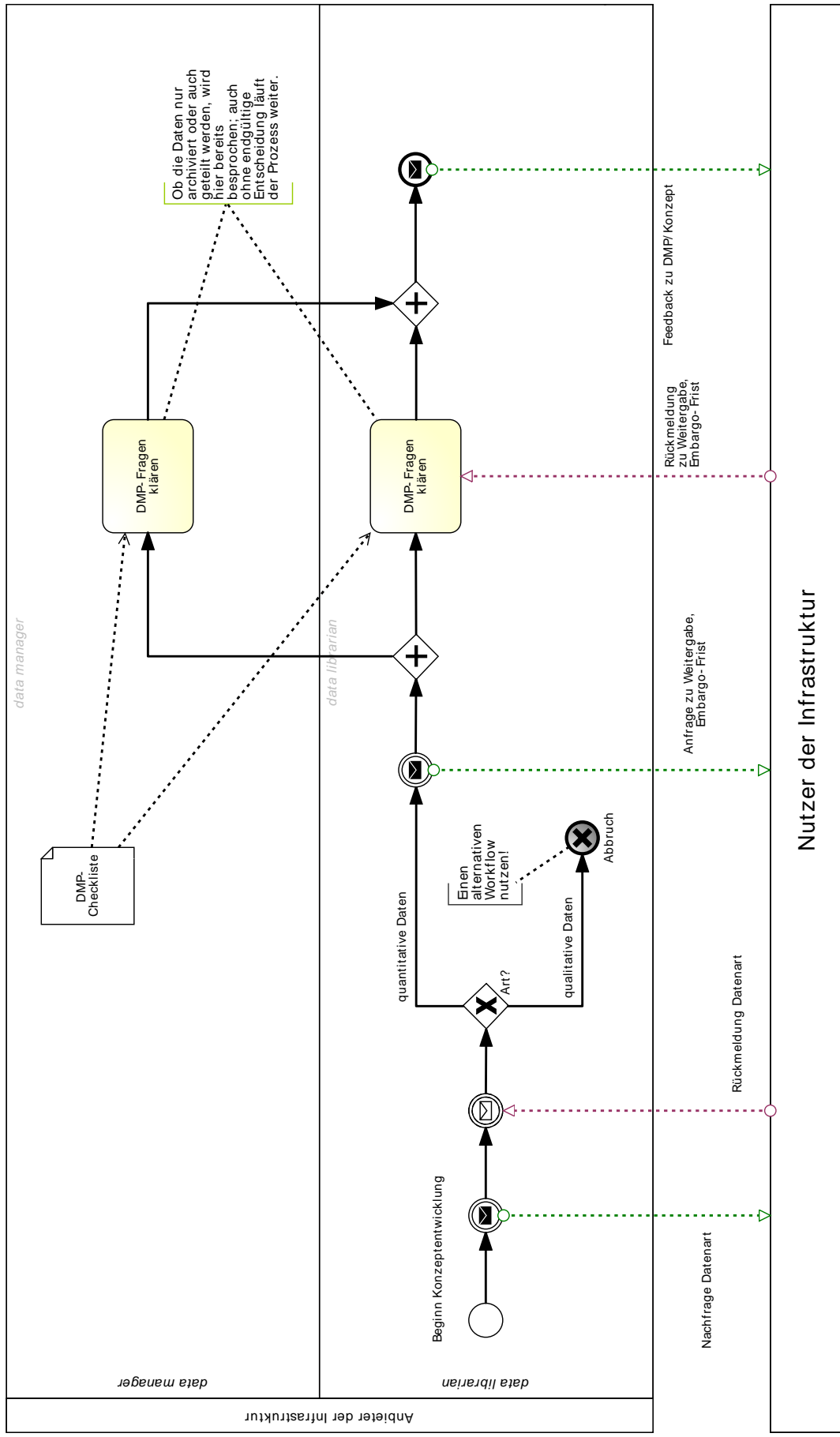
## Anhang F) Geschäftsprozessdiagramm

Auf beiliegender DVD sind folgende Diagramme (XML-Export aus dem *Signavio Process Editor*) enthalten:

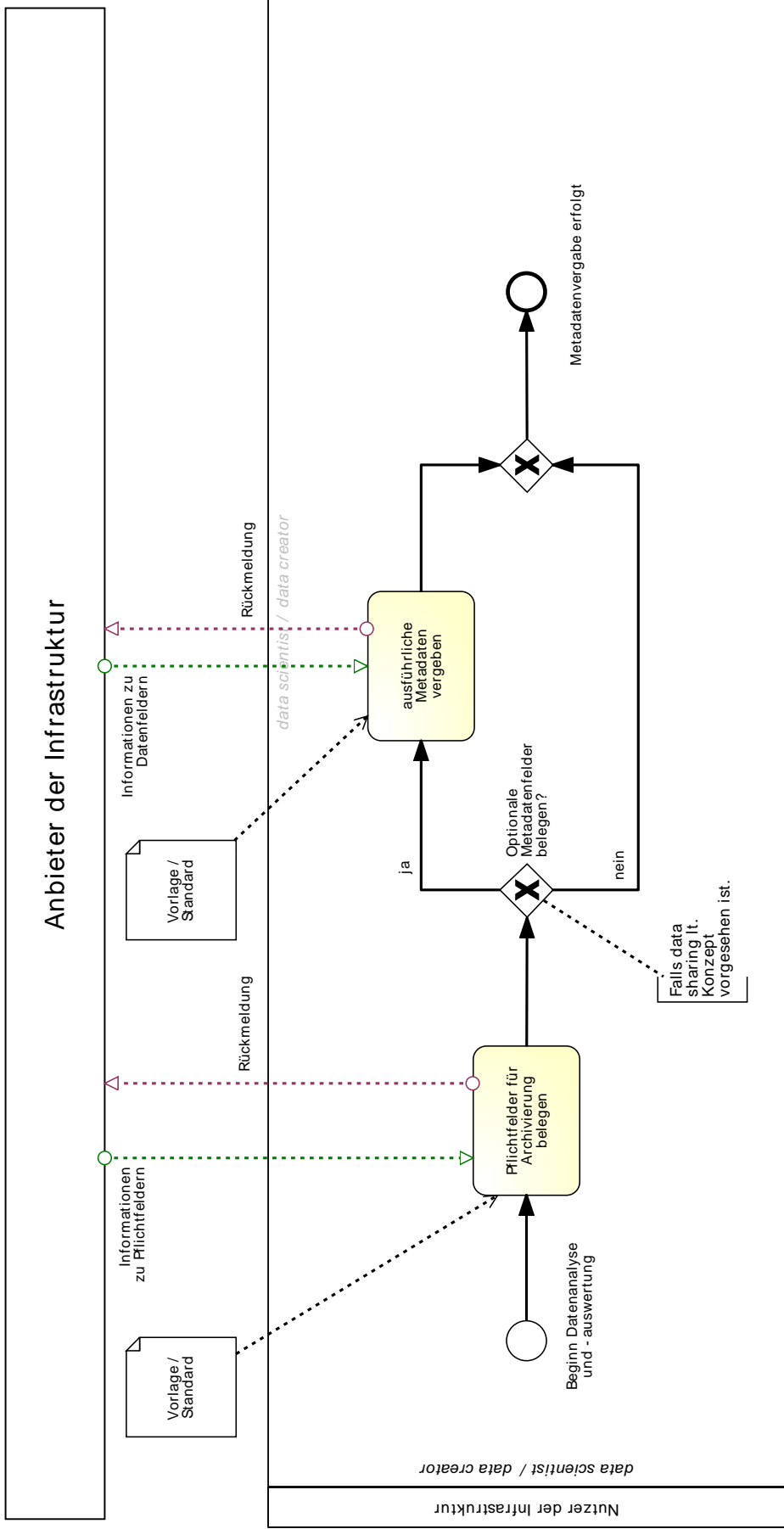
- Gesamtprozess: FDM-Workflow.bpmn
- Unterprozess: Konzeptentwicklung\_unterstützen.bpmn
- Unterprozess: Metadatenvergabe.bpmn
- Unterprozess: Qualitätskontrolle\_formal\_durchführen.bpmn



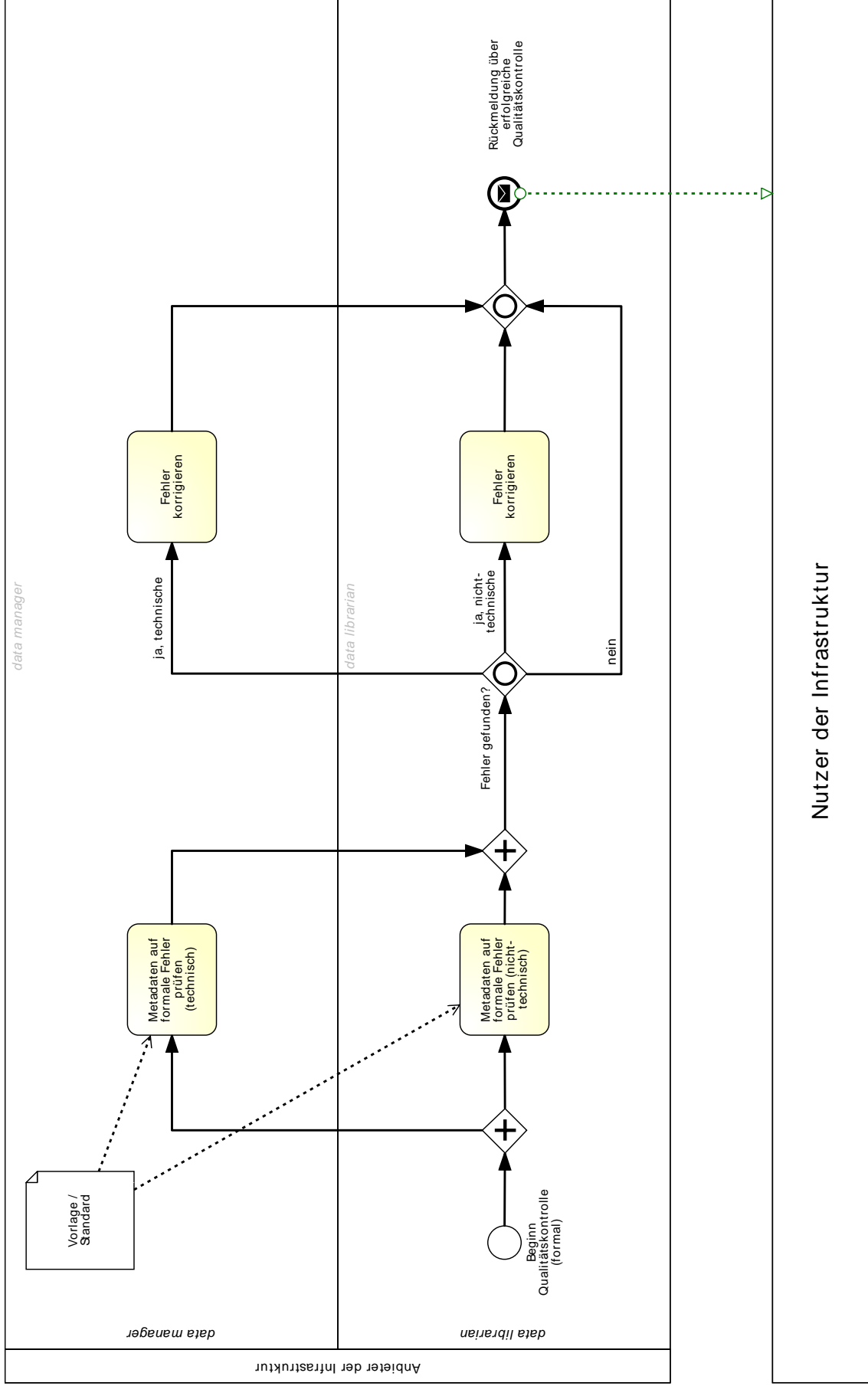
FDM-Workflow



Unterprozess  
Konzeptentwicklung unterstützen



Unterprozess  
Metadatenvergabe



Unterprozess  
Qualitätskontrolle (formal) durchführen

## **Eidesstattliche Erklärung**

Hiermit versichere ich, Christiane Behnert, an Eides statt, dass ich die vorliegende Masterarbeit mit dem Titel „Workflow-Aspekte des Forschungsdatenmanagements - Eine qualitative Anforderungsanalyse am Wissenschaftszentrum Berlin für Sozialforschung“ selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinne nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist weder veröffentlicht, noch in anderer Form als Prüfungsleistung vorgelegt worden.

Potsdam, den 13.08.2013