

SCHLUSSBERICHT ZUM VERBUNDPROJEKT

„SICHERUNG UND NACHNUTZUNG DER FORSCHUNGSDATEN DES RAHMENPROGRAMMS ZUR FÖRDERUNG DER EMPIRISCHEN BILDUNGSFORSCHUNG“

FÖRDERKENNZEICHEN 01JS1302A-C

Zuwendungsempfänger:	DIPF – Bildungsforschung und Bildungsinformation, Frankfurt GESIS – Leibniz-Institut für Sozialwissenschaften e. V., Köln IQB – Institut zur Qualitätsentwicklung im Bildungswesen, Berlin
Verfasser/innen:	Alexia Meyermann, Doris Bambey, Thomas Ebel, Marcus Eisentraut, Malte Jansen, Poldi Kuhl, Reiner Mauer, Claudia Neuendorf, Lisa Pegelow, Maike Porzelt, Marc Rittberger, Thomas Schwager, Petra Stanat, Jessica Trixa
Projektlaufzeit:	01.10.2013 bis 30.09.2016

Schlussbericht zu Nr. 3.2. BNBEST-BMBF 98

INHALTSVERZEICHNIS

1. EINLEITUNG	4
2. AUSGANGSPUNKT	4
3. SICHERUNG UND BEREITSTELLUNG VON FORSCHUNGSDATEN BEIM VFDB (TEILZIELE 1 UND 2)	6
3.1. VERTEILTE ARCHIVIERUNG MIT ZENTRALEM ZUGANG	6
3.1.1. <i>Zusammenarbeit auf der Ebene der Arbeitsorganisation</i>	7
3.1.2. <i>Eingesetzte Technologien</i>	9
3.1.3. <i>Zusammenarbeit auf der rechtlichen Ebene</i>	10
3.2. SICHERUNG DER FORSCHUNGSDATEN: SICHTBARMACHUNG UND ARCHIVIERUNG	10
3.2.1. <i>Datenakquisition: Kontaktaufnahme und Vertragsabschluss</i>	11
3.2.2. <i>Information und Dokumentation (IuD)</i>	16
3.2.3. <i>Dateneingang</i>	22
3.2.4. <i>Datenbereitstellung</i>	27
3.2.5. <i>Stellungnahmen der Projektnehmer/innen gegenüber dem DLR Projektträger</i>	30
3.3. HÜRDEN DER SICHERUNG UND BEREITSTELLUNG DER FORSCHUNGSDATEN AUS DEM EBF-RAHMENPROGRAMM	31
3.3.1. <i>Fehlende datenschutzrechtliche Voraussetzungen</i>	31
3.3.2. <i>Ausschließende behördliche Genehmigungen</i>	34
3.3.3. <i>Fehlende urheberrechtliche Erlaubnis</i>	35
3.3.4. <i>Unzureichende Ressourcenausstattung</i>	36
3.3.5. <i>Einstellungen von Forschenden gegenüber Data Sharing</i>	37
3.3.6. <i>Sonstige Faktoren</i>	39
4. FORSCHUNGSDATENMANAGEMENT: INFORMATION, BERATUNG UND SCHULUNG (TEILZIEL 3)	40
4.1. SCHULUNGEN	41
4.2. INFORMATIONSMANGEBOTE DER WEBSITE	42
4.3. BERATUNGSDIENSTE.....	45
4.4. AUSBLICK	45
5. ZUSAMMENFASSUNG UND AUSBLICK.....	46
6. LITERATUR	51
7. ANHANG	54
7.1. DIE PROJEKTE DES EBF-RAHMENPROGRAMMS.....	54
7.2. KURZBESCHREIBUNG DES ABLAUFS DER ARCHIVIERUNG BEIM VFDB.....	59
7.3. ARBEITSBEREICH DATENAKQUISITION: ERSTELLTE DOKUMENTE.....	61
7.4. ARBEITSBEREICH DOKUMENTATION UND DATENÜBERNAHME: ERSTELLTE DOKUMENTE	61
7.5. VERTRAGLICHE VEREINBARUNGEN UND RECHTLICHE RAHMENBEDINGUNGEN	62
7.6. ERFOLGTE ODER GEPLANTE VERÖFFENTLICHUNGEN DER ERGEBNISSE	62
7.7. KURZEVALUATION.....	64
7.7.1. PLANUNG UND ABLAUF DES VORHABENS	64
7.7.2. ZIELERREICHUNGSGRAD	64
7.7.3. AUSSTATTUNG UND MITTELVERWENDUNG.....	66
7.7.4. PROJEKTINTERNE PROZESSE UND STRUKTUREN	66
7.7.5. ZUSAMMENARBEIT MIT ANDEREN STELLEN / KOOPERATIONEN	68

ABBILDUNGSVERZEICHNIS

ABBILDUNG 1: STRUKTUR DER VERTEILTEN ARCHIVIERUNG MIT ZENTRALEM ZUGANG 7

ABBILDUNG 2: ORGANIGRAMM DES VFDB 9

ABBILDUNG 3: SCREENSHOTS DES ONLINEFORMULARS ZUR DATENMELDUNG 16

ABBILDUNG 4: ABLAUF DER METADATENPRÜFUNG 18

ABBILDUNG 5: SCREENSHOT DER STUDIENSUCHE IM ONLINE-PORTAL FORSCHUNGSDATEN-BILDUNG.DE 22

TABELLENVERZEICHNIS

TABELLE 1: AUFGABENVERTEILUNG DES BEREICHS DATENSICHERUNG IM VFDB 11

TABELLE 2: BEARBEITUNGSSTAND DER KONTAKTIERUNG DER PROJEKTE 14

TABELLE 3: ÄNDERUNGEN KONTAKT 15

TABELLE 4: ANGABEN, DIE ÜBER DAS ONLINE-FORMULAR ERFASST WERDEN 17

TABELLE 5: ANZAHL GEMELDETER DATEN NACH DATENTYP DURCH 133 STUDIEN; MEFHRFACHNENNUNGEN MÖGLICH 20

TABELLE 6: EINGEGANGENE FORSCHUNGSDATEN 25

TABELLE 7: ERGEBNISSE DER EINGANGSPRÜFUNGEN FÜR N=38 DATENBESTÄNDE* 25

TABELLE 8: ZUGANGSSTUFEN DER FDZ IM VERBUND FORSCHUNGSDATEN BILDUNG 27

TABELLE 9: NUTZUNGSWEGE DER FDZ IM VERBUND FORSCHUNGSDATEN BILDUNG 28

TABELLE 10: BEREITGESTELLTE FORSCHUNGSDATEN 29

TABELLE 11: SPERRFRISTEN 30

TABELLE 12: ANGABEN LT. STELLUNGNAHME AN DEN DLR PROJEKTTRÄGER 30

TABELLE 13: AUFGABENVERTEILUNG DES BEREICHS DATENMANAGEMENT IM VFDB 41

TABELLE 14: INFORMATIONEN ANGEBOTE FÜR FORSCHENDE 42

TABELLE 15: INFORMATIONEN FÜR DATENGEBER/INNEN DES EBF-RAHMENPROGRAMMS 43

TABELLE 16: ZUGRIFFE IN DEN JAHREN 2014-2016, EINDEUTIGE BESUCHER 44

TABELLE 17: TOP 5 DER SEITENBESUCHE (EINMALIG) IM INFORMATIONSBEREICH 44

TABELLE 18: TOP 10 EINMALIGE DOWNLOADS 45

TABELLE 19: FORSCHUNGSSCHWERPUNKTE IM EBF-RAHMENPROGRAMM 55

TABELLE 20: ANZAHL PROJEKTE NACH UNTERSCHIEDLICHER ZÄHLWEISE 58

TABELLE 21: ANZAHL PROJEKTE NACH FORSCHUNGSSCHWERPUNKTEN 58

TABELLE 22: PROJEKTE JE PERSON 59

TABELLE 23: ERSTELLTE DOKUMENTE DES BEREICHS DATENAKQUISITION 61

TABELLE 24: ERSTELLTE DOKUMENTE DES BEREICHS DOKUMENTATION UND DATENÜBERNAHME 61

TABELLE 25: VERTRAGLICHE VEREINBARUNGEN UND RECHTLICHE RAHMENBEDINGUNGEN 62

TABELLE 26: TEILZIELE, ARBEITSPAKETE UND JEWEILIGER ZIELERREICHUNGSGRAD 65

TABELLE 27: FÖRDERSUMMEN JE TEILPROJEKT* 66

1. Einleitung

Für den Erkenntnisgewinn, die Qualitätssicherung und letztlich den Erfolg wissenschaftlicher Forschung ist es zentral, für Wissenschaftler/innen einen nachhaltigen und grundsätzlich offenen Zugang zu Forschungsdaten und -ergebnissen zu gewährleisten. Die verlässliche Zugänglichkeit von Forschungsdaten sorgt für Transparenz und Nachvollziehbarkeit und fördert den kumulativen Erkenntnisgewinn. Wissenschaftspolitische Akteure und Zuwendungsgeber fordern und fördern daher seit einigen Jahren verstärkt den Aufbau entsprechender Infrastrukturen zur Archivierung und Nachnutzung qualitätsgesicherter Forschungsdaten.¹ Seit 2008 unterstützt auch das BMBF durch entsprechende Förderauflagen im *Rahmenprogramm zur Förderung der empirischen Bildungsforschung* (EBF-Rahmenprogramm) gezielt die Nachhaltigkeit von Forschungsdaten der Bildungsforschung. So ist es Bestandteil der Förderbedingungen, dass Projektnehmer/innen die in den Projekten generierten Daten und Instrumente archivieren und für die wissenschaftliche Nachnutzung zur Verfügung stellen.² Vor diesem Hintergrund haben sich die Datenzentren der drei Institute DIPF, GESIS und IQB zum Verbund Forschungsdaten Bildung (VFDB) zusammengeschlossen. Ziel des Projektes war es, die Forschungsdaten aus dem EBF-Rahmenprogramm zu sichern und im Rahmen der bestehenden Möglichkeiten zur Nachnutzung bereitzustellen. Die Herausforderung, der sich der VFDB gegenüber sah, bestand darin, eine Vielzahl an Projekten (über 300) aus dem interdisziplinären Forschungsgebiet der Bildungsforschung mit einer entsprechend großen methodischen Vielfalt und mit heterogenen Daten zu betreuen. Da dieses Vorhaben nur kooperativ zu bewältigen war, hat der VFDB in der Förderperiode 10/2013 bis 09/2016 eine Struktur der *verteilten Archivierung* aufgebaut, die eine zentrale Meldekomponente für Forscher/innen mit einer dezentralen, intensiven Kuratierung der Forschungsdaten koppelt.

Im vorliegenden Schlussbericht des Projektes werden die Prozesse und Ergebnisse im Hinblick auf die Meldung, Sicherung und Bereitstellung von Forschungsdaten aus dem EBF-Rahmenprogramm geschildert (vgl. Kapitel 3; Teilziele 1 und 2 des Projektantrags) sowie in einem zweiten Teil über die Informations-, Beratungs- und Schulungsangebote des VFDB zum Forschungsdatenmanagement berichtet (vgl. Kapitel 4; Teilziel 3 des Projektantrags). Voranstehend wird in Kapitel 2 die Ausgangssituation des Vorhabens geschildert.

2. Ausgangspunkt

Im Rahmenprogramm für Empirische Bildungsforschung wurden in den Jahren zwischen 2005 und 2017 303 empirische Projekte gefördert, für die der VFDB zuständig war. Laut Förderbestimmung waren diese Projekte verpflichtet, die im Projekt gewonnenen Daten nach Abschluss des Projektes weiterzugeben. Die im EBF-Rahmenprogramm geförderten Projekte waren zu elf unterschiedlichen Förderschwerpunkten zusammengefasst. Dazu gehörten unterschiedliche thematische Schwerpunkte wie beispielsweise „Professionalisierung pädagogischen Personals“, „Sprachliche Bildung und Mehrsprachigkeit“ oder „Entwicklungsstörungen schulischer Fertigkeiten“ (vgl. Anhang 7.1). Diese Multidisziplinarität des EBF-Rahmenprogramms ging mit einer entsprechend großen methodischen und terminologischen Vielfalt einher. Hieraus resultierte eine große Bandbreite an Datentypen. Zu

¹ vgl. bspw. Allianz der Wissenschaftsorganisationen (2010), Wissenschaftsrat (2012), Kommission Zukunft der Informationsinfrastruktur (2011).

² vgl. bspw. www.bmbf.de/foerderungen/bekanntmachung.php?B=339 (Zugegriffen: 04. Mai 2016); im Jahr 2012 wurden die Auflagen bezüglich Forschungsdaten noch darüber hinausgehend erweitert, vgl. bspw. www.bmbf.de/foerderungen/bekanntmachung.php?B=774 (Zugegriffen: 04. Mai 2016).

den produzierten Datentypen zählten quantitative ebenso wie qualitative Daten und vielfältige Formate wie Tabellen, Bilder (Netzwerkkarten), Videodaten (Unterrichtsbeobachtungen), Audioaufnahmen (Experteninterview) oder Verhaltensdaten (Reaktionszeitmessungen).

Zum Zeitpunkt der Antragstellung existierte kein Infrastrukturangebot, das in der Lage war, die Vielfalt und Vielzahl der Daten aus dem EBF-Rahmenprogramm zu archivieren. Die sozial-, wirtschafts- und verhaltenswissenschaftliche Dateninfrastruktur innerhalb Deutschlands war stark diversifiziert und die Angebote für die Bildungsforschung lagen verteilt an unterschiedlichen Standorten vor. Die Mehrheit der Datenzentren fokussierte auf die Archivierung von Large-Scale-Surveys und/oder die Archivierung institutseigener Daten. Bestimmte Datentypen, wie qualitative Daten, waren unterrepräsentiert. Nur eine Minderheit der Datenzentren bot Angebote zur Archivierung von Daten aus der universitären Einzelforschung. Hierzu gehörten die Datenzentren von DIPF, GESIS und IQB, die Archivierungsangebote für qualitative Daten, für Surveydaten und für Daten aus Schulleistungsstudien anboten. Diese Diversität der Forschungsdateninfrastruktur, die diese bis heute kennzeichnet, geht mit Vor- und mit Nachteilen einher. Von Vorteil ist, dass die vorhandenen FDZ eine fachspezifische Kuratierung von Daten und eine enge Betreuung von Nutzer/innen gewährleisten konnten. Von Nachteil ist, dass die Diversität mit einer entsprechenden Unübersichtlichkeit und Komplexität für Nutzer/innen einhergeht. Diese müssen gegebenenfalls mehrfach und archivspezifisch nach Archivierungs- und Sekundärnutzungsangeboten recherchieren. Auch gab es zum damaligen Zeitpunkt kaum Beratungsangebote zum Forschungsdatenmanagement insgesamt und unserem Wissen nach keine Angebote, die speziell auf die Bedarfe der Bildungsforschung zugeschnitten waren.

Diese Ausgangssituation stellte und stellt besondere Anforderungen an das durch das vorliegende Projekt aufzubauende Dateninfrastrukturangebot: Für die Vielzahl an Daten gilt es, entsprechende (organisatorische, personelle, technische) infrastrukturelle Kapazitäten anzubieten; die *Vielfalt* der Daten bedarf entsprechender Fachexpertise in der erforderlichen Breite. Vor diesem Hintergrund schlossen sich die Datenzentren der drei Institute DIPF, GESIS und IQB zusammen, um bestehende Ressourcen zu bündeln und ein gemeinsames Angebot für das EBF-Rahmenprogramm zu entwickeln. Es erschien nicht sinnvoll, ein eigenes FDZ für die Projekte des EBF-Rahmenprogramms aufzubauen und so zur weiteren Ausdifferenzierung der Forschungsdateninfrastruktur (FDI) beizutragen. Ziel war es stattdessen, die Vorteile der bestehenden Diversität (fachspezifische Kuratierung und Betreuung) zu erhalten und gleichzeitig deren Nachteile (Unübersichtlichkeit, fehlende Nutzerfreundlichkeit) zu vermeiden. Um Kenntnisse im Bereich des Forschungsdatenmanagements stärker zu verbreiten und das Bewusstsein für die Vorteile des Data Sharing zu schärfen, war es weiterhin erforderlich, das Archivierungsangebot durch ein breites Informations- und Schulungsangebot zu begleiten.

Die Kooperation unter Forschungsdatenzentren auf operativer Ebene und die Entwicklung eines gemeinsamen Archivierungsangebots war und ist ein innovatives und exploratives Unterfangen. In der jüngeren Vergangenheit sind weitere Kooperationsprojekte entstanden (z. B. SowiDataNet, GERDI³), die eine stärkere Integration der ausdifferenzierten Landschaft fördern. Diese konzentrieren sich jedoch im Gegensatz zum stärker fachspezifisch orientierten Ansatz des VFDB stärker auf generische Aspekte des Forschungsdatenmanagements. Diese Entwicklungen entsprechen den jüngsten Empfehlungen des Rates für Informationsinfrastrukturen (RfII). Der RfII (2016) konstatiert

³ vgl. GERDI <http://www.gerdi-project.de/> und SowiDataNet <https://sowidatanet.de/>

eine „schwach koordinierte“, durch „projektförmige Initiativen“(ebd.: 1) geprägte und als „Zerfaserung“(ebd.: 36) charakterisierte Forschungsdateninfrastruktur mit geringer Nachhaltigkeit und empfiehlt ein stärker koordiniertes und kooperatives Vorgehen.

3. Sicherung und Bereitstellung von Forschungsdaten beim VFDB (Teilziele 1 und 2)

Ausgehend von den oben beschriebenen Anforderungen und Rahmenbedingungen haben die drei Institute DIPF, GESIS und IQB im VFDB das Kooperationsmodell der „verteilten Archivierung“ realisiert. Das FDZ Bildung am DIPF ist hierbei für die qualitativen Daten zuständig, das GESIS Datenarchiv für Surveydaten und das FDZ am IQB für Daten aus Kompetenz- und Leistungsmessungen. Dieses Modell einer föderierten Infrastruktur (vgl. bspw. RfII 2016: 43f. oder rdn, o.J.) verbindet die Vorteile einer dezentralen Kuratierung mit den Vorteilen eines zentralen Zugangs. In Kapitel 3.1 werden das Modell sowie die dazugehörigen Kooperationsebenen organisatorischer, technologischer und rechtlicher Art vorgestellt. Kapitel 3.2 beschreibt den Prozess der Sicherung der Forschungsdaten aus dem EBF-Rahmenprogramm selbst und berichtet über die Ergebnisse. Im abschließenden Kapitel 3.3 werden bestehende Hürden der Sicherung von Forschungsdaten des EBF-Rahmenprogramms diskutiert.

3.1. Verteilte Archivierung mit zentralem Zugang

Abbildung 1 veranschaulicht den Aufbau des Verbundes Forschungsdaten Bildung und das Modell der verteilten Archivierung. Die Primärforscher/innen konnten Projektinformationen (Metadaten) und Daten ihres Projektes an zentraler Stelle an den VFDB übermitteln. Dort wurden die Metadaten in Datenbanken verwaltet und durch Mitarbeiter/innen des VFDB aufgewertet, um sie über ein Online-Portal an zentraler Stelle zu präsentieren. Die Forschungsdaten selbst wurden an die jeweiligen Einrichtungen weitergeleitet, bei denen die Kuratierung, Archivierung und Bereitstellung der Daten erfolgte.

Abbildung 1: Struktur der verteilten Archivierung mit zentralem Zugang



Dieses Modell der verteilten Archivierung ermöglicht methoden- und datentypspezifische Kompetenzen der unterschiedlichen Datenzentren synergetisch zu nutzen und gleichzeitig gegenüber den Forschenden einen einheitlichen und nutzerfreundlichen Auftritt zu realisieren. Auf der einen Seite trägt die Existenz eines gemeinsamen, FDZ-übergreifenden Portals zur Sicht- und Auffindbarkeit der verteilt aufbewahrten Forschungsdaten bei. Auf der anderen Seite ermöglichen es die fachlichen Profile der VFDB-Partner, bedarfsnah zu agieren: Die Angebote können communityspezifisch gestaltet werden, und die Kommunikation erfolgt adressatengerecht über einschlägige Kanäle. Services können sehr gezielt entwickelt werden, beispielsweise im Hinblick auf konkrete Forschungsfelder oder die Intensität der Nutzerbetreuung bei der Sekundärnutzung von Daten. Durch die Zusammenführung der Angebote unterschiedlicher Datenzentren kann der Diversität und Heterogenität von Methoden und Daten der Bildungsforschung, adäquat fachspezifisch begegnet werden.

Um diesen gemeinsamen Workflow umzusetzen, haben die drei Partner im Verbund eine Zusammenarbeit sowohl auf organisatorischer als auch auf technologischer und rechtlicher Ebene umgesetzt.

3.1.1. Zusammenarbeit auf der Ebene der Arbeitsorganisation

Der VFDB hatte für die Umsetzung seiner Aufgabe im Projekt folgende Arbeitsbereiche eingerichtet: 1) Akquisition und Betreuung von Datengebern (Pre-Ingest), 2) Information und Dokumentation (IuD), 3) Dateneingang und -aufbereitung (Ingest), 4) Datenbereitstellung sowie 5) Datensicherung/Hosting. Diese Organisationsstruktur orientierte sich am ISO-Standard für digitale Langzeitarchive, dem Open Archival Information System (vgl. nestor-Arbeitsgruppe OAIS-Übersetzung/Terminologie 2013). Für die unterschiedlichen Arbeitsbereiche wurden jeweils

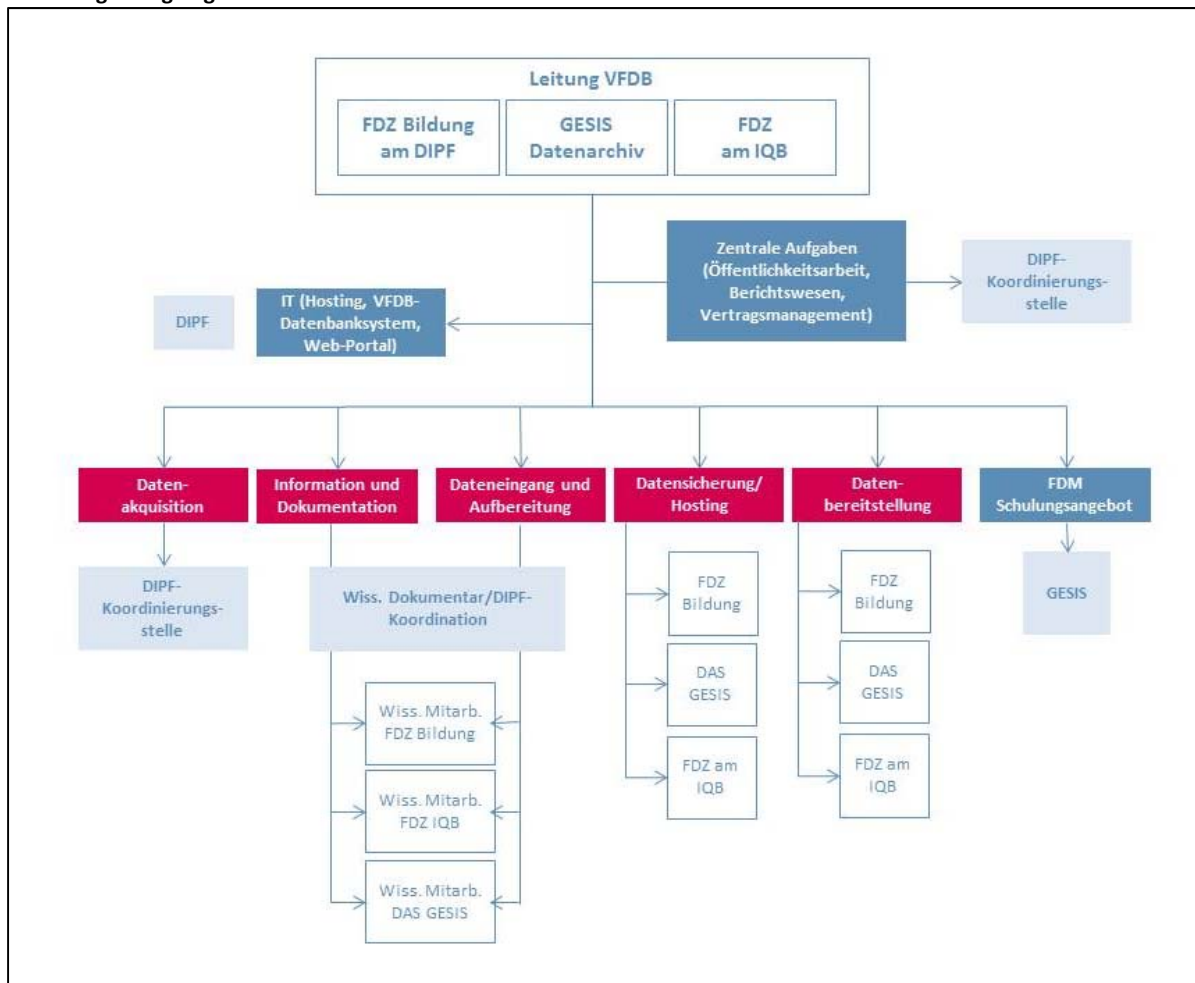
Workflows entwickelt und soweit möglich und erforderlich untereinander abgestimmt. Der erste Arbeitsbereich beinhaltete die Akquisition von Projekten bzw. den dort generierten Forschungsdaten sowie deren Betreuung beginnend mit der Erstsprache bis hin zur Datenübermittlung (vgl. Kapitel 3.2.1). Dabei wurden in erheblichem Umfang Unterstützungs- und Beratungsleistungen sowohl zu prozessualen Fragen der Metadaten- und Datenmeldung als auch zu rechtlichen, forschungsethischen und dokumentarischen Fragestellungen erbracht.

Durch den Arbeitsbereich IuD (vgl. Kapitel 3.2.2) wurden die von den Projekten über ein Online-Formular zum jeweiligen Projekt gelieferten Angaben geprüft, soweit möglich in standardisierte Metadaten überführt und angereichert. Zusätzlich wurden Verknüpfungen mit Literaturdatenbanken erstellt. Ziel dieses Arbeitsbereichs war es, eine Übersicht über die erfassten Projekte und Forschungsdaten zu liefern und diese sicht- und auffindbar zu machen. Dabei wurden gängige Standards für die Vergabe von Metadaten berücksichtigt, um Interoperabilität zwischen und Vergleichbarkeit über verschiedene Datenbank- und Portalsysteme hinweg zu ermöglichen. Dies geschah mit Blick auf einen späteren Ausbau des VFDB-Portals über das EBF-Rahmenprogramm hinaus.

Im Rahmen des Arbeitsbereichs Dateneingang und -aufbereitung (Ingest) wurden die beim VFDB eingegangenen Daten geprüft und zur Sicherung und/oder Nachnutzung aufbereitet (vgl. Kapitel 3.2.3). Datenkuratierung und -bereitstellung erfolgten dezentral durch die jeweils zuständigen Einrichtungen. Die Prüfung der Forschungsdaten erfolgte auf Basis harmonisierter Kriterien (vgl. Kapitel 3.2.3.2). Das Ziel dieses Arbeitsbereichs bestand darin, Forschungsdaten so zu beschreiben und aufzubereiten, dass sie durch Dritte interpretierbar und nachnutzbar sind.

Abbildung 2 zeigt das Organigramm des VFDB für die zentralen Arbeitsbereiche:

Abbildung 2: Organigramm des VFDB



Die Akquisition wurde durch die Koordinierungsstelle am DIPF durchgeführt, die datentypübergreifenden Tätigkeiten im Rahmen des Dateneingangs wurden durch die wissenschaftlichen Dokumentare ebenfalls am DIPF durchgeführt, die datentypspezifischen Tätigkeiten durch die wissenschaftlichen Dokumentar/innen und Mitarbeiter/innen der jeweils zuständigen Einrichtungen. Archivierung und Bereitstellung der Daten für Dritte sowie die Verwaltung der Datenzugänge wurden durch die jeweiligen Datenzentren durchgeführt.

Für eine über die beteiligten Institute vergleichbare und transparente Bearbeitung der Daten wurden Workflows und Leitlinien definiert (vgl. Anhang 7.4). Zur Steuerung der internen Prozesse und um extern auskunftsfähig zu sein, wurde ein institutsübergreifendes Berichtswesen installiert.

3.1.2. Eingesetzte Technologien

In der ersten Projektphase wurde eine technologische, datenbankgestützte und .NET-basierte Infrastruktur für die Dokumentation und Bearbeitung der Projekte des EBF-Rahmenprogramms entwickelt und genutzt. Über ein Eingabetool konnten Forscherinnen und Forschern webbasiert ihre Projekt- und Studienangaben an den VFDB übermitteln. Diese Informationen wurden in einer .NET-Datenbank gespeichert, die die standardgerechte und standortübergreifende Verwaltung der Projekte ermöglichte. Alle drei Partner konnten vom jeweiligen Standort zugreifen, um Projekte zu verwalten und die eigene Arbeit zu dokumentieren. Somit wurde der Anforderung der verteilten Kuratierung und Archivierung Rechnung getragen. Die Metadaten lassen sich über eine JSON-basierte

Schnittstelle aus der Datenbank auslesen und exportieren. Durch die Orientierung an gängigen Metadatenstandards wie DDI 3.2. und dem Metadatenschema von da|ra ist ausreichend Interoperabilität gewährleistet. Über eine erstellte Indexdatei werden die Metadaten außerdem in das zentrale, php-basierte Online-Portal des VFDB, www.forschungsdaten-bildung.de, eingespeist. Die Studien- und Datennachweise sind im Online-Portal über die Software Lucene durchsuchbar. Das Online-Portal bietet der interessierten Öffentlichkeit einen Überblick über Studien des EBF-Rahmenprogramms und einen zentralen Recherchezugang zu den dezentral archivierten Forschungsdaten.

Das Hosting erfolgte durch die Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG). Diese stellte einen LINUX-Server samt Lizenzen sowie ausreichend Speicherkapazitäten von 30 Terabyte inklusive Backup bereit. Die Administration dieser Server erfolgte durch das DIPF. Der Upload erfolgte über einen eigenen LINUX-Server über SFTP-Clients. Die Datengeber/innen erhielten einen eigenen Zugang und konnten ihre Daten in geschützte Verzeichnisse hochladen. Nach Abschluss des Uploads einer Datei wurde diese asynchron über ein Verschlüsselungsskript mit einem General Key verschlüsselt und ein Backup erstellt. Dieses Vorgehen garantierte, dass nur ausreichend geschützte Daten auf den Servern und Speichersystemen des Dienstleisters gelangten. Der passende Private Key war ausschließlich den VFDB-Partnern zugänglich. Bei jedem Zugriff auf die hochgeladenen Forschungsdaten musste zunächst eine dezentrale Entschlüsselung beim jeweiligen VFDB-Partner erfolgen. So kann und konnte eine hohe Sicherheit gewährleistet werden.

3.1.3. Zusammenarbeit auf der rechtlichen Ebene

Zur Regelung der Zusammenarbeit wurde eine Kooperationsvereinbarung zwischen den drei Partnern im VFDB geschlossen. Darin wurden unter anderem die verschiedenen Verantwortungsbereiche geregelt sowie die Rolle und Aufgaben des DIPF als Koordinierungsstelle festgelegt. Das DIPF trat zusätzlich als Vertragspartner des Datengebers auf, wenn dieser die Daten an den VFDB übermittelte, da der VFDB keine eigene Rechtspersönlichkeit besitzt. Die Aufgabe des DIPF bestand darin, die Daten nach Übermittlung durch die Datengeber an die jeweils zuständigen Partner IQB und GESIS weiterzuleiten. Dies wurde in der sog. Übergabvereinbarung zwischen Datengeber und DIPF geregelt (vgl. Kapitel 3.2.1). Die Übergabvereinbarung regelte die Datenübermittlung an das DIPF, die temporäre Sicherung der Daten am DIPF, die Weiterleitung der Daten durch das DIPF an die Partner IQB und GESIS sowie die Dokumentation der Daten im Portal forschungsdaten-bildung.de. Die Herausforderung bestand darin, dass die Übergabvereinbarung sehr unterschiedliche Fälle abdecken sollte, wie die Übergabe anonymisierten und nicht-anonymisierten Datenmaterials, den Verbleib der Daten beim DIPF oder deren Weitergabe. Die Übergabvereinbarung wurde anschließend durch bilaterale Vereinbarungen zwischen den datenhaltenden Einrichtungen und dem Datengeber ersetzt bzw. ergänzt. In diesen bilateralen Vereinbarungen wurden die rechtlichen Voraussetzungen für die langfristige Archivierung geschaffen sowie Fragen der Bereitstellung geklärt wie etwaige Sperrfristen und Zugangsbedingungen.

3.2. Sicherung der Forschungsdaten: Sichtbarmachung und Archivierung

Im Folgenden werden die Ergebnisse und Erfahrungen bezüglich der Archivierung von Forschungsdaten dargestellt. Dies erfolgt anhand der einzelnen Schritte des Archivierungsprozesses. Hierzu gehören die Kontaktphase und die Vertragsvorbereitung (vgl. Kapitel 3.2.1), die Dokumentation der Projekte im Portal forschungsdaten-bildung.de (vgl. Kapitel 3.2.2), der Eingang

der Forschungsdaten und deren Aufbereitung (Kapitel 3.2.3) sowie die Bereitstellung der Forschungsdaten für Dritte (Kapitel 3.2.4). Tabelle 1 zeigt die Aufgabenverteilung im VFDB.

Tabelle 1: Aufgabenverteilung des Bereichs Datensicherung im VFDB

Aufgabe im VFDB	Verantwortliche Einrichtung
Datenakquisition (Kapitel 3.2.1)	DIPF
Information und Dokumentation (IuD) (Kapitel 3.2.2)	Koordination: DIPF Bearbeitung: DIPF, GESIS, IQB je Datentyp
Dateneingang (Kapitel 3.2.3)	DIPF, GESIS, IQB je Datentyp
Datenbereitstellung (Kapitel 3.2.4)	DIPF, GESIS, IQB je Datentyp

Eine Beschreibung der Population der Projekte des EBF-Rahmenprogramms, die Gegenstand der Arbeit des VFDB war und die den nachfolgend berichteten Statistiken zugrunde liegt, findet sich im Anhang (vgl. Kapitel 7.1). Da die tatsächliche Anzahl empirischer Projekte mit Primärdatenerhebungen aufgrund von fehlenden Antworten unbekannt ist (d. h. die Bruttostichprobe), können keine Angaben zu relativen Häufigkeiten (Prozentzahlen) der durch den VFDB dokumentierten Projekte oder archivierten Daten gemacht werden.

3.2.1. Datenakquisition: Kontaktaufnahme und Vertragsabschluss

3.2.1.1. Kontaktaufnahme und Vertragsabschlüsse

Am 21.08.2014 wurden die Projektnehmer/innen durch das BMBF bzw. den DLR Projektträger schriftlich über den VFDB und seine Ziele unterrichtet. (In Ausnahmefällen wurden Projekte später unterrichtet.) Seit Ende des Jahres 2014 wurden die Projektnehmer/innen aus dem Rahmenprogramm nach und nach durch den VFDB per E-Mail kontaktiert und darum gebeten, online Angaben zu ihrem Projekt und den durchgeführten Studien und Datenerhebungen zu machen. In dieser E-Mail wurde den Projektnehmer/innen die Möglichkeit offeriert, die im Projekt generierten Forschungsdaten beim VFDB zu sichern und zur Nachnutzung bereitzustellen. Die in dieser Projektphase manuell ausgeführten Tätigkeiten im Rahmen der Kommunikation und Dokumentation sollen in der zweiten Förderphase durch zu entwickelnde technische Tools unterstützt werden. Die Kontaktaufnahme wurde verantwortet und bearbeitet durch die DIPF Koordinierungsstelle.

Die Übergabvereinbarung war die rechtliche Basis für die Datenübermittlung. Für die Sicherung der Forschungsdaten durch die jeweiligen Institute bzw. die GWDG wurde ein entsprechender Vertrag mit den Datengebern benötigt. Der Archivverbund benötigte Nutzungsrechte, um

- Forschungsdaten und Metadaten zu sichern (bei den jeweiligen Archiven bzw. beim Host GWDG) und um
- Forschungsdaten in öffentlich zugänglichen Onlineportalen über Metadaten beschreiben zu dürfen.

In der Übergabvereinbarung sicherte der Datengeber dem VFDB zu, dass die zu übermittelnden Daten von Ansprüchen Dritter frei sind. Dies beinhaltete sowohl urheber- als auch datenschutzrechtliche Aspekte. Der Vertrag eignete sich sowohl für personenbezogene Daten als auch für die verschiedenen Datentypen, die in den Projekten vorlagen. Das DIPF schloss im Namen des Verbundes die Übergabvereinbarung mit den Datengebern ab. In der Anlage der Übergabvereinbarung wurden die Bestandteile des zu übermittelnden Datenpakets, das Gegenstand des Vertrags ist, spezifiziert. Dies ermöglichte dem Team der wissenschaftlichen Dokumentare (vgl. Kapitel 3.2.3) die hochgeladenen Daten mit den im Vertrag spezifizierten Daten

abzugleichen. Auf Basis der Übergabvereinbarung waren die Partnereinrichtungen des VFDB berechtigt, die Daten zu sichern, aber *nicht* berechtigt, die Daten an Dritte zur Sekundärnutzung bereitzustellen. Hierzu waren zusätzliche, bilaterale Verträge zwischen Datengeber/in und datenhaltender Einrichtung zu schließen. Innerhalb dieser Verträge konnte der/die Datengeber/in besondere Zugangsrestriktionen für die Daten abstimmen wie Sperrfristen oder Informationsrechte.

Das zweistufige Vorgehen (zunächst Abschluss der Übergabvereinbarung und Übermittlung der Forschungsdaten, anschließend Abschluss eines bilateralen Vertrags zur Archivierung und Bereitstellung der Forschungsdaten) hatte sich im bisherigen Verlauf als praktikabel erwiesen. Es ermöglichte dem VFDB, Forschungsdaten zu prüfen, *bevor* weitergehende Absprachen hinsichtlich der Art und Weise der Archivierung und Bereitstellung mit dem Datengeber zu treffen waren. Im Rahmen der Prüfung konnte der VFDB eine Einschätzung zu den Archivierungsmöglichkeiten und der Nachnutzungsfähigkeit⁴ des jeweiligen Datenbestands treffen und auf Basis dieser Kenntnisse mit dem Datengeber Fragen zu Zugangsbeschränkungen (bspw. Anonymisierung, Sperrfristen, Informationsrechte des Datengebers) klären. In die anschließend bilateral zu schließenden Verträge⁵ konnten die Ergebnisse dieser Prüfungen und Verhandlungen einfließen. Ein weiterer Vorteil des zweistufigen Vorgehens bestand darin, dass Forschungsdaten, die aus bestimmten Gründen Dritten nicht verfügbar gemacht werden konnten bzw. sollten, dennoch an den VFDB zur Sicherung übermittelt werden konnten. Hierdurch konnte der Datengeber seine Pflichten gegenüber dem Förderer sowie hinsichtlich der Transparenz und Nachvollziehbarkeit seiner Forschungstätigkeit im Sinne der guten wissenschaftlichen Praxis erfüllen.

Die Laufzeit der Übergabvereinbarung ist zunächst auf zehn Jahre begrenzt. Die Laufzeiten der jeweiligen Partnerverträge, die die Nachnutzung der Daten regeln, ist unbegrenzt bzw. abhängig vom Bestand des jeweiligen Instituts. IQB und DIPF haben bislang keine Verfahren der Langzeitarchivierung implementiert, das GESIS Datenarchiv verfügt über ein entsprechendes Langzeitarchivierungsangebot⁶. Zu beachten ist, dass die am VFDB beteiligten Institute gegenüber den Datengebern des EBF-Rahmenprogramms langfristige Verpflichtungen übernommen haben, obwohl diesen nur eine kurzfristige (dreijährige) Finanzierung des VFDB gegenüberstand.

3.2.1.2. Ergebnisse der Kontaktierung und Vertragsabschlüsse

Bis zum 30.09.2016 wurden von den 303 empirischen Projekten 283 kontaktiert (vgl.

⁴ Hierzu wurde ein Schema zur Bewertung der Nachnutzungsfähigkeit von Forschungsdaten entwickelt (vgl. Kapitel 3.2.4.2).

⁵ Bei GESIS als „Archivierungsvertrag“ bezeichnet, beim FDZ IQB „Nutzungsvertrag Bereitstellung“, beim FDZ Bildung „Zusatzvereinbarung“.

⁶ <http://www.gesis.org/unser-angebot/archivieren-und-registrieren/datenarchivierung/langzeitarchivierung/> (Zugegriffen: 29. Okt. 2015)

Tabelle 2). 121 dieser Projekte konnten final abgeschlossen werden; dabei konnte es sich sowohl um eine Zusage und Archivierung der Daten als auch um eine begründete Absage handeln. 103 Projekte mussten unvollständig abgeschlossen werden, da keine Rückmeldung der Projektnehmer/innen erfolgte.

Tabelle 2: Bearbeitungsstand der Kontaktierung der Projekte

	Anzahl absolut	In % der empirischen Projekte
Noch nicht bearbeitet	22	7
In Bearbeitung	57	19
unvollständig abgeschlossen	103	34
final abgeschlossen	121	40
emp. Projekte insgesamt	303	100

Basis: Quartalsbericht; Stand 30.09.2016

Bislang wurden 54 Verträge zur Übermittlung der Forschungsdaten an den VFDB abgeschlossen. In drei Fällen gab es Änderungswünsche an der Übergabvereinbarung von Seiten des Datengebers, denen in zwei Fällen entsprochen werden konnte.

Während der Akquisitionsphase trat eine Vielzahl von Fragen und Klärungsbedarfen der Projektnehmer/innen auf, die im Folgenden benannt werden.

Klärungsbedarfe administrativer Art

Die Bearbeitungseinheit stellte die Fördereinheit des BMBF dar (identifizierbar über das Förderkennzeichen). Jeder Zuwendungsempfänger (bspw. Universität X) und jede Zuwendung erhält ein eigenes Förderkennzeichen, unabhängig davon ob es sich um ein Einzel- oder ein Teilprojekt eines Verbundes handelt oder um Projekte, die über mehrere Förderphasen laufen. Wenn Datenbestände über zwei Förderphasen oder mehrere Teilprojekte hinweg erstellt wurden, konnte es aus datenarchivarischer Perspektive sinnvoll sein, diese nicht getrennt sondern gemeinsam zu bearbeiten. Dies musste je Einzelfall geklärt werden. Diesen Klärungsbedarf gab es im EBF-Rahmenprogramm bei 156 Projekten, die Teil von 67 Projektverbänden waren und bei 29 Projekten, die über mehr als eine Förderphase liefen (Stand 30.09.2016).

Art der Rückfragen

Rückfragen zum Ablauf bezogen sich darauf, in welcher Form Forschungsdaten übermittelt werden konnten. Hier waren Fragen der Aufbereitung, Anonymisierung und Dokumentation zu beantworten. Klärungsbedarf gab es weiter hinsichtlich des Zeitpunkts der Übermittlung und dazu, welche Daten zu übermitteln waren (sämtliche oder Teile hiervon). Des Weiteren gab es Fragen zum konkreten Ablauf des Archivierungsprozesses, zu den Vertragsinhalten oder zum Upload. Auch Fragen zur Möglichkeit, Sperrfristen einzurichten, wurden übermittelt. Zur Bündelung derartiger Anfragen hat der VFDB umfangreiche Informationen sowie FAQ auf der Website www.forschungsdaten-bildung.de veröffentlicht.

Zeitpunkt der Ansprache

Einige der Projekte baten um einen zeitlichen Aufschub der Übermittlung der Daten, beispielsweise weil bestimmte Arbeiten an/mit den Daten noch nicht abgeschlossen waren oder zum aktuellen Zeitpunkt keine freien Kapazitäten im Projekt vorhanden waren. Eine Vielzahl der empirischen Projekte aus dem EBF-Rahmenprogramm war bereits seit über drei Jahren abgeschlossen (N=121 mit Projektende vor dem 30.09.2013). Deren Bearbeitung war besonders schwierig, da Ansprechpartner/innen fehlten oder diese schlecht erreichbar waren. Es ist anzunehmen, dass die Bearbeitung eines bereits seit über drei Jahren abgeschlossenen Projektes für die Forschenden von niedriger Priorität war.

Der Aufwand der Akquise und Betreuung der Projekte verteilte sich ungleich. Eine geringe Anzahl an Projekten war sehr beratungsintensiv, viele Projekte hatten leicht zu klärende Anfragen und mit einer Vielzahl an Projekten fand kein bilateraler Kontakt jenseits der Standardkommunikation statt.

Projektleitung und Ansprechpartner

In 118 Fällen kam es zu Änderungen hinsichtlich des Kontakts (vgl. Tabelle 3). Die Anzahl der Projekte, in denen es einen oder mehrere weitere Ansprechpartner/innen gab, betrug 94. Teilweise gab es mehr als einen Ansprechpartner, teilweise wechselten die Ansprechpartner innerhalb der Kontaktphase. Bei insgesamt 15 Projekten kam es zu einer Änderung der Affiliation der Projektleitung (vor oder während der Kontaktphase). Neun Projekte konnten nur eingeschränkt bearbeitet werden, weil die Projektleitung inzwischen nicht mehr im Dienst war.

Tabelle 3: Änderungen Kontakt

Änderung	Anzahl Projekte	Anzahl Personen
Anderer Ansprechpartner	94	-
Andere Affiliation	15	10
emeritiert / verstorben / dauerhaft erkrankt	9	8
insgesamt	118	

Basis: Akquisitionstabelle; Stand 30.09.2016

Aufwandseinschätzung Vertragsabwicklung

Die Übergabvereinbarung war zum Download über die Website www.forschungsdaten-bildung.de verfügbar⁷. Um den Aufwand für den Datengeber möglichst gering zu halten, wurde die Möglichkeit implementiert, dass die Übergabvereinbarung neben der postalischen Abwicklung auch elektronisch abgeschlossen werden konnte. Zukünftig soll es hier weitere Vereinfachungen geben (vgl. Abschnitt 3.2.1.3).

Weiterer Aufwand entstand für den oder die Datengeber/in dadurch, dass diese/r ggf. zwei Verträge abschließen musste, da der Vertrag zur Sicherung der Daten und Übermittlung an den VFDB getrennt von dem Vertrag zur Archivierung und Bereitstellung durch eine Partnereinrichtung des VFDB abgewickelt wurde. Daraus folgte, dass Forscher/innen, die unterschiedliche Datentypen generiert hatten, mehr als zwei Verträge abschließen mussten, wenn die Daten bei zwei oder mehr Datenzentren archiviert wurden. Der gesamte Vorgang konnte sich über mehrere Wochen ziehen.

3.2.1.3. Zusammenfassung und Ausblick

Der größte Teil der Kommunikation fand per E-Mail statt und erfolgte überwiegend nach Ablauf der Projektlaufzeit. Unserer Einschätzung nach besteht hier Optimierungsbedarf. Zukünftig soll eine frühere Ansprache der Projekte erfolgen. Wenn möglich und erforderlich sollen Projekte persönlich angesprochen und individueller betreut werden. Gleichzeitig gilt es in der zweiten Förderphase technische Tools zu entwickeln, die den Prozess der Akquisition unterstützen. Da sich das zweistufige Vertragsverfahren als praktikabel erwiesen hat, soll grundsätzlich hieran festgehalten werden. Der damit einhergehende Nachteil höheren Aufwands kann durch eine Veränderung des Verfahrens reduziert werden. Es ist geplant, das bisherige Verfahren zum Abschluss der Übergabvereinbarung durch ein elektronisches Zustimmungsverfahren zu ersetzen. Beim diesem Opt-In-Verfahren sind die Projektnehmer/innen aufgefordert, jeweils vor dem Hochladen der Dateien den vertraglichen Bedingungen per Mausklick zuzustimmen. Somit wird die bislang erforderliche Unterschrift beider

⁷ Siehe www.forschungsdaten-bildung.de/ebf (Zugegriffen: 29. Aug. 2016).

Vertragsparteien (Projektnehmer/innen und DIFP in Stellvertretung des Verbunds) ersetzt. Die Rechtmäßigkeit dieses Verfahrens wurde bereits geprüft. Da Änderungswünsche an der Übergabvereinbarung bisher nur in Einzelfällen aufgetreten sind, ist eine solche Vereinfachung umsetzbar.

3.2.2. Information und Dokumentation (IuD)

Der VFDB-Workflow sah vor, dass Forschende die von ihnen durchgeführten Studien über ein eigens hierfür entwickeltes Online-Formular meldeten (vgl. Kapitel 3.2.2.1). Die Metadaten wurden dokumentarisch geprüft und angereichert (vgl. Kapitel 3.2.2.2).

3.2.2.1. Melden von Projekt, Studie und Daten: Das Online-Formular

Über ein vom *Verbund Forschungsdaten Bildung* entwickeltes Online-Formular konnten die Projektnehmer/innen Angaben zu ihrem Projekt, zu den durchgeführten Studien und den erhobenen oder sekundär genutzten Daten machen. Abbildung 3 zeigt zwei Screenshots des Online-Formulars, das über www.forschungsdaten-bildung.de/melden erreichbar war. Die Eingabe von Informationen war nur Forscher/innen des EBF-Rahmenprogramms unter Angabe von Benutzerkennung und Passwort möglich.

Abbildung 3: Screenshots des Onlineformulars zur Datenmeldung

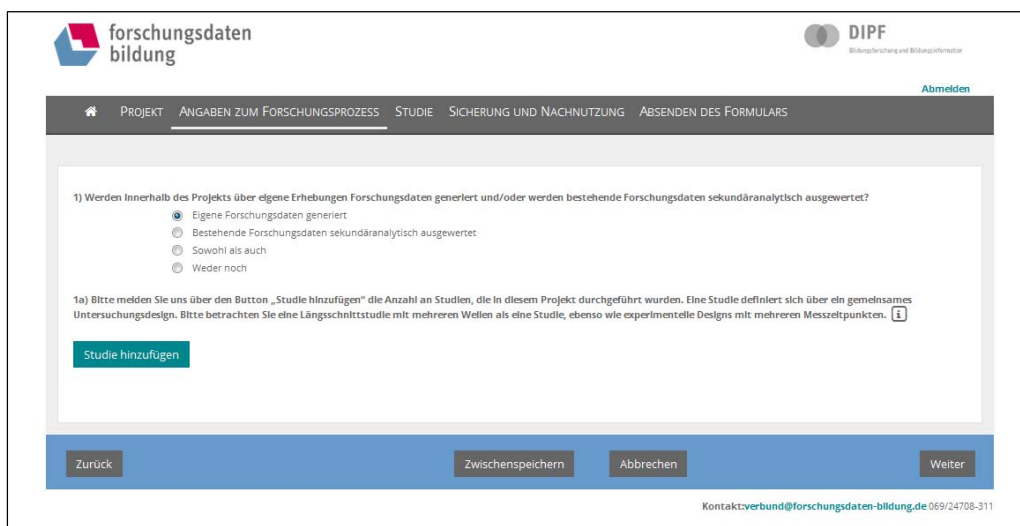
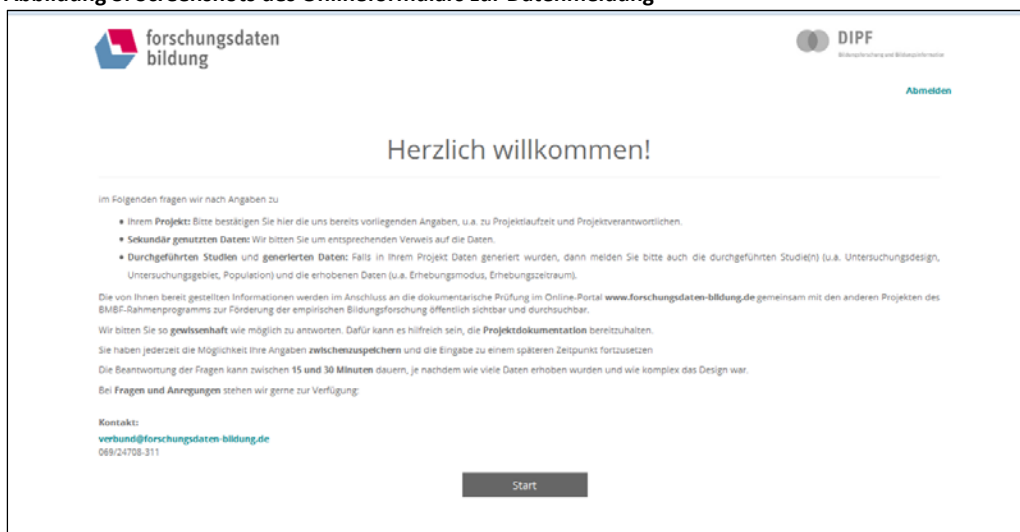


Tabelle 4 listet auf, welche Angaben im Online-Formular erfasst wurden. Die Ausfülldauer variierte stark je nach Komplexität und Umfang des Untersuchungsdesigns eines Projektes.

Tabelle 4: Angaben, die über das Online-Formular erfasst werden

Inhalte	Fragetyp
I Angaben zum Projekt	
Förderkennzeichen	vorausgefüllt
Titel des Forschungsprojektes	vorausgefüllt
Untertitel	vorausgefüllt
BMBF-Forschungsschwerpunkt	vorausgefüllt
Finanzielle Förderung	vorausgefüllt
Projektlaufzeit von bis	vorausgefüllt
Projektleitung	vorausgefüllt
Institutionen	vorausgefüllt
Projektbeschreibung	vorausgefüllt
Schlagwörter	vorausgefüllt
Änderungswünsche oder Ergänzungen zu den voranstehenden Angaben	Freitext
Filter: Erhebung eigener Forschungsdaten, Sekundärdatennutzung oder beides oder keine Daten (ggf. Angabe der genutzten Sekundärdaten)	Multiple Choice (Einfachauswahl); Filterfrage
Gefiltert: Falls Sekundärnutzung: Angaben zu den genutzten Daten	Freitext
II Angaben zur Studie	
Titel der Studie	Freitext
Studienlaufzeit von bis	Freitext
Primärforscher dieser Studie	Drop-down
Kurze Zusammenfassung der Studie	Freitext
Untersuchungsdesign	Multiple Choice (Einfachauswahl)
Gefiltert: Spezifizierung des Untersuchungsdesign	Freitext
Geographisches Untersuchungsgebiet der Studie	Multiple Choice (Einfachauswahl)
Gefiltert: Weitere geographische Angaben	
Grundgesamtheit, Zielpopulation, Stichprobengröße	Freitext
Auswahlverfahren bzw. Feldzugang	Freitext
III Angaben zu den generierten Datentypen	
Arten von Forschungsdaten	Multiple Choice (Mehrfachauswahl); Filterf.
Umfrage- und Aggregatdaten (verantwortlich: GESIS Datenarchiv)	
Zeitraum/Zeiträume der Erhebung (Feldphase/n) von bis	Freitext
Erhebungsmodi	Drop down, Auswahlmenü
Daten aus Kompetenz-/Leistungsmessungen (verantwortlich: FDZ am IQB)	
Zeitraum/Zeiträume der Erhebung (Feldphase/n) von bis	Freitext
Beschreibung der fassten Kompetenzen	Freitext
Skalendokumentation vorhanden	Multiple Choice (Einfachauswahl)
Erhebungsverfahren	Multiple Choice (Mehrfachauswahl)
Gefiltert: Sonstige Beschreibung der Methodik	Freitext
Gefiltert: Beschreibung der Inhalte	Freitext
Anonymisierung	Multiple Choice (Mehrfachauswahl)
Qualitatives, nicht oder gering standardisiertes Datenmaterial (verantwortlich: FDZ Bildung am DIPF)	
Zeitraum/Zeiträume der Erhebung (Feldphase/n) von bis	Freitext
Erhebungsverfahren	Dropdown
Datenmaterial nach Typen	Dropdown
Weitere Materialien zur Datenerhebung und –auswertung	Dropdown
Anonymisierung	Multiple Choice (Mehrfachauswahl)
Sonstiges Datenmaterial	
Zeitraum/Zeiträume der Erhebung (Feldphase/n) von bis	Freitext
Erhebungsverfahren	Freitext
Datenmaterial nach Typen	Freitext
Weitere ergänzende Angaben	Freitext
IV. Sonstiges	
Wurden Daten in nachnutzbarer Form schon verfügbar gemacht	Multiple Choice (Einfachauswahl)
Fragen zur Datenübermittlung	Freitext
Sonstige Anmerkungen, Fragen, Kritik	Freitext

Um den Aufwand für die Projektnehmer/innen zu reduzieren, waren die „Angaben zum Projekt“ im ersten Abschnitt von Tabelle 4 im Formular bereits ausgefüllt. Diese konnten aus einer bereits bestehenden EBF-Projektgedatenbank in die aktuelle VFDB-Datenbank nach einer umfassenden Aktualisierung und Prüfung der Angaben importiert werden. Forschende konnten diese Angaben im Online-Formular einsehen und etwaige Änderungs- oder Ergänzungswünsche in ein Freitextfeld eintragen. Diese wurden anschließend von den Dokumentaren des VFDB geprüft und ggf. umgesetzt. Auf Studienebene („Angaben zur Studie“) wurden Informationen zum Design erfasst (Untersuchungsdesign etc.). Jeder Studie konnten mehrere Datentypen zugeordnet werden, die im Rahmen dieser Studie erhoben wurden, und sie konnten methodisch näher beschrieben werden. Eine Studie war über ein einheitliches Untersuchungsdesign definiert, mehrere Wellen einer Längsschnittstudie oder mehrere Messzeitpunkte eines Experiments sollten als eine Studie erfasst werden.

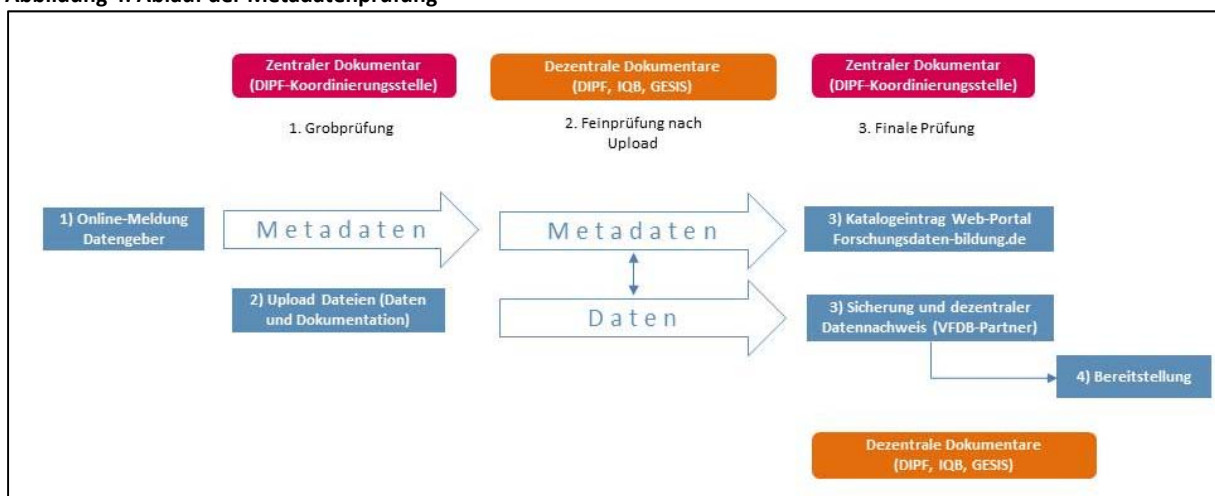
Im VFDB-Onlineformular wurden methodische Angaben zu den generierten Daten nicht detailliert erfasst. Die feingranulare Beschreibung der Daten erfolgte bei den jeweiligen Datenzentren, durch die die Daten archiviert und bereitgestellt wurden (dezentraler Datennachweis). Dieses Vorgehen war sinnvoll, da dort die erforderlichen, spezifischen Fachkenntnisse vorliegen. Die Erfassung detaillierterer Angaben auf Datenebene durch Datengeber/innen bereits im Online-Formular hätte für diese einen aus unserer Sicht zu hohen Aufwand bedeutet.

Der modulare Aufbau des Online-Formulars mit der Möglichkeit, mehrere Studien einem Projekt und mehrere Datentypen einer Studie zuzuordnen, ermöglichte es, die Heterogenität der Projekte im EBF-Rahmenprogramm hinsichtlich Projektstrukturen, Forschungsdesigns, Methoden und Daten in einem Portal abzubilden.

3.2.2.2. **Metadatenprüfungen, Anreicherungen und Kontextualisierung**

Die Einträge im Online-Formular wurden automatisch in die VFDB-Datenbank geschrieben. Auf dieser Basis führten die Dokumentar/innen die Metadatenprüfungen durch. Abbildung 4 zeigt eine Übersicht über den Ablauf Metadatenprüfungen im VFDB.

Abbildung 4: Ablauf der Metadatenprüfung



Eine *erste Grobprüfung* der Metadaten wurde durch die DIPF-Koordinierungsstelle durchgeführt. Diese fand nach Eingang des Online-Formulars und damit vor Abschluss einer etwaigen

Übergabvereinbarung und vor einem etwaigen Upload der Forschungsdaten statt. Die Grobprüfung der Metadaten erfolgte vor Vertragsabschluss und Upload, damit die Möglichkeit bestand, vorab offene Fragen zu klären. Bei der Grobprüfung wurden die von den Projektnehmer/innen getätigten Angaben in den Freitextfeldern (Anmerkungsfeldern) sowie die Strukturierung von Studien und Daten im jeweiligen Projekt nach definierten dokumentarischen Kriterien geprüft. Zuständigkeiten zwischen den Partnern wurden vergeben.

Die *Feinprüfung* der Metadaten fand *nach* dem Upload der Forschungsdaten statt und wurde durch die jeweils zuständigen Partner durchgeführt (vgl. Kapitel 3.2.3.2). Im Rahmen der Feinprüfung konnten die Metadaten mit den Informationen aus den hochgeladenen Forschungsdaten und Begleitmaterialien verglichen und ggf. ergänzt oder korrigiert werden. Im Rahmen dieser Prüfungen erfolgten auch Anreicherungen der Metadaten (Abstracting, Verschlagwortung u. a.). Im Ergebnis entstanden aus den geprüften und angereicherten Metadaten ein Katalogeintrag im Webportal www.forschungsdaten-bildung.de und ein dezentraler Datennachweis in den Portalen der jeweiligen Einrichtungen im Verbund Forschungsdaten Bildung. Im Portal erscheinen Metadaten auf Projekt- und auf Studienebene mit Verweis auf die Datenverfügbarkeit bei den jeweiligen Partnern. Auf den Websites der Partnereinrichtungen werden die Daten noch detaillierter beschrieben. Die Metadaten zu Studien mit ausschließlicher Sekundärdatennutzung werden im Portal direkt angezeigt. Metadaten zu den Studien, die Forschungsdaten beim VFDB sichern lassen, werden erst nach Upload dieser veröffentlicht.

Die Prüfungen der qualitativen Studien wurden vom FDZ Bildung am DIPF durchgeführt, die quantitativen Studien mit Umfrage- und Aggregatdaten vom GESIS Datenarchiv und die Studien mit Leistungs-/Kompetenzmessungen vom FDZ am IQB. Bei Projekten mit mehr als einem Datentyp wurde eine Hauptzuständigkeit vergeben. Für die einzelnen Schritte der Metadatenprüfung (Grob- und Feinprüfung) ebenso wie zur Benutzung der Verwaltungsdatenbank wurden *Leitfäden* erstellt (vgl. auch Anhang 7.4). In Kasten 1 findet sich eine Auflistung der Bestandteile der Metadatenprüfungen im Rahmen der Studiendokumentation.

Kasten 1

Bestandteile der Metadatenprüfungen im Rahmen der Studiendokumentation

- Prüfung der Anmerkungsfelder hinsichtlich Ergänzungs- oder Korrekturwünschen oder besonderen Hinweisen
- Prüfung, ob die Zuordnung von Studien und Daten, die Anzahl erfasster Studien pro Projekt korrekt ist
- Zuweisung der Zuständigkeiten: Das DIPF weist die Kontrollaufgaben nach Datentypen zu. Die Zuständigkeiten lassen sich nach Absprache ändern.
- Rechtschreib- und Grammatikprüfung der Eintragungen
- Plausibilitätsprüfung: Prüfung, ob bestimmte Angaben inhaltlich logisch und plausibel sind. Beinhaltet z. B. der Zeitraum der Projektlaufzeit die Zeiträume der Studienlaufzeiten? Sind Untersuchungspersonen 250 Jahre alt? (Vergleiche mit externen Quellen erfolgen nur in Ausnahmefällen.)
- Konsistenzprüfung: Die Informationen im Online-Formular sollten untereinander konsistent sein. Konsistenz sollte innerhalb eines einzelnen Projektes gegeben sein, wird aber soweit möglich auch zwischen den verschiedenen Projekten angestrebt. (z. B.: Stimmt die angegebene Anzahl an Studien und die angegebene/n Art/en von Daten mit den im weiteren Verlauf gemachten Angaben überein?)
- Vollständigkeit der Angaben: Wurden notwendige Metadaten vollständig übermittelt? Nicht ausgefüllte Felder können je nach deren Bedeutung beim Datengeber nachträglich ermittelt oder durch eigene Recherche selbst ergänzt werden. Bsp.: Datensatzname/DOI bei Sekundärdatennutzung.
- Prüfung auf datenschutzrechtlich sensible Inhalte

- Anreicherung und Harmonisierung der Metadaten: ggf. Kürzungen der Projektbeschreibungen, Verschlagwortung, Standardisierung der Metadaten nach Leitfäden und kontrollierten Vokabularien, Referenzierungen der Daten zu anderen Datenbanken
- Bereinigung der offenen Angaben/Freitextfelder (Vereinheitlichung der Metadaten): Nach dem offiziellen Leitfaden zur Nutzung der Datenbank müssen die Werte und Eintragungen den Vorgaben angepasst und somit standardisiert werden. Bsp: Die Freitextantworten sind in ein (semi-)kontrolliertes Vokabular zu überführen.
- Dokumentation der Ergebnisse der Prüfung in der Projektadministration.

Die Dokumentar/innen besprachen in 14-tägigen Sitzungen offene Fragen gemeinsam mit der/dem Verantwortlichen für den Bereich Akquisition. Bei Bedarf wurde der Forscher/die Forscherin um weitere Informationen gebeten.

Die Erfassung der Metadaten im Rahmen der Studiendokumentation stellt einen zentralen Bestandteil der Archivierung von Forschungsdaten im Allgemeinen dar, denn zur Archivierung von Forschungsdaten gehört deren Sichtbarmachung (Katalogisierung) in Portalen und das Erzeugen von Auffindbarkeit. Des Weiteren dient die Prüfung von Metadaten der Qualitätssicherung. Eine hinreichende Qualität und Vergleichbarkeit von Metadaten ist zentral, da diese *sowohl* die Auffindbarkeit der Daten gewährleisten *als auch* die Basis für die Entscheidungen zukünftiger Nachnutzer/innen für/wider eine Sekundärnutzung dieses Datenbestands bilden. Nachnutzende sollten das Analysepotential der Forschungsdaten einschätzen können, um beurteilen zu können, ob diese zur Beantwortung der eigenen Forschungsfrage geeignet sind sowie, um die methodische Beschaffenheit der Daten im Hinblick darauf beurteilen können, welche Implikationen dies für die Wahl der eigenen Analyse und die Interpretation der Ergebnisse hat. Aufgrund der Bedeutung der Metadaten für den langfristigen Erhalt und die langfristige Auffindbarkeit von Forschungsdaten bedurfte diese Arbeit entsprechender, dokumentarischer Expertise, die im VFDB vorlag.

3.2.2.3. Ergebnisse: Übermittelte Online-Formulare

Bis zum 30.09.2016 hatten 133 Projektnehmer/innen Projekte über das Online-Formular gemeldet. Aufgrund von Teilprojekten, für die nicht jeweils eine eigene Meldung erforderlich war, lagen dem VFDB 151 Meldungen für Projekte vor. Von den 151 Projekten nutzten 21 Projekte ausschließlich oder zusätzlich zur Primärerhebung Sekundärdaten.

Von den gemeldeten Projekten mit Primärdatenerhebung wurden in 38 Projekten qualitative Daten erhoben, in 46 Projekten Surveydaten/Umfragedaten, in 63 Projekten Kompetenz- und Leistungsdaten und in elf Projekten andere Datenarten (vgl. Tabelle 5). In 47 Projekten wurde mehr als ein Datentyp generiert, wobei es sich in 14 Projekten um eine Kombination aus quantitativen und qualitativen Daten handelte (sog. Mixed Methods Studien).

Tabelle 5: Anzahl gemeldeter Daten nach Datentyp durch 133 Studien; Mehrfachnennungen möglich

Datentypen	Qualitative	Quantitative (Surveydaten)	Quantitative mit Tests/Leistungs-messungen	Sonstige	Insg.
Anzahl	38	46	63	11	158

Basis: Verwaltungsdatenbank; Stand 30.09.2016

Die von den Forscher/innen über das Meldeformular zur Verfügung gestellten Informationen wiesen überwiegend einen ausreichend bis guten Detailliertheitsgrad auf. Eine Anschlusskommunikation mit Datengebern oder Datengeberinnen war meist nicht erforderlich. Einzelne Felder wurden jedoch

teilweise unvollständig ausgefüllt (Auswahlverfahren, Untersuchungsgebiet). Wurden in Studien keine eigenen Daten generiert, sondern bereits existierende Daten verwendet, waren die Angaben zu den verwendeten Daten sehr unterschiedlich. Teilweise wurden eindeutige Identifikatoren angegeben, teilweise unspezifische Datensatzbezeichnungen. Möglich ist, dass dies weniger mit einer unkorrekten Angabe des Forschers oder der Forscherin zusammenhing, sondern damit, dass viele in der Bildungsforschung genutzten Datenbestände schlicht nicht eindeutig identifizierbar sind, weil persistente Identifikatoren oder eindeutige Datensatzbezeichnungen fehlen.

Projektnehmer/innen hatten im Meldeformular die Möglichkeit, Änderungswünsche hinsichtlich der vorausgefüllten Angaben auf Projektebene zu äußern. Hier gab es überwiegend nur leichte Korrekturwünsche (insb. hinsichtlich Schlagwörtern, Projektbeschreibungen und Projektteams). Insgesamt wurden die offenen Anmerkungsfelder von den Projektnehmern und Projektnehmerinnen nur in geringem Ausmaß genutzt. Gelegentlich wurden diese genutzt, um auf Probleme bezüglich Datenschutz oder Datenübermittlung hinzuweisen.

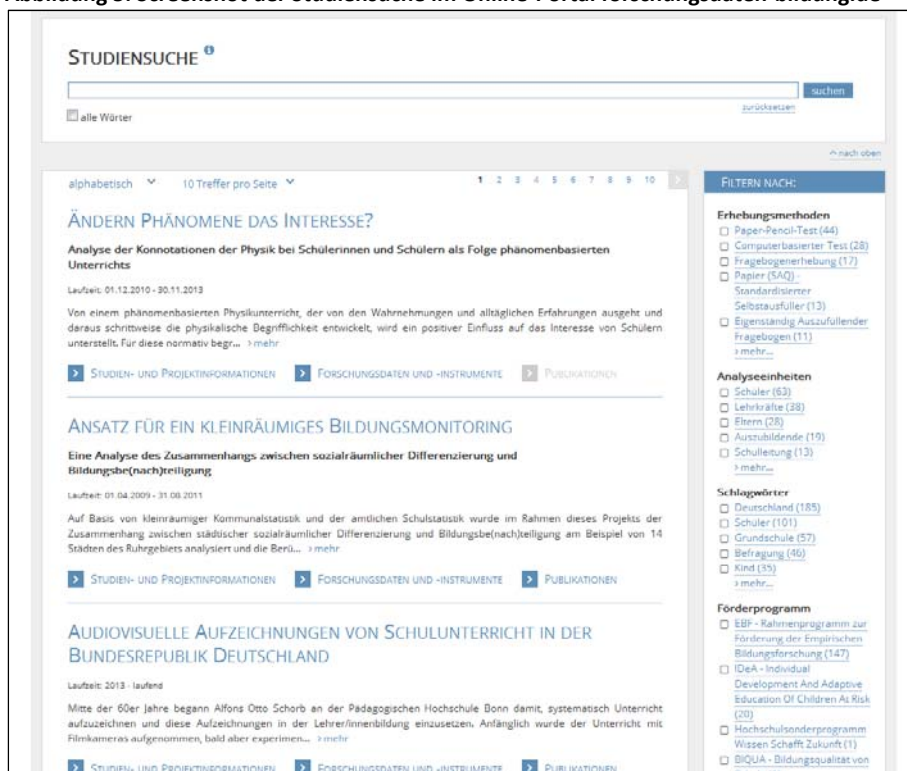
Wie oben erläutert, war das Meldeformular modular aufgebaut mit der Möglichkeit, jedem Projekt mehrere Studien zuzuordnen zu können und jeder Studie mehrere Datenkollektionen. Mit dem modularen und flexiblen Aufbau des Meldeformulars kamen die Projektnehmer/innen unterschiedlich gut zurecht. Teilweise wurde keine „Studie“ angelegt. Dies führte zur Wiedervorlage des Formulars bzw. zur Kontaktierung der Datengeber/in. Teilweise mussten die Dokumentare und Dokumentarinnen die durch den/die Datengeber/in angelegte Struktur anpassen.

Vereinzelte wurden Begriffe und methodisches Vokabular von den Projektnehmer/innen unterschiedlich verwendet. So existierte zum Beispiel eine unterschiedliche Verwendung des Begriffs „Zufallsstichprobe“. Das war aufgrund der disziplinären Vielfalt und den damit verbundenen, unterschiedlichen Traditionen hinsichtlich der Begrifflichkeiten zu erwarten gewesen. Die Bearbeiter/innen nahmen hier die erforderlichen Anpassungen vor.

In der Vergangenheit unvollständig oder fehlerhaft ausgefüllte Formulare wurden als Anlass genommen, das *Design des Online-Formulars* neu zu gestalten. Nach dem Re-Design (Mai/Juni 2015) nahm die Anzahl der Fälle, in denen beispielsweise vergessen wurde, Angaben auf Studienebene zu machen, stark ab.

Die Prüfung der Meldeformulare erfolgte in der Regel innerhalb einer Woche nach deren Eintreffen durch die wissenschaftlichen Dokumentar/innen. Für die Prüfung der Metadaten (des Meldeformulars) benötigte das Dokumentationsteam in der Regel ca. 30 Minuten. Die Dauer variierte je nachdem, wie umfangreich das Projektdesign war (ein oder mehrere Studien) und wie umfassend die Projektnehmer/innen Angaben gemacht hatten. Bei unzureichend dokumentierten Studien fielen Bearbeitungszeiten von über einer Stunde an. Zusätzlich kam die Dauer der durch die DIPF-Koordinierungsstelle erfolgten Grobprüfung hinzu (jeweils ca. 5 Minuten). Problematisch war, dass sich die Metadatenprüfungen über mehrere Monate hinziehen konnten: beginnend mit den Prüfungen nach Projektmeldung und vor Dateneingang bis hin zu den Prüfungen und Anpassungen, die nach Dateneingang erfolgten. Bis zum 30.09.2016 wurden durch die Dokumentar/innen des VFDB 151 Projekte final geprüft. Diese werden im Online-Portal angezeigt. Abbildung 5 zeigt einen Screenshot der Studiensuche im Portal.

Abbildung 5: Screenshot der Studiensuche im Online-Portal forschungsdaten-bildung.de



Stand: 31.09.2016

3.2.2.4. Das Portal www.forschungsdaten-bildung.de: Zusammenfassung und Ausblick

Die Online-Meldung wurde von den meisten Projekten ausgefüllt. Die Angaben waren überwiegend von ausreichendem Informationsgehalt. In wenigen Fällen wurde um ergänzende Informationen gebeten. Durch die dokumentarische Prüfung der Angaben, deren Standardisierung sowie die Anreicherung der Metadaten unter anderem durch Verknüpfungen zu Literaturdatenbanken entstand ein besonderer Mehrwert für Nutzer/innen, die sich einen Überblick über die empirische Bildungsforschung verschaffen möchten.

Zukünftig soll der Meldeprozess für die Datenproduzierenden insbesondere durch Anpassungen des Layouts weiter vereinfacht werden. Dies geschieht auf Basis der in der Pilotphase gemachten Erfahrungen. Usability Tests oder eine Evaluation sind vorgesehen. Unabhängig von den Bemühungen der FDZ, die Aufwände für Datenproduzierende gering zu halten, bleiben diese gefordert, möglichst vollständige Angaben zu machen. Nur so kann eine qualitativ hochwertige Dokumentation erzeugt werden. Bei der Dokumentation von Studien und Daten galt es, eine Balance zu finden zwischen den dokumentarischen Anforderungen an Genauigkeit und Vollständigkeit und den Aufwänden, die dies für Projektnehmer/innen mit sich bringt. Herausforderungen für beide Seiten stellen hierbei nach wie vor die Interdisziplinarität des Forschungsfeldes, die terminologische Vielfalt sowie die Heterogenität der Forschungsdesigns und Methoden innerhalb der Bildungsforschung dar. Neben dem Thema Usability wird in der Ausbauphase des VFDB verstärkt das Thema Information Retrieval (Suchoptimierung und Navigation) adressiert.

3.2.3. Dateneingang

Als Datenübernahme oder Ingest wird in der einschlägigen Literatur (vgl. z. B. nestor-Arbeitsgruppe OAIÜbersetzung/Terminologie 2013) die Übernahme digitalen Archivguts bezeichnet. Im Rahmen

der Datenübernahme wird das übermittelte Datenpaket (das sog. *Submission Information Package*) für die Archivierung vorbereitet und durch eine Reihe von Arbeitsschritten ein zu archivierendes Datenpaket (sog. *Archival Information Package*) erstellt, „das den Standards des Archivs für Datenformate und Dokumentation entspricht“ (nestor-Arbeitsgruppe OAIS-Übersetzung/Terminologie 2013: 33).⁸

Der VFDB-Workflow sah vor, dass Forschende nach Meldung ihrer Daten und Vertragsabschluss die Daten sowie die dazugehörige Dokumentation online an den VFDB übermitteln (vgl. Kapitel 3.2.3.1). Die Daten wurden durch die wissenschaftlichen Dokumentar/innen und wissenschaftlichen Mitarbeiter/innen im VFDB geprüft sowie zur Archivierung und – soweit möglich – zur Bereitstellung vorbereitet. Im Folgenden werden die dazugehörigen Arbeitsschritte erläutert und die jeweiligen Ergebnisse und Erfahrungen mit der Bearbeitung der Projekte des EBF-Rahmenprogramms berichtet.

3.2.3.1. Übermittlung der Forschungsdaten - Der Uploadvorgang

Forscher/innen konnten ihre Daten hochladen, sobald der Vertrag fertiggestellt war und sie vom VFDB die hierfür erforderlichen Zugangsdaten erhalten hatten. Das Hochladen der Dateien erfolgte über eine gesicherte Verbindung. Hierzu mussten sich die Projektnehmer/innen einen SFTP-Client (z. B. WinSCP oder FileZilla) installieren. Nach Eingabe einer Benutzerkennung konnten Sie mittels Drag & Drop Forschungsdaten und dazugehörige Materialien hochladen. Der SFTP-Client erlaubte das Hochladen großer Datenvolumina. Der Uploadvorgang konnte unterbrochen und zu einem späteren Zeitpunkt fortgesetzt werden, was bei instabilen oder langsamen Internetverbindungen von Vorteil ist. Als Hilfestellung für den Upload wurden entsprechende Hinweise erstellt (vgl. Tabelle 15). Die Forscher/innen waren aufgefordert, ihre Daten – soweit möglich und nötig – getrennt nach Datentyp hochzuladen, da dies die weitere Bearbeitung durch die Verbundpartner erleichterte.

Die Dateien wurden direkt zu dem externen Hosting-Partner des VFDB (der GWDG) hochgeladen und dort automatisch verschlüsselt. Über den Dateneingang wurde der zentrale Dokumentar am DIPF automatisch per E-Mail benachrichtigt. In dieser E-Mail waren weitere sog. Logging-Informationen (Anzahl und Art der hochgeladenen Dateien, Größe, Format, Datum u. a.) enthalten. Hierdurch waren die jeweiligen Mitarbeiter/innen zeitnah über den Uploadvorgang informiert sowie die Uploadinformationen dokumentiert.

3.2.3.2. Dateneingangsprüfungen

Nach Eingang der Daten erfolgten durch die jeweils zuständigen Partnerinstitute die Eingangsprüfungen der hochgeladenen Dateien bzw. Forschungsdaten. Die Prüfungen dienten dem Zweck, den Erhalt der Daten zu gewährleisten. Die Sicherung wurde zunächst für einen Zeitraum von mindestens zehn Jahren zugesagt. Die Forschungsdaten sollten interpretierbar sein, d. h. vollständig, verständlich und (aus archivarischer Perspektive) fehlerfrei. Ergaben sich bei der Eingangsprüfung Unklarheiten, waren diese durch Rücksprache mit dem/der Datengeber/in zu klären. Sämtliche Prüfungsinhalte waren zwischen den Partnereinrichtungen abgestimmt, um ein hohes Maß an Vergleichbarkeit zu erreichen. Die formal-technischen Prüfungen waren bei den Partnern identisch, die inhaltlichen Prüfungen unterschieden sich je Datentyp, waren aber soweit wie möglich äquivalent zueinander.

⁸ Der VFDB hat sich hier auch an den Vorgehensweisen weiterer einschlägiger Datenzentren orientiert (vgl. z. B. UK Data Archive 2014; vgl. auch Harvey 2010, Kapitel 12).

Im ersten Schritt der Eingangsprüfung erfolgte die *technisch-formale Prüfung* der Dateien: Hier wurde geprüft, ob die übertragenen Dateien mit der in der Übergabvereinbarung spezifizierten Dateiliste übereinstimmte, es wurden Checksummen erstellt und Virenskans durchgeführt. Anschließend wurden Kopien der Originaldateien erstellt, die Originale verschlüsselt und in der Folge nicht mehr bearbeitet. Nun wurden die Dateinamen der Kopien geprüft und gegebenenfalls bereinigt (u. a. durch Entfernen von Sonderzeichen und Leerzeichen), und die einzelnen Dateien wurden auf Funktionsfähigkeit überprüft (ließen sich diese öffnen und waren sie lesbar). Der VFDB hat für diese Prüfung eine unter den Partnern abgestimmte Checkliste erstellt, die sich an international gängigen Ingest-Checklisten orientiert (vgl. z. B. Harvey 2010; vgl. Anhang 7.4).

Im zweiten Schritt wurden *Prüfungen der Daten und der Dokumentation* durchgeführt. Diese waren im Vergleich zur technisch-formalen Prüfung, die für alle Datentypen gleichermaßen angewandt werden konnte, datentypspezifisch und basierten auf den Kriterien des zuständigen Partnerinstituts im VFDB. Hier wurde geprüft, ob die übermittelten Daten mit der dazugehörigen Dokumentation und den bereits vorliegenden Metadaten (aus dem Online-Formular, vgl. Kapitel 3.2.2.1) übereinstimmten, und ob die Daten vollständig waren. Bei quantitativen Daten wurde u. a. geprüft, ob die Variablen im Datensatz mit den Items im Fragebogen übereinstimmten, ob die Variablen- und Wertelabels vorhanden und verständlich sowie Codebücher vorhanden waren. Bei qualitativen Daten wurde beispielsweise geprüft, ob die vorliegenden Beobachtungsprotokolle oder Transkripte audiovisuellen Dateien (z. B. Unterrichtsaufzeichnungen) zuzuordnen waren, ob die Transkription oder Protokollierung vollständig oder in Ausschnitten erfolgte, und ob weitere (soziodemographische) Informationen zu den Teilnehmer/innen oder zur Interview-/Beobachtungssituation vorhanden waren. Weiter wurden die datenschutz- und urheberrechtlichen Bedingungen der Forschungsdaten geprüft. Waren die Daten anonym oder personenbezogen, lagen Einwilligungen zu den geplanten Zwecken der Datenverarbeitung vor, wie sensibel war das Material und wie hoch waren Identifikationsrisiken?

Nach der Überprüfung der Daten und nach Absprache mit dem/der Datengeber/in wurde die Archivierungsentscheidung getroffen. Bei Daten, die die Mindestanforderungen im Bereich der Dokumentation nicht erfüllten, d. h., wenn die Verständlichkeit der übergebenen Materialien noch nicht ausreichte, um Nachvollziehbarkeit zu gewährleisten, wurde eine Bereitstellung auch von der Möglichkeit und Bereitschaft der Datengeber abhängig gemacht, weitere Dokumentationsanstrengungen zu unternehmen. War ein großes Interesse der Datengeber an einer Bereitstellung erkennbar, konnten in der Regel gemeinsame Lösungen gefunden werden, um die Qualität der Dokumentation auf ein Mindestniveau anzuheben. Dies hing vor allem vom Engagement der ehemaligen Projektleiter sowie ihrer Mitarbeiter ab, die (häufig unentgeltlich) neben ihrer aktuellen Tätigkeit noch diese zusätzlichen Aufträge erfüllten.

Nach den Prüfschritten wurden die zugehörigen Datenbankinformationen auf Studien- und Datenebene vom zuständigen VFDB-Partner aktualisiert. Falls die Daten bereitgestellt wurden, wurden entsprechende Referenzierungen (i. d. R. eine DOI) und Angaben zum Datenzugang ergänzt. Nach Abschluss der Eingangsprüfung wurde in einem finalen Schritt der komplette Katalogeintrag vom zentralen, wissenschaftlichen Dokumentar (DIPF) redigiert. Bei Bedarf wurden in Rücksprache mit den Verbundpartnern letzte Unklarheiten beseitigt. Abschließend wurde der Katalogeintrag vom DIPF freigeschaltet und im Portal www.forschungsdaten-bildung.de inklusive Referenzierungen (falls vorhanden) angezeigt (vgl. auch Kapitel 3.2.2).

3.2.3.3. Ergebnisse

Bis zum 30.09.2016 lagen dem VFDB 46 einzelne Datenbestände aus dem EBF-Rahmenprogramm vor. Tabelle 6 zeigt, wie sich die hochgeladenen Forschungsdaten auf die einzelnen Einrichtungen verteilen. In der ersten Förderphase sind aus 13 Studien qualitative Daten aufgenommen worden, aus acht Studien quantitative Umfragedaten und aus 25 Studien quantitative Daten mit Kompetenz- und Leistungsmessungen. Vier Projekte haben quantitative und qualitative Daten hochgeladen. Der relative Anteil eingegangener Forschungsdaten ist auf Basis der uns vorliegenden Informationen nicht anzugeben (vgl. die Populationsbeschreibung im Anhang 7.1).

Tabelle 6: Eingegangene Forschungsdaten

	Eingegangene Forschungsdaten
FDZ Bildung am DIPF	13
GESIS Datenarchiv	8
FDZ am IQB	25
Insgesamt	46

Stand 30.09.2016

Nach dem Dateiupload erfolgten die Eingangsprüfungen durch die jeweiligen FDZ. Die Ergebnisse dieser Prüfungen über alle FDZ hinweg sind in Tabelle 7 aufgeführt. Bis zum 30.09.2016 waren Bewertungen von 38 der 46 Datenbestände erfolgt. Es zeigt sich, dass die Bewertungen überwiegend gut bis sehr gut waren und nur in wenigen Fällen eine Sicherung oder Bereitstellung der Daten verhinderten. Insbesondere hinsichtlich der technischen Eignung wurden keine Probleme registriert. Die positiven Ergebnisse in der Kategorie „rechtliche Eignung“ ergeben sich vor allem daraus, dass Datenbestände mit fehlender rechtlicher Eignung häufig gar nicht erst übermittelt wurden und daher nicht Gegenstand der Bewertung waren.

Tabelle 7: Ergebnisse der Eingangsprüfungen für N=38 Datenbestände*

Noten	A = sehr gut	B = gut	C = mäßig	D = unzureichend	
Kriterien	Anzahl	Anzahl	Anzahl	Anzahl	Summe
Rechtliche Eignung	16	18	4	0	38
Technische Eignung	36	2	0	0	38
Dokumentarische Eignung	10	17	8	3	38
Gesamturteil	14	16	6	2	38

Stand 30.09.2016

* Eingangsprüfungen durch das IQB wurden durchgeführt für sämtliche Daten; Eingangsprüfungen durch GESIS und DIPF wurden nur durchgeführt für solche Daten, die zur potentiellen Nachnutzung übermittelt worden waren.

Der Bereitstellung quantitativer Daten durch das FDZ am IQB standen mitunter datenschutzrechtliche Probleme entgegen: Zum einen konnten Forschungsdaten aus Kompetenz- und Leistungsmessungen mit Personenbezug in der Regel nicht vollständig übermittelt werden. Zum anderen musste regelmäßig geprüft werden, ob der realisierte Anonymisierungsgrad ausreichend war. Das hieß, dass Daten durch das Datenzentrum ggf. noch nachanonymisiert werden mussten, indem bspw. Variablen geleert und/oder Kategorien zusammengefasst wurden. So konnten dann auch diese Daten datenschutzrechtlich korrekt bereitgestellt werden. Die Überprüfung von Anonymisierungsgrad und Re-Identifikationsrisiko war sowohl bei quantitativen als auch bei

qualitativen Daten einer der zeitintensivsten Datenprüfprozesse. Für audiovisuelle Forschungsdaten (z. B. videographierte Unterrichtsbeobachtungen oder Schüler- und Lehrerinterviews) zeigte sich, dass diese aufgrund ihres Personenbezuges meist gar nicht erst übermittelt werden konnten. Stattdessen wurden häufig „nur“ die anonymisierten Transkripte dieser Aufzeichnungssituationen übermittelt.

Die technischen Eigenschaften der übermittelten Daten erfüllten bisher alle die archivseitigen Mindestanforderungen. Bisher kamen Eingänge unbekannter Datenformate nur in Einzelfällen vor. Dabei handelte es sich um Neurodaten/bildgebende Verfahren. Diese Daten konnten lediglich gesichert, nicht aber geprüft werden.

Der Grad der Dokumentationstiefe von Datengeberseite aus variierte sehr stark. Bei qualitativen Studien war zu konstatieren, dass nur in seltenen Fällen eine umfassende Dokumentation des Erhebungskontextes sowie Zusatzmaterialien geliefert wurden, die eine Interpretation der Daten wesentlich erleichtert hätten. Ein Sonderproblem qualitativer Daten trat auf, wenn nur Teile des erhobenen Datenmaterials zur Verfügung gestellt wurden. Beispielsweise war dies der Fall bei Studien, in denen sowohl Videoaufzeichnungen als auch darauf bezogene Interviews durchgeführt wurden und uns aus datenschutzrechtlichen Gründen nur die Interviewtranskripte, nicht aber die Videos übermittelt werden konnten. Aufgrund von engen Bezügen der unterschiedlichen Daten zueinander kann sich das Analysepotential erheblich verringern, wenn nur die Interviewtranskripte vorliegen. Bei quantitativen Daten variierte die Qualität der Dokumentation sehr stark, deren Vollständigkeit war nur begrenzt feststellbar.

Aufwandseinschätzung

Die Aufwände, die im Rahmen des Dateneingangs aufgebracht wurden, wurden im letzten Projektjahr pilothaft erfasst anhand der einzelnen Arbeitsschritte Eingangskontrolle, Vertragsabschluss/Kommunikation, Aufbereitung und Dokumentation/Bereitstellung. Ergebnisse der Messungen lagen zum Projektende jedoch erst für eine geringe und noch nicht aussagekräftige Menge an bearbeiteten Daten vor. Festgehalten werden kann, dass die einzelnen Schritte im Prozess des Dateneingangs und der -aufbereitung von Projekt zu Projekt stark variierten.

3.2.3.4. Zusammenfassung und Ausblick

Die Prüfung der eingehenden Daten, die durch die jeweiligen Partnereinrichtungen erfolgte, ist ein zentraler Bestandteil der Tätigkeit von professionellen Forschungsdatenzentren. Hier wird deutlich, welche besonderen Leistungen zu erbringen sind, um Forschungsdaten über längere Zeiträume hinweg erhalten zu können und zukünftige Nachnutzungen zu ermöglichen. Die Arbeiten im VFDB standen vor der besonderen Herausforderung, über die einzelnen Einrichtungen hinweg vergleichbare und hohe Qualitätsstandards der Arbeitsprozesse zu implementieren. Diese Ziele sollen in der folgenden Förderphase weiterverfolgt werden. Daneben werden einige Änderungen auf operativer Ebene erfolgen.

In der nächsten Förderphase soll das temporäre Hosting der eingehenden Forschungsdaten nicht wie in der Pilotphase des Projektes über einen externen Dienstleister (1. Förderphase: GWDG) erfolgen, sondern wird durch das DIPF übernommen. Die Dateien werden dann direkt auf DIPF-Server hochgeladen und gespeichert, auf die die jeweiligen Partnereinrichtungen Zugriff zum Download haben. Für dieses Vorhaben sind in der nächsten Förderphase entsprechende Mittel eingestellt. Zusätzlich zum SFTP-Upload sollen den Forschenden weitere Uploadmöglichkeiten angeboten

werden, die sich für kleinere Datenmengen (und anonymisiertes Material) eignen. Hier ist beispielsweise die „CryptShare“-Software zu nennen, mit welcher das GESIS Datenarchiv bereits seit längerem arbeitet und positive Erfahrung gemacht hat, da sie maximale Sicherheit bei der Datenübertragung gewährleistet, gleichzeitig aber auch selbsterklärend und daher nutzerfreundlich ist. Auch eine Übermittlung auf verschlüsselten Datenträgern (CD-ROM, DVD, USB-Stick) sowie auf einen Onlineserver (vergleichbar mit Dropbox) wäre möglich. Die neuen Uploadmöglichkeiten sollen auch die Aufwände für Datengeber/innen weiter reduzieren.

3.2.4. Datenbereitstellung

3.2.4.1. Zugangswege und Restriktionen

Das dritte zentrale Aufgabengebiet des VFDB war die Bereitstellung oder Zugänglichmachung von Forschungsdaten für Dritte. Die über den VFDB eingesammelten Forschungsdaten wurden nach Aufbereitung und Dokumentation dieser durch die jeweiligen Verbundpartner ebenfalls dezentral über die jeweiligen Zugangswege bereitgestellt. Im Rahmen der Bereitstellung wurden die Zugangsbedingungen mit den Datengebern abgestimmt und vertraglich fixiert.

VFDB-Datenzentren bieten derzeit folgende Zugangsstufen an:

Tabelle 8: Zugangsstufen der FDZ im Verbund Forschungsdaten Bildung

GESIS Datenarchiv	FDZ am IQB	FDZ Bildung am DIPF
0 Zugang für jedermann (nach Registrierung)	Zugang nach Registrierung und bewilligtem Antrag (Voraussetzung: Nutzung der Daten für wissenschaftliche Zwecke sowie Anbindung des Datennutzers an eine Forschungseinrichtung)	Zugang für jedermann nach Registrierung
A Zugang für akademische Forschung und Lehre nach Registrierung		Zugang für akademische Forschung nach Registrierung und Antragstellung
B Daten sind für Forschung und Lehre freigegeben, wenn keine Veröffentlichung der Ergebnisse erfolgt (Publikationen dürfen nur mit einer Genehmigung des Datenarchivs erfolgen)		
C: Datenweitergabe erst nach schriftlicher Genehmigung des Datengebers		

Zugangsstufen unterscheiden sich im Ausmaß der Zugangsrestriktionen, die für einen bestimmten Datenbestand gegeben sind. Forschungsdaten können frei zugänglich sein, zugänglich für Wissenschaftler/innen nach Registrierung oder zugänglich für Wissenschaftler/innen nach Registrierung und Antragstellung. Das FDZ Bildung am DIPF prüft für sensible Datenbestände beispielsweise die Identität einer Antragstellerin per Post-Ident-Verfahren und das konkrete Forschungsvorhaben. Für jedes Forschungsvorhaben, d. h. für jede Nutzung eines Datenbestands ist ein gesonderter Antrag zu stellen. Eine registrierte Nutzerin erhält in diesem Fall keinen Zugriff auf sämtliche Daten im Bestand des FDZ Bildung, sondern nur auf diejenigen, für die ein Nutzungsantrag gestellt und genehmigt wurde. Welche Zugangsrestriktionen für einen bestimmten Datenbestand festgelegt werden, entscheidet sich in der Regel nach dem Sensibilitäts- und Anonymitätsgrad der Daten.

Für die Nachnutzung von Forschungsdaten sind verschiedene Nutzungswege möglich. Daten können an den Nutzer per DVD oder Download herausgegeben werden. Per Remote-Access können Nutzer/innen vom eigenen Arbeitsplatzrechner über eine gesicherte Verbindung auf die Daten zugreifen, es ist aber technisch nicht möglich, die Daten zu speichern oder zu übertragen. Bei der kontrollierten Datenfernverarbeitung (KDFV) schickt der oder die Nutzer/in Auswertungsprogramme an das FDZ, FDZ-Mitarbeiter/innen führen die Auswertung der Daten durch und schicken die Ergebnisse (nach datenschutzrechtlicher Unbedenklichkeitsprüfung) an den oder die Nutzer/in zurück. Auf besonders sensible Daten erhalten Nutzer/innen ausschließlich bei einem Gastaufenthalt an besonders geschützten Rechnern in den Räumen des FDZ Zugriff.

Für die Daten aus dem EBF-Rahmenprogramm wurden bisher sog. Scientific Use Files (SUF) für die (ausschließlich) wissenschaftliche Nachnutzung erstellt sowie durch das FDZ am IQB sog. IZA-Files für die kontrollierte Datenfernverarbeitung.

Tabelle 9: Nutzungswege der FDZ im Verbund Forschungsdaten Bildung

Nutzungswege	GESIS Datenarchiv	FDZ am IQB	FDZ Bildung am DIPF
Scientific Use Files (Herausgabe der Daten per Datenträger oder Download)	Ja, per DVD und Download	Ja, per Download (via Datenaustauschportal)	Ja, per Download
Remote Access	Nein	Nein	Nein
Datenfernverarbeitung	Nein	Ja	Nein
On-site Nutzung (Gastaufenthalt)	Ja	Ja	Nein

3.2.4.2. Bewertung der Daten und Erstellung der Datenpakete

Die eingegangenen Daten aus dem EBF-Rahmenprogramm wurden archivarisch-dokumentarisch nach festgelegten Kriterien geprüft und v. a. hinsichtlich ihrer Nutzbarkeit bewertet. Das Datenpaket, das für die Nachnutzung zur Verfügung gestellt wird, unterscheidet sich vom *Archival Information Package* (vgl. Kapitel 3.2.3) und wird daher im OAIS-Funktionsmodell als *Dissemination Information Package* bezeichnet (vgl. nestor-Arbeitsgruppe OAIS-Übersetzung/Terminologie 2013). Dieses enthält möglicherweise nur eine bestimmte Auswahl der insgesamt archivierten Daten wie beispielsweise ausschließlich stark anonymisierte Daten oder Daten einer Subpopulation der Studie. Zum Teil wurden an den Instituten auch unterschiedliche Datenpakete erstellt, die je nach Zugangsstufe unterschiedliche Anonymisierungsstufen erfüllen (zum Beispiel Scientific Use Files (SUF) oder weniger stark anonymisierte Versionen die in der kontrollierten Datenfernverarbeitung oder an speziellen Arbeitsplätzen vor Ort genutzt werden können (IZA-Files beim IQB).

Gründe, die gegen eine vollständige oder teilweise Bereitstellung der Daten sprachen, waren das Vorliegen anderer als quantitativer oder qualitativer Daten der Sozialforschung (z. B. medizinische Daten), das Fehlen erforderlicher Genehmigungen von Landes- oder Schulbehörden, besonders sensible und schutzwürdige Populationen, fehlende oder unzureichende Einverständniserklärungen, persönliche Verpflichtungen des Forschenden gegenüber den Probanden, die Unvollständigkeit der Daten oder eine unvollständige Dokumentation.

Zur Bewertung der Studien wurde vom VFDB ein Schema zur Bewertung der Nachnutzungsfähigkeit von Forschungsdaten entwickelt, das die grundlegende Auswahl von Forschungsdaten zur Aufbereitung und Bereitstellung anleiten, transparent gestalten und interpersonell vergleichbar machen soll. Eine solche Einschätzung ist bei der Priorisierung der Bearbeitung von Studien hilfreich

und dient einem effizienten Ressourceneinsatz. Die Bewertung der Daten erfolgte bei den Datenzentren auf Grundlage der Prüfungen nach Dateneingang (vgl. Kapitel 3.2.3.2). Der Entwurf des Bewertungsschemas enthielt die Basiskategorien „rechtliche Anforderungen“, „technische Anforderungen“ und „dokumentarische Anforderungen“.

- rechtliche Anforderungen: Für die Datennachnutzung gibt es keine rechtlichen Einschränkungen bspw. datenschutzrechtlicher oder urheberrechtlicher Art
- technische Anforderungen: Die Dateiformate sind mit gängiger Software zu bedienen und können in für die Langzeitarchivierung geeignete Formate überführt werden.
- dokumentarische Anforderungen: Die Daten sind aufbereitet und Daten sowie Dokumentation liegen vollständig und korrekt vor.

Das Schema war beschränkt auf die Bewertung der *Nachnutzungsfähigkeit* aus dokumentarisch-archivarischer Perspektive. Aussagen zum *Nachnutzungspotential* aus wissenschaftlicher Perspektive im Sinne des inhaltlichen Analysepotentials wurden im Rahmen des Schemas von den Bearbeiter/innen nicht getroffen. Jede dieser Basiskategorien wurde vom Dokumentationsteam auf einer vierstufigen Skala bewertet. Die Skala unterscheidet Anforderungen der Nachnutzung zwischen „vollständig erfüllt“, „überwiegend erfüllt“, „teilweise erfüllt“ und „nicht erfüllt“. Die drei Basiskategorien sind gleichrangig zueinander. In der zweiten Förderphase wird das Bewertungsschema weiterentwickelt und publiziert werden.

3.2.4.3. Ergebnisse

Tabelle 10 zeigt eine Übersicht über die Forschungsdaten, die bereits zur Nachnutzung bereitgestellt werden konnten und über die Forschungsdaten, die ausschließlich zur Sicherung angenommen wurden.

Tabelle 10: Bereitgestellte Forschungsdaten

	eingegangene Forschungsdaten insgesamt	davon noch in Bearbeitung	Bearbeitung abgeschlossen	davon zur Nachnutzung**	davon zur Sicherung* (keine Nachnutzung möglich)
FDZ Bildung am DIPF	13	2	11	3	8
GESIS Datenarchiv	8	0	8	7	1
FDZ am IQB	25	4	21	13	8
<i>insgesamt</i>	46	6	40	23	17

Basis: Meldung der jeweiligen Einrichtung; Stand 30.09.2016

* Sicherung: Forschungsdaten werden bei dem jeweiligen Archiv langfristig gesichert. Sie werden nicht an Dritte weitergegeben. Eine Nachnutzung ist nicht möglich.

** Nachnutzung oder Sekundärnutzung: Forschungsdaten werden an nicht an der Erhebung beteiligte Forschende weitergegeben.

In 17 von 40 bereits final bearbeiteten Datenbeständen erfolgte nur eine Sicherung der Daten (vgl. Tabelle 10: Spalte 6). Das IQB konnte in vier Fällen Daten aus rechtlichen Gründen nicht für die Nachnutzung verfügbar machen. GESIS war eine Bereitstellung in einem Fall und dem IQB in sechs Fällen nicht möglich, da die Dokumentation nicht ausreichte und der Datensatz nicht nachvollziehbar war. Die an das DIPF zur Sicherung übermittelten Studien konnten überwiegend aus datenschutzrechtlichen Gründen nicht zur Nachnutzung verfügbar gemacht werden.

Tabelle 11: Sperrfristen

	Forschungsdaten zur Nachnutzung (bereits veröffentlicht)	davon Anzahl Daten mit Sperrfristen
FDZ Bildung am DIPF	3	0
GESIS Datenarchiv	7	0
FDZ am IQB	13	8
<i>Insgesamt</i>	<i>23</i>	<i>8</i>

Stand 30.09.2016

Insgesamt wurden acht Datenbestände mit Sperrfristen versehen. Die Sperrfristen laufen (Stand September 2016) im längsten Fall bis November 2017. Ergebnisse zur Verwendung des Bewertungsschemas werden nicht berichtet, da dieses in der 1. Förderphase zunächst erprobt wurde.

3.2.5 Stellungnahmen der Projektnehmer/innen gegenüber dem DLR Projektträger

Zum Abschluss eines jeden Projekts hatte die Projektleitung gegenüber dem DLR Projektträger Bericht zu erstatten. Dazu gehörte auch die Abgabe der „Stellungnahme des Zuwendungsempfängers zur Erfüllung der Nebenbestimmungen im Hinblick auf die im Rahmen des Projektes gewonnenen Daten“. Auf Grundlage dieser Stellungnahme prüfte der DLR Projektträger die Einhaltung der Nebenbestimmung. Die Projektnehmer/innen waren aufgefordert, ihre Bereitschaft dazu zu erklären, die im Projekt gewonnenen Forschungsdaten einer geeigneten Einrichtung zur Sicherung und Nachnutzung zur Verfügung zu stellen. Eine Ablehnung oder Einschränkung dieser Bereitschaft war begründungspflichtig.

Aus den Stellungnahmen ging somit hervor, wie viele Projekte mit welchen Begründungen eine Weitergabe ihrer Daten ablehnten bzw. dieser nur mit Einschränkungen zustimmten. Dem VFDB liegen (Stand 30.09.2016) 91 Stellungnahmen datenerhebender Projekte vor (vgl. Tabelle 12). Aus diesen geht hervor, dass 59 Projekte sowohl der Sicherung als auch der Nachnutzung ihrer Forschungsdaten zustimmten, zwölf Projekte ausschließlich der Sicherung ihrer Forschungsdaten zustimmten und deren Nachnutzung durch Dritte ablehnten, sowie 20 Projekte sowohl die Sicherung ihrer Forschungsdaten als auch deren Nachnutzung ablehnten. In Tabelle 12 sind ebenfalls die Begründungen kategorisiert aufgeführt, die von den Projektnehmer/innen angegeben wurden. Dabei ist zu berücksichtigen, dass innerhalb einer Stellungnahme mehrere Begründungsinhalte aufgeführt werden konnten.

Tabelle 12: Angaben lt. Stellungnahme an den DLR Projektträger

Angabe lt. Stellungnahme	Anzahl	datenschutz- rechtliche und ethische Aspekte*	nicht abgeschlossene Arbeiten*	behördliche Genehmigungen*	urheberrechtliche Einschränkungen*
a. Sicherung und Nachnutzung	59	6	18	3	5
b. Sicherung ohne Nachnutzung	12	7	3	5	2
c. Weder noch	20	17	4	9	0
d. Sonstiges	0	0	0	0	0
Insgesamt	91	30	25	18	7

Basis: Eigene Erfassung und Auswertung; Falls inhaltlich erforderlich, wurden die standardisierten Antworten umcodiert. Die offenen Angaben wurden inhaltlich nach den vier Kategorien codiert; Stand 30.09.2016.

*Mehrfachnennungen bei der Angabe von Gründen möglich.

Als Begründungen für eine vollständige oder teilweise Ablehnung der Datenübermittlung wurden in den Stellungnahmen 1) datenschutzrechtliche und ethische Aspekte, 2) nicht abgeschlossene Arbeiten (Qualifizierungsarbeiten, Analysen, Publikationen), 3) behördliche Auflagen und 4) urheberrechtliche Einschränkungen aufgeführt. In einigen Projekten wurde angegeben, dass die Sicherung der Forschungsdaten in der eigenen Einrichtung erfolgen wird. Die in den Stellungnahmen aufgeführten Begründungen sind unterschiedlich detailliert und reichen von zwei Sätzen bis zu drei Seiten.

3.3. Hürden der Sicherung und Bereitstellung der Forschungsdaten aus dem EBF-Rahmenprogramm

Aus den in den Stellungnahmen dargelegten Begründungen sowie den persönlichen Rückmeldungen der Forscherinnen und Forscher lassen sich Rückschlüsse ziehen, welche Hürden im Hinblick auf die Ziele Sicherung und Nachnutzung von Forschungsdaten existieren und wie diese einzuordnen sind. Diese werden in den folgenden Kapiteln dargestellt.

3.3.1. Fehlende datenschutzrechtliche Voraussetzungen

Die Bundes- und Landesdatenschutzgesetze schreiben vor, dass zur Verarbeitung personenbezogener Daten in der Regel das Einverständnis der betroffenen Personen einzuholen ist, außer es liegen rechtliche Ausnahmetatbestände vor (vgl. hierzu weiterführend RatSWD 2017). Dabei muss es sich um ein sog. informiertes Einverständnis handeln, das heißt, die Betroffenen müssen hinreichend über die Art der Daten, die verarbeitet werden, und die unterschiedlichen Zwecke der Verarbeitung informiert sein. In der Praxis sind hohe Anforderungen an die rechtsgültige Erstellung von Einverständniserklärungen gestellt (vgl. Verbund Forschungsdaten Bildung 2015a). Eine rechtsgültige Einverständniserklärung sollte möglichst konkret und laienverständlich benennen, welche Art von Daten erhoben und wie und durch wen diese verarbeitet werden sollen. Je nach Projekt kann eine unterschiedliche Rechtsquelle zu beachten sein, da der Datenschutz in den Landesdatenschutzgesetzen, im Bundesdatenschutzgesetz oder in Spezialgesetzen wie dem Schulgesetz geregelt ist.

Im Rahmen dieses Projekts wurden uns mehrere Einwilligungserklärungen zur Ansicht vorgelegt. Typische Fälle, die eine Weitergabe der Daten an ein Archiv ausschließen, waren demnach die folgenden:

(1) Information über Weitergabe an Dritte fehlt: In fast allen der uns vorgelegten Einwilligungen fehlte die Information darüber, dass die Daten nach Beendigung des Projektes an ein Archiv weitergegeben werden sollen. Die Betroffenen waren somit nicht vollständig über die Verwendungszwecke ihrer Daten informiert. In solchen Fällen – wenn die Weitergabe nicht explizit ausgeschlossen wird – ist die Weitergabe an ein Archiv nur unter engen Voraussetzungen möglich. Denn der Gesetzgeber hat für bestimmte Ausnahmetatbestände die Möglichkeit geschaffen, Daten trotz fehlender Einwilligung zu verwenden. Das sog. Forschungsprivileg besagt, dass Daten auch ohne Einwilligung für Forschungszwecke verarbeitet werden dürfen, „wenn es zur Durchführung wissenschaftlicher Forschung erforderlich ist, das wissenschaftliche Interesse an der Durchführung des Forschungsvorhabens das Interesse des Betroffenen an dem Ausschluss der Zweckänderung erheblich überwiegt und der Zweck der Forschung auf andere Weise nicht oder nur mit

unverhältnismäßigem Aufwand erreicht werden kann“ (§ 14 BDSG). Das Forschungsprivileg ist somit an enge Voraussetzungen geknüpft, deren Vorliegen im Einzelfall geprüft werden muss.

(2) Eine die Weitergabe der Daten explizit ausschließende Formulierung: In einigen Fällen wurde den Betroffenen in der Einwilligungserklärung zugesichert, dass die erhobenen Forschungsdaten ausschließlich im Rahmen des erhebenden Forschungsprojektes verwendet und nicht an Dritte weitergegeben werden. Kritisch war hier oft, ob in der Einwilligungserklärung die Art der Daten, auf die sich die Formulierung bezog, näher bestimmt worden war. Denn bezieht sich die Formulierung auf „personenbezogene Daten“ oder auf „Daten“ ganz allgemein, ist eine Weitergabe an Dritte zur Archivierung i. d. R nicht möglich. Dies liegt daran, dass von Betroffenen als datenschutzrechtlichen Laien nicht erwartet werden kann, dass diese den Unterschied zwischen personenbezogenen und nicht personenbezogenen (anonymisierten) Daten kennen. Daraus folgt, dass in einem solchen Fall die Weitergabe sowohl der personenbezogenen als auch der anonymisierten Daten durch die vorliegende Einwilligung nicht gedeckt wäre. Erforderlich ist eine Formulierung, die die konkreten Datenkategorien benennt, die weitergegeben und nicht weitergegeben werden. Ein Rückgriff auf gesetzliche Ausnahmetatbestände wie das Forschungsprivileg, das unter bestimmten Voraussetzungen die einwilligungslose Verarbeitung personenbezogener Daten erlauben würde, ist hier nicht möglich. Denn wer eine Einwilligung einholt, gibt zu erkennen, dass er sich an Inhalt und Umfang dieser halten wird.

Festzuhalten ist, dass bislang häufig keine Einwilligung zur Archivierung der Forschungsdaten eingeholt wurde oder die Archivierung sogar explizit ausgeschlossen wurde. Üblicherweise bezog sich die Einwilligung ausschließlich auf die Datenverarbeitung innerhalb des Projektes. Das Vorliegen rechtsgültiger und die Archivierung der Forschungsdaten beinhaltender Einwilligungen ist jedoch eine Grundvoraussetzung für die Archivierbarkeit von Forschungsdaten. *Ex post* Datenerhebung können solche Datenbestände nicht bzw. nur unter hohen Kosten (nachträgliche, nochmalige Ansprache von Studienteilnehmer/innen; Rechtsexpertisen) archiviert werden. Vor diesem Hintergrund ist ein hoher Beratungsbedarf im Hinblick auf Einwilligungen zu konstatieren. Dabei sind hohe Anforderungen an die Beratung zur Erstellung einer Einwilligungserklärung gestellt. Beratende Personen sollten datenschutzrechtliche Kenntnisse, Kenntnisse der Arbeitsprozesse der empirischen Sozialforschung und der Abläufe in empirischen Forschungsprojekten sowie Kenntnisse der Methoden der empirischen Sozialforschung im Hinblick auf Teilnahmebereitschaft und Motivation von Beforschten aufweisen (vgl. bspw. Homuth et al. 2017). Die bisherige Beratungssituation ist unseres Erachtens unzureichend. So traten die beschriebenen Probleme auch in solchen Fällen auf, in denen Einwilligungserklärungen mit den universitären Justitiariaten abgestimmt waren. Auch zeigte sich in manchen Fällen, dass es zu divergierenden Einschätzungen der beteiligten Juristen kam.

Aufgrund der hohen Anforderungen an das Erstellen rechtsgültiger Einwilligungserklärungen und der derzeitigen Situation hat der VFDB eine Checkliste mit Hinweisen zur Erstellung rechtsgültiger Einwilligungserklärungen in Zusammenarbeit mit der Rechtsanwaltskanzlei Goebel & Scheller erstellt (vgl. Verbund Forschungsdaten Bildung 2015a). Diese Checkliste ergänzt bereits verfügbare Mustereinwilligungen wie bspw. die des RatSWD (Liebig et al. 2014). Einwilligungen sind stets vor dem Hintergrund des jeweiligen Projektes zu beurteilen, Mustereinwilligungen sind daher nur begrenzt über unterschiedliche Projekte hinweg einsetzbar.

Ein weiterer Baustein zur Einhaltung datenschutzrechtlicher Anforderungen ist die *Anonymisierung* der Forschungsdaten. Dies setzt voraus, dass in der Einwilligung nicht zugesichert wurde, dass auch die anonymisierten Daten nicht weitergegeben werden (vgl. voranstehend). Aus datenschutzrechtlicher Sicht ist eine faktische Anonymisierung der Daten ausreichend. Forschungsdaten gelten als faktisch anonym, wenn der Aufwand zur Identifizierung „unverhältnismäßig“ groß ist. Zu den Aufwänden bzw. Kosten eines potentiellen Angreifers zählt dabei nicht nur der direkte Aufwand, der erforderlich ist, einen bestimmten Betroffenen zu identifizieren. Dieser entsteht beispielsweise durch die erforderlichen Recherchen im Internet oder in sonstigen Quellen. Zusätzlich trägt der Angreifer das Risiko, dass das eigene Fehlverhalten entdeckt und geahndet wird sowie das Risiko, dass seine Versuche zur Identifizierung einer bestimmten Person misslingen. Ein Angreifer handelt schon durch die Versuche der Identifikation vertragswidrig, wenn er sich im Rahmen einer Nutzungsvereinbarung mit der datenbereitstellenden Einrichtung zur Unterlassung von Identifizierungsversuchen verpflichtet hat. Für diesen Fall sehen Nutzungsverträge von Datenzentren typischerweise Konsequenzen beispielsweise in Form von Geldzahlungen oder einem Ausschluss von der zukünftigen Nutzung einer Einrichtung vor. Identifizierungsrisiken sind auch von der Sensibilität der Daten abhängig und ob den Betroffenen durch Re-Identifizierung ein Schaden entstehen könnte sowie davon, ob es sich um kommerziell verwertbare Daten handelt.

Bildungsforschungsdaten weisen häufig eine besondere Sensibilität auf: so bspw. Daten aus Kompetenzmessungen von Schülerinnen, Schülern oder Lehrkräften, Daten zur sozialen Herkunft und familiären Situation, zum Freizeitverhalten oder zu Einstellungen und Werten. Zudem stehen Schüler/in und Schule sowie Lehrkraft und Schule in besonderen Abhängigkeitsverhältnissen zueinander, die besonderen Vertrauensschutz erfordern. Auch ist das De-Anonymisierungsrisiko bei Mehrebenen- oder verknüpften Daten, wie sie Daten zu Lehrer/innen, Schüler/innen, Eltern und Schulen darstellen, vergleichsweise hoch. Das Arbeitgeber-Arbeitnehmerverhältnis zwischen Lehrkraft und Schule ist ebenfalls besonders schutzbedürftig.

Ein dritter Baustein des Schutzes der Betroffenen - neben Einwilligung und Anonymisierung der Daten – stellen *Zugangs- und Zugriffsbeschränkungen* auf die Daten dar. Forschungsdatenzentren prüfen im Rahmen der Aufbereitung der Daten deren Identifikationsrisiko und gestalten dementsprechend die Zugangsbedingungen zu den betreffenden Daten. Kontrolle und Reglementierung des Zugriffs erhöhen den Aufwand für potentielle Angreifer zur Identifikation von Betroffenen möglichst auf ein unverhältnismäßig hohes Niveau und tragen hierdurch zur faktischen Anonymität der Daten bei (vgl. auch Smolle 2015). Daraus folgt im Umkehrschluss, dass das Vorliegen personenbezogener oder schwach anonymisierter Daten eine Archivierung von Forschungsdaten nicht per se ausschließt, sondern der Schutz der Betroffenen, der durch Zugangs- und Zugriffsbeschränkungen erreicht wird, ebenfalls zu berücksichtigen ist. Dennoch scheint es, dass die Anonymität von Daten in der Praxis aufgrund der steigenden Online-Verfügbarkeit von Informationen und dem technologischen Fortschritt zunehmend schwieriger zu bestimmen ist. Daher steigt die Bedeutung zusätzlicher Schutzmechanismen. Datenzentren haben sowohl ein hohes Eigeninteresse am Schutz der Betroffenen als auch die erforderlichen Kompetenzen und Ressourcen (z. B. hinsichtlich Datensicherheit), diesen umzusetzen. In der Praxis gehen die Schutzmaßnahmen von Archiven häufig über die gesetzlichen Vorschriften hinaus, bspw. dann, wenn auch anonymisiertes Material nur zugriffsgeschützt zugänglich ist. Datenzentren sind nicht nur datenschutzrechtlich sondern auch ethisch gegenüber den Betroffenen *und* den Datengebern verpflichtet. Der Vertrauensschutz spielt eine besondere Rolle.

Es bleibt festzuhalten, dass eine Archivierung von Forschungsdaten ohne Einwilligung oder mit unzureichender Anonymisierung stets mit hohen Aufwänden, rechtlichen Unsicherheiten und Restrisiken verbunden ist. Die Beurteilung, ob das Forschungsprivileg zum Tragen kommt, die Beurteilung der Gültigkeit der vorliegenden Einwilligung ebenso wie die Beurteilung der Anonymität sind häufig erst nach einer detaillierten Prüfung des Einzelfalls möglich. Hierzu müssen die Originalunterlagen (u. a. die Einwilligungserklärungen) und die Forschungsdaten selbst hinzugezogen werden. Letztlich obliegen die Entscheidung, ob Daten zur Archivierung weitergegeben werden oder nicht, sowie das Risiko einer Fehlentscheidung, dem Datengeber bzw. der verantwortlichen Stelle im Sinne des Datenschutzes. Maßnahmen sollten daher darauf zielen, dass *ex ante* Datenerhebung den Forschenden die Bedingungen der Archivierung bekannt sind und beim Einholen der Einwilligungen sowie bei der Durchführung von Anonymisierungsmaßnahmen berücksichtigt werden.

Zwei Publikationen, in denen detailliert auf die datenschutzrechtlichen Besonderheiten bei der Archivierung der Daten aus dem EBF-Rahmenprogramm, sind in derzeit in Vorbereitung (vgl. Anhang 7.6). Der VFDB wird sein Beratungsangebot in den nächsten Jahren ausbauen und weitere Informationsmaterialien erarbeiten.

3.3.2. Ausschließende behördliche Genehmigungen

Erhebungen, die an Schulen durchgeführt werden, bedürfen in der Regel einer Genehmigung der Schulaufsichtsbehörde oder des Kultusministeriums. Die gesetzlichen Regelungen, die für Schulbefragungen relevant sind, finden sich in den jeweiligen Landesdatenschutz- und Schulgesetzen. Teilweise basieren die Bestimmungen auf amtlichen Verwaltungsvorschriften und Erlassen – beispielsweise des Kultusministeriums – oder werden in den Schulordnungen der einzelnen Schularten spezifiziert (vgl. Forschungsdaten Bildung o.J.). Aus Sicht der Nachhaltigkeit von Forschungsdaten ist es problematisch, wenn diese Genehmigungen eine Archivierung und Nachnutzung der Forschungsdaten ausschließen. Dem VFDB wurden mehrere Fälle zur Ansicht vorgelegt, die mit den beratenden Rechtsexperten besprochen werden konnten. Die Genehmigungen in diesen Fällen enthielten die Formulierung, dass eine Weitergabe der Daten an Dritte ausgeschlossen ist. Wenn sich die Vorgabe explizit nur auf die personenbezogenen Daten bezieht, könnte eine Weitergabe der anonymisierten Daten möglich sein. Generell ist festzuhalten, dass es einer Einzelfallprüfung unter Vorlage sämtlicher Unterlagen bedarf, um beurteilen zu können, ob eine erteilte Genehmigung die Archivierung der Forschungsdaten ausschließt oder nicht. Bei Unklarheiten kann auch der erneute Kontakt mit der genehmigenden Behörde gesucht werden.

Gegenwärtig stehen Forscher/innen widersprüchlichen Anforderungen gegenüber: zum einen der Anforderung der Geldgeber hinsichtlich der Archivierung und Nachnutzung der Forschungsdaten und zum anderen den Anforderungen, die sich aus den behördlichen Auflagen für Erhebungen an Schulen ergeben. Zu vermuten ist, dass in der Regel die Anforderungen einer späteren Archivierung durch die Projektnehmer/innen im Rahmen der Beantragung sowie die genehmigenden Stellen schlicht nicht berücksichtigt wurden und diese daher kein Bestandteil der erteilten Genehmigung waren. Es liegt uns ein einziger Fall vor, in dem die Beantragung die Archivierung der Forschungsdaten beinhaltet und die Genehmigung diese nicht ausschließt. Wie die genehmigenden Stellen tatsächlich zur Archivierung der Forschungsdaten stehen, ist noch ungeklärt. Deren Interessen könnten jedoch auch im Rahmen einer Archivierung der Daten gewahrt werden. Aus unserer Sicht besteht ein erheblicher Informationsbedarf sowohl auf Seiten der Forschenden als auch auf Seiten der genehmigenden Stellen. Das Ziel sollte es sein, dass behördliche Genehmigungen generell die Weitergabe der

Forschungsdaten an ein Forschungsdatenzentrum ermöglichen. Um dieses Ziel zu erreichen, erscheint eine Abstimmung mit den Behörden dringend erforderlich. Auch sollten die Forschenden im Rahmen des Genehmigungsverfahrens eines konkreten Projektes auf die durch die Mittelgeber geforderte Archivierung der Forschungsdaten deutlich hinweisen. In der nächsten Förderphase strebt der VFDB daher an, die relevanten Gremien, wie beispielsweise die Kultusministerkonferenz und genehmigende Behörden auf die Thematik aufmerksam zu machen und Lösungsvorschläge einzubringen, so dass mittelfristig die Rahmenbedingungen zur Archivierung von Forschungsdaten verbessert werden können.

3.3.3. Fehlende urheberrechtliche Erlaubnis

Ablehnungen der Datenweitergabe aufgrund von urheberrechtlichen Aspekten sind ebenfalls vorgetragen worden, allerdings in weit geringerem Ausmaß verglichen mit datenschutzrechtlichen Aspekten. In den 92 Stellungnahmen (vgl. Kapitel 3.2.5) wurde dies in sechs Fällen angegeben. Dabei handelte es sich in allen Fällen um das Urheberrecht an in den Projekten entwickelten Testinstrumenten.

Für die Archivierung von Forschungsdaten und Instrumenten ist es erforderlich, dass der oder die Urheber/in dem Archiv Nutzungsrechte einräumt. Urheberrechte sind üblicherweise dann betroffen, wenn es sich bei dem vorliegenden Material um ein „Werk“ mit entsprechender Schöpfungshöhe im urheberrechtlichen Sinn handelt. Da die Beurteilung des Werkscharakters und der Schöpfungshöhe von Forschungsdaten nicht trivial ist und im Einzelfall zu prüfen wäre, gehen Datenzentren regelmäßig davon aus, Datengeber als Urheber zu behandeln und sich entsprechende Nutzungsrechte vertraglich einräumen zu lassen. Die Nutzung der Daten durch das Archiv kann dann im Rahmen der Nutzungsrechteinräumung erfolgen. Auch kann es sinnvoll sein, Daten oder Instrumente mit bekannten Creative Commons-Lizenzen zu versehen.

In der Psychologie kommt es häufig vor, dass (sog. ausschließliche) Nutzungsrechte für im Rahmen von Forschungsprojekten entwickelte Testinstrumente an einen Verlag übertragen werden (bspw. Hogrefe), der diese dann kommerziell verwertet. Dies schließt jedoch eine Archivierung der Forschungsdaten nicht per se aus. Denn es besteht die Möglichkeit, die Forschungsdaten *exklusive* der Test-Items zu veröffentlichen.⁹ In diesem Fall müssten Sekundärnutzende den entsprechenden Test beim Verlag käuflich erwerben, um die Forschungsdaten vollständig interpretieren und nutzen zu können. Anzumerken ist an dieser Stelle, dass eine kommerzielle Verwertung von Forschungsdaten und Instrumenten allgemein anerkannten Leitlinien (vgl. z. B. OECD 2007; Allianz der deutschen Wissenschaftsorganisationen 2014) widerspricht.

Der VFDB unterstützt entsprechend des „Appells zur Nutzung offener Lizenzen in der Wissenschaft“ der Allianz der deutschen Wissenschaftsorganisationen die Verwendung der Lizenz Creative Commons Namensnennung CC-by. Diese ermöglicht die Weiterverwendung der Daten unter Nennung des Urhebers („by“).¹⁰

⁹ Der VFDB hat Hinweise zur Dokumentation von Daten, die mit urheberrechtlich geschützten Instrumenten entstanden sind, erstellt: www.forschungsdaten-bildung.de/doku-instrumente (Zugegriffen: 30. Sep. 2016).

¹⁰ Weitere Informationen hier: <https://de.creativecommons.org/index.php/was-ist-cc/> (Zugegriffen: 09. Mai 2017).

3.3.4. Unzureichende Ressourcenausstattung

Projektnehmer/innen äußerten in bilateralen Gesprächen Bedenken hinsichtlich des Aufwands der Datenaufbereitung und -dokumentation. In den Projekten selbst seien nicht in ausreichendem Maß Kapazitäten vorhanden, um die anstehenden Arbeiten im Zuge der Datenarchivierung durchzuführen. Teilweise war dies darin begründet, dass die während der Projektlaufzeit für das Datenmanagement zuständigen Mitarbeiter/innen nach Projektende nicht mehr verfügbar oder mit neuen Aufgaben betraut waren.

Aufwände entstanden den Projektnehmer/innen

- durch die Kommunikation mit dem VFDB im Rahmen der Archivierung von Daten,
- bei der Online-Meldung,
- durch den Vertragsabschluss,
- für das Zusammenstellen des Datenpakets und
- für den Upload der Forschungsdaten.

Zudem entstanden Aufwände in denjenigen Fällen, in denen die Daten und Dateien vor ihrer Übermittlung noch bearbeitet (anonymisiert, dokumentiert oder aufbereitet) werden mussten, um nachvollziehbar zu sein. Bei quantitativen Daten geht es beispielsweise um die eindeutige Bezeichnung von Variablen und Werten in den Datensätzen, so dass Fragebögen eindeutig den Variablen im Datensatz zuordenbar sind, um die Beschreibung von generierten Variablen (für die es keine Entsprechung im Fragebogen gibt) oder um das Erstellen von Skalenhandbüchern, Codebüchern und Methodenberichten, in denen Zusatzinformationen zur Datenerhebung dokumentiert sind. Bei qualitativen Daten geht es beispielsweise um die Verwendung einheitlicher Layouts für Transkripte oder Beobachtungsprotokolle, um das Dokumentieren von Transkriptionssystemen oder um die Beschreibung der Vorgehensweisen bei der Kodierung.

In persönlichen Gesprächen mit Projektmitarbeiter/innen zeigte sich allerdings auch, dass die Aufwände, die mit der Datendokumentation und -aufbereitung verbunden sind, teilweise überschätzt wurden. Wenn Forschende bislang selbst noch keine Erfahrung mit der sekundäranalytischen Nutzung von Datensätzen gemacht haben, ist es möglicherweise schwierig zu beurteilen, welche Informationen für Sekundärnutzer/innen erforderlich sind und wie die Sekundärnutzung der eigenen Daten aussehen würde.

Der VFDB ist der Problematik unzureichender Ressourcenausstattung bei den Projekten auf mehreren Ebenen begegnet. Die noch laufenden Projekte wurden zu Projektbeginn des VFDB über das neue Angebot informiert. Der Zeitpunkt der Datenübermittlung wird mit den Projektnehmer/innen bilateral abgestimmt und kann daher von diesen in den üblichen Projektverlauf eingepasst werden. Schließlich ist der VFDB bestrebt, die den Projektnehmer/innen entstehenden Aufwände für Datendokumentation und -aufbereitung im Vorfeld der Datenarchivierung gering zu halten. Informationsblätter und die Website forschungsdaten-bildung.de informieren Forschende darüber, in welcher Form Forschungsdaten an den VFDB übermittelt werden sollten¹¹. Diese haben überwiegend Empfehlungscharakter und stellen keine verpflichtenden Vorgaben dar, um den unterschiedlichen Situationen in den Forschungsprojekten Rechnung tragen zu können. Die Prozesse des VFDB werden auch in der zweiten Förderphase kontinuierlich weiterentwickelt mit dem Ziel, die

¹¹ Eine Übersicht über die Informationsangebote des VFDB findet sich in Kapitel 4.2.

Aufwände für die Projektnehmer/innen weiter zu verringern. Gleichzeitig ist an dieser Stelle anzumerken, dass auch der VFDB und die kooperierenden Forschungsdatenzentren über begrenzte Ressourcen verfügen. Die Möglichkeiten bilateraler Absprachen und von auf einzelne Projekte zugeschnittenen Lösungen sind daher begrenzt. Abweichungen von den Standardverfahren bedeuten immer einen erheblichen Mehraufwand auf Seiten des VFDB. Eine in diesem Zusammenhang zu nennende, prospektiv wirkende Maßnahme des VFDB besteht darin, Forschende gezielt über die Möglichkeit zu informieren, bei Neuanträgen an das BMBF die erforderlichen Mittel für das Datenmanagement einzuwerben. Erste Anfragen dazu, welche Ressourcen hierfür zu kalkulieren sind, sind bereits eingegangen.

Die Ressourcenknappheit von datenerhebenden Projekten ist ein allgemeines, strukturell bedingtes Problem in den aktuellen wissenschaftlichen Projektstrukturen. Die Projekte des EBF-Rahmenprogramms haben Laufzeiten von in der Regel drei Jahren. Hinsichtlich der Laufzeit unterscheiden sich die Projekte, in denen ausschließlich Sekundäranalysen durchgeführt werden, nicht von denjenigen Projekten, in denen Daten generiert werden. Die Arbeitsbelastung in Primärforschungsprojekten ist jedoch entsprechend höher, da zusätzlich zu den Aufgaben der Analyse und Publikation die Aufgaben im Zusammenhang mit der Datenerhebung zu bewältigen sind. Gleichzeitig sind die methodischen Anforderungen an Erhebung und Analyse (bspw. durch Weiterentwicklungen statistischer Analyseverfahren) in den letzten Jahren gestiegen. Letztlich ist anzustreben, dass in den Projekten angewandtes systematisches Forschungsdatenmanagement den Einsatz von Ressourcen optimiert und die Qualität und Nachnutzbarkeit der produzierten Daten erhöht.

3.3.5. Einstellungen von Forschenden gegenüber Data Sharing

Auf Seiten der Forschenden werden teilweise erhebliche Bedenken gegenüber der Archivierung der eigenen Forschungsdaten geäußert. Die Erfahrungen des VFDB decken sich hier mit den auch in anderen Projekten in der Literatur berichteten Erfahrungen (vgl. bspw. Borgman 2010; Haas und Park 2010; Häussler 2010; RIN 2008).

Kasten 2

Beispiele für von Forschenden geäußerte Bedenken in der bilateralen Kommunikation

- Bleibt die Anonymität der Beforschten trotz Datenweitergabe gewahrt?
- Wie und durch wen werden die Daten nachgenutzt? Werden sie frei ins Netz gestellt?
- Was machen die Personen mit den Forschungsdaten? Was ist, wenn diese falsch interpretiert werden?
- Wie aufwendig ist die Vorbereitung der Forschungsdaten für die Archivierung? Welche Aufwände kommen auf mich zu?
- Sind meine Daten überhaupt interessant für andere?

Ein Teil der Bedenken (vgl. Kasten 2) bezieht sich auf den mit der Herausgabe der Daten verbundenen Kontrollverlust über die eigenen Daten: Die Forschenden sind unsicher, ob trotz Datenweitergabe die Anonymität der Befragten zu jeder Zeit gewahrt bleibt und wie, durch wen und zu welchen Zwecken die „eigenen“ Daten nachgenutzt werden.

Die Befürchtungen hinsichtlich der Wahrung der Anonymität der Studienteilnehmer/innen sind teilweise erheblich (vgl. Kapitel 3.3.1). Beispielsweise wird befürchtet, dass im Rahmen der Archivierung beim VFDB Forschungsdaten für jeden frei zugänglich online gestellt werden. Hinzu

kommt, dass – wie oben beschrieben – das Identifizierungspotential teilweise überschätzt wird. Diese Bedenken werden verstärkt von qualitativ forschenden Wissenschaftler/innen geäußert.

Die Feldarbeit qualitativ forschender Wissenschaftler/innen beruht häufig auf einer persönlichen Vertrauensbeziehung zwischen Forscher/in und Beforschtem/r. Daher sehen sich diese in einer persönlichen Verpflichtung, die Beforschten vor Re-Identifikation zu schützen und dafür Sorge zu tragen, dass die Forschungsdaten „richtig“ verwendet und interpretiert werden.

Auch quantitativ Forschende fürchten einen Kontrollverlust hinsichtlich der Verwendung ihrer Daten durch Dritte. Zur Nachvollziehbarkeit von Datenanalysen reicht das Vorliegen der Forschungsdaten häufig nicht aus, sondern zusätzlich sind Aufbereitungs- und Analysesyntax erforderlich. Forschende könnten befürchten, dass sich die eigenen Forschungsbefunde nicht ohne Zusatzinformationen wie die verwendete Syntax und eine ausreichende Dokumentation dieser durch Dritte replizieren lassen, und die Befunde der Primärforschenden im Nachhinein angezweifelt werden könnten. Dem könnte entgegengewirkt werden, indem die Forschenden die Syntax ebenfalls archivieren lassen und zumindest zu Replikationszwecken herausgeben. In einigen Fällen wurde von Datengebern Aufbereitungssyntax übermittelt. Die Übermittlung von Analysesyntax ist bislang noch nicht üblich. Die Bedenken, die hier unter dem Begriff „Kontrollverlust“ subsumiert werden, gehen somit über rein datenschutzrechtlich-ethische Verpflichtungen gegenüber den Beforschten hinaus und beziehen sich auch auf die Sorge um die eigene wissenschaftliche Reputation. Diese könnte Schaden nehmen, wenn die eigenen Forschungsbefunde nicht replizierbar sind. Gleichzeitig handelt es sich bei Syntaxen ebenso wie bei den Erhebungsinstrumenten sowie beim Forschungsdesign selbst, um kodifiziertes Expertenwissen, das im wissenschaftlichen Wettbewerb (um Publikationen, Stellen und Drittmittel etc.) relevant ist.

Im Übrigen werden sowohl von qualitativen als auch von quantitativen Forscher/innen Befürchtungen hinsichtlich der zusätzlichen Aufwände, die mit der Datenarchivierung verbunden sind, geäußert (wie in Kapitel 3.3.4 beschrieben). Auch besteht bei den Forschenden Unsicherheit darüber, ob die eigenen Daten überhaupt ein ausreichendes Nachnutzungspotential aufweisen, d. h., für Dritte interessant sind. Weitere Befürchtungen von Forschenden bestehen in der Sorge um den Verlust der eigenen Investitionen.

Eine der Ursachen für die genannten Bedenken scheint in fehlenden oder unzureichenden Informationen über die Konditionen der Archivierung und die Prozesse und Strukturen von Datenzentren zu liegen. Forschenden ist teilweise unbekannt, welche Möglichkeiten und Maßnahmen des Schutzes von Datengeber/innen und Betroffenen durch Datenzentren existieren. Bei einem Großteil der Bildungsforschenden besteht noch wenig Erfahrung mit der Archivierung eigener Daten über Datenzentren (auch jenseits des VFDB). Zugleich existieren die Angebote des VFDB und anderer Datenzentren erst seit wenigen Jahren. Auch fehlen möglicherweise eigene Erfahrungen mit der Sekundärnutzung von Daten.

Ein wichtiger Bestandteil der Arbeit des VFDB in den Jahren der Pilotphase bestand daher darin, Informationen über die Arbeit von und die Prozesse in Datenzentren zu distribuieren und die Forschenden bei der Durchführung dieses Verfahrens auch in bilateralen Gesprächen zu unterstützen. Ein zentraler Aspekt besteht und wird in Zukunft darin bestehen, in den *scientific communities* Vertrauen aufzubauen und als VFDB ein verlässlicher und kompetenter Partner für die Archivierung von Forschungsdaten zu sein. Ein weiteres Ziel des VFDB muss es daher sein, verstärkt

über die sekundäranalytische Nutzung von Forschungsdaten zu informieren und diese in den Disziplinen der Bildungsforschung zu fördern. Je mehr praktische, eigene Erfahrungen im Forschungsfeld mit der Nachnutzung von Forschungsdaten bestehen, desto eher können Befürchtungen in dieser Hinsicht ausgeräumt werden. Die bisherigen Erfahrungen des VFDB lassen vermuten, dass Forschende, die in der Vergangenheit auf Sekundärdaten zurückgegriffen haben und auf diese angewiesen waren, eher bereit sind, eigene Daten zu teilen. Dass in vielen Disziplinen der Bildungsforschung bislang nur geringe Erfahrungen mit der sekundäranalytischen Nutzung von Daten bestehen, erschwert daher die Aufgabe des VFDB.

3.3.6. Sonstige Faktoren

Neben den voranstehend genannten Faktoren gab es eine Reihe prozessbedingter, insbesondere zeitlicher Hindernisse der Datenarchivierung, die an uns herangetragen wurden.

Qualifizierungsarbeiten, Promotionen und Habilitationen: Die häufigste zeitliche Hürde der Forschungsdatenarchivierung, die genannt wurde, bestand in noch ausstehenden Qualifizierungsarbeiten (wie Promotionen oder Habilitationen) und Publikationen. Ebenso wie das Memorandum des Fachkollegiums Erziehungswissenschaften (vgl. Stanat 2014) gehen wir davon aus, dass die Anfangsinvestitionen der datenerhebenden Forscher/innen zu schützen sind. Die Fertigstellung von Qualifizierungsarbeiten und Publikationen sollte nicht gefährdet werden. Forschende wünschen sich in diesen Fällen häufig einen Aufschub der Datenarchivierung. Sperrfristwünsche sind an uns herangetragen worden (vgl. Tabelle 11).

Aus archivarischer Sicht sind mehrjährige Fristen zu vermeiden, da sie das Risiko bergen, dass mit dem Aktualitätsverlust eine Verminderung des Analysepotentials der betreffenden Daten einhergeht. Für die Forschenden ist es von Vorteil die Datenweitergabe bereits frühzeitig durchzuführen, damit der DOI-registrierte Datenbestand durch die in den Projekten entstandenen Qualifikationsarbeiten zitiert werden kann. Alternativen zur Einrichtung von Sperrfristen, die ebenfalls die Anfangsinvestitionen der Datengeber schützen und dennoch die zeitnahe Archivierung der Daten ermöglichen, existieren. Beispielsweise können konkrete Forschungsfragen oder Teile des Datenbestands (bestimmte Variablen), die von den Primärforscher/innen bearbeitet werden, für die Sekundärnutzung gesperrt werden. In diesem Fall können die Forschungsdaten zeitnah zugänglich gemacht werden, ohne jedoch die Arbeiten der Primärforschenden zu beeinträchtigen. Bislang wurde diese Option im Rahmen der im VFDB bearbeiteten Studien noch nicht genutzt.

Nochmaliger Einsatz von Testinstrumenten, die bis dahin nicht anderweitig eingesetzt werden dürfen: Sollen Testinstrumente nochmals im Feld eingesetzt werden, ist es wichtig, dass diese den späteren Studienteilnehmer/innen nicht zuvor bekannt sind. Eine Veröffentlichung der Instrumente kann somit nur unter restriktiven Zugangsbedingungen erfolgen. In solchen Fällen bedarf es entsprechender Restriktionen des Zugangs zu den Daten und Instrumenten. D. h. Zugang nur für wissenschaftliche Zwecke nach Stellen eines Nutzungsantrags, auch dürfen die Instrumente nicht frei zum Download verfügbar sein. Im Nutzungsantrag ist zudem festzuhalten, dass die Instrumente in einem bestimmten Zeitraum nicht für eigene Erhebungen verwendet werden dürfen.

Ende der Projektlaufzeit / Priorisierung der Datengeber/innen und Fortgang der Mitarbeiter/innen: Da das EBF-Rahmenprogramm seit November 2007 läuft, war die Mehrheit der Datenerhebungen zu Beginn des VFDB (im Oktober 2013) bereits abgeschlossen und viele Projekte bereits ausgelaufen. Dies beeinträchtigte die Möglichkeiten der Archivierung der im EBF-Rahmenprogramm generierten

Forschungsdaten erheblich. Nach Ablauf eines Forschungsprojektes waren in der Regel keine für die Abwicklung der Forschungsdatenarchivierung ausreichenden Kapazitäten vorhanden. Tätigkeiten im Zusammenhang mit ausgelaufenen Projekten haben im Vergleich zu gegenwärtig laufenden Projekten zudem in der Regel keine hohe Priorität mehr. Neben diesen Kapazitäts- und Prioritätsaspekten ist häufig schlicht das notwendige Knowhow im Zusammenhang mit den generierten Forschungsdaten nicht mehr verfügbar, da einst verantwortliche Mitarbeiter/innen nicht mehr im Projekt tätig sind. Auch die Pensionierung oder der Wechsel der Institutszugehörigkeit erschwerte oder verhinderte in einzelnen Fällen eine Archivierung der erhobenen Forschungsdaten (vgl. Tabelle 3).

Kommunikation und Abstimmung in Projektverbänden

In Projektverbänden, die aus mehreren Teilprojekten bestehen, die gemeinsam einen Datenbestand generiert haben, war eine Abstimmung der Projektleitungen untereinander erforderlich. Die Projektleitungen mussten Einigkeit darüber erzielen, ob und wie die Daten gesichert werden sollen und wer hierfür verantwortlich sein wird. Außerdem musste Einigkeit erzielt werden sowohl hinsichtlich der Inhalte des zu übermittelnden Datenpakets als auch hinsichtlich der Qualität bzw. Ausgestaltung desselben. In vielen Fällen führten die erforderlichen Abstimmungsprozesse zu erheblichen Zeitverzögerungen. In einzelnen Fällen wurde dem VFDB bekannt, dass keine Einigung über den zu übermittelnden Datenbestand (Qualität, Ausgestaltung) erzielt werden konnte, und dies die Archivierung der betreffenden Daten verhinderte.

4. Forschungsdatenmanagement: Information, Beratung und Schulung **(Teilziel 3)**

Die Angebote für Datengeber/innen und Datennutzer/innen unter dem Dach des VFDB wurden begleitet von Informations-, Beratungs- und Schulungsaktivitäten zum Thema Datenmanagement. Unter Forschungsdatenmanagement (FDM) versteht man den systematischen und planvollen Umgang mit Forschungsdaten über den gesamten Lebenszyklus der Daten hinweg (vgl. z. B. Jensen 2012; Meyermann 2012; Verbund Forschungsdaten Bildung 2015b). Vor dem Hintergrund der Anforderung an die Archivierung und Verfügbarmachung von Forschungsdaten sind in den letzten Jahren auch Anforderungen an das Datenmanagement von Forschungsdaten in den datengenerierenden Projekten gestiegen. Datenmanagementpläne oder zumindest Aussagen zum Datenmanagement werden bei der Antragstellung zunehmend von Förderern eingefordert. Dies resultiert daher, dass sich die *Nachnutzbarkeit* von Forschungsdaten zum großen Teil nicht durch die Arbeit von Forschungsdatenzentren *ex post* Datenproduktion herstellen lässt, sondern im Aufgaben- und Verantwortungsbereich der datenerhebenden Forscher/innen liegt. Daten sollten so dokumentiert und aufbereitet sein, dass sie für Dritte, nicht an der Erhebung Beteiligte, ggf. auch nach mehreren Jahren, interpretierbar sind. Weiter ist die Nachnutzung personenbezogener und -beziehbarer Daten über ein FDZ aus datenschutzrechtlicher Sicht ein Nutzungszweck, für den in der Regel das Einverständnis der Betroffenen einzuholen ist (vgl. z. B. RatSWD 2017). Aufgrund der gestiegenen Anforderungen und da FDM bisher kein regulärer Bestandteil in der universitären Ausbildung ist, bot der VFDB Informationen, Beratungen und Schulungen zum Thema an, die es Forschenden ermöglichten, sich umfassend und frühzeitig im Projektverlauf zum Thema Forschungsdatenmanagement zu informieren und zu qualifizieren.

Der VFDB wurde in der Distribution von Wissen zum Forschungsdatenmanagement aktiv durch

- (1) die Durchführung von Workshops in unterschiedlichen Formaten,
- (2) das Erstellen von Informationsmaterialien sowie
- (3) das Angebot von Beratungsleistungen.

Tabelle 13 zeigt die Aufgabenverteilung im VFDB im Arbeitsbereich Forschungsdatenmanagement.

Tabelle 13: Aufgabenverteilung des Bereichs Datenmanagement im VFDB

Aufgabe im VFDB	Verantwortliche Einrichtung
Schulungen	GESIS
Informationsmaterialien	Allgemeiner Art: DIPF Besonderer Art: DIPF, GESIS, IQB je Datentyp
Beratung	Allgemeiner Art: DIPF Besonderer Art: DIPF, GESIS, IQB je Datentyp

4.1. Schulungen

Ein Schwerpunkt der ersten Förderphase des VFDB bestand darin, ein Schulungsprogramm zum Forschungsdatenmanagement für die Bildungsforschung zu erstellen. Ziel der Workshops war die Vermittlung von Kenntnissen und Fähigkeiten zu systematischem Forschungsdatenmanagement als Voraussetzung für nachvollziehbare und den Kriterien guten wissenschaftlichen Arbeitens entsprechende Forschung, die eine Archivierung und Nachnutzung der generierten Forschungsdaten ermöglicht. Die zweitägigen Workshops bestanden aus sechs Sessions zu den Themen

- Forschungsdatenmanagement,
- Data Sharing,
- Lizenzierung,
- Metadaten und Dokumentation,
- Datenadministration und -speicherung sowie
- Ethikfragen und Datenschutz.

Die Workshopinhalte sind beschrieben in Trixa und Ebel 2015.

Den Teilnehmenden wurde in den Workshops ein grundlegendes Verständnis für die Thematik vermittelt. Hierzu zählten u. a. die nachhaltige Dokumentation ebenso wie die (langfristige) Sicherung und Archivierung von Forschungsdaten, die Beachtung forschungsethischer und rechtlicher Standards im Rahmen des Datenschutzes, der informierten Einwilligung der Befragten und des Urheberrechts sowie die selbstständige Erstellung eines Datenmanagementplan. Insgesamt wurden in den Jahren 2015 und 2016 vier zweitägige Workshops bei der GESIS in Köln für jeweils ca. 15 Teilnehmer/innen durchgeführt (Mai und November 2015, Mai und Juni 2016). Die Teilnahme war jeweils kostenlos, da die Entwicklung des Trainingskonzepts durch das BMBF gefördert wurde. Für den Workshop im November 2015 wurde zudem ein Rechtsanwalt zum Thema Datenschutz als ein externer Referent eingeladen.

Es gab eine hohe Nachfrage, so dass die Workshops frühzeitig ausgebucht waren.¹² Dies unterstreicht die Relevanz des Themas und den Bedarf. Insgesamt bewerteten die Teilnehmenden die Veranstaltungen durchweg positiv und zielführend für ihre Forschungsprojekte. Neben den Vorträgen und den vermittelten Inhalten lobten sie vor allem die Möglichkeiten, konkrete Probleme

¹² Die im Folgenden angegebene Bewertung ebenso wie die Zitate beziehen sich auf die kumulierten Ergebnisse einer Kurzevaluation am Ende der jeweiligen Workshops der Jahre 2015.

offen diskutieren und Anregungen für neue Lösungsansätze gewinnen zu können. Unter den Anmerkungen und Verbesserungsvorschlägen findet sich neben inhaltlichen Anregungen auch der Wunsch, „dass [der] Workshop ins[besondere] zu Studienbeginn (besser schon während der Studienplanung / Antrag schreiben) am sinnvollsten“ sei. So war die Aussage, dass „viele Informationen für mich zu spät“ kamen im Rahmen der einzelnen Sitzungen leider keine Seltenheit. Wiederholt geäußert wurde auch die Bitte, dass ein derartiger „Kurs (...)“ häufiger stattfinden [sollte], als nur 2x im Jahr, da ein großer Bedarf für die Thematik besteht.“

Die Entwicklung des Workshopkonzepts und die Durchführung der Workshops erfolgten federführend durch GESIS. Die Verbundpartner kooperierten bei der Erstellung des Workshopkonzepts sowie der Publikationen in diesem Themenbereich eng.

Zusätzlich zu den genannten zweitägigen Workshops leistete der VFDB kürzere Einführungen in das Forschungsdatenmanagement und Vorträge zu ausgewählten Themen. Hier sind insbesondere die zweistündige Kurzeinführung in das Forschungsdatenmanagement und die Archivierung von Daten im Verbund Forschungsdaten Bildung bei den GEBF-Tagungen 2015 und 2016 herauszugreifen sowie eine Einführung in das Forschungsdatenmanagement qualitativer Daten bei einer Veranstaltung des Forschungsschwerpunktes „Sprachliche Bildung und Mehrsprachigkeit“ im EBF-Rahmenprogramm in Hamburg im Jahr 2015 (vgl. Anhang 7.6).

4.2. Informationsangebote der Website

Die Website www.forschungsdaten-bildung.de enthält Informationen zum Forschungsdatenmanagement, die fortlaufend aktualisiert und zukünftig weiter ausgebaut werden. Zum Ende der ersten Förderphase standen folgende Informationsangebote zum Forschungsdatenmanagement auf der Website oder zum Download über die Website zur Verfügung (Stand 30.09.2016) (vgl. Tabelle 14):

Tabelle 14: Informationsangebote für Forschende¹³

Dokument/Website	Link (Stand 30.09.2016, Zugriff 30.09.2016)
Informationen zu Data Sharing	www.forschungsdaten-bildung.de/data-sharing
Allgemeine Hinweise zum Datenmanagement	www.forschungsdaten-bildung.de/datenmanagement
Checkliste zur Erstellung eines Datenmanagementplans	www.forschungsdaten-bildung.de/datenmanagementplan ; (Link aktuell: www.forschungsdaten-bildung.de/files/fdbinfo_2.pdf , Zugriffen: 09. Mai 2017)
Hinweise zur Datensicherheit und Datensicherung	www.forschungsdaten-bildung.de/datensicherheit
Informationsangebote für Forschende, die Daten in der eigenen Einrichtung sichern lassen	Nicht online, wird bei Bedarf herausgegeben.
Dateibenennung und Datenorganisation	www.forschungsdaten-bildung.de/datei-benennung
Hinweise zur Aufbereitung quantitativer Daten	www.forschungsdaten-bildung.de/aufbereitung ; www.forschungsdaten-bildung.de/files/fdb-informiert-nr-4.pdf
Hinweise zur Erstellung eines Feldberichts	www.forschungsdaten-bildung.de/files/fdbinformiert_Feldbericht.pdf (Link aktuell: www.forschungsdaten-bildung.de/files/fdbinfo_3.pdf ,

¹³ Aufgrund der Vielzahl der Informationsangebote wurden die Website und die Publikationsorgane seit Abschluss des Projektes (nach 30.09.2016) umstrukturiert. Die genannten Links sind daher nicht alle aktuell und die Inhalte haben sich seit dem 30.09.2016 in einigen Fällen verändert.

	Zugegriffen: 09. Mai 2017)
Hinweise zur Anonymisierung von qualitativen Daten	www.forschungsdaten-bildung.de/files/fdb-informiert-nr-1.pdf
Hinweise zur Anonymisierung von quantitativen Daten	www.forschungsdaten-bildung.de/files/fdb-informiert-nr-3.pdf
Hinweise zu urheberrechtlichen Aspekten der Sicherung und Nachnutzung von Forschungsdaten	www.forschungsdaten-bildung.de/copyright
Hinweise zu datenschutzrechtlichen Aspekten der Sicherung und Nachnutzung von Forschungsdaten	www.forschungsdaten-bildung.de/datenschutzrecht
Hinweise zu ethischen Aspekten der Sicherung und Nachnutzung von Forschungsdaten	www.forschungsdaten-bildung.de/ethik
Hinweise und Checkliste zur Erstellung informierter Einwilligungserklärungen	www.forschungsdaten-bildung.de/einwilligung und www.forschungsdaten-bildung.de/files/FDB_Einwilligung_Checkliste.pdf
Übersicht über die länderspezifischen Besonderheiten für Befragungen an Schulen	www.forschungsdaten-bildung.de/genehmigungen
Dokumentation von urheberrechtlich geschützten Instrumenten	www.forschungsdaten-bildung.de/doku-instrumente
Informationen zu den Forschungsdatenmanagement-Workshops des CESSDA-Trainingscenters im Auftrag des VFDB	www.forschungsdaten-bildung.de/files/fdb-informiert-nr-2.pdf
Informationen zur Datenkuratierung beim VFDB	www.forschungsdaten-bildung.de/data-curation
Welche „Daten“?	www.forschungsdaten-bildung.de/welchedaten
Frequently Asked Questions	www.forschungsdaten-bildung.de/hilfe#Frequently-Asked-Questions

Sämtliche Materialien mit Informationen zu datenschutz- oder urheberrechtlichen Aspekten wurden in Kooperation mit den Rechtsanwälten Goebel & Scheller erstellt.

Speziell für die Schritte der Datenmeldung und -übermittlung durch die Projektnehmer/innen des EBF-Rahmenprogramms wurden die folgenden Handreichungen erstellt (vgl. Tabelle 15):

Tabelle 15: Informationen für Datengeber/innen des EBF-Rahmenprogramms

<i>Dokument/Website (Stand 30.09.2016)</i>	<i>Link (Zugriffe 30.09.2016)</i>
Übersicht zum Ablauf der Dokumentation, Sicherung und Nachnutzung	www.forschungsdaten-bildung.de/ebf
Empfohlene Dateiformate für die Übermittlung der Forschungsdaten-Dateien	www.forschungsdaten-bildung.de/formate
Instruktionen zum Upload	www.forschungsdaten-bildung.de/files/Upload.pdf
Instruktionen zum Zippen von Dateien für den Upload	www.forschungsdaten-bildung.de/files/Software_7Zip.pdf
Instruktionen zur Nutzung von Win-SCP für den Upload	www.forschungsdaten-bildung.de/files/Software_WinSCP.pdf

Nutzung der Website / Downloadraten, Sessions

Für die Analyse der Websitenutzung wurde das Analysetool PIWIK eingesetzt. Es stehen Daten zur Verfügung für den Zeitraum 09/2014 bis 09/2016. Erfasst werden eindeutige Besucher, das heißt ein Besucher wird pro Tag nur einmal erfasst. Es zeigt sich, dass die Zugriffszahlen im Jahr 2016 bereits nach neun Monaten höher waren als im gesamten Jahr 2015 (vgl. Tabelle 16).

Tabelle 16: Zugriffe in den Jahren 2014-2016, eindeutige Besucher

Zeitraum	Summe je Jahr	Durschnitt je Monat
2014 (Sep bis Dez)	3501	875
2015	18615	1551
2016 (bis Sep)	29031	3226

Basis: PIWIK

Weiter wurde betrachtet, welche Informationsseiten (Seiten unter dem damaligen Menüpunkt „Tipps & Materialien“) am häufigsten besucht wurden und welche Materialien am häufigsten heruntergeladen wurden.

Tabelle 17 zeigt für den Projektzeitraum die fünf am häufigsten besuchten Seiten des Webangebots zum Forschungsdatenmanagement an. Zu beachten ist, dass im Jahr 2014 erst ein geringer Teil an Informationsseiten vorlag.

Tabelle 17: Top 5 der Seitenbesuche (einmalig) im Informationsbereich

Rang	2014 (ab 09/2014)	2015	2016 (bis 09/2016)
1	www.forschungsdaten-bildung.de/datenmanagement , Forschungsdatenmanagement (103)	www.forschungsdaten-bildung.de/anonymisierung , Hinweise zur Anonymisierung qualitativer Daten (2151)	www.forschungsdaten-bildung.de/genehmigungen , Genehmigungsverfahren der Bundesländer (4362)
2	www.forschungsdaten-bildung.de/dokumentation , Datendokumentation (103)	www.forschungsdaten-bildung.de/genehmigungen , Genehmigungsverfahren der Bundesländer (1592)	www.forschungsdaten-bildung.de/anonymisierung , Hinweise zur Anonymisierung qualitativer Daten (4026)
3	www.forschungsdaten-bildung.de/data-sharing , Data Sharing (44)	www.forschungsdaten-bildung.de/einwilligung , Informierte Einwilligung (595)	www.forschungsdaten-bildung.de/einwilligung , Informierte Einwilligung (1767)
4	www.forschungsdaten-bildung.de/data-curation , Datenkuratierung (33)	www.forschungsdaten-bildung.de/aufbereitung , Hinweise zur Aufbereitung quantitativer Daten (438)	www.forschungsdaten-bildung.de/aufbereitung , Hinweise zur Aufbereitung quantitativer Daten (759)
5	www.forschungsdaten-bildung.de/einwilligung , Informierte Einwilligung (23)	www.forschungsdaten-bildung.de/datenmanagement , Forschungsdatenmanagement (281)	www.forschungsdaten-bildung.de/datenmanagement , Forschungsdatenmanagement (363)

Basis: PIWIK

Tabelle 18 zeigt die zehn am häufigsten heruntergeladenen Dateien aus dem Informationsbereich des Webangebots von forschungsdaten-bildung.de an über den Projektzeitraum. Hier zeigte sich ebenfalls eine hohe Nachfrage nach Informationen zu rechtlichen Sachverhalten. So wurden im Berichtszeitraum vor allem Dokumente mit Muster-Einwilligungserklärungen und Hinweisen zur

Erstellung einer Einwilligungserklärungen heruntergeladen. Weiterhin waren auch die Hinweise zur Anonymisierung von Daten, speziell von qualitativen Daten, nachgefragt. Die Website forschungsdaten-bildung.de verweist auch auf Angebote anderer Einrichtungen wie des RatSWD oder von Qualiservice und trägt so zu deren Verbreitung bei.

Tabelle 18: Top 10 einmalige Downloads

Rang	2014 (ab 09/2014)	2015	2016 (bis 09/2016)
1	RatSWD Muster 2 Einwilligungserklärung (11)	Hinweise zur Anonymisierung qualitativer Daten (206)	RatSWD Muster 1 Einwilligungserklärung (423)
2	RatSWD Muster 1 Einwilligungserklärung (10)	RatSWD Muster 1 Einwilligungserklärung (141)	Hinweise zur Anonymisierung qualitativer Daten (305)
3	Flyer VFDB (8)	RatSWD Muster 2 Einwilligungserklärung (95)	Qualiservice Muster Einwilligungserklärung (305)
4	Kommentierung zu den RatSWD Muster- Einwilligungserklärungen (6)	Qualiservice Muster Einwilligungserklärung (93)	RatSWD Muster 2 Einwilligungserklärung (237)
5	Hinweise zur Anonymisierung qualitativer Daten (1)	Checkliste zur Erstellung rechtskonformer Einwilligungserklärungen (84)	Checkliste zur Erstellung rechtskonformer Einwilligungserklärungen (218)
6	-	Übergabevereinbarung VFDB (77)	Checkliste Datenmanagementplan (101)
7	-	Upload-Instruktionen (49)	Kommentierung zu den RatSWD Muster-Einwilligungserklärungen (56)
8	-	Hinweise zur Aufbereitung quantitaiver Daten (28)	Hinweise zur Anonymisierung von quantitativen Daten (38)
9	-	Checkliste Datenmanagementplan (25)	Upload-Instruktionen (37)
10	-	Flyer VFDB (15)	Hinweise zur Aufbereitung Quantitaiver Daten (33)

Basis: PIWIK

4.3. Beratungsdienste

Der VFDB hat in der Projektlaufzeit Forschende des EBF-Rahmenprogramms zur Archivierung von Forschungsdaten sowie zum Forschungsdatenmanagement beraten. Hinzu kamen Beratungsfragen von Personen und Projekten außerhalb des EBF-Rahmenprogramms. Hier wurden in 15 Fällen Beratungsdienstleistungen erbracht, die länger als einen halben Tag in Anspruch nahmen. (Beratungsdienstleistungen von kürzerer Dauer wurden nicht standardisiert erfasst.) Inhalte der Beratungsdienste waren u. a. das Erstellen von Einwilligungserklärungen und allgemeine Fragen zum Datenschutz, die Erstellung von Datenmanagementplänen, die Beantragung von Mitteln für das Datenmanagement sowie Fragen zum Ablauf der Datenarchivierung oder den Anforderungen an die Dokumentation.

4.4. Ausblick

Die Beratungstätigkeiten werden voraussichtlich mittel- oder langfristig einen positiven Einfluss auf das in den Projekten durchgeführte Forschungsdatenmanagement haben. Wir erwarten perspektivisch eine steigende Qualität der übermittelten Materialien sowie geringere, rechtliche Hürden (durch fehlende Einverständniserklärungen) für die Archivierung und Nachnutzung. Dies

würde die archivseitigen Aufwände reduzieren. Nach wie vor schwankt der Aufwand für Dokumentation und Aufbereitung zwischen den Studien stark. Besondere Informations- und Beratungsbedarfe werden auch zukünftig im Bereich rechtlicher Fragen (Datenanonymisierung etc.) gesehen.

Aufgrund der steigenden Nachfrage wird der VFDB seine bereits implementierten Angebote in der zweiten Förderphase weiter ausbauen. Um die Reichweite der Qualifizierungsangebote zu erhöhen, wird das bisherige Schulungsprogramm zukünftig durch niedrigschwelligere Angebote wie Webinare und begleitende Diskussionsforen im Web ergänzt. Zusätzlich wird mit der Ausbildung von Multiplikatoren („Train-The-Trainer“) begonnen.

5. Zusammenfassung und Ausblick

Wie dargestellt, hat der VFDB in kurzer Zeit ein funktionsfähiges Angebot zur Datenarchivierung für die Gesamtheit der Projekte des EBF-Rahmenprogramms geschaffen. Seit Anfang 2015 haben Projektnehmer/innen die Möglichkeit, ihre Daten über den VFDB sichern zu lassen. Der VFDB hat in der Pilotphase ein gemeinsames Portal mit Informationen zu Studien der Bildungsforschung, einen zentralen Recherchezugang zu Studien und Daten der Bildungsforschung sowie eine zentrale Möglichkeit der Datenübergabe für die Vielzahl und die disziplinäre Vielfalt der Projekte des EBF-Rahmenprogramms geschaffen. Die entwickelten Services ermöglichen es den Forscherinnen und Forschern, ihre Daten online zu melden und an den VFDB zu übermitteln. Die Forschungsdaten werden anschließend in die beteiligten Datenzentren eingespeist, dort aufbereitet und soweit möglich für Dritte verfügbar gemacht. Die Studien der empirischen Bildungsforschung des EBF-Rahmenprogramms sowie die verteilt archivierten Forschungsdaten sind über das Portal forschungsdaten-bildung.de an zentraler Stelle sichtbar und auffindbar.

Trotz der geschaffenen Infrastruktur konnten bis zum Projektende nicht alle der im EBF-Rahmenprogramm generierten Forschungsdaten gesichert werden, und nur zu einem geringen Teil konnten Daten zur Nachnutzung verfügbar gemacht werden. In Kapitel 3.3 wurden daher die Hürden beschrieben, die eine vollumfassende Sicherung und Bereitstellung der im Rahmenprogramm generierten Datenbestände erschwerten und zu einem nicht unerheblichen Ausmaß verhinderten. Hervorzuheben sind vor allem die fehlenden, rechtlichen Voraussetzungen zur Archivierung der Forschungsdaten (vgl. Kapitel 3.3.1; 3.3.2; 3.3.3) und die in den Projekten fehlenden Ressourcen zur Abwicklung der Archivierung selbst (vgl. Kapitel 3.3.4). Ein Teil dieser Probleme ist dem im Vergleich zum Beginn des EBF-Rahmenprogramms deutlich nachgelagerten Projektbeginn des VFDB geschuldet. Es ist davon auszugehen, dass diese Probleme in geringerem Ausmaß auftreten, wenn Projektnehmer/innen frühzeitiger – mit Antragstellung bzw. vor Datenerhebung – mit dem VFDB in Kontakt treten und sich über die Voraussetzungen der Forschungsdatenarchivierung informieren. Neben diesen objektiven Archivierungshürden sind subjektive Aspekte zu berücksichtigen. So bestehen teilweise erhebliche Bedenken auf Seiten der Forschenden, die die Archivierung erschweren (vgl. Kapitel 3.3.5).

Der VFDB hat für die meisten der genannten Problemfelder bereits in der Pilotphase Maßnahmen implementieren können, um sie gezielt aufzugreifen. So wurden Informationen über die Website und über Vorträge distribuiert. Daneben gab es die Schulungs- und Beratungsangebote zum Forschungsdatenmanagement. Allerdings werden die entwickelten Maßnahmen vor allem prospektiv

wirksam. Um zukünftig möglichst viele Projektnehmer/innen möglichst frühzeitig zu erreichen, sollten mehr Ressourcen in die bilaterale Beratung und in den Ausbau der Informationsangebote fließen. Hierzu ist der VFDB auch auf die Unterstützung der Fördereinrichtungen angewiesen, die über die Information verfügen, welche Projekte Datenerhebungen planen und durchführen, und die die Projekte im Rahmen der Antragstellung zu Forschungsdatenmanagement und Datenweitergabe verpflichten oder entsprechende Empfehlungen aussprechen.

Anreize und Verpflichtungen zum Forschungsdatenmanagement und zur Datenweitergabe sind aufgrund des Nachhaltigkeitsanspruchs an Forschungsdaten begrüßenswert. Sie sind jedoch zwingend von flankierenden Maßnahmen zu begleiten, um erfolgreich zu sein. Eine reine Verpflichtung ohne eine flankierende Unterstützung der Forschenden *und* der Dateninfrastrukturen stößt an vielfältige Grenzen. Das haben die Erfahrungen in der Pilotphase des VFDB gezeigt. Vor dem Hintergrund der neuen Anforderungen sowohl an Forschende als auch an die Forschungsdateninfrastruktur werden im Folgenden einige, bestehende Desiderate herausgegriffen und kurz vorgestellt:

(1) Ressourcenbedarf in den Projekten: Es besteht Bedarf an einem Austausch mit und innerhalb der Bildungsforschung darüber, ob und ggf. in welchem Umfang ein zusätzlicher Ressourcenbedarf in den datenerhebenden Projekten für die Arbeiten des Forschungsdatenmanagements besteht. Gut dokumentierte, aufbereitete und nachvollziehbare Forschungsdaten zu erzeugen, entspricht den üblichen Anforderungen einer guten wissenschaftlichen Praxis. Grundlegende Maßnahmen des Forschungsdatenmanagements sollten daher unabhängig von einer eventuell nachfolgenden Archivierung in einem Datenzentrum und der Nachnutzung durch Dritte standardmäßig in jedem Projekt durchgeführt werden. Ressourcen für diese Maßnahmen sollten Projekte grundsätzlich bei der Beantragung von Fördermitteln berücksichtigen. Aussagen darüber, welche Kosten Projekten darüber hinaus – mit Blick auf die Erstellung langfristig nachnutzbarer Daten – entstehen, lassen sich pauschal nur sehr schwer treffen, da sie einerseits von den konkret erzeugten Daten und andererseits von den anvisierten Nachnutzungsszenarien abhängen.

Aktuell befassen sich einige Projekte mit Kostenmodellen der Archivierung digitaler Objekte. Bislang lassen sich hieraus jedoch noch keine Empfehlungen oder konkreten Aussagen ableiten, welche Kosten damit für die Forschenden einhergehen. Der VFDB empfiehlt Forschungsprojekten, sich dieser Frage im Rahmen der Planung des Forschungsdatenmanagements bei Antragstellung zu widmen und sich gegebenenfalls beraten zu lassen. Anhand eines Datenmanagementplans können die Aufwände für die einzelnen Tätigkeiten kalkuliert werden. Dabei sind nicht nur zusätzliche Personalressourcen sondern ggf. auch zusätzliche Software, Hardware oder Mittel für externe Dienstleistungen zu berücksichtigen. Die bereits bekannten Arbeitsschritte des Forschungsdatenmanagements, wie sie in jedem datenerhebenden Projekt unabhängig von der Archivierung der Forschungsdaten nach Projektende anfallen, sind unter Umständen auszuweiten oder anzupassen und zusätzliche Arbeitsschritte wie bspw. die vollständige/re Anonymisierung der Forschungsdaten, das Einholen der Archivierung beinhaltender Einverständniserklärungen und die Inanspruchnahme von Beratungsleistungen zu berücksichtigen.

(2) Ressourcenbedarf in den FDZ: Bislang ist unklar welcher Ressourcenbedarf zukünftig bei den FDZ anfallen wird, um die Masse an Forschungsdaten, die generiert wird, zu bearbeiten und wie dieser Ressourcenbedarf finanziert werden kann. Daher ist es erforderlich, entsprechende Geschäfts- und

Gebührenmodelle zu entwickeln (vgl. bspw. Palaiologk et al. 2012). Derzeit bieten die meisten FDZ ihre Leistungen kostenfrei an und sind meist ebenfalls projektförmig finanziert. Um ein nachhaltiges Angebot an qualitätsgesicherten Forschungsdaten bieten zu können, benötigen FDZ jedoch eine verlässliche Finanzierungsgrundlage. Ergänzend zu einer direkten Finanzierung der Infrastrukturen durch öffentliche Mittel für die Basisdienstleistungen zur Grundsicherung ist vorstellbar, besondere Dienstleistungen der FDZ zu bepreisen und so zumindest teilweise durch diejenigen finanzieren zu lassen, die diese in Anspruch nehmen. Letzteres würde zu einer größeren Transparenz hinsichtlich der in einem konkreten Projekt entstehenden Kosten beitragen. Darüber hinaus würde so die konkrete Entscheidung darüber, in welche Daten Mittel in welcher Höhe investiert werden soll, stärker von den Dateninfrastrukturen hin zu den auf Fachgutachten basierenden Entscheidungen der Förderinstitutionen verlagert. Die mit verschiedenen Geschäftsmodellen (mit mehr oder weniger fixen oder variablen Komponenten) einhergehenden Vor- und Nachteile sind in ihren Wirkungen zu untersuchen und mit allen Beteiligten zu diskutieren.

(3) *Selektion von Forschungsdaten:* Vor dem Hintergrund begrenzter Ressourcen können die archivarischen Dienstleistungen nicht allen Projekten auf dem gleichen Niveau bzw. in der gleichen Tiefe angeboten werden. Stattdessen ist es erforderlich, differenzierte Angebote selektiv, das heißt für unterschiedliche Gruppen von Forschungsdaten und Studien anzubieten. Hierzu sind entsprechende Kriterien und Verfahren der Selektion (vgl. z. B. Whyte und Wilson 2010) auf der einen und gestufte Dienstleistungsmodelle und Serviceangebote auf der anderen Seite zu entwickeln. Ziel ist es, auf fundierter Basis in einem transparenten und nachvollziehbaren Verfahren entscheiden zu können, für welche Daten welche Art von Leistung zu erbringen ist. Dabei ist anzumerken, dass Selektionsentscheidungen nicht alleine von Archiven getroffen werden können, sondern der Rückkopplung mit der Bildungsforschung bedürfen. Auch könnte eine solche Entscheidung Teil des Begutachtungsverfahrens bei Antragstellung sein.

(4) *Kenntnisse des Forschungsdatenmanagements:* Forschungsdatenmanagement ist bislang noch kein Bestandteil universitärer Ausbildung oder der Promovendenausbildung im Rahmen von Graduiertenschulen. Aktuell bedarf es der Eigeninitiative der Forschenden, sich zusätzliche Kenntnisse des FDM anzueignen und Schulungsangebote wahrzunehmen. Zu klären ist, wie das erforderliche Knowhow in die *scientific communities* und die einzelnen Forschungsprojekte hineingetragen werden kann. Bislang sind die Angebote in diesem Bereich noch zu wenig ausgebaut, und es bedarf erheblicher Anfangsinvestitionen, da in den Archiven die entsprechenden Kompetenzen und Kapazitäten auf- und ausgebaut werden müssen.

(5) *Auswirkungen auf die empirische Forschung:* Bislang gibt es kaum empirische Befunde zu den Effekten der Archivierung und Nachnutzung von Forschungsdaten auf die Teilnahmebereitschaft im Speziellen als auch auf den Prozess der Datengenerierung im Allgemeinen.¹⁴ Dies gilt insbesondere für die qualitative Sozialforschung. Die Befürchtungen unter qualitativen Forscher/innen sind groß, Teilnehmer/innen durch die Anforderungen zur Datenarchivierung abzuschrecken und das Vertrauensverhältnis zwischen Forscherin und Beforschter nachhaltig zu zerstören. Qualitative Forscher/innen befürchten, dass die Archivierungsanforderung Feldforschung verunmöglichen könnte (vgl. bspw. Hirschauer 2014). Hier besteht dringender Forschungsbedarf.

¹⁴ Eine Ausnahme ist bspw. Kuula 2011.

(6) Anforderungen an die Dokumentation: Im Themenfeld Dokumentation sind Fragen im Hinblick auf Art und Tiefe der Dokumentation qualitativen ebenso wie quantitativen Materials von Relevanz. Es gibt wenige empirische Studien dazu, welche Auswirkungen die Art und Tiefe der Dokumentation auf die Möglichkeiten der Nachnutzung hat. Forschungsdatenzentren standardisieren die Dokumentation zwar, diese basiert aber auf der bereits vorhandenen, d. h. in den datenerhebenden Projekten entstandenen Dokumentation. Wenn diese nicht an Standards orientiert ist (bzw. sein kann, weil es solche nicht gibt), beeinflusst das die Interpretierbarkeit und Vergleichbarkeit der Forschungsdaten nachhaltig. Die Festlegung von Standards ist zunächst kein archivarisches, sondern ein methodologisches Problem: Welche Indikatoren zur Interpretation und Vergleichbarkeit von Forschungsdaten sind standardmäßig wie festzuhalten und zu berichten? Es ist zu erwarten, dass durch die neuen Anforderungen zur Datenweitergabe und die steigende Nachfrage nach Sekundärdaten der Diskurs zu dieser Frage an Bedeutung gewinnt (vgl. RatSWD 2014).

(7) Rechtliche Regelungslücken: Wie oben in den vorherigen Kapiteln aufgeführt, gibt es ungeklärte rechtliche Aspekte, die starke Auswirkungen auf die Nachhaltigkeit von Forschungsdaten haben. So sehen sich Forschende widersprüchlichen Anforderungen von Fördereinrichtungen und genehmigenden Behörden gegenüber. Die Erfahrungen der Pilotphase haben gezeigt, dass die derzeitige behördliche Genehmigungspraxis in vielen Fällen eine Archivierung und Nachnutzung der erhobenen Daten verhindert. Hier sind die Entscheidungsträger gefragt, Rahmenbedingungen zu schaffen, die die neuen Anforderungen an Data Sharing berücksichtigen. Mit Blick auf die datenschutzrechtlichen Aspekte der Archivierung von Forschungsdaten bleibt abzuwarten, welche Änderungen sich im Zuge der bis 2018 umzusetzenden EU-Datenschutzgrundverordnung ergeben werden. Mit Blick auf urheberrechtliche Aspekte ist zu konstatieren, dass die kommerzielle Verwertung von Messinstrumenten, die durch öffentliche Mittel finanziert wurden, dem Gedanken des Open Access entgegensteht. Fehlende Lizenzinformationen bei frei verfügbaren Instrumenten führen zu Rechtsunsicherheit bei der Entwicklung von Fragebögen und der Planung von Studien. Hier sind unter anderem entsprechende Qualifizierungsangebote erforderlich.

(8) Anreize zur Archivierung und zum Schutz der Investitionen der Primärforschenden: Die gegenwärtigen Anreizstrukturen innerhalb der Bildungsforschung unterstützen nicht die Investition in die Erzeugung qualitativ hochwertiger Daten und deren Weitergabe. Belohnt werden vor allem Publikationen von Forschungsartikeln. Dies liegt daran, dass Forschungsdaten nicht den „Status eines eigenständigen, wertvollen, wissenschaftlichen ‚Gutes‘“ (Meyermann 2012: 14) genießen, und es an einer „Publikationskultur für Forschungsdaten“ (Büttner et al. 2011: 19) mangelt. Gleichzeitig sind die Investitionen zur Erzeugung qualitativ hochwertiger Daten hoch (Risiko des Misserfolgs, finanzielle und personelle Mittel, Kompetenzen). Diese werden derzeit vor allem auch deshalb geleistet, da der exklusive Zugang zu Forschungsdaten einen nicht unerheblichen Vorteil im Wettbewerb um die begrenzten Publikationsplätze in Journals bedeutet. Ein generelles Archivierungsgebot zeitnah nach Abschluss der Projektlaufzeit würde diesen Wettbewerbsvorteil gefährden. Es besteht das Risiko, dass die Anreize, Primärforschungen durchzuführen, zu gering werden könnten. Auch Stanat (2014) fordert die Möglichkeit von Sperrfristen. Sperrfristen sollten sich im besten Fall nur auf Teile eines Datenbestands oder bestimmte Forschungsfragen beziehen. So könnten die Interessen der Primärforscher/innen gewahrt und gleichzeitig aktuelle Daten verfügbar gemacht werden.

Die voranstehend genannten Desiderate sind von den beteiligten Akteuren zu adressieren, damit sich ein nachhaltiger, qualitätsorientierter und nutzerfreundlicher Umgang mit Forschungsdaten in der

Bildungsforschung etablieren kann. Hierzu sind gemeinsame Anstrengungen der Dateninfrastruktur, der Forschungsförderung und der Forschung zu unternehmen.

Der VFDB befindet sich derzeit in seiner zweiten Förderphase (10/2016-09/2019). In den kommenden Jahren werden die bestehenden Services weiter ausgebaut und auch für Projekte außerhalb des EBF-Rahmenprogramms zugänglich sein. Die Arbeit des VFDB verfolgt dabei nach wie vor die drei Schwerpunkte: (1) Archivierung und Bereitstellung von Forschungsdaten anbieten, (2) Übersicht über Studien und Datenbestände schaffen, (3) Schulungen und Beratungen zum Forschungsdatenmanagement durchführen. Um seine Aufgaben optimal zu erfüllen, wird der VFDB die Zusammenarbeit mit anderen Forschungsdatenzentren ebenso wie den Austausch mit Bildungsforscherinnen und -forschern verstärken. Der VFDB kann hierbei die Rolle eines infrastrukturellen Knotenpunkts für die Bildungsforschung einnehmen, der wiederum Teil einer nationalen Forschungsdateninfrastruktur ist. Der VFDB hat hierzu in den letzten Jahren pilothaft ein spezifisches Modell einer föderierten Infrastruktur erprobt, das voranstehend als Modell der „verteilten Archivierung“ bezeichnet wurde. Das gemeinsam zu verfolgende Ziel besteht darin, ein nutzerfreundliches Angebot zu erstellen und Forschende dabei zu unterstützen, Forschungsdaten zu erzeugen und nachzunutzen und damit letztlich den wissenschaftlichen Erkenntnisgewinn zu fördern. Der VFDB unterstützt dabei eine Vision von *Open Data*, in der die datenschutzrechtlichen und ethischen Verpflichtungen gegenüber den Betroffenen gewahrt und die *first user* Ansprüche der Datenproduzierenden respektiert werden.

6. Literatur

- Allianz der Wissenschaftsorganisationen. 2010. Grundsätze zum Umgang mit Forschungsdaten. www.allianzinitiative.de/fileadmin/user_upload/www.allianzinitiative.de/Grundsaeetze_Forschungsdaten_2010.pdf (Zugegriffen: 09. Mai 2017).
- Allianz der deutschen Wissenschaftsorganisationen. 2014. Appell zur Nutzung offener Lizenzen in der Wissenschaft. www.allianzinitiative.de/fileadmin/user_upload/www.allianzinitiative.de/Appell_Offene_Lizenzen_2014.pdf (Zugegriffen: 08. Sep. 2015).
- Borgman, Christine L. 2010. Research Data: who will share what, with whom, when, and why? *RatSWD Working Paper Series* 161. www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_161.pdf (Zugegriffen: 08. Mai 2017).
- Büttner, Stephan, Hans-Christoph Hobohm und Lars Müller. 2011. *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock und Herchen.
- Forschungsdaten Bildung. o.J. Übersicht über die länderspezifischen Besonderheiten für Befragungen an Schulen. www.forschungsdaten-bildung.de/genehmigungen (Zugegriffen: 18. Apr. 2016).
- Haas, Martine R., und Sangchan Park. 2010. To Share or Not to Share? Professional Norms, Reference Groups, and Information Withholding Among Life Scientists. *Organization Science* 21 (4): 873–891. <http://orgsci.journal.informs.org/cgi/content/abstract/21/4/873> (Zugegriffen: 08. Mai 2017).
- Häussler, Carolin. 2010. Information-Sharing in Academia and the Industry. A Comparative Study. *RatSWD Working Paper Series* 154. www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_154.pdf (Zugegriffen: 08. Mai 2017).
- Harvey, Ross. 2010. *Digital Curation: A How-To-Do-It Manual*. New York, NY: Neal-Schuman Publishers.
- Hirschauer, Stefan. 2014. Sinn im Archiv? Zum Verhältnis von Nutzen, Kosten und Risiken der Datenarchivierung. *Soziologie* 43 (3): 300-312.
- Homuth, Christoph, Monja Schmitt, Christian Lorenz und Daniel Mann. 2017. Warum ein erneutes Genehmigungsverfahren im laufenden Längsschnitt weitreichende Folgen für die Datenqualität hat. *JERO* 9 (1): 7-31. www.j-e-r-o.com/index.php/jero/article/viewFile/731/302 (Zugegriffen: 05. Mai 2017).
- Jensen, Uwe. 2012. Leitlinien zum Management von Forschungsdaten, Sozialwissenschaftliche Umfragedaten. *gesis Technical Reports* 7. www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-07.pdf (Zugegriffen: 09. Mai 2017).
- Kommission Zukunft der Informationsinfrastruktur. 2011. Gesamtkonzept für die Informationsinfrastruktur in Deutschland. Empfehlungen der Kommission Zukunft der Informationsinfrastruktur im Auftrag der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder. https://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Infrastruktur/KII_Gesamtkonzept.pdf (Zugegriffen: 05. Mai 2017).
- Kuula, Arja. 2011. Methodological and Ethical Dilemmas of Archiving Qualitative Data. *IASSIST Quarterly* 35 (1-2): S. 12–17. www.iassistdata.org/sites/default/files/iqvol34_35_kuula.pdf (Zugegriffen: 05. Mai 2017).
- Liebig, Stefan, Tobias Gebel, Matthis Grenzer, Julia Kreuzsch, Heidi Schuster, Ralf Tscherwinka, Oliver Watteler und Andreas Witzel. 2014. Datenschutzrechtliche Anforderungen bei der Generierung und Archivierung

qualitativer Interviewdaten. *RatSWD Working Paper Series* 238. www.ratswd.de/dl/RatSWD_WP_238.pdf (Zugegriffen: 17. Jun. 2016).

Meyermann, Alexia. 2012. Datenmanagement. Eine zentrale Voraussetzung für den Erfolg der Data Sharing-Idee in den Sozialwissenschaften. *DSZ-BO Working Paper* 2. www.uni-bielefeld.de/dsz-bo/pdf/2012-07-23_WP2_Datenmanagement_v2.pdf (Zugegriffen: 05. Mai 2017).

nestor-Arbeitsgruppe OAIS-Übersetzung/Terminologie. 2013. Referenzmodell für ein Offenes-Archiv-Informationssystem – Deutsche Übersetzung 2.0. *nestor-Materialien* 16. <http://d-nb.info/104761314X/34> (Zugegriffen: 08. Mai 2017).

OECD. 2007. OECD Principles and Guidelines for Access to Research Data from Public Funding. www.oecd.org/sti/sci-tech/38500813.pdf (Zugegriffen: 30. Aug. 2016).

Palaiologk, Anna S., Anastasios A. Economides, Heiko D. Tjalsma, und Laurents B. Sensink. 2012. An activity-based costing model for long-term preservation and dissemination of digital research data: the case of DANS. *International Journal on Digital Libraries* 12 (4): 195-214.

RatSWD. 2014. Qualitätsstandards zur Entwicklung, Anwendung und Bewertung von Messinstrumenten in der sozialwissenschaftlichen Umfrageforschung. *RatSWD Working Paper Series* 230. www.ratswd.de/dl/RatSWD_WP_230.pdf (Zugegriffen: 09. Mai 2017).

RatSWD. 2017. Handreichung Datenschutz. *Output Paper* 5. www.ratswd.de/dl/RatSWD_Output5_HandreichungDatenschutz.pdf (Zugegriffen: 12. Apr. 2017).

rdn (research data netherlands). o.J. A federated data infrastructure for the Netherlands: the front-office – back-office model. http://www.researchdata.nl/fileadmin/content/RDNL_algemeen/Documenten/RDNL_FOBOModel-UK-web.pdf (Zugegriffen: 09. Mai 2017).

RIN (Report commissioned by the Research Information Network). 2008. To Share or Not To Share: Publication and Quality Assurance of Research Data Outputs. Main Report. https://eprints.soton.ac.uk/266742/1/Published_report_-_main_-_final.pdf (Zugegriffen: 05. Mai 2017).

RfII (Rat für Informationsinfrastrukturen). 2016. Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. www.rfii.de/?wpdmml=1998 (Zugegriffen: 05. Mai 2017).

Smolle, Michael. 2015. Datenschutzrechtliche Anmerkungen zum Artikel „vom Datenangreifer zum zertifizierten Wissenschaftler“ von Ulrich Rendtel. *AStA Wirtschafts- und Sozialstatistisches Archiv* 9: 73-78.

Stanat, Petra. 2014. Bereitstellung und Nutzung quantitativer Forschungsdaten in der Bildungsforschung: Memorandum des Fachkollegiums „Erziehungswissenschaft“ der DFG. http://dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/forschungsdaten_memorandum_fk_109.pdf (Zugegriffen: 08. Mai 2017).

Trixa, Jessica, und Thomas Ebel. 2015. „Einführung in das Forschungsdatenmanagement in der empirischen Bildungsforschung“. Zusammenfassung des Workshops im Rahmen des CESSDA-Trainingsprogramms. *forschungsdaten bildung informiert* 2. www.forschungsdaten-bildung.de/forschungsdaten-bildung-informiert (Zugegriffen: 08. Mai 2017).

UK Data Archive. 2014. Data Ingest Processing Procedures. Version 07.00. www.data-archive.ac.uk/media/54782/ukda079-ds-dataprocessingstandards.pdf (Zugegriffen: 08. Mai 2017).

Verbund Forschungsdaten Bildung. 2015. Checkliste zur Erstellung rechtskonformer Einwilligungserklärungen mit besonderer Berücksichtigung von Erhebungen an Schulen. Version 1.1. fdbinfo Nr. 1.

www.forschungsdaten-bildung.de/files/fdbinfo_1.pdf (Zugegriffen: 08. Mai 2017).

Verbund Forschungsdaten Bildung. 2015. Checkliste zur Erstellung eines Datenmanagementplans in der empirischen Bildungsforschung. Version 1.1. fdbinfo Nr. 2. [www.forschungsdaten-](http://www.forschungsdaten-bildung.de/files/fdbinfo_2.pdf)

[bildung.de/files/fdbinfo_2.pdf](http://www.forschungsdaten-bildung.de/files/fdbinfo_2.pdf) (Zugegriffen: 09. Mai 2017)

Whyte, Angus, und Andrew Wilson. 2010. How to Appraise and Select for Curation. *A Digital Curation Centre und Australian National Data Service 'working level' guide.*

www.dcc.ac.uk/sites/default/files/documents/How%20to%20Appraise%20and%20Select%20Research%20Data.pdf (Zugegriffen: 02. Jul. 2014).

Wissenschaftsrat. 2012. Empfehlungen zur Weiterentwicklung der wissenschaftlichen

Informationsinfrastrukturen in Deutschland bis 2020. www.wissenschaftsrat.de/download/archiv/2359-12.pdf

(Zugegriffen: 05. Mai 2017).

7. Anhang

7.1. Die Projekte des EBF-Rahmenprogramms

Das BMBF-Rahmenprogramm förderte im EBF-Rahmenprogramm Projekte in verschiedenen, thematisch zusammenhängenden Förderschwerpunkten. Der VFDB war für 316 Projekte aus neun dieser Förderschwerpunkte zuständig sowie zusätzlich für den Schwerpunkt ASCOT.¹⁵

- Professionalisierung des pädagogischen Personals (ProPäda)
- Lehr-Lern-Forschung unter neurowissenschaftlicher Perspektive (NIL)
- Sprachliche Bildung und Mehrsprachigkeit
- Chancengerechtigkeit und Teilhabe
- Steuerung im Bildungssystem (SteBiS)
- Entwicklungsstörungen schulischer Fertigkeiten (ESF)
- Sprachdiagnostik/Sprachförderung (FiSS)
- Jedem Kind ein Instrument (JeKI)
- Promotionsförderung

Eine inhaltliche Beschreibung der Forschungsschwerpunkte findet sich in nachfolgender Tabelle (vgl. Tabelle 19) (sowie auf <http://www.empirische-bildungsforschung-bmbf.de/de/1381.php>.) Es ist festzustellen, dass die disziplinäre und methodische Vielfalt im Rahmenprogramm sehr groß ist. Im Rahmenprogramm vertretene Disziplinen sind neben den Erziehungswissenschaften/der Pädagogik, die Sozial- und Wirtschaftswissenschaften, die Psychologie und Neurowissenschaften, Didaktik-Institute diverser Fachbereiche (Mathematik, Biologie, Musik u. a.) sowie Sprachwissenschaften. Mit den Sprachwissenschaften ist auch eine geisteswissenschaftliche, mit den Neurowissenschaften und (teilweise) der Psychologie auch eine naturwissenschaftliche Disziplin vertreten. Die disziplinäre Vielfalt spiegelt sich auch in der Vielfalt der eingesetzten Methoden und Forschungsdesigns wider. So sind neben reinen Surveys, Leistungs- und Kompetenzmessungen auch qualitative Verfahren wie AV-gestützte Unterrichtsbeobachtungen oder neurowissenschaftliche Messverfahren vertreten; neben Querschnittsdesigns gibt es experimentelle Designs (z. B. bei Evaluationsstudien).

¹⁵ Der Förderschwerpunkt NEPS des EBF-Rahmenprogramms ist nicht Gegenstand der Akquisition des Verbunds.

Tabelle 19: Forschungsschwerpunkte im EBF-Rahmenprogramm

Schwerpunkte	Kurzbeschreibung	Disziplinen	Methoden
ProPäda: Entwicklung von Professionalität des pädagogischen Personals in Bildungseinrichtungen Förderzeitraum: 2009-2016	„Die Fördermaßnahme „Entwicklung von Professionalität des pädagogischen Personals in Bildungseinrichtungen“ fokussiert im Kontext des Rahmenprogramms zur Förderung der empirischen Bildungsforschung auf die empirische Untersuchung der Entwicklungsbedingungen professioneller Kompetenzen des pädagogischen Personals. So soll langfristig eine empirisch fundierte Adjustierung von Standards der Aus- und Fortbildung pädagogischen Personals sowie dessen Professionswissen erreicht werden. Zudem wird durch diese Maßnahme eine evidenzbasierte Outputkontrolle im Bildungssystem angestrebt.“ (URL: http://www.empirische-bildungsforschung-bmbf.de/de/1945.php (Zugriff am 13.08.15))	Erziehungs- wissenschaften, Psychologie, diverse Didaktik- Institute mehrerer Fachbereiche	Kompetenzmessungen, Schülerleistungstests, Unterrichtsbeobachtungen, Mixed Methods, Experteninterviews, Experimente, Videoanalysen
Chancengerechtigkeit und Teilhabe: Sozialer Wandel und Strategien der Forschung Förderzeitraum: 2011 - 2015	„Ziel des Förderschwerpunkts ist es, Ausmaß, Erscheinungsformen und Ursachen ungleicher Bildungsteilhabe sowie die Wirkungsweise von Maßnahmen, Programmen und Reforminstrumenten zur Verringerung von Bildungsungleichheit zu untersuchen, um Förderstrategien für die pädagogische Praxis weiterentwickeln zu können. Die Weiterentwicklung der Förderstrategien setzt empirische Kenntnisse über die beteiligten Akteure in ihren sozialen Lebenswelten, ihre Bildungsstrategien und -einstellungen voraus. Akteure sind v.a. die Kinder und Jugendlichen selbst, aber auch das pädagogische Personal und die Eltern.“ (URL: http://www.empirische-bildungsforschung-bmbf.de/de/1387.php (Zugriff am 13.08.15))	Erziehungswissen schaften, Psychologie, Soziologie	Qualitative Interviews, quantitative Befragungen, Meta-Analysen, Längsschnittuntersuchungen, Kompetenzmessungen, Schülerleistungstests, Videoanalysen
ESF: Entwicklungsstörungen schulischer Fertigkeiten Förderzeitraum: 2010 - 2017	„Lesen, Schreiben und Rechnen sind wichtige kulturelle Fähigkeiten, ohne die eine aktive und selbstbestimmte Teilhabe am gesellschaftlichen Leben nur erschwert möglich ist. Zahlreiche Studien belegen die schlechteren Bildungs- und Berufschancen für Menschen mit Defiziten in diesen Leistungsbereichen. Eine evidenzbasierte Förderung der betroffenen Kinder, Jugendlichen und Erwachsenen ist daher von zentraler Bedeutung für die Gewährleistung gleicher Möglichkeiten individueller Entfaltung. Allerdings sind Entwicklungsstörungen schulischer Fertigkeiten wie auch Fördermaßnahmen in den genannten Bereichen bislang noch unzureichend wissenschaftlich untersucht. [...] Diese BMBF-Fördermaßnahme zielt darauf ab, durch Forschungsprojekte empirisches Wissen zu generieren, das dazu beitragen kann, Kindern, Jugendlichen und Erwachsenen, die von Störungen im Bereich des Lesens, Schreibens und Rechnens betroffen sind, eine individuelle, ursachenbezogene Diagnostik und evidenzbasierte Förderung zu ermöglichen.“ (URL: http://www.empirische-bildungsforschung-bmbf.de/de/1389.php (Zugriff am 18.08.15))	Psychologie, Neurowissenschaf ten, Erziehungswissen schaften	Längsschnittuntersuchungen, Experimente, Kompetenzmessungen, IQ-Tests, neurowissenschaftliche Untersuchungen
FISS: Sprachdiagnostik / Sprachförderung Förderzeitraum: 2009 - 2016	„Die Fördermaßnahme ermöglicht vorzugsweise interdisziplinär angelegte empirische Forschungsvorhaben zu Fragen des Erwerbs, der Diagnose und der Förderung sprachlicher Kompetenzen, insbesondere im Hinblick auf die spezifischen sprachlichen Anforderungen in den Bildungsinstitutionen. Gefördert werden grundlagen- und anwendungsbezogene Forschungsvorhaben zu Verfügbarkeit, Struktur, Nutzung und Erwerb sprachlicher Kompetenzen, Studien zur Messung sprachlicher Kompetenzen und Kompetenzzuwächse, Interventionsstudien zu konkreten Angeboten der Sprachförderung, ihrer Implementierung in Bildungsinstitutionen und (differenziellen) Wirksamkeitsüberprüfungen sowie Interventionsstudien zu Angeboten der Qualifizierung pädagogischer Kräfte für die Aufgaben der Erfassung von Sprachständen und der Sprachförderung.“ (URL: http://www.empirische-bildungsforschung-bmbf.de/de/1395.php (Zugriff am 18.08.15))	Erziehungswissen schaften, Psychologie, Sprachwissenschaf ften	Videoanalysen, Interviews, Längsschnittuntersuchungen, Experimente, Kompetenzmessungen, computergestützte Reaktionsmessungen
JeKi: Jedem Kind ein Instrument	„Das Forschungsprogramm soll wissenschaftliche Erkenntnisse erarbeiten und bereitstellen, die bezogen sind auf die Ziele der beiden Projekte "Jedem Kind ein Instrument" (JeKi), die in Hamburg und in Nordrhein-Westfalen im Raum Essen und Ruhrgebiet durchgeführt werden. Beide Projekte sollen dazu beitragen, Kinder	Musik- wissenschaft, Musikpädagogik,	Dokumentenanalyse, Interviews, quantitative Erhebungen,

<p>Förderzeitraum: 2009 - 2015</p>	<p>bereits im Grundschulalter über das Erlernen eines Musikinstrumentes aktiv an Musik heranzuführen, ihnen die Freude am Musizieren zu vermitteln und ihnen – ggf. auch ihren Familien – damit den Zugang zur Kultur in besonderer Weise und nachhaltig zu eröffnen. [...] Darüber hinaus dient das Forschungsprogramm dem Ziel, bisherige Forschungsergebnisse und Forschungshypothesen zum Musik- bzw. Instrumentalunterricht auf der Grundlage einer sehr umfangreichen Stichprobe weiter zu entwickeln und einer umfassenden empirischen Prüfung zu unterziehen.“ (URL: http://www.empirische-bildungsforschung-bmbf.de/de/1397.php (Zugriff am 19.08.15))</p>	<p>Erziehungswissenschaften, Psychologie</p>	<p>Videoanalysen, Längsschnittuntersuchungen, Mehrebenenanalysen, neurowissenschaftliche Untersuchungen</p>
<p>Sprachliche Bildung und Mehrsprachigkeit Förderzeitraum: 2013 - 2017</p>	<p>„Ausreichende Sprachkompetenzen sind die Grundvoraussetzung für Bildungserfolg und Chancengleichheit. Insofern muss das Bildungswesen möglichst gute Rahmenbedingungen dafür schaffen, dass vor allem Kinder, deren Muttersprache nicht Deutsch ist, möglichst frühzeitig die deutsche Sprache lernen können, ohne dabei ihre Familiensprachen zu verlieren, und alle Kinder und Jugendlichen beim Erwerb weiterer Sprachen erfolgreich unterstützt werden. [...] Mit der Maßnahme des Bundesministeriums für Bildung und Forschung (BMBF) zur Förderung von Forschung im Bereich "Sprachliche Bildung und Mehrsprachigkeit" wird der Fokus auf die Potenziale gesellschaftlicher und individueller Mehrsprachigkeit gerichtet. Ziel ist es, Wissen über erfolgreiche Prozesse der Entwicklung von Mehrsprachigkeit in Kindertageseinrichtungen sowie in Schulen der Primarstufe und der Sekundarstufe I zu erarbeiten und bereitzustellen.“ (URL: http://www.empirische-bildungsforschung-bmbf.de/de/1788.php (Zugriff am 19.08.15))</p>	<p>Sprachwissenschaften, Erziehungswissenschaften, Psychologie</p>	<p>Längsschnittuntersuchungen, Interventionsstudien, Videoanalysen, Experimente, quantitative Erhebungen, Interviews</p>
<p>NIL: Lehr-Lern-Forschung unter neurowissenschaftlicher Perspektive Förderzeitraum: 2008 - 2012</p>	<p>„Systematische Kooperationsstrukturen zwischen den unterschiedlichen Disziplinen der Lehr-Lern-Forschung und den Neurowissenschaften waren 2005 zu Beginn der Fördermaßnahme „Lehr-Lern-Forschung unter neurowissenschaftlicher Perspektive (NIL)“ noch nicht etabliert. Das BMBF verfolgte daher mit der Fördermaßnahme das Ziel, eine neurowissenschaftliche Perspektive in die Lehr-Lern-Forschung einzubeziehen und dadurch einen zusätzlichen evidenzbasierten Ansatz für die Weiterentwicklung institutionalisierter Lerngelegenheiten zu schaffen. Die Fördermaßnahme zielte insbesondere darauf ab, durch Forschungsprojekte empirisches Wissen hervorzubringen, das dazu beitragen kann, Kindern, Jugendlichen und Erwachsenen, die von Störungen im Bereich des Lesens, Rechtschreibens und Rechnens betroffen sind, eine individuelle, ursachenbezogene Diagnostik und evidenzbasierte Förderung zu ermöglichen.“ (URL: http://www.empirische-bildungsforschung-bmbf.de/de/1919.php (Zugriff am 19.08.15))</p>	<p>Neurowissenschaften, Psychologie, Sprachwissenschaften</p>	<p>neurowissenschaftliche Untersuchungen, computergestützte Verfahren, Leistungstests, Längsschnittuntersuchungen</p>
<p>SteBis: Steuerung im Bildungssystem Förderzeitraum: 2010 - 2016</p>	<p>In den Ländern der Bundesrepublik Deutschland wurden in jüngerer Zeit zahlreiche Maßnahmen initiiert, die dem Ziel dienen, die Qualität des Bildungssystems und seiner Ergebnisse weiterzuentwickeln und nachhaltig zu sichern. [...] Diese Innovationen mögen Grundlagen und förderliche Bedingungen für eine verbesserte Systemsteuerung schaffen; ob sie aber erfolgreich im Sinne der definierten Ziele sind, ist eine empirische Frage. Innovationen sind keine "Selbstläufer", die den erhofften Ertrag automatisch erbringen. Vielmehr müssen sie im Bildungssystem auf verschiedenen Systemebenen verstanden, aufgegriffen und in neue Entwicklungsmaßnahmen umgesetzt werden. Angesichts ihrer Bedeutung im Rahmen von Umstrukturierungsprozessen auf allen Ebenen des Systems soll durch Forschungen im Kontext des Rahmenprogramms zur Förderung der empirischen Bildungsforschung das Wissen um Einstellungen und Erfahrungen, Vorgänge, Prozesse und Wirkungen empirisch fundiert werden. Das durch diese empirische Forschung generierte Wissen soll einen Beitrag leisten zur Verbesserung von Steuerung im Bildungssystem. (URL: http://www.empirische-bildungsforschung-bmbf.de/de/1391.php (Zugriff am 19.08.15))</p>	<p>Erziehungswissenschaften, Wirtschaftswissenschaften, Soziologie</p>	<p>quantitative Erhebungen, Interviews, Längsschnittuntersuchungen, Leistungstests</p>

Quelle: Eigene Erstellung

Die folgende, detaillierte Beschreibung der Population der Projekte, für die der VFDB zuständig war, dient dazu, die berichteten Statistiken einordnen zu können. Zusammenfassend ist festzuhalten, dass es auf Basis der uns vorliegenden Informationen nicht möglich ist, die tatsächliche Anzahl empirischer Projekte mit Primärdaten zu bestimmen. Aus diesem Grund können keine Angaben zu relativen Anteile im Sinne von Erfolgsquoten oder Ähnlichem gemacht werden.

Projektdemographie

Insgesamt wurden im Rahmenprogramm 315 Projekte eigenständig gefördert. Jedes dieser Projekte ist eindeutig identifizierbar über das Förderkennzeichen (FKZ). Zwölf dieser Projekte sind keine empirischen Projekte, d. h. in diesen Projekten, wurden weder Forschungsdaten primär erhoben noch sekundär ausgewertet. Dies sind insbesondere die Koordinierungsstellen der jeweiligen Forschungsschwerpunkte. 303 Projekte bilden somit die Brutto-Grundgesamtheit, die vom Verbund Forschungsdaten Bildung zu kontaktieren war.

Die Angabe der Netto-Grundgesamtheit ist leider mit folgenden Schwierigkeiten behaftet. Bei den 303 empirischen Projekten handelt es sich zum Teil um „Einzelprojekte“, in denen jeweils ein eigener, abgegrenzter Datenbestand erhoben wurde. Zum Teil handelt es sich aber auch um Teilprojekte aus Forschungsverbänden. Dies ist relevant, da in diesen Fällen häufig ein gemeinsamer Datenbestand je Verbund und nicht ein einzelner Datenbestand je Teilprojekt erhoben wurde.¹⁶ Vergleichbares gilt für die Förderphasen: Die BMBF-Förderung erfolgt je Förderphase getrennt. Ein Projekt, das über zwei Förderphasen läuft, erhält somit zwei unterschiedliche FKZ und ist zweimal in der Grundgesamtheit von 303 Projekten enthalten.

Insgesamt wurden 67 empirische Projektverbände gefördert (gegenüber 237 Einzelvorhaben). Diese bestehen aus zwei bis fünf Teilprojekten. Ein Verbundvorhaben besteht aus fünf Teilprojekten, 6 Verbundvorhaben aus jeweils vier und 9 Verbände aus jeweils drei Teilprojekten. Die meisten Verbundvorhaben bestehen somit aus jeweils zwei Teilprojekten.

Berücksichtigt man die Zusammenschlüsse von Teilprojekten zu Verbänden, reduziert sich die Anzahl der Projekte im EBF-Rahmenprogramm von 303 auf 212. Berücksichtigt man hingegen mehrere Förderphasen nur einfach, ergibt sich eine Zahl von 278 Projekten. Leider lässt sich nicht abschließend sagen, ob ein gemeinschaftlicher Datenbestand über mehrere Teilprojekte und/oder Förderphasen hinweg erhoben wurde, da es hierzu der Auskunft der betreffenden Projektnehmer/innen bedürfte. Bei Projekten, die sich nicht beim VFDB melden, wird dauerhaft Unklarheit darüber bestehen, ob sie getrennt zu erfassen sind oder nicht. Die Angabe der Netto-Grundgesamtheit ist daher nicht möglich. In der Folge können die absoluten Zahlen (bspw. Anzahl Projekte, die eine Online-Meldung durchgeführt haben, oder Anzahl Projekte, die Daten übermittelt haben) nicht relativ, d.h. prozentual, angegeben werden.

Bislang ist bekannt, dass 18 Projekte keine Primärdatenerhebungen durchgeführt, sondern ausschließlich sekundäranalytisch gearbeitet haben. Somit beträgt die potentielle Grundgesamtheit an Projekten mit Primärdaten N=285 (vgl. Tabelle 20).

¹⁶ Die administrative Förderlogik richtet sich nach dem Zuwendungsempfänger. Jeder Zuwendungsempfänger wird eigenständig administriert und erhält eine eigene FKZ. Kooperationsprojekte über mehrere Standorte, d. h. Zuwendungsempfänger, hinweg, werden dieser Logik folgend getrennt erfasst.

Tabelle 20: Anzahl Projekte nach unterschiedlicher Zählweise

	Anzahl insgesamt	Anzahl empirisches Projekt	Anzahl Projekte mit Primärdaten	Anzahl Projekte bei Zusammenfassung mehrerer Förderphasen	Anzahl Projekte bei Zusammenfassung von Teilprojekten zu Verbänden
Projekte insgesamt	315	303	285	278	212

Stand 31.08.2016

Tabelle 21 gibt die Anzahl der Projekte nach Forschungsschwerpunkten differenziert wieder. Die Größe der Forschungsschwerpunkte variiert zwischen 10 Projekten (NIL) und 41 Projekten (Chancengerechtigkeit). Im Rahmen der Promotionsförderung wurden 73 Promotionsprojekte seit 2009 gefördert. Drei Forschungsschwerpunkte verfügen nicht über eine Koordinierungsstelle, weshalb sich bei diesen die Anzahl empirischer Projekte nicht von der Gesamtzahl geförderter Projekte im Schwerpunkt (ASCOT, Chancengerechtigkeit, NIL) unterscheidet. Die Forschungsschwerpunkte ASCOT, Chancengerechtigkeit, JEKI, FiSS und NIL wurden nur eine Förderphase lang gefördert. Der Schwerpunkt Mehrsprachigkeit befindet sich derzeit noch in seiner Förderphase.

Tabelle 21: Anzahl Projekte nach Forschungsschwerpunkten

	Anzahl insgesamt
Projekte insgesamt	315
<i>Projekte nach Forschungsschwerpunkten</i>	
ASCOT	21
Chancengerechtigkeit	41
ESF	34
FiSS	39
JEKI	19
Mehrsprachigkeit	17
NIL	10
ProPäda	37
SteBis	25
<i>Promo</i>	73

Jedes der 303 empirischen Projekte wurde vom VFDB angeschrieben und darum gebeten, Angaben zu Projekt und Studie im Online-Formular zu machen. Hieraus wurden standardisierte und kontrollierte Metadaten erstellt. Ziel war es, einen möglichst vollständigen Überblick der im Rahmenprogramm geförderten Projekte im Online-Portal forschungsdaten-bildung.de zu geben. Durch die Angaben im Online-Portal war ersichtlich, ob es sich bei den Projekten um solche handelt, die Primärdaten erheben oder solche, die ausschließlich sekundär bestehende Daten auswerten. Ausschließlich die primärdatenerhebenden Projekte bildeten die Grundgesamtheit für das zweite Ziel des VFDB: die Sicherung der im Rahmenprogramm generierten Forschungsdaten (vgl. Kapitel 3.2.3).

Tabelle 22 zeigt die Anzahl unterschiedlicher Personen, die im Rahmenprogramm ein oder mehrere Förderungen erhalten haben. 315 Projekten stehen 209 Projektleitungen gegenüber. 146 Personen leiten ein Projekt, 63 Personen leiten zwei oder mehr (bis zu sieben) Projekte.

Tabelle 22: Projekte je Person

	Insgesamt	mit 1 Projekt	mit 2 Projekten	mit 3 Projekten	mit 4 Projekten	mit 5 Projekten	mit 6 Projekten	mit 7 Projekten
Anzahl Personen:	209	146	38	15	5	2	2	1

Die empirischen Projekte des EBF-Rahmenprogramms wurden gefördert mit Beginn 01.04.2008 bis aktuell 31.10.2017 (aktuell spätestes Laufzeitende der empirischen Projekte). Projektnehmer/innen wurden vor Ende der Laufzeit angeschrieben. Projekte, die Ende 2016 ausgelaufen sind oder in 2017 auslaufen werden, werden erst in der zweiten Förderphase des VFDB kontaktiert.

7.2. Kurzbeschreibung des Ablaufs der Archivierung beim VFDB

In der Pilotphase des Projektes konnte ein standardisierter Ablauf für die Sicherung und Bereitstellung von Forschungsdaten beim VFDB entwickelt werden. Dieser Ablauf ergab sich aufgrund der spezifischen, gegebenen Anforderungen und Rahmenbedingungen des Projektes. Der Ablauf ist im Folgenden beschrieben:

1. Schritt: Der VFDB schreibt die Projekte an und bittet darum, dass diese in einem hierfür speziell entwickelten Online-Formular¹⁷ Angaben zu ihrem Projekt (zu Studien und ggf. generierten Daten) machen.

In dem vom VFDB entwickelten Online-Formular werden Metadaten auf potentiell drei Ebenen erfasst: Ebene 1 enthält Informationen zum Projekt (administrative Metadaten, Abstract, Publikationen). Diese wurden aus einer bestehenden Datenbank in die Verwaltungsdatenbank des VFDB importiert. Diese Angaben sind vorausgefüllt im Online-Formular und müssen von den Projektnehmer/innen nur noch bestätigt oder korrigiert werden. Auf Ebene 2 werden Metadaten zur Studie oder den Studien erfasst (Untersuchungsdesign, Stichprobenauswahl etc.), auf Ebene 3 Metadaten zu den in den Studien generierten Forschungsdaten (Art der Daten, Erhebungsmethode etc.). Eine solch modulare Struktur wurde entwickelt, da sich die Projekte im Rahmenprogramm hinsichtlich ihrer Strukturen und in ihrer Komplexität stark unterscheiden. So gibt es Projekte, in denen ein Datentyp generiert wurde und solche, in denen mehrere Studien und mehrere Datentypen generiert wurden. Im Online-Formular können beliebig viele Studien- und Datentypen eingetragen werden.

Nach dem Ausfüllen und Absenden des Online-Formulars durch den/die Projektnehmer/in werden die eingetragenen Metadaten intern überprüft und ggf. angereichert (vgl. Kapitel 3.2.2).

Ziel der Erfassung der Projekte in diesem Schritt ist es, sämtliche Projekte des Rahmenprogramms über das Online-Portal forschungsdaten-bildung.de sichtbar zu machen. Interessierte können sich anhand des Portals einen Überblick über die Forschungsaktivitäten im Rahmenprogramm verschaffen. Perspektivisch soll das Portal ausgebaut werden und auch Studien außerhalb des EBF-Rahmenprogramms erfassen.

¹⁷ vgl. <http://melden.forschungsdaten-bildung.de/User/Login>

2. Schritt: Projekte, die lt. Online-Formular Primärdaten erhoben haben, werden gefragt, ob sie ihre Forschungsdaten beim VFDB sichern und verfügbar machen möchten. Die Übermittlung der Forschungsdaten zur Sicherung wird vertraglich geregelt.

Die Sicherung von Forschungsdaten im VFDB erfolgt auf der rechtlichen Basis der sogenannten *Übergabvereinbarung*. Dies ist ein zwischen den Partnern abgestimmter Vertrag, der die Sicherung der Forschungsdaten beim VFDB regelt. Die Projektnehmer/innen erklären in dieser Vereinbarung, dass der Vertragsgegenstand (die Forschungsdaten) frei von datenschutz- und urheberrechtlichen Ansprüchen Dritter ist.

3. Schritt: Übermittlung der Daten an den VFDB: Nach Vertragsabschluss werden die Daten und Materialien an den VFDB übermittelt und durch diesen in geschützter Umgebung aufbewahrt.

Derzeit erfolgt der Upload direkt zu einem externen Host, der Gesellschaft für wissenschaftliche Datenverarbeitung mbH (GWDG, Göttingen). Dort werden die Forschungsdaten verschlüsselt und gesichert. Der Upload erfolgt über eine sichere Verbindung (mittels SFTP-Client) und passwortgeschützt.

Nach Eingang der Forschungsdaten werden diese anhand archivarisch-dokumentarischer Kriterien durch die jeweils zuständigen Verbundpartner geprüft (vgl. Kapitel 3.2.3).

4. Schritt: Bereitstellung der Forschungsdaten zur Nachnutzung durch Dritte

Eine Bereitstellung der Forschungsdaten zur Nachnutzung durch Dritte wird dezentral von den jeweiligen Verbundpartnern bearbeitet. Ob Daten zur Sekundärnutzung bereitgestellt werden, wird nach Prüfung der datenschutzrechtlichen und ethischen Aspekte sowie der archivarischen Qualität durch die einzelnen Verbundpartner in Abstimmung mit den Projektnehmer/innen entschieden. Sie beruht auf einer weiteren vertraglichen Vereinbarung zwischen dem jeweiligen, datenhaltenden Institut (DIPF, GESIS oder IQB) und dem/der Projektnehmer/in.¹⁸

Entsprechend der Bewilligung des Projekts „Verbund Forschungsdaten Bildung“ werden im Rahmen der ersten Förderphase nur ausgewählte Datenbestände zur Sekundärnutzung bereitgestellt (DIPF 3-4, IQB 3-5 Studien, GESIS sämtliche, eingehende Studien).

Für die Projektnehmer/innen des EBF-Rahmenprogramms besteht die Möglichkeit, ihre Daten an anderer Stelle als dem VFDB zu sichern und bereitzustellen oder Daten nur zu sichern und nicht bereitzustellen. In jedem dieser Fälle sollte der Datenbestand im Portal www.forschungsdaten-bildung.de dokumentiert werden. Ziel ist, dass das Portal über den Gesamtbestand der Daten des Rahmenprogramms an zentraler Stelle informiert.

¹⁸ Die Sicherung und die Bereitstellung in einem Vertrag zu regeln, erwies sich aus rechtlichen Gründen als nicht umsetzbar (der Verbund besteht aus drei unterschiedlichen juristischen Personen). Die Bereitstellung der Forschungsdaten wird dezentral bei den einzelnen Instituten administriert.

In der folgenden Förderphase wird der Ablauf umfassend überarbeitet. Ziel ist es, die Datenmeldung und Übermittlung nutzerfreundlicher zu machen und die Aufwände sowohl bei den Forschenden als auch bei den Forschungsdatenzentren zu reduzieren.

7.3. Arbeitsbereich Datenakquisition: erstellte Dokumente

Tabelle 23: Erstellte Dokumente des Bereichs Datenakquisition

<i>Dokumente</i>	<i>(Stand 30.09.2016)</i>	<i>Inhaltsbeschreibung</i>
Workflow-Kommunikation		Beschreibt den Ablauf der Kommunikation mit Projektnehmer/innen, enthält E-Mail-Vorlagen und Verweise auf relevante Dateien (incl. Checkliste)
Akquisitionsliste		xls-Datei zur Dokumentation der Akquise
Akquisefälle [intern]		Beschreibt die eingetretenen Fälle und Vorgehensweisen der Bearbeitung
Liste Stellungnahmen		xls-Datei zur Dokumentation der eingegangenen Stellungnahmen
Hinweise Stellungnahme		Dokument mit Hinweisen zur Interpretation der Stellungnahmen für den DLR Projektträger
Fälle Rechtsprüfungen		Eine Zusammenfassung der eingetretenen Fälle (Prüfungen von Einverständniserklärungen und behördlichen Genehmigungen), die mit der Rechtsanwaltskanzlei Goebel & Scheller besprochen wurden

7.4. Arbeitsbereich Dokumentation und Datenübernahme: erstellte Dokumente

Die Dokumente beziehen sich auf Arbeitsprozesse, die institutsübergreifend relevant sind und untereinander abgestimmt wurden.

Tabelle 24: Erstellte Dokumente des Bereichs Dokumentation und Datenübernahme

<i>Dokumente</i>	<i>(Stand 30.09.2016)</i>	<i>Inhaltsbeschreibung</i>
Beschreibung des Workflows „Dateneingang und Datentransfer“		Zusammenfassung der Arbeitsabläufe (Schwerpunkt Technik und Dokumentation)
Workflow „Eingangsprüfung“		Beschreibung der Eingangsprüfungen von Daten
Checkliste „Technisch-formale Eingangsprüfung“ und FDZ-spezifische bzw. Datentyp-spezifische Checklisten „inhaltliche Eingangsprüfung“		Faktoren, die bei der Eingangsprüfung durch die Bearbeiter/innen geprüft werden
Leitfaden „Funktionalitäten Verbunddatenbank“		Leitfaden zur Benutzung der Verbunddatenbank bei der Dokumentation der Projekte (Datenmanagementsoftware)
Metadatenset		Im VFDB abgestimmtes Metadatenset Projekt/Studie/Datenkollektion inkl. DDI Mapping
Liste kontrollierter Vokabularien		Zusammenstellung der kontrollierten Vokabularien für die Studiendokumentation
Übersicht Bearbeitungsstand der Dokumentation der VFDB-Projekte		Statistik zu allen bearbeiteten Projekten (eingegangenen Meldeformularen) mit Zuständigkeit, Datentyp, Status
Funktionskonzept externer Dienstleister GWDG (Hosting)		Technische Umsetzung des Hostings, Upload, Verschlüsselung
Leitfaden Abstracting		Leitfaden zum Abstracting der Studien im Web-Portal nach dokumentarischen Standards

7.5. Vertragliche Vereinbarungen und rechtliche Rahmenbedingungen

Tabelle 25: Vertragliche Vereinbarungen und rechtliche Rahmenbedingungen

Dokument (Stand 30.09.2016)	Link	(Zugriffe am 30.09.2016)
Übergabvereinbarung VFDB	www.forschungsdaten-bildung.de/files/VFDB_UebergabeVereinb_V7.docx	
Datengebervereinbarung GESIS-Datenarchiv	www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Archivierungsvertrag_GESIS_Datenarchiv_annotiert.pdf	
Datengebervereinbarung FDZ Bildung am DIPF	www.forschungsdaten-bildung.de/files/FDZBildung_Weitergabevertrag_V1.pdf	
Datengebervereinbarung FDZ am IQB	www.forschungsdaten-bildung.de/files/IQB_Nutzungsvertrag_Bereitstellung_VFDB.pdf	
Zugangsbedingungen FDZ Bildung am DIPF	www.fachportal-paedagogik.de/forschungsdaten_bildung/medien_informationen.php	
Zugangsbedingungen GESIS-Datenarchiv	www.gesis.org/unser-angebot/daten-analysieren/datenservice/benutzungsordnung/ (aktueller Link, Zugriff am 12.04.2017: www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/bgordnung_bestellen/Benutzungsordnung.pdf)	
Zugangsbedingungen IQB	www.iqb.hu-berlin.de/fdz/application	

7.6. Erfolgte oder geplante Veröffentlichungen der Ergebnisse

Publikationen

Ebel, Thomas, und Alexia Meyermann. 2015. Hinweise zur Anonymisierung von quantitativen Daten. *forschungsdaten bildung informiert* 3. www.forschungsdaten-bildung.de/files/fdb-informiert-nr-3.pdf (Zugegriffen: 08. Mai 2017).

Meyermann, Alexia, und Maike Porzelt. 2014. Hinweise zur Anonymisierung qualitativer Daten *forschungsdatenbildung informiert* 1. www.forschungsdaten-bildung.de/files/fdb-informiert-nr-1.pdf (Zugegriffen: 08. Mai 2017).

Pegelow, Lisa. 2016. *Forschungsdaten im Spannungsverhältnis von Datenschutz und Nachnutzung*. Unveröffentlichte Masterarbeit, Humboldt-Universität zu Berlin.

Trixa, Jessica, und Thomas Ebel. 2015. Hinweise zur Aufbereitung quantitativer Daten. *forschungsdaten bildung informiert* 4. www.forschungsdaten-bildung.de/files/fdb-informiert-nr-4.pdf (Zugegriffen: 08. Mai 2017).

Trixa, Jessica, und Thomas Ebel. 2015. „Einführung in das Forschungsdatenmanagement in der empirischen Bildungsforschung“. Zusammenfassung des Workshops im Rahmen des CESSDA-Trainingprogramms. *forschungsdaten bildung informiert* 2. www.forschungsdaten-bildung.de/files/fdb-informiert-nr-2.pdf (Zugegriffen: 08. Mai 2017).

Verbund Forschungsdaten Bildung. 2015. Checkliste zur Erstellung eines Datenmanagementplans in der Empirischen Bildungsforschung. www.forschungsdaten-bildung.de/files/fdbinfo_2.pdf (Zugegriffen: 08. Mai 2017).

Verbund Forschungsdaten Bildung, und Goebel & Scheller. 2015. Checkliste zur Erstellung rechtskonformer Einwilligungserklärungen bei der Erhebung personenbezogener oder -beziehbarer

Daten im Rahmen der empirischen Bildungsforschung. https://www.forschungsdaten-bildung.de/files/fdbinfo_1.pdf (Zugegriffen: 08. Mai 2017).

Verbund Forschungsdaten Bildung. 2015. Hinweise zur Erstellung eines Methoden-/Feldberichts. Version 1.1. *fdbinfo* 3. https://www.forschungsdaten-bildung.de/files/fdbinfo_3.pdf (Zugegriffen: 08. Mai 2017).

Publikationen in Vorbereitung:

Meyermann, Alexia, Reiner Mauer und Alexander Schuster (in Vorbereitung): Zur Auswahl von Forschungsdaten in den Sozialwissenschaften. Ein Diskussionsbeitrag.

Meyermann, Alexia, und Maike Porzelt (in Vorbereitung): Datenschutzrechtliche Anforderungen in der Empirischen Bildungsforschung – Erfahrungen aus der Praxis des Verbund Forschungsdaten Bildung.

Meyermann, Alexia, und Maike Porzelt (in Vorbereitung): Zum Einsatz informierter Einwilligungen in der empirischen Bildungsforschung – Erfahrungen aus der Praxis des Verbund Forschungsdaten Bildung.

Vorträge

Bambey, Doris: Datensicherung und -nutzung beim Verbund Forschungsdaten Bildung und dem Forschungsdatenzentrum Bildung“. Beitrag im Rahmen des Workshop des Forschungsschwerpunktes „Sprachliche Bildung und Mehrsprachigkeit“, Hamburg, 26.02.2014.

Doris Bambey, Poldi Kuhl, Reiner Mauer und Sebastian Wurster: Pilotprojekt „Verbund Forschungsdaten Bildung“. Beitrag beim FDI-Ausschuss des RatSWD, Berlin, 10.10.2014.

Gerecht, Marius, und Doris Bambey: Forschungsdatenzentren als Arche Noah für Forschungsdaten?! Runder Tisch bei der 3. Frankfurter Tagung zu Videoanalysen in der Unterrichts- und Bildungsforschung, Frankfurt, 18.02.-19.02.2015

Gerecht, Marius, und Alexia Meyermann: Workshop im Forschungsschwerpunkt „Sprachliche Bildung und Mehrsprachigkeit“: Aufbereitung, Archivierung, Nachnutzung qualitativer Daten. Hamburg, 10.07.2015.

Kuhl, Poldi: Forschungsdatenmanagement, Vortrag beim FoWiN-Vernetzungstreffen, Berlin, 06.05.2015.

Meyermann, Alexia, und Reiner Mauer: Zur Frage der Nachnutzungswürdigkeit von Forschungsdaten in der Bildungsforschung. 79. Tagung der Arbeitsgruppe für Empirische Pädagogische Forschung. Hamburg, 15. - 17.09.2014 (*Beitrag eingereicht und angenommen, Präsentation musste leider kurzfristig krankheitsbedingt abgesagt werden.*)

Meyermann, Alexia: Datenschutz und empirische Sozialforschung. Beitrag beim 2. Vernetzungstreffen des BMBF-Förderschwerpunkt „Förderung von Entwicklungs- und Erprobungsvorhaben zur pädagogischen Weiterbildung von Kunst- und Kulturschaffenden“, Berlin, 23.-24.03.2015.

Meyermann, Alexia: Vorstellung des Verbund Forschungsdaten Bildung. Beitrag auf der Zwischentagung des Forschungsschwerpunkts „Steuerung im Bildungssystem“, Berlin, 16.10.2015.

Meyermann, Alexia: Datenschutzrechtliche Aspekte in empirischen Forschungsprojekten. Beitrag beim 2. Vernetzungstreffen des BMBF-Förderschwerpunkt: „Förderung von Entwicklungs- und Erprobungsvorhaben zur pädagogischen Weiterbildung von Kunst- und Kulturschaffenden“, Berlin, 23.-24.03.2015.

Meyermann, Alexia: Archivierung und Nachnutzung qualitativer Forschungsdaten. Gießener Methoden-Werkstatt. Bildung-Kultur-Gesellschaft, Gießen, 20.02.2016.

Netscher, Sebastian, und Poldi Kuhl. 2016. Einführung in das Forschungsdatenmanagement. GEBF Tagung 2016: Erwartungswidrige Bildungserfolge über die Lebenspanne, Berlin, 09.03.2016.

Trixa, Jessica, und Poldi Kuhl: Einführung in das Forschungsdatenmanagement. GEBF Tagung 2015, Bochum, 11.03.2015.

7.7. Kurzevaluation

7.7.1. Planung und Ablauf des Vorhabens

Die Arbeiten erfolgten weitestgehend im Rahmen des Zeitplans. Verzögerungen im Ablauf ergaben sich vor allem bei der technischen Entwicklung. Zur Überbrückung wurden alternative Lösungen realisiert, so dass die Bedarfe der Forschenden bedient werden konnten.

Es bedurfte projektinterner Anpassungen des Ressourceneinsatzes aufgrund unerwarteter Aufwände in einzelnen Service- und Entwicklungsbereichen. Höhere Aufwände fielen für die Bereiche Betreuung und Beratung der Projektnehmer/innen, Dokumentation der Studien im Portal sowie der Bearbeitung rechtlicher Themen an.

Die Aufgabe des VFDB bestand darin, die Daten der Projekte des EBF-Rahmenprogramms zu sichern. Da der VFDB erst nach Beginn des Rahmenprogramms seine Arbeit aufnahm, war in vielen Fällen nur eine retrospektive Archivierung möglich. Das heißt, viele Projekte und Datenerhebungen waren zum Zeitpunkt des Beginns des VFDB bereits abgeschlossen. Die Ergebnisse des VFDB zeigen, dass eine retrospektive Archivierung nur bedingt erfolgreich sein konnte, da in den Projekten nicht die Voraussetzungen geschaffen wurden und werden konnten, die für eine Archivierung von Forschungsdaten erforderlich sind (vgl. auch Kapitel 5).

7.7.2. Zielerreichungsgrad

Die Vorgehensweisen und Ergebnisse der im Antrag formulierten Teilziele 1 „Datenrettung und Basiserschließung“ und 2 „Pilothafte Umsetzung von Maßnahmen zur Datennachnutzung und strategisches serviceorientiertes Nutzungskonzept“ wurden im Kapitel 3 des vorliegenden Berichts dargestellt. Teilziel 3 des Antrags befasste sich mit dem Thema „Transfer und Schulung zum Datenmanagement“ und wurde im Schlussbericht in Kapitel 4 behandelt. In der folgenden Tabelle sind die im Antrag genannten einzelnen Arbeitspakete je Teilziel aufgelistet und hinsichtlich des Zielerreichungsgrads beurteilt.

Tabelle 26: Teilziele, Arbeitspakete und jeweiliger Zielerreichungsgrad

Teilziele lt. Antrag	Siehe Kapitel	Zielerreichung
Teilziel 1: Datenrettung und Basiserschließung	Kapitel 3	
AP 1: Zentraler Nachweis der verteilt archivierten Datenbestände	3.2.2 „Information und Dokumentation (IuD)“	erreicht
AP 2: Gesamtverzeichnis aller Forschungsdaten	3.2.2, „Information und Dokumentation (IuD)“	erreicht
AP 3: Zentrale Koordination der Workflows, Verhandlungen mit den Projekten, datenbankgestützte Steuerung und Dokumentation der Prozesse	3.2.1, „Datenakquisition“	erreicht
AP 4: Entwicklung von Instrumenten und Musterlösungen zu rechtlichen Regelungsfragen und Aufbau einer Internet-Informationssite	4.2 (siehe Teilziel 3, Arbeitspaket 2)	erreicht
Teilziel 2: Pilothafte Umsetzung von Maßnahmen zur Datennachnutzung und strategisches serviceorientiertes Nutzungskonzept	Kapitel 3	
AP 1: Entscheidungsgrundlage für die pilothafte Umsetzung der Nachnutzung von ausgewählten Datenpaketen nach einem Jahr	3.2.3.2, 3.2.4.2, „Dateneingang“ und „Datenbereitstellung“	teilweise erreicht
AP 2: Vorlage eines strategischen Nachnutzungskonzeptes für eine breite, forschungsunterstützende Nachnutzung der Forschungsdaten nach zwei Jahren	„Datenbereitstellung“, verschoben in das Folgeprojekt	nicht erreicht
AP 3: Für Sekundäranalysen aufbereitete Datenpakete: GESIS: „vollständige Umsetzung“ DIPF: 3-4 AV-Daten IQB: 3-5 Datensätze	3.2.4.3, „Datenbereitstellung“	erreicht
Teilziel 3: Transfer und Schulung zum Datenmanagement	Kapitel 4	
AP 1: Entwicklung eines Trainingskonzepts für Wissenschaftler/innen in der empirischen Bildungsforschung	4.1	erreicht
AP 2: Entwicklung von Schulungsmaterialien und weiteren Informationsangeboten (bspw. Broschüren, Website)	4.1, 4.2 (zusammengefasst mit Teilziel 1, Arbeitspaket 4)	erreicht
AP 3: Durchführung von insgesamt vier Workshops in Jahr 2 bzw. 3	4.2	erreicht
AP 4: Präsentationen auf Tagungen der Fachgesellschaften, Tagungen BMBF-geförderter Projekte im Bereich der Bildungsforschung und weiteren passenden Veranstaltungen	4.2, 7.6	erreicht

Begründung bei fehlender Zielerreichung; Arbeitspakete 1 und 2 des Teilziels 2

Da sämtliche Daten, die an den VFDB übermittelt wurden und die rechtlichen, technischen und dokumentarischen Anforderungen erfüllten, zur Nachnutzung bereitgestellt werden konnten, war ein Konzept zur Auswahl von Forschungsdaten in dieser Förderphase nicht erforderlich. Es wurden jedoch Vorarbeiten zur Bewertung der Nachnutzungsfähigkeit von Forschungsdaten geleistet (vgl. Kapitel 3.2.4.2). Im Übrigen stand in der ersten Förderphase des VFDB die Akquisition von Forschungsdaten im Fokus der Arbeiten. Eine Reihe von Hürden konnten diesbezüglich identifiziert werden (vgl. Kapitel 3.3). Das darüber hinausgehende Thema der Nachnutzung von Daten wird Bestandteil verschiedener Arbeitspakete der kommenden Förderphase sein.

7.7.3. Ausstattung und Mittelverwendung

Die Fördersumme für das Verbundvorhaben betrug insgesamt 1,15 Millionen EUR (vgl. Tabelle 27). Der Großteil der Ausgaben (770T EUR) fiel am DIPF an, da dort die Koordinierungsstelle angesiedelt war. Durch das DIPF erfolgte die Auftragsvergabe zur Entwicklung der IT-Infrastruktur, bestehend aus dem Datenbanksystem (ca. 90T EUR), dem Hosting eingehender Forschungsdaten (ca. 70T EUR) sowie dem Webauftritt (Webdesign, Frontend- und Backendentwicklung: ca. 7T EUR Auftragsmittel, ca. 31T EUR Softwareentwicklung durch eigene Entwickler). Die Aufwände für die IT-Infrastruktur - ohne Personalkosten für die IT-Koordination - betrugen somit ca. ein Viertel (26%) der Gesamtmittel. Die Personalkosten des DIPF setzen sich insbesondere zusammen aus Personalmitteln für die Projektkoordination, die IT-Koordination, die zentrale Studiendokumentation sowie die Bearbeitung der eingegangenen Forschungsdaten. Das DIPF vergab einen Auftrag zur Rechtsberatung in Höhe von 46T EUR. Die Rechtsberatung wurde in Anspruch genommen für die Ausgestaltung der Verträge sowie für die juristische Beratung zu datenschutz- und urheberrechtlichen Aspekten im Zusammenhang mit Forschungsdaten.

Bei GESIS fielen Personal- (59T EUR) und Sachmittel (25T EUR inclusive Hilfskraftmittel) vor allem für die Durchführung der Workshops zum Forschungsdatenmanagement an (ca. 41% der Gesamtmittel). Beim IQB wurde der Großteil der Personal- (ca. 100T EUR) und Sachmittel (ca. 25T EUR Hilfskraftmittel) für die Bearbeitung der Forschungsdaten d. h. die Aufbereitung und Dokumentation eingehender Datensätze genutzt (ca. 70% der Gesamtmittel). Darüber hinaus wurde am IQB ein Programmierauftrag in Höhe von 17T EUR für die Anpassung der eigenen IT-Infrastruktur an die neuen Anforderungen vergeben.

Bei sämtlichen Partnern fielen Reisetätigkeiten in geringem Umfang an (DIPF: 7T EUR, GESIS: 2,4T EUR, IQB: 3T EUR). Die Reisemittel wurden für persönliche Projekttreffen sowie zur Vorstellung des Projektes bei Fachveranstaltungen verwendet.

Tabelle 27: Fördersummen je Teilprojekt*

	DIPF	GESIS	IQB
Personalmittel	483	162	123
Sach-/Auftragsmittel**	287	45	52
Fördersumme je Teilprojekt	770	207	175
		Fördersumme insgesamt	1152

*Angaben in 1000 EUR, gerundet

** incl. Programmpauschale und Hilfskraftmittel

7.7.4. Projektinterne Prozesse und Strukturen

Der VFDB als Kooperationsprojekt

Der VFDB war als Kooperationsprojekt mit der Aufgabe gestartet, ein institutsübergreifendes Infrastrukturangebot für Bildungsforscher/innen zu schaffen. Aufgrund der Neuartigkeit der durch den VFDB zu erfüllenden Aufgabe wurde eine enge Zusammenarbeit der drei Partner DIPF, IQB und GESIS angestrebt und realisiert. Die verschiedenen Arbeitspakete wurden jeweils im engen Austausch bearbeitet. Das Projekt hat dabei ein spezifisches Modell einer förderierten Infrastruktur, das Modell der „verteilten Archivierung“, erfolgreich erprobt. Es ist geplant, die Kooperationen mit weiteren

Forschungsdatenzentren in der kommenden Förderphase auszubauen und hierbei weitere Kooperationsformen umzusetzen.

Organisationsaufbau und -ablauf, Arbeitsprozesse

Der organisatorische Aufbau des Projektes ist in Kapitel 3.1.1 des Schlussberichts dargelegt. Innerhalb des VFDB waren die Zuständigkeiten hinsichtlich der Betreuung von Datenbeständen nach Datentypen (quantitativ mit und ohne Schulleistungstests, qualitativ) organisiert. Eine Besonderheit stellten Fälle dar, in denen Studien/Datenbestände aus mehreren Datentypen (z. B. Mixed-Methods-Studien bestehend aus quantitativen und qualitativen Daten) bestanden. Diese machten eine fallspezifische Abstimmung zwischen unterschiedlichen FDZ erforderlich, da die Zuständigkeit der FDZ datentypabhängig ist. In unserem Modell ist das FDZ Bildung für qualitative Daten und die Datenzentren der GESIS und des IQB für quantitative Daten zuständig. Daraus folgt, Datengeber/innen, die sowohl quantitative als auch qualitative Daten aus Mixed-Methods-Studien archivieren möchten, haben zwei Ansprechpartner und Vertragspartner auf Seiten der FDZ; spätere Datennutzer/innen müssen sich zur Sekundärnutzung des gesamten Datenbestands bei zwei FDZ registrieren. Hierin zeigt sich ein potentieller Konflikt zwischen dem Ziel der datentypgerechten Betreuung und dem Ziel der Nutzerfreundlichkeit, für das zukünftig Lösungen gefunden werden müssen. Im Förderzeitraum trat diese Problematik jedoch nur in sehr geringem Umfang auf (vier Fälle).

Eine weitere Herausforderung bestand darin, die Prozesse der Bearbeitung der Forschungsdaten (Dokumentation, Eingangsprüfung/Ingest) zu harmonisieren, um ein vergleichbares, hohes Niveau zu schaffen. Hierzu wurden gemeinsam Leitlinien der Bearbeitung definiert und vergleichbare sowie transparente Standards erarbeitet. Dabei zeigten sich Grenzen der Harmonisierung durch die datentypspezifisch unterschiedlichen Anforderungen und unterschiedliche organisatorische Rahmenbedingungen in den drei Instituten. Diese hatten jedoch keine negativen Folgen für die Zielerreichung im Projekt. Positive Effekte der Zusammenarbeit ergaben sich beispielsweise dann, wenn auf bestehendes Knowhow einzelner Institute zurückgegriffen werden konnte.

Die Forschungsdateninfrastruktur hat sich in den letzten Jahren sehr dynamisch entwickelt und verfügt in Teilen über einen geringen Grad an Standardisierung und Professionalisierung. Eine einheitliche Verwendung von Terminologien hat sich noch nicht durchgesetzt, ebenso wenig wie eine einheitliche oder vergleichbare Gestaltung von Arbeitsabläufen oder Organisationsstrukturen. Vor diesem Hintergrund ist die Zusammenarbeit der drei Datenzentren erfolgreich zu bewerten und trägt zur Verbreitung hoher professioneller Standards von FDZ-Tätigkeiten bei.

Die zukünftige Herausforderung für den VFDB wird in der Vergrößerung des Netzwerks und der Aufnahme weiterer Partner bestehen. Erfolgskritisch für die Arbeit des VFDB wird die weitere Professionalisierung der beteiligten FDZ hin zu sog. „vertrauenswürdigen Archiven“ sein, die unterschiedlich fortgeschritten ist.¹⁹

¹⁹ Das GESIS Datenarchiv besitzt das Data Seal of Approval und bereitet derzeit seine Zertifizierung mit dem nector Siegel vor, vgl. <http://www.gesis.org/unser-angebot/archivieren-und-registrieren/datenarchivierung/langzeitarchivierung/> (Zugegriffen: 09. Jan. 2017).

Personalplanung

Für die Aufgabenstellung des vorliegenden Projektes wurde eine Vielzahl unterschiedlicher Kompetenzen benötigt: sozial-, bildungs- und verhaltenswissenschaftliches Fach- und Methodenwissen (z. B. für die Kommunikation mit den Forschenden und die Interpretation der Daten), Kompetenzen im Datenmanagement (z. B. für die Datenaufbereitung), dokumentarische und bibliothekarische (z. B. für Katalogisierung, Metadaten) sowie informationswissenschaftliche Kompetenzen (z. B. für das Web-Retrieval), informationstechnologische Kompetenzen der Softwareentwicklung, der Systemarchitekturplanung und Systemadministration sowie juristische Expertise (für Fragen des Urheber-, Vertrags- und Datenschutzrechts). In Teilen können die geforderten Kompetenzen durch externe Auftragsvergabe abgedeckt werden (bspw. die juristische Expertise, IT-Expertise). Darüber hinaus stellt gerade die interdisziplinäre Zusammenarbeit eine besondere Herausforderung dar. Das vorliegende Projekt war mit der aktuellen Personalbesetzung exzellent aufgestellt, da diese das Spektrum der genannten Kompetenzen abdeckt. Zukünftig wird ein weiterer Aufbau von Kompetenzen angestrebt. Da die erforderlichen Kompetenzen derzeit auf dem Arbeitsmarkt nur sehr begrenzt verfügbar sind und gleichzeitig die Arbeitsnachfrage in diesem Gebiet ansteigt, stellt die Personalbindung für das Projekt einen besonders erfolgskritischen Faktor dar. Ausbildungsprogramme zum Data Scientist oder Data Manager existieren noch nicht im nennenswerten Umfang, so dass das vorliegende Projekt auf Mitarbeiter/innen mit entsprechender Berufserfahrung angewiesen ist.

Auftragsvergabe externe Dienstleister (Softwareentwicklung, Hosting, Rechtsanwälte)

Externe Aufträge für die IT-Softwareentwicklung des Datenbankmanagementsystems wurden an die Firma OPIT, für das Hosting von Dateien an die Firma GWDG sowie für die datenschutzrechtliche Beratung vergeben. Die datenschutzrechtliche Beratung durch die Rechtsanwaltskanzlei Goebel & Scheller erwies sich als sehr hilfreich. Unserer Einschätzung nach ist eine solche Rechtsexpertise für Projekte dieser Art unerlässlich. Aus diesem Grund wird eine solche Beratung auch in Zukunft in Anspruch genommen werden. Die an die externen Dienstleister OPIT und der Gesellschaft für wissenschaftliche Datenverarbeitung mbH (GWDG) übergebenen Aufträge wurden entsprechend der Zielvorgaben umgesetzt.

7.7.5. Zusammenarbeit mit anderen Stellen / Kooperationen

Die für die Projektentwicklungen maßgebliche Zusammenarbeit des VFDB fand im RatSWD statt. Das FDZ Bildung am DIPF sowie das FDZ am IQB sind Mitglied im Forschungsdateninfrastruktur-Ausschuss des RatSWD. In der Förderphase fand eine aktive Mitarbeit in verschiedenen Arbeitsgruppen statt: in der RatSWD-Task Force Forschungsdatenmanagement (DIPF, IQB), RatSWD-Arbeitsgruppe Handreichung Datenschutz (DIPF) sowie in der Unterarbeitsgruppe Evaluation und Reakkreditierung (DIPF).

Berichtsblatt

1. ISBN oder ISSN ---	2. Berichtsart (Schlussbericht oder Veröffentlichung) Schlussbericht
3. Titel Schlussbericht zum Verbundprojekt „Sicherung und Nachnutzung der Forschungsdaten des Rahmenprogramms zur Förderung der empirischen Bildungsforschung“	
4. Autor(en) [Name(n), Vorname(n)] Meyermann, Alexia, Bambey, Doris, Ebel, Thomas, Eisentraut, Marcus, Jansen, Malte, Harzenetter, Karoline, Kuhl, Poldi, Mauer, Reiner, Neuendorf, Claudia, Pegelow, Lisa, Porzelt, Maike, Rittberger, Marc, Schwager, Thomas, Stanat, Petra, Trixa, Jessica;	5. Abschlussdatum des Vorhabens 30.09.2016
	6. Veröffentlichungsdatum ---
	7. Form der Publikation ---
8. Durchführende Institution(en) (Name, Adresse) DIPF – Bildungsforschung und Bildungsinformation, Schloßstraße 29, 60486 Frankfurt GESIS – Leibniz-Institut für Sozialwissenschaften, Unter Sachsenhausen 6-8, 50667 Köln IQB (Institut zur Qualitätsentwicklung im Bildungswesen), Humboldt Universität zu Berlin, Unter den Linden 6, 10099 Berlin	9. Ber. Nr. Durchführende Institution ---
	10. Förderkennzeichen 01JS1302A-C
	11. Seitenzahl 68
12. Fördernde Institution (Name, Adresse) Bundesministerium für Bildung und Forschung (BMBF) 53170 Bonn	13. Literaturangaben 31
	14. Tabellen 27
	15. Abbildungen 5
16. Zusätzliche Angaben ---	
17. Vorgelegt bei (Titel, Ort, Datum) ---	
18. Kurzfassung Ziel des Verbundprojektes war es, die Forschungsdaten aus dem Rahmenprogramm zur Förderung der empirischen Bildungsforschung des BMBF zu sichern und im Rahmen der bestehenden Möglichkeiten zur Nachnutzung verfügbar zu machen. Die Aufgabe bestand darin, eine Vielzahl an Projekten aus dem interdisziplinären Forschungsgebiet der Bildungsforschung mit einer entsprechend großen methodischen Vielfalt und heterogenen Daten zu betreuen. Für das Vorhaben haben sich die Datenzentren der drei Institute DIPF, GESIS und IQB kooperativ zum Verbund Forschungsdaten Bildung (VFDB) zusammengeschlossen und in der ersten Förderperiode (10/2013 bis 09/2016) eine Struktur der verteilten Archivierung aufgebaut. Diese Archivierungsform verbindet einen nutzerfreundlichen Zugang für Forschende mit einer fachspezifischen Betreuung der in der Bildungsforschung generierten, heterogenen Forschungsdaten. Die in der Pilotphase entwickelten Services ermöglichen es den Forscherinnen und Forschern, ihre Daten online zu melden und an den VFDB zu übermitteln. Die Forschungsdaten werden anschließend in den beteiligten Datenzentren archiviert, dort aufbereitet und soweit möglich für Dritte verfügbar gemacht. Die Studien der empirischen Bildungsforschung des Rahmenprogramms sowie die verteilt archivierten Forschungsdaten sind über das Portal forschungsdaten-bildung.de an zentraler Stelle sichtbar und auffindbar. Die Angebote zur Sicherung und Bereitstellung der Daten wurden im Projekt begleitet von Informationsangeboten, Beratungsleistungen sowie Schulungsaktivitäten zum Thema Forschungsdatenmanagement. Diese ermöglichten es Forschenden, sich umfassend und frühzeitig im Projektverlauf zum Thema Forschungsdatenmanagement zu informieren und zu qualifizieren. Im Schlussbericht des Projektes wird zunächst die Ausgangssituation des Vorhabens ausgeführt. Anschließend werden die Prozesse und Ergebnisse im Hinblick auf die Meldung, Sicherung und Bereitstellung der Forschungsdaten sowie die Informations-, Beratungs- und Schulungsangebote des VFDB zum Forschungsdatenmanagement beschrieben. Hürden für die Sicherung und Bereitstellung der Forschungsdaten im EBF-Rahmenprogramm, mit denen sowohl die Forschenden als auch die Forschungsdatenzentren im Projektverlauf konfrontiert waren, werden diskutiert. Im Ausblick wird abschließend auf bestehende Desiderate und die zukünftige Fortführung des Projektes eingegangen.	
19. Schlagwörter Empirische Bildungsforschung; Forschungsdaten; Infrastruktur; Forschungsdatenzentrum; Archivierung; Dokumentation; Nachnutzung; Data Sharing; Sekundäranalyse; Forschungsdatenmanagement	
20. Verlag ---	21. Preis ---

Document Control Sheet

1. ISBN or ISSN ---	2. type of document (e.g. report, publication) Final report
3. title Final report of the cooperation project „Archiving and Re-use of Research Data from the Framework Programme for the Promotion of Empirical Educational Research“	
4. author(s) (family name, first name(s)) Meyermann, Alexia, Bambey, Doris, Ebel, Thomas, Eisentraut, Marcus, Jansen, Malte, Harzenetter, Karoline, Kuhl, Poldi, Mauer, Reiner, Neuendorf, Claudia, Pegelow, Lisa, Porzelt, Maike, Rittberger, Marc, Schwager, Thomas, Stanat, Petra, Trixa, Jessica;	5. end of project 30 th of Sep 2017
	6. publication date ---
	7. form of publication ---
8. performing organization(s) (name, address) DIPF – Bildungsforschung und Bildungsinformation, Schloßstraße 29, 60486 Frankfurt GESIS – Leibniz-Institut für Sozialwissenschaften, Unter Sachsenhausen 6-8, 50667 Köln IQB (Institut zur Qualitätsentwicklung im Bildungswesen), Humboldt Universität zu Berlin, Unter den Linden 6, 10099 Berlin	9. originator's report no. ---
	10. reference no. 01JS1302A-C
	11. no. of pages 68
12. sponsoring agency (name, address) Bundesministerium für Bildung und Forschung (BMBF) 53170 Bonn	13. no. of references 31
	14. no. of tables 27
	15. no. of figures 5
16. supplementary notes ---	
17. presented at (title, place, date) ---	
18. abstract This joint-venture project was set up to secure research data that were generated in the context of the BMBF Framework Programme for Empirical Educational Research – subject to existing possibilities the project also targeted the provision of these data for re-usage. The project team was commissioned to manage a large number of projects from the interdisciplinary field of educational research with its variety of methods and heterogeneous data. To this end, the data centres from three institutes, i.e. DIPF, GESIS and IQB joined in a co-operative alliance, the “Verbund Forschungsdaten Bildung“(VFDB). In the first funding phase (10/2013 until 09/2016), a distributed archiving structure was established. This type of archiving links user-friendly access for researchers with subject-specific management of heterogeneous data from educational research. Services were developed in the pilot phase that enable researchers to register their data online and transfer the data to the VFDB. The research data are subsequently archived by the participating research data centres, processed and if possible made available to third parties. A central access point, the portal forschungsdaten-bildung.de , ensured that the empirical educational studies from the Framework Programme and the distributed archived research data are visible and findable. Data storage and provision services were complemented by information, counselling and training activities on the topic of research data management. Researchers were thus enabled to become well-informed and qualified regarding research data management at an early stage of their research. The final report on the project sets out by describing the situation as it was. This is followed by a summary of processes and results regarding the registration, storage and provision of research data and the information, counselling and training services regarding research data management. The report also discusses obstacles the researchers but also the data centres faced while storing and providing the research data. The outlook finally highlights remaining desiderata and the envisaged continuation of the project.	
19. keywords empirical educational research; research data; infrastructure; research data centre; archiving; documentation; re-usage; data sharing; secondary analysis; research data management	
20. publisher ---	21. price ---