

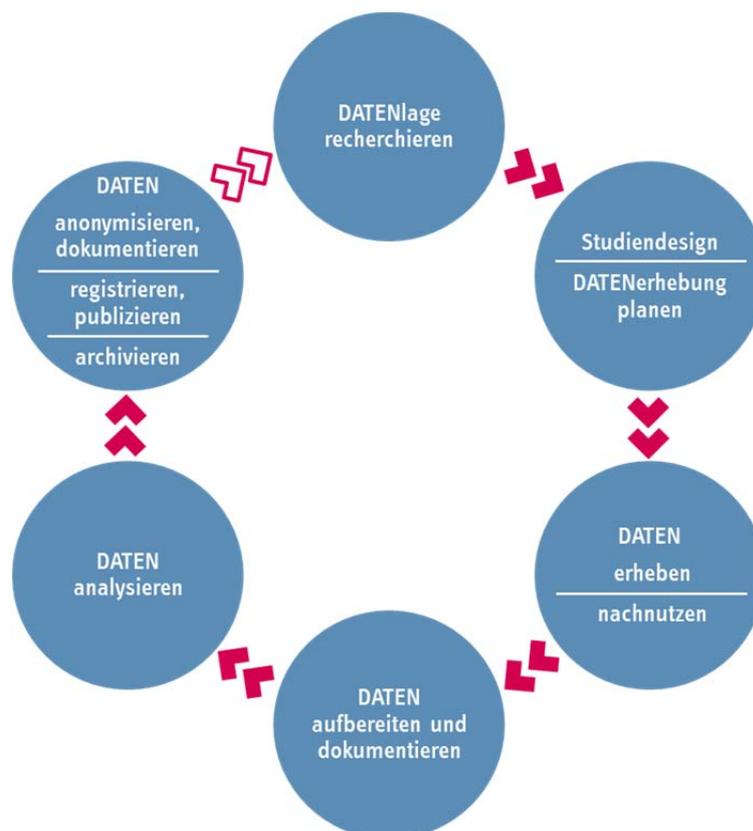
Checkliste zur Erstellung eines Datenmanagementplans in der empirischen Bildungsforschung

fdbinfo Nr. 2 // August 2015 // Version 1.1.

Bitte zitieren als: Verbund Forschungsdaten Bildung (2015): Checkliste zur Erstellung eines Datenmanagementplans in der empirischen Bildungsforschung. Version 1.1. fdbinfo Nr. 2.

1. Vorbemerkungen

Der Begriff Datenmanagement bezeichnet alle im Zusammenhang mit Forschungsdaten durchzuführenden Tätigkeiten, die sich entlang des Datenlebenszyklus ergeben. Die Tätigkeiten des Datenmanagements reichen von der Datenerhebung im Forschungsprojekt über die Datenbearbeitung (insb. Auswertungen) bis hin zur Archivierung der Forschungsdaten, wie sie beispielsweise durch ein Datenarchiv erfolgt. Berücksichtigt werden auch die Zuweisung von Verantwortlichkeiten und eine entsprechende Ressourcenplanung für die Implementation des Datenmanagementplans.



Forschungsdatenzyklus

Als Bestandteile des Datenmanagements werden im Allgemeinen unterschieden: Datenerhebung, Datendokumentation, Klärung der rechtssicheren und forschungsethisch unbedenklichen Verwendung von Forschungsdaten, die Datenweitergabe an Dritte zur Sekundärnutzung und die Auswahl der Daten zum längerfristigen Erhalt, zur Archivierung.

Bei der Erstellung eines Datenmanagementplans (DMP) geht es darum, diese Tätigkeiten möglichst zielorientiert, systematisch und effizient zu bestimmen. Der Nutzen einer derartigen, formalisierten Planung – unter Zuhilfenahme von Checklisten wie dieser – besteht darin, über eine frühzeitige und umfassende Planung, Beschränkungen in der späteren Datennutzung zu vermeiden. Solche Beschränkungen treten beispielsweise dann auf, wenn bestimmte Arbeiten mit/an den Daten (Analysen, Aufbereitung, Anonymisierung, Dokumentation, Archivierung, Verknüpfungen, Sekundärnutzungen etc.) nicht durchgeführt werden können, weil entweder

- » die Ressourcen fehlen,
- » die Qualität der Daten nicht ausreicht,
- » die rechtlichen Voraussetzungen nicht gegeben sind,
- » die Daten nicht interpretierbar sind oder
- » die Daten verloren gehen oder beschädigt sind.

Eine solche Datenmanagementplanung verfolgt *erstens* das Ziel einer effizienten, rechtssicheren und qualitätsorientierten Nutzung der Daten in den datengenerierenden Forschungsprojekten selbst. Ein weiteres Ziel besteht *zweitens* darin, die Replizierbarkeit bzw. Transparenz und Nachvollziehbarkeit der Forschungsergebnisse, die auf diesen Daten beruhen, zu gewährleisten und so wissenschaftlichen Standards, wie sie auch die allgemein anerkannten Richtlinien der DFG zur „Sicherung guter wissenschaftlicher Praxis“¹ beschreiben, umzusetzen. *Drittens* ist ein geplantes Datenmanagement eine notwendige Voraussetzung für die Nachhaltigkeit der Forschungsdaten und den Erhalt des kulturellen Wertes, den diese darstellen. Neue Forschungspotentiale ergeben sich für Sekundäranalysen mit neuen Forschungsfragen oder -methoden, für Zeit- und/oder Stichprobenvergleiche oder Verknüpfungen der Daten mit anderen Quellen. Ganz besonders relevant wird der Erhalt der Forschungsdaten für den Fall historisch einmaliger, nicht-wiederholbarer Datenerhebungen oder Datenerhebungen über schwer zugängliche Gruppen und seltene Phänomene.

Datenmanagementpläne werden zunehmend von Förderinstitutionen bei der Antragstellung gefordert. Auch stellen die DFG ebenso wie das BMBF inzwischen explizit Mittel für Tätigkeiten des Datenmanagements bereit.²

Best Practices: Um für die Forschung hilfreich sein zu können, sollten Datenmanagementpläne (DMP) **frühzeitig** erstellt werden, d. h. am besten bereits bei Projektplanung und Antragstellung. DMP sind keine statischen, sondern **dynamische** Dokumente, die an den laufenden Forschungsprozess anzupassen sind. Ein DMP sollte möglichst **umfassend** sein, da sein Vorteil darin liegt, die verschiedenen datenrelevanten Tätigkeiten im Datenlebenszyklus in ihren Wechselwirkungen zu erfassen. Eine

¹ Die DFG-Grundsätze „Sicherung guter wissenschaftlicher Praxis“ beinhalten eine zehnjährige Aufbewahrungspflicht (1. Fassung 1998, 2. Fassung 2013).

http://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/gwp/, Zugriff am 14.07.2015

² Vgl. BMBF-Bekanntmachung vom 10.10.2012, www.bmbf.de/foerderungen/20319.php, Zugriff am 07.05.2015 und DFG-Vordruck 54_01 – 06/14, S.5f., www.dfg.de/formulare/54_01/54_01_de.pdf, Zugriff am 01.07.2015.

Orientierung an allgemeinen Richtlinien ist möglich, die konkrete Ausgestaltung ist jedoch stets von den Bedingungen des jeweiligen Forschungsprojekts abhängig. DMP sollen keine „reinen Pflichtveranstaltungen“ sein, sondern den Forschenden eine Hilfestellung bieten. Workshops zum Thema Datenmanagement bietet GESIS CESSDA Training³, für Information und Beratung stehen auch die Mitarbeiter/-innen von Forschungsdaten Bildung zur Verfügung und Informationstexte und Material zum Download sind auf www.forschungsdaten-bildung.de verfügbar.

Der vorliegende DMP wurde erstellt in Anlehnung an

- DCC. 2013. Checklist for a Data Management Plan. v.4.0. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/data-management-plans>;
- ICPSR. o.J. Guidelines for Effective Data Management Plans. <http://www.icpsr.umich.edu/files/datamanagement/DataManagementPlans-All.pdf>, Zugriff am 27.05.2015.

2. Administrativ-formale Angaben

Ziel: Den administrativen und inhaltlichen Hintergrund des datenerhebenden Forschungsprojekts beschreiben.

Element	Beschreibung
ID	Ein permanenter Identifikator, wie er vom Förderer oder der Institution vergeben wurde. Falls mit anderen Projekten gemeinschaftlich ein DMP erstellt wird: Nennung aller Förderkennzeichen.
Förderer bzw. Eigenfinanzierung	Angabe des Förderers, falls vorhanden.
Projekttitle	Offizieller Name des Projektes, ggf. zusätzlich Akronym und Kurztitel.
Projektbeschreibung	Inhalt und Art des Projektes. Welche Forschungsfragen werden adressiert? Zu welchem Zweck werden Daten erhoben?
Forscher	Name des/der Projektleiter/s oder Primärforscher/s; ggf. zusätzlich Forscher-ID, wie ORCID (http://orcid.org).
Kontakt	Name (falls unterschiedlich zu Projektleitung/Primärforscher), Telefon und Email.
Erstellungsdatum	Datum der ersten Version dieses DMPs.
Änderungsdatum	Datum der letzten Änderung dieses DMPs.
Zu beachtende Vorgaben von Dritten	Verweis auf etwaige Richtlinien zum Umgang mit Forschungsdaten von der eigenen Einrichtung, dem Berufsverband, dem Fachverband, dem Förderer, der datenhaltenden Einrichtung, dem Archiv etc. Eine Auflistung der Anforderungen aller relevanten Parteien (Universität, Institut, Förderer, Öffentliche Stellen, Auftragnehmer etc.) in Zusammenhang mit Datenmanagement und

³ www.gesis.org/en/archive-and-data-management-training-and-information-center/training-center-home/, Zugriff am 10.09.2015

 Datenbereitstellung.

3. Datenerhebung: Daten generieren und bearbeiten

Ziel: Erfassen, welche Daten im Projekt generiert und verarbeitet werden und welche Anforderungen sich hieraus für das Datenmanagement ergeben.

Leitfragen: Welche Art von Daten werden erhoben, wie werden diese erhoben und wie verarbeitet? Zu welchen Zwecken werden Daten erhoben?

Element	Beschreibung
Datenbeschreibung	Beschreibung des Datenmaterials, das erhoben wird.
Vorhandene Daten	Recherche existierender Datenbestände, die für das Projekt relevant sind. Diskussion, ob und wie diese Daten genutzt werden könnten. <i>Hinweis: In diesem Vergleich zeigt sich der Wert der geplanten Datenerhebung.</i>
Qualitätssicherung	Beschreibung der Maßnahmen, die getroffen werden, um eine hohe Datenqualität zu erreichen und zu erhalten. <i>“Producing data of high quality is essential to the advancement of science”(ICPSR⁴)</i>
(geplante) Datenauswertung und Datenverarbeitung	Beschreibung der geplanten Auswertungen und des Zwecks der Datenverarbeitung. Beschreibung der Verarbeitungsschritte der Rohdaten (z. B. Transkription, Bereinigung, Anonymisierung). <i>Hinweis: Verwendungszwecke personenbezogener Daten werden beispielsweise für die Erstellung von Einwilligungserklärungen benötigt.</i>
Publikation	Sofern möglich, kurze Beschreibung der geplanten Publikation der Daten und Datenauswertungen (z. B. Zitate aus Transkripten in Forschungsartikeln, Ausschnitte aus Videos in Lehre und Weiterbildung).

⁴ Für dieses und die folgenden Zitate des ICPSR gilt folgende Quelle:

<http://www.icpsr.umich.edu/files/datamanagement/DataManagementPlans-All.pdf>, Zugriff am 13.06.2015

4. Datenspeicherung: Daten und Dateien speichern, organisieren und sichern

Ziel: Physischer Erhalt der Daten und deren Unversehrtheit (Schutz vor Verlust und Manipulation)

Leitfrage: Welche Maßnahmen sind zu ergreifen, um die Daten vor versehentlicher oder absichtlicher Manipulation, vor Verlust oder unbefugten Zugriffen zu schützen?

Element	Beschreibung
Speicherung und Backup	<p>Angabe der Speicher- und Backup-Verfahren, der Orte der Speicherung, erwartete Speicherkapazitäten;</p> <p><i>Hinweis: Wo und wie werden Kopien der Dateien aufbewahrt? Wie viele werden aufbewahrt und wie werden diese synchronisiert?</i></p>
Datenorganisation	<p>Beschreibung, wie die Daten im Projekt verwaltet werden, Festlegen von Richtlinien zur Benennung von Dateien, zur Versionskontrolle etc.</p> <p><i>Hinweis: Derartige Maßnahmen erleichtern das Identifizieren, Auffinden und Nachvollziehen von Dateien; insbesondere bei Arbeiten über längere Zeiträume in Teams von Bedeutung.</i></p>
Formate	<p>Angabe der Formate, in denen die Daten generiert, aufbewahrt und verfügbar gemacht werden sollen;</p> <p>Erläuterung, inwieweit diese Formate für die Arbeit mit den Daten und deren längerfristige Aufbewahrung geeignet sind;</p>
Sicherheit	<p>Beschreibung der technischen und organisatorischen Schutzmaßnahmen der Daten (Verschlüsselung, Passwortschutz, Firewall u. a.), insbesondere der vertraulichen Informationen;</p> <p>Beschreibung des Rechtemanagements (Zugriffsrechte und -restriktionen, Embargos);</p> <p><i>Ggf. Beratung mit dem IT-Service der Einrichtung.</i></p>



5. Datendokumentation: Daten dokumentieren und Metadaten erstellen

Ziel: Sicherstellen der Auffindbarkeit und Identifizierbarkeit der Forschungsdaten sowie deren Interpretierbarkeit (über das Projektende hinaus und auch durch nicht am Projekt beteiligte Dritte).

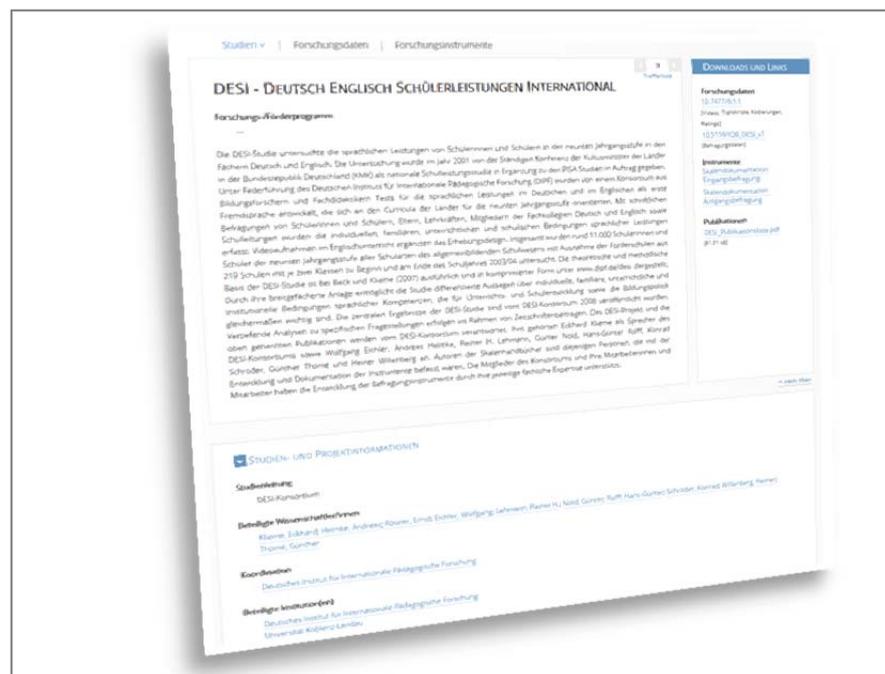
Leitfrage: Welche Informationen werden benötigt, um Daten finden und nachnutzen zu können?

„Metadata are often the only form of communication between the secondary analyst and the data producer“ (ICPSR)

Element	Beschreibung
Dokumente	<p>Auflistung aller Dokumente, die erstellt werden und zum Verständnis der Daten erforderlich sind. Dazu gehören beispielsweise:</p> <ul style="list-style-type: none"> » Codebücher, Transkriptionsmanuale, Interviewschemas » Erhebungsinstrumente (Fragebögen, Leitfäden, Protokolle) » Intervieweranweisungen, Kontaktprotokolle, Kontextnotizen » Einverständniserklärungen, Anonymisierungsmaßnahmen » QDA-Dateien, Syntax
Standardisierte Metadaten (optional)	<p>Falls gegeben: Angabe, welche Metadaten mit den generierten Daten erstellt werden (z. B. welches standardisierte Formular zur Metadatenerfassung genutzt wird; Formulare werden bspw. von Archiven oder Registrierungsagenturen herausgegeben).</p>

Hinweis: Anhand von Metadaten werden Forschungsdaten näher beschrieben. Standardisierte Metadaten werden erstellt, um Forschungsdaten in Datenbanken und Portalen verwalten zu können und auffindbar zu machen. Sie sind in der Regel maschinenlesbar und interoperabel.

Datennachweis anhand von Metadaten im Portal forschungsdaten-bildung.de



6. Legitimität: Übereinstimmung mit ethischen und rechtlichen Normen

Ziel: Rechtskonformität der Datennutzung und Schutz der Persönlichkeitsrechte aller Beteiligten; datenschutzrechtliche und ethische Anforderungen einhalten; urheberrechtliche Ansprüche wahren.

Leitfrage: Welche rechtlichen Anforderungen bestehen? Wie können diese eingehalten werden?

Element	Beschreibung
Urheberrechte	<p>Angabe, ob und welche Personen Urheberrechte an den Daten oder dazugehörigen Materialien besitzen. Angaben, wie Urheberrechte, falls erforderlich, geschützt werden. Nennung der Materialien, die dem Urheberrechtsschutz unterliegen.</p> <p><i>Hinweis: Urheberrechte können beispielsweise für eingesetzte oder entwickelte Instrumente existieren. Um Urheberrechte zu wahren und Nachnutzungen zu regeln, können Forschungsdaten und -instrumente lizenziert werden.</i></p>
Ethik und Datenschutz (Anonymität)	<p>Erläuterung, wie die informierte Einwilligung der Beforschten zur geplanten Datennutzung eingeholt wird. Erläuterung, wie die Beforschten vor einer möglichen Deanonymisierung geschützt werden. Beschreibung besonderer ethischer Fragen (bspw. bei sensiblen Forschungsthemen oder besonders schutzwürdigen Populationen) und wie damit umgegangen wird.</p> <p><i>Hinweis: Beachten Sie die Anforderungen an das Einholen informierter Einwilligungen, für den Fall, dass Forschungsdaten archiviert und nachgenutzt werden sollen.</i></p> <p><i>Ggf. Beratung mit dem Datenschutzbeauftragten der Einrichtung.</i></p>
Behördliche Genehmigungen	<p>Erläuterungen, ob behördliche Genehmigungen für die Durchführung der Erhebung (bspw. an Schulen) eingeholt werden müssen und falls ja, wie dies erfolgen wird (Ressourcen, Zeitrahmen, Arbeitsschritte).</p> <p><i>Hinweis: Achten Sie – sofern möglich – darauf, dass die Genehmigung nicht den DFG-Regeln guter wissenschaftlicher Praxis zur Aufbewahrung der Daten für 10 Jahre oder den sonstigen, geplanten Verwendungszwecken der Forschungsdaten widerspricht.</i></p>

7. Data Sharing - Daten teilen: Daten Dritten zugänglich machen

Ziel: Wege der Nachnutzung finden; Den Anforderungen an Replizierbarkeit, Transparenz und Nachvollziehbarkeit der eigenen Forschung nachkommen; Neue Forschung mit den generierten Daten ermöglichen.

Leitfrage: Welche Schritte sind zu unternehmen, um Dritten Zugang zu den Forschungsdaten zu ermöglichen?

Element	Beschreibung
Verfügbarkeit	Erläuterung, wie die Forschungsdaten Dritten zugänglich gemacht werden (Ort der Bereitstellung, Zugangsprozeduren, Sperrfristen, technische Umsetzung, offener oder restriktiver Zugang). Zeitraumen für das Verfügbarmachen der Daten.
Nachnutzungspotential	Einschätzung des Nachnutzungspotentials der Daten: Für welche Analyse Zwecke, welche Nutzergruppen könnten die Forschungsdaten jetzt oder in Zukunft interessant sein? <i>Hinweis: Sind die Daten beispielsweise für den Einsatz in der Lehre oder für die Bildungspraxis wichtig, sind entsprechende Vorkehrungen frühzeitig zu treffen.</i>



8. Datenerhalt: Daten auswählen und archivieren

Ziel: Langfristigen Erhalt der Forschungsdaten gewährleisten; Erhalt als kulturelles Erbe.

Leitfragen: Sollen die Forschungsdaten langfristig erhalten bleiben? Welche der vorliegenden Forschungsdaten sollen erhalten bleiben und wie?

„Digital data need to be actively managed over time to ensure that they will always be available and usable. This is important in order to preserve and protect our shared scientific heritage as technologies change.”(ICPSR)

Element	Beschreibung
Selektion und Aufbewahrungsdauer	Begründung, welche Daten archiviert werden sollen und welche nicht und wie lange die Daten aufbewahrt werden sollen. Pläne für etwaige Löschung der Daten in der Zukunft.
Archivierung und Erhaltung	Geplante Maßnahmen der Langzeitarchivierung und des dauerhaften Erhalts der Daten. Ausweichplan für den Fall, dass sich die ausgewählte, archivierende Einrichtung auflöst.

9. Verantwortlichkeiten und Ressourcen

Ziel: Implementation des Datenmanagementplans; Zuweisung ausreichender Ressourcen.

Leitfragen: Welche Ressourcen sind erforderlich für die Umsetzung des Datenmanagements? Wer ist für welche Maßnahmen verantwortlich?

Element	Beschreibung
Verantwortlichkeiten	Namen der Personen, die für das Datenmanagement im Forschungsprojekt verantwortlich sind.
Budget	Kosten der Vorbereitung der Daten (Aufbereitung, Anonymisierung, Dokumentation) für die Archivierung und wie diese Kosten getragen werden. Gegebenenfalls Informationen zur geplanten Beschaffung der erforderlichen Mittel.

Hinweis: Informieren Sie sich über die Möglichkeiten, Mittel für das Datenmanagement zu beantragen. Die DFG ebenso wie das BMBF ermöglichen die Beantragung solcher Mittel.
