



# Datenmanagementplan

## Volltextversion zum gleichnamigen Vortrag des Projektes FoDaKo<sup>1</sup> (Forschungsdatenmanagement in Kooperation)

Torsten Rathmann  
Bergische Universität Wuppertal  
Universitätsbibliothek  
Gaußstr. 20  
D-42119 Wuppertal  
ORCID: 0000-0001-5880-1546

### Zusammenfassung

Diese Informationsschrift beschreibt, wie ein Datenmanagementplan das Forschungsdatenmanagement erleichtern kann, welche Themen behandelt werden sollten und wie Datenmanagementplan, Meilensteinplan und Kostenplan zusammenhängen. Ein Abschnitt ist Software-Werkzeugen zur Aufstellung von Datenmanagementplänen gewidmet.

### Einführung

Ein Datenmanagementplan (DMP) ist ein Dokument, das beschreibt, wie Daten während eines Forschungsprojekts und nach Abschluss des Projekts behandelt werden<sup>2</sup>. Wie kann ein Datenmanagementplan helfen? Er kann eine verbindliche Grundlage für den gemeinsamen Umgang mit Forschungsdaten festlegen. Insbesondere, wenn mehrere Projektpartner zusammenarbeiten müssen, kann ein solches gemeinsames Dokument Verantwortlichkeiten definieren und Zugriffsrechte regeln. Er kann auch die Koordination der Projektpartner und die Definition der Arbeitsabläufe erleichtern. Darüber hinaus kann ein Datenmanagementplan dazu beitragen, Probleme in Bezug auf das Datenmanagement zu erkennen und Lösungen zu entwickeln. Er kann helfen, Datenverlust und Sicherheitslücken zu vermeiden.

Ein weiterer Zweck, der immer mehr an Bedeutung gewinnt, ist die Verwendung des Plans als Anhang zu einem Antrag auf Forschungsförderung. In den letzten Jahren erwarten immer mehr Institutionen, die Forschungsgelder anbieten, eine Datenmanagementstrategie zusammen mit dem Förderantrag. Eine solche Datenmanagementstrategie kann die erste Version eines Daten-

---

<sup>1</sup> <https://www.fodako.de/>

<sup>2</sup> [https://en.wikipedia.org/wiki/Data\\_management\\_plan](https://en.wikipedia.org/wiki/Data_management_plan), aufgerufen am 02.10.2018

managementplans sein. Ein Datenmanagementplan erleichtert so nicht nur den Umgang mit Forschungsdaten, sondern erhöht auch die Chance auf eine Förderung.

## Hauptziele eines Datenmanagementplans

Datenmanagementpläne sind so verschieden wie die Forschungsdaten selbst. Hauptziele können zum Beispiel sein<sup>3</sup>:

- eine Beschreibung der erwarteten Daten
- Strategien zum Schutz der Mitwirkenden und Verfahren zur Einholung der Einwilligung nach Aufklärung
- Auflistung von Rechten Dritter an den Daten (z.B. im Zusammenhang mit dem Datenschutz, Rechte an geistigem Eigentum)
- Bedingungen für die Nachnutzung
- eine Spezifikation der entsprechenden Metadaten
- Ordnungs- und Namenskonventionen
- Beschreibung der Workflows für Erzeugung, Ingest, gemeinsame Nutzung und Veröffentlichung von Forschungsdaten
- eine Planung der Langzeitarchivierung, z.B. ein Konzept für die Bitstream-Erhaltung und gegen unberechtigten Zugriff sowie ein Katalog von Aufgaben für die Datenpflege
- eine Spezifikation der Verantwortlichkeiten

## Ein kleines Beispiel: Wie Sie es nicht machen sollten

*Liebe Kolleginnen und Kollegen vom BMBF,*

*ich darf Ihnen versichern, dass Ihre Zusatzanforderung, unsere Datenmanagementstrategie schriftlich zu formulieren und unseren Antrag auf Forschungsförderung um einen zusätzlichen Anhang zu bereichern, in unserer Arbeitsgruppe große Freude ausgelöst hat. Hier die Details:*

*Unsere Primärdaten werden vom Messgerät in einem proprietären und kaum dokumentierten Binärformat erzeugt. Die — spärlichen — Metadaten stehen mit im Header dieser Dateien, so dass Daten und Metadaten auf das Innigste miteinander verbunden sind. Die Header bzw. Metadaten vervollständigen wir nicht und nennen die Gesamtheit der Header stattdessen „Sparse Density Header Matrix“ passend zu den anderen Worthülsen in unserem „Action Planning“.*

*Die Namen der Output-Dateien — ca. 50000 an der Zahl — werden vom Gerät automatisch in Form fortlaufender Nummern erzeugt und sind entsprechend vielsagend, zeigen die Lücken in der Nummerierung doch — bis auf Verwechslungen — wie viele Fehlversuche und Mühe es gekostet hat, zu diesen Daten zu kommen. Passend zu dieser Nummerierung mit arabischen Ziffern benennen wir die für Verständnis und Reproduzierbarkeit so wichtigen Steuer- und Kalibrierdateien mit Namen, die außer arabischen Ziffern auch arabische Schriftzeichen enthalten. Da Sie danach gefragt haben, an welche Normen wir uns halten, die Zeichenkodierung folgt strikt ISO*

---

<sup>3</sup> Teilweise übernommen von cessda, Research data management plan, [https://www.cessda.eu/content/download/239/2381/file/CESSDA%20User%20Guide%20for%20data%20management\\_2\\_Research%20data%20management%20plan.pdf](https://www.cessda.eu/content/download/239/2381/file/CESSDA%20User%20Guide%20for%20data%20management_2_Research%20data%20management%20plan.pdf), aufgerufen am 02.10.2018

8859-6 (Arabisch). Das ist für uns ganz praktisch, weil unser Programmierer Araber ist und wenig Deutsch und Englisch spricht.

Die Steuer- und Kalibrierdateien können von Nummer 1024 an aufwärts gefunden werden. Sollte sich ein potentieller Nachnutzer so weit vorgearbeitet haben: wir verwenden Steuerdaten, die im Format CSV (Comma Separated Values) vorliegen, und versehen diese mit zusätzlichen Kommas, damit sich Nachnutzer mit unserem Datenmodell eingehend auseinandersetzen und diese Daten nicht etwa nur maschinell einlesen.

Unsere Forschungsdaten speichern wir selbst auf Bandkassetten, die wir vor ca. 10 Jahren angeschafft haben. Von denen haben wir noch mehr als genug. Das Lesegerät lässt sich zwar nicht mehr in moderne Desktop-Computer einbauen, aber das ist kein Problem, da wir über einen großen Pool von Altrechnern verfügen.

Im Projekt ist auf Ihren Wunsch hin eine Qualitätskontrolle vorgesehen. Gegenüber dem Controller werden wir jeden Fehler zäh verteidigen, bis sich dieser beim besten Willen nicht mehr länger abstreiten lässt. Korrigierte Datensätze benennen wir ohne Änderungsmitteilung wie die vorangehende Version. Das ist zeitsparend und beunruhigt potentielle Nachnutzer nicht.

Unsere Zeitplanung ist so einfach wie effektiv. Das Forschungsdatenmanagement legen wir grundsätzlich in die Urlaubszeit der Mitarbeiter. Diese können dann selbst entscheiden, ob sie sich dann um die Daten kümmern oder lieber in Urlaub fahren. Auch die Beschäftigung damit nach Ende der Anstellung bei uns wird gern gesehen.

Sollte es trotz dieser wegweisenden Datenmanagementstrategie zu einer Datenveröffentlichung kommen, legen wir den Daten eine Lizenz bei, die Nachnutzer darauf festlegt, diese Forschungsdaten nur zu verwenden, um unsere bereits publizierten Ergebnisse zu reproduzieren.

Selbstverständlich ist das Satire. Aber dennoch, abgesehen von der Anrede und dem ersten Absatz gehören die angesprochenen Themen auch in einen realen Datenmanagementplan:

- Welche Daten werden erzeugt oder nachgenutzt ◇
- Datenformat
- Namenskonventionen
- Verwendete Standards ◇
- Wo und wie die Daten gespeichert werden ◇
- Qualitätssicherung ◇
- Versionskontrolle
- Zeitplan ◇
- Bereitstellung, Datenveröffentlichung und Lizenz ◇

Die roten Rauten kennzeichnen die Themen, die von der DFG empfohlen sind<sup>4</sup>. Für DFG-Anträge ist die Liste fast komplett, nur die Punkte *Rechte Dritter* und *in welchem Umfang die Daten relevant in anderen Forschungskontexten sein könnten* fehlen im Satirebeispiel. Natürlich kann ein Datenmanagementplan viel mehr enthalten, aber auch weniger. Wenn ein Thema Ihre

---

<sup>4</sup> [http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien\\_forschungsdaten.pdf](http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf),  
aufgerufen am 01.10.2018

Forschung nicht betrifft, wenn dieses Detail noch unklar ist oder wenn Sie einfach nicht ins Detail gehen wollen, können Sie das Thema auslassen. Ein unvollständiger Datenmanagementplan ist besser als gar keiner, und auch wenn eine Reihe von Details noch unklar sind, sollten die bereits geklärten Punkte zumindest innerhalb des Forschungsprojekts kommuniziert werden.

Ein Datenmanagementplan ist normalerweise kein statisches Dokument. Er kann erweitert und verändert werden. Beispielsweise kann eine erste Version dem Antrag auf Forschungsförderung beigelegt werden. Darauf aufbauend kann eine zweite Version mit weiteren Details während des Projekts und eine dritte Version mit der Planung für die Langzeitarchivierung nach dem Ende des Projekts entwickelt werden. Die beste Zeit, um mit der Arbeit am Datenmanagementplan zu beginnen, liegt vor dem Einreichen des Förderantrags, da zu dieser Zeit Datenmanagementplan, Meilensteinplan und Kostenplan zusammen ausgearbeitet und koordiniert werden können.

## Zeitplan

Die Forschung selbst verbraucht normalerweise mehr Zeit, aber der Aufwand für das Forschungsdatenmanagement sollte nicht unterschätzt werden. Nicht nur die Daten müssen verarbeitet werden, auch Metadaten müssen erzeugt und ein Datenarchiv ausgesucht werden. Daten, die mit anderen geteilt werden sollen, sollten möglichst durch zusätzliche Informationen ergänzt werden:

- Datendokumentation
- Sinnvolle Datensatz-, Datei- und Verzeichnisnamen
- Qualitätskontrollbericht
- Lizenz mit Regelungen für die Nachnutzung
- Versionsinformation

Im Falle von Fehlern in den Daten kann eine neue Datenversion erforderlich werden. Es sollten schon vor oder zu Beginn des Projekts Vorkehrungen getroffen werden, um mehrere verschiedene Versionen zu ermöglichen, z.B. eine zusätzliche Ebene im Verzeichnisbaum. Eine Versionsinformation sollte zusammenfassen, was geändert wurde. Fehler in den Daten sollten kurz in einem Erratum beschrieben werden.

Auch die Verwendung eines kontrollierten Vokabulars, d.h. eine Liste zulässiger Einträge für einen Teil der Metadaten, Datei-Header, Einheitennamen usw., sollte früh beschlossen werden. Dies betrifft auch Namenskonventionen für Dateien und Verzeichnisse. Alle diese Themen sollten im Datenmanagementplan näher ausgeführt werden.

Ein späteres Datenmanagement mit all den Komponenten ist prinzipiell möglich, aber viel schwieriger und zeitaufwendiger, da die Erinnerung an Details schnell verblasst und einige der Datenproduzenten womöglich das Projekt verlassen und ihr Wissen mitgenommen haben. Daher würde ein spätes Datenmanagement dazu führen, dass der Zeitplan ins Wanken gerät. Best Practice ist eine gemeinsame Entwicklung von Meilensteinplan und Datenmanagementplan, bevor das Projekt gestartet wurde. Zu dieser Zeit können Datenmanagementschritte leicht integriert werden. Ein weiterer positiver Effekt: Die Arbeit an nachfolgenden Datensätzen kann so organisiert werden, dass die Daten nicht alle gleichzeitig ankommen, sondern gestaffelt, siehe Abbildung 1. Auf diese Weise kann eine Überlastung von Mitarbeitern und Geräten vermieden werden.

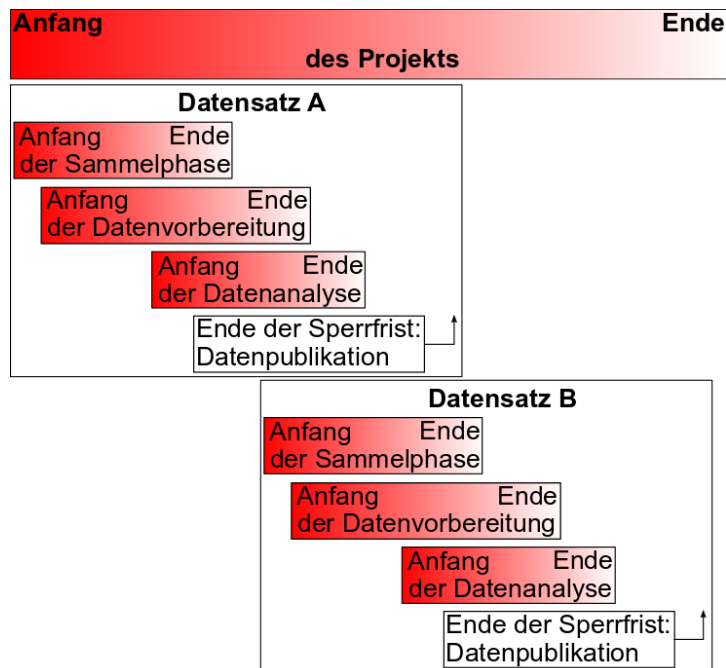


Abbildung 1: Zeitplan für das Forschungsdatenmanagement

## Kosten

Kosten sind selten in Datenmanagementplänen enthalten, da Einrichtungen, die Forschungsförderung gewähren, in der Regel eine Kostenaufstellung im Formularteil des Förderantrags erwarten und die Kosten für das Forschungsdatenmanagement dort enthalten sein sollten. Ausgenommen sind erziehungswissenschaftliche Projekte, die vom BMBF gefördert werden. In dem Fall müssen die Kosten für das Forschungsdatenmanagement in den Datenmanagementplan aufgenommen werden. Unabhängig davon, welches Dokument die Kostenaufstellung enthält, ist es sinnvoll, die Kosten des Forschungsdatenmanagements aufzulisten, da diese Kosten oft erstattet werden können.

Nur ein Absatz mit einigen Fakten über die Kosten des Forschungsdatenmanagements: Arbeitskosten sind viel höher als die Materialkosten. In Bezug auf die Schritte im Datenlebenszyklus ist die Speicherung in der Regel der kostengünstigste Schritt. Auswahl und Ingest kosten normalerweise viel mehr, siehe Abbildung 2. Das gleiche gilt für die Kosten für die Bereitstellung, wenn die Nachnutzung erlaubt ist, z.B. die Kosten für die Entwicklung und den Betrieb eines Webportals für den Download und — falls die serverbasiert angeboten wird — Prozessierung von Daten. Abbildung 2 zeigt die Kostenverteilung aus Archivsicht<sup>5</sup>. Wenn Sie die Daten selbst speichern, haben Sie wahrscheinlich eine ähnliche Kostenverteilung.

<sup>5</sup> Projekt Radieschen (Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur), Kostenverteilung und Risiken, [http://doi.org/10.2312/RADIESCHEN\\_008](http://doi.org/10.2312/RADIESCHEN_008), aufgerufen am 02.10.2018

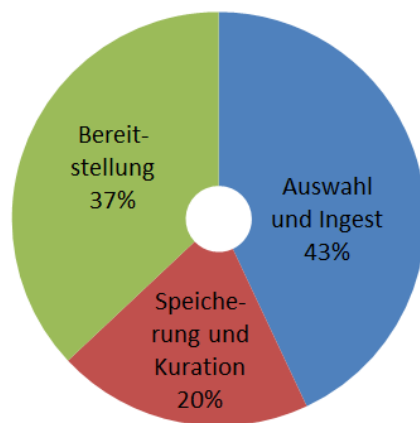


Abbildung 2: Kostenverteilung im Hinblick auf die Schritte im Datenlebenszyklus

## Software Tools

Natürlich kann ein Datenmanagementplan mit einem Textverarbeitungsprogramm geschrieben werden. Es gibt jedoch auch spezielle Software-Werkzeuge, die die Erstellung und Erweiterung eines Datenmanagementplans erleichtern können. Ein Vorteil all dieser Werkzeuge sind die Fragebögen, die einen guten Ausgangspunkt darstellen und verhindern können, dass etwas Wichtiges vergessen wird.

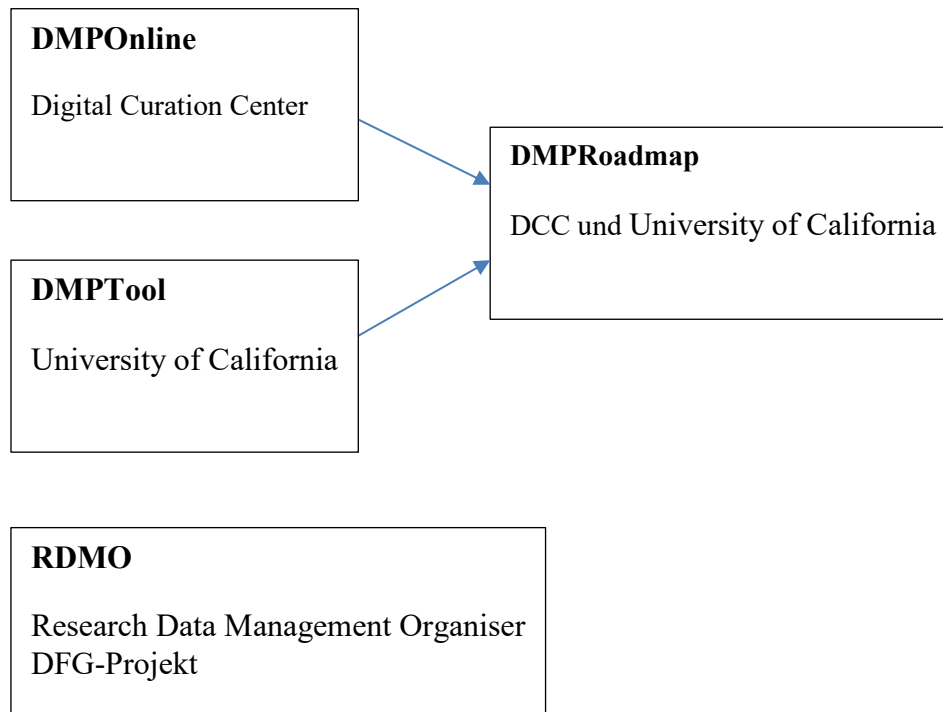


Abbildung 3: Generische Tools zur Erstellung von Datenmanagementplänen

Generische Open-Source-Tools sind DMPonline<sup>6</sup>, DMPTool<sup>7</sup> und RDMO (Research Data Management Organizer)<sup>8</sup>. Alle drei sind im Internet gehostet, d.h. nur ein Webbrowser ist notwendig, um sie zu verwenden. DMPonline wurde vom britischen Digital Curation Centre (DCC)<sup>9</sup>, DMPTool vom University of California Curation Centre (UC3) entwickelt. Beide Tools sind etabliert und haben viele Benutzer. Nach der Registrierung kann man Datenmanagementpläne erstellen, die auf einem britischen oder amerikanischen Server gespeichert sind. DCC und UC3 haben beschlossen, zusammenzuarbeiten und ein gemeinsames Tool, DMPRoadmap<sup>10</sup>, zu entwickeln, welches sich noch in einem frühen Entwicklungsstadium befindet.

RDMO ist ebenfalls noch in der Entwicklung, kann aber bereits verwendet werden. Die Software wird vom gleichnamigen DFG-Projekt RDMO entwickelt. Die Software wird vom AIP Potsdam gehostet, kann aber auch von lokalen Anbietern installiert werden. Die Entwickler von RDMO wollen ihre Software sogar zu einem echten Organisier erweitern, wie das auch Namensbestandteil von RDMO ist. Angekündigt und teilweise sogar schon entwickelt sind

- Metadaten-Extraktion, d.h. Metadaten können aus dem Datenmanagementplan extrahiert werden und in das Daten-Repository eingefüllt werden
- Definition von Aufgaben, wegen der — wenn sie anstehen — eine Erinnerung per E-Mail gesendet wird

Schon jetzt kann der ausführliche RDMO-Fragebogen eine Hilfe sein.

Darüber hinaus haben einige wissenschaftliche Gemeinschaften an fachspezifischen Werkzeugen gearbeitet. Beispiele sind der DMPTY Wizard<sup>11</sup> von CLARIN-D<sup>12</sup> für die Geisteswissenschaften und DataWiz<sup>13</sup> für die Psychologie. Auch für die Arbeit mit Textverarbeitungsprogrammen gibt es bereits eine Mustervorlage<sup>14</sup>.

## Weiterführende Links

**Weitere FoDaKo-Lesson:** Torsten Rathmann, Anforderungen der Forschungsförderer an das Forschungsdatenmanagement, <https://doi.org/10.5281/zenodo.1464993>

**Datenmanagementpläne allgemein:** [forschungsdaten.info](https://forschungsdaten.info)

**Ausführliche Checkliste für Datenmanagementpläne** in [Jens Ludwig, Harry Enke \(Hrsg.\), Leitfaden zum Forschungsdaten-Management, Verlag Werner Hülsbusch, Glückstadt 2013](#)

---

<sup>6</sup> <https://dmponline.dcc.ac.uk/>, aufgerufen am 02.10.2018

<sup>7</sup> <https://dmptool.org/>, aufgerufen am 02.10.2018

<sup>8</sup> <https://rdmo.aip.de/>, aufgerufen am 02.10.2018

<sup>9</sup> <http://www.dcc.ac.uk/>, aufgerufen am 02.10.2018

<sup>10</sup> <https://github.com/DMPRoadmap>, aufgerufen am 02.10.2018

<sup>11</sup> <https://www.clarin-d.net/de/aufbereiten/datenmanagementplan-entwickeln>, aufgerufen am 02.10.2018

<sup>12</sup> <https://www.clarin-d.net/de/>, aufgerufen am 02.10.2018

<sup>13</sup> <https://datawiz.leibniz-psychology.org/DataWiz/>, aufgerufen am 02.10.2018

<sup>14</sup> <https://www.esiencences.uni-trier.de/?p=989>, aufgerufen am 15.10.2018

## Beispiele für Datenmanagementpläne:

<https://esrc.ukri.org/files/research/international/esrc-dfid-example-data-management-plan/>

[https://www.gla.ac.uk/media/media\\_418168\\_en.pdf](https://www.gla.ac.uk/media/media_418168_en.pdf)

[https://www.dataone.org/sites/all/documents/DMP\\_MaunaLoa\\_Formatted.pdf](https://www.dataone.org/sites/all/documents/DMP_MaunaLoa_Formatted.pdf)

[https://www.fdm.hhu.de/fileadmin/redaktion/ZIM/FD\\_LZA/Musterplan-DFG.pdf](https://www.fdm.hhu.de/fileadmin/redaktion/ZIM/FD_LZA/Musterplan-DFG.pdf)

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung



Dieses Werk ist lizenziert unter  
einer [Creative Commons Namens-  
nennung 4.0 International Lizenz](https://creativecommons.org/licenses/by/4.0/).

Dieser Volltext ist im Wesentlichen eine deutsche Übersetzung der FoDaKo-Information *Data Management Plan*, <https://doi.org/10.5281/zenodo.1465001>