

Bausteine Forschungsdatenmanagement
Empfehlungen und Erfahrungsberichte für die Praxis von
Forschungsdatenmanagerinnen und -managern

Kosten und Aufwände von Forschungsdatenmanagement

Jan Leendertseⁱ Dirk von Suchodoletzⁱⁱ

2020

Zitiervorschlag

Leendertse, Jan und Dirk von Suchodoletz. 2020. Kosten und Aufwände von Forschungsdatenmanagement. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 1/2020: S. 1-7. DOI: [10.17192/bfdm.2020.1.8246](https://doi.org/10.17192/bfdm.2020.1.8246).

Dieser Beitrag steht unter einer
[Creative Commons Namensnennung 4.0 International Lizenz \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

ⁱAlbert-Ludwigs-Universität Freiburg. ORCID: [0000-0001-5676-493X](https://orcid.org/0000-0001-5676-493X)

ⁱⁱAlbert-Ludwigs-Universität Freiburg. ORCID: [0000-0002-4382-5104](https://orcid.org/0000-0002-4382-5104)

1 Abstract

Der Beitrag basiert auf dem Input der Plenumsdiskussion des Workshops zu Kosten- und Betriebsmodellen für Forschungsdaten-Services "Wie lassen sich Kosten und Aufwände ermitteln?". Er widmet sich der Frage, in welchem Umfang und an welcher Stelle Kosten durch Forschungsdatenmanagement (FDM) anfallen und wie diese zu verteilen sind. Benannt werden auch die Vorteile von FDM für Datenproduzierende selbst wie für die Wissenschaft. Dabei werden Handlungsfelder, Akteurinnen und Akteure identifiziert und mögliche Handlungsempfehlungen formuliert, die in der universitären Praxis an den verschiedenen Stellen des Forschungsprozesses Anwendung finden könnten. Kosten und Aufwände lassen sich zeitlich über den Lebenszyklus der Daten eines Projektes verorten und können zugleich Metrik für die Planung von Datenmanagement und die langfristige Speicherung von Daten sein. Die zunehmende Ausdifferenzierung der Fachgebiete und der nachhaltige Umgang mit Forschungsdaten erfordern eine neue Arbeitsteilung. Aufwände an Sach- und Personalmittel im Bereich FDM müssen bei der Beantragung von Fördergeldern berücksichtigt werden.

2 Ausgangspunkt der Überlegungen

Forderungen nach FDM ergeben sich aus den Ansprüchen an die gute wissenschaftliche Praxis¹ und werden vermehrt durch Drittmittelgeber erhoben. Sowohl Universitäten und Einrichtungen, die die Umgebung von Forschung gestalten, als auch Forschende, die innerhalb dieser Umgebung Wissenschaft betreiben, müssen auf diese Forderungen reagieren, wenn sie langfristig Drittmittel für ihre Vorhaben erfolgreich einwerben wollen. Mit der zunehmenden Wahrnehmung und Bedeutung von FDM kommen die damit verbundenen Kosten und Aufwände in den Blick. An dieser Stelle müssen die Produzenten von Daten und datenverarbeitende Infrastrukturen sich um eine Lösung bemühen. Für alle stellt sich die Frage, wie sich Aufwände und Nutzen von FDM ausbalancieren und wie anfallende Kosten verteilt werden sollte. Bislang gibt es kaum praktische Erfahrungen bei der Erfassung und Bewertung von Kosten, weshalb zunächst eher strukturelle Überlegungen anzustellen sind. Dieser Beitrag basiert auf einer Zusammenfassung der Plenumsdiskussion "Wie lassen sich Kosten und Aufwände ermitteln?", die im Rahmen des zweitägigen Workshops zu Kosten- und Betriebsmodellen für Forschungsdaten-Services in Trier² stattfand. Sie komplementierte die Break-Out-Sessions "Was heißt Nachhaltigkeit?" und "Was sind sinnvolle

¹Deutsche Forschungsgemeinschaft, Hrsg. "Leitlinien zur Sicherung guter wissenschaftlicher Praxis (Kodex)", August 2019. https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf.

²Zweitägiger Workshop "Wer soll das bezahlen? - Kosten- und Betriebsmodelle für nachhaltige Forschungsinfrastrukturen und FDM-Services" an der Universität Trier Mitte Juni 2019. Zu den Vorträgen des Workshops s. https://www.forschungsdaten.org/index.php/Wer_soll_das_bezahlen%3F_Kosten-_und_Betriebsmodelle_f%C3%BCr_nachhaltige_Forschungsinfrastrukturen_und_FDM-Services. Zugegriffen am 9.03.2020.

dauerhafte Kooperationsformen?”. Weiterhin spielt das Thema in Projekten wie dem Science Data Center BioDATEN³ oder der Nationalen Forschungsdaten-Infrastruktur (NFDI) eine Rolle, wo nach langfristig tragfähigen Kosten- und Betriebsmodellen gesucht wird. Bei der NFDI wird die Diskussion als “Querschnittsthemen” (“cross-cutting topic”) beispielsweise unter den Überschriften “collaborative governance and general NFDI framework” bzw. “technical infrastructure and concepts” geführt.⁴

3 Begriffsdefinition und Verortung

“Kosten” und “Aufwand” sollen hier als betriebswirtschaftliche Begriffe verstanden werden, die sich in einem gewissen Umfang abstrakt auf Forschung und Forschungsvorhaben übertragen lassen. Kosten sind eine kalkulatorische Größe, die zur quantitativen Erfassung dient. Unter Aufwände sei das Einbringen von Ressourcen zu verstehen. Im akademischen Bereich geschieht dies häufig ohne eine Quantifizierung. So können in diese Kategorie auch weitergehende Aktivitäten wie Aus- und Weiterbildung der Forschenden oder des Laborpersonals fallen oder der Nutzungsanteil von Infrastrukturen, die auch für andere Projekte bereitgestellt werden. FDM wird trotz der vorausgesetzten Anbindung an die allgemeinen Prinzipien “einer guten wissenschaftlichen Praxis” zu einer differenzierten Ausgestaltung der Forschungsdokumentation führen: Anlegen und Führen von Laborbüchern, umfassende technische Beschreibung von Grabungsstätten, Aufbewahrung von Fragebögen, erschließbare Bereitstellung von Sensordaten aller Art, um nur einige Beispiele anzudeuten. Während viele Praktiken der analogen Welt auf eine lange Entwicklung zurückschauen, muss sich die gute wissenschaftliche Praxis im Digitalen noch durchsetzen. Die Diskussion zeigte die Schwierigkeiten, eine einfache Trennung zwischen “Forschung” und “Forschungsdatenmanagement” aufzumachen: Zählen beispielsweise Kosten der Datenerhebung oder -beschaffung bereits zum Datenmanagement oder wird der Fokus primär auf Archivierung und Nachnutzung gelegt? Oder in den Geisteswissenschaften: Der Umgang mit Sekundärliteratur ist wesentlicher Teil der Primärforschung.⁵ Forschung ist in den meisten Disziplinen eine arbeitsteilige Tätigkeit, so dass die Aufgabenverteilung in einer Organisation oder in einem Projekt über die Kosten- und Aufwandsverteilung mitentscheidet. Vergleichbares gilt zwischen Projekten und den einzelnen Forschenden; hier unterscheiden sich die Interessenlagen ebenfalls. Mit der Annahme eines Projektes gehen je nach Fördergeber bestimmte Verpflichtungen bezüglich Publikation und Dokumentation von Ergebnissen einher, die weit über die Ver-

³Homepage <https://biodaten.info>; gefördert durch das Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg im Rahmen der eScience-Initiative.

⁴Glöckner, Frank Oliver, Michael Diepenbroek, Janine Felden, Jörg Overmann, Aletta Bonn, Birgit Gemeinholzer, Anton Güntsch, u. a. “Berlin Declaration on NFDI Cross-Cutting Topics”, 22. September 2019. <https://doi.org/10.5281/zenodo.3457213>.

⁵DARIAH-EU, und Digital Humanities@TCD, (Hrsg.) “Open Data for Humanists, A Pragmatic Guide”. Zugegriffen 5. November 2019. <https://dh.tcd.ie/dh/wp-content/uploads/2018/12/Open-Data-for-Humanists-A-Practical-Guide.pdf>.

verantwortlichkeit eines einzelnen Forschenden hinausgehen können. Eine Ausgangsbasis für die Ermittlung von Kosten und Aufwänden auf der Zeitachse bietet das Domänenmodell, welches die Phasen "private", "group", "persistent" und "public"⁶ unterscheidet. Die ersten beiden Phasen fallen in die aktive Laufzeit des Vorhabens, die weiteren Phasen beschreiben den Übergang und die Zeit nach Projektabschluss. Mit diesem Modell, das Kontext und Ambition der am Forschungsdatenmanagement Beteiligten berücksichtigt, lassen sich komplexe Beziehungen von Aufwänden und Ergebnissen analytisch fassen.

4 Planung und Anfall von Kosten

Aus Sicht vieler Projekte besteht das Hauptproblem in der Planung im "Ende" des Datenlebenszyklusses. Kosten, die beispielsweise für Speicher- und Verarbeitungssysteme während eines Vorhabens anfallen, sind den meisten Akteuren bewusst und aus ihrer Erfahrung her kalkulierbar. Hierbei liegt der Fokus stark auf beantragbaren Kosten. Datenmanagementpläne (DMP) werden zu genaueren Budgetierungen führen, weil sie dank ihrer Systematik die Berücksichtigung aller Aspekte im Lebenszyklus von Forschungsdaten sicherstellen. Schon bei der Erstellung von DMPs⁷ sollten umfassende Kostenüberlegungen für die Durchführung von Vorhaben erfolgen. Um den Anforderungen der Fördergeber zu genügen, müssen Projekte die Dauerhaftigkeit von Forschungsdaten via angemeldeter Kosten für die langfristige Perspektive planen. Teilweise existieren bereits Finanzierungsmodelle für Speicherplatz nach Projektende, die über die Budgets der Universitätsbibliotheken laufen. Kosten korrelieren mit Datenformaten und -typen; hochauflösende Bilder oder Messungen mit schneller Sample-Rate benötigen viel Speicher, komplexe Datenformate erfordern vielfach spezielle Software, die auch über lange Zeiträume mitgeführt werden muss. Deshalb können Kosten zwischen (Teil-)Disziplinen sehr weit auseinander liegen.

5 Spezialisierung bei Services und Ressourcen

Digitale oder digitalisierte Workflows erlauben ganz neue Erkenntnisgewinne, gehen jedoch mit gestiegenen Anforderungen an die dafür notwendige Qualifikation einher.

⁶Siehe Eifert, Thorsten, Stephan Muckel, und Dominik Schmitz, 2016. Introducing research data management as a service suite at RWTH Aachen university. In 9. DFN-Forum Kommunikationstechnologien. Gesellschaft für Informatik eV sowie Klar, Jochen, und Harry Enke. "Projekt RADIESCHEN. Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur - Report „Organisation und Struktur“". JL Archivbox 3. Zugriffen 24. Mai 2017. http://gfzpublic.gfz-potsdam.de/pubman/item/escidoc:117051:3/component/escidoc:124050/ProjektRadieschen_Organisation_und_Struktur.pdf.

⁷Siehe u.a. Leendertse, Jan, Susanne Mocken, und Dirk von Suchodoletz. 2019. "Datenmanagementpläne Zur Strukturierung Von Forschungsvorhaben". Bausteine Forschungsdatenmanagement, Nr. 2 (Mai), 4-9. <https://doi.org/10.17192/bfdm.2019.2.8003>.

Zunehmend wird spezialisiertes Personal zur Ausgestaltung von digitalen Workflows notwendig, so wie dieses für die Bedienung von Geräten und Messinstrumenten beispielsweise in den Naturwissenschaften unausweichlich wurde. Dieses Personal kann sowohl im Bereich der Administration der Hardware-Software-Stacks (HPC-Cluster) oder für die Interpretation und Analyse von Daten (Big-Data, Workflow-Design) notwendig werden. Vergleichbare Entwicklungen lassen sich zunehmend bei der Beantragung von INF-Projekten⁸ im Rahmen von Sonderforschungsbereichen (SFB) beobachten.⁹ Im Projektbereich könnten Datenspezialisten (data stewards, curators) zukünftig Bestandteil von Forschungsanträgen werden. Spezialisierte Arbeitsteilung wird zu der Situation führen, dass nicht jeder im FDM notwendige Schritt den Aufwand eines Vollzeitäquivalents oder eines antragsgerechten Teilzeitäquivalents erfordert, besonders wenn es auf die Ebene eines limitierten Projekts geht. In solchen Fällen ist eine sinnvolle Arbeitsteilung mit anderen Projekten oder Hochschulen zu finden. Das kann jedoch bedeuten, dass Einrichtungen für ihre Forschenden in Vorleistung gehen, da nicht alle Projekte gleichzeitig starten und die gleiche Laufzeit haben. Herausforderungen ergeben sich dann durch das Arbeitsrecht, das beispielsweise die Befristung begründungspflichtig macht. Auf der Ebene der Einrichtung spielen zunehmend Support-Services für Begleitung von Projekten, Weiterbildung und Schulung eine Rolle. Der zunehmende Beratungsaufwand im erweiterten Forschungsumfeld schlägt sich in Kosten nieder, die sich nicht direkt einzelnen Projekten oder Forschungsvorhaben zuordnen lassen.

6 Steuerung durch Kostenbetrachtung

Selbst wenn keine Refinanzierung bestimmter Aufwendungen geplant ist, hilft eine Kosten- und Aufwandsanalyse, um langfristig qualifizierte Entscheidungen für den Umgang mit Forschungsdaten treffen zu können. Die Angabe von Kosten liefert Anhaltspunkte für eine Ressourcenplanung. Weiterhin kann sie Nutzern die Motivation liefern, nicht auf den vielleicht aufwändigen Schritt einer Datenauswahl und -kuratierung zu verzichten. Unkuratierte und unstrukturierte Massendaten können selbst groß angelegte Speichersysteme weit über der erwarteten Datenrate befüllen. Ebenfalls regt es zur Abschätzung an, ob bestimmte Daten bei Bedarf nicht zu geringeren Kosten gleichwertig neu erzeugt (erneute Messung) oder wiederholt berechnet werden können. Weiterhin liefern Zugriffe auf Daten-Repositories Hinweise auf häufiger und weniger häufig nachgefragte Daten. Insbesondere im längerfristigen Verlauf

⁸Als Abkürzung für "Informationsmanagement und Informationsstrukturen in Sonderforschungsbereichen.

⁹Vgl. beispielsweise: Söring, Sibylle, Germaine Götzelmann, Philipp Hegel, Michael Krewet, und Danah Tonne. 2019. "An der Schnittstelle von Fach- und Informationswissenschaft:". Bausteine Forschungsdatenmanagement, Nr. 2 (Oktober), 89-95. <https://doi.org/10.17192/bfdm.2019.2.8083> bzw. Springer, Pina, und Hela Mehrrens. 2019. "Forschungsdatenmanagement im SFB 754: Klima - Biogeochemische Wechselwirkungen im Tropischen Ozean". Bausteine Forschungsdatenmanagement, Nr. 2 (Oktober), 84-88. <https://doi.org/10.17192/bfdm.2019.2.8073>.

werden Anhaltspunkte bzw. Metriken benötigt, damit Governance-Gremien entscheiden können, ob bestimmte Daten noch über die geforderten Basishaltezeiten (z.B. bei DFG-Projekten zehn Jahre) weiter aufgehoben und bereitgestellt werden. Kosten und Nutzen von Forschungsdatenmanagement sind nicht auf Personen, Projekte oder Einrichtungen eindeutig abgrenzbar, und dies sowohl in der räumlichen als auch zeitlichen Dimension. Das bereits benannte Domänenmodell gibt bereits Ansatzpunkte für den analytischen Zugriff. Forschungsdaten können Nutzen bei denjenigen stiften, die sie zu einem späteren Zeitpunkt oder in anderen organisatorischen Zusammenhängen replizieren oder nachnutzen. Für sie sind die Kosten bzw. Aufwände im Bereich der Datenproduzenten und der datenverarbeitenden Infrastrukturen bei der Planung irrelevant. Auch das Nutzungsverhalten hat Einfluss auf die Kostenstruktur. Die fallweise Replikation von archiviertem Material nach Anforderung aus einem Bandarchiv kann günstiger sein als eine permanente Online-Bereitstellung im Direktzugriff für ausgewählte Gruppen oder die Öffentlichkeit allgemein. Die Zuordnung von Kosten erfordert also in der sich neu sortierenden Arbeits- und Serviceteilung von den datenverarbeitenden Infrastrukturen, diese intern zuzuordnen oder als externe Anforderungen anzusehen.

7 Handlungsempfehlungen

Aus den angestellten Überlegungen werden für vier Gruppen Empfehlungen abgeleitet, wie Aufwände und die Bezifferung in Kosten für die Steuerung von FDM nutzbar gemacht werden können: Forschungseinrichtungen, Leitung von solchen, Drittmittelgeber und Forschende selbst. Forschungseinrichtungen sollten sich in der Pflicht sehen, ihre Mitglieder zur Umsetzung ihrer Forschungsdaten-Policy¹⁰ anzuhalten und hierfür geeignete Mittel und Infrastrukturen bereitzustellen. Sie sollten einen geeigneten organisatorischen Rahmen festlegen, in dem FDM in den zentralen Einrichtungen oder in den Fakultäten organisiert wird, was in den meisten Hochschulen über den etablierten Aufgabenkanon hinausgeht. Das tägliche Management von FDM ist also neu auszurichten. Eine mögliche Variante ist, INF-Projekte von SFBs mit beantragten Teilstellen oder Datenmanager in Exzellenz-Clustern zu einer eigenen FDM-Einheit zu bündeln und diese zumindest teilweise über direkt eingeworbene Mittel und nicht nur die Projektpauschalen zu bezahlen. In vielen Einrichtungen gibt es Sparpotenziale, da im Moment viele Projekte derzeit noch Aufwände duplizieren. Übergreifende Aktivitäten wie im Rahmen der Science Data Center können Erfahrungen in einer Community wie beispielsweise der Bioinformatik im SDC BioDATEN liefern, die sich auf andere wissenschaftliche Communities verallgemeinern lassen. Für Projektleitungen von Forschungsprojekten, Fakultätsleitungen oder Institutsvorständen besteht die Aufgabe darin, die eventuell zusätzlichen Aufwendungen für das Forschungsdatenmana-

¹⁰Vgl. hierzu beispielsweise: Kramer, Claudia, und Robert Ulrich. 2019. "Leitlinien zum Forschungsdatenmanagement am Karlsruher Institut für Technologie (KIT)". Bausteine Forschungsdatenmanagement, Nr. 2 (April), 1-3. <https://doi.org/10.17192/bfdm.2019.2.8016>.

gement zu quantifizieren und nach geeigneten Formen der Kompensation zu suchen. Hierbei müssen sie abschätzen, welche Finanzierungsquellen in Frage kommen. Ebenso sollten sie überlegen, wie die notwendige Expertise für die Umsetzung der guten wissenschaftlichen Praxis bereitgestellt wird. Diese Expertise kann durch Schulungen bei den Nachwuchsforschenden selbst aufgebaut oder durch qualifizierte Spezialisten als Consulting bereitgestellt werden. Selbst kleinere Projekte könnten Stellenanteile beantragen, wenn die Forschungseinrichtung einen geeigneten Rahmen vorgibt, in denen solche Mittel genutzt werden können. Die Drittmittelgeber haben vielfach die Entwicklung des Forschungsdatenmanagements durch ihre Forderungen an die gute wissenschaftliche Praxis vorangebracht. Sie sollten Forschende dadurch unterstützen, dass sie berechnete Mehraufwendungen für das FDM fördern oder auf geeignete Ressourcen verweisen. Die Nationale Forschungsdateninfrastruktur wird die Kosten- und Aufwandsdiskussion mit hoher Wahrscheinlichkeit voranbringen, neue Standards definieren und die Entwicklung fachspezifischer FDM-Ressourcen antreiben. Den einzelnen Forschenden obliegt es, den Regeln der guten wissenschaftlichen Praxis ihres Fachgebiets zu folgen und ihr Handeln entsprechend auszurichten. Das darf nicht bedeuten, dass sie mit dieser Aufgabe alleine gelassen werden. Sie sollten FDM als Instrument ihrer Profilierung begreifen und ihre Ansprüche, passende Angebote vorzufinden, offensiv formulieren. Denn die Sichtbarkeit der Nachfrage wird die Bewertung von Kosten und Aufwänden zu einem wirksamen Steuerungsinstrument machen.