

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

Etablierung von Forschungsdatenmanagement-Services in geowissenschaftlichen Sonderforschungsbereichen am Beispiel des SFB/Transregio 32, SFB 1211 und SFB/ Transregio 228

Constanze Curdtⁱ Dirk Hoffmeisterⁱⁱ Tanja Krammⁱⁱⁱ
Ulrich Lang^{iv} Georg Bareth^v

2019

Zitiervorschlag

Curdt, Constanze, Dirk Hoffmeister, Tanja Kramm, Ulrich Lang und Georg Bareth. 2019. Etablierung von Forschungsdatenmanagement-Services in geowissenschaftlichen Sonderforschungsbereichen am Beispiel des SFB/Transregio 32, SFB 1211 und SFB/Transregio 228. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 2/2019: S. 61-67. DOI: [10.17192/bfdm.2019.2.8103](https://doi.org/10.17192/bfdm.2019.2.8103).

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

ⁱUniversität zu Köln. ORCID: [0000-0002-9606-9883](https://orcid.org/0000-0002-9606-9883)

ⁱⁱUniversität zu Köln. ORCID: [0000-0002-8178-557X](https://orcid.org/0000-0002-8178-557X)

ⁱⁱⁱUniversität zu Köln. ORCID: [0000-0003-2684-7157](https://orcid.org/0000-0003-2684-7157)

^{iv}Universität zu Köln.

^vUniversität zu Köln.

1 Einleitung

Ein aktiver Austausch von Forschungsideen, -ergebnissen und -daten ist die Grundlage für ein erfolgreiches Verbundforschungsprojekt wie einem Sonderforschungsbereich (SFB), der institutionsübergreifend, interdisziplinär und langfristig angelegt ist. Dieser Austausch unter den Forschenden kann durch regelmäßige Projekttreffen, sowie durch gemeinsame Geländekampagnen oder Feldaufenthalte zur Datenerhebung gefördert werden. Gerade bei standort- oder institutionsübergreifenden Projekten ist es essentiell, den Forschenden zum Austausch und zur nachhaltigen Sicherung von Informationen oder Daten (u.a. projektinterne Berichte, Konferenzbeiträge, Publikationen, oder Forschungsdaten) verschiedene Services und technische Infrastrukturen zur Verfügung zu stellen.

In diesem Beitrag werden am Beispiel von drei SFBs bzw. SFB/Transregio (TRR) bereits etablierte und derzeit im Aufbau befindliche, unterstützende Services mit Fokus auf das Forschungsdatenmanagement (FDM) präsentiert

2 Projektbeschreibung der SFBs und ihrer Daten

Der SFB/TRR 32 (<https://www.tr32.de/>) „Muster und Strukturen in Boden-Pflanzen-Atmosphären-Systemen: Erfassung, Modellierung und Datenassimilation“ (Förderung 1/2007 – 12/2018) hat zum Ziel, mit Hilfe innovativer Monitoring- und Modellierungsansätze die Herkunft von und die Wechselbeziehungen zwischen den zeitlichen und räumlichen Mustern innerhalb des Boden-Vegetation-Atmosphäre-Systems besser zu verstehen. Innerhalb der letzten 12 Jahre wurde eine Vielzahl von heterogenen Daten in verschiedenen räumlichen Skalen und mit unterschiedlicher zeitlicher Auflösung erstellt. U.a. wurden Daten in Feldkampagnen (z.B. durch meteorologische oder hydrologisches Monitoring, Flugzeugkampagnen), Laboruntersuchungen (z.B. Messung der Pflanzenbiomasse) oder durch Datenmodellierungsansätze erzeugt. Bereits in der Planung des SFB/TRR 32 wurde in der 1. Phase (1/2007 – 12/2010) ein zentrales Teilprojekt zum Datenmanagement initiiert, das für den Aufbau einer nachhaltigen FDM-Infrastruktur verantwortlich war. In den Förderphasen zwei und drei wurde dieses als Teilprojekt Informationsinfrastruktur (INF-Projekt) weitergeführt. Neben dem Aufbau von FDM-Services war dieses Teilprojekt ebenfalls für sogenannte Geodaten-Services zuständig, wie die Beschaffung und Bereitstellung von Geo- und Wetterdaten, sowie für die Erstellung von Landnutzungsanalysen des Projektgebietes.

Der SFB 1211 (<https://sfb1211.uni-koeln.de/>) „Evolution der Erde und des Lebens unter extremer Trockenheit“ (Förderung seit 07/2016) hat zum Ziel, die Forschung im Bereich der wechselseitigen Beziehung zwischen biologischer Evolution und Landschaftsevolution maßgeblich voranzutreiben. Es sollen die Schlüsselmerkmale biologischer Aktivität in extrem wasserlimitierten Habitaten der Erde identifiziert und Erdoberflächenprozesse, die unter nahezu wasserfreien Bedingungen ablaufen, charak-

terisiert werden. Dazu wird gleichfalls eine Vielzahl unterschiedlicher Daten erhoben (z.B. von Bohrungen bis zu Daten von Wetterstationen). Bereits ab der ersten Förderphase wurde das Teilprojekt „Datenmanagement und raumbezogene Analyse“ als zentrales INF-Projekt im SFB verankert. Neben dem Aufbau von zentralen FDM-Services und Infrastruktur, besteht eine weitere Aufgabe u.a. in der Analyse raumbezogener Daten für das Forschungsgebiet des SFBs.

Der Zusammenhang von sozial-ökologischen Landnutzungswandel und Zukunftsgestaltung im ländlichen Afrika wird im SFB/Transregio 228 (<https://www.crc228.de>) „Zukunft im ländlichen Afrika: Zukunft-Machen und sozial-ökologische Transformation“ (Förderung seit 01/2018) untersucht. Entsprechend sind die im Projekt erstellten Daten sehr vielfältig und werden in zahlreichen Feld- und Geländekampagnen erhoben. Beispielsweise werden lokale und überregionale qualitative und quantitative Interviews durchgeführt, sowie Bodenuntersuchungen und Biodiversitäts-Monitoring. Bereits ab der ersten Förderphase wurde das Teilprojekt „Data Management“ als zentrales INF-Projekt im SFB verankert. Im Gegensatz zu den vorher beschriebenen Projekten liegt der Fokus im INF-Projekt nur auf dem Aufbau von zentralen FDM-Services und Infrastruktur.

3 Ermittlung des Bedarfes in den SFBs

In allen drei SFB bzw. SFB/TRR wurden jeweils zu Beginn der Projektförderung sowie einer neuen Förderphase Bedarfsermittlungen im Hinblick auf FDM-Unterstützung und aufzubauende Services durchgeführt. Es wurden jeweils Umfragen erstellt, die mindestens auf Teilprojektebene und ergänzend sogar auf Ebene von Einzelforschenden (u.a. PhD, PostDoc, PI) beantwortet wurden. Auf Grundlage der ersten Umfrage im SFB/TRR 32 im Jahr 2007, wurde der Fragenkatalog sukzessiv mit neuen Fragen und in Richtung eines Daten Management Plans (DMP) erweitert. Die letzte Bedarfserhebung wurde in 2018 im SFB/TRR 228 durchgeführt und beinhaltete 47 Fragen aus sieben Fragenkategorien, wie beispielsweise Fragen zu allgemeinen Informationen zum Teilprojekt, Erhebung von Projektdaten, Benutzung der FDM-Infrastruktur (die sogenannte „Projektdatenbank“), Datenaustausch und -abgabe, Verwendung von Software, sowie zum gewünschten Unterstützungsangebot im SFB. Die Standardfragen der Umfrage wurden jeweils mit Fokus auf die Besonderheiten des SFBs angepasst, sowie im Hinblick auf die spezifischen Unterstützungsangebote und ergänzende Aufgaben des INF-Projektes. Des Weiteren wurden Bedarfe für FDM-Unterstützung in Einzelgesprächen während der diversen Projekttreffen ermittelt.

Zusammenfassend war der Bedarf nach unterstützenden FDM-Services in allen drei SFBs ähnlich. Der Aufbau einer zentralen FDM-Infrastruktur wurde gewünscht, die in der Lage ist die Vielzahl an heterogenen Projektdaten, die durch die Forschenden erzeugt wird nachhaltig zu speichern und zur Verfügung zu stellen. Neben den Forschungsdaten selber, sollten auch immer weitere Daten (wie Publikationen, Bilder, Be-

richte) mit integriert werden. Ebenfalls wurde der Bedarf nach unterstützenden FDM-Seminaren und praktischen Trainings deutlich.

4 Unterstützung der SFBs mit FDM-Services

Im Rahmen des SFB/TRR 32 wurde 2007 begonnen, unterstützende FDM-Services aufzubauen und diese innerhalb einer FDM-Infrastruktur zu vereinen ¹. Diese Erfahrung wurde genutzt um nach dem Vorbild des SFB/TRR 32 für die zeitlich nachfolgenden SFB 1211 und TRR 228 FDM-Services aufzubauen. In allen drei SFBs sind die folgenden generellen FDM-Services etabliert und über die aufgebaute FDM-Infrastruktur (in den Projekten auch als „Projektdatenbank“ bezeichnet) verfügbar:

- **Beratung und Training:** Unterstützung der Forschenden u.a. durch Online-Tutorials, FAQs und ergänzende praktische und theoretische Trainingseinheiten
- **Interne Datenaustauschplattform:** Möglichkeit zum internen Datenaustausch, u.a. für temporäre oder vorläufige Daten ohne Verpflichtung zur Beschreibung mit Metadaten
- **Speicherung und Backup:** Möglichkeit zur zentralen Speicherung inkl. automatisches Backup von Daten wie Forschungsdaten, Publikationen, Bilder, Berichte, etc.
- **Dokumentation der Daten:** Möglichkeit zur Beschreibung von Daten entsprechend gängiger Metadatenstandards (z.B. DataCite, Dublin Core, INSPIRE)
- **Veröffentlichung der Daten:** Möglichkeit auf Anfrage zur Veröffentlichung der Daten mit DOI
- **Datensuche:** Suche nach Daten über verschiedene Suchmasken und vordefinierte Listen
- **Daten Download und Zugriffsrechte:** Möglichkeit zum Herunterladen von Daten je nach Status des Interessierten (z.B. Projektmitglied, Gast) und gewähltem Datenzugriffsrecht, festgelegt durch den Datenbereitstellenden
- **Statistiken:** Verfügbarkeit von automatisch generierten Statistiken für die gesamte „Projektdatenbank“, sowie auf Datensatzebene
- **Kartenfunktionen:** Möglichkeit der räumlichen Suche nach Daten

Ergänzend zu den generellen FDM-Services wird jede SFB „Projektdatenbank“ nach den individuellen Bedürfnissen der Projekte sowie Forschenden angepasst. Beispielsweise enthält die Struktur des SFB/TRR 32 eine spezifische, interne Sektion, in der die beschafften Wetterdaten (u.a. des DWD) organisiert und dokumentiert bereitgestellt werden, sowie ein Tool zur Visualisierung der Daten. Im SFB 1211 gibt es ebenfalls einen eigenen Bereich, in dem die Wetterdaten zur Verfügung gestellt werden. Zusätzlich besteht für beide Projekte ein internes Web-GIS zur Visualisierung der Geodaten.

¹Siehe hierzu auch Curdt, Constanze. 2019. "Supporting the Interdisciplinary, Long-Term Research Project 'Patterns in Soil-Vegetation-Atmosphere-Systems' by Data Management Services." Data Science Journal no. 18 (1):5. <https://doi.org/10.5334/dsj-2019-005>.

Innerhalb der SFB/TRR 228 wurde auf Wunsch das „Transdisciplinary Diary“, als interne Plattform zum Sammeln und Austauschen von Erlebnissen, Erkenntnissen und Geschichten der Forschenden aus den Geländeaufenthalten aufgesetzt.

5 Aufbau und Implementierung der technischen FDM-Infrastruktur

Eine erste Version der SFB/TRR 32 „Projektdatenbank“ (TR32DB, <https://www.tr32db.de>) wurde seit 2007 aufgebaut und ist bereits im Februar 2008 online gegangen. Die vorhandene FDM-Infrastruktur wurde im SFB 1211 und SFB/TRR 228 nachgenutzt und jeweils an die Bedürfnisse der Projekte angepasst. So war es zeitnah nach Projektbeginn möglich, die Forschenden zu unterstützen. Entsprechend ist die „Projektdatenbank“ des SFB 1211 (CRC1211DB, <https://www.crc1211db.uni-koeln.de>) seit Ende 2016 und die des SFB/TRR 228 (TRR228DB, <https://www.trr228db.uni-koeln.de>) seit Ende 2018 online verfügbar. Alle drei Infrastrukturen bzw. Systeme werden seitdem gemeinschaftlich und kontinuierlich weiterentwickelt.

Hauptziel bei der Entwicklung der technischen FDM-Infrastruktur war die Erfassung, Beschreibung, Speicherung, Sicherung und der Austausch aller im erstellten Projekt erzeugten Daten. Neben den Forschungsdaten lag der Fokus hierbei ebenfalls auch bei der Aufbewahrung von Publikationen, Berichten, Vorträge und Postern, die als Ergebnis der Forschenden erstellt werden. Diese Infrastruktur sollte den Datenaustausch zwischen allen beteiligten Projektmitgliedern während der Projektlaufzeit und die Wiederverwendung aller Daten über die Projektfinanzierung hinaus unterstützen. Bei der Erstellung wurde neben den Bedürfnissen der Forschenden ebenfalls die Anforderungen der DFG berücksichtigt (z.B. die Kooperation mit einem lokalen Rechenzentrum, Gute Wissenschaftliche Praxis). Entsprechend wurde das FDM-System in Zusammenarbeit mit dem Regionalen Rechenzentrum (RRZK) der Universität zu Köln aufgebaut, wo es auch physisch gehostet wird. Es wurde eine dreistufige Architektur² entworfen, die auf den verfügbaren Hard- und Softwarekomponenten des RRZK basiert und damit auch die langfristige nachhaltige Verfügbarkeit und damit die Wiederverwendbarkeit aller Projektdaten über die Projektförderung hinaus sicherstellt. Alle Projektdaten werden in einem dateibasierten Datenspeicher (aktuell: Andrew File System) abgelegt. Entsprechende Metadaten und Verwaltungsdaten werden in einer Datenbank (aktuell: MariaDB) verwaltet. Eine Weboberfläche mit Web-Mapping-Komponenten (z.B. Web-GIS) ermöglicht den Benutzerzugriff. Die Weboberfläche wurde mit gängigen Webtechnologien und –sprachen selbst entwickelt (z.B. PHP, HTML,

²Siehe hierzu auch Curdt, Constanze, und Dirk Hoffmeister. 2015. "Research data management services for a multidisciplinary, collaborative research project: Design and implementation of the TR32DB project database." Program no. 49 (4):494-512. <https://doi.org/10.1108/PROG-02-2015-0016>.

XML, JavaScript, CSS) und mit Kartenfunktionen ergänzt (z.B. ArcGIS Enterprise, Google Maps API). Das Herzstück des Systems ist das selbst entwickelte mehrstufige Metadaten-System, dem ein eigenes Metadaten-Schema³ zugrunde liegt. Dieses Schema unterstützt die genaue, interoperable Beschreibung aller betrachteten Datentypen (z.B. Geodaten, Daten, Publikationen, Berichte, Bilder). Dublin Core und später das DataCite Metadaten-Schema dienen als Basisschema. Diese wurden mit fachspezifischen Metadaten-Schemata wie z.B. ISO 19115 und INSPIRE gemappt, sowie mit projektspezifischen (z.B. Keywords) und datentypspezifischen Eigenschaften (z.B. Instrumente, Parameter) erweitert. Ein benutzerfreundlicher, selbst entwickelter Metadaten-Wizard ermöglicht den Forschenden die Eingabe und Änderung aller Metadaten über die Web-Oberfläche.

6 Gewonnene Erfahrungen

Um das FDM in einem Forschungsverbund erfolgreich zu etablieren, hat sich in den drei Projekten SFB/TRR 32, SFB 1211 und SFB/TRR 228 gezeigt, dass ein regelmäßiger Austausch mit den Forschenden essentiell ist. Dieses wird durch die regelmäßige Teilnahme an Projekttreffen, gemeinsame Geländekampagnen, Projektklausurtagungen oder Workshops gefördert. Hierbei kann der Bedarf für neue FDM-Services relativ einfach bei den Forschenden ermittelt werden, es können gemeinsam Ideen für Services erarbeitet werden, sowie vorhandene Angebote an den Forschenden vermittelt werden. Ebenfalls trägt der Austausch dazu bei, die Workflows der Forschenden und ihrer Daten besser zu verstehen und diese Erkenntnis in die FDM-Services zu integrieren.

Im Hinblick auf die technische Umsetzung einer FDM-Infrastruktur hat sich gezeigt, dass die Nutzung bestehender, nachhaltiger Services und die Zusammenarbeit mit einem Rechenzentrum förderlich sind. Zum einen war es so möglich, die FDM-Systeme der beiden jüngeren SFBs zeitnah zur Verfügung zu stellen. Zum anderen ermöglicht die Nutzung der RRZK-Infrastruktur und der vorhandenen „IT-Basisdienste“ (wie Speicher, Backup, Webprojekte, Identity Management) eine langfristige Verfügbarkeit der FDM-Systeme und Daten. Bei der Umsetzung der FDM-Services ist darauf zu achten, dass diese entsprechend des Bedarfs der Forschenden aufgebaut und regelmäßig evaluiert und angepasst werden. Ebenfalls hat sich die Bildung von Schnittstellen zwischen verschiedenen Infrastrukturen als positiv herausgestellt. Beispielsweise wurde jeweils eine Schnittstelle zwischen den etablierten „Projektbanken“ und den offiziellen Webauftritten des SFB/TRR 32 und SFB 211 erstellt, um direkt Angaben zu Teilprojekt-Publikationen oder Forschungsdaten zu integrieren.

Gleichfalls hat sich die Etablierung eines breiten Unterstützungsangebotes in allen drei SFBs als erfolgreich herausgestellt. Neben den auf jeder Projekt-Webseite ver-

³Siehe hierzu auch Curdt, Constanze. 2014. TR32DB Metadata Schema for the Description of Research Data in the TR32DB. <https://doi.org/10.5880/TR32DB.10>.

fürbaren FAQs, auf denen allgemeine und spezifische Fragen zum FDM und zur FDM-Infrastruktur beantwortet werden, wurden unterschiedliche PDF-Tutorials erstellt. In allen Projekten werden zusätzlich und regelmäßig verschiedene Schulungs- und Trainingsformate angeboten. Falls möglich, sind diese Trainingseinheiten innerhalb der Graduiertenschulen der SFBs verankert.

Um das FDM innerhalb der SFBs noch etwas sichtbarer und verbindlicher zu gestalten, ist die Erstellung und Etablierung einer FDM-Richtlinie essentiell. Entsprechend wurde in allen drei SFBs eine Daten-Policy verabschiedet, in der der Umgang mit den Daten im Projekt sowie deren Nachnutzung geregelt wird (u.a. Datenbereitstellung, Datenbeschreibung, Zitierung). Diese Policy wurde auf Wunsch der Forschenden bewusst schlank gehalten. Falls möglich, wurde diese Policy ebenfalls innerhalb der Satzung des SFBs verankert, sowie in der Satzung der Graduiertenschule des SFBs. In allen drei SFBs ist die Verpflichtungen zur Datenabgabe und zum Austausch von Forschungsdaten in die jeweilige „Projektdatenbank“ relativ „weich“ geregelt. Im Fall des SFB/TRR 32 wurde daher noch ergänzend die Vereinbarungen in der Graduiertenschule getroffen, dass Promovierende verpflichtet werden, ihre Forschungsdaten in der TR32DB abzulegen, um das Zertifikat der Graduiertenschule zu erhalten.

Danksagung

Die Autorinnen und Autoren danken allen Mitarbeitenden, die an der Konzeption und Umsetzung der Projekte beteiligt sind und waren. Darüber hinaus sind die Autorinnen und Autoren dankbar für die finanzielle Unterstützung durch den SFB/TRR 32, SFB 1211 und SFB/TRR 228, gefördert durch die Deutsche Forschungsgemeinschaft (DFG).