

Research Data Management Organiser

Ulrike Wuttke¹, Heike Neuroth¹, Jochen Klar², Harry Enke³, Janine Straka¹, Kerstin Wedlich-Zachodin⁵, Claudia Kramer⁵, Olaf Michaelis³, Jens Ludwig⁴

¹ Fachhochschule Potsdam (FHP)

² Berater und Softwareentwickler

³ Leibniz-Institut für Astrophysik Potsdam (AIP)

⁴ Stiftung Preußischer Kulturbesitz

⁵ Karlsruher Institut für Technologie (KIT)

Zusammenfassung

Das Erstellen eines Datenmanagementplans (DMP) und die Planung des Datenmanagements erfordern ein frühzeitiges Nachdenken darüber, wie mit Daten während der Forschungsaktivitäten und nach Projektende umgegangen werden soll. Es stellen sich Fragen zu Nutzungsrechten, zur Datenpublikation und zu Archivierungsmöglichkeiten. Der Research Data Management Organiser (RDMO) ist eine webbasierte Open Source-Software, die Forschungsprojekte bei der Planung, Umsetzung und Verwaltung aller Aufgaben des Forschungsdatenmanagements über den gesamten Datenlebenszyklus unterstützt. Zusätzlich ermöglicht RDMO die textuelle Ausgabe eines DMP nach den Vorgaben unterschiedlicher Förderer und kann als Teil des Infrastrukturangebots einer Einrichtung an deren Bedarfe (z. B. eigene Frage- und Antwortvorlagen) und deren eigene Corporate Identity angepasst werden.

In diesem Beitrag wird das DFG-Projekt Research Data Management Organiser (RDMO) vorgestellt, das sich momentan in der zweiten Projektphase befindet. In dieser wird durch die Projektpartner des Leibniz-Instituts für Astrophysik Potsdam, der Fachhochschule Potsdam und der Bibliothek des Karlsruher Instituts für Technologie die Software weiterentwickelt und verbessert. Außerdem werden in enger Zusammenarbeit mit den Nutzenden neue Funktionalitäten hinzugefügt, die Community bei der Integration von RDMO in lokale Infrastrukturen unterstützt, Einführungs- und Schulungsmaterialien erstellt und ein Nachhaltigkeitskonzept für RDMO erarbeitet.

Abstract

Data management planning and creating a Data Management Plan (DMP) requires early consideration of how to handle the data during the project and after the end of the project. Questions arise concerning usage rights, data publication, and long-term

storage. The Research Data Management Organiser (RDMO) is a web-based open source software that supports research projects in planning, implementing, and managing all research data management tasks over the entire data lifecycle. In addition, RDMO enables the textual output of a DMP according to the specifications of different funders. It can be adapted as an infrastructural service of an institution according to its requirements (e.g. individual question and answer templates) and its own corporate identity.

The article presents the DFG funded project Research Data Management Organiser (RDMO), which is in its second project phase. Currently, the software is being further developed and improved by the project partners Leibniz Institute for Astrophysics Potsdam, University of Applied Sciences Potsdam and the library of the Karlsruhe Institute of Technology. Furthermore, new functionalities will be added in close cooperation with the users, the community will be supported in integrating RDMO into local infrastructures, introductory and training materials will be created and a sustainability concept for RDMO is being developed.

1. Einleitung

1.1 Aktuelle Entwicklungen

Im Jahr 2018 hat das Bundesministerium für Forschung und Entwicklung (BMBF) nach einer Re-Strukturierung des Ministeriums ein eigenes Referat 421 für Forschungsdaten ins Leben gerufen¹. Parallel dazu wurde in Deutschland das Förderprogramm Nationale Forschungsdateninfrastruktur (NFDI) gestartet. Bereits 2016 legte dafür der Rat für Informationsinfrastrukturen (RfII) mit der Veröffentlichung des Berichts „Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland“ (RfII 2016) den Grundstein für die Strategie eines nationalen Ansatzes des Forschungsdatenmanagements. Darauf folgende Veröffentlichungen des RfII gaben weitere Impulse für die Ausgestaltung der NFDI.² Neben verschiedenen Fachverbänden nahmen aus Bibliotheks- und Infrastrukturperspektive verschiedene Verbände und Institutionen das Thema mit Nachdruck auf und legten weitere, die

¹ <https://www.bmbf.de/pub/orgplan.pdf>. Alle Links wurden am 15.04.2019 geprüft.

² Vgl. die weiteren Diskussionsimpulse des RfII, u. a. RfII 2017a, RfII 2017b, RfII 2018a und RfII 2018b.

Wichtigkeit des Managements von Forschungsdaten im Rahmen der nationalen Strategieentwicklung unterstreichende, Impuls- und Strategiepapiere vor.³

Im internationalen Kontext sei in diesem Zusammenhang auf die Tätigkeiten der Initiative „Research Data Alliance - Research Data Sharing without barriers“⁴ verwiesen, die ein breites Spektrum im Rahmen von Arbeitsgruppen abdeckt. Auf europäischer Ebene wurden im Rahmen der „European Open Science Cloud“ (EOSC)⁵ im November 2018 zwei wichtige Berichte veröffentlicht: 1.) „Prompting an EOSC in practice“ (Muscella et al. 2018), der sich hauptsächlich mit Regularien für verschiedene Beteiligungsszenarien an der EOSC beschäftigt, und 2.) „Turning FAIR into reality“ (European Commission Expert Group on FAIR Data 2018). Der zuletzt genannte Bericht beschreibt „the broad range of changes required to turn „FAIR data into reality““ (ebd., S. 13). Alle Initiativen, Entwicklungen und Diskussionen sowie konkreten Förderprogramme basieren auf den FAIR-Prinzipien („Findable – Accessible – Interoperable – Reusable“)⁶, die in Zukunft einen menschen- und vor allem maschinenlesbaren Zugang zu Forschungsdaten und deren Nachnutzung realisieren sollen.

Die im November 2018 veröffentlichte „Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI)“ (GWK 2018) weist in ihrer Präambel auf die Notwendigkeit von potentiellen NFDI-Konsortien als „wissenschaftlich breit nutzbare Datenschätze mit gesellschaftlichem Mehrwert“ hin und macht in ihren Förderkriterien darauf aufmerksam, dass „ein stimmiges Konzept zu Datennutzung und -zugang sowie Auffindbarkeit und Nachnutzbarkeit der Daten, welches entlang der FAIR-Prinzipien ausgerichtet ist“ (ebd., S. 4), vorliegen muss. Dafür stehen im Zeitraum von 2019 bis 2028 bis zu 90 Mio. Euro pro Jahr im Endausbau für die Projektförderung der NFDI zur Verfügung. Flankierend eröffnete der Stifterverband ein eigenes Förderprogramm für neue Lehr- und Lernkonzepte im Bereich „Data Literacy“.⁷

Zusammenfassend kann festgestellt werden, dass Forschungsdaten eines der wichtigen Themen der Zukunft darstellen und der Umgang mit und das Management von Forschungsdaten weiter in den Mittelpunkt der Aufmerksamkeit rücken. Dazu

³ Vgl. u. a. DFG 2018a, Kapitel B-2d, DBV 2018, DFG 2018b.

⁴ <https://www.rd-alliance.org/>.

⁵ <https://eosc-pilot.eu>.

⁶ Vgl. z. B. <https://www.go-fair.org/fair-principles/>.

⁷ <https://www.stifterverband.org/data-literacy-education>.

gehört auch die Entwicklung und Bereitstellung geeigneter Werkzeuge und Dienste, die alle beteiligten Akteure aus der Wissenschaft, Infrastruktur-Dienstleister etc. darin unterstützen, den Umgang mit Forschungsdaten gemäß den FAIR-Prinzipien zu ermöglichen.

1.2 Hintergrund des DFG-Projekts

Seit 2015 fördert die DFG das Projekt Research Data Management Organiser: Zunächst in einer ersten Förderphase für 18 Monate unter dem Namen „Entwicklung und Implementierung eines Werkzeugs für die Planung, Umsetzung und Kontrolle des Forschungsdatenmanagements (FDMP-Werkzeug)“ und seit November 2018 in einer zweiten Förderphase für 30 Monate unter dem jetzigen Titel und Akronym „Research Data Management Organiser (RDMO)“. Seit der ersten Phase waren die beiden Einrichtungen Leibniz-Institut für Astrophysik Potsdam (AIP) und der Fachbereich Informationswissenschaften an der Fachhochschule Potsdam (FHP) für die Durchführung verantwortlich, ab der zweiten Phase hat sich die Bibliothek des Karlsruher Instituts für Technologie (KIT) angeschlossen. In beratender Funktion war stets Jens Ludwig, (ehemals Staatsbibliothek zu Berlin, jetzt: Strategien & Technologien Koordination Digitale Transformation, Stiftung Preußischer Kulturbesitz) als Mitinitiator des Projekts beteiligt.

Nach dem Ende der ersten Förderphase stand ein Werkzeug für die strukturierte Planung, Umsetzung und Verwaltung des Forschungsdatenmanagements zur Verfügung. Da sich schon nach relativ kurzer Zeit zeigte, dass Forschungsdatenmanagement ein aktiver, dynamischer Prozess ist, der mit der Planung von Forschungsvorhaben beginnen sollte und kontinuierlich angepasst und dokumentiert werden muss, wurde das Werkzeug von FDMP zu RDMO umbenannt, was die Schwerpunktverschiebung von der Planung zur Organisation des Forschungsdatenmanagements verdeutlichte. Die Umbenennung sollte auch unterstreichen, dass viele Akteure am Forschungsdatenmanagement beteiligt sind und das Werkzeug auch für die mittel- bis langfristige strategische Planung (z. B. im Technologiebereich, für die Finanzkalkulation einer gesamten Einrichtung etc.) verwendet werden kann.

Ein weiteres wesentliches Merkmal von RDMO ist, dass die Software von jeder Einrichtung eigenständig implementiert und so dem eigenen „Look & Feel“ angepasst

werden kann. RDMO versucht so weit wie möglich, gängige Schnittstellen zur Verfügung zu stellen, so dass das Werkzeug mit bereits vorhandenen Diensten interoperabel verknüpft werden kann. Weiterhin verbleiben alle Daten in der Hoheit der Einrichtung. Die RDMO-Community kann sich in kollaborativ orchestrierten Prozessen an der Weiterentwicklung und Pflege der Software und den Fragenkatalogen etc. beteiligen.

Schwerpunkte der zweiten Förderphase sind einerseits die technologische Stabilisierung des Werkzeugs inklusive Bereitstellung weiterer Schnittstellen für die einfache Integration in lokale Technologieumgebungen, andererseits die stabile Verankerung in der Community. Je mehr Einrichtungen und Forschergruppen das Werkzeug einsetzen und sich aktiv an der Weiterentwicklung beteiligen, desto nachhaltiger kann der Betrieb gesichert werden. Im Folgenden werden die Aktivitäten rund um RDMO näher beschrieben.

2. Research Data Management Organiser (RDMO)

2.1 Ziele

In Deutschland werden Datenmanagementpläne (DMP) in der Regel momentan nur einmal, als Teil des Antragstexts, erstellt und im Verlauf des Forschungsvorhabens nicht weiter aktualisiert. Im Mittelpunkt des RDMO-Projekts steht die Weiterentwicklung des Organisers zu einem multilingualen Research Data Management Organiser (RDMO), der es ermöglicht, das Forschungsdatenmanagement während des gesamten Projekts zu organisieren und entsprechend den jeweiligen Bedarfen anzupassen. Unter Einbeziehung des Nutzerfeedbacks und entsprechender Bedarfe der Zielgruppe und aufbauend auf den Ergebnissen des Vorgängerprojekts werden im aktuellen Projekt vier Teilziele verfolgt.

Erweiterung: RDMO ist so aufgebaut, dass alle relevanten Daten zum Projektstart erfasst und daraus ein Datenmanagementplan erstellt werden kann, mit dessen Hilfe das Datenmanagement während des gesamten Projektverlaufs organisiert werden kann (aktives Datenmanagement). Diese mit RDMO erstellten aktiven DMPs reflektieren immer den aktuellen Stand aller für das Datenmanagement wichtigen Aspekte und enthalten Angaben zu konkreten Aufgaben und Rollen (Stichwort *actionable data management plan*) (vgl. Miksa et al. 2019). Dazu kommt die besondere Unterstützung zentraler Aufgaben, wie der Kostenabschätzung, des Ingest-Prozesses

und der Interoperabilität mit z. B. Nachweissystemen wie re3data⁸ durch die Entwicklung entsprechender Module und eines Metadatenmodells. Die Implementierung von Schnittstellen und Identifier-Systemen zur Verlinkung zwischen RDMO und den tatsächlichen Daten und zur Übernahme von Metadaten aus Forschungsinformationssystemen (FIS) wird außerdem beabsichtigt.

Integration in die Infrastruktur: Ein weiteres Ziel des Projekts ist die Vereinfachung der Integration von RDMO in die jeweiligen Infrastrukturen von Infrastrukturanbietern durch eine standardisierte Installation (z. B. über Docker-Container), die Entwicklung eines integrierten Update-Mechanismus zur besseren Unterstützung der Wartung, sowie die Erweiterung der Unterstützung von Authentifizierungs- und Autorisierungsverfahren. Die vereinfachte Integration ist eine wichtige Voraussetzung für die Etablierung von RDMO als verlässliches Service-Angebot von Infrastrukturanbietern.

Etablierung in der Community: Das dritte Ziel von RDMO ist die Etablierung in verschiedenen Fachdisziplinen (wie Biodiversität, Psychologie etc.), universitären und außeruniversitären Wissenschaftseinrichtungen (wie Universitäten und Leibniz-Instituten) sowie Wissenschaftsorganisationen (wie der Fraunhofer-Gesellschaft), als zentrales Tool für aktives Datenmanagement. Außerdem steht in diesem Bereich die internationale Anschlussfähigkeit im Mittelpunkt (mit z. B. DMPonline⁹ des Digital Curation Centers in Großbritannien oder dem DMPTool¹⁰ der University of California in den USA) sowie die Förderung der konkreten Nutzung des Tools vor Ort in den verschiedenen Communities. Hierfür sind sowohl die zentrale Bereitstellung von Schulungsmaterialien und Online-Tutorials als auch dezentrale Ansätze sinnvoll.

Nachhaltigkeit / Verstetigung: Zentrale Aufgabe in diesem Bereich ist die Erhaltung von RDMO als eigenständiges Tool, damit sich die in den Communities eingesetzten RDMO-Instanzen in der Zukunft austauschen und spezifische Entwicklungen integriert werden können und RDMO insgesamt beständig weiterentwickelt wird. Hierfür arbeitet das Projekt mit verschiedenen Initiativen wie E-Science Baden-Württemberg¹¹ sowie (außer-)universitären Forschungseinrichtungen zusammen.

⁸ <https://www.re3data.org/>.

⁹ <https://dmponline.dcc.ac.uk/>.

¹⁰ <https://dmptool.org/>.

¹¹ <https://mwk.baden-wuerttemberg.de/de/forschung/forschungslandschaft/e-science/>.

2.2 Community

Für eine Software wie RDMO sind Community Building und Bestrebungen zur Sicherung der Nachhaltigkeit eng miteinander verzahnt. Die Nutzung von RDMO als Werkzeug zum Datenmanagement in den Communities, d. h. den verschiedenen Fachdisziplinen und Institutionen, kann die Nachhaltigkeit des Tools sicherstellen. Ziel ist es letztendlich, dass sich die verschiedenen RDMO-Instanzen selbst tragen und auf diese Weise RDMO insgesamt weiterentwickelt wird. Daher ist eine der Hauptaufgaben des RDMO-Projekts die Etablierung einer aktiven Community und der Aufbau einer dezentralen Struktur.

Für das Community-Building bedient sich das Projekt verschiedener Instrumente. Neben klassischer Dissemination und Training spielen insbesondere der Aufbau von Kooperationsstrukturen und die Förderung von Aktivitäten in der Community eine wichtige Rolle. Zentrale Pfeiler sind hierbei technische Dokumentationslösungen und die Erleichterung der Integration von RDMO in die eigene Infrastruktur, sowie die Etablierung verschiedener Foren zum Austausch. Neben den Projekt-eigenen digitalen Kommunikationskanälen (wie z. B. GitHub, Slack, allgemeine Mailingliste), haben sich vor allem die bisher an verschiedenen Orten durchgeführten RDMO-Anwendertreffen bewährt.¹² In diesen konnten sich (potentielle) Anwender*innen zu Fragen der Installation und Anwendung mit dem RDMO-Team und untereinander austauschen. Außerdem konnten weitere Bedarfe gesammelt werden, die nach Prüfung durch das Projektteam ggf. in die Verfeinerung bzw. Anpassung der Projektaufgaben einfließen.

Zur Halbzeit des Projekts sind durch diese unterschiedlichen Aktivitäten verschiedene Kooperationsszenarien etabliert bzw. befinden sich in der Entwicklung. Der Einsatz von RDMO in verschiedenen Zusammenhängen, wie z. B. in Projekten, in Universitäten bzw. Universitätskooperationen, in DFG-Sonderforschungsbereichen, an Leibniz-Instituten, oder regionalen Verbünden zeigt, dass RDMO bereits eine gute Aufnahme gefunden und sich eine Anwender-Community gebildet hat.¹³

¹² Die lokalen Workshops fanden sowohl bei den unmittelbaren RDMO-Projektpartnern als auch bei RDMO-Kooperationspartnern statt. Siehe die Übersicht auf der RDMO-Webseite: <https://rdmorganiser.github.io/workshops/>.

¹³ Siehe die interaktive Visualisierung der Projektkooperationen in Form einer Karte auf der RDMO-Projektwebseite: <https://rdmorganiser.github.io/kooperationen/>.

Darüber hinaus ist der Einsatz von RDMO im Projektantrag „Domain-Data-Protokolle für die empirische Bildungsforschung – DDP-Bildung“ zum BMBF-Call „Kuration und Qualitätsmerkmale von Forschungsdaten“ als Bestandteil des Vorhabens vorgesehen. RDMO wurde als Pilot-Service in die Informationsplattform forschungsdaten.info¹⁴ eingebettet, das Hosting und technische Betreuung erfolgen beim RDMO-Projektpartner KIT (Karlsruher Institute of Technologie). Weitere Kooperationen sind geplant, wie z. B. die Kooperation zwischen RDMO und dem RADAR-Projekt¹⁵ für die Nutzung von RDMO als ein Interface für das RADAR-Angebot¹⁶ und die Integration von RDMO in die erweiterte Cloud-basierte Forschungsdatenmanagement-Umgebung Sciebo¹⁷.

Alle RDMO-Instanzen sind auf der RDMO-Webseite als interaktive Karte visualisiert, wobei zwischen Produktiv- und Testinstanzen unterschieden wird.¹⁸ Während Testinstanzen nur für eine kleine Zahl Testnutzer*innen zugänglich sind, um die Funktionalitäten auszuprobieren, sind Produktivinstanzen allen Anwender*innen der jeweiligen anbietenden Institution zugänglich und können aktiv für das Datenmanagement eingesetzt werden. Es folgt eine Kategorisierung des Einsatzes von RDMO anhand konkreter Beispiele (siehe Tab. 1-3). Die Anordnung erfolgt dabei auf Basis des Umfangs der jeweiligen direkten bzw. potentiellen Anwender-Community, von Projektzusammenhängen zu regionalen Verbünden.

Einsatz von RDMO im Rahmen von Forschungsprojekten	
Name	Standort(e)
EmiMin - Verbundvorhaben Emissionsminderung Nutztierhaltung (BMEL) ¹⁹	RDMO-Instanz ZB Med/Publisso (Köln)
Domain Data Protokolle in der empirischen Bildungsforschung ²⁰	Kooperationsprojekt, Leitung GESIS (Köln), Beteiligung AIP (Potsdam)

¹⁴ <https://rdmo.forschungsdaten.info/>.

¹⁵ <https://www.radar-service.eu/de>.

¹⁶ Diese geplante Kooperation war Thema eines Lightning Talks während der Heidelberger E-Science Tage 2019: <https://www.e-science-tage.de/de/programm>.

¹⁷ <https://www.sciebo.de/>.

¹⁸ <https://rdmorganiser.github.io/kooperationen/>.

¹⁹ <https://www.ktbl.de/themen/emimin/>.

²⁰ Noch keine Projekt-Webseite.

FoDaKo - Forschungsdaten in Kooperation (BMBF) ²¹	Universitäten Düsseldorf-Siegen-Wuppertal
MaMoMaR - Management Molekularer Daten im Research Data Life Cycle (DFG) ²²	Fachhochschule Potsdam und Robert Koch-Institut, Berlin
RADAR ²³	FIZ Karlsruhe

Tab. 1: RDMO-Kooperationsszenarien: Projekte

RDMO als Service einer Infrastruktureinrichtung (z. B. Bibliothek oder Rechenzentrum) an einer Hochschule bzw. außeruniversitären Einrichtung	
Name	Standort(e)
Alfred-Wegner-Institut (AWI), Helmholtz-Zentrum für Polar- und Meeresforschung ²⁴	Bremen
Deutsches Zentrum für Luft- und Raumfahrt (DLR) ²⁵	Weßling
Fachhochschule Potsdam (FHP) ²⁶	Potsdam
Forschungszentrum Jülich ²⁷	Jülich
Friedrich-Alexander Universität Erlangen-Nürnberg ²⁸	Erlangen, Nürnberg
Göttingen eResearch Alliance ²⁹ der Georg-August-Universität Göttingen	Göttingen
Helmholtz-Zentrum Berlin für Materialien und Energie (HZB) ³⁰	Berlin

²¹ <https://www.fodako.de/>.²² Noch keine Projekt-Webseite.²³ <https://www.radar-service.eu/de>.²⁴ <https://www.awi.de/>.²⁵ <https://www.dlr.de/>.²⁶ <https://www.fh-potsdam.de/>.²⁷ <https://www.dlr.de/>.²⁸ <https://fau.de>.²⁹ <http://www.eresearch.uni-goettingen.de>.³⁰ <https://www.helmholtz-berlin.de/>.

Karlsruher Institut für Technologie (KIT), Helmholtz-Gemeinschaft	Karlsruhe
Leibniz-Institut für Astrophysik Potsdam (AIP) ³¹	Potsdam
Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) ³²	Gatersleben
Max Planck Digital Library ³³	München
Physikalisch-Technische Bundesanstalt (PTB) ³⁴	Braunschweig
Ruhr Universität Bochum ³⁵	Bochum
RWTH Aachen ³⁶	Aachen
Stiftung Preußischer Kulturbesitz ³⁷	Berlin
Technische Universität Braunschweig ³⁸	Braunschweig
Technische Universität Darmstadt ³⁹	Darmstadt
Technische Universität Hamburg ⁴⁰	Hamburg
Universität Bayreuth ⁴¹	Bayreuth
Universität Darmstadt ⁴²	Darmstadt
Universität Duisburg-Essen ⁴³	Duisburg, Essen
Universität Hamburg ⁴⁴	Hamburg

³¹ <https://www.aip.de/de>.

³² <https://www.ipk-gatersleben.de/>.

³³ <https://www.mpg.de/>.

³⁴ <https://www.ptb.de/>.

³⁵ <https://www.ruhr-uni-bochum.de/de>.

³⁶ <https://www.rwth-aachen.de/>.

³⁷ <https://www.preussischer-kulturbesitz.de/>.

³⁸ <https://www.tu-braunschweig.de/>.

³⁹ <https://www.tu-darmstadt.de/>.

⁴⁰ <https://www.tuhh.de/tuhh/startseite.html>.

⁴¹ <https://www.uni-bayreuth.de/de/index.html>.

⁴² <https://www.tu-darmstadt.de/>.

⁴³ <https://www.uni-due.de>.

⁴⁴ <https://www.uni-hamburg.de/>.

Universität Hildesheim ⁴⁵	Hildesheim
Universität Stuttgart ⁴⁶	Stuttgart
Universität Trier ⁴⁷	Trier
Westfälische Wilhelms-Universität Münster ⁴⁸	Münster

Tab. 2: RDMO-Kooperationsszenarien: Infrastruktureinrichtungen

RDMO im Rahmen regionaler Verbünde	
Name	Standort(e)
bwFDM-Info (Koordiniertes Forschungsdatenmanagement in Baden-Württemberg) ⁴⁹	Baden-Württemberg, Karlsruher Institut für Technologie, Universität Konstanz
HeFDI (Hessische Forschungsdateninfrastrukturen) ⁵⁰	Hessen, Projektleitung Philipps-Universität Marburg
Landesinitiative NFDI der Digitalen Hochschule NRW ⁵¹	Nordrhein-Westfalen, Projektleitung Universität Duisburg-Essen

Tab. 3: RDMO-Kooperationsszenarien: Regionale Verbünde

2.3 Software

RDMO ist eine interaktive Webanwendung mit der die Nutzer*innen primär über einen Webbrowser interagieren. Als Basis für den serverseitigen Teil der Software dient das in der Programmiersprache Python geschriebene Django-Framework⁵². Über die Python-eigene Paketverwaltung pip wird auf eine Reihe weiterer, etablierter Programmbibliotheken aus der Python- und Django-Community zurückgegriffen.

⁴⁵ <https://www.uni-hildesheim.de/>.

⁴⁶ <https://www.uni-stuttgart.de/>.

⁴⁷ <https://www.uni-trier.de/>.

⁴⁸ <https://www.uni-muenster.de/>.

⁴⁹ <http://bwfdm.scc.kit.edu/>. Einbindung von RDMO in das zentrale Angebot Forschungsdaten.info unter <https://rdmo.forschungsdaten.info>.

⁵⁰ <https://www.uni-marburg.de/de/forschung/kontakt/forschungsdatenmanagement/projekte/hefdi-hessische-forschungsdateninfrastrukturen>.

⁵¹ <https://fdm-nrw.de/>.

⁵² <https://www.djangoproject.com/>.

Als interaktive Webanwendung verfügt RDMO auch über ausgeprägte clientseitige Funktionalitäten, die in der Programmiersprache JavaScript und mit Hilfe des AngularJS 1⁵³-Frameworks realisiert sind. Client und Server stehen dabei über REST-Schnittstellen in Verbindung. Für die Unterstützung verschiedener Ausgabeformate wird pandoc⁵⁴ verwendet. Bei der Auswahl dieser Komponenten wurde insbesondere auf einen ausreichenden Reifegrad der jeweiligen Projekte und der Verbreitung nicht nur in der akademischen Webentwicklungs-Community geachtet.

RDMO wurde vom ersten Tag an als Open Source-Software entwickelt und unter der Apache2-Lizenz⁵⁵ bereitgestellt. Als zentrale Plattform für die Entwicklung, aber auch die Interaktion mit der Community in technischen Fragen, wird die Internetplattform GitHub⁵⁶ verwendet. GitHub dient hierbei nicht nur als Repository, sondern ist auch durch Integrationen mit Werkzeugen der Continuous Integration (Travis CI, Coveralls) und mit Zenodo zur Vergabe von DOIs für Releases verbunden.

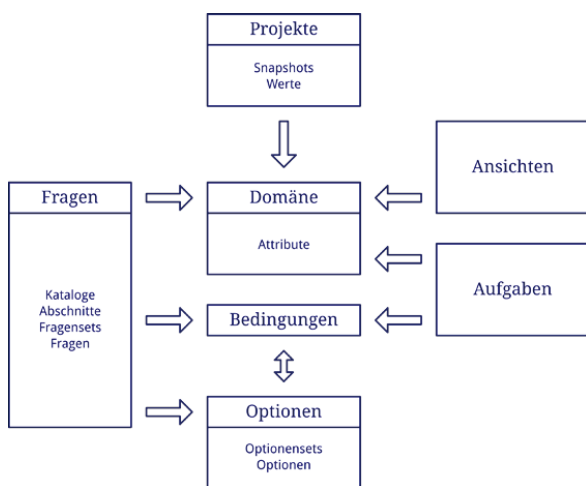


Abb. 1: Überblick über das RDMO-Datenmodell und die verschiedenen Module und ihre Abhängigkeiten.

Die Applikation organisiert sich anhand einer Reihe von die verschiedenen Teilfunktionalitäten von RDMO abbildenden Modulen (siehe Abb. 1). Primärer

⁵³ <https://angularjs.org/>.

⁵⁴ <https://pandoc.org/>.

⁵⁵ <https://www.apache.org/licenses/LICENSE-2.0>.

⁵⁶ RDMO auf GitHub: <https://github.com/rdmorganiser>.

Interaktionspunkt mit den Nutzer*innen ist das Fragen-Modul in dem verschiedene Fragenkataloge konfiguriert werden können, die sich wiederum in Abschnitte, Fragensets und schließlich in einzelne Fragen gliedern. Die Fragen werden durch die Nutzer*innen in einem interaktiven, strukturierten Interview beantwortet. Das Optionen-Modul stellt kontrollierte Vokabulare als Antwortoptionen zur Verfügung. Die Bedingungen legen fest, welche Fragen auf Grund schon gegebener Antworten übersprungen werden können. Nach der Beantwortung der Fragen können die getätigten Eingaben einerseits in identischer Form oder aber, durch das Ansichten-Modul, anhand von hinterlegten Vorlagen, in der von Förderern verlangten textuellen Form neu zusammengesetzt ausgegeben werden. Der Download unterstützt hierbei gängige Formate, wie Microsoft Word, Open Office oder LaTeX. Zusätzlich können auf den Nutzereingaben basierende Aufgaben konfiguriert werden und steht eine API für den programmatischen Zugriff, beispielsweise durch andere Tools, zur Verfügung.

Eingabe- und Ausgabeteil sind durch das RDMO-Domänenmodell verbunden. Dieses besteht aus einem internen, baumförmigen Vokabular sogenannter Attribute, die Fragen und Antworten miteinander verbinden. Die gleichen Attribute werden auch für die Ansichten und Aufgaben verwendet, so dass sich eine Entkopplung zwischen Eingabe und Ausgabe ergibt. Dadurch, dass die Nutzereingaben mit Bezug auf die Attribute gespeichert werden, können Projekte auch zwischen verschiedenen RDMO-Instanzen über ein spezielles XML-Format exportiert bzw. importiert werden (so lange das gleiche Domänenmodell verwendet wird). Eine ähnliche Funktionalität erlaubt auch eine Weitergabe von Fragenkatalogen, Ansichten und anderen Inhalten zwischen den Betreibenden von RDMO-Instanzen (siehe auch unten).

Besonderes Augenmerk gilt der Anpassbarkeit von RDMO. Die beschriebenen Inhalte (Fragenkatalog, Ansichten, Domäne etc.) können alle über ein Webinterface bearbeitet werden. Hierfür sind kein Dateizugriff auf den Server oder Programmierkenntnisse nötig. Auch die Software ist so aufgebaut, dass sich leicht Anpassungen am Aussehen der Seite vornehmen lassen (Corporate Identity), die eigene RDMO-Instanz jedoch weiter kompatibel zur zentral gepflegten RDMO-Bibliothek bleibt. Die Installation selber setzt nur einen Linux-Server und normale Administrationskenntnisse voraus. RDMO besitzt eine eigene Nutzerverwaltung, integriert aber auch OAUTH2 Provider (z. B. ORCID) und LDAP-System oder lässt sich innerhalb einer Shibboleth-Föderation betreiben.

2.4 RDMO Metadatenmodell

Wie bereits oben beschrieben, erfolgt die Eingabe in RDMO in Form eines strukturierten Interviews (Fragenkatalog) über ein interaktives Webinterface. Die eingegebenen Informationen können auf verschiedene Art und Weise aggregiert, kombiniert und ausgegeben werden: Ausgabe der Fragen und Antworten als Text; Ausgabe über spezielle DMP-Vorlagen, wie z. B. Horizon 2020; Ausgabe bestimmter Aufgaben um zukünftige Handlungsbedarfe anzuzeigen, wie z. B. Termine zur Datenveröffentlichung; Informationen zur Kostenabschätzung für das Datenmanagement; sowie Nachnutzung durch andere Software-Werkzeuge über programmierbare Schnittstellen (APIs) (vgl. auch Neuroth et al. 2018, S. 62). Diese verschiedenen Optionen unterstreichen die Funktionalität von RDMO, nicht – wie viele andere Tools für das Forschungsdatenmanagement – die Generierung eines textuellen DMP in den Vordergrund zu rücken. RDMO unterstützt das Forschungsdatenmanagement über den ganzen Projektzeitraum unter Einbeziehung aller daran beteiligten Akteure (wie z. B. Wissenschaftler*innen, Datenmanager*innen, IT-Expert*innen, Institutsleiter*innen etc.).

Die Grundlage für die komplexen Abfrage- und Ausgabefunktionalitäten bildet das interne Datenmodell von RDMO (siehe Abb.1, vgl. Neuroth et al. 2018, S. 62).⁵⁷ Die Fragen (und die dazu eingegebenen Antworten) des Fragenkatalogs sowie die verschiedenen Ansichten und Aufgaben referenzieren auf ein baumförmiges Domänenmodell. Das jeweils aktuelle Domänenmodell steht auf GitHub im XML-Format zum Download zur Verfügung.⁵⁸ Innerhalb des RDMO-Domänenmodells werden die Fragen und Antworten auf sogenannte Attribute referenziert, wobei die Attribute einzelne Informationseinheiten (Frage- und Antwortoptionen) bezeichnen. Die einzelnen Elemente des Domänenmodells sind mit eindeutigen Bezeichnern versehen, um die semantische Interoperabilität der einzelnen RDMO-Instanzen untereinander, aber auch mit Datenmodellen wie DataCite⁵⁹, oder perspektivisch CERIF⁶⁰ (Common European Research Information Format) oder dem Kerndatensatz

⁵⁷ Vgl. auch Heger 2016. Diese Arbeit ist jedoch aufgrund der Überarbeitung des RDMO-Domänenmodells im Jahr 2018 in einigen Teilen überholt.

⁵⁸ <https://github.com/rdmorganiser/rdmo-catalog/blob/master/rdmorganiser/domain/rdmo.xml>.

⁵⁹ <https://datacite.org/>.

⁶⁰ <https://www.eurocris.org/cerif/main-features-cerif>.

Forschung⁶¹ zu ermöglichen. Die semantische Interoperabilität der verschiedenen RDMO-Instanzen untereinander erlaubt den Datenexport für die Übertragung der in einer Instanz vorhandenen Daten auf eine andere Instanz, z. B. im Falle eines Umzugs an eine neue Institution. Darüber hinaus bildet die Interoperabilität mit anderen Datenmodellen, wie zum Beispiel DataCite, die Voraussetzung für die breitere internationale Anschlussfähigkeit von RDMO und die Standardisierung des Datenmanagements (Stichwort *Machine Actionable Datamanagement Plans*).

Im Vorläuferprojekt wurden bereits im Rahmen der Anforderungsanalyse etwa 70 User Stories bezüglich inhaltlicher Aspekte und der Usability von RDMO aus der Sicht verschiedener Stakeholder/Akteure definiert (ein Ansatz der agilen Software-Entwicklung) (vgl. Neuroth et al., S. 58-59). Einige dieser User Stories betreffen Szenarien in denen z. B. Datenmanager*innen oder Institutsleiter*innen Statistiken über verschiedene Einrichtungen oder eine spezifische Einrichtung erstellen oder Infrastrukturanbieter bestimmte Antworten von DMPs abfragen wollen, um Planungen, z. B. zur rechtzeitigen Bereitstellung von Speicherplatz, vornehmen zu können (vgl. ebd., S. 59). Die Anforderungen an RDMO zur Erfüllung dieser beispielhaften Szenarien betreffen die Fragenstruktur und -logik bzw. die Antwortmöglichkeiten – so erhöht der weitgehende Verzicht auf Freitextfelder die Vergleichbarkeit der Antworten über verschiedene Projekte hinweg – und ist eng mit dem RDMO-Domänenmodell zusammenhängenden Bereichen wie Exportmöglichkeiten und Schnittstellen verbunden (vgl. ebd., S. 59).

Die Bereitstellung des RDMO-Quellcodes als Open Source-Software erlaubt es interessierten Institutionen RDMO ausgiebig zu testen und dem Projekt agil auf Feedback aus der Community zu reagieren. So war Feedback aus der Community der Anlass zur Vereinfachung des RDMO-Domänenmodells in der ersten Phase der aktuellen Projektlaufzeit (2018). Im Zuge dieser Überarbeitung wurden die Import- und Exportfunktionen erweitert und einige grundsätzliche Änderungen vorgenommen, mit dem Ziel einer höheren Austauschfähigkeit. Zusätzlich reflektieren diese Änderungen aus der Teilnahme an der RDA-Interest Group Active Data Management Plans⁶² und

⁶¹ <https://kerndatensatz-forschung.de/>. Der Kerndatensatz Forschung enthält momentan nur rudimentäre Ansätze zu mit Forschungsdaten verbundenen Fragen.

⁶² <https://rd-alliance.org/groups/active-data-management-plans.html>.

Feedback im Rahmen der Vorstellung des RDMO-Datenmodells beim RDA-Treffen 2018 in Berlin⁶³ gewonnene Erkenntnisse.

In der zweiten Phase des Projekts steht die weitere Vereinfachung des RDMO-Domänenmodells mit Hinsicht auf ein zukünftiges definierte RDMO-Metadatenmodell im Mittelpunkt. Hierbei wird besondere Aufmerksamkeit der Unterstützung der Aufgaben Kostenabschätzung und Rolleneinteilung, sowie des Ingest-Prozesses und der Interoperabilität mit Nachweissystemen, insbesondere DataCite, d. h. dem DataCite-Metadata Schema⁶⁴, sowie der Einbeziehung der Ergebnisse der RDA-Interest Group Active Metadata Plans⁶⁵, gewidmet. Weiterhin ist es vorgesehen, das RDMO-Datenmodell (sowie ggf. das zukünftige RDMO-Metadatenmodell), ganz im Sinne der FAIR-Prinzipien, offen für die Community bereit zu stellen.

Referenzen

- DBV - Deutscher Bibliotheksverband (2018): Wissenschaftliche Bibliotheken 2025. Verfügbar unter: https://www.bibliotheksverband.de/fileadmin/user_upload/Sektionen/sektion4/Publikationen/2018_02_27_WB2025_Endfassung_endg.pdf.
- DFG - Deutsche Forschungsgemeinschaft (2018a): Stärkung des Systems wissenschaftlicher Bibliotheken in Deutschland. Ein Impulspapier des Ausschusses für Wissenschaftliche Bibliotheken und Informationssysteme der Deutschen Forschungsgemeinschaft. Verfügbar unter: http://www.dfg.de/download/pdf/foerderung/programme/lis/180522_awbi_impulspapier.pdf.
- DFG - Deutsche Forschungsgemeinschaft (2018b): Förderung von Informationsinfrastrukturen für die Wissenschaft. Ein Positionspapier der Deutschen Forschungsgemeinschaft. Verfügbar unter: http://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier_informationsinfrastrukturen.pdf.
- European Commission Expert Group on FAIR Data (2018): Turning FAIR into Reality. Final Report and Action Plan from the European Commission Expert Group on FAIR Data. Verfügbar unter: <https://doi.org/10.2777/1524>.
- GWK - Gemeinsame Wissenschaftskonferenz (2018): Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI). Verfügbar unter: <https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf>.
- Heger, Martin (2016): Datenmodellierung für Forschungsdatenmanagementpläne. Masterarbeit an der Fachhochschule Potsdam im Fachbereich 5 Informationswissenschaften. Verfügbar unter: http://rdmorganiser.github.io/docs/Heger_MA.pdf.
- Miksa, Tomasz, Simms, Stephanie, Mietchen, Daniel, Jones, Sarah (2019): Ten principles for machine-actionable data management plans. PLoS Comput Biol 15(3): e1006750. Verfügbar unter: <https://doi.org/10.1371/journal.pcbi.1006750>.

⁶³ <https://www.rd-alliance.org/dmp-common-standards-wg-rda-11th-plenary-meeting>.

⁶⁴ Momentan Version 4.2. vom 20.03.2019, vgl. <https://schema.datacite.org/>, Zugriff 25.03.2019.

⁶⁵ <https://rd-alliance.org/groups/active-data-management-plans.html>. Die Ergebnisse der Interest Group sollen während des RDA-Treffens in Philadelphia (02.-04.04.2019) vorgestellt werden.

- Muscella, Silvana, Campos Plasencia, Isabel, Komatsoulis, George A., Mortensen, Andreas, Răim, Toivo, Robida, François, Strick, Linda, Tochtermann, Klaus, Turk, Žiga, Wilkinson, Ross (2018): Prompting an EOSC in practice: final report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC). Luxembourg: European Commission. Verfügbar unter: <https://doi.org/10.2777/112658>.
- Neuroth, Heike, Engelhardt, Claudia, Klar, Jochen, Ludwig, Hens, Enke, Harry (2018). Aktives Forschungsdatenmanagement. In: ABI Technik 38:1, S. 55-64. Verfügbar unter: <http://doi.org/10.1515/abitech-2018-0008>.
- RfII - Rat für Informationsinfrastrukturen (2016): Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. Verfügbar unter: <http://www.rfii.de/?p=1998>.
- RfII - Rat für Informationsinfrastrukturen (2017a): Schritt für Schritt – oder: Was bringt wer mit? Ein Diskussionsimpuls zu Zielstellung und Voraussetzungen für den Einstieg in die Nationale Forschungsdateninfrastruktur (NFDI). Diskussionspapier. Verfügbar unter: <http://www.rfii.de/?wpdmdl=2269>.
- RfII - Rat für Informationsinfrastrukturen (2017b): Entwicklung von Forschungsdateninfrastrukturen im internationalen Vergleich. Bericht und Anregungen. Verfügbar unter: <http://www.rfii.de/?wpdmdl=2346>.
- RfII - Rat für Informationsinfrastrukturen (2018a): Zusammenarbeit als Chance. Zweiter Diskussionsimpuls zur Ausgestaltung einer Nationalen Forschungsdateninfrastruktur (NFDI) für die Wissenschaft in Deutschland. Verfügbar unter: <http://www.rfii.de/?wpdmdl=2529>.
- RfII - Rat für Informationsinfrastrukturen (2018b): Zusammenarbeit als Chance. Zweiter Diskussionsimpuls zur Ausgestaltung einer Nationalen Forschungsdateninfrastruktur (NFDI) für die Wissenschaft in Deutschland. Verfügbar unter: <http://www.rfii.de/?wpdmdl=2529>.