

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

Erfahrungsbericht: Auf dem Weg zum ganzheitlichen Forschungsdatenmanagement

Luca Leipoldⁱ Janine Strakaⁱⁱ Kyanoush Yahosseiniⁱⁱⁱ

2021

Zitiervorschlag

Leipold, Luca, Janine Straka und Kyanoush Yahosseini. 2021. Erfahrungsbericht: Auf dem Weg zum ganzheitlichen Forschungsdatenmanagement. Version 2.0. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 3/2021: S. 27-38. DOI: [10.17192/bfdm.2021.3.8362](https://doi.org/10.17192/bfdm.2021.3.8362).

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

ⁱORCID: [0000-0003-1351-1240](https://orcid.org/0000-0003-1351-1240)

ⁱⁱORCID: [0000-0002-0695-1689](https://orcid.org/0000-0002-0695-1689)

ⁱⁱⁱORCID: [0000-0003-0094-5971](https://orcid.org/0000-0003-0094-5971)

1 Zusammenfassung

Forschungsdatenmanagement im institutionellen Kontext spielt eine immer größere Rolle. Allerdings existieren bisher nur wenige technische Plattformen, die eine Verwaltung der Forschungsdaten gemäß den FAIR-Prinzipien ermöglichen und dabei den gesamten Forschungsdatenlebenszyklus abdecken.

In diesem Beitrag stellen wir das Projekt „Management molekularer Daten im Research Data Life Cycle“ (MaMoDaR) vor. MaMoDaR hat zum Ziel, eine technische Plattform zu entwickeln, die den effizienten und nachhaltigen Umgang mit molekularen Forschungsdaten ermöglicht. Mit der Software DataLinker wurde ein Werkzeug geschaffen, das die Wissenschaftler:innen dabei unterstützt, im Research Data Management Organiser (RDMO) hinterlegte Projekte aufzufinden, Forschungsdaten direkt mit diesen Projekten zu verknüpfen und in öffentlichen Repositorien nach den FAIR-Prinzipien zu veröffentlichen.

Für das Gelingen des institutionellen Forschungsdatenmanagements muss der Mehrwert der technischen Lösung für die Wissenschaftler:innen direkt erfahrbar werden. Neben unseren Erfahrungen bei der Entwicklung der Software erörtern wir daher im Beitrag auch, dass es für den Erfolg einer Plattform nicht ausreicht, eine nutzerfreundliche Softwarelösung bereitzustellen, die sich allein auf das Forschungsdatenmanagement beschränkt. Vielmehr ist es notwendig, die Forschenden auch bei den begleitenden organisatorischen und bürokratischen Prozessen zu unterstützen. Aus diesem Anspruch heraus wurde der Projektfokus von MaMoDaR um die Abbildung entsprechender Prozesse in der Software erweitert.

2 Einleitung

Als Bundesinstitut hat das Robert Koch-Institut (RKI) den gesetzlichen Auftrag, wissenschaftliche Erkenntnisse als Basis für gesundheitspolitische und operative Public-Health-Entscheidungen zu gewinnen. Hierfür betreibt das RKI intensive Forschungsaktivitäten. Die dabei erzeugten, sehr heterogenen Forschungsdaten sollen Wissenschaftler:innen im In- und Ausland zugänglich gemacht werden. Dieses Ziel soll durch das Forschungsdatenmanagement (FDM) unterstützt werden. Im Sinne der FAIR-Prinzipien¹ sollen Forschungsdaten dabei auffindbar, zugänglich, interoperabel und wiederverwendbar gemacht werden. Dafür bedarf es unter anderem einer geeigneten IT-Infrastruktur mit passenden Werkzeugen. Auch wenn es hierfür bereits zahlreiche Einzellösungen gibt, fehlt es bisher an Plattformen, die den gesamten Lebenszyklus der Forschungsdaten abdecken.

Aus diesem Grund hat sich das Projekt „Management Molekularer Daten im Research Data Life Cycle“ (MaMoDaR) zum Ziel gesetzt, einen Prototypen einer integrierten FDM-Plattform für molekulare Daten zu entwickeln.

¹Mark D. Wilkinson et al., „The FAIR Guiding Principles for scientific data management and stewardship,“ *Scientific Data* 3 (2016): 160018. doi:[10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).

Im Folgenden beschreiben wir das Vorgehen innerhalb des Projekts, zeigen Hindernisse auf und teilen unsere Erfahrungen. Insbesondere beleuchten wir dabei den Prozess, die von uns entwickelte Plattform innerhalb und außerhalb des Instituts zu etablieren. Die im Rahmen des Projekts entwickelten Konzepte und technischen Lösungen sind öffentlich verfügbar² und können somit als Grundlage für andere wissenschaftliche Einrichtungen mit ähnlichen Organisationsstrukturen dienen.

3 Ausgangssituation

Im RKI gibt es bisher keinen umfassenden Überblick über aktuelle, abgeschlossene und geplante Forschungsprojekte und die darin erzeugten Daten. Forschungsdaten werden in den jeweiligen Fachabteilungen in über die Jahre gewachsenen Strukturen abgelegt. Dies sind Gruppenlaufwerke im lokalen Netzwerk, Datenbanken, elektronische oder analoge Aktenordner und Laborbücher sowie weitere Speicherorte. Es wäre allerdings von außerordentlicher Relevanz, dass die zum Teil einzigartigen Forschungsdaten, die im Institut erzeugt werden, mit ausreichend Metadaten angereichert, systematisch gespeichert und veröffentlicht werden, sodass eine disziplinübergreifende und externe Nachnutzung ermöglicht wird.

Der Mehrwert der Datennachnutzung lässt sich zum Beispiel an der Disziplin der Digitalen Epidemiologie aufzeigen: Durch die Erweiterung von Standardwerkzeugen der Infektionsepidemiologie durch digitale Methoden trägt die Disziplin dazu bei, dass Wissenschaftler:innen ein besseres Verständnis von aktuellen, aber vor allem auch von zukünftigen Krankheitsausbrüchen und -verläufen in der Bevölkerung erlangen können.³ Dafür müssen große Mengen gesundheitsrelevanter Daten, die unstrukturiert oder strukturiert vorliegen, analysiert werden, um in Echtzeit Aufschluss über den Gesundheitsstatus der Bevölkerung zu geben.⁴ Ein gutes FDM erlaubt es, die für die Digitale Epidemiologie notwendigen Forschungsdaten mit geringem zeitlichem Verzug weltweit zugänglich zu machen. Gerade in gesundheitspolitischen Fragestellungen spielt die internationale Zusammenarbeit eine immer wichtigere Rolle, da sich durch die Globalisierung Epidemien schneller und weitreichender ausbreiten können. Hierfür sind eine ausreichende Datenbasis und offene Daten von zentraler Bedeutung.

Die beschriebene Ausgangssituation macht einen ganzheitlichen Ansatz für ein umfassendes FDM am RKI notwendig. Dafür müssen die einzelnen Organisations-

²Kyanoush Yahosseini, Janine Straka und Luca Leipold, „MaMoDaR: Functional account representing the MaMoDaR project,“ <https://github.com/mamodar> (letzter Zugriff am 11.6.2021).

Kyanoush Yahosseini, Janine Straka und Luca Leipold, „Welcome to MaMoDaR's documentation!“, <https://mamodar-docs-en.readthedocs.io/en/latest/> (letzter Zugriff am 11.6.2021).

Kyanoush Yahosseini, Janine Straka und Luca Leipold, „Index of /doc,“ <http://datalinker.h2888668.stratoserver.net/doc/> (letzter Zugriff am 11.6.2021).

³Robert Koch-Institut, „Evidenz erzeugen, Wissen teilen, Gesundheit schützen und verbessern“ (2017). doi:10.17886/rkipubl-2017-003.

⁴Dirk Brockmann, „Digitale Epidemiologie,“ *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* 63, Nr. 2 (2020): 166–175. doi:10.1007/s00103-019-03080-z.

einheiten des Instituts miteinander verzahnt werden und es muss eine hohe Akzeptanz für die Öffnung von Forschungsdaten und die Nutzung der Plattform bei den Wissenschaftler:innen erreicht werden.

4 Entwicklung eines umfassenden FDM-Tools im Projekt MaMoDaR: Die Plattform DataLinker

Im Rahmen des Projekts MaMoDaR wurde eine Plattform entwickelt, die den Forschenden eine vereinfachte Verwaltung ihrer Forschungsdaten ermöglichen soll. Die Software-Lösung deckt alle Schritte des Forschungsdatenlebenszyklus ab.

Am Beginn eines (Forschungs-)Projekts steht die Planung, die idealerweise mit der Erstellung eines Datenmanagementplans (DMP) einhergeht. Es gibt bereits zahlreiche etablierte Software-Tools für die Erstellung von DMPs. In unserem Projekt MaMoDaR verwenden wir das Tool Research Data Management Organiser (RDMO).⁵ RDMO geht über das statische Erstellen eines DMPs hinaus und bietet kollaboratives Arbeiten, Hilfetexte, Aufgabenverwaltung, das Überführen der Metadaten in exportierbare Vorlagen und weitere Funktionen. Dies soll dazu anregen, von einem DMP als passivem Dokument hin zu einem aktiven FDM zu gelangen.⁶ Für ein ganzheitliches FDM ist damit aber noch nicht der gesamte Lebenszyklus der Forschungsdaten abgedeckt. So hat man in RDMO zum Beispiel ausschließlich Einblick in die eigenen Projekte, und deren Datensätze sind lediglich formal beschrieben. Deshalb strebten wir mit der Entwicklung des DataLinkers an, das RDMO-Tool um weitere Funktionalitäten zu erweitern. Dies beinhaltet, dass die in RDMO hinterlegten Projekte auffindbar werden, Forschungsdaten mit den Projekten verknüpft werden können und die Publikation von Forschungsdaten in öffentlichen Repositorien unterstützt wird.

4.1 Funktionalitäten des DataLinkers

a) Forschungsdaten auffinden

Mit dem DataLinker wurde eine Plattform entwickelt, die die bereits existierenden Daten am RKI leicht und schnell auffindbar macht. Dafür werden die in RDMO erfassten (Projekt-)metadaten über eine Schnittstelle an den DataLinker übergeben. Für die Übermittlung der Projektmetadaten werden im Sinne der Interoperabilität international geltende Metadatenstandards herangezogen, z. B. DataCite und DublinCore. Die Metadaten werden darüber hinaus auf das jeweilige Metadatenchema gemappt, um

⁵Harry Enke, Jochen Klar, Claudia Kramer, Jens Ludwig, Olaf Michaelis, Heike Neuroth, Janine Straka, Robert Ulrich, Kerstin Wedlich-Zachodin und Ulrike Wuttke, „RDMO: Research Data Management Organiser“, <https://rdmorganiser.github.io> (zuletzt geprüft am 11.6.2021).

⁶Heike Neuroth, Claudia Engelhardt, Jochen Klar, Jens Ludwig und Harry Enke, „Aktives Forschungsdatenmanagement,“ *ABI Technik* 38, Nr. 1 (2018): 55–64. doi:[10.1515/abitech-2018-0008](https://doi.org/10.1515/abitech-2018-0008).

über Schnittstellen in andere interne oder externe Systeme überführt werden zu können. Im DataLinker stellen wir eine Suchfunktionalität zur Verfügung, mittels derer alle Beschäftigten die in RDMO hinterlegten Projekte mit den dazugehörigen Metadaten auffinden können. Dafür werden eine Stichwortsuche und eine facettierte Suche angeboten. Um nicht ausschließlich Projekte, sondern auch die zugehörigen Forschungsdaten aufzufinden, werden im DataLinker Forschungsdaten mit dem dazugehörigen Projekt verknüpft. In der Trefferanzeige des Projekts werden dann alle verknüpften Ressourcen angezeigt.

b) Ressourcen verknüpfen

Die Nutzenden können im DataLinker ihre Projekte mit Ressourcen verknüpfen. Ressourcen können alle Arten von projektbezogenen Daten sein, z. B. Forschungsdaten, Dokumentationen zu den Forschungsdaten, Visualisierungen, Berichte, Manuskripte, Websites, Projektordner und jegliche Art von Veröffentlichungen. Der DataLinker verzichtet bewusst auf eine direkte Integration von Forschungsdaten und versteht sich ausschließlich als Zeiger für Speicherorte. Zugriffskontrolle findet daher auch weiterhin auf der Ebene der entsprechenden Systeme statt.

Diese Zugriffskontrolle ist vor allem deshalb relevant, weil im RKI auch sensible Daten erzeugt werden, die ohne Begründung nicht institutsweit geteilt werden können und lediglich zweckgebunden herausgegeben werden (z. B. Schutz personenbezogener Gesundheitsdaten, Schutz vor Missbrauch im Bioterrorismus). Entsprechend gibt der DataLinker bei offenen, unkritischen Forschungsdaten ihren jeweiligen Speicherort an; bei geschlossenen Daten nennt der DataLinker die jeweilige Ansprechperson, damit Kontakt aufgenommen und über eine mögliche interne Nachnutzung gesprochen werden kann. Auch können Angaben über die Nutzungs- und Lizenzbedingungen im DataLinker hinterlegt werden, um den Umgang mit den Daten für die Nutzenden transparent zu machen. Durch die öffentliche Verfügbarkeit der Metadaten wird die Sichtbarkeit auch nicht offener Forschungsdaten erhöht. Die Metadaten erlauben das Nachverfolgen von Forschungsaktivitäten und geben Aufschluss darüber, welche Arten von Daten erzeugt werden. Diese Information würde bei einem System verloren gehen, welches sich ausschließlich auf die Integration offener Forschungsdaten stützt.

c) Forschungsdaten veröffentlichen

Ein wichtiger Aspekt von FDM nach den FAIR-Prinzipien ist die Bereitstellung von Forschungsdaten in öffentlichen Repositorien. Bei diesem Veröffentlichungsprozess werden Wissenschaftler:innen durch den DataLinker unterstützt. Von den Nutzenden im DataLinker ausgewählte Forschungsdaten werden mit Metadaten angereichert. Diese Metadaten werden auf das Metadatenschema des Zielrepositoriums gemappt und automatisch über eine Schnittstelle an dieses übermittelt. Persistente Identifikatoren werden dann zurück an den DataLinker übergeben. Durch diese vereinfachte Möglichkeit der Publikation von wissenschaftlichen Daten wird den Forschenden die Bereitstellung der Daten erleichtert. Im Sinne von Open Science wird dadurch die Sichtbarkeit und Verfügbarkeit von Forschungsdaten auf nationaler wie internationaler Ebene erhöht und somit der wissenschaftliche Austausch unterstützt.

4.2 Einbindung der Nutzenden

Bei der Entwicklung des DataLinkers wurden Forschende, die besonders datenintensiv arbeiten, eng eingebunden. Für einen ersten Überblick wurde eine Umfrage mit diesen Forschenden durchgeführt. Anschließend wurden Expert:innen-Interviews geführt, um den Forschungsalltag der Anwender:innen besser zu verstehen und daraus ein bedarfsgerechtes Software-Konzept zu entwickeln. Auf Basis der Ergebnisse wurde ein Anforderungskatalog an die technische Lösung erstellt.

Eine wichtige Erkenntnis war dabei, dass die Umsetzung von (Forschungs-)Projekten im RKI stark an bürokratische Prozesse gekoppelt ist. So fiel auf, dass vor und während der Durchführung von Projekten, in denen Forschungsdaten erzeugt werden, von den Wissenschaftler:innen viele Anträge gestellt und Genehmigungen eingeholt werden müssen:

- Entscheiden sich Wissenschaftler:innen dazu, ein Drittmittelprojekt zu beantragen, so wird eine Drittmittelanzeige bei der Forschungscoordination notwendig. Diese prüft die Anträge (z. B. hinsichtlich der Personal- und Sachmittelkalkulation), bevor sie an die Förderorganisationen weitergegeben werden.
- Von den Fördermittelgebenden wird teils die Einreichung eines DMPs verlangt.
- Sofern im Forschungsprojekt eine Kooperation angestrebt wird, wird durch das Rechtsreferat ein Forschungsvertrag aufgesetzt, der die genauen Verantwortlichkeiten der jeweiligen Partner:innen regelt.
- Wird bei einer Kooperation biologisches Material ausgetauscht, so muss ein Material Transfer Agreement geschlossen werden.
- Bei sensiblen Daten muss durch die:den Datenschutzbeauftragte:n eine Prüfung vorgenommen werden, inwiefern die Daten verarbeitet und weitergegeben werden dürfen.
- Im Gesundheitsbereich kommt darüber hinaus auch die Ethikkommission zum Einsatz.
- Sollen Manuskripte oder Forschungsdaten veröffentlicht werden, so wird eine Veröffentlichungsanzeige durch die Bibliothek geprüft.
- Bei Bedarf kann die Abteilung Organisationsentwicklung involviert werden, um Projekte im gesamten Prozess mit konkreten Projektmanagementtools und Beratungsleistungen zu unterstützen.

Die Erkenntnis führte im Projekt zu der Zielsetzung, den Wissenschaftler:innen eine Software-Lösung anzubieten, die auch als Organisations- und Steuerungsinstrument auf übergeordneter Ebene funktionieren kann, indem z. B. Mehrfacherfassung von Metadaten vermieden wird. Die Entwicklung der Plattform wurde daher sowohl an den Anforderungen der Wissenschaftler:innen sowie an den aus den bürokratischen Prozessen abgeleiteten Vorgaben ausgerichtet. Kontinuierlich wurden während der Entwicklung Usability-Tests mit Nutzer:innen durchgeführt und in Feedbackschleifen die Software stetig verbessert. Darüber hinaus wurden Gespräche mit den Querschnittsabteilungen geführt, um die organisatorischen Prozesse sinnvoll im System abzubilden. Auf diese Weise konnte schließlich ein erster funktionierender Prototyp fertiggestellt

werden. Dieser bildet den kompletten Datenlebenszyklus ab: von der Erstellung eines Projekts in RDMO bis hin zur Veröffentlichung aus dem DataLinker heraus in Zenodo. Für den Aufbau einer Community wurde am 6. Oktober 2020 ein digitaler Workshop durchgeführt.⁷ In diesem interaktiven Workshop wurde der Prototyp DataLinker mit RDMO sowohl Mitarbeiter:innen des RKI als auch Partner:innen aus anderen wissenschaftlichen Einrichtungen vorgestellt. Dabei konnten die Teilnehmer:innen das Gesamtsystem live nutzen und bewerten. Dieses Feedback diente als Grundlage für die Weiterentwicklung des DataLinker und die Anpassung und Verfeinerung der Prozessgestaltung in RDMO.

Die Erkenntnisse aus dem Workshop und den nutzerzentrierten Prozessen wurden dazu genutzt, die Ergebnisse und Visionen aus der Perspektive des FDMs den Querschnittsabteilungen im RKI zu präsentieren. Dabei wurden auch die Vorteile für die jeweiligen Stakeholder aufgezeigt. Ziel war es insbesondere, für die Idee von gemeinsamen Prozessen für Forschungsprojekte und Forschungsdaten zu werben. In einer offenen Diskussion wurden die Herausforderungen und Wünsche der verschiedenen Stakeholder erfasst.

Daran anschließend wurden alle Erfahrungen aus dem Projekt MaMoDaR zusammengetragen und in einer Rückschau bewertet, um die Erkenntnisse in einen neuen Prozess zu überführen. Dieser beinhaltet die Entwicklung eines Systems, das sich als Andockstelle versteht, um Metadaten aus allen angrenzenden Systemen an einer zentralen Stelle vorzuhalten und zwischen den verschiedenen Systemen auszutauschen. Die angrenzenden Systeme umfassen dabei Expertensysteme aus den jeweiligen administrativen Abteilungen, die das Management von Forschungsprojekten in allen Belangen unterstützen. Weiterhin werden auch Systeme angebunden, die direkt Forschungsdaten und zugehörige Metadaten vorhalten.⁸

5 Schlussfolgerungen für die Ausrichtung des institutionellen FDM

5.1 Integration von organisationalen Prozessen

Die zentrale Erkenntnis im Projektverlauf war, dass FDM im institutionellen Kontext als eng verzahnt mit organisationalen Prozessen zu betrachten ist. Daher müssen FDM-Plattformen diese Prozesse integrieren und abbilden oder zumindest kennen und berücksichtigen. Wir folgern, dass Forschende in beiden Bereichen unterstützt werden müssen, um ihnen das Befolgen der institutionellen Data Policies zu ermöglichen. FDM ist kein alleinstehendes Modul, das an andere Prozesse als zusätzliche

⁷Der Workshop folgte dem Ansatz von Katarzyna Biernacka et al., „Train-the-Trainer Konzept zum Thema Forschungsdatenmanagement“ (2020). doi:[10.5281/zenodo.4322849](https://doi.org/10.5281/zenodo.4322849).

⁸Kyanoush Yahosseini, Luca Leipold und Janine Straka, „Institutional research data management - Findings from the development and introduction of holistic research data management tools,“ in *E-Science-Tage 2021: Share your research data* (Heidelberg: heiBOOKs), i. Dr.

Komponente angegliedert werden kann. Es muss in alle vorhandenen (z. B. administrativen) Prozesse hineingreifen und über alle Prozesse hinweg existieren. Wie dabei alle am FDM Beteiligten eingebunden sowie Verantwortlichkeiten und Kommunikationswege festgelegt werden können, wurde in entsprechenden Referenzmodellen wie DIAMANT⁹ und RISE-DE¹⁰ beschrieben. Dafür werden zunächst die IST- und SOLL-Zustände aller FDM-Bereiche definiert und schließlich ein FDM-Servicekatalog abgeleitet.

Der Aufwand für das FDM liegt vor allem bei den Wissenschaftler:innen selbst, während der Nutzen durch Arbeitsentlastung und Einsparung von Geldern vordergründig in Querschnittsabteilungen wie zum Beispiel bei der IT deutlich wird. Daher muss der Mehrwert einer technischen Lösung für Wissenschaftler:innen stärker herausgearbeitet werden, um die Akzeptanz zu erhöhen. Er muss für alle Beteiligten direkt erfahrbar sein. Die Erkenntnis, dass organisationale Strukturen in das FDM zu integrieren sind, machten wir uns zunutze und bildeten die bürokratischen Prozesse in unseren Tools mit ab. So kann das entwickelte FDM-Tool zusätzlich auch als Leitsystem fungieren, um die Nutzer:innen durch die notwendigen Prozessschritte zu führen. Hierbei gibt es keine feststehende Prozessschrittabfolge, sondern die Schritte ergeben sich flexibel aus der Projektart und der Art der entstandenen oder nachgenutzten Forschungsdaten.

5.2 Das Spannungsfeld zwischen Akzeptanz und Nachnutzung

Mit der übergeordneten Perspektive, strukturelle und organisatorische Prozesse in das FDM-Tool einzubinden, sind wir letztlich von dem Vorhaben abgerückt, eine Lösung ausschließlich für molekulare Forschungsdaten zu entwickeln, denn die identifizierten organisatorischen Prozesse betreffen alle Projekte im Institut, in denen Forschungsdaten erzeugt werden. Demgegenüber steht der Grundgedanke, dass sich die FDM-Dienstleistung möglichst passgenau an der eigenen Fachdisziplin orientieren und sich möglichst exakt in die gängigen Arbeitsprozesse einfügen soll, um eine möglichst hohe Verbindlichkeit und Akzeptanz bei den Wissenschaftler:innen zu erzeugen. Da aber gleichzeitig der Anspruch existiert, die zu entwickelnden Systeme nicht auf Einzellösungen zu beschränken und ein entsprechender Bedarf in allen Fachdisziplinen vorliegt, haben wir entschieden, an einer institutsweiten Lösung zu arbeiten und diese generisch auszurichten, um möglichst vielfältige Anwendungs- und Nutzungsmöglichkeiten zu schaffen. Neben molekularen Daten werden im Institut beispielsweise auch sozialwissenschaftliche Daten für das Gesundheitsmonitoring erzeugt. Diese Erweiterung auf andere Datentypen erleichtert auch die Nachnutzung für Einrichtungen mit anderen Fachschwerpunkten.

⁹Marina Lemaire, Lea Gerhards, Stefan Kellendonk, Katharina Blask und André Förster, „Das DIAMANT-Modell 2.0: Modellierung des FDM-Referenzprozesses und Empfehlungen für die Implementierung einer institutionellen FDM-ServiceLandschaft,“ *Universität Trier eSciences Working Papers* 5 (2020). doi:[10.25353/ubtr-xxxx-f5d2-ffff](https://doi.org/10.25353/ubtr-xxxx-f5d2-ffff).

¹⁰Niklas K. Hartmann, Jacob Boris und Nadin Weiß, „RISE-DE – Referenzmodell für Strategieprozesse im institutionellen Forschungsdatenmanagement“ (2019). doi:[10.5281/zenodo.3585556](https://doi.org/10.5281/zenodo.3585556).

Durch das Feedback von externen Teilnehmenden im Workshop stellte sich allerdings heraus, dass das entwickelte System trotz des generischen Ansatzes bereits zu stark an die Arbeitsabläufe des RKI gekoppelt ist. Die Entwicklungen stellten sich als zu rigide für die Nachnutzung an anderen Einrichtungen heraus. So sind die Organisationsprozesse im RKI im Einzelnen andere als in anderen Einrichtungen: Bevor eine Veröffentlichung von Forschungsdaten möglich ist, muss im RKI z. B. eine Veröffentlichungsanzeige durch die Wissenschaftler:innen erstellt und durch die Querschnitts-abteilungen geprüft werden. Um Wissenschaftler:innen am RKI zu unterstützen, wurden solche Prozessschritte in das System integriert, da dies den Aufwand von Verwaltungsakten herabsetzt und durch die Arbeitserleichterung die Akzeptanz des Systems bei den Nutzenden erhöht. Je mehr aber das System individualisierte Anpassungen enthält, desto eher wird seine Nachnutzbarkeit in anderen Einrichtungen herabgesetzt. Weiterhin wird die Wartung dadurch erschwert, dass an einer zentralen Stelle viele Prozesse gebündelt werden, die durch unterschiedlichste Ansprechpersonen verantwortet werden.

Eine zentrale Herausforderung besteht also darin, die für das FDM entwickelten Systeme möglichst generisch zu halten und dabei gleichzeitig spezifische Anforderungen umsetzen zu können. Ein generisches System ermöglicht eine breite Anwendung und eine Abdeckung aller relevanten Funktionalitäten zur Unterstützung des gesamten Forschungsdatenlebenszyklus. Allerdings wird erst durch spezifische Detailanpassungen an die Strukturen einer Organisation oder Fachdisziplin ein realer Mehrwert geschaffen.

5.2.1 Modulare Systementwicklung

Ein Lösungsansatz für dieses Problem ist die Modularisierung der Systementwicklung. So können erst übergeordnet generische Module, die an allen Einrichtungen Bestandteil der Lösung sein sollen, zur Verfügung gestellt werden, z. B. das Modul der Veröffentlichung in öffentlichen Repositorien. Das Modul kann dann aber an die örtlichen Gegebenheiten angepasst werden und disziplinspezifische Kriterien können zur Anwendung kommen.

Der Grundgedanke, die organisatorischen Prozesse mit dem FDM zu verzahnen, ist nach wie vor gegeben, allerdings stellte es sich als unvorteilhaft heraus, diese Prozesse direkt in die Lösung zu integrieren. Anstatt ein zentrales System zu entwickeln, das alle Bedarfe abdeckt, sollten existierende Komponenten eingebunden und strategisch erweitert werden. Daraus folgt, dass alle Expertensysteme ihre Berechtigung haben. Es sollte eine Anbindung aller Expertensysteme über Schnittstellen angestrebt werden. Solch ein System kann wie in [Abbildung 1](#) dargestellt umgesetzt werden.

So können Metadaten maschinell systemübergreifend miteinander ausgetauscht werden. Dies ist vor allem für die Akzeptanz und Nutzung entscheidend, da FDM für viele Nutzer:innen ein Prozess von vielen ist und nicht notwendigerweise den zentralen Einstiegspunkt für die Beschreibung und Verwaltung von Forschungsdaten darstellt. Das im FDM zu entwickelnde System sollte daher als Referenzsystem fungieren und nicht

die Forschungsdaten selbst vorhalten, sondern auf die Speicherorte in den genannten Expertensystemen verweisen.

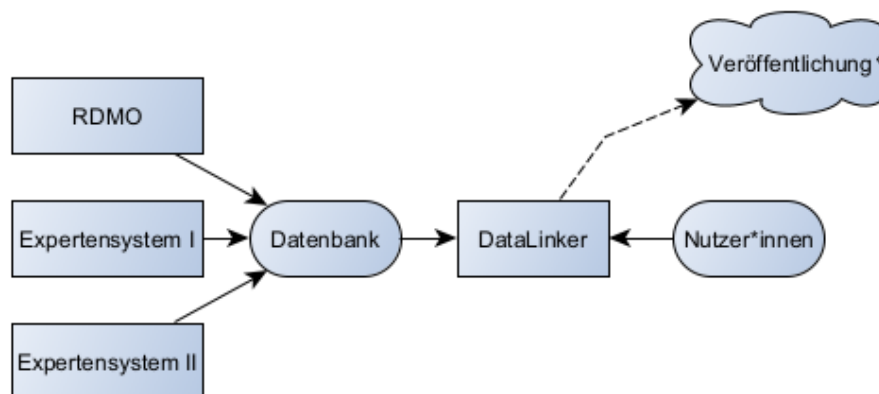


Abbildung 1: Ein Transparenzsystem, welches drei Systeme (RDMO und zwei Expertensysteme) in einer zentralen Datenbank aggregiert und diese über den DataLinker für die Nutzer:innen durchsuchbar macht. Gleichzeitig können die Nutzer:innen eine Veröffentlichung anstoßen.

5.2.2 Stellenwert der Akzeptanz durch die Stakeholder

Bisher haben wir uns in diesem Beitrag vor allem auf die bestehenden Prozesse und die Integration dieser in unser Konzept konzentriert. Um an einer gemeinsam funktionierenden Lösung zu arbeiten, ist es jedoch ebenso wichtig, nicht nur die Prozesse abzubilden, sondern auch die Menschen dahinter einzubinden. Die Komplexität der Perspektiven aller Stakeholder wird deutlich durch die verschiedenen Ebenen, die hier gebündelt werden müssen. Jede Abteilung hat ihre eigene Perspektive auf identische oder ähnliche Abläufe – die Sichtweise des FDM ist dabei ausschließlich eine von vielen. So kommt es vor, dass ein unterschiedliches Verständnis von Begrifflichkeiten existiert, dass abweichende Arbeitsweisen vorherrschen und dass unterschiedliche Gewichtungen bei der Relevanz von Maßnahmen vorgenommen werden, da alle Stakeholder schwerpunktmäßig andere Ziele verfolgen. Aus diesen unterschiedlichen Perspektiven muss möglichst ein gemeinsames Ziel formuliert werden.

Insbesondere im Pandemiegeschehen von SARS-CoV-2 wurde das Projekt MaMoDaR depriorisiert, da alle vorhandenen Ressourcen im Institut gebündelt wurden, um gemäß dem gesetzlichen Auftrag kontinuierlich die aktuelle pandemische Lage zu erfassen, Informationen zu bewerten, eine Risikoeinschätzung für die Bevölkerung vorzunehmen und Empfehlungen für die Fachöffentlichkeit abzugeben. Somit wurden der langfristige Nutzen und die Reichweite des geplanten Systems erst mit Einbindung aller Stakeholder für sie offenkundig. In einer offenen Diskussion wurden die Herausforderungen und Wünsche der verschiedenen Stakeholder zusammengetragen, um

diese für die weitere Entwicklung des Systems ausreichend zu berücksichtigen. Das Verständnis für den Nutzen eines solchen Systems auch von Stellen wie dem Datenschutzteam und der IT-Abteilung ist fundamental, um ausreichend Unterstützung zu erhalten und benötigte IT-Ressourcen in ausreichendem Maße zur Verfügung gestellt zu bekommen.

6 Ausblick und nächste Schritte

Der Beitrag hat gezeigt, dass es für die Umsetzung eines ganzheitlichen institutionellen FDM wichtig ist, FDM und organisatorische Prozesse eng zu verzahnen. Allerdings sollte dabei darauf geachtet werden, nicht alle Prozesse in ein System zu integrieren, sondern vielmehr ein Referenzsystem mit Schnittstellen zu den jeweiligen Expertensystemen aufzubauen und modular anzubinden. Dies hat den Vorteil, dass bereits etablierte Systeme, die auf spezifische Kontexte in der jeweiligen Fachdisziplin ausgerichtet sind, weiterverwendet werden können und die Nutzenden in ihrem Arbeitsalltag abgeholt werden. Neu zu etablierende Module, die für die Verwaltung von Forschungsdaten notwendig sind, sollten zunächst generisch entwickelt und erst dann einrichtungs- oder fachspezifisch erweitert werden, um eine hohe Nachnutzbarkeit zu gewährleisten.

Wir haben unsere Erkenntnisse aus dem Projekt MaMoDaR bewusst in die Weiterentwicklung von RDMO eingebracht, sodass Features und eine verbesserte Nutzer:innenführung der gesamten Community zugutekommen. Aus den Usability-Tests zu RDMO und dem DataLinker ergaben sich überwiegend Änderungswünsche für RDMO. Durch die generische Ausrichtung des RDMO-Tools wird es von vielen unterschiedlichen Einrichtungen über viele Disziplinen hinweg genutzt. Für die Weiterentwicklung haben wir deshalb darauf geachtet, dass Funktionalitäten zunächst generisch implementiert werden, sodass sie von anderen Einrichtungen leichter nachgenutzt und individuell angepasst werden können. Beispielsweise wird auf unsere Veranlassung ein Plugin eingebaut, das eine Verschlagwortung nach international gültigen Standards erlaubt. Dies ist eine Funktionalität, welche die Interoperabilität von Metadaten unterstützt. In einem zweiten Schritt wird dieses Plugin für medizinische und gesundheitsbezogene Forschungsdaten spezialisiert, sodass dort die MeSH (Medical Subject Headings) eingebunden werden. Das Plugin kann dann ebenfalls für die Einbindung anderer Standards verwendet werden. Die aktuell im Entwicklungsprozess befindlichen Features und Verbesserungen werden nach Abschluss über die regelmäßigen Update Releases von RDMO der gesamten Community automatisch zur Verfügung gestellt.

Der DataLinker wird nicht in den Produktivbetrieb im RKI überführt, sondern dient als Prototyp und Vorstudie. Aus den geschilderten Erfahrungen hat sich ergeben, dass wir in einem nächsten Schritt ein neues Referenzsystem entwickeln werden. Dieses deckt sich im Wesentlichen mit den Grundfunktionalitäten des DataLinkers. Allerdings soll es bestehende Expertensysteme besser anbinden, einen Metadaten-

Austausch über alle Systeme hinweg ermöglichen sowie auf die verschiedenen Speicherorte referenzieren. Dabei werden nicht mehr die administrativen Prozesse (z. B. Projektmanagement, Personalmittelkalkulation) in das FDM-Tool mitintegriert und an einer zentralen Stelle zusammengefasst, sondern die Prozesse verbleiben in ihren jeweiligen Systemen, die viel präziser auf den jeweiligen Anwendungsfall und die Nutzenden ausgerichtet sind. Allein die Metadaten werden aus diesen Systemen extrahiert, gemappt und im Referenzsystem durchsuchbar gemacht. Das System soll von Anfang an die Belange des gesamten Instituts berücksichtigen und alle Stakeholder einbeziehen.

Die Implementierung der FDM-Software hängt nicht allein von ihrer Funktionalität und ihrem Mehrwert ab, sondern ein Schwerpunkt liegt auch auf der Steigerung von Akzeptanz und dem dahinterstehenden Kulturwandel in Bezug auf Open Data, der gemeinsam angestrebt werden muss.

Danksagung

Die Autor*innen danken Piotr Wojciech Dabrowski, Bernhard Renard, Henriette Senst, Heike Neuroth und Thilo Muth für ihren umfassenden Einsatz als Ideengeber, geistige Urheber und durch ihre unterstützenden Recherchen sowie Linus Grabenhenrich für seine wertvollen Ratschläge bei der Erstellung des Beitrags.