

FDM

Ivanov, Maier, Nied, Pilch, Schilirò,
Schimmer, Schwarz, Weyd, Weyland

Forschungsdatenmanagement Auch das noch? Und was ist das überhaupt?

**Story-Skript:
Felix D. Müller und die Forschungsdaten**



FDM@HAW.rlp

Dieses Dokument wurde im Rahmen des Verbundprojekts
FDM@HAW.rlp – Nachhaltiger und qualitätsgesicherter Kompetenzaufbau
zu FDM an HAW in Rheinland-Pfalz erarbeitet.

Das Projekt wird vom BMBF in der „Förderrichtlinie zur Förderung von Projekten zum Thema
Nachnutzung und Management von Forschungsdaten an Fachhochschulen“ gefördert.

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung



Finanziert von der
Europäischen Union
NextGenerationEU

Impressum:

Forschungsdatenmanagement. Auch das noch? Und was ist das überhaupt? Story Skript: Felix
D. Müller und *die Forschungsdaten* ist von Igor Ivanov, Daniela Maier, Anja Nied,
Shanna Pilch, Darina Schilirò, Thomas M. Schimmer, Anja Schwarz, Gery Weyd
und Maximilian Weyland unter der Creative Commons Attribution 4.0 International (CC BY 4.0)
<https://creativecommons.org/licenses/by/4.0/deed.de> lizenziert.



Redaktion und Layout:
Die Autorinnen und Autoren

FDM@HAW.rlp
2023

Forschungsdatenmanagement

Auch das noch? Und was ist das überhaupt?

Story Skript: Felix D. Müller und die Forschungsdaten

Igor Ivanov¹, Daniela Maier², Anja Nied³, Shanna Pilch⁴, Darina Schilirò⁵,
Thomas M. Schimmer², Anja Schwarz⁶, Gery Weyd⁷, Maximilian Weyland⁸

1 Hochschule für Wirtschaft und Gesellschaft Ludwigshafen, 2 Hochschule Mainz,
3 Hochschule Trier, 4 Hochschule Worms, 5 Katholische Hochschule Mainz,
6 Hochschule Koblenz, 7 Technische Hochschule Bingen, 8 Hochschule Kaiserslautern

Vorbemerkung

Dieses Story-Skript wurde für die hochschulübergreifende und landesweite Veranstaltung „Forschungsdatenmanagement. Auch das noch? Und was ist das überhaupt? Der FDM-Talk an HAW in RLP“ am 29. März 2023 erarbeitet. Zentrale Aspekte des FDM sollten so allgemeinverständlich an Forschende, Hochschulangehörige und Studierende vermittelt werden. Eine kurze Einführung skizzierte den Kontext von Open Science, Open Access und die Neubewertung von Forschungsdaten durch die Wissenschaftsgemeinschaften und die Wissenschaftspolitik. Das Story-Skript erzählt die Geschichte von Felix D. Müller und seinen Forschungsdaten anhand des Forschungsdatenlebenszyklus.¹ Das Story-Skript ist als Ergänzung zum ebenfalls veröffentlichten Foliensatz zu sehen (<https://doi.org/10.5281/zenodo.7970226>). Es darf gemäß der CC BY-SA Lizenz mit der Nennung der Autoren verwendet, verändert und zu gleichen Bedingungen weitergegeben werden.

Prolog: Felix D. Müller und die Forschungsdaten [Folie 8]

[Folie 9] Felix D. Müller hat gerade seinen Abschluss im Bereich der Wirtschaftswissenschaften gemacht und freut sich: Denn sein Professor hat ihm angeboten, bei ihm im Rahmen des frisch eingeworbenen Verbundprojekts „*Consuming and Covid: Consumption Patterns During the Pandemic*“ zu promovieren. Der Verbund ist interdisziplinär ausgerichtet und in ihm arbeiten Forschende aus den Wirtschaftswissenschaften, den Sozialwissenschaften und aus der Mathematik gemeinsam an der Frage, wie sich das Konsumverhalten während der Pandemie entwickelt hat und welche Probleme es aus wirtschafts- und sozialwissenschaftlicher Sicht gegeben hat. Ziel ist eine Analyse verschiedener Aspekte mit unterschiedlichen Schwerpunkten, u. a. im Bereich der Sportartikel, um krisenresiliente und adaptive Geschäftsmodelle zu entwickeln. Diese sollen sowohl mit abflauerender als auch mit einer ad hoc steigenden Nachfrage zurechtkommen. Aufgrund seiner bisherigen Interessen in seinem Studium passt Felix D. Müller perfekt in dieses Projekt, denn schon in seiner Abschlussarbeit hat er sich mit wirtschaftlichen Aspekten im Sportbereich beschäftigt. Dort ging er der Frage nach, welche Rolle die Social Media-Auftritte einschlägiger Sportschuhhersteller dabei spielen, ob ein Sportschuh von professionellen Sportler:innen gekauft wird oder – obwohl es ein teures Profimodell ist – als Modeerscheinung trendet.

Der Verbund, an dem Felix D. Müllers Professor als Co-Applicant beteiligt ist, besteht aus einer rheinland-pfälzischen Hochschule für angewandte Wissenschaften (HAW), einer Katholischen Hochschule aus Hessen und der Universität Hamburg als Lead. Zwar gibt es noch kein flächendeckendes Promotionsrecht für Absolventen von HAWn, aber durch die Kooperation mit der Universität Hamburg sind kooperative Promotionen mit den beteiligten Hochschulen

1 Zum Verständnis von FDM als Narrativ siehe Lemaire, Marina/Gerhards, Lea/Kellendonk, Stefan/Blask, Katharina/Förster, André: Das DIAMANT-Modell 2.0. Modellierung des FDM-Referenzprozesses und Empfehlungen für die Implementierung einer institutionellen FDM-ServiceLandschaft. Universität Trier eSciences Working Paper, Nr. 05, Trier 2020, S. 9 (<https://ubt.opus.hbz-nrw.de/frontdoor/index/index/docId/1432>, Stand: 19.03.2023).

möglich. Da zum Förderformat explizit die starke Vernetzung des wissenschaftlichen Nachwuchses gehört, wird Felix D. Müller viel mit seinen Mitstreiterinnen und Mitstreitern zusammenarbeiten, da alle Promotionen im Rahmen des Projekts auf den gleichen oder ähnlichen Forschungsdaten aufbauen werden. Alle lernen sich beim Kickoff des Projekts kennen und kommen in einen Austausch. Felix D. Müller kommt unter anderem in ein Gespräch mit der Sozialwissenschaftlerin Antje Schmidt, die an der Katholischen Hochschule in Hessen arbeitet, mit der er sich auch gleich schon über die geplante Erhebung und Auswertung der quantitativen und qualitativen Daten berät, die beide für ihre Promotionsvorhaben benötigen. Auch in großer Runde werden nochmal die anstehenden Themen benannt: Datenerhebungen zu Verkaufszahlen in unterschiedlichen Phasen der Pandemie, die Durchführung von Interviews mit Konsumenten, Produzenten und (Online-)Händlern und vieles mehr. Angesichts des Umfangs, vereinbaren die Promovierenden ein Arbeitstreffen, um die Herausforderungen gemeinsam und koordiniert anzugehen.

Phase I: Planung des Forschungsprojekts [Folie 10]

Nachdem alle Formalitäten geregelt sind, ist es Zeit, aktiv in die Planung des Forschungsvorhabens einzusteigen: [Folie 11] An dem verabredeten Arbeitstreffen der Promovierenden nimmt auch der leitende Professor teil und erläutert nochmals die Inhalte des Antrags und die Ziele des Projekts. Hierbei erwähnt er u. a., dass im Antrag auch bereits Angaben zum Forschungsdatenmanagement gemacht werden mussten. Er weist die Promovierenden darauf hin, diese Punkte im weiteren Projektverlauf unbedingt zu beachten und in ihre Planungen und Arbeitsschritte zu integrieren.

Am Ende verabreden die Promovierenden, sich zeitnah zu einem weiteren virtuellen Arbeitstreffen zusammenzufinden, um die Grundlagen der Zusammenarbeit zu regeln. Bis zu diesem Treffen eine Woche später wollen sich alle den Antrag noch einmal genau ansehen, besonders den Abschnitt zum FDM, also zum Forschungsdatenmanagement.

Bei der Durchsicht des Projektantrags sind die Promovierenden in Zusammenhang mit den verpflichtenden Angaben zum FDM auf den Begriff *FAIR Data* gestoßen. Da dies nicht allen etwas sagt, erläutert Felix D. Müller den anderen kurz, was es damit auf sich hat: Die im Projektverlauf erhobenen Daten sollen den FAIR-Prinzipien entsprechen, also später für andere auffindbar (*findable*), zugänglich (*accessible*), kompatibel (*interoperable*) und wiederverwendbar bzw. nachnutzbar (*reusable*) sein.

Als nächstes verschaffen sich die Promovierenden einen Überblick darüber, welche Art von Daten in dem Projekt erhoben werden sollen. Es werden sowohl quantitative Daten (Verkaufszahlen), als auch qualitative Daten (Interviews) benötigt. Die Sozialwissenschaftlerin Antje Schmidt von der katholischen Hochschule weist direkt darauf hin, dass somit personenbezogene Daten entstehen werden, man sich also mit Aspekten des Datenschutzes sowie des sicheren Datentransfers beschäftigen muss. Sie bietet an, bis zum nächsten Treffen für die anstehenden Erhebungen eine Probanden-Einwilligung zu entwerfen, die auch die besonderen Vorgaben ihrer Hochschule berücksichtigt, die durch die kirchliche Trägerschaft entstehen.

Als nächstes wird besprochen, wie die Daten und andere Materialien innerhalb der Projektgruppe geteilt werden sollen (kollaboratives Arbeiten), so dass alle jederzeit Zugriff haben und mit den Daten arbeiten können. Die Option, die Daten per E-Mail hin und herzuschicken, wird direkt verworfen, weil das schnell unübersichtlich wird und sensible Daten dabei sind.

Schnell ist klar, dass man eine Cloud-Lösung braucht. Einer der Promovierenden schlägt daher vor, Google Drive zu nutzen. Felix D. Müller erinnert sich an einen Vortrag der Datenschutzbeauftragten seiner Hochschule bei der Einführungsveranstaltung für neue Mitarbeiten-

de. Dort hieß es, dass sensible Daten nur bei Anbietern gespeichert werden dürfen, deren Server in Deutschland oder Europa stehen. Daten, die auf Servern in den USA gespeichert werden, wie es bei Google Drive der Fall ist, entsprechen nicht den Vorgaben der DSGVO. Nach Möglichkeit sollten landeseigene Lösungen genutzt werden. Das Problem ist aber für die Promovierenden, dass jedes Bundesland einen eigenen Service hat: Nextcloud in Hamburg, Seafile in Rheinland-Pfalz und die Hessenbox in Hessen. Was also tun bei ihrem länderübergreifenden Projekt? Da die Uni Hamburg den Lead hat, würde man die Daten gerne dort ablegen. Die beiden Promovierenden aus Hamburg werden sich bis zur nächsten Sitzung erkundigen, ob die Nextcloud aus Hamburg mit einem Gastzugang auch von Externen genutzt werden kann. Das Thema müssen sie also noch klären, können sich aber wieder der Diskussion über die zu erhebenden Daten zuwenden.

Ein Mathematiker aus der Gruppe wirft in die Runde, dass die Datenmenge auch eine Rolle spielen könnte. Wie groß werden die Datensätze mit den Verkaufszahlen? Haben Seafile oder Nextcloud da irgendwelche Begrenzungen? Muss man eventuell das lokale Rechenzentrum um extra Platz bitten? Felix D. Müller hat früher schon mit ähnlichen Datensätzen zu tun gehabt und kann beruhigen: Es geht um maximal ein paar Dutzend GB, das ist standardmäßig vorhanden.

Aber das Team soll nicht nur neue Daten erheben: Im Antrag lesen die Promovierenden, dass der antragstellende Professor in „Vor-Corona“ Zeiten bereits eine ähnliche Studie publiziert hat und die Daten daraus als „Kontrollgruppe“ verwendet werden sollen. Die Promovierenden aus Hamburg fragen entsprechend bei ihrem Professor nach und bekommen eine Excel-Datei zugeschickt. Schnell merken die Promovierenden aber, dass sie die Daten nicht verstehen: Die Variablennamen bestehen aus kryptischen Abkürzungen, die Werte selbst sind nirgendwo erklärt und es ist auch nicht festgehalten, welche Auswertungsschritte genau gemacht wurden: „Daten über Daten“, sogenannte Metadaten, fehlen demnach. Nach Rückfrage gibt der Professor zu, dass die Datenerhebung damals durch einen Promovierenden erfolgte, der aber nicht mehr an der Hochschule arbeitet. Der Kontakt ist abgebrochen, daher kann nicht mehr nachvollzogen werden, wie der Datensatz aufgebaut ist. Nach mehreren Wochen Arbeit und vielem hin und her müssen unsere Promovierenden leider feststellen, dass sie den Datensatz nicht nutzen können. Diese Erfahrung motiviert sie, die eigenen Forschungsdaten sauber zu dokumentieren. Sie einigen sich auf Benennungskonventionen bei den Variablen und werden alles nochmal ausführlich in einem Codebook festhalten, so dass alle Informationen zu den Daten jederzeit einsehbar sind. Es war aber nicht alles vergebens: Um doch noch irgendwie eine „Kontrollgruppe“ zu erhalten, führen die Promovierenden eine Literaturrecherche durch in der Hoffnung, doch noch an geeignete Daten zu kommen. Sie stoßen dabei tatsächlich auf eine passende Publikation und interessanterweise wurden die zugrundeliegenden Daten in einer digitalen Datenbank online veröffentlicht. Sogar mit ausreichend Metadaten und zwar mit dem Ziel, dass andere Forschergruppen den Datensatz für die eigene Forschung nachnutzen können. Das bringt unser Team auf die Idee, dass sie die eigenen Daten ja auch später zusammen mit den eigentlichen Publikationen veröffentlichen könnten, damit die nächste Generation Forschender nicht die gleichen Hürden vorfindet wie sie. Damit hätte man auch gleich die Vorgabe des Projektträgers adressiert, die Daten „FAIR“ zu machen. Und damit macht sich das Projektteam an die Datenerhebung ...

Phase II: Datenerhebung & Probleme [Folie 12]

[Folie 13] Nachdem die Planungen abgeschlossen sind geht das Verbundprojekt in die zweite Phase des Forschungsdatenlebenszyklus über. Die Promovierenden werden zwei Sets von Daten haben: qualitative und quantitative. **[Folie 14]** Da Verkaufszahlen keine personenbezo-

genen Daten beinhalten, können sie problemlos verarbeitet werden und sollen auf mehreren Excel-Sheets gespeichert werden. Die Sozialwissenschaftlerin Antje von der Katholischen Hochschule hat, wie versprochen, bis zur nächsten Sitzung begonnen, in Zusammenarbeit mit dem Justitiariat ihrer Hochschule die Einverständniserklärung für die Interviews zu erarbeiten. **[Folie 15]** Als katholische Institution sind sie an das kirchliche Datenschutzgesetz (KDG – Gesetz über den kirchlichen Datenschutz) gebunden, das sich in manchen Belangen von der DSGVO unterscheidet. Daher benötigen sie ein erweitertes Einverständniserklärungsformular, in dem alle relevanten Gesetzestexte genannt werden. Auch muss gewährleistet sein, dass alle Promovierenden Zugriff auf die Daten haben, was ebenfalls in die Erklärung mit aufgenommen werden muss. Des Weiteren fällt ihnen auf, dass sie neben der Einwilligung in die Verarbeitung der Daten in ihrem Projekt eine zusätzliche Einwilligung für die weitere Nutzung zu anderen Forschungszwecken benötigen, damit die Daten nachgenutzt werden dürfen. Für die Nachnutzung benötigen sie aber nicht nur die Erlaubnis in Form der Einverständniserklärung, es muss auch gewährleistet sein, dass die Daten so aufbereitet werden, dass sie die Kriterien beider Datenschutzrichtlinien erfüllen. Um möglichst große Spielräume zu haben, entscheidet die Gruppe, die Daten zu anonymisieren. Im Anonymisierungsschritt entfernen sie daher sämtliche persönlichen Informationen, die Rückschlüsse auf die Identität der Interviewpartner erlauben würden wie z. B. Hinweise auf Geschlecht oder Name.

[Folie 16] Nun müssen sich die Promovierenden noch auf eine Plattform für die kollaborative Zusammenarbeit einigen, um mit der Datenerhebung zu beginnen: Die Hamburger Promovierenden hatten zwischenzeitlich herausgefunden, dass Nextcloud auch mit einem Gastzugang benutzt werden kann, weshalb sie sich auf diese Cloud einigen. Sie einigen sich darauf, alle Datensheets in die Cloud hochzuladen und eine Client-Software zu verwenden. Dies bedeutet, dass alle Änderungen auch automatisch auf die Cloud übertragen werden. Das sieht nach einer guten Lösung aus, sorgt aber dann doch für Probleme. Denn ein Promovierender aus der Mathematik löscht im Laufe des Projekts versehentlich eines der Datensheets bei sich, was augenblicklich auch auf der Cloud aktualisiert wird. So verschwindet auch dort das Datensheet. Und da die Promovierenden davon ausgegangen waren, dass die Cloudvariante als Backup ausreichen würde, hatten sie keine anderen Back-Up Methoden implementiert. Das Datensheet kann nur deswegen gerettet werden, da ein Mitglied der Gruppe zuvor eine Kopie lokal auf einem anderen Rechner gespeichert hatte. Danach sichern die Forschenden ihre Daten nur noch nach der „3-2-1 Methode“, um allen möglichen Eventualitäten vorzubeugen. **[Folie 17]** Hierbei haben sie drei Kopien aller Datensheets, aber auch ihrer eigenen Auswertungen, auf zwei Dateiträgern gespeichert, wovon sie einen an einem anderen Ort aufbewahren. Sie sind sich schnell einig, dass ein USB-Stick kein geeignetes Speichermedium darstellt, da er viel zu schnell verloren gehen kann und dass sie Nextcloud nur noch als kollaboratives Tool verwenden werden und nicht mehr als Backup-Methode. Ab diesem Zeitpunkt speichern sie die Dateien auf einem externen Rechner und auf einer externen Festplatte, die in einem anderen Institut aufbewahrt wird.

[Folie 18] Bei einer ersten Vorstellungsrunde der Ergebnisse des ersten Analyseschritts stellen die Promovierenden in Bezug auf die quantitativen Daten fest, dass sie normalerweise mit sehr unterschiedlicher Software und verschiedenen Formaten arbeiten. Die Mathematiker nutzen R, die Sozialwissenschaftlerin SPSS und Felix D. Müller als Wirtschaftswissenschaftler schwört auf STATA. Damit alle die Daten bzw. die Ergebnisse der Analyseschritte für ihre Arbeit nutzen können, einigt man sich darauf, immer auch eine Version der Ergebnisse in Excel zu exportieren, denn dies ist aus allen drei Programmen möglich. Außerdem gibt es an jeder beteiligten Institution eine Office-Lizenz, so dass alle Zugriff auf Excel haben. Ohne es direkt zu beabsichtigen erfüllen sie mit dem Export der Daten in Excel zumindest unter sich ein FAIR-Kriterium, denn sie machen ihre Daten interoperabel. Richtig FAIR sind ihre Daten

aber damit nicht, denn bei Excel-Dateien handelt es sich um ein sogenanntes proprietäres Dateiformat – wer kein Excel hat, kann sie nicht ohne weiteres öffnen. Die Daten sind daher nicht „accessible“.

[Folie 19] Vor der Auswertung der Analyseschritte werden die Daten aller Promovierenden zusammengeführt, die zu den unterschiedlichen Themensetzungen aufbereitet wurden. Dabei kommt es zu Irritationen, weil die Dateibenennung keiner einheitlichen Systematik und Versionierung folgen, so dass nicht jedem Team-Mitglied klar ist, was die Datensätze beinhalten. Nach ihrer Erfahrung mit den nicht nachnutzbaren Daten hatten sie sich zwar entschieden, ihre Dateien zu versionieren und etablierten ein grobes Schema. Bei einem Workshop hatte Felix D. Müller schon vor längerem gehört, dass gute Dateinamen objektiv, intuitiv und personenunabhängig nachvollziehbar sein sollten. Daher hatten sie eine Dokumentationsdatei angelegt, in der genau aufgeschlüsselt wurde, wie die Dateien zu benennen sind. Nun stellt sich aber heraus, dass nach einer anfänglichen Phase der sorgfältigen Führung die Pflege der Dokumentation eingeschlafen ist. Stattdessen etablierte jedes Team-Mitglied eigene Systematiken. Damit gibt es zwar systematische Benennungen, diese sind aber nicht für alle konsistent. Auch hatten sie nicht bedacht, verschiedenen Kontexten Rechnung zu tragen. So gibt es beispielsweise fünf Dateien mit dem Titel „.final“ oder Variationen dieses Titels, einige von derselben Person. Daher müssen die Promovierenden mühsam anhand von handschriftlichen Notizen zurückverfolgen, welche Datei zu welchem Schritt gehört und alle Dateien einheitlich mit Datumsangaben und selbsterklärenden Namen versehen, damit sie mit ihnen arbeiten können. Dies kostet viel Zeit und ist auch nicht in jedem Fall möglich. Die bereits zu Beginn des Projekts formulierte detaillierte Beschreibung der Erhebungssituation macht es jedoch möglich, viele Schritte zu rekonstruieren. Im Folgenden bemühen sie sich, ihre Systematik weiter zu pflegen. Eine Beachtung dieser Punkte hilft nicht nur den Promovierenden selbst, sondern auch einer weiteren Doktorandin aus den Wirtschaftswissenschaften, die einige Monate später zum Projekt ~~dazu~~ stößt und sich so einen ersten Überblick verschaffen kann.

Mit der Einführung eines guten Datenmanagements enden die Probleme der Promovierenden jedoch leider nicht. Sie hatten sich geeinigt, die Daten auch in Excel zu exportieren. Die verwendeten Excel-Tabellen sind jedoch unterschiedlich formatiert und es kommt zu Fehlern. So hat das Programm beim Export Zahlen in Datumsangaben umgewandelt, was bedauerlicherweise auch erst spät bemerkt wurde. Die Promovierenden müssen nun mehrere Schritte der Auswertung wiederholen und verlieren Zeit, die sie für andere Schritte gut gebrauchen hätten können.

[Folie 20] Nachträglich macht eine FDM-erfahrene Kollegin die Promovierenden darauf aufmerksam, dass ein Datenmanagementplan (DMP) ihnen viele dieser Probleme erspart hätte. Nach längerer Diskussion entscheiden sie aber, dass es in diesem späten Stadium wenig Sinn macht, einen DMP zu erstellen, da sie zwar Zeit verloren haben, aber nun gut zurechtkommen. Außerdem hat auch der Forschungsförderer dies nicht explizit verlangt. **[Folie 21]** Schließlich haben alle ihre Ergebnisse und können anfangen auszuwerten, ihre Dissertationen zu präsentieren und auszuarbeiten.

Phase III: Zielgerade **[Folie 22]**

[Folie 23] Nach drei Jahren ist alles ganz gut gelaufen und die Zielgerade ist erreicht. **[Folie 24]** Die Dissertation ist in der Begutachtungsphase und Felix D. Müller sichtet Publikationswege für sein Werk. Dabei begegnet ihm der Begriff Open Access, von dem er zwar schon gehört hat; ihm ist aber nicht klar, was genau das bedeutet. Er recherchiert und neue Begriffe wie „goldener Weg“ und „APCs“ kommen hinzu. So richtig durchdringt er das Thema aber nicht. Als er schon fast aufgeben will, findet er auf der Website der Hochschule den Hinweis, dass

die Hochschulbibliothek zum Thema Open Access-Publikation berät. Er nimmt Kontakt auf und schnell vereinbart er einen Termin mit der Bibliothekarin.

[Folie 25] Diese erklärt ihm erst einmal, was man genau unter Open Access versteht: also dass wissenschaftliche Publikationen der *scientific community* und der Allgemeinheit mittlerweile unentgeltlich über das Internet zur Verfügung gestellt werden können. Dabei gibt es den Goldenen und den Grünen Weg als zwei Veröffentlichungsstrategien. **[Folie 26]** Beim Goldenen Weg wird das Werk im Open Access erstveröffentlicht. Beim Grünen Weg erfolgt die Zweitveröffentlichung eines bereits publizierten Textes als Open Access-Version.

Felix D. Müller entscheidet sich, seine Dissertation direkt als Open Access-Monografie zu veröffentlichen und wählt damit den Goldenen Weg. Dadurch verbleiben die Verwertungsrechte an seinem Werk in diesem Fall bei ihm und er muss sie nicht an einen Verlag abtreten. Seine Dissertation wird durch die Open Access-Veröffentlichung nun für alle Interessierten online frei verfügbar und kann weiterverbreitet werden. Für die Veröffentlichungskosten, die er als Autor tragen muss, bekommt er eine Förderung durch den hochschulinternen Publikationsfonds. Was er zu diesem Zeitpunkt noch nicht ahnt: Nur wenige Monate nach der Open Access-Veröffentlichung wird Felix D. Müller schon in mehreren Fachaufsätzen zitiert, da sich mit der freien Zugänglichkeit auch die Reichweite seiner Forschung erhöht.

Nun erinnert er sich aber zunächst daran zurück, wie schwierig es war, brauchbare Daten für den Aufbau einer Kontrollgruppe zu finden. **[Folie 27]** Die Daten, die der betreuende Professor ihm damals weitergegeben hatte, entpuppten sich ja als Reinform. Wie gut, dass die Doktorandengruppe dann geeignete Daten im Internet auf der Plattform *Zenodo* gefunden hatte. Da ihm diese veröffentlichten Daten damals so sehr geholfen haben, möchte Felix sich nun bei der Wissenschaftscommunity revanchieren. **[Folie 28]** Er fragt die anderen Doktoranden, ob sie nicht auch ihre zusammengeführten Datensätze mit einer guten Dokumentation veröffentlichen möchten, damit andere Forschende die Daten dann für weitere Forschungszwecke verwenden können. Grundsätzlich finden alle die Idee gut. Die Promovierenden haben sich aber in den letzten Jahren auch Gedanken gemacht, wie sie die Daten für eigene Zwecke weiterverwenden könnten, und möchten im Anschluss an die Promotion ein Startup im Sportartikelbereich gründen. Daher stimmen nicht alle der Veröffentlichung der Daten zum aktuellen Zeitpunkt zu, weil sie die Daten vielleicht noch exklusiv für ihr Startup brauchen. Sobald sie aber mit ihrem Unternehmen erfolgreich sind, möchten sie die Daten publizieren.

Sie beschließen, dass sich Felix D. Müller aber zum jetzigen Zeitpunkt schon einmal nach Veröffentlichungsmöglichkeiten für die Daten umzusehen soll. **[Folie 29]** Er erinnert sich, dass es an seiner Hochschule auch einen Forschungsdatenmanagement-Referenten gibt. „Wenn sich jemand mit der Veröffentlichung von Forschungsdaten auskennt, dann wohl der“, denkt sich Felix. Er sucht den Referenten im Personenverzeichnis der Hochschule und schickt ihm eine kurze E-Mail-Anfrage. Am nächsten Tag hat er bereits eine Antwort, in der der FDM-Referent ein kurzes Treffen für eine erste Besprechung des Falls vorschlägt. Felix stimmt zu und so treffen sich die beiden einige Tage später über BigBlueButton. **[Folie 30]** Der FDM-Referent lässt sich dann erst einmal erklären, welche Daten Felix und seine Mitstreiterinnen und Mitstreiter veröffentlichen möchten und zu welchem Zeitpunkt das passieren soll, um einen geeigneten Publikationsweg zu empfehlen. Im Laufe des Gespräches stellt sich auch heraus, dass bestimmte Daten nur archiviert, aber nicht veröffentlicht werden können. So dürfen zum Beispiel die Originalaufnahmen der durchgeführten Interviews aus datenschutzrechtlichen Gründen nicht veröffentlicht werden.

Der Referent erzählt ihm, dass Daten grundsätzlich über drei verschiedene Publikationswege veröffentlicht werden können und zwar als **[Folie 31]**

1. Supplement zu einem im Verlag veröffentlichten Fachartikel
2. Publikation von Informationen über publizierte Daten in Data Journals
3. Eigenständiges Informationsobjekt in einem Repository

Um Felix einen besseren Überblick zu geben, erklärt ihm der FDM-Referent die verschiedenen Wege **[Folie 32]**:

1. Daten können als Supplement zu einem wissenschaftlichen Artikel über einen Verlag veröffentlicht werden, um die im Artikel beschriebenen Forschungsergebnisse zu verdeutlichen. Hierfür eignen sich vor allem aggregierte Daten, wie z. B. Bilder oder Tabellen.
2. **[Folie 33]** Die Veröffentlichung von Informationen über schon publizierte Daten in Data Journals: Dabei sind die Daten selbst in einer digitalen Datenbank, einem sogenannten Repositorium, abgelegt. Bei den Informationen handelt es sich um ausführliche Dokumentationen der Daten, deren Eigenschaften und Angaben zur potentiellen Nachnutzung. Die Daten im Repositorium und ihre Dokumentation in der Zeitschrift werden mittels eines DOI miteinander verknüpft und sind somit eindeutig auffindbar. Einige dieser Zeitschriften bieten ein Peer-Review-Verfahren an, bei dem der Datensatz sowie seine Dokumentation begutachtet werden. Da es in Felix D. Müllers Projekt aber keine Mittel für diese Art der Datenpublikation gibt und auch an seiner Hochschule kein Fond dafür existiert, ist es für die Gruppe kein geeigneter Weg.
3. **[Folie 34]** Daten können als ein eigenständiges Informationsobjekt in einem Repositorium publiziert werden. Repositorien sind Datenbanken, in denen digitale Objekte wie Datensätze archiviert, dokumentiert und publiziert werden können. Da es unterschiedliche Modelle von Repositorien gibt, erläutert der FDM-Referent Felix die Unterschiede.
 - ◇ Als institutionelle Repositorien werden Dokumentenserver bezeichnet, die von Institutionen betrieben werden, so z. B. von Hochschulen oder Hochschulverbänden.
 - ◇ In disziplinspezifischen Repositorien ist es für die community einfacher, die Daten zu finden. Jedoch existieren noch nicht für alle Disziplinen geeignete Repositorien. Der Referent macht Felix auf *re3data* aufmerksam, eine Suchmaschine für Repositorien, wo man relativ einfach nachsehen kann, welche Fachrepositorien für die Gruppe in Frage kommen könnten.
 - ◇ Schließlich gibt es noch generische Repositorien wie *Zenodo*, in denen unterschiedliche Daten publiziert werden können.

Der FDM-Referent weist darauf hin, dass Felix' Hochschule ein eigenes institutionelles Repositorium hat. **[Folie 35]** Alle publizierten Forschungsdaten bekommen bei der Veröffentlichung einen *Digital Object Identifier* (DOI), also eine Nummer, die die veröffentlichten Forschungsdaten langfristig auffindbar macht. Über die Angabe der DOI können die Forschungsdaten auch mit den dazugehörigen Textpublikationen verknüpft werden.

[Folie 36] Der Referent erwähnt auch, dass die Nutzung des institutionellen Repositoriums für Forschende kostenfrei ist. Felix D. Müller findet die Empfehlung des institutionellen Repositoriums gut. Als er später mit den anderen ehemaligen Promovierenden über die Veröffentlichung spricht, entscheiden sie sich, die geeigneten Daten mit einer Sperrfrist von 2 Jahren zu veröffentlichen. Die Forschungsdaten für die Publikation können sie schon in den nächsten Wochen im Repositorium ablegen, aber erst nach zwei Jahren werden sie einsehbar und zur Nachnutzung für Dritte freigegeben. Die restlichen, nicht veröffentlichten Forschungsdaten wollen sie gemeinsam archivieren.

Um zu regeln, wie genau die veröffentlichten Daten weiterverwendet werden dürfen, müssen sie zunächst noch eine Lizenz vergeben. **[Folie 37]** Der FDM-Referent gibt Felix den Tipp, eine kostenfreie Creative Commons-Lizenz (CC) zu nutzen. Offene Lizenzen sind standardisierte Nutzungsverträge, die den Gebrauch von urheberrechtlich geschützten Text- und Datenpublikationen regeln, also Wegweiser für die Frage: „Was kann ich mit diesem Werk

machen?“. Der Referent erklärt, dass es unterschiedliche CC-Lizenzen gibt. Aber wie kann man die Lizenzen lesen und verstehen? Es gibt 4 Grundelemente (BY = Namensnennung, SA = Weitergabe unter gleichen Bedingungen, NC = Nicht-kommerziell, ND = Keine Bearbeitung). Abhängig von den gewünschten Nachnutzungsbedingungen kann man verschiedene Elemente zusammensetzen und einen der 6 Lizenztypen auswählen (CC BY, CC BY-SA, CC BY-NC, CC BY-ND, CC BY-NC-SA, CC BY-NC-ND).

[Folie 38] Aber wie kann man die Lizenz auf eigene Werke anwenden? Welche CC-Lizenz ist die richtige für uns, fragt sich Felix. Um die CC-Lizenz auszuwählen, die ihren Zielen entspricht, schaut er sich die Infografik an, die ihm der FDM-Referent als Entscheidungshilfe an die Hand gibt. Der Referent macht ihn auch darauf aufmerksam, dass die Gruppe ihre Wahl durch die Abbildung des gewünschten Lizenztyps klar kommunizieren muss. Am besten nehmen sie zusätzlich in der Veröffentlichung einen Link zu der gewählten Lizenz auf, damit Interessierte gleich zu den richtigen Informationen weitergeleitet werden. Felix D. Müller schlägt nach seiner Recherche den anderen Doktoranden eine CC BY-Lizenz vor, die die maximale Nachnutzbarkeit der Daten gewährleistet, denn das war ja ihre Absicht. **[Folie 39]** Also kuratieren sie sorgfältig die Daten für die Publikation und stellen sie mit der Embargofrist von 2 Jahren in das institutionelle Repositorium von Felix D. Müllers Hochschule ein. Die weniger relevanten oder nicht für die Veröffentlichung geeigneten Daten archivieren sie; und schon ist wieder ein Schritt geschafft und eine Phase des Datenlebenszyklus abgeschlossen.

Phase IV: Wertvolle Erfahrungen **[Folie 40]**

[Folie 41] Das Start-Up ist ein Erfolg und Herr Dr. Felix D. Müller sammelt wichtige Erfahrungen. Nachdem er mit seinen ehemaligen Kommilitonen das Unternehmen einige Jahre erfolgreich geführt hat, möchte er sich beruflich verändern und will seine vielfältigen Erfahrungen weitergeben. Er überlegt sich, zurück an die Hochschule zu gehen und in Lehre und Forschung einzusteigen. Er wird auf eine ausgeschriebene Professur im Bereich Wirtschaft und Unternehmensführung an einer HAW aufmerksam. Dr. Felix D. Müller bewirbt sich und wird berufen. Er verpflichtet sich in den Berufungsverhandlungen mit Begeisterung, neben der Lehre auch zu forschen. Gerne möchte er an sein eigenes Forschungsthema anknüpfen und hat auch schon eine Idee, in welche Richtung er seine Forschung weiter vorantreiben möchte. Er legt gleich los, formuliert sein Forschungsthema und begibt sich sogleich auf die Suche nach einem passenden Förderauftrag.

Aus seinen vorherigen Projekten kennt er bereits verschiedene Drittmittelgeber **[Folie 42]** und beginnt im Internet zu recherchieren. Er erinnert sich noch gut an seine Erfahrungen im Umgang mit den Forschungsdaten während seiner Promotion, und auch in seinem Unternehmen hat er die ein oder andere Erfahrung im Datenmanagement gemacht. Daher schaut er sich neben den Förderinhalten insbesondere auch die Anforderungen einschlägiger Drittmittelgeber an das Forschungsdatenmanagement an und muss dabei feststellen, dass es noch keine einheitlichen Verfahren gibt. Die EU verlangt im Rahmen des Förderprogramms *Horizon Europe* spätestens nach sechs Monaten ein DMP und bei Projekten, die länger als 12 Monate dauern, einen Mid-Term-DMP. Nach Projektende muss dann der finale DMP folgen und dieser muss dann schließlich auch veröffentlicht werden. Bei Projekten, die eventuell von den Bundesministerien gefördert werden könnten, stellt er fest, dass es wichtig ist, auf den konkreten Förderauftrag zu achten. So werden z. T. bereits bei einigen Förderaufträgen der Bundesministerien Datenmanagementpläne gefordert, manchmal wird aber auch nur ein „Konzept zum Umgang mit Forschungsdaten“ verlangt. Die Bandbreite der Anforderungen sind seines Erachtens teilweise sehr konkret bis hin zu sehr vage. Sein Fazit lautet somit: „Es gibt viele Drittmittelgeber mit sehr unterschiedlichen Anforderungen an das FDM“.

Schließlich findet er eine passende Förderinitiative für sein Projekt bei der Volkswagen-Stiftung [Folie 43]. Die Ausschreibung umfasst mehrere Seiten und unter den Informationen zur Antragstellung in der Förderinitiative findet er sogar einen kurzen Absatz über den Umgang mit den Forschungsdaten mit Stand vom Dezember 2022. Dort steht:

„Werden in Ihrem Vorhaben Daten verwendet, neu erhoben und/oder verarbeitet, geben Sie die wesentlichen Informationen zum Umgang mit diesen Daten an. Bitte füllen Sie dazu entweder den Datenmanagementplan eines schon identifizierten Zielrepositoriums aus oder machen Angaben im Formular Basis-Datenmanagementplan. Werden in Ihrem Projekt keine Daten in relevanten Umfang genutzt oder erzeugt, geben Sie dies bitte ausdrücklich an.“²

Felix erinnert sich, dass er sich bei seiner Dissertation zusammen mit seinen Kolleginnen und Kollegen explizit gegen die Erstellung eines Datenmanagementplans entschieden hat. Jetzt wird er aufgrund der Anforderungen der VW-Stiftung der Erstellung eines DMP aber wohl nicht mehr ausweichen können und so fragt er sich aufs Neue: Was ist eigentlich ein Datenmanagementplan?

[Folie 44] „Ein Datenmanagementplan (DMP) ist ein Dokument, in dem der beabsichtigte Umgang mit Forschungsdaten beschrieben wird.“³ [Folie 45]. Damit ist er schon einmal schlauer und schaut auf den Basis-DMP der VW-Stiftung. Dieser besteht aus sieben Teilfragen. Der frisch berufene Professor erinnert sich wieder an seine eigene Forschungsarbeit und entwickelt anhand des Fragenkatalogs ein Konzept für den Umgang mit den neu erhobenen Forschungsdaten. Das fällt ihm jetzt sehr viel leichter, weil er ja schon wertvolle Erfahrungen im Forschungsdatenmanagement gesammelt hat. Und so stellen die Fragen für ihn kein Problem dar, weil er aus eigener Erfahrung Hürden und Probleme kennt und diese nun von vornherein ausschließen kann. Dabei muss er sich u. a. mit folgenden Themen beschäftigen [Folie 46]:

1. Nachnutzung vorhandener Daten (Zugänglichkeit, Urheber- und Persönlichkeitsrechte)
2. Datentypen, Datenformate, zu erwartendes Datenvolumen
3. Organisation der Daten (Verwendung von Metadaten, Metadatenstandards, Speicherung der Daten bzw. Metadaten, Kuratierung, Persistenter Identifikator)
4. Lizenzierung (Verwertungs- und Nutzungsrechte) während des Projektes bei kollaborativen Projekten [Folie 46]
5. Lizenzierung (Verwertungs- und Nutzungsrechte) nach Projektende für die Wissenschaft
6. (Langzeit-)Archivierung: Wahl eines entsprechenden Repositoriums, Dauer der Speicherung und Zugänglichkeit
7. Benennung eines Datenverantwortlichen

Er stellt fest, dass die Anforderungen der VW-Stiftung sich sehr nah an einem Standard-DMP orientieren und er beschließt für sein kommendes Projekt einen vollumfänglichen DMP aufzustellen. Für seinen DMP entscheidet er sich für folgende Kapitel [Folie 47]:

1. Allgemeine Angaben zur Forschung: Aufbauend auf seiner Forschungsfrage zu krisenresilienten Vertriebsmodellen, möchte er nun herausfinden, inwiefern sich das Einkaufsverhalten nach der Corona-Pandemie bei erwarteter Verknappung von Wirtschaftsgütern geändert hat.

2 https://www.volkswagenstiftung.de/sites/default/files/documents/MB_131a_d.pdf (Stand: 29.03.2023).

3 Biernacka, Katarzyna/Buchholz, Petra/Danker, Sarah Ann/Dolzycka, Dominika/Engelhardt, Claudia/Helbig, Kerstin/Jacob, Juliane/Neumann, Janna/Odebrecht, Carolin/Petersen, Britta/Slowig, Benjamin/Trautwein-Bruns, Ute/Wiljes, Cord/Wuttke, Ulrike: Train-the-Trainer-Konzept zum Thema Forschungsdatenmanagement, Version 4 [Computer software]. Zenodo, <https://doi.org/10.5281/zenodo.5773203>, S. 49.

2. Verantwortlichkeiten und Ressourcen: Als Projektverantwortlicher ist er auch Verantwortlicher für das Datenmanagement. Die Aktualisierung des DMP überträgt er einer:m Wissenschaftliche:n Mitarbeiter:in. Da auch in seinem beantragten Projekt wieder personenbezogene Daten erhoben werden sollen, bespricht er sich zudem mit dem Datenschutzbeauftragten und stimmt mit ihm eine Einwilligungserklärung zur Weiterverwendung der erhobenen Daten ab. Während des Projekt wird die hochschuleigene Infrastruktur verwendet.
3. Nachnutzung existierender Daten: Da er die Forschungsdaten aus seiner eigenen Promotion wiederverwendet und diese den FAIR-Prinzipien entsprechen, spricht einer Nachnutzung nichts entgegen.
4. Zu erhebende Daten: Hier legt er die Art der Daten (Bild- und Tonaufnahmen, Interviewbögen, usw.), die Datenformate, sowie die Datenmenge fest. Es wird sich teamintern auf die Software R zur Analyse der Daten geeinigt.
5. Forschungsethik: Dieser Punkt ist in diesem Fallbeispiel nicht relevant.
6. Rechtliche Aspekte (Datenschutz, Urheberrecht): Da die Möglichkeit bestünde, die Ergebnisse auf sein Unternehmen, welches er im Nebengewerbe führt, zu übertragen, entschließt er sich für ein Datenembargo von zwei Jahren. Im DMP legt er zusätzlich das Urheberrecht fest, regelt weitere Formalien zur Nachnutzung und legt die Creative Commons-Lizenz fest. Weiterhin muss er die Daten anonymisieren, um die Kriterien der DSGVO zu erfüllen. **[Folie 48]**
7. Datenorganisation (Speicherung, Organisation, Sicherung, Dokumentation, Dateinamenskonvention): Er erinnert sich an den nicht verwendbaren Datensatz aus seiner früheren Forschungsarbeit und entschließt sich, im DMP ein Codebook zu integrieren, damit eine saubere Dokumentation der erhobenen Daten erfolgt. Um eine Nachnutzung zu gewährleisten, werden zusätzlich offene Dateiformate, wie .csv oder .txt, verwendet.
8. Datenaustausch und -zugang: In Abstimmung mit dem Rechenzentrum seiner Hochschule legt er die internen Zugriffs- und Nutzungsrechte innerhalb Projektes fest zur Gewährleistung des kollaborativen Arbeitens.
9. Langzeitarchivierung und Publikation der Daten: Da an seiner jetzigen Hochschule kein institutionelles Repositorium existiert, entscheidet er sich für ein generisches Repositorium und wählt Zenodo.
10. Qualitätssicherung: Dieser Punkt trifft auf sein geplantes Projekt nicht zu.
11. Zugriff und Nachnutzung: Wie schon in 6. beschrieben, legt er die Nachnutzungsrechte fest.

Nachdem er nun den Umfang eines DMP kennt, stellt er sich die Frage, wie viele Arbeitsstunden für die Erstellung eingeplant werden sollen und wie es sich mit den Kosten für die Erstellung eines DMPs verhält **[Folie 49]**. Er stellt fest, dass sogar in seinem Förderantrag die Frage nach den Kosten für das Forschungsdatenmanagement gestellt wird. Er fragt sich, ob diese Kosten mit beantragt werden können bzw. welche Kosten und in welcher Höhe anfallen. Prof. Dr. Felix D. Müller begibt sich auch hier wieder auf die Suche nach Antworten. Und ja, er findet heraus, dass auch Kosten für das Forschungsdatenmanagement förderfähig sind. Dazu gehören z. B. Personalkosten für das FDM, Materialkosten, z. B. für Hardware, Software bzw. Gerätezeiten, Dienstleistungskosten (also Kosten für nachgenutzte Daten oder Kosten für das Repositorium) und auch indirekte Kosten wie z. B. bei der Bereitstellung von Infrastruktur, also Miete, Heizkosten und Strom. Er findet auch etwas über die Kostenmerkmale heraus und weiß nun, dass förderfähige Kosten für das FDM projektbezogen sein müssen und nicht durch die Grundausstattung der Hochschule abgedeckt sein dürfen. Jetzt fehlen ihm nur noch Zahlen

über die Höhe der förderfähigen Kosten und die Information, wo diese beantragt werden können. **[Folie 50]**. Er recherchiert hierzu und findet heraus, dass es auch hier sehr unterschiedliche Regelungen gibt. So können in manchen Förderprogrammen Kosten für das FDM direkt mit beantragt werden, während es z. B. bei der VW-Stiftung hierfür eigene Förderinitiativen gibt. Weiterhin findet er heraus, dass eine Höhe von 5 % der Projektkosten für das FDM als Richtgröße genannt wird.

Nachdem er seinen DMP nun komplett ausgearbeitet hat, sendet er ihn vor Antragstellung dem/der FDM-Referent:in zu und bittet um eine Begutachtung und ggf. Beratung zur Verbesserung.

