

Das Projekt FORDATIS – Aufbau einer Forschungsdateninfrastruktur in der Fraunhofer-Gesellschaft

Andrea Wuchner

Fraunhofer IRB Stuttgart

Einführung

Forschungsdaten und Forschungsdatenmanagement spielen in der Wissenschaft eine immer bedeutendere Rolle. 2010 verabschiedete die Allianz der Wissenschaftsorganisationen die Grundsätze zum Umgang mit Forschungsdaten. Diese erkennen Forschungsdaten als einen „Grundpfeiler wissenschaftlicher Erkenntnis“ an und legen die Grundsätze des Umgangs mit ihnen fest (Allianz der Wissenschaftsorganisationen, 2010). Zunächst war es ausreichend, die Regeln der guten wissenschaftlichen Praxis (DFG 2013) zu befolgen und Daten mindestens 10 Jahre aufzubewahren. Inzwischen wird gefordert, dass Daten nicht nur konserviert, sondern auch öffentlich zugänglich gemacht werden. So hat die Deutsche Forschungsgemeinschaft im Jahr 2015 Leitlinien zum Umgang mit Forschungsdaten (DFG 2015) veröffentlicht. Das europäische Förderprogramm Horizon 2020 (European Commission 2016) fordert zudem seit 2017 die Teilnahme aller geförderter Projekte am Open Research Data Pilot (European Commission 2016). In beiden Fällen gilt, dass Forschungsdaten aus geförderten Projekten veröffentlicht werden müssen. Auch auf politischer Ebene wird der freie Zugang zu Daten immer stärker eingefordert (Rfll 2016). Mit der European Open Science Cloud (European Commission, 2018) und der Nationale Forschungsdateninfrastruktur NFDI (Rfll 2016) sollen auf europäischer und nationaler Ebene übergreifende Infrastrukturen entstehen, deren Ausgestaltung noch offen ist.

Auch die wissenschaftlichen Verlage erkennen das Vermarktungspotenzial von Forschungsdaten zunehmend. Durch das Herausbringen eigener Datenjournale, die ausschließlich der Publikation von Forschungsdaten dienen, und durch die Verpflichtung der Datenabgabe bei klassischen Publikationen, versuchen sich die Verlage Rechte an wertvollen Forschungsdaten zu sichern.

Als Reaktion auf all diese Entwicklungen arbeiten viele außeruniversitäre Forschungseinrichtungen und Universitäten an der Entwicklung von

Infrastruktur-Angeboten und Services für das Forschungsdatenmanagement in ihrer Institution. So wurde beispielsweise in der Leibniz-Gemeinschaft ein Projekt zur Konzepterstellung eines institutionellen Forschungsdatenmanagements für die Leibniz Universität Hannover ab 2014 durchgeführt (Meyer, Janna, & Soßna, 2017). Innerhalb der Max-Planck-Gesellschaft werden zentrale Services rund um das Forschungsdatenmanagement von Max Planck Digital Library (Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V., 2019) und der Max Planck Computing and Data Facility (Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V., 2013) angeboten. Auch für die Universitäten stellt Forschungsdatenmanagement ein wichtiges Thema dar: Vielfältige Services und Dienstleistungsangebote werden von Rechenzentren und Bibliotheken erbracht. So wurde beispielsweise am Karlsruher Institut für Technologie KIT das Serviceteam Research Data Management gegründet. Dieses bietet Beratung und Dienste, „um in allen Phasen des Forschungszyklus ein effektives Forschungsdatenmanagement gemäß den FAIR-Prinzipien (Findable – Accessible – Interoperable – Reusable) zu fördern. Weiterhin baut das Serviceteam im Rahmen von gemeinsamen Projekten mit Forscherinnen und Forschern auch fachspezifische Informationsinfrastrukturen auf.“ (Karlsruher Institut, kein Datum) Daneben betreiben diverse Universitäten in Deutschland institutionelle Forschungsdaten-Repositoryen.

Begleitet werden diese Infrastruktur-Bemühungen von verschiedenen Initiativen zur Information über und zur Standardisierung von Forschungsdatenmanagement. Auf internationaler Ebene etwa durch die Research Data Alliance (Research Data Alliance, kein Datum), deren deutsche Mitglieder sich inzwischen im Verein RDA Deutschland e.V. (RDA Deutschland e.V., kein Datum) engagieren oder auch national durch die DINI-AG/Nestor AG Forschungsdaten (Deutsche Initiative für Netzwerkinformation e. V., kein Datum). Es gibt eine Vielzahl von Workshops und Konferenzen, die das Thema aufgreifen, beispielsweise die eScience-Tage Heidelberg (Universität Heidelberg, 2018), organisiert vom Projekt bwFDM-Info II (KIT Karlsruhe, 2018) unter Beteiligung des Karlsruher Instituts für Technologie, der Universität Konstanz und der Universität Heidelberg. Als Beratungsangebot im deutschen

Raum wird die Webseite forschungsdaten.info betrieben (Universität Konstanz, 2019), die sich an Wissenschaftler richtet. Hinzu kommt die Seite forschungsdaten.org (Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum GFZ, 2013), die sich an Infrastruktur-Dienstleister richtet.

Die Fraunhofer-Gesellschaft stellt mit 72 Instituten, 25.000 Wissenschaftlerinnen und Wissenschaftler und einem jährlichen Gesamtbudget von 2,3 Milliarden Euro Europas größte Forschungsorganisation für anwendungsorientierte Forschung dar. Im Rahmen der am Menschen ausgerichteten Forschungsfelder "Gesundheit und Umwelt", "Schutz und Sicherheit", "Mobilität und Transport", "Produktion und Dienstleistung", sowie "Kommunikation und "Wissen" entstehen an den Fraunhofer-Instituten große Mengen heterogener Forschungsdaten – Forschungsdaten, die sich hinsichtlich Erhebungsmethoden, verwendeter Standards und Geräten sowie Format stark unterscheiden. Daneben nimmt die Fraunhofer-Gesellschaft innerhalb der deutschen Wissenschaftslandschaft eine Sonderstellung ein: Sie ist lediglich zu 30% grundfinanziert, 70% des Budgets müssen durch öffentliche Forschungsprojekte sowie durch Kooperationsprojekten mit der Industrie eingeworben werden. Als Folge dessen, spielen Rechte am geistigen Eigentum¹ eine besondere Rolle für die Fraunhofer-Gesellschaft: Es ist also im Einzelfall zu prüfen, welche Daten für eine Veröffentlichung geeignet sind (Küsters & Klages, 2019).

Die qualitätsgesicherte Publikation von Forschungsdaten und die Sicherstellung der Erfüllung der internen und externen Anforderungen wurden von der Fraunhofer-Gesellschaft als wichtiges Themenfeld identifiziert. Es gab daher 2014 erste Überlegungen und Bedarfsermittlungen. Es wurde die Frage geklärt, welche Anforderungen an eine Forschungsdateninfrastruktur innerhalb der Fraunhofer-Gesellschaft bestehen und wie diese unter Einbeziehung aller relevanten Stakeholder aufgebaut werden kann. Ein wichtiger Punkt dabei war die Gestaltung des Zusammenspiels der IT in Fraunhofer-Zentrale und – Instituten, des Fachinformationsmanagements der Institute und der zentralen

¹ engl. IPR = Intellectual Property Rights

Dienste. Zu dem bestand an den Instituten ein erheblicher Informations- und Beratungsbedarf im Umgang mit Forschungsdaten.

Im Oktober 2015 verabschiedete der Vorstand der Fraunhofer-Gesellschaft die Fraunhofer-Open-Access-Strategie 2020. Der Aufbau eines institutionellen Forschungsdatenrepositoriums gehörte zu den grundlegenden Maßnahmen zur Umsetzung der Strategie: „Das bestehende zentrale Open-Access-Repositorium »Fraunhofer-ePrints« der Fraunhofer-Gesellschaft wird um ein Forschungsdaten-Repositorium »Fordatis« ergänzt und als integrierte Infrastruktur bereitgestellt“ (Fraunhofer-Gesellschaft, 2015).

Die Umsetzung war Teil des Responsible Research & Innovation-Ansatzes, den die Fraunhofer-Gesellschaft im Rahmen des EU-Projekt JERRI – Joining Efforts for Responsible Research and Innovation (JERRI - Joining Efforts for Responsible Research & Innovation, 2017) verfolgt. Hier soll durch die tiefgreifende Verankerung „Deep Institutionalization“ und Umsetzung verschiedener Dimensionen² in der Fraunhofer-Gesellschaft, darunter auch die Dimension Open Access zu einer verantwortungsbewussten Forschung und Innovation beigetragen werden. In der Dimension Open Access wurde dies durch das Infrastruktur-Projekt “FORDATIS – Forschungsdateninfrastruktur angestrebt, das von Juni 2016 bis Mai 2018 am Stuttgarter Fraunhofer Competence Center »Research Services & Open Science« durchgeführt wurde. Ziele des Projektes waren der Aufbau eines eigenen Forschungsdatenrepositoriums für veröffentlichte und zu veröffentlichende Forschungsdaten der Fraunhofer-Gesellschaft, die Etablierung eines Schulungsangebots für die Fraunhofer-Wissenschaftler und weitere Interessensgruppen zum Thema “Forschungsdatenmanagement” sowie die Einrichtung eines Support-Angebots für das Thema Forschungsdatenmanagement. Eingeleitet wurde das Projekt durch eine Fraunhofer-weite Umfrage zum Thema Forschungsdatenmanagement, die wichtige Fragen in Bezug auf das Thema klären sollte und Einschätzungen zum konkreten Bedarf lieferte.

² Die Dimensionen sind Ethik, Gender, Wissenschaft & Ausbildung, öffentliches Engagement und Open Access

Im folgenden Abschnitt wird die Umsetzung des Projekts beschrieben.

Darstellung der Methodik

Um festzustellen, welche Angebote und Bedarfe zum Thema Forschungsdatenmanagement an den Instituten bereits existieren und um ein Bild des Umgangs mit Forschungsdaten der Wissenschaftler innerhalb der Fraunhofer-Gesellschaft zu gewinnen, wurde im März 2017 eine Fraunhoferweite Umfrage bei IT-Managern³, Fachinformationsmanagern⁴ und Wissenschaftlern der Fraunhofer-Gesellschaft durchgeführt. Als Grundlage für die Konzeption der Umfrage wurden dafür bereits an anderen Forschungsorganisationen durchgeführte Umfragen, wie beispielsweise die Umfrage der HU-Berlin (Simucovic, Kindling, & Schirnbacher, 2013) und die Umfrage der Universität Marburg (Projekt Forschungsdatenmanagement und -Archivierung der Universität Marburg, 2015) als Grundlage genommen und angepasst. Die Umfrage wurde mit Hilfe eines zentralen Online-Tools durchgeführt mit je einem Abschnitt für Wissenschaftler, IT-Manager und Fachinformationsmanager. Die folgende Grafik gibt Auskunft über die Beteiligung an der Umfrage:

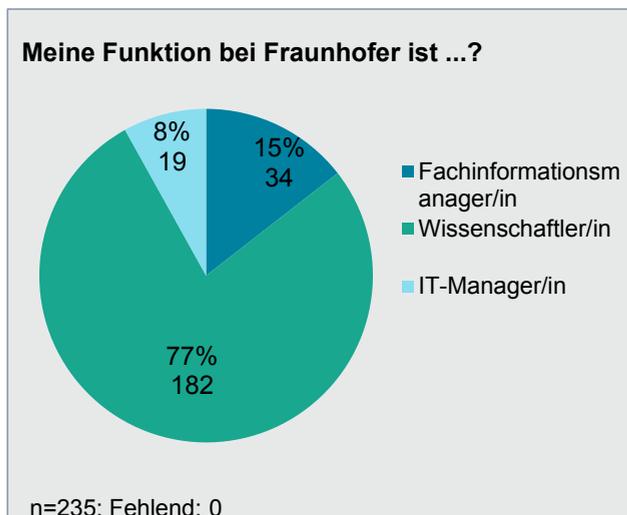


Abbildung 1: Rücklaufquote der Umfrage

³ IT-Manager: Für die IT verantwortlichen Personen an den Fraunhofer-Instituten

⁴ Fachinformationsmanager: Rolle an den Fraunhofer-Instituten, für Fachinformationsversorgung und weitere bibliothekarischen Dienstleistungen am Institut zuständige Person

Forschungsdatenmanagement wurde von allen drei befragten Gruppen als relevantes Thema wahrgenommen. Sowohl IT-Manager als auch Fachinformationsmanager hatten bereits begonnen, sich mit dem Thema auseinander zu setzen. Beide Gruppen wurden von Wissenschaftlern mit Anfragen zum Thema konfrontiert und gaben an, schon Services anbieten zu können.

Für den Umgang mit Forschungsdaten an den Instituten gab es nur in wenigen Fällen festgelegte Prozesse und Verantwortlichkeiten. In den meisten Fällen wurden die Forschungsdaten in Standard-Formaten, wie Microsoft-Formate, csv und jpeg generiert, selbstgeschriebene Software wurde für die Erstellung nur selten verwendet. 17% der befragten Wissenschaftler gaben an, bereits Forschungsdaten veröffentlicht zu haben, 53% wäre nach eigenen Angaben dazu grundsätzlich bereit. 30% der befragten Wissenschaftler, gaben an, Infrastruktur für Speicherung und Veröffentlichung zu benötigen. Ein Schulungs- und Beratungsangebot wurde von allen drei befragten Gruppen gewünscht. Als Antwort auf diese Umfrage, wurden verschiedene Maßnahmen umgesetzt.

Zunächst wurde ergänzend zu bestehenden Strukturen für den Publikationssupport ein zentraler Support für das Thema Forschungsdaten aufgebaut. Dazu wurde eine zentrale Mailing-Adresse eingerichtet, um vor allem die Fachinformationsmanager an den Instituten bei der Beantwortung von Fragen zu unterstützen. Auch die direkte Beratung der Wissenschaftler wurde damit ermöglicht. Bis heute (Stand März 2019) wurden zahlreiche Anfragen per E-Mail und telefonisch bearbeitet. Hierbei handelt es sich um 2nd-Level-Support-Fragen, deren Beantwortung komplexer ist und ausführliche Recherche erfordert. Nachgefragte Themen waren dabei Datenmanagementpläne, Anforderungen der Forschungsförderer, Fragen zur Nutzung von »Fordatis« und zunehmend auch politische Entwicklungen wie die Nationale Forschungsdateninfrastruktur. Der Support berät auch bei der für die Fraunhofer-Gesellschaft wichtigen Fragen, welche Daten veröffentlicht werden sollen.

Ergänzend zum Support wurde 2018 ein eintägiges Schulungs-Angebot für den Themenbereich Forschungsdatenmanagement etabliert, das von Fraunhofer-Instituten gebucht werden kann. Ziel ist es, diese Schulung vor Ort an den einzelnen Fraunhofer-Instituten durchzuführen. Inhaltlich werden dabei Grundlagen des Forschungsdatenmanagements, Anforderungen der Förderorganisationen, rechtliche Rahmenbedingungen, Erstellung von Datenmanagementplänen und Veröffentlichungen von Forschungsdaten behandelt. Bislang wurde die Veranstaltung 10-mal an verschiedenen Fraunhofer-Standorten durchgeführt und hat 140 Teilnehmer erreicht.

Ein weiteres Ziel des FORDATIS-Projekts war der Aufbau einer Fraunhofer-weiten Infrastruktur zur Veröffentlichung von Forschungsdaten. Dafür sollte die bibliografische Datenbank »Fraunhofer-Publica« mit ihrem Volltext-Repository »Fraunhofer-ePrints« um ein Forschungsdatenrepository »Fordatis« für veröffentlichte Forschungsdaten erweitert werden.

Aus Projektantrag und Umfrage ließen sich folgende Hauptanforderungen ableiten:

- Eindeutige Referenzierbarkeit aller publizierten Datensätze: Die Forschungsdaten werden über einen eindeutigen Identifier referenziert. Damit sind sie jederzeit eindeutig auffindbar, auch wenn sich der physikalische Speicherort der Daten ändert.
- Login: Alle Wissenschaftler bei Fraunhofer können sich am System anmelden und dieses benutzen. Ein manuelles Account-Management soll verhindert werden.
- Verknüpfung zwischen Forschungsdatenpublikationen und anderen Outputformaten und Ressourcen: »Fordatis« ermöglicht die Verknüpfung zwischen Forschungsdaten und anderen Ressourcen wie wissenschaftlicher Software und Publikationen.
- FAIR-Compliance: »Fordatis« ermöglicht die Veröffentlichung von Forschungsdaten gemäß den FAIR-Prinzipien (FORCE 11, 2016): Daten werden so veröffentlicht werden, dass sie auffindbar "findable", zugänglich "accessible", interoperabel "interoperable" und nachnutzbar "reusable" sind.
- Einheitliche, standardisierte Erfassung der Metadaten: »Fordatis« bietet allen Fraunhofer-Instituten eine einheitliche, standardisierte Erfassung der Metadaten zu ihren lokal erhobenen heterogenen

Forschungsdatenbeständen. Fraunhofer-spezifische Merkmale, wie zum Beispiel die Zugehörigkeit zu Fraunhofer-Verbänden werden abgebildet.

- Sicherung der Datenhoheit: »Fordatis« wird auf Fraunhofer-eigenen Servern betrieben. So wird sichergestellt werden, dass die Hoheit über die Daten und Metadaten bei der Fraunhofer Gesellschaft selbst verbleibt.
- Google-Indexierung der Metadaten: Die in »Fordatis« gespeicherten Metadaten werden von Google indexiert und sind dort auffindbar. Eine Integration in die Google Dataset Search ist zu prüfen.
- Vernetzung mit anderen Repositorien und Portalen: Durch den Einsatz verbreiteter Metadaten-Standards wird die Vernetzung mit anderen Repositorien, Portalen und Aggregatoren wie zum Beispiel OpenAIRE oder Re3data angestrebt. Damit werden die Anforderungen von Förderorganisationen (zum Beispiel der europäischen Kommission durch die Umsetzung des Open Research Data Pilots in Horizon2020) erfüllt. Auch ist die Sichtbarmachung der Daten in den Fachcommunities durch Harvesting durch die entsprechenden Repositories möglich.
- Bindeglied für Forschungsdaten in den internen Fraunhofer-Datenraum: »Fordatis« stellt das Bindeglied für Forschungsdaten in den Fraunhofer-Datenraum dar. Beim Fraunhofer-Datenraum handelt es sich um einen geschlossenen Datenraum, in dem verschiedene Datenarten wie Publikationsdaten, Patentdaten aber auch Forschungsdaten übergreifend aufgefunden und abgefragt werden sollen. Die veröffentlichten Daten können so Bestandteil übergreifender Abfragen sein.
- Qualitätsgesicherte Erfassung: Analog zu den Publikationsdaten in der »Fraunhofer-Publica« wird eine qualitätsgesicherte Erfassung etabliert. Zu veröffentlichende Forschungsdaten durchlaufen einen Workflow, in dem die beschreibenden Metadaten mehrmals überprüft werden.

Zu Projektbeginn Mitte 2016 zeichnete sich bereits ab, dass die »Fraunhofer-Publica«, die bislang mit der proprietären Software »STAR« betrieben wurde, mittelfristig auf eine moderne Software-Architektur migriert werden soll. Vor diesem Hintergrund wurde entschieden, das Forschungsdaten-Repository nicht in die bestehende »Publica« zu integrieren, sondern zunächst als eigenständiges System zu realisieren, das mittelfristig mit der erneuerten »Publica« zusammengeführt wird.

Um die Entscheidung für die einzusetzende Software vorzubereiten, wurden die oben beschriebenen Anforderungen zunächst in ein Lastenheft formuliert.

<p>Status der Items</p> <p>Die Items sollen in verschiedene Status versetzt werden können:</p> <ul style="list-style-type: none"> a) Gespeichert: Item ist nur für die eingebende Person bzw. für die Nutzerrolle „Bearbeiter“ sichtbar. b) In Kontrolle: Rolle QM hat das Item zur Kontrolle „In Kontrolle“ genommen. c) Freigegeben: nach erfolgreicher Kontrolle wurde das Item freigegeben. Item ist über die Suche recherchierbar und Metadaten werden in das Set für das Harvesting aufgenommen. d) In Bearbeitung: müssen noch Ergänzungen im Datensatz vorgenommen, wird das Item von der Rolle QM zum „Bearbeiter“ zurückgeschickt. Das Item hat den Status „In Bearbeitung“ e) Gelöscht: Wird ein Freigegebenes Item gelöscht, erhält es den Status „gelöscht“, der Metadatensatz wird zusammengedampft und mit dem Vermerk „Dieses Item wurde gelöscht“ versehen. Es soll weiterhin über den Handle aufrufbar sein. Items, die nur gespeichert sind, sollen „spurlos“ gelöscht werden können.
<p>Bearbeitung von Items</p> <p>Items sollen von der Rolle Moderator jederzeit bearbeitet werden können. Metadaten sollen verändert werden können, sowie Dateien hochgeladen bzw. entfernt werden. Freigegebene Items sollen nach der Bearbeitung nur in den Status „Freigegeben“ versetzt werden können.</p>
<p>DOI-Vergabe</p> <p>Die DOI soll von der Itemanzeige und von der Itemkurzanzeige aus möglich sein. Sie soll je Forschungsdatensatz (= Item) möglich sein. Die DOI soll anschließend automatisch in die Metadaten geschrieben werden. Die DOI-Vergabe soll für Moderatoren und Administratoren für Datensätze im Status „Freigegeben“ möglich sein. Während der DOI-Vergabe soll ein Formular angehakt werden, in dem der Mitarbeiter versichert, dass der betreffende Datensatz noch keine DOI hat und dass es sich um ein wissenschaftliches Objekt handelt.</p>
<p>Löschen von Items</p> <p>Freigegebene Items sollen von der Rolle Moderator gelöscht werden können. Dazu wird der Metadatensatz zusammengedampft und mit dem Vermerk „Dieses Item wurde gelöscht“ versehen. Es soll weiterhin über den Handle aufrufbar sein. Gespeicherte Items soll vom Depositor gelöscht werden können. Sie sollen nachhaltig aus dem Repository entfernt werden.</p>

Abbildung 2: Ausschnitt aus dem Lastenheft von Fordatis

Diese wurde durch ein umfangreiches Use-Case-Dokument ergänzt.

Abschnitt	Inhalt
Bezeichner	F-1-5
Name	Einloggen eines Nutzers (Fraunhofer-Wissenschaftler, Data-Curator)
Autoren	wua
Priorität	5
Kritikalität	5
Quelle	wua
Verantwortlicher	wua
Kurzbeschreibung	Der Nutzer loggt sich in Fordatis ein.
Auslösendes Ereignis	Der Nutzer möchte entsprechend seiner Berechtigung in Fordatis arbeiten.
Akteure	Nutzer, Fordatis
Vorbedingung	Account des Nutzers ist bereits angelegt.
Nachbedingung	Der Nutzer ist in Fordatis eingeloggt.
Ergebnis	Login des Nutzers
Hauptszenario	<ol style="list-style-type: none"> 1.) Nutzer ruft Fordatis auf. 2.) Nutzer klickt auf die Schaltfläche „Login“. Die Loginmaske öffnet sich. 3.) Der Nutzer gibt seine Login-Daten ein und klickt auf die Schaltfläche „Login“ 4.) Der Bereich „Meine Datensätze“ mit Schaltflächen für die weiteren den Rechten entsprechenden möglichen Aktionen erscheint. Der Nutzer ist eingeloggt.
Alternativszenarien	
Ausnahmeszenarien	Fordatis ist nicht erreichbar. Der Login funktioniert nicht.
Qualitäten	

Abbildung 3: Ausschnitt aus dem Use-Case-Dokument für Fordatis

Der nächste Schritt war die Auswahl einer geeigneten Software zur Realisierung von »Fordatis«. Folgende Kriterien waren dafür entscheidend:

- Kompatibilität mit der erneuerten »Fraunhofer-Publica« bzw. »Fraunhofer-ePrints« und weiteren Systemen wie dem »Fraunhofer-FIS« (Küsters & Klages, 2019): Zum Projektstart war es abzusehen, dass die bibliografische Nachweisdatenbank »Fraunhofer-Publica« und

das Volltext-Repositorium »Fraunhofer-ePrints« von der proprietären Software STAR abgelöst werden. Bei der einzusetzenden Software für »Fordatis« sollte es sich um eine Software-Lösung handeln, mit der beide Systeme betrieben werden können, damit Synergie-Effekte und Know-How für beide Systeme genutzt werden.

- Open-Source-Produkt: Bei der einzusetzenden Software sollte es sich um ein Open-Source-Produkt handeln, um die Nachteile einer proprietären Software nicht mehr in Kauf nehmen zu müssen. Darunter fallen Lizenz-Gebühren. Auch kann von Weiterentwicklungen einer großen Open-Source-Community profitiert werden. Die Fraunhofer-Gesellschaft als öffentliche Einrichtung kann eigene Entwicklungen ebenfalls weitergeben.
- Große Nutzercommunity: Um die Vorteile eines Open-Source-Produkts möglichst gut auszunutzen, sollte einen Open-Source-Software mit möglichst großer Community und einer großen Anzahl an erfolgreich betriebenen Installationen eingesetzt werden.

Die Entscheidung fiel nach Sichtung von Invenio und Fedora zu Gunsten der Software DSpace. Diese wird in der Version DSpace 6.3 JSPI eingesezt.

Um die Umsetzung des Repositoriums möglichst effizient und mit dem nötigen technischen Know-How durchzuführen, wurde das Vorhaben von einem externen Dienstleister unterstützt. Insgesamt fanden drei Workshops zur Spezifizierung von Funktionalitäten und Metadatenschema und zur Erarbeitung eines Layouts statt, das mit dem Fraunhofer-Corporate-Design konform ist. Die Implementierung wurde im September 2018 begonnen und dauert bis zum heutigen Tag an.

Die oben beschriebenen Anforderungen wurden folgendermaßen realisiert:

- Eindeutige Referenzierbarkeit der Forschungsdaten: Hierfür sollen Digital Object Identifier (DOIs) eingesetzt werden. Diese werden von der TIB-Hannover bereitgestellt.
- Login: Um für alle Wissenschaftler der Fraunhofer-Gesellschaft einen Account mit Rechten zur Dateneingabe bereit zu stellen, wird die Anbindung an den Fraunhofer-Verzeichnisdienst realisiert.
- Verknüpfung zwischen Forschungsdaten und anderen Ressourcen: Die Verknüpfung von Forschungsdaten mit anderen Ressourcen wie Publikationen oder Software erfolgt über ein Metadatenfeld, in das der

Unique Identifier der zu verknüpfenden Ressource eingetragen wird. Die Art der Beziehung kann über ein Auswahlmenü spezifiziert werden.

- FAIR-Compliance: Vom Fraunhofer FIT (Beyan, et al., 2018) wurde eine Evaluierung der FAIR-Compliance von »Fordatis« durchgeführt. Hierfür wurde die FAIR-Metrik (Wilkinson, 2018) der FAIRMetrics Group (Fair Metrics Group, 2018) eingesetzt. Die Kriterien konnten erfüllt werden. Zusätzlicher Handlungsbedarf besteht hinsichtlich eines Konzepts für die Langlebigkeit von Metadaten sowie die Einbindung Community-spezifischer Standards.
- Einheitliche, standardisierte Erfassung der Metadaten: Das Application Profile von »Fordatis« basiert auf dem Dublin-Core-Standard (Dublin Core Metadata Initiative, 2013). Es bietet eine standardisierte, einheitliche Erfassung aller Forschungsdaten. Gleichzeitig werden Fraunhofer-spezifische Merkmale, wie zum Beispiel die Zugehörigkeit zu Fraunhofer-Verbänden abgebildet. Hierfür wurden eigene Felder unter dem Namespace fordatis hinzugefügt.
- Sicherung der Datenhoheit: »Fordatis« wird auf Fraunhofer-eigenen Servern, der »Fraunhofer-Cloud« betrieben. So soll sichergestellt werden, dass die Hoheit über die Daten und Metadaten bei der Fraunhofer Gesellschaft selbst verbleibt. Die Server stehen an mehreren Standorten in Deutschland. Zudem sind sie skalierbar.
- Google-Indexierung der Metadaten: Die in »Fordatis« gespeicherten Metadaten sollen von Google indexiert werden und dort auffindbar sein. DSpace bringt einige Werkzeuge mit, um die Inhalte indexieren zu lassen. Hierfür wird die URL von Fordatis zunächst manuell bei Google registriert. Des Weiteren bietet DSpace ein Sitemap-Feature an, das es Suchmaschinen-Crawler erleichtert, die Inhalte zu indexieren. Die Sitemap enthält eine Liste aller Items, Sammlungen und Communities und verhindert, dass Crawler jede Seite einzeln aufrufen müssen. Diese Sitemap ist sowohl im Footer als auch in der robot.txt verlinkt, damit Google optimal Zugriff darauf hat. Ein weiterer Schritt ist die Anpassung der robots.txt-Datei, welche Einstellungen zum Crawler-Zugriff enthält. Hier ist es wichtig, dass die Datei direkt auf höchster Ebene (Top Level) in DSpace eingebunden wird. Des Weiteren muss sichergestellt sein, dass verschiedene URLs indexiert werden können. Um die Indexierbarkeit auch bei komplexen Layout-Anpassungen zu gewährleisten, beinhaltet das <head>-Element der Item-HTML-Seite Item-Metadaten. Weiterhin soll die Weiterleitung von direkten Download-Zugriffen auf Dateien zu den Landing-Pages vermieden werden, ebenso die Generierung von PDF-Deckblättern, welche DSpace anbietet.

- Vernetzung mit anderen Repositorien und Portalen: DSpace enthält eine OAI-PMH-Schnittstelle und setzt intern das Dublin Core-Metadatenformat ein. Dadurch ist das Harvesten der Metadaten in andere Repositorien möglich. Für die Lieferung an OpenAIRE wird ein Mapping von Dublin Core zum Standard DataCite (DataCite, 2014), Version 3.1 eingesetzt.
- Bindeglied für Forschungsdaten in den internen »Fraunhofer-Datenraum«: »Fordatis« soll das Bindeglied für Forschungsdaten in den »Fraunhofer-Datenraum« darstellen. Beim »Fraunhofer-Datenraum« handelt es sich um einen geschlossenen Datenraum, der sich derzeit im Aufbau befindet und auf den alle Fraunhofer-Institute Zugriff haben sollen. Dort werden verschiedene Datenarten, die innerhalb der Fraunhofer-Gesellschaft entstehen, wie Publikationsdaten, Patentdaten aber auch Forschungsdaten, zusammengeführt, so dass sie übergreifend aufgefunden und abgefragt werden können. Die veröffentlichten Daten können so Bestandteil übergreifender Abfragen sein. Die Anbindung an den Datenraum soll über eine RESTAPI gewährleistet werden.
- Qualitätsgesicherte Erfassung: Analog zu den Publikationsdaten in der »Fraunhofer-Publica« soll eine qualitätsgesicherte Erfassung etabliert werden. Dafür durchlaufen die Forschungsdaten innerhalb von »Fordatis« einen Workflow, in dem die beschreibenden Metadaten mehrmals überprüft werden und abschließend freigegeben werden.

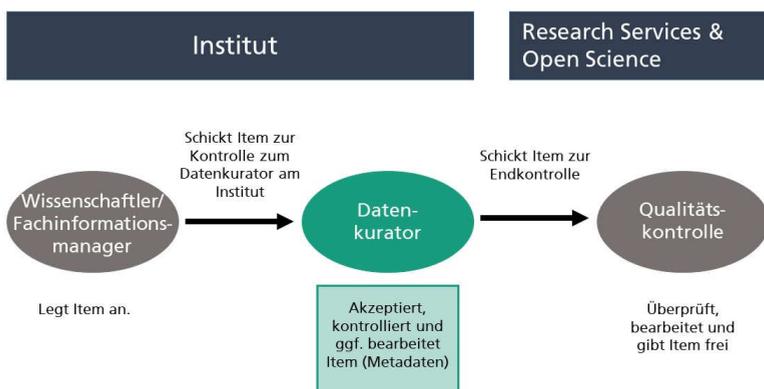


Abbildung 4: Workflow zur Qualitätssicherung

Die Daten werden vom Wissenschaftler bzw. vom Datenkurator in »Fordatis« eingegeben. Anschließend werden sie von einem zuvor bestimmten Datenkurator im Institut, der oftmals in der Institutsbibliothek angesiedelt ist, überprüft, gegebenenfalls bearbeitet und zur zentralen Qualitätskontrolle an die Abteilung »Research Services & Open Science« weitergeleitet. Hier werden die Datensätze nochmal überprüft und dann freigegeben. Erst dann sind sie für die Öffentlichkeit in »Fordatis« sichtbar. Die Kontrolle der Daten erstreckt sich nur auf die Metadaten, hierfür wurde ein eigenes Regelwerk erstellt, das festlegt, welche Daten in welcher Form in welches Feld einzutragen sind. Eine inhaltliche Kontrolle der Forschungsdaten kann nicht geleistet werden, hierfür sind die Wissenschaftlerinnen und Wissenschaftler selbst verantwortlich.

- **Rechtliche Compliance:** Im Zusammenhang mit Forschungsdaten und dem Betrieb eines Repositoriums gibt es verschiedene rechtliche Voraussetzungen zu beachten. Zunächst stellt sich die Frage, wem die Forschungsdaten „gehören“ und wer über ihre Veröffentlichung bestimmen darf. Hierbei ist zwischen urheberrechtlich gestützten und nicht urheberrechtlich geschützten Forschungsdaten zu unterscheiden. Um in dieser Frage Klarheit zu erhalten, hat die Fraunhofer-Gesellschaft ein rechtliches Gutachten in Auftrag gegeben. Des Weiteren sind für den Betrieb eines Forschungsdaten-Repositoriums die notwendigen, rechtlichen Voraussetzungen zu schaffen: Da über Shibboleth eine Verknüpfung zu personenbezogenen Daten besteht und auch Autorennamen erfasst werden, ist es notwendig ein Verarbeitungsverzeichnis gemäß DSGVO Artikel 30 zu erstellen, sowie eine Genehmigung des Gesamtbetriebsrats einzuholen. Ferner werden gut sichtbare Datenschutzhinweise sowie ein rechtssicheres Impressum benötigt. Um die Forschungsdaten öffentlich bereitzustellen und sie gegebenenfalls später für die Langzeitarchivierung in ein anderes Format umwandeln zu dürfen, ist es notwendig, von den Datenautoren eine Deposit-Lizenz einzuholen. In »Fordatis« wird hierfür ein PDF eingebunden, das elektronisch signiert werden muss und anschließend an die »Abteilung Research Services & Open Science« geschickt werden muss. Anschließend wird die Deposit-Lizenz dort als Bitstream zu den Forschungsdaten mithochgeladen.

Im November 2018 wurde die Fordatis Sandbox einem ersten Nutzerkreis vorgestellt.

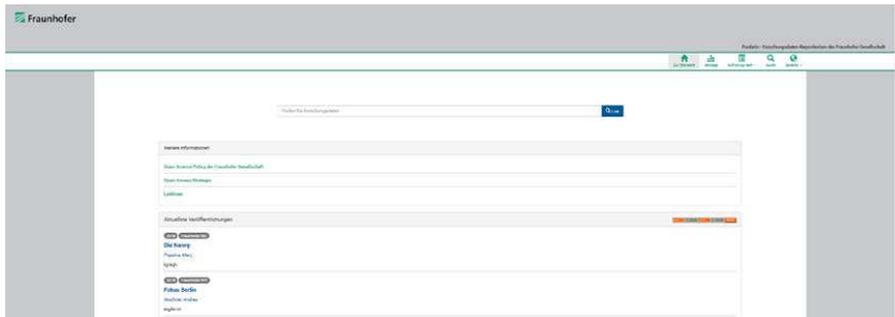


Abbildung 5: Layout von Fordatis

Von der technischen Entwicklung abgesehen, stellt die organisatorische Einbettung des Forschungsdaten-Repositorys eine weitere Herausforderung dar. An 72 Instituten müssen Erfassungswflows etabliert werden, ein Datenkurator muss benannt werden. Sowohl Datenkuratoren als auch Wissenschaftler müssen befähigt werden mit Fordatis zu arbeiten. Um an den Instituten Erfassungswflows zu etablieren, wurden die Fachinformationsmanager gezielt angeschrieben und informiert. Damit sollte gewährleistet werden, dass die Rolle des Datenkurators in bestehende Fachinformations-Strukturen an den Instituten integriert wird und Synergieeffekte genutzt werden. Die Wissenschaftler und Fachinformationsmanager werden durch verschiedene Angebote bei ihrer Arbeit mit Fordatis unterstützt: Anleitungen zur Benutzung von Fordatis, Screencasts, Informationen zu den angebotenen Lizenzen. Des Weiteren wurde zur Kommunikation von »Fordatis« für verschiedene Stakeholder (vorrangig innerhalb der Fraunhofer-Gesellschaft) ein Kommunikationskonzept erstellt.

Begleitend zum Go-Live wird die Fraunhofer-Gesellschaft eine Forschungsdaten-Policy veröffentlichen. Diese sieht die Veröffentlichung Forschungsdaten vor, die zur Überprüfbarkeit von Forschungsergebnissen notwendig sind. Um die Interessen von Kooperationspartnern aus der Industrie hinsichtlich Geheimhaltung zu berücksichtigen, sind Forschungsdaten, die Geheimhaltungsvereinbarungen unterliegen, von der Veröffentlichungspflicht ausgenommen.

Der Go-Live von »Fordatis« ist für Sommer 2019 geplant.

Diskussion

Mit dem Go-Live wird ein wichtiger Meilenstein des Projekts »FORDATIS« erreicht. Es stehen am Repository selbst noch weitere Aufgaben an: Ein

wichtiges Thema ist die Frage der Langzeitarchivierung der Forschungsdaten in »Fordatis«, für die ein Konzept zur Langzeitarchivierung erstellt und umgesetzt werden soll.

Mittelfristig soll die bibliografische Datenbank »Fraunhofer-Publica« auf DSpace CRIS migriert werden. Beide Systeme sollen so zusammengeführt werden, dass Normdaten, wie Projektdaten oder Autoredaten gemeinsam genutzt werden können. Nach außen soll dieser Verbund als »Fraunhofer-Open-Science-Cloud« sichtbar werden und geharvestet werden. Es wird in der Zukunft notwendig sein, »Fordatis« auf zukünftigen DSpace-Versionen zu aktualisieren und Anpassungen an die Anforderungen der Stakeholder vorzunehmen.

Das Projekt »FORDATIS« leistet einen wichtigen Beitrag zur Weiterentwicklung der Bibliothekare an den Institutsbibliotheken bzw. den Fachinformationsmanagern an den Instituten: Mit der Rolle „Datenkurator“ eröffnet sich ein in die Zukunft gerichtetes Betätigungsfeld, um die bestehenden Services am Institut mit dem neuen Aufgabenbereich des „Forschungsdatenmanagement“ zu erweitern. Langfristig können in diesem Bereich auch weitere Aufgaben, zum Beispiel die Beratung zu Anforderungen oder Datenmanagementplänen übernommen werden. Einige Institute nutzen diese Gelegenheit, um den „Datenkurator“ fest in der Institutsbibliothek zu verankern.

Das im Projekt entwickelte Beratungs- und Schulungsangebot wird sehr gut angenommen: Es wurden seit 2018 zahlreiche Anfragen zu Forschungsdaten und Forschungsdatenmanagement beantwortet. Häufige Themen waren dabei Datenmanagementpläne und die Anforderungen der Förderorganisationen. Mit den Arbeiten an »Fordatis« wurden auch zunehmend Fragen zum Repositorium gestellt. Das Schulungs-Angebot wird ebenfalls gut angenommen und regelmäßig von den Instituten nachgefragt.

Eine Herausforderung ist es, die Nutzung des Forschungsdatenrepositoriums über die kommenden Jahre im Institutsalltag zu verankern und Mehrwerte für die Wissenschaftler und die Fraunhofer-Gesellschaft zu erzeugen. Zunächst muss das Angebot bei Fraunhofer bekannt gemacht und ausgerollt werden. Da

es in direkter Konkurrenz zu anderen teilweise kostenlosen Repositorien wie Zenodo (CERN, 2019) und Figshare (Figshare, 2019) steht, und die Wissenschaftler im Hinblick auf ihr Veröffentlichungsmedium freie Wahl haben, muss es durch Mehrwerte überzeugen. Auch die Fraunhofer Forschungsdaten-Policy wird zur Nutzung des institutionellen Repositoriums beitragen, da sie die Relevanz des Themas unterstreicht und der Fraunhofer-Gesellschaft die Möglichkeit gibt, an zentraler Stelle den wichtigen Forschungsoutput der Forschungsdaten sichtbar zu machen.

Innerhalb der Fraunhofer-Gesellschaft stellen die Ergebnisse des Projekts »FORDATIS« den Anknüpfungspunkt zu einer weiteren Entwicklung dar: Im Rahmen des BMBF-Projekts HEFE – Heterogene Forschungsdaten im Stadtkontext wurde eine Data Governance konzipiert – ein System aus Regeln, wie mit Forschungsdaten innerhalb der Fraunhofer-Gesellschaft umgegangen werden soll. In diesem Rahmen wurde ein Domänenkonzept entwickelt, das in einem ersten Schritt für verschiedene Datenarten, wie Interviewdaten oder Befragungen Leitlinien und Checklisten erarbeitet. Langfristig möchte das Projekt den Anstoß zu einer digitalen Unterstützung der Data Governance über den gesamten Forschungsprozess hinweg geben. Hierbei soll »Fordatis« zum Veröffentlichenden der Forschungsdaten dienen.

Im Hinblick auf eine Fortführung und Fortentwicklung der Forschungsdateninfrastruktur sind zwei Aspekte von besonderer Bedeutung: Ein wichtiger Impuls können die Beantwortung der bislang offenen rechtlichen Fragestellungen darstellen, beispielsweise wem die Forschungsdaten „gehören“ bzw. wer für die Qualität und die Weiterverwendung Forschungsdaten „haftet“ darstellen. Zum anderen sind die Entwicklungen der European Open Science Cloud und der Nationalen Forschungsdateninfrastruktur (NFDI) zu beobachten. »Fordatis« soll so weiterentwickelt werden, dass es an beide Infrastrukturen anschlussfähig ist.

Um die Arbeiten des Projekts fortzuführen, wurde der Betrieb des Repositoriums und der Service-Angebote zunächst bis Ende 2020 verstetigt. Das Projekt FORDATIS stellt im Kontext der Open Science einen wichtigen Schritt für die Umsetzung von Open Data in der Fraunhofer-Gesellschaft dar.

Daneben werden in der Fraunhofer-Gesellschaft Open Access sowie in ersten Ansätzen Open Software gefördert. Besonders im Hinblick auf das Spannungsfeld zwischen Forschung und Wirtschaft, in dem sich die Fraunhofer-Gesellschaft befindet, müssen innerhalb der Fraunhofer-Gesellschaft jeweils passende Umsetzungsstrategien entwickelt werden. (Küsters & Klages, 2019).

Literatur

Beyan, O., Wuchner, A., Eisengräber-Pabst, D., Quix, C., Zschke, C., & Schumacher, O. (2018). Research Data in the Fraunhofer Digital Project : Creating a FAIR Research Data Infrastructure and Culture.

Büttner, S., Rumpel, S., & Hobohm, H.-C. (2011). Informationswissenschaftler im Forschungsdatenmanagement. In S. Büttner, H.-C. Hobohm, & L. Müller, Handbuch Forschungsdatenmanagement (S. 2010 - 2018). Bad Honnef: Bock + Herchen Verlag.

CERN. (2019). Zenodo. Abgerufen am 01. 04 2019 von <https://www.zenodo.org/>

DataCite. (2014). DataCite Standard 3.1. Von <https://schema.datacite.org/meta/kernel-3.1/> abgerufen

Deutsche Forschungsgemeinschaft. (2013). Sicherung Guter Wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“, . Von <http://doi.org/10.1002/9783527679188.oth1> abgerufen

Deutsche Forschungsgemeinschaft. (2015). Leitlinie zum Umgang mit Forschungsdaten. Online verfügbar unter http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf, zuletzt geprüft am 13.03.2017.

Deutsche Initiative für Netzwerkinformation e. V. (kein Datum). Webseite der DINI/Nestor AG Forschungsdaten. Abgerufen am 15. 04 2019 von <https://dini.de/ag/dininestor-ag-forschungsdaten/>

Dublin Core Metadata Initiative . (2014). Dublin Core Metadata Terms. Von <http://www.dublincore.org/specifications/dublin-core/dcmi-terms/> abgerufen

Dublin Core Metadata Initiative. (2013). Dublin Core Element Set. Abgerufen am 28. 03 2019 von <http://dublincore.org/documents/dces/>

European Commission. (2016). Guidelines on FAIR Data Management in Horizon 2020. Abgerufen am 31. 03 2019 von http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European Commission. (2016). Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 . Abgerufen am 31. 03 2019 von http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi

- European Commission. (2018). Implementation Roadmap for the European Science Cloud. Von https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf#view=fit&pagemode=none abgerufen
- Fair Metrics Group . (2018). Von FAIR Metrics Org: <http://fairmetrics.org/about.html> abgerufen
- Figshare. (2019). Figshare. Abgerufen am 01. 04 2019 von <https://figshare.com/>
- FORCE 11. (2016). The fair data principles. Abgerufen am 31. 03 2019 von <https://www.force11.org/group/fairgroup/fairprinciples>
- Fraunhofer-Gesellschaft. (2015). Fraunhofer Open-Access-Strategie. Abgerufen am 19. Februar 2019 von <https://www.openaccess.fraunhofer.de/de/open-access-strategie.html>
- Hapke, H. (2016). Data Librarian: Das moderne Berufsbild. b.i.t online, S. 159 - 164.
- Hauck, R., Kaps, R., Krojanski, H. G., Meyer, A., Neumann, J., & Soßna, V. (2016). Der Umgang mit Forschungsdaten an der Leibniz Universität Hannover: Auswertung einer Umfrage und ergänzender Interviews 2015/16. Abgerufen am 31. 03 2019 von <https://doi.org/10.15488/265>
- Heidelberg, R. d. (Letzter Zugriff 26.03.2019). Webseite der E-Science-Tage Heidelberg. Von <https://www.e-science-tage.de/> abgerufen
- Heidelberg, U. (2017). E-Science-Tage Heidelberg. Abgerufen am 27. 03 2019 von <https://www.e-science-tage.de>
- Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum GFZ. (2013). Forschungsdaten.org. Abgerufen am 27. 03 2019 von <http://www.forschungsdaten.org>
- JERRI - Joining Efforts for Responsible Research & Innovation. (2017). Abgerufen am 19. Februar 2019 von <https://www.jerri-project.eu/jerri/index.php>
- Karlsruher Institut für Technologie. (kein Datum). Research Data Management (RDM) am KIT. Abgerufen am 15. 04 2019 von <http://www.rdm.kit.edu/>
- Karlsruher Institut. (kein Datum). Webseite des Karlsruher Institut für Technologie. Abgerufen am 15. 04 2019 von <http://www.kit.edu/forschen/13557.php>
- KIT Karlsruhe. (2018). Projekt bwFDM-Info II. Abgerufen am 27. 03 2019 von <https://bwfdm.scc.kit.edu/bwFDM-Info.php>
- Küstern, U., & Klages, T. (2019). Fostering Open Science@Fraunhofer. *Procedia computer science*(146), S. 39 -52. doi:10.1016/j.procs.2019.01.078
- Leibniz Universität Hannover. (2011). Open-Access-Resolution. Abgerufen am 31. 03 2019 von <https://www.uni-hannover.de/fileadmin/luh/content/webredaktion/universitaet/ziele/openaccess-resolution.pdf>
- Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. (2013). Webseite der Max Planck Computing & Data Facility. Abgerufen am 27. 03 2019 von <https://www.mpcdf.mpg.de>
- Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. (2019). Webseite der Max Planck Digital Library. Abgerufen am 27. 03 2019 von <https://www.mpd.l.mpg.de/>

Meyer, A., Janna, N., & Soßna, V. (2017). Service durch Kompetenzbündelung - Das institutionelle Konzept zum Forschungsdatenmanagement der Leibniz Universität Hannover. Abgerufen am 2019. 03 31 von <https://doi.org/10.11588/heibooks.285.377>

Projekt Forschungsdatenmanagement und -Archivierung der Universität Marburg. (2015). Forschungsdatenmanagement an der Philipps-Universität Marburg. Die Ergebnisse der Umfrage zum Forschungsdatenmanagement im November 2014. Von <https://10.17192/es2015.0019> abgerufen

RDA Deutschland e.V. (kein Datum). RDA Deutschland. Abgerufen am 31. 03 2019 von <https://www.rda-deutschland.de/>

Research Data Alliance. (kein Datum). Webseite der Research Data Alliance. Abgerufen am 15. 04 2019 von <https://www.rd-alliance.org/>

Rfll - Rat für Informationsinfrastrukturen. (kein Datum). Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. Abgerufen am 31. 03 2019 von <urn:nbn:de:101:1-201606229098>

Simucovic, E., Kindling, M., & Schirmbacher, P. (2013). Umfrage zum Umgang mit digitalen Forschungsdaten an der Humboldt-Universität zu Berlin. Umfragebericht, Version 1.0. Von <urn:nbn:de:kobv:11-100213001> abgerufen

Swan, A., & Brown, S. (2008). Skills, Role & Career Structure of Data Scientists. Report to the JISC.

Universität Heidelberg. (2018). EScience-Tage Heidelberg. Abgerufen am 27. 03 2019 von <https://www.e-science-tage.de>

Universität Konstanz. (2019). Forschungsdaten.info. Abgerufen am 27. 03 2019 von <https://www.forschungsdaten.info>

Wilkinson, M. D. (2018). A design framework and exemplar metrics for FAIR. Scientific Data.