
Open Science bei Fraunhofer – Serviceentwicklung und Realisierung einer Forschungsdateninfrastruktur für Open Data

Tina Klages¹, Andrea Wuchner²

1,2 Fraunhofer-Informationszentrum Raum und Bau

Zusammenfassung. Die Digitalisierung hat als Megatrend inzwischen die Wissenschaft erreicht und führt in diesem Zusammenhang sowohl zu vielen Potenzialen, die sich durch die vorhandenen und zukünftigen technischen Möglichkeiten ergeben, aber auch zu Herausforderungen für Wissenschaftsorganisationen weltweit. In diesem Kontext ist Open Science ein wichtiges Schlagwort, da es inzwischen möglich ist, den gesamten Wissenschaftsprozess von der ersten Idee bis hin zur Veröffentlichung der Forschungsergebnisse offen zu legen. Ziel ist es, neben Transparenz und Nachvollziehbarkeit wissenschaftlicher Ergebnisse und Daten auch deren Nachnutzbarkeit zu gewährleisten, wo dies möglich ist. Kern und Ausgangspunkt von Open Science sind der offene Zugang zu Publikationen (Open Access) und vor allem zu den Forschungsdaten (Open Data). Die Fraunhofer-Gesellschaft hat sich des Themas Open angenommen und steht dabei als Wissenschaftsorganisation der angewandten Forschung vor der Herausforderung, dass ein Großteil der Forschungsprojekte gemeinsam mit Industrieunternehmen durchgeführt wird. Daher sind vor allem Geheimhaltungsinteressen der Unternehmen bei der Ausgestaltung von Open Science für die Fraunhofer-Gesellschaft zu berücksichtigen. Auf das Thema Open Access wirkt sich diese Situation nicht aus, da Publikationen, ob offen zugänglich oder nicht, ohnehin publiziert sind. In Bezug auf das Thema Forschungsdaten ist die Situation eine andere. Hier fallen die Geheimhaltungsinteressen der Unternehmen stark ins Gewicht, was vor allem den offenen Zugang zu Forschungsdaten angeht, der inzwischen z.B. im Zuge des Förderprogramms Horizon2020 von der europäischen Kommission gefordert wird. Daher verfolgt die Fraunhofer-Gesellschaft das Prinzip, Forschungsdaten so offen wie möglich, aber auch so geschlossen wie nötig zu behandeln. Die Umsetzung einer fraunhofer-spezifischen Lösung stößt auf verschiedene Herausforderungen: Kultur, Information, Qualifikation, Prozessentwicklung, Serviceentwicklung, Nachweisbarkeit und Interoperabilität, Disziplinspezifische Kulturen, rechtliche Rahmenbedingungen und Qualität. Die technische Infrastruktur für den Nachweis und die Veröffentlichung von Forschungsdaten soll mit der bestehenden Publikationsinfrastruktur der Fraunhofer-Gesellschaft verknüpft und in bestehende Publikations- und Beratungsprozesse integriert werden. Darüber hinaus soll die Infrastruktur die Erfüllung von Anforderungen der Forschungsförderer zu Open Data automatisiert erbringen.

Durch die bedarfsgerechte Entwicklung und Etablierung einer Forschungsdateninfrastruktur, die an die heterogenen Ansprüche der 70 Institute der Fraunhofer-Gesellschaft angepasst ist, wird ein wichtiger Beitrag zu der Umsetzung von Open Science erbracht. Die Infrastruktur wird durch entsprechende Services und Unterstützungsangebote für die Fraunhofer-Wissenschaftler begleitet. Durch eine Fraunhofer-Policy zum Umgang mit Forschungsdaten wird das Thema auf strategischer Ebene ebenfalls in die Fraunhofer-Gesellschaft getragen. Somit wird ein ganzheitlicher Ansatz in Bezug auf Forschungsdaten bei Fraunhofer verfolgt.

Schlagwörter. Fraunhofer Gesellschaft, Forschungsdatenmanagement, Forschungsdateninfrastruktur

Ausgangssituation

Die Digitalisierung hat als Megatrend inzwischen neben der Wirtschaft auch die Wissenschaft erreicht und führt in diesem Zusammenhang zu großen Veränderungen im wissenschaftlichen Arbeiten. Durch die Möglichkeiten des Internets und anderen Innovationen im Bereich der Information und Kommunikation arbeiten Wissenschaftler weltweit zusammen an Forschungsfragen und kooperieren dabei über Plattformen, teilen Forschungsergebnisse und beschleunigen dadurch den wissenschaftlichen Fortschritt. In diesem Kontext ist Open Science ein wichtiges Konzept, das es ermöglicht, den gesamten Wissenschaftsprozess von der ersten Idee bis hin zur Veröffentlichung der Forschungsergebnisse offen zu legen. Ziel ist es, neben Transparenz und Nachvollziehbarkeit wissenschaftlicher Ergebnisse bzw. Daten auch deren Nutzbarkeit zu gewährleisten, sofern dies z.B. aus rechtlichen Gründen oder Geheimhaltungsründen möglich ist. Gleichzeitig sollen dadurch Innovationszyklen verkürzt werden, um den Forschungs- und Wirtschaftsstandort Europa zu stärken¹. Das Konzept der Open Science umfasst mehrere Teilbereiche, wie z. B. Open Source, Citizen Science². Kern von Open Science ist jedoch der offene Zugang zu Publikationen (Open Access) und zu Forschungsdaten (Open Access to Research Data). Dieser Paradigmenwechsel von einem analogen und geschlossenen Forschungsprozess zu dessen Digitalisierung und Öffnung ändert das wissenschaftliche Arbeiten von Grund auf. Dadurch ergeben sich viele Potenziale, aber auch Herausforderungen in Bezug auf die Ausgestaltung von Open Science, denen sich Wissenschaftsorganisationen weltweit stellen müssen.

Auch die Fraunhofer-Gesellschaft hat sich des Themas Open Science angenommen. Die Fraunhofer-Gesellschaft ist die führende Organisation für angewandte Forschung in Europa. In Kooperation mit anderen Wissenschaftsorganisationen und Wirtschaftsunternehmen erforscht sie Themen, die sich eng an den Bedarfen der Gesellschaft orientieren: Gesundheit, Sicherheit, Kommunikation, Mobilität, Energie und Umwelt. Mit insgesamt 69 Instituten und 24 500 Mitarbeitern werden dabei vom jährlichen Forschungsvolumen von 2,1 Milliarden Euro ca. 1,9 Milliarden Euro über Vertragsforschung erbracht. Davon werden insgesamt ca. 70 Prozent durch Kooperationsprojekte mit Partnern aus der Wirtschaft und durch öffentlich geförderte Forschungsprojekte erwirtschaftet³. Ziel ist neben der Forschung für die Gesellschaft die Beförderung der Innovationsfähigkeit des Wirtschaftsstandortes Deutschland bzw. Europas. Durch ihre besondere Position im Innovationssystem als Bindeglied zwischen Grundlagenforschung und Entwicklung sieht sich die Fraunhofer-Gesellschaft bei der Ausgestaltung von Open Science dabei als Wissenschaftsorganisation der angewandten Forschung vor die Herausforderung gestellt, dass ein Großteil der Forschungsprojekte gemeinsam mit Industrieunternehmen durchgeführt wird. Ziel dieser Kooperationsprojekte ist die gemeinsame Wertschöpfung durch die Erarbeitung von individuellen Lösungen für Fragestellungen aus der Praxis. Ergebnis dieser Aktivitäten sind Innovationen, die auf technologischem Fortschritt basieren und die Marktposition der Kooperationspartner stärken. Aus diesen Gründen ist es essentiell, die Geheimhaltungs- bzw. Verwertungsinteressen der Unternehmen zu berücksichtigen. Die Öffnung von Forschungsprozessen sowie die offene Bereitstellung von Forschungsergebnissen in Form von Publikationen und Forschungsdaten sind in vielen Fällen

1 „EU Digital Market Strategy“, accessed March 29th 2017, https://ec.europa.eu/commission/priorities/digital-single-market_en

2 Homepage der AG Open Science, accessed March 29th, 2017, <http://www.ag-openscience.de/>

3 „Fraunhofer Zahlen und Fakten“, accessed March 29th 2017, <https://www.fraunhofer.de/de/ueber-fraunhofer/profil-selbstverstaendnis/zahlen-und-fakten.html>

bei Fraunhofer somit nicht ohne weiteres möglich und müssen fallweise entschieden werden. Die Voraussetzungen für Open Science sind für Organisationen der angewandten Forschung andere als für Einrichtungen, die öffentlich finanzierte Grundlagenforschung betreiben, da deren Ergebnisse meist ohne Einschränkung für die Öffentlichkeit bestimmt sind.

Betrachtet man im Kontext von Open Science den offenen Zugang zu Publikationen und Forschungsdaten, so sind Publikationen in diesem Kontext nicht problematisch, da sich diese grundsätzlich an die Öffentlichkeit richten. Hier wurde die Entscheidung der Veröffentlichung bereits im Vorfeld getroffen. In Bezug auf Daten ist die Situation eine andere: Hier fallen die Geheimhaltungsinteressen der Unternehmen stark ins Gewicht, was vor allem den offenen Zugang zu Forschungsdaten betrifft, der inzwischen z. B. im Zuge des Förderprogramms Horizon 2020⁴ von der Europäischen Kommission gefordert wird. Daher verfolgt die Fraunhofer-Gesellschaft das Prinzip, Forschungsdaten so offen wie möglich, aber auch so geschlossen wie nötig zu behandeln.⁵

Herausforderungen und institutioneller Lösungsansatz

Der offene Zugang kann nicht ohne weiteres durch jeden Wissenschaftler einzeln sichergestellt werden. Auch das Management der Forschungsdaten kann, je nach Umfang und Anforderungen, nicht jeder Wissenschaftler im Rahmen seiner Forschungstätigkeit eigenständig leisten. Für Forschungsdaten besteht weiterhin die Anforderung der Nutzbarkeit, die durch ein gewisses Maß an Standardisierung über Disziplinen hinweg ermöglicht werden muss. Hier ergibt sich somit der Bedarf einer Bündelung von Kompetenzen im Umgang mit Forschungsdaten in der Fraunhofer-Gesellschaft.

Darüber hinaus ist der Aufbau einer institutionellen Infrastruktur für den Nachweis von Forschungsdaten notwendig, die sowohl die Veröffentlichung als auch den Nachweis von Datensätzen erlaubt, soweit dies im Kontext von Open Data gewünscht ist. Hierbei stellt die direkte Verknüpfung von Datensätzen mit den zugehörigen Publikationen eine wesentliche Anforderung dar. Die Infrastruktur muss dazu mit der bestehenden Publikationsinfrastruktur der Fraunhofer-Gesellschaft verknüpft und in bestehende Publikations- und Beratungsprozesse integriert werden. Darüber hinaus soll die Infrastruktur die Erfüllung von Anforderungen der Forschungsförderer zu Open Data über Schnittstellen automatisiert erbringen, wie beispielsweise zum OpenAIRE Portal der Europäischen Kommission⁶. Ziel ist es, den Umgang mit und den Nachweis bzw. die Publikation von Forschungsdaten durch die Wissenschaftler bestmöglich zu unterstützen. Die Ausgestaltung dieser Vorhaben wird durch das Competence Center Research Services & Open Science geleistet. Dabei handelt es sich um einen zentralen Dienstleister der Fraunhofer-Gesellschaft, der sich u. a. mit Themen beschäftigt, die sich aus der Digitalisierung der Wissenschaft ergeben. Die Entwicklung einer Fraunhofer-spezifischen Lösung birgt eine Reihe von Herausforderungen, die einerseits im Bereich des Umgangs mit Forschungsdaten im Allgemeinen begründet sind, sich andererseits jedoch auch durch die spezifische Struktur der Fraunhofer-Gesellschaft ergeben.

4 Open Data Pilot im Forschungsprogramm Horizon2020, accessed March 29th 2017, <https://www.openaire.eu/opendatapilot>

5 EARTO Paper on OpenX, accessed March 29th 2017, http://www.earto.eu/fileadmin/content/Website_2/EARTO_Paper_on_Open_X_-_13_November_2015_-_Final.pdf

6 OpenAIRE Portal der Europäischen Kommission, accessed March 29th, 2017 <https://www.openaire.eu>

Kultur:

Eine Infrastruktur ist zwar eine notwendige, jedoch keine hinreichende Bedingung für den verantwortungsvollen Umgang mit Forschungsdaten. Dazu ist vor allem die Entwicklung einer Kultur des nachhaltigen Umgangs mit Forschungsdaten und deren Management erforderlich. Durch die Digitalisierung und die Entwicklung hin zu Open Science bzw. Open Data hat sich die Bedeutung von Forschungsdaten, offen oder geschlossen, stark gewandelt. Sie stellen inzwischen eine wertvolle Ressource für Wissenschaft und Wirtschaft dar, die einen sorgfältigen Umgang erfordert, um alle damit einhergehenden Potenziale voll ausschöpfen zu können. Hier ist also ein Umdenken der Wissenschaft notwendig, das durch die Schaffung von Awareness unterstützt werden sollte.

Information:

Umfangreiche Informationsaktivitäten sind essentiell, um die veränderte Bedeutung von Forschungsdaten im Kontext von Open Science deutlich machen. Hierbei ist es wichtig, sowohl die damit einhergehenden Herausforderungen als auch die Potenziale von Forschungsdatenmanagement und Open Data zu kommunizieren. Darüber hinaus müssen die Wissenschaftler über den Umgang mit Forschungsdaten informiert werden.

Qualifikation:

Um einen qualifizierten Umgang mit Forschungsdaten in den unterschiedlichen Disziplinen in Einklang mit den Anforderungen der wissenschaftlichen Integrität zu erreichen, bedarf es Maßnahmen für die Qualifizierung der Fraunhofer-Wissenschaftler⁷. Neue Kompetenzprofile sind notwendig, um die Wissenschaftler für die Herausforderungen im Umgang mit Daten, vor allem in Bezug auf große Datenmengen, vorzubereiten bzw. Spezialisten in diesem Bereich auszubilden. Hier existieren unterschiedliche Konzepte und Berufsbilder, z. B. Data Curator, Data Manager, Data Scientist, Data Analyst und Data Librarian.

Prozessentwicklung:

Die Heterogenität der Disziplinen spiegelt sich innerhalb der Fraunhofer-Institute auch in der Heterogenität von Prozessen wieder, dies wird u. a. in Bezug auf Publikationen deutlich. Für einen nachhaltigen Umgang mit Forschungsdaten ist es daher von großer Bedeutung, entsprechende Prozesse innerhalb der Institute zu etablieren. Teil dieser Prozesse müssen qualifizierte Personen sein, die das Thema wissenschaftlich begleiten und die Wissenschaftler beraten.

Serviceentwicklung:

Die Entwicklung und Bereitstellung von Unterstützungsangeboten für die Fraunhofer-Institute und deren Wissenschaftler zu allen Fragestellungen rund um die Themen Forschungsdaten und deren Management bzw. Open Data sowie den damit einhergehenden politischen Anforderungen ist essentiell. Hierzu gehört ein Forschungsdatensupport, der z. B. in Hinblick auf Datenmanagementpläne (wie in Horizon 2020 seit 2017 verpflichtend gefordert) und Forschungsdatenmanagement im Zusammenhang mit Projekten berät.

7 DFG: Leitlinien zur guten wissenschaftlichen Praxis, accessed March 29th 2017, http://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/gwp/index.html

Nachweisbarkeit und Interoperabilität:

Um Forschungsdaten nachzuweisen und diese auch auffindbar zu machen, ist es unerlässlich ein Metadaten-Profil festzulegen, mithilfe dessen es möglich ist, die Datensätze nachhaltig zu beschreiben und in Nachweissystemen zu erfassen. Hier ist es sinnvoll, bereits bestehende und verbreitete Standards nachzunutzen, und diese den Anforderungen der Fraunhofer-Gesellschaft entsprechend anzupassen. So ist die Interoperabilität zu anderen Nachweis-Systemen unter Berücksichtigung der Fraunhofer-spezifischen Besonderheiten gewährleistet.

Disziplinspezifische Kulturen:

Die einzelnen wissenschaftlichen Fachdisziplinen zeichnen sich durch einen unterschiedlichen Umgang mit Forschungsdaten aus. So existiert eine Vielzahl verschiedener Datenformate und vorherrschender Standards, welche disziplinspezifisch zum Einsatz kommen. Auch die anfallenden Datenmengen sind zwischen den jeweiligen Fachdisziplinen unterschiedlich. Diese Heterogenität, die sich bei Fraunhofer widerspiegelt, muss beim Aufbau der Infrastruktur und bei der Entwicklung von Serviceangeboten berücksichtigt werden.

Rechtliche Rahmenbedingungen der Nachnutzung von veröffentlichten Daten:

Um Forschungsdaten nachnutzen zu können, bedarf es das Urheberrecht betreffend klarer Regelungen, die im Rahmen der derzeitigen Rechtsprechung noch nicht geklärt sind. Dennoch hat der Erzeuger der Daten (z. B. über die Vergabe von CC Lizenzen) die Möglichkeit, eindeutige Nachnutzungsrechte zu gewähren.

Qualität:

Die Fraunhofer-Publikationsinfrastruktur, bestehend aus der Publikationsnachweisdatenbank Fraunhofer-Publica und dem Repository Fraunhofer-ePrints, unterliegt in Bezug auf den Nachweis von Publikationen und Volltexten einer Qualitätskontrolle, um sicherzustellen, dass die Möglichkeit der Auffindbarkeit der Inhalte besteht. Dies ist auch für den Nachweis von Forschungsdaten unerlässlich, um einen gesicherten Standard festzulegen, der für alle Datennachweise und Datensätze gleichermaßen gilt.

Projekt FORDATIS

Um die bei Fraunhofer generierten Forschungsdaten standardisiert nachzuweisen und zu veröffentlichen, wurde das Fraunhofer-interne Projekt „FORDATIS – Forschungsdateninfrastruktur“ beauftragt.

Das Data Curation Continuum Model⁸, das an der Monash University in Australien entwickelt wurde, eignet sich, um zu verdeutlichen, auf welchen Ausschnitt der generierten Forschungsdaten das Projekt abzielt.

8 Treloar, A. & Harboe-Ree, C., 2008. Data management and the curation continuum. How the Monash experience is informing repository relationship., accessed March 29th 2017, http://www.valaconf.org.au/vala2008/papers2008/111_Treloar_Final.pdf

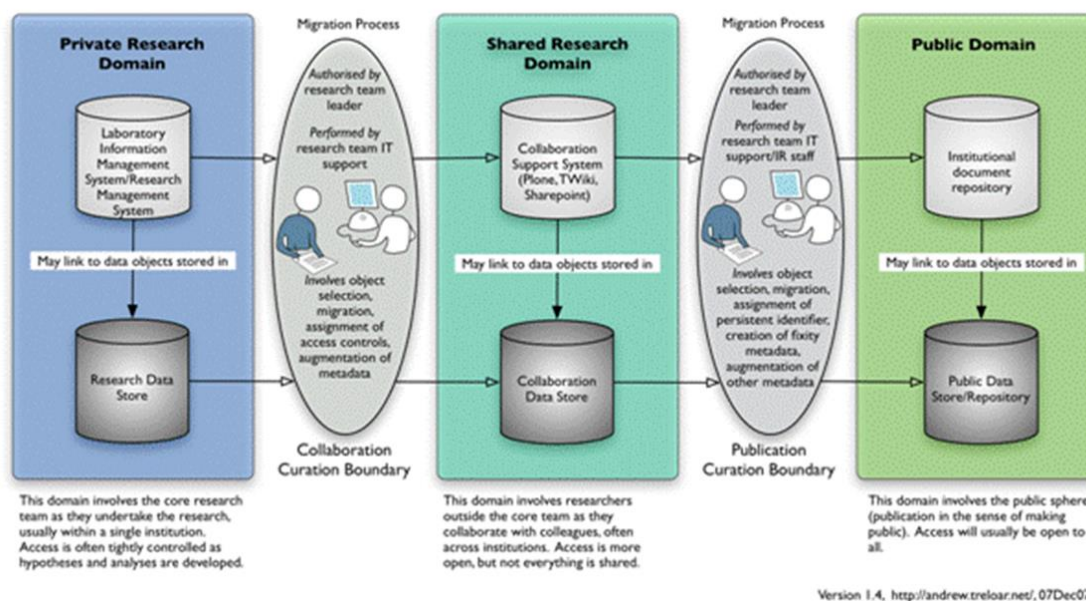


Abbildung 1. Data Curation Continuum (Treloar, A. & Harboe-Ree, 2008)

Das Modell beschreibt die verschiedenen Domänen, in denen Forschungsdaten sich im Verlauf ihres Lebenszyklus bewegen. Erzeugt und erstmalig ausgewertet, werden sie in der Private Research Domain. Dort befinden sich die Daten in sogenannten Research Management-Systemen. Sie werden in die Shared Research Domain überführt, wenn sie anderen Forschern oder Vorgesetzten zugänglich gemacht werden. In dieser Phase sind also Systeme notwendig, die eine Kollaboration von Wissenschaftlern anhand der Forschungsdaten ermöglichen. Mit dem Abschluss der Forschungsarbeiten erfolgt die Veröffentlichung ausgewählter Daten in die sogenannte Public Domain. Darunter wird in diesem Fall verstanden, dass die Daten einer breiten Öffentlichkeit zugänglich gemacht werden. Die Forschungsergebnisse, zu denen auch die Daten zählen, sind erstmalig ausgewertet, können veröffentlicht und für die Nachnutzung zur Verfügung gestellt werden. Dazu werden die Forschungsergebnisse in Repositories migriert. Zusätzlich muss eine Verlinkung mit Persistent Identifier und mit Metadaten versehenen Datenobjekten erfolgen, die sich in einem öffentlichen Forschungsdaten-Repository befinden⁹.

An dieser Stelle setzt das Projekt „FORDATIS“ an: Ziel ist es, eine Forschungsdateninfrastruktur für veröffentlichte und zu veröffentlichende Forschungsdaten aufzubauen. Die technische Infrastruktur soll von verschiedenen Services, wie z. B. Support und Schulungen begleitet werden¹⁰, was langfristig zu einer Erhöhung der Awareness im Umgang mit Forschungsdaten führen soll. Betrachtet man den internen Datenfluss bei Fraunhofer, ergeben sich Parallelen zum Data Curation Continuum:

Die Daten werden in Forschungsprojekten an den einzelnen Instituten erzeugt und erstmalig ausgewertet. Hierbei handelt es sich um verschiedene Ausprägungen von Forschungsdaten:

- Forschungsdaten, die unveränderlich sind und in Beziehung zu einer Publikation stehen,
- Forschungsdaten, die unveränderlich für sich stehen, oder aber
- Forschungsdaten, die sich in dieser Phase noch verändern.

9 Büttner, Stephan, Hans-Christoph Hobohm, and Lars Müller, ed. 2011. Handbuch Forschungsdatenmanagement. Bad Honnef: Bock + Herchen, S. 29-32

10 Eine Erläuterung der Services findet sich auf S. 8f.

Zur Fraunhofer-internen Nachnutzung können unveränderliche Forschungsdaten, sowohl eigenständig als auch im Zusammenhang mit einer Publikation bereitgestellt werden. Das gibt den verschiedenen Fraunhofer-Instituten die Möglichkeit der frühen und internen Nachnutzung. Allerdings haben die verschiedenen Fraunhofer-Institute Kompetenz-Portfolios, die sich teilweise überschneiden, so dass eine interne Veröffentlichung der Daten unter Umständen mit verschiedenen Vorbehalten verbunden ist.

Nach der Erstauswertung der Daten und Abschluss des Forschungsprojektes können diese entweder in Zusammenhang mit Publikationen oder als eigenständige Publikation veröffentlicht werden, was die Sichtbarkeit der Ergebnisse erhöht und es gleichzeitig ermöglicht, der Nachweispflicht bei Förderanforderungen nachzukommen. Um diese Veröffentlichung zu ermöglichen, soll das Forschungsdatenrepositorium „FORDATIS“ an die Publikationsnachweisdatenbank Fraunhofer-Publica angebunden werden. Dadurch wird die direkte Verknüpfung von Forschungsdaten mit den dazugehörigen Publikationen ermöglicht.

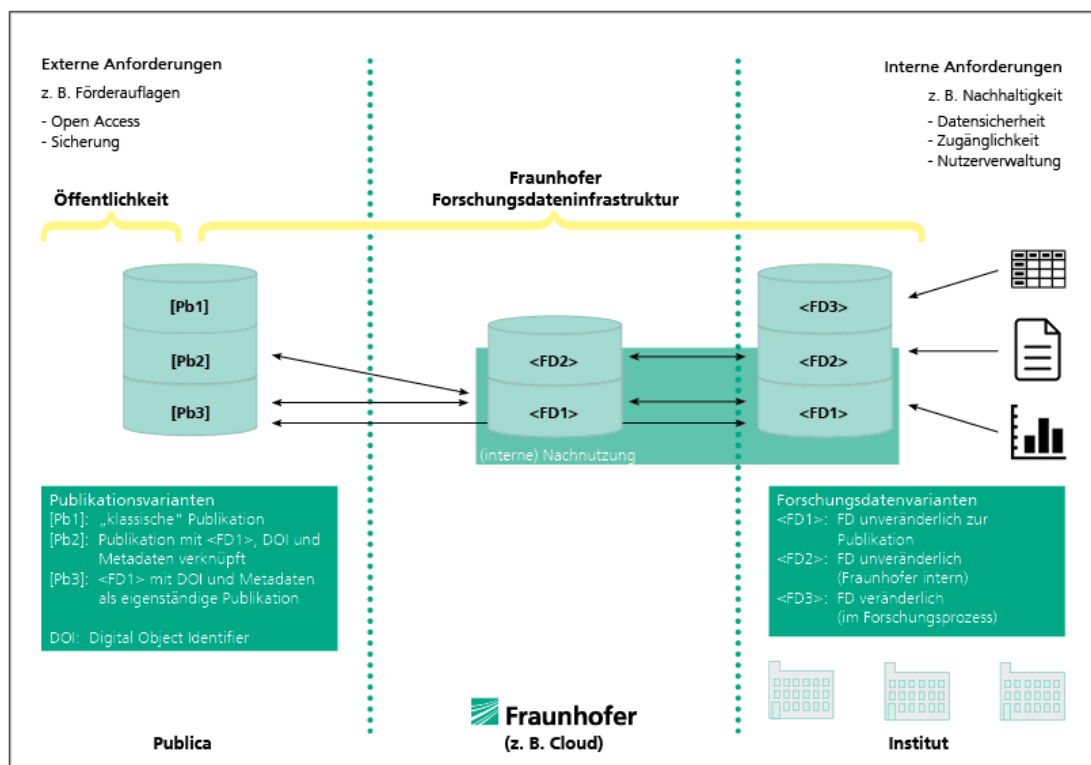


Abbildung 2. Forschungsdaten bei Fraunhofer (Spiecker, Claus 2016)

Umsetzung

Das Projekt „FORDATIS“ wird in dem Zeitraum von Juni 2016 bis Juni 2018 am Fraunhofer-Informationszentrum Raum und Bau (Fraunhofer IRB) bearbeitet.

Dazu beinhaltet das Projekt acht Arbeitspakete:

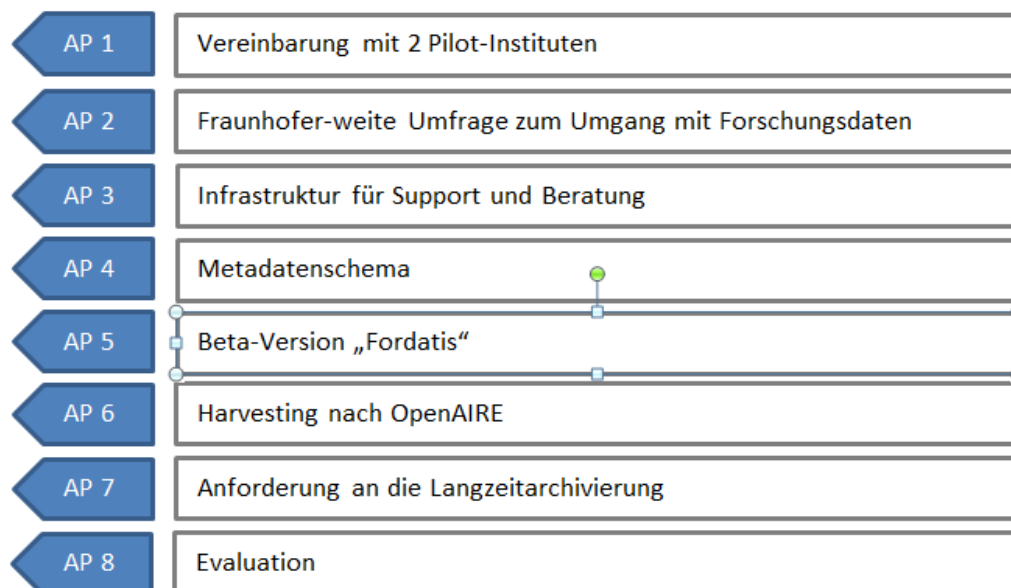


Abbildung 3. Arbeitspakete in FORDATIS

AP1: Vereinbarung mit 2 Pilot-Instituten: Zwei Pilot-Institute werden bei der Umsetzung der Aufgaben für Forschungsdaten begleitet, beraten und unterstützt. Die gewonnenen Erkenntnisse und Umsetzungen dienen als Basis für Fraunhofer-weite Lösungen in der zweiten Phase.

AP2: Fraunhofer-weite Umfrage: Die Umfrage soll nähere Erkenntnisse der Institute zum Umgang mit Forschungsdaten bringen und so weitere Bedarfe konkretisieren.

AP3: Infrastruktur für Support und Beratung: Elemente des Basisangebots sind Informationsbereitstellung, Schulung, Beantwortung von Anfragen, Beratung speziell für EU-Antragsteller. Hierzu ist die Zusammenarbeit mit internen und externen Multiplikatoren und Netzwerken sowie anderen Wissenschaftsorganisationen notwendig.

AP4: Application Profile: In diesem Arbeitspaket geht es um die Erarbeitung eines Metadatenprofils als Grundlage für eine standardisierte Erfassung von Forschungsdaten.

AP5: Beta-Version „Fordatis“: Programmierung, Test und Realisierung eines Prototyps des erweiterten Forschungsdaten-Repositorys. Daneben wird ein Meldeworkflow für eine Forschungsdaten-Registrierung entworfen. Im Rahmen dieses Meldeworkflows findet auch die Qualitätskontrolle statt, die für standardisierte Metadaten und Daten sorgt.

AP6: Harvesting nach OpenAIRE: Die Compliance-Anforderungen der OpenAIRE-Plattform zum Forschungsdatennachweis bei EU-Projekten werden überprüft und realisiert.

AP7: Anforderung an die Langzeitarchivierung: In diesem Arbeitspaket sollen die Anforderungen an die Langzeitarchivierung definiert werden.

AP8: Evaluation: Das Projekt soll evaluiert und weiterer Handlungsbedarf und Entwicklungsdesiderate sollen abgestimmt und zur möglichen Projektweiterführung in Phase 2 vorbereitet werden.

Verschiedene Arbeitspakete werden dabei auch im Rahmen des EU-Projekts „JERRI“, Joining Efforts for Responsible Research and Innovation, mitbearbeitet. Das Projekt hat das Ziel fünf Dimensionen von Responsible Research and Innovation an der Fraunhofer-Gesellschaft und der Niederländischen Organisation für Angewandte Naturwissenschaftliche Forschung (TNO) zu betrachten, weiterzuentwickeln und zu implementieren. Das Fraunhofer IRB ist dabei mit der

Dimension „Open Access“ vertreten. Open Access wird dabei als offener Zugang zu Publikationen und Forschungsdaten verstanden.

Kern von FORDATIS ist der technische Prototyp der Forschungsdateninfrastruktur. Dieser soll folgende Funktionalitäten zur Verfügung stellen:

- DOI-Vergabe, um Forschungsdaten eindeutig referenzieren zu können. Die Registrierung der Forschungsdaten wird über die Technische Informationsbibliothek (TIB) Hannover automatisiert erfolgen.
- Forschungsdaten sollen analog zu Publikationen nachgewiesen werden und recherchierbar sein. Der Nachweis von Forschungsdaten erfolgt entsprechend dem Metadaten-Standard DataCite4.0. Dieser Standard wird von der EU zur Erfüllung der Veröffentlichungspflicht von Forschungsdaten im Rahmen des Förderprogramms Horizon2020 empfohlen. Er ist auch die Grundlage für die Lieferung der Metadaten für die DOI-Vergabe.
- Die Metadaten der Forschungsdaten sollen an das OpenAIRE-Repository über die OAI-PMH-Schnittstelle geharvestet werden.
- Die Forschungsdaten sollen analog zu den Volltexten abgelegt und mit den Metadaten verknüpft werden.
- Die Forschungsdaten sollen mit den bibliographischen Angaben der Publikationsnachweisdatenbank Fraunhofer-Publica verknüpft werden. Dies soll über einen Identifier innerhalb der Publikationsnachweise geschehen.

Darüber hinaus soll der technische Prototyp von verschiedenen Service- und Beratungsangeboten flankiert werden. Diese werden bei Fraunhofer traditionell sehr bedarfsorientiert entwickelt. Denkbar sind folgende Angebote:

- Etablierung von Meldeworkflows, Qualitätskontrolle und individuellen Forschungsdaten-Prozessen an den Instituten, welche darauf abzielen, den Umgang mit Forschungsdaten zu begleiten und damit die Veröffentlichung und Langzeitarchivierung zu vereinfachen.
- Beratung zu Forschungsdatenmanagement bei Projektantragsstellung: Die Förderorganisationen stellen verschiedene Anforderungen an den Umgang mit Forschungsdaten. Um den einzelnen Wissenschaftler zu entlasten sollen Beratung und Support während der Projektantragsphase etabliert werden.
- Beratung zum Umgang mit Forschungsdaten im Rahmen von Kooperationsprojekten mit Industriepartnern: Bei Projekten, die Fraunhofer-Wissenschaftler mit Industriepartnern oder im Auftrag aus der Industrie durchführen, gibt es meist ein Geheimhaltungsinteresse, da wissenschaftliche Ergebnisse kommerziell verwertet werden sollen. Dieses Geheimhaltungsinteresse erstreckt sich auch auf Forschungsdaten, die in diesem Fall nicht veröffentlicht werden können. Wissenschaftler sollen im Rahmen von Best-Practices sensibilisiert und beraten werden, wie in diesem Fall mit Forschungsdaten zu Verfahren ist, um einerseits den Anforderungen der Industriepartner nachzukommen, aber auch gemeinsam mit ihnen mögliche Potenziale offener Daten auszuloten.
- Bereitstellung von Templates für Datenmanagementpläne auf Deutsch und Englisch für Projekte des Förderprogramms „Horizon 2020“. Im Rahmen dieses Förderprogramms müssen Datenmanagementpläne spätestens 6 Monate nach Projektstart eingereicht und während der Projektlaufzeit fortgeschrieben werden. Dazu werden Templates bereitgestellt.
- Schulungen der Fraunhofer-Wissenschaftler im Bereich „Forschungsdaten“: Analog zu bereits etablierten Schulungen im Bereich „Publikationsmanagement“ soll ein Schulungsan-

gebot zum Thema „Forschungsdaten“ etabliert werden. Dieses soll sich an die Fraunhofer-Wissenschaftler wenden.

Fazit und Ausblick

Die Fraunhofer-Gesellschaft kann mit dem Projekt „FORDATIS“ einem Großteil der zuvor skizzierten Herausforderungen im Zusammenhang mit Forschungsdatenmanagement bzw. Open Data begegnen:

Information: Durch Supportangebote ist es möglich, den Wissenschaftlern die Herausforderungen, aber auch die Potenziale der Veröffentlichung von Forschungsdaten zu kommunizieren und sie in der Praxis zu beraten.

Prozessentwicklung: Im Rahmen des Projekts kann die Entwicklung von Prozessen zum Umgang mit Forschungsdaten an den Instituten unterstützt werden. Dies kann exemplarisch mit den Pilotinstituten erfolgen.

Serviceentwicklung: Das Projekt sieht die Entwicklung und Durchführung oben skizzierter Services im Bereich „Forschungsdaten“ vor. Diese sollen langfristig verstetigt werden.

Nachweisbarkeit und Interoperabilität: Durch die qualitätsgesicherte Erfassung von Metadaten wird die Nachweisbarkeit von Forschungsdaten gewährleistet. Die Metadaten basieren auf dem Standard DataCite 4.0, wodurch die Interoperabilität der Daten abgedeckt ist.

Qualität: Die Qualität der Metadaten wird durch die manuelle Qualitätskontrolle der erfassten Metadaten gewährleistet.

Qualifikation: Dieser Herausforderung kann im Rahmen von FORDATIS nur zu einem Teil entsprochen werden. Für Wissenschaftler werden Schulungen zum Thema Forschungsdatenmanagement angeboten. Die Qualifizierung von Data Scientists und Data Analysts muss in einem anderen Rahmen erfolgen.

Rechtliche Rahmenbedingungen der Nachnutzung: Die EU macht diesbezüglich Empfehlungen hin zu den Lizenzen CC-BY und CC-O.¹¹ Für Fraunhofer könnten rechtliche Rahmenbedingungen in einer Forschungsdaten-Policy festgelegt werden.

Disziplinspezifische Kulturen: Durch die Zusammenarbeit mit verschiedenen Pilotinstituten soll der disziplinspezifische Umgang mit Forschungsdaten bei Fraunhofer näher untersucht werden. Die Ergebnisse werden in die Ausgestaltung von FORDATIS mit einbezogen.

Die Arbeiten an Fraunhofer-FORDATIS werden nach Abschluss der Projektphase 1 (Juni 2018) voraussichtlich noch nicht beendet sein. Der Projekt-Antrag sieht eine weitere Phase vor, in der die gewonnenen Erkenntnisse aus Umfragen und Zusammenarbeit mit Pilot-Instituten umgesetzt und die aufgebaute Infrastruktur zum Beispiel durch Anbindung an weitere, im Aufbau befindlichen Infrastrukturen, wie die Nationale Forschungsdateninfrastruktur¹² oder Europäische Open Science Cloud¹³ erweitert werden soll. In diesem Rahmen sind weitere Standardisierungsvorgaben „Top-Down“ zu erwarten. Zentrale Services, wie Beratung oder Schulungen sollen optimiert und verstetigt werden. Die Feinplanung der Phase 2 ist abhängig von den Erkenntnissen

11 Open Research Data Pilot, accessed 29th March 2017, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

12 RfII – Rat für Informationsinfrastrukturen: Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, Göttingen 2016, S. 2

13 European Open Science Cloud, accessed 29th March 2017, <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

aus Phase 1, insbesondere der Umfrage und den Erfahrungen mit den Pilotinstituten sowie den Erkenntnissen aus JERRI.

Ein weiterer, großer Themenbereich für Phase 2 ist die Langzeitarchivierung der Forschungsdaten. Die Anforderungen, die in Phase 1 definiert werden, sollen dabei umgesetzt werden. Angestrebt ist eine Langzeitarchivierung entsprechend dem OAIS-Modell¹⁴. Es müssen Fragen hinsichtlich der Datenmenge und der Dauer der Archivierung geklärt werden.

Literaturangaben

„EU Digital Market Strategy“. Online verfügbar unter https://ec.europa.eu/commission/priorities/digital-single-market_en. Zuletzt geprüft am 29.03.2017.

Homepage der AG Open Science. Online verfügbar unter <http://www.ag-openscience.de/>. Zuletzt geprüft am 29.03.2017.

„Fraunhofer Zahlen und Fakten“. Online verfügbar unter <https://www.fraunhofer.de/de/ueber-fraunhofer/profil-selbstverstaendnis/zahlen-und-fakten.html>. Zuletzt geprüft am 29.03.2017.

“Open Data Pilot im Forschungsprogramm Horizon2020”. Online verfügbar unter <https://www.openaire.eu/opendatapilot>. Zuletzt geprüft am 29.03.2017

“EARTO Paper on OpenX”. Online verfügbar unter http://www.earto.eu/fileadmin/content/Website_2/EARTO_Paper_on_Open_X_-_13_November_2015_-_Final.pdf. Zuletzt geprüft am 29.03.2017.

“OpenAIRE Portal der Europäischen Kommission”. Online verfügbar unter <https://www.openaire.eu>. Zuletzt geprüft am 29.03.2017.

„DFG: Leitlinien zur guten wissenschaftlichen Praxis“. Online verfügbar unter http://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/gwp/index.html. Zuletzt geprüft am 29.03.2017.

“Data management and the curation continuum. How the Monash experience is informing repository relationship”. Online verfügbar unter http://www.valaconf.org.au/vala2008/papers2008/111_Treloar_Final.pdf. Zuletzt geprüft am 29.03.2017.

Büttner, Stephan, Hans-Christoph Hobohm, and Lars Müller (Hrsg.). 2011. *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen: 29-32.

“Open Research Data Pilot”. Online verfügbar unter http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf. Zuletzt geprüft am 29.03.2017.

14 Nestor Handbuch Forschungsdatenmanagement, accessed 29th March 2017, http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_368.pdf

RfII – Rat für Informationsinfrastrukturen. 2016. *Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*. Göttingen: 2.

European Open Science Cloud. Online verfügbar unter <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>. Zuletzt geprüft am 29.03.2017.

Nestor Handbuch Forschungsdatenmanagement. Online verfügbar unter http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_368.pdf. Zuletzt geprüft am 29.03.2017.