

DIESE SEITE WURDE ABSICHTLICH LEER GELASSEN

Kapitel 1

Einführung

Die Veröffentlichung, langfristige Verfügbarkeit sowie Nachnutzbarkeit von Forschungsdaten (FDs) nach den FAIR-Prinzipien (siehe Abschnitt 2.1.1) gewinnen in der Wissenschaft zunehmend an Bedeutung. Dies wird, zumindest in Teilen, durch ein wachsendes Bewusstsein für die Bedeutung von Transparenz und Nachvollziehbarkeit wissenschaftlicher Ergebnisse und den damit einhergehend zunehmenden Forderungen nach Open Access und Open Data Initiativen, getrieben. So rückt das Thema Forschungsdatenmanagement (FDM) zunehmend in den Fokus der Wissenschaft.

Entsprechend dieser stetig wachsenden Bedeutung von FDM sind in den Bibliotheks- und Informationswissenschaften folgende Themen hierzu vermehrt Gegenstände präskriptiver und deskriptiver Forschung: (i) wie FDs langfristig gelagert und zugänglich gemacht werden können und wie diese bisher tatsächlich gelagert und zugänglich gemacht werden, (ii) welche Metadatenklassifikationssysteme hierzu genutzt werden sollten und welche in der Praxis bisher genutzt werden, (iii) welche Datenbanken für langfristige Lagerung und Zugänglichkeit geeignet sind und welche nicht, (iv) welche Möglichkeiten es gibt, den richtigen Umgang mit FDs durch die Forschenden zu fördern und (v) wie eine entsprechende Infrastruktur aufgebaut werden kann und welche Infrastruktur bereits existiert.

*Bisherige
Forschung*

Da sie für den Aufbau und die Bearbeitung von sehr hoher Relevanz ist, ist hier spezifisch zu erwähnen, dass, als Teil der deskriptiven Forschung zu FDM, sich für die Klassifikation von FDs in der Literatur eine hierarchische Klassifikationsmöglichkeiten nach Publikationsart herauskristallisiert hat. Diese evaluiert, inwiefern die FDs mit dem dazugehörigen Text verknüpft bzw. in diesen integriert worden sind. So lässt sich zwischen drei verschiedenen Publikationsformen für FDs unterscheiden [1, S. 36ff.]:

Klassifikation

1. vollständig integrierte Daten (z.B. Tabellen und Grafiken, die in der PDF-Datei eingebettet worden sind),
2. beigefügte Daten (z.B. Dateien, die gemeinsam mit der PDF-Datei der Publikation auf den Publikationsserver hochgeladen worden sind)
3. Daten, die separat zu der dazugehörigen Publikation auf einer Plattform für FD hochgeladen werden (z.B. fachspezifische FD-Repositories)

Details zu dieser Klassifikationshierarchie werden in Kapitel 2 näher erläutert.

Die meiste Literatur zum Thema FDM ist eher allgemeiner Natur oder hat ein Hauptaugenmerk auf jene FDs, welche mit Artikeln in wissenschaftlichen Journals produziert wurden. Meist zumeist weniger beachteter Aspekt sind jene Daten, die im Rahmen des Verfassens einer Hochschulschriften generiert worden sind und daher nicht primär mit einem Journalartikel assoziiert werden. Dies bedeutet allerdings nicht, dass zu

Wissenslücke

diesem Thema eine absolute Untersuchungsarmut besteht. Jedoch sind Publikationen zu diesem Thema zumindest bisher meist eher präskriptiver statt deskriptiver Natur: So geben im wissenschaftlichen Kontext z.B. präskriptive Publikationen aus dem DFG-Förderprojekt „eDissPlus“ [2, 3, 4] sowie die „Policy für dissertationsbezogene FDs“ der Deutschen Nationalbibliothek [5] vermehrt Richtlinien für den Umgang mit FDs in Hochschulschriften. Es fehlen bisher allerdings umfassende Studien zur Wirksamkeit bzw. Durchsetzung dieser Richtlinien bei Studierenden und Nachwuchsforschenden (z.B. durch entsprechende Prüfungsordnungen und Beratungsangebote zu diesem Thema durch Universitätsbibliotheken). Hier existiert bisher größtenteils nur eher spezialisierte oder fachspezifische Untersuchungen.

Motivation

Dabei sind Hochschulschriften—insbesondere Promotionsvorhaben—von besonderer Relevanz für das lokale Angebot von wissenschaftlichen Bibliotheken: Dissertationen sind, neben Habilitationsschriften, die wichtigsten Hochschulschriften, die an Institutionen von Forschenden produziert werden. Noch dazu werden diese Schriften zunehmend in institutionellen Repositorien veröffentlicht, welche wiederum zumeist entweder von der wissenschaftlichen Bibliothek betrieben oder zumindest mitbetreut werden. Entsprechend ist eine fachgerechte, langfristige sowie zugängliche FD-Infrastruktur, die sich an relevante internationale Standards hält, sowie passende Beratungsgebote für Forschende zum Thema FDM unabdingbar. Um dies jedoch effizient bewerkstelligen zu können, benötigen Bibliothekare ausreichende empirisch-deskriptive Daten sowie präskriptive Richtlinien, an denen sie sich orientieren und von denen sie ihr lokales Beratungs- und Infrastrukturangebot ableiten können. Daher ist es unabdingbar, diese Lücke in der wissenschaftlichen Literatur sukzessive zu minimieren, um die Effizienz und Relevanz bibliothekarischer Arbeit in diesem Bereich zu optimieren.

Forschungsfragen

Um dieser sukzessiven Bemühung beizusteuern, stellt diese Arbeit zwei thematisch zentrale Forschungsfragen, welche sich jeweils in zwei und fünf untergeordnete Forschungsfragen aufgliedern lassen. Hierbei konzentriert sich diese Arbeit speziell auf das Unterthema Dissertationen und Promotionsvorhaben, da diese durch ihre wissenschaftlich hohe Relevanz und Frequenz, den wichtigsten Teil der Hochschulschriften ausmachen. Die Forschungsfragen lauten wie folgt:

1. Inwiefern wird der Umgang mit FD für Promotionsvorhaben in Deutschland bereits in verbindlichen verwaltungsrechtliche Dokumente geregelt?
 - a) Inwiefern wird der Umgang mit FD in den allgemeinen Richtlinien einer Institution geregelt, welche auch Promovierende und ihr Forschungsvorhaben betreffen?
 - b) Inwiefern wird der Umgang mit FD in Promotions- und Prüfungsordnungen geregelt?
2. Auf welche Art und Weise haben Promovierende, welche ihre Dissertationen 2012–2023 in einem institutionellen Repositorium publiziert haben, ihre FD in ihre Publikation integriert?

-
- a) Für welchen Anteil an Dissertationen wurden FD in welcher Form nach dem Klassifikationssystem von REILLY u. a. [1] veröffentlicht?
 - b) Inwiefern hat sich das Publikationsverhalten zu FD in Dissertationen in den letzten zwölf Jahren verändert?
 - c) Inwiefern unterscheiden sich die verschiedenen wissenschaftlichen Fachrichtungen in ihrem Publikationsverhalten in Bezug auf FD aus Dissertationen?
 - d) Wie wird in den Metadaten von FD aus Dissertationen sichtbar gemacht, dass es eine dazugehörige Dissertation gibt?
 - e) Wie wird in den Metadaten von Dissertationen sichtbar gemacht, dass es dazugehörige FD gibt?

Die erste Forschungsfrage behandelt die deutschlandweite Situation zu FD in Promotionsvorhaben aus verwaltungsrechtlicher Sicht, während die zweite zentrale Forschungsfrage dieser Arbeit das tatsächliche Publikationsverhalten von Promovierenden in Bezug auf FD in den letzten zwölf Jahren untersucht. Da zur zweiten Forschungsfrage eine umfassende und differenzierte Auswertung aller in diesem Zeitraum in Deutschland erschienenen Dissertationen im Rahmen einer Masterarbeit nicht stemmbar wäre, konzentriert sich diese Arbeit explizit nur auf jene Dissertationen, die an der Leibniz Universität Hannover (LUH) im institutionellen Repositorium der Leibniz Universität Hannover (LUH-Repo) der Technischen Informationsbibliothek (TIB) veröffentlicht wurden.

Diese Masterarbeit verfolgt einen mehrstufigen methodischen Ansatz zur Beantwortung der einzelnen Forschungsfragen. Die Arbeit ist hierfür in mehrere Kapitel gegliedert, die jeweils spezifische Aspekte der Forschungsfragen untersuchen und beantworten. In Kapitel 2 wird zunächst der aktuelle Stand der Forschung zu FDs und FDMs, insbesondere in Bezug auf Promotionsvorhaben, dargestellt. Dieses Kapitel beleuchtet dabei die bibliothekarische Klassifikation von FDs sowie die Veröffentlichungsraten und -arten in verschiedenen wissenschaftlichen Disziplinen. Kapitel 3 analysiert verwaltungsrechtliche Dokumente einer repräsentativen Stichprobe promotionsberechtigter Institutionen in Deutschland. Der Fokus liegt dabei auf Richtlinien und Anforderungen, die den Umgang mit FDs betreffen, einschließlich allgemeiner und promotionsspezifischer Dokumente wie Promotions- und Prüfungsordnungen. In Kapitel 4 wird das LUH-Repo und die sich darin befindenden Dissertationen in Bezug auf primäre FDs, die während des dazugehörigen Promotionsvorhabens generiert worden sind, untersucht. Es wird geprüft, wie sich die Fakultäten im Umgang mit FDs unterscheiden und wie sich diese Praxis über einen Zeitraum von zwölf Jahren entwickelt hat. Zusätzlich wird in diesem Kapitel untersucht, inwiefern die darin geleistete Arbeit sich dafür eignet, via maschinelles Lernen einen automatischen binären Klassifikator zu entwickeln. Abschließend werden in Kapitel 5 die Erkenntnisse aus den vorherigen Kapiteln zusammengeführt. Es werden allgemeine Schlussfolgerungen zu den Forschungsfragen der Masterarbeit gezogen und spezifische Empfehlungen für die LUH und die TIB formuliert. Der Aufbau der einzelnen Kapitel wird jeweils zu Beginn des jeweiligen Kapitels detailliert erläutert.

*Ziele und
Dokument-
struktur*

DIESE SEITE WURDE ABSICHTLICH LEER GELASSEN

Kapitel 2

Aktueller Stand der Forschung

In diesem Kapitel werden die, für diese Masterarbeit, notwendigen und wichtigsten Aspekte zum aktuellen Stand der Forschung zusammengefasst, um den Beitrag dieser Abschlussarbeit in dem breiteren wissenschaftlichen Kontext einordnen zu können.

Thema des Kapitels

Hierzu werden in Abschnitt 2.1 allgemeine Informationen zu Forschungsdaten (FDs) und Forschungsdatenmanagement (FDM) dargestellt. Dies umfasst, wie FDs definiert und klassifiziert werden, welche Prinzipien und Standards nach heutiger Auffassung in Relation zu FDs befolgt werden sollten und wie und von wem FDs aus empirischer Sicht bisher veröffentlicht werden. Darauf folgend werden in Abschnitt 2.2 bisherige Studien zu dem Thema FD-Richtlinien in Deutschland zusammengefasst. Schließlich wird in Abschnitt 2.3 eine Übersicht gegeben, inwiefern das Publizieren von FDs im Rahmen von Dissertationen bereits erforscht wurde und was die dazugehörigen Ergebnisse waren.

Aufbau des Kapitels

2.1 Grundlagen zu Forschungsdaten

Seit etwa zehn Jahren rückt der korrekte und nachhaltige Umgang mit FDs zunehmend in den Fokus der deutschen Wissenschaft. Dies lässt sich z.B. durch die wissenschaftspolitische Entwicklungen der letzten Jahre erkennen.

Wichtigkeit

So betonte bereits 2014 die Hochschulrektorenkonferenz (HRK), dass das exponentielle Wachstum und die Komplexität digitaler Forschungsdaten neue Methoden und Werkzeuge erfordern. Hier forderte sie Hochschulleitungen explizit dazu auf, Leitlinien zu entwickeln, Informationskompetenz zu fördern und effizientes Forschungsdatenmanagement zu unterstützen. Dies solle von Bund und Länder durch übergreifende Maßnahmen koordiniert und zumindest in Teilen finanziert werden [6]. Nur wenig später hat auch der von der Gemeinsamen Wissenschaftskonferenz (GWK) beschlossene Rat für Informationsinfrastrukturen (RfII) entsprechende Empfehlungen veröffentlicht und vorgeschlagen, dass eine Nationale Forschungsdateninfrastruktur (NFDI) aufgebaut werden sollte [7], was auch von der Deutschen Forschungsgemeinschaft (DFG) unterstützt wurde [8]. Dies kulminierte 2018 in einer Vereinbarung zwischen Bund und Ländern, eine entsprechende Infrastruktur aufzubauen [9], was wiederum zu der Gründung des NFDI e.V. in 2020 führte [10].

NFDI

Darüber hinaus wird zum Zeitpunkt dieser Abschlussarbeit vom Bundesministerium für Bildung und Forschung (BMBF) angestrebt, dass ein FD-Gesetz verabschiedet wird, welches den Zugang auf Daten öffentlicher Hand erleichtern, Datenschutzregelung zu Gunsten besser FD-Handhabung verändern, FDs durch neue Metadatenkataloge der Forschungseinrichtungen auffindbarer machen und ein *Micro Data Center* für

FD-Gesetz

Statistik- und Registerdaten aufbauen soll [11]. Dieses Vorhaben wird auch sowohl von der DFG [12] wie auch von dem RfII [13] unterstützt.

Meinung der
Wissenschaftler

Entsprechend ändert sich auch langsam die Bereitwilligkeit von Forschenden, ihre FDs anderen Personen zugänglich zu machen. So zeigte z.B. eine Studie in 2019, dass nur 11 % älterer Forschenden im Bereich der Biologie gewillt sind, Daten auf Anfrage anderer Forschenden zur Verfügung zu stellen, während dieselbe Rate bei 72 % für Nachwuchsforschende in derselben Disziplin liegt [14].

Aufbau des
Abschnitts

Der korrekten wissenschaftlichen Handhabung von FD und FDM wird also sehr hohe Bedeutung beigemessen. In diesem Abschnitt werden hierzu die grundlegenden Informationen zum aktuellen Stand wiedergegeben. In Abschnitt 2.1.1 werden die Regeln der guten wissenschaftlichen Praxis (GWP) und die FAIR-Prinzipien, auf die in der GWP referenziert werden, erklärt. In Abschnitt 2.1.2 werden die unterschiedlichen Publikationsarten für FDs dargelegt. Eine von diesen, die Nutzung von externen FD-Repositorien, wird in Abschnitt 2.1.3 eingehender erklärt. In Abschnitt 2.1.4 werden dann die am häufigst genutzten Metadaten-Schemata, die für FDs verwendet werden, grob umrissen.

2.1.1 Gute wissenschaftliche Praxis und FAIRe Forschungsdaten

Als Grundlage für das wissenschaftliche Verhalten in Deutschland gelten die Regeln der GWP der DFG [15]. Diese spezifizieren korrektes wissenschaftliches Verhalten für Forschungsvorhaben und wie deren Umgebung gestaltet werden sollte. Dies umfasst organisatorische Angelegenheiten wie die Leitung wissenschaftlicher Einrichtungen bis zu operativen Anforderungen wie die phasenübergreifende Qualitätssicherung innerhalb eines Forschungsprojektes.

Da FDs ein elementarer Bestandteil des wissenschaftlichen Prozesses sind, geben die Regeln der GWP entsprechende Leitlinien für deren Produktion, Bearbeitung und Publikation vor. So sollten z.B., gemäß *Leitlinie 7* der GWP, die Generierung, Prozessierung und Analyse von FDs unter Einhaltung fachspezifischer Standards und Methoden vollzogen und deren Umfang und Art ausreichend dokumentiert werden. Von besonderem Interesse für diese Abschlussarbeit ist dabei *Leitlinie 13* der GWP, welche besagt, dass öffentlicher Zugang zu den Forschungsergebnissen, gemäß den sogenannten FAIR-Prinzipien, hergestellt werden sollte.

FAIRe Daten

Die FAIR-Prinzipien besagen, dass FDs *Findable* (dt. Auffindbar), *Accessible* (dt. Zugänglich), *Interoperable* (dt. Interoperabel) und *Reusable* (dt. Wiederverwendbar) sein sollten [16]. Diese Richtlinien lassen sich wiederum in einzelne Unterempfehlungen aufgliedern, wie das entsprechende Ziel erreicht werden sollte oder was notwendig ist, um diese Ziel erreichen zu können. Diese Unterempfehlungen werden durch den jeweiligen Anfangsbuchstaben des FAIR-Akronyms und einer aufsteigenden Ziffer voneinander differenziert.

Findable

Unter *Findable* wird Forschern empfohlen, dass Daten und dazugehörige Metadaten so beschrieben werden sollten, dass sie sowohl für Menschen als auch für Maschinen leicht auffindbar sind. Um dies zu erreichen sollten Daten und Metadaten eine

global eindeutige und dauerhafte Kennung erhalten (F1), Daten mit umfangreichen Metadaten beschrieben werden (F2), Metadaten eindeutig und explizit die Kennung der beschriebenen Daten enthalten (F3) und (Meta-)Daten in einer durchsuchbaren Ressource registriert oder indiziert worden sein (F4).

Unter *Accessible* wird verstanden, dass, einmal auffindbar, die Daten auch leicht zugänglich sein sollten. Dies bedeutet, dass die Daten so zugänglich gemacht werden, dass diese unter Angabe der (Meta-)Datenkennung nach einem standardisierten Kommunikationsprotokoll abgerufen werden können (A1) und auf die Metadaten auch dann zugegriffen werden kann, wenn die Daten nicht mehr verfügbar sein sollten (A2). Hierbei sollte das Protokoll offen sowie universell implementierbar sein (A1.1) und, bei Bedarf, ein Authentifizierungs-/Autorisierungsverfahren ermöglichen (A1.2).

Accessible

Unter *Interoperable* wird spezifiziert, dass Daten und Metadaten in Formaten vorliegen sollten, die die Integration und das Zusammenspiel mit anderen Daten und Anwendungen ermöglicht. Hierfür soll eine zugängliche, gemeinsame, formale und allgemein anwendbare Sprache für die Wissensrepräsentation in den Daten und Metadaten genutzt werden (I1), Vokabulare verwendet werden, die den FAIR-Prinzipien folgen (I2) und die (Meta-)Daten sollen qualifizierte Verweise auf andere (Meta-)Daten enthalten (I3).

Interoperable

Schlussendlich besagt *Reusable*, dass die Daten so aufbereitet und dokumentiert sein sollten, dass sie von anderen Forschern unter den angegebenen Bedingungen wieder- und weiterverwendet werden können. Dies erfordert detaillierte Beschreibungen der Datenquelle, des Datenerstellungsprozesses und der zugrunde liegenden Methoden (R1.2), einer eindeutigen und zugänglichen Nutzungslizenz (R1.1) und die Nutzung domänenrelevanter Community-Standards (R1.3).

Reusable

Hierbei sollte jedoch erwähnt werden, dass freier Zugriff auf Forschung und deren Ergebnisse unter ethischen und rechtlichen Einschränkungen zwar prinzipiell empfehlenswert ist [17], und auch einen gewissen Zitationsvorteil für die Publizierenden bietet [18, 19], jedoch nicht Teil der FAIR-Prinzipien ist. Entsprechend müssen FAIR FDs nicht unbedingt *Open Data* sein.

2.1.2 Publikationsarten

In der Literatur zu FDs wird im Allgemeinen zwischen drei verschiedenen Publikationsarten unterschieden, basierend auf der Integrationsstufe der FDs in Relation zu eines dazugehörigen wissenschaftlichen Schriftwerkes [1]. Diese Stufen werden mit der sogenannten *Publikationspyramide* dargestellt, welche von oben nach unten die drei Publikationsarten mit abnehmender Integrationsstärke zeigt bevor die unterste Stufe schließlich unveröffentlichte Datensätze repräsentiert [1]. Diese Publikationspyramide ist in Abbildung 2.1 dargestellt.

Publikationspyramide

Entsprechend der Datenpublikationspyramide, sind die integriertesten FDs jene, welche als integrierter Bestandteil eines Artikels bzw. eines wissenschaftlichen Schriftstückes veröffentlicht werden, ohne dass eine Trennung zwischen Dokument und FDs stattfindet. Sie besitzen keine eigene Kennung und sind nicht über eigene Metadaten

Integrierte FDs

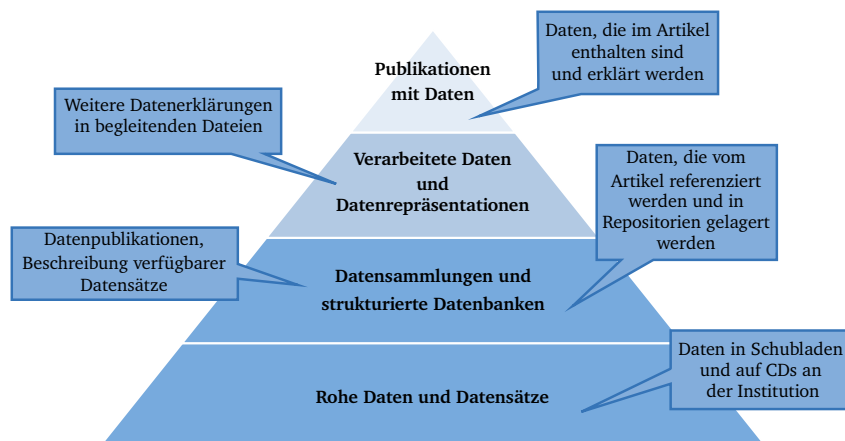


Abbildung 2.1: Die Datenpublikationspyramide nach REILLY u. a. [1], basierend auf der Datenqualitätspyramide nach GRAY [20]. Eigene Übersetzung.

weiter erschlossen [1]. Entsprechend sind sie nach den FAIR-Prinzipien prinzipiell wenig *findable*, *accessible*, *interoperable* oder *reusable* [vgl. 1]. In dieser Abschlussarbeit werden solche FDs als *integrierte FDs* bezeichnet.

Begleitende FDs Die nächste Stufe der Publikationsart ist es, wenn FDs als begleitende Dateien eines Artikels bzw. eines wissenschaftlichen Schriftstückes auf derselben Plattform publiziert werden. Hierbei befinden sich die FDs auf derselben Plattform und besitzen typischerweise keine eigene Kennung oder erschließende Metadaten, können diese jedoch in wenigen Fällen vorweisen [1]. Während die situation hier bereits ist als bei den integrierten Daten, entsprechen diese Daten meist auch unzureichend den FAIR-Prinzipien [vgl. 1]. In dieser Abschlussarbeit werden solche FDs als *begleitende FDs* bezeichnet.

Externe FDs Die dritte und letzte Publikationsart ist es, wenn FDs separat zu dem dazugehörigen Schriftstück auf einer anderen Plattform oder zumindest unter einer anderen Kennung auf derselben Plattform publiziert werden. In diesen Fällen werden die Daten im wissenschaftlichen Schriftstück als externer Datensatz zitiert und befinden sich idealerweise in dedizierten FDs-Repositorien (siehe Abschnitt 2.1.3) oder, was jedoch seltener der Fall ist, in FDs-Journalen, welche eine eingehendere Beschreibung und Indexierung der Daten über entsprechende Metadaten erlaubt (siehe Abschnitt 2.1.4). In solchen Fällen, verweisen die Metadaten des Datensatzes wie auch die des Schriftstückes auf die Existenz des entsprechenden Konterparts [1]. Sie entsprechen, zu unterschiedlich starken Graden, den FAIR-Prinzipien [vgl. 1]. In dieser Abschlussarbeit werden solche FDs als *externe FDs* bezeichnet.

Nicht publiziert Die letzte Stufe der Pyramide betrifft dann jene FDs, die nicht publiziert worden sind und damit auch keine (offizielle) Beziehung zu einem wissenschaftlichen Schriftstück besitzen. Diese Daten befinden sich meistens auf lokalen Servern oder anderen Datenträgern einer wissenschaftlichen Institution [1].

2.1.3 Forschungsdatenrepositorien

Um eine FAIRe Publizierung von FDs zu erlauben, wurden über die Jahre einige Repositorien gegründet, deren dedizierter Zweck es ist, FDs langfristig zu lagern, findbar und verfügbar zu machen. Während diese Repositorien früher stark verstreut waren und es schwer war, einen einigermaßen vollständigen Überblick über diese zu erhalten, wurde vor zwölf Jahren ein Großteil dieser Repositorien mit dem offenen *re3data*-Register mit einem zentralen Index auffindbar und nach eigenen Bedürfnissen filterbar gemacht [21].

FD-Repositorien lassen sich in drei Kategorien einteilen: fachspezifische Repositorien, institutionelle Repositorien und allgemeine Repositorien.

Unter fachspezifische Repositorien werden jene Repositorien verstanden, auf denen nur FDs einer bestimmten wissenschaftlichen Disziplin hochgeladen werden dürfen. Diese machen mit ungefähr zwei Drittel ($n = 2\,212$) aller Repositorien auf *re3data* ($n = 3\,274$) den Großteil aller Repositorien aus [22].

Unter institutionellen Repositorien werden jene Repositorien verstanden, auf denen nur FDs hochgeladen werden, welche an der dazugehörigen Forschungsinstitution produziert wurden oder in Kooperation mit besagter Institution entstanden sind. Diese sind hierbei jedoch meist disziplinübergreifend. Institutionelle Repositorien machen ungefähr ein Fünftel ($n = 761$) aller Repositorien auf *re3data* ($n = 3\,274$) aus [22]. Von diesen sind wiederum 149 institutionelle Repositorien aus Deutschland [23], womit Deutschland, nach den Vereinigten Staaten von Amerika, die zweithöchste Anzahl an institutionellen Repositorien aller Länder hat.

Unter allgemeinen Repositorien werden jene Repositorien verstanden, welche disziplinübergreifend FDs aufnehmen, jedoch nicht an eine bestimmte Forschungsinstitution gebunden sind. Die bekanntesten Beispiele hierfür sind *Zenodo*, *Dryad* und *Figshare*.

Als Basis für FD-Repositorien dienen sowohl speziell dafür entwickelte Software wie auch allgemeine Datenbanksysteme. So ist die am häufigsten genutzte Software für FD-Repositorien *Dataverse*, dicht gefolgt von *DSpace* [22]. Diese machen respektiv 11,11 % ($n = 119$) und 10,27 % ($n = 110$) aller Repositorien auf *re3data*, die Informationen zu der genutzten Software besitzen ($n = 1\,071$), aus (eigene Berechnung auf Basis von Daten aus [22]). Das am häufigsten genutzte allgemeine Datenbanksystem—und das dritthäufigst genutzte System insgesamt—ist *MySQL* mit 8,03 % ($n = 86$; eigene Berechnung auf Basis von Daten aus [22]).

Allen FD-Repositorien gemein ist, dass sie durch Metadaten erlauben, die hochgeladenen FDs weiter zu beschreiben und zu erschließen, was eine gezieltere Suche nach bestimmten FDs erlaubt.

2.1.4 Metadaten-Schemata

Für die Erschließung von Metadaten gibt es einige Standards, welche es über multiple Repositorien hinweg erlauben sollen, Daten uniform beschreiben und suchen zu können. Die am häufigsten genutzten Metadatenschemata für diesen Zweck sind

hierbei *Dublin Core* [24] und *DataCite Metadata Schema* [25], welche respektiv 44,37 % ($n = 595$) und 32,44 % ($n = 435$) aller *re3data*-Repositorien mit Einträgen zu Metadatenschemata mit einem Eigennamen ($n = 1\,341$) [22, 26]. Eine Nutzung mehrerer Metadatenschemata ist möglich.

Der *Dublin Core* Metadata Standard ist ein Metadatenschema, das entwickelt wurde, um die Beschreibung von Ressourcen im Internet zu vereinheitlichen und zu vereinfachen. Das Schema ist hierbei so konzipiert, dass es nur aus grundlegende Elementen besteht, die vielseitig einsetzbar sind. Es besteht aus einem Satz von 15 grundlegenden Elementen (z.B. Titel, Autor, Thema etc.), die auf eine Vielzahl von Ressourcen angewendet werden können. Im *DataCite Metadata Schema* werden Einträge zwischen FDs und den dazugehörigen Schriftstücken jeweils durch *dcterms:isPartOf* und *dcterms:hasPart* verzeichnet.

Das *DataCite Metadata Schema*, hingegen, wurde explizit für FDs entwickelt und beinhaltet eine differenziertere Anzahl an möglichen Metadateneinträgen (insgesamt 22) mit einem kontrollierten Vokabular für viele seiner Einträge. Im *DataCite Metadata Schema* werden Einträge zwischen FDs und den dazugehörigen Schriftstücken typischerweise jeweils durch ein *relatedIdentifier*-Eintrag mit dem *relationType IsSupplementTo* und *IsSupplementedBy* verzeichnet [27, 28]—idealerweise unter Angabe der entsprechenden Digital Object Identifiers (DOIs). Allerdings wird in einigen Fällen auch der *relationType IsReferencedBy* bzw. *References* genutzt, insofern dies auf die FDs und dem dazugehörigen Schriftstück zutrifft [28].

Eine Zuordnung zwischen den beiden Metadatenschemata wurde von der *DataCite Metadata Working Group* erstellt [29]. Hiermit ist eine Konversion zwischen den beiden Standards möglich—allerdings besteht die Möglichkeit, dass einige Informationen verloren gehen, wenn von *DataCite Metadata Schema* zu *Dublin Core* konvertiert wird, da letzteres weniger Informationen codieren kann [29].

2.2 Richtlinien zu Forschungsdaten

FDM-Policies [30, 31, 32] FDM-Policies NUR BIELEFELD [6]

2.3 Forschungsdaten in Dissertationen

DISSERTATION-FD EMPIRICAL [2, 4, 5] [33] [34]

forschungsdatengesetz

–open– Budapest Open Access Initiative Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities European Commission’s Horizon 2020 Open Access Mandate

ALLGEMEIN [17]

[35, 36, 37]

DIESE SEITE WURDE ABSICHTLICH LEER GELASSEN

Literaturverzeichnis

- [1] REILLY, Susan; Wouter SCHALLIER; Sabine SCHRIMPF; Eefke SMIT und Max WILKINSON. *Opportunities of Data Exchange: Report on Integration of Data and Publications*. 2011. Verfügbar unter: <https://hdl.handle.net/10013/epic.40198.d001>.
- [2] WEISBROD, Dirk; Ben KADEN und Michael KLEINEBERG. eDissPlus – Optionen für die Langzeitarchivierung dissertationsbezogener Forschungsdaten aus Sicht von Bibliotheken und Forschenden. In: *{E-Science-Tage: Forschungsdaten managen}*. Humboldt-Universität zu Berlin, 2017., S. 189–198. Verfügbar unter: <https://doi.org/10.18452/22310>.
- [3] KLEINEBERG, Michael und Ben KADEN. Zur Veröffentlichung dissertationsbezogener Forschungsdaten: Perspektiven und Kompetenzen von Promovierenden an Berliner Universitäten. *Bausteine Forschungsdatenmanagement*, 2018-10, Nr. 1, S. 64–69. Verfügbar unter: <https://doi.org/10.17192/bfdm.2018.1.7938>. Verfügbar unter: <https://bausteine-fdm.de/article/view/7938>.
- [4] WEISBROD, Dirk. Pflichtablieferung von Dissertationen mit Forschungsdaten an die DNB – Anlagerungsformen und Datenmodell. *o-bib. Das offene Bibliotheksjournal*, Bd. 5 (2018), Nr. 2, S. 72–78. Verfügbar unter: <https://doi.org/10.5282/o-bib/2018H2572-78>.
- [5] DEUTSCHE NATIONALBIBLIOTHEK [DNB]. *Policy der Deutschen Nationalbibliothek für dissertationsbezogene Forschungsdaten*. Deutsche Nationalbibliothek, 2017. Verfügbar unter: <https://d-nb.info/114060242X/34>.
- [6] HOCHSCHULREKTORENKONFERENZ [HRK]. *Management von Forschungsdaten – eine zentrale strategische Herausforderung für Hochschulleitungen: Empfehlung der 16. Mitgliederversammlung der HRK am 13. Mai 2014 in Frankfurt am Main*. 2014. Verfügbar unter: https://www.hrk.de/fileadmin/_migrated/content_uploads/HRK_Empfehlung_Forschungsdaten_13052014_01.pdf.
- [7] RAT FÜR INFORMATIONSIINFRASTRUKTUREN [RFII]. *Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*. Göttingen: 2016-05. Verfügbar unter: <http://nbn-resolving.de/urn:nbn:de:101:1-201606229098>.
- [8] DEUTSCHE FORSCHUNGSGEMEINSCHAFT [DFG]. *Förderung von Informationsinfrastrukturen für die Wissenschaft: Ein Positionspapier der Deutschen Forschungsgemeinschaft*. 2018. Verfügbar unter: http://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier_informationsinfrastrukturen.pdf.

- [9] GEMEINSAME WISSENSCHAFTSKONFERENZ [GWK]. *Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI)*. 2018-11. Verfügbar unter: <https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf>.
- [10] NATIONALE FORSCHUNGSDATENINFRASTRUKTUR. *Satzung*. 2020-11. Verfügbar unter: <https://www.nfdi.de/wp-content/uploads/2021/05/Satzung-NFDI-eV.pdf>.
- [11] BUNDESMINISTERIUM FÜR BILDUNG UND FORSCHUNG [BMBF]. *Eckpunkte BMBF Forschungsdatengesetz*. 2024-02. Verfügbar unter: https://www.bmbf.de/SharedDocs/Downloads/de/2024/240306_eckpunktepapier-forschungsdaten.pdf.
- [12] DEUTSCHE FORSCHUNGSGEMEINSCHAFT [DFG]. *Die Wissenschaft braucht ein Forschungsdatengesetz!: Positionierung der Deutschen Forschungsgemeinschaft (DFG)*. Version v1. Zenodo, 2023-04. Verfügbar unter: <https://doi.org/10.5281/zenodo.8223970>.
- [13] RAT FÜR INFORMATIONSFRASTRUKTUREN [RFII]. *Anforderungen an die Ausgestaltung eines Forschungsdatengesetzes und eines Dateninstituts*. Göttingen: 2023-07. Verfügbar unter: <https://rfii.de/download/10-07-2023-pressemitteilung-diskussionsimpuls-zum-forschungsdatengesetz-und-dateninstitut/>.
- [14] CAMPBELL, Hamish A.; Mariana A. MICHELI-CAMPBELL und Vinay UDYAWER. Early Career Researchers Embrace Data Sharing. *Trends in Ecology & Evolution*, Bd. 34 (2019), Nr. 2, S. 95–98. Verfügbar unter: <https://doi.org/10.1016/j.tree.2018.11.010>.
- [15] DEUTSCHE FORSCHUNGSGEMEINSCHAFT [DFG]. *Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“*. Weinheim: Wiley-VCH, 2013. ISBN 978-3-527-67918-8. Verfügbar unter: <https://doi.org/10.1002/9783527679188.oth1>.
- [16] WILKINSON, Mark D.; Michel DUMONTIER; IJsbrand Jan AALBERSBERG; Gabriëlle APPLETON; Myles AXTON u. a. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, Bd. 3 (2016), Nr. 1, Artikelnummer 160018. ISSN 2052-4463. Verfügbar unter: <https://doi.org/10.1038/sdata.2016.18>.
- [17] HOPF, David; Sarah DELLMANN; Christian HAUSCHKE und Marco TULLNEY. *Wirkungen von Open Access. Literaturstudie über empirische Arbeiten 2010-2021*. Hannover: Technische Informationsbibliothek [TIB], 2022. Verfügbar unter: <https://doi.org/10.34657/7666>.
- [18] PIOWAR, Heather A. und Todd J. VISION. Data reuse and the open data citation advantage. *PeerJ*, Bd. 1 (2013), e175. ISSN 2167-8359. Verfügbar unter: <https://doi.org/10.7717/peerj.175>.

- [19] BAUTISTA-PUIG, Nuria; Carmen LOPEZ-ILLESCAS; Felix de MOYA-ANEGON; Vicente GUERRERO-BOTE und Henk F. MOED. Do journals flipping to gold open access show an OA citation or publication advantage? *Scientometrics*, Bd. 124 (2020), Nr. 3, S. 2551–2575. ISSN 1588-2861. Verfügbar unter: <https://doi.org/10.1007/s11192-020-03546-x>.
- [20] GRAY, Jim. *Data Quality Pyramid*. 2009-04. Verfügbar unter: <https://www.slideshare.net/JamesWGray/data-quality-pyramid>.
- [21] PAMPEL, Heinz; Paul VIERKANT; Frank SCHOLZE; Roland BERTELMANN; Maxi KINDLING u. a. Making Research Data Repositories Visible: The re3data.org Registry. *PLOS ONE*, Bd. 8 (2013), Nr. 11, e78080. Public Library of Science. Verfügbar unter: <https://doi.org/10.1371/journal.pone.0078080>.
- [22] KHAN, Aasif Mohammad; Fayaz Ahmad LOAN; Umer Yousuf PARRAY und Sozia RASHID. *Information Discovery and Delivery*, Bd. 52 (2024), Nr. 1. Emerald Insight. ISSN 2398-6247. Verfügbar unter: <https://doi.org/10.1108/IDD-07-2022-0069>.
- [23] RE3DATA.ORG. 2024-06. Suchanfrage für alle institutionellen Repositorien aus Deutschland. Verfügbar unter: <https://www.re3data.org/search?query=&types%5B%5D=institutional&countries%5B%5D=DEU>. [Zugriff am 2024-06-06]. Suchergebnis archiviert unter <https://www.doi.org/10.5281/zenodo.11401021>.
- [24] DUBLIN CORE™ METADATA INITIATIVE. *DCMI Metadata Terms*. 2020-01. Verfügbar unter: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>. [Zugriff am 2024-06-06].
- [25] DATACITE METADATA WORKING GROUP. *DataCite Metadata Schema for the Publication and Citation of Research Data and Other Research Outputs*. Version 4.5. DataCite e.V., 2024-03. Verfügbar unter: <https://doi.org/10.14454/g8e5-6293>.
- [26] RE3DATA.ORG. 2024-06. Suchanfrage für alle Repositorien mit Informationen zu einem Metadatenschemata mit Eigennamen. Verfügbar unter: [55](https://www.re3data.org/search?query=%22ABCD+-+Access+to+Biological+Collection+Data%22%7C%22AVM+-+Astronomy+Visualization+Metadata%22%7C%22CF+%28Climate+and+Forecast%29+Metadata+Conventions%22%7C%22CIF+-+Crystallographic+Information+Framework%22%7C%22CIM+-+Common+Information+Model%22%7C%22CSMD-CCLRC+Core+Scientific+Metadata+Model%22%7C%22DCAT+-+Data+Catalog+Vocabulary%22%7C%22DDI+-+Data+Documentation+Initiative%22%7C%22DIF+-+Directory+Interchange+Format%22%7C%22Darwin+Core%22%7C%22DataCite+Metadata+Schema%22%7C%22Dublin+Core%22%7C%22EML+-+Ecological+Metadata+Language%22%7C%22FGDC%2FCSDGM+-+Federal+Geographic+Data+Committee+Content+Standard+for+Digital+Geospatial+Metadata%22%7C%22FITS+-+Flexible+Image+Transport+System%22%7C%22Genome+</p></div><div data-bbox=)

- Metadata%22%7C%22ISA - Tab%22%7C%22ISO+19115%22%7C%22International+Virtual+Observatory+Alliance+Technical+Specifications%22%7C%22MIBBI+ - +Minimum+Information+for+Biological+and+Biomedical+Investigations%22%7C%22MIDAS - Heritage%22%7C%22OAI - ORE+ - +Open+Archives+Initiative+Object+Reuse+and+Exchange%22%7C%22PROV%22%7C%22QuDEX+ - +Qualitative+Data+Exchange+Format%22%7C%22RDF+Data+Cube+Vocabulary%22%7C%22Repository - Developed+Metadata+Schemas%22%7C%22SDMX+ - +Statistical+Data+and+Metadata+Exchange%22%7C%22SPASE+Data+Model%22. [Zugriff am 2024-06-06]. Suchergebnis archiviert unter <https://www.doi.org/10.5281/zenodo.11401021>.
- [27] STARR, Joan und Angela GASTL. isCitedBy: A metadata scheme for DataCite. *D-Lib Magazine*, Bd. 17 (2011), 1/2. ISSN 1082-9873. Verfügbar unter: <https://doi.org/10.1045/january2011-starr>.
- [28] COUSIJN, Helena; Patricia FEENEY; Daniella LOWENBERG; Eleonora PRESANI und Natasha SIMONS. Bringing Citations and Usage Metrics Together to Make Data Count. *Data Science Journal*, Bd. 18 (2019), Nr. 9, S. 1–7. Verfügbar unter: <https://doi.org/10.5334/dsj-2019-009>.
- [29] DATACITE METADATA WORKING GROUP. *DataCite to Dublin Core Mapping 4.4*. DataCite e.V., 2021. Verfügbar unter: <https://doi.org/10.14454/qn00-qx85>.
- [30] BRINEY, Kristin; Abigail GOBEN und Lisa ZILINSKI. Do You Have an Institutional Data Policy?: A Review of the Current Landscape of Library Data Services and Institutional Data Policies. *Journal of Librarianship and Scholarly Communication*, Bd. 3 (), Nr. 2, eP1232. Verfügbar unter: <https://doi.org/10.7710/2162-3309.1232>.
- [31] HIEMENZ, Bea Maria und Monika KUBEREK. *Evaluation institutioneller Forschungsdaten-Policies in Deutschland: Analyse und Abgleich mit internationalen Empfehlungen*. Technische Universität Berlin, 2018-12. Verfügbar unter: <https://doi.org/10.14279/depositonce-7324.2>.
- [32] HIEMENZ, Bea Maria und Monika KUBEREK. Leitlinie? Grundsätze? Policy? Richtlinie? – Forschungsdaten-Policies an deutschen Universitäten. *o-bib. Das offene Bibliotheksjournal*, Bd. 5 (2018), Nr. 2, S. 1–13. Verein Deutscher Bibliothekarinnen und Bibliothekare [VDB]. Verfügbar unter: <https://doi.org/10.5282/o-bib/2018H2S1-13>.
- [33] WÜNSCHE, Stephan. *Forschungsdaten in musikwissenschaftlichen und musikpädagogischen Dissertationen: Eine Stichprobe anhand der im Jahr 2015 in Deutschland angenommenen Arbeiten*. Masterarbeit. Humboldt-Universität zu Berlin, Philosophische Fakultät, 2018. Verfügbar unter: <https://doi.org/10.18452/19457>.

- [34] SCHÖPFEL, Joachim; Južnič PRIMOŽ; Hélène PROST; Cécile MALLERET; Ana ČEŠAREK u. a. Dissertations and Data. In: {*GL17 International Conference on Grey Literature*. Amsterdam, Niederlande: 2015-12.}, S. 1–33. Verfügbar unter: <https://hal.science/hal-01285304v1>.
- [35] MARTIN, Christiane Laura. Wissenschaftliche Bibliotheken als Akteure im Forschungsdatenmanagement. *LIBREAS. Library Ideas: Forschungsdaten. Metadaten. Noch mehr Daten. Forschungsdatenmanagement*, Bd. 23 (2013), S. 12–21. Verfügbar unter: <https://doi.org/10.18452/9039>.
- [36] TENOPIR, Carol; Sanna TALJA; Wolfram HORSTMANN; Elina LATE; Dane HUGHES u. a. Research Data Services in European Academic Research Libraries. *LIBER Quarterly: The Journal of the Association of European Research Libraries*, Bd. 27 (2017), Nr. 1, S. 23–44. Verfügbar unter: <https://doi.org/10.18352/lq.10180>.
- [37] TRÖGER, Beate. Forschungsdatenmanagement an der Universität Münster. *Bibliotheksdienst*, Bd. 50 (2016), Nr. 7, S. 616–622. Verfügbar unter: <https://doi.org/doi:10.1515/bd-2016-0076>.