

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

Die Publikation von Forschungsdaten in den Repositorien von FU Berlin, HU Berlin und TU Berlin. Ansätze zur Analyse von Metadaten

Dorothea Streckerⁱ Maxi Kindlingⁱⁱ
Anastasiia Iarkaevaⁱⁱⁱ

2025

Zitiervorschlag

Strecker, Dorothea, Kindling, Maxi, Iarkaeva, Anastasiia . 2025. Die Publikation von Forschungsdaten in den Repositorien von FU Berlin, HU Berlin und TU Berlin. Ansätze zur Analyse von Metadaten. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 1/2025: S. 2-21. DOI: [10.17192/bfdm.2025.1.8724](https://doi.org/10.17192/bfdm.2025.1.8724).

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

ⁱInstitut für Bibliotheks- und Informationswissenschaft, Humboldt-Universität zu Berlin. ORCID: [0000-0002-9754-3807](https://orcid.org/0000-0002-9754-3807)

ⁱⁱOpen-Access-Büro Berlin, Universitätsbibliothek, Freie Universität Berlin. ORCID: [0000-0002-0167-0466](https://orcid.org/0000-0002-0167-0466)

ⁱⁱⁱQUEST Center, Berlin Institute of Health, Charité - Universitätsmedizin Berlin. ORCID: [0000-0002-7370-1663](https://orcid.org/0000-0002-7370-1663)

Abstract

In diesem Beitrag werden die Metadaten von Forschungsdaten analysiert, die in den institutionellen Repositorien von Freie Universität Berlin, Humboldt-Universität zu Berlin und Technische Universität Berlin publiziert wurden. Ausgehend von den wissenschaftspolitischen Rahmenbedingungen in Berlin und der davon geprägten Entwicklung der drei Repositorien werden die dort indexierten Metadatensätze mit zwei Ansätzen analysiert. Der erste Ansatz wertet Metadaten, die über die OAI-PMH-Schnittstellen in einem DSpace-Austauschformat bereitgestellt werden, mit der Skriptsprache R aus. Der zweite Ansatz nutzt F-UJI, ein automatisiertes Tool zur Bewertung des Erfüllungsgrads der FAIR Data Principles.

Die Analyseergebnisse zeigen, dass die drei Repositorien insgesamt eine gute Beschreibungstiefe für Forschungsdaten erreichen. Die Auswertung mit F-UJI erzielt zwar hohe Werte in der Dimension Auffindbarkeit, aber deutlich weniger in der Dimension Interoperabilität. Eine tiefere Auseinandersetzung mit der Operationalisierung der FAIR Data Principles in F-UJI zeigt, dass die Unterschiede zumindest teilweise auf die verwendeten Metriken zurückzuführen sind. Die beschriebenen Analysemethoden helfen dabei, einen Überblick über in den Repositorien verfügbaren Forschungsdaten anhand ihrer Metadatenbeschreibung zu gewinnen. Zudem liefern sie Ansatzpunkte für die Verbesserung der formal-technischen Metadatenqualität. Werkzeuge zur Bewertung des Erfüllungsgrads der FAIR Principles wie F-UJI sollten jedoch mit Bedacht eingesetzt werden, indem die ausgegebenen Werte kritisch reflektiert und im Kontext des individuellen Repositoriums interpretiert werden.

1 Hintergrund

Institutionelle Repositorien sind ein wesentlicher Bestandteil der Infrastrukturlandschaft zur Unterstützung des Forschungsdatenmanagements und der offenen Zugänglichkeit von Forschungsdaten in Deutschland. Durch kuratierende Maßnahmen ermöglichen sie das Auffinden, Verstehen und Nachnutzen von Datensätzen. Der Erstellung und Pflege von strukturierten Metadatensätzen zur Beschreibung von Forschungsdaten kommt dabei eine besondere Bedeutung zu, da sie die Grundlage zentraler Funktionen von Repositorien bilden. Dieser Beitrag befasst sich mit Metadaten institutioneller Repositorien der drei multidisziplinär ausgerichteten Berliner Universitäten. Er skizziert zunächst den wissenschaftspolitischen Rahmen und die damit verbundene Entwicklung der Repositorien. Anschließend werden die dort indexierten Metadatensätze anhand zweier Ansätze analysiert: (1) eine Auswertung basierend auf Metadaten, die über die OAI-PMH-Schnittstellen bezogen wurden und (2) eine Auswertung des Erfüllungsgrads der FAIR Data Principles mit F-UJI. Die Ergebnisse der gewählten Vorgehensweisen werden gemeinsam betrachtet und reflektiert.

1.1 Wissenschaftspolitische Rahmenbedingungen für den offenen Zugang zu Forschungsdaten im Land Berlin

2015 verabschiedete das Berliner Abgeordnetenhaus die Open-Access-Strategie Berlin [12]. Sie hat die Verbesserung des freien Zugangs auf und die Nachnutzung von digitalen Wissensressourcen zum Ziel. Dazu wurden in der Strategie mehrere Maßnahmen aufgeführt, die dazu dienen, Angehörige der Berliner wissenschaftlichen Einrichtungen bei der Bereitstellung von Open-Access-Publikationen und offenen Forschungsdaten zu unterstützen. Dazu zählt der Ausbau von Publikationsinfrastrukturen, die das elektronische Publizieren befördern. Speziell im Maßnahmenteil für Publikationen wird die „flächendeckende Versorgung mit einer service-orientierten und qualitativ hochwertigen Repositorien-Infrastruktur“ adressiert [12, S. 6]. Für die bessere Nachvollziehbarkeit und Nachnutzung von Forschungsdaten wird in Aussicht gestellt, dass das Land Berlin „den kooperativen Aufbau von Forschungsdaten-Repositorien in enger Abstimmung mit den Wissenschaftlerinnen und Wissenschaftlern der betreffenden Disziplinen [unterstützt], wobei nationale und internationale Entwicklungen zu berücksichtigen und möglichst integrierte Lösungen anzustreben sind“ [12, S. 7]. Zudem sollten die wissenschaftlichen Einrichtungen in ihren jeweiligen Open-Access-Policies entsprechende Grundsätze für den Umgang mit Forschungsdaten integrieren.

Seit der Veröffentlichung der Berliner Open-Access-Strategie hat sich insbesondere das Handlungsfeld „Open Access zu Forschungsdaten“ deutlich weiterentwickelt [10]. Die Berliner wissenschaftlichen Einrichtungen haben seither neben Open-Access-Policies auch Leitlinien zum Umgang mit Forschungsdaten verabschiedet, die die Praxis der möglichst offenen Bereitstellung von Forschungsdaten thematisieren. Nach der Humboldt-Universität zu Berlin (HU Berlin), die bereits zum Zeitpunkt der Verabschiedung der OA-Strategie eine Forschungsdaten-Policy veröffentlicht hatte [7], haben inzwischen auch die Technische Universität Berlin (TU Berlin) [13] und die Freie Universität Berlin (FU Berlin) [4] eine Policy zum Umgang mit Forschungsdaten verabschiedet. Diese Policies empfehlen, Forschungsdaten offen und gemäß der FAIR Data Principles zugänglich zu machen, sofern dem keine rechtlichen oder ethischen Gründe entgegenstehen. Die 2016 veröffentlichten FAIR Data Principles [14] haben sich inzwischen zu einem Quasi-Standard etabliert, der von Seiten der Forschungsförderorganisationen für das Forschungsdatenmanagement eingefordert wird [6, S. 105]. Das 2023 verabschiedete Leitbild für eine Offene Wissenschaft der Berlin University Alliance¹, der die drei Universitäten sowie die Charité - Universitätsmedizin (Charité) angehören, baut wiederum auf den vorhandenen Policies der Einrichtungen auf. Im Berliner Hochschulgesetz wurde 2021 erstmals festgehalten, dass die Berliner Hochschulen den „uneingeschränkten und langfristigen Zugang zu wissenschaftlichen Texten, Forschungsdaten, Software [...] fördern“[1]. Für die Veröffentlichung werden u.a. offene Lizenzen unter Berücksichtigung der Gegebenheiten in verschiedenen Fächerkulturen empfohlen. Die Forschenden der Hochschulen sollen für die Umsetzung u.a.

¹Leitbild für eine Offene Wissenschaft der Berlin University Alliance: <https://www.berlin-university-alliance.de/commitments/research-quality/open-science/Leitbild-fuer-OS/index.html>

die durch die Einrichtungen betriebenen Publikationsdienste wie z.B. Repositorien nutzen. In Form von Hochschulverträgen werden zwischen Land und Hochschulen Vereinbarungen dazu getroffen, wie das Land Berlin die Hochschulen finanziert und welche Leistungen die Hochschulen erbringen müssen. In den Hochschulverträgen 2024 bis 2028 wurde der „freie Zugang zu Ergebnissen im Sinne einer offenen Wissenschaft (Open Research)“ und der Ausbau des freien Zugangs zu Forschungsergebnissen bzw. Publikationen und Forschungsdaten hervorgehoben [2]. Die Berliner Hochschulen und die Charité sollen darüber hinaus die „nachhaltige Nutzung erzielter Forschungsergebnisse und der damit verbundenen Daten“ anstreben und diese Bemühungen in die geplante Open-Research-Strategie Berlin integrieren [8].

1.2 Institutionelle Repositorien für die Veröffentlichung von Forschungsdaten

An den drei Universitäten und der Charité stehen den Forschenden institutionelle Repositorien zur Verfügung: Die HU Berlin betreibt den edoc-Server², die FU Berlin betreibt federführend das Refubium³, das durch FU Berlin und Charité genutzt wird und die TU Berlin das Repository DepositOnce⁴. Die drei Repositorien bauen auf einer langen Tradition im Kontext des elektronischen Publizierens von Dissertationen und der Veröffentlichung von Open-Access-Dokumenten auf. Außerdem können in allen drei Repositorien Forschungsdaten publiziert werden, die beispielsweise im Rahmen von Forschungsprojekten sowie Dissertationen und weiteren Qualifikationsarbeiten entstehen.

In DepositOnce können seit 2014 Forschungsdaten veröffentlicht werden. 2015 wurden Textpublikationen aus einer bestehenden OPUS-Instanz des Repositoriums ergänzt. In den Repositorien edoc-Server und Refubium sind Datenpublikationen seit 2018 möglich.

Die drei Repositorien sind innerhalb der internationalen Repositoriengemeinschaft vernetzt, z.B. durch die Umsetzung der OpenAIRE Guidelines⁵ oder im Rahmen der Confederation of Open Access Repositories (COAR)⁶. Gleichzeitig orientieren sie sich an national etablierten Kriterien des DINI-Zertifikats⁷, einem Qualitätssiegel für Open-Access-Publikationsdienste wie Repositorien. Der edoc-Server und Refubium erhielten das DINI-Zertifikat zuletzt 2019. Auch untereinander kooperieren die betreibenden Einrichtungen. Alle drei Repositorien basieren auf der Open-Source-Software DSpace⁸. Die betreibenden Einrichtungen Charité, FU Berlin, HU Berlin und TU Berlin sind Mitglieder

²edoc-Server: <https://edoc.hu-berlin.de/>

³Refubium: <https://refubium.fu-berlin.de/>

⁴DepositOnce: <https://depositonce.tu-berlin.de>

⁵OpenAIRE Guidelines: <https://guidelines.openaire.eu>

⁶COAR: <https://coar-repositories.org/>

⁷DINI Zertifikat für Open-Access-Publikationsdienste: <https://dini.de/dienste-projekte/dini-zertifikat>

⁸DSpace: <https://dspace.org>

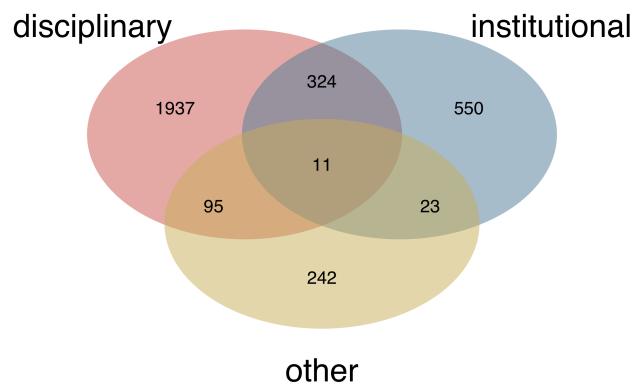


Abbildung 1: Verteilung von Forschungsdatenrepositorien in re3data auf Repositorientypen (Stand: 2024-05-02; N = 3.182)

im DSpace-Konsortium Deutschland. Zudem haben sie gemeinsam ein harmonisiertes Metadatenschema erarbeitet, das die Interoperabilität der Metadaten zwischen den Einrichtungen verbessern soll [5].

Die institutionellen Repositorien sind nicht die einzige Anlaufstelle für Angehörige der Einrichtungen, die ihre Daten publizieren möchten. In vielen Disziplinen sind internationale, disziplinspezifische Forschungsdatenrepositorien etabliert und werden entsprechend in Policies und Leitlinien empfohlen. Sie ermöglichen eine gute Sichtbarkeit innerhalb der jeweiligen wissenschaftlichen Communities. Ein Blick in das Verzeichnis re3data⁹, das Forschungsdatenrepositorien weltweit erfasst, zeigt, dass die Mehrheit der Repositorien eine disziplinäre Ausrichtung hat (s. Abbildung 1).

Es ist davon auszugehen, dass Datensätze, die in den multidisziplinär ausgerichteten, institutionellen Repositorien veröffentlicht wurden, nur einen kleinen Anteil der publizierten Datensätze von Forschenden der Einrichtungen ausmachen. Bislang ist es aber nur sehr eingeschränkt möglich, einen Gesamtüberblick über die Datenpublikationen Angehöriger einer bestimmten Institution zu gewinnen. Dafür gibt es entweder nicht ausreichend Informationen bzw. Nachweise über Datenpublikationen oder die Datenqualität in Nachweisdatenbanken ist nicht zufriedenstellend. Ein Beispiel, wie Datenpublikationen einer Einrichtung erfasst werden können, stellt das Charité Dashboard on Responsible Research¹⁰ des QUEST Center am Berlin Institute of Health (BIH) dar. Die Nachweise über Datenpublikationen unter Charité-Autor*innenschaft werden in Artikel-Volltexten mit Hilfe des Text-Mining-Algorithmus ODDPub detektiert [11]. Institutionelle Repositorien bilden im Fall der Charité einen so geringen Anteil, dass sie nicht in eine separate Kategorie unterteilt werden. Neben der Einschätzung des offenen oder eingeschränkten Zugangs der Datensätze wird die Umsetzung der FAIR Data Principles mithilfe automatisierter Tools evaluiert. Die Ergebnisse dieser Bewertung

⁹re3data: <https://www.re3data.org/>

¹⁰Charité Dashboard on Responsible Research: <https://quest-dashboard.charite.de/#tabStart>

sind im Detail im zugehörigen FAIR Dashboard ¹¹ zu finden.

Um mehr über die Nutzung der institutionellen Repositorien und die Einhaltung von Community-Standards wie der FAIR Data Principles durch die institutionellen Repositorien zu erfahren und somit insgesamt ein vollständigeres Bild der Datenpublikationen zu erhalten, wurden zwei Ansätze zur Analyse der Metadaten der Forschungsdatenpublikationen in den Repositorien von FU Berlin, HU Berlin und TU Berlin ausgewählt. Der erste Ansatz wertet Metadaten, die über die OAI-PMH-Schnittstellen bereitgestellt werden, mit der Skriptsprache R aus. Der zweite Ansatz nutzt F-UJI, ein automatisiertes Tool zur Bewertung des Erfüllungsgrads der FAIR Data Principles. Anschließend werden beide Ansätze reflektiert.

2 Vorgehen

2.1 Analyse der via OAI-PMH ausgelieferten Metadaten

Um die Metadaten der drei Repositorien zu analysieren, wurden diese jeweils über eine OAI-PMH-Schnittstelle der drei Repositorien bezogen. Eingeschränkt wurde die Abfrage auf die OAI-PMH-Sets, die Forschungsdaten beinhalten (s. Tabelle 1), sowie den Zeitraum vom 01.01.2020, 00:00:00 Uhr bis 31.12.2024, 23:59:59 Uhr. Durch die Nutzung der OAI-PMH-Sets konnte auf die Zuschreibung durch die Repositorien Bezug genommen werden, welche Objekte als „Forschungsdaten“ gelten. Die drei Repositorien bieten ihre Metadaten über die OAI-PMH-Schnittstellen in verschiedenen Formaten an. Die Metadaten wurden im Format *DIM* (DSpace Intermediate Metadata)¹² bezogen, da es von allen drei Repositorien angeboten wird und umfangreiche Informationen zu den genutzten Metadatenelementen und -standards enthält. Das resultierende Sample umfasst 78 Datensätze im Refubium, 31 im edoc-Server und 378 in DepositOnce (s. Tabelle 1).

2.2 Analyse von FAIR-Metriken nach F-UJI

Um in Repositorien geteilte Daten sichtbar und nachnutzbar zu machen, sollten sie sowohl disziplinären Qualitätsstandards als auch weiteren formal-technischen Kriterien genügen [9]. Die für die Datenpublikationen genutzten Infrastrukturen wie Repositorien können dies unterstützen, indem sie die etablierten FAIR Data Principles umsetzen [14]. Bei den FAIR Data Principles handelt es sich um Minimalanforderungen an die Datenaufbereitung mit dem Ziel der Maschinenlesbarkeit: Sie sollen auffindbar (**f**indable), zugänglich (**a**ccessible), interoperabel (**i**nteroperable) und nachnutzbar (**r**eusable) bereitgestellt werden.

Um zu bestimmen, inwieweit die Datensätze in Repositorien die FAIR Data Principles

¹¹FAIR Dashboard: <https://quest-dashboard.charite.de/#tabFAIR>

¹²DIM: <https://wiki.lyrasis.org/plugins/viewsource/viewpagesrc.action?pagelId=22021312>

Repositorium	OAI-PMH base URL	Bezeichnung des OAI-PMH-Sets	Anzahl Forschungsdatensätze
Refubium (FU Berlin/Charité)	https://refubium.fu-berlin.de/oai	doc-type:ResearchData	78
edoc-Server (HU Berlin)	https://edoc.hu-berlin.de/oai	researchData	31
DepositOnce (TU Berlin)	https://api-depositonce.tu-berlin.de/serve/oai	col_11303_13955	378

Tabelle 1: Beschreibung der genutzten OAI-PMH-Schnittstellen

erfüllen, können automatisierte Tools verwendet werden. Mithilfe der Tools werden die Metadaten ausgewählter Datensätze („data objects“) und die Datensätze selbst überprüft. Bei der Auswahl des automatisierten Tools wurde ein Fokus auf die effiziente Nachnutzung verfügbarer Ressourcen gelegt. Jedes der automatisierten Tools stellt unterschiedliche Anforderungen an die Nutzung und erfordert eine gewisse Einarbeitung, beispielsweise durch eine lokale Installation oder API-Konfigurationen. Das Tool F-UJI¹³, das vom BIH QUEST Center auch für das FAIR Dashboard der Charité eingesetzt wird¹⁴, hat den Vorteil, dass es für Zwecke wie die hier vorliegende Analyse ohne vorherige Einarbeitung direkt einsatzbereit war. Darüber hinaus hat F-UJI eine gewisse Anerkennung in der Fachcommunity erlangt. So listet beispielsweise die FAIRsharing-Community¹⁵ mindestens zehn Tools auf, die sich automatisiert auf Datensätze bzw. digitalen Objekte anwenden lassen. Dennoch wird in den Ressourcen von FAIR-IMPACT von EOSC¹⁶ ausschließlich F-UJI als automatisiertes Tool empfohlen. Inzwischen zählt F-UJI zu den stabileren und regelmäßig weiterentwickelten Tools. F-UJI basiert auf 16 Metriken, die sich auf zwölf der insgesamt 15 FAIR Data Principles in den vier Dimensionen Findability, Accessibility, Interoperability und Reusability beziehen.

Die in den drei Repositorien als Datensätze identifizierten Objekte wurden einer Prüfung mittels F-UJI v3.2.0¹⁷ unterzogen, basierend auf Version 0.5 der FAIR-Metriken [3].

¹³F-UJI: <https://github.com/pangaea-data-publisher/fuji>

¹⁴Methode: <https://quest-dashboard.charite.de/#tabMethods>

¹⁵Fairassist: <https://fairassist.org>

¹⁶FAIR Implementation resources available in EOSC: <https://catalogue.fair-impact.eu/resources>

¹⁷F-UJI Release-Version v3.2.0: <https://github.com/pangaea-data-publisher/fuji/releases/tag/v3.2.0>

3 Ergebnisse

3.1 Entwicklung des Publikationsaufkommens

Sowohl das Gesamtaufkommen an veröffentlichten Datensätzen als auch die Zuwachsraten in den drei Repositorien haben sich im betrachteten Zeitraum unterschiedlich entwickelt (s. Abbildung 2). Alle drei Repositorien verzeichnen im dargestellten Zeitraum Zuwächse; im Refubium waren zum Zeitpunkt der Datenerhebung 78 Datensätze nachgewiesen, 31 im eDoc-Server und 378 in DepositOnce.

3.2 Nutzung freier Lizenzen

Rechteinformationen sind für jeweils alle Datenpublikationen in den drei Repositorien verfügbar. Abbildung 3 zeigt den Status der Lizenzen, die für Datensätze in einem Repository vergeben wurden. Der Status wird unterteilt in:

- frei im engeren Sinn – CC0 und Softwarelizenzen, die laut SPDX License List¹⁸ von FSF oder OSI als „frei“ angesehen werden
- überwiegend frei – CC BY und CC BY-SA
- nicht frei – copyright und nicht-freie Lizenzen (restriktivere CC-Lizenzen oder CD-LA)
- ohne Standardlizenz – Nutzungsbedingungen des Repositoriums

Abbildung 3 zeigt, dass jeweils 48,7 % (Refubium), 67,7 % (edoc-Server) bzw. 76,2 % (DepositOnce) der Datensätze unter freien Lizenzen im engeren Sinn bzw. überwiegend freien Lizenzen veröffentlicht wurden. Im Refubium wird im Feld *rights* häufig auf die Nutzungsbedingungen des Repositoriums verwiesen (34,6 %) – es werden also nicht immer Standardlizenzen wie beispielsweise die aus der Creative-Commons-Lizenzfamilie verwendet.

3.3 Nutzung von Metadatenelementen

Die Auswertung der Metadaten im DIM-Format zeigt, dass alle drei Repositorien Dublin-Core-Metadatenelemente zur Beschreibung ihres Forschungsdatenbestands nutzen. Alle drei nutzen ergänzend Elemente aus lokal entwickelten Metadatenschemata; Refubium und edoc-Server nutzen zusätzlich auch *dcterms*.

In den Metadaten vom Refubium werden durchschnittlich 14,2 Metadatenelemente zur Beschreibung eines Datensatzes genutzt (10,2 aus Dublin Core), im edoc-Server 12,7 (10,7 aus Dublin Core), und in DepositOnce 11,6 (9,6 aus Dublin Core).

Abbildung 4 zeigt, wie häufig die Metadatenelemente in jeder Sammlung verwendet

¹⁸SPDX License List: <https://spdx.org/licenses/>

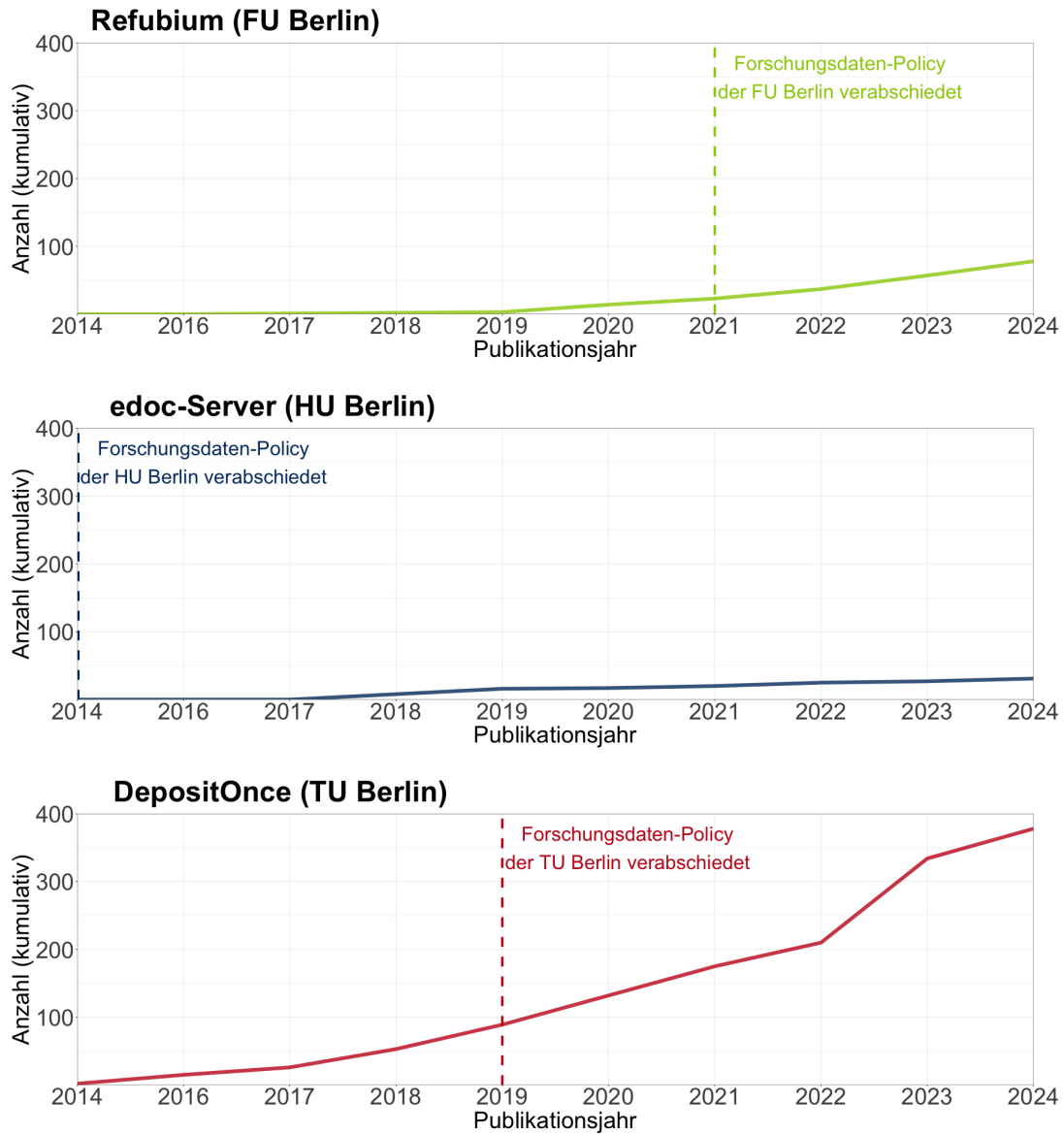


Abbildung 2: Entwicklung des Publikationsaufkommens

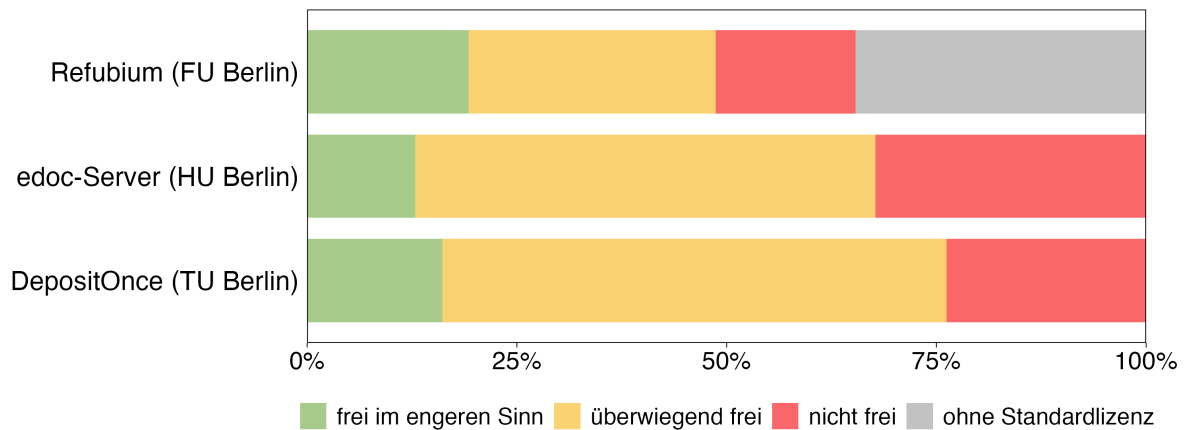


Abbildung 3: Verwendete Lizenzen, zusammengefasst nach Status (n = 561)

werden. Hier zeigen sich einige Unterschiede in der Beschreibungspraxis. Die Metadaten von edoc-Server und DepositOnce beinhalten beispielsweise häufiger Beziehungen zu anderen Forschungsoutputs. Im Gegensatz zu den anderen Repositorien wird das Dublin-Core-Element „publisher“ bei DepositOnce nicht verwendet. Angaben zur zeitlichen und/oder geografischen Abdeckung der Datensätze werden bei Refubium am häufigsten gemacht. Neben dem Dublin-Core-Element „coverage“ werden dafür auch die Elemente „spatial“ und „temporal“ aus dem lokalen Schema genutzt. Metadaten des edoc-Servers beinhalten häufig Angaben zum „department“, während Refubium vergleichsweise oft Angaben zum „funding“ macht. Angaben zu Urheber*innen der Datensätze werden an allen drei Repositorien über das Dublin-Core-Element „contributor“ mit dem qualifier „author“ gemacht.

3.4 Auswertung mit F-UJI

Bei der Bewertung des Erfüllungsgrads der FAIR Data Principles mit F-UJI erreicht der edoc-Server durchschnittlich den höchsten F-UJI score über alle Dimensionen der FAIR Data Principles (65,5%), gefolgt von DepositOnce (65 %) und Refubium (54,6%). Die sogenannte FAIR-Reife („maturity“), die die Werte zwischen 0 und 3 annehmen kann, liegt im Durchschnitt bei 1,3 für Refubium, 1,9 für edoc-Server und 1,9 für DepositOnce.

Abbildung 5 zeigt die Verteilung der F-UJI scores aufgeschlüsselt nach den vier Dimensionen der FAIR Data Principles. Hier wird deutlich, dass an allen drei Repositorien für *findable* die höchsten F-UJI scores erzielt werden, während die Werte für *interoperable* deutlich geringer ausfallen.

Eine genaue Betrachtung der Metriken, die F-UJI zur Bewertung des Erfüllungsgrads der FAIR Data Principles anwendet, zeigt, dass die erreichten Werte in einigen Metriken vergleichsweise gering ausfallen. In insgesamt vier Metriken erreichen mindestens zwei der drei untersuchten Repositorien im Durchschnitt über den untersuchten

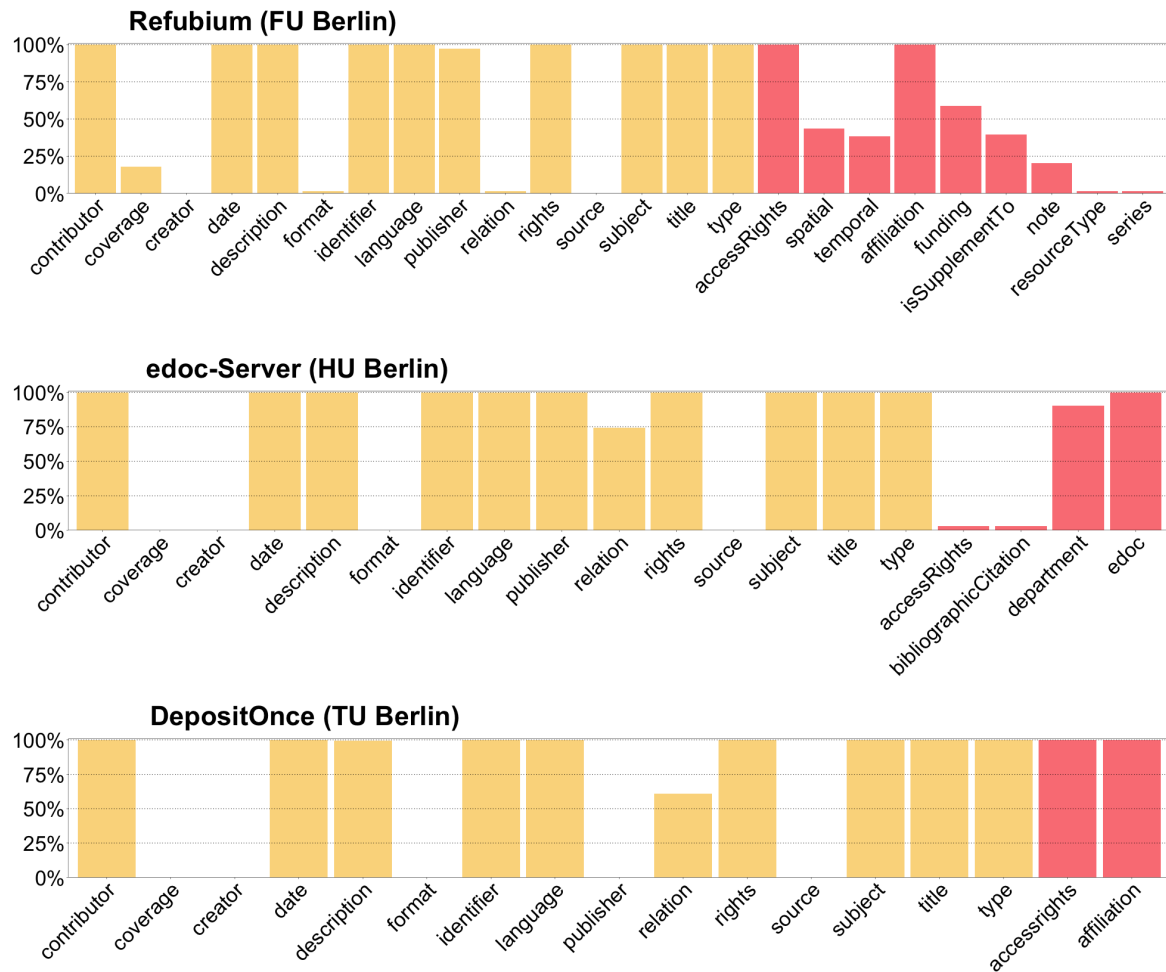


Abbildung 4: Nutzung der Metadatenelemente; gelb: Dublin Core ; rot: andere Metadatenschemata

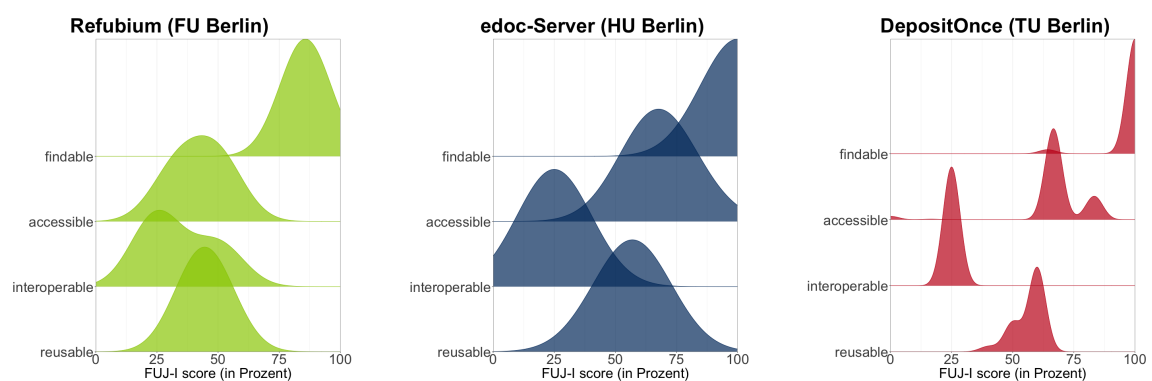


Abbildung 5: Verteilung der F-UJI scores (x-Achse, in Prozent) der Datensätze in den Repositorien nach Dimension der FAIR Data Principles (y-Achse)

Metadaten und Datensätzen einen Erfüllungsgrad von <50 % (s. Tabelle 2).

Metrik	Beschreibung	Erfüllungsgrad (Durchschnitt über die untersuchten Datensätze)	Test erfüllt (pass/fail)
F3-01M	Metadata includes the identifier of the data it describes.	Refubium: 0 % edoc-Server: 100,0 % DepositOnce: 96,6 %	0/78 31/0 365/13
A1-01M	Metadata contains access level and access conditions of the data.	Refubium: 26,9 % edoc-Server: 4,8 % DepositOnce: 10,2 %	42/36 3/28 77/301
A1-03D	Data is accessible through a standardized communication protocol.	Refubium: 0 % edoc-Server: 100,0 % DepositOnce: 96,6 %	0/78 31/0 365/13
I1-01M	Metadata is represented using a formal knowledge representation language.	Refubium: 50,0 % edoc-Server: 50,0 % DepositOnce: 50,0 %	78/0 31/0 378/0
I2-01M	Data is accessible through a standardized communication protocol.	Refubium: 0 % edoc-Server: 0 % DepositOnce: 0 %	78/0 31/0 378/0
I3-01M	Metadata includes links between the data and its related entities.	Refubium: 37,2 % edoc-Server: 0 % DepositOnce: 0 %	29/49 0/31 0/378
R1-01MD	Metadata specifies the content of the data.	Refubium: 25,0 % edoc-Server: 25,0 % DepositOnce: 25,0 %	78/0 31/0 378/0
R1.2-01M	Metadata includes provenance information about data creation or generation.	Refubium: 50,0 % edoc-Server: 50,0 % DepositOnce: 50,0 %	78/0 31/0 378/0
R1.3-02D	Data is available in a file format recommended by the target research community.	Refubium: 0 % edoc-Server: 77,4 % DepositOnce: 86,8 %	0/78 24/7 328/50

Tabelle 2: Erfüllungsgrad von F-UJI-Metriken

Für jede Metrik führt F-UJI zwischen einem und acht Tests durch. Die Tests sind voneinander unabhängig – bei mehreren Tests wird der nächste ausgeführt, auch wenn der vorherige fehlgeschlagen ist. Wenn ein Test erfolgreich ist, gilt die gesamte Metrik als erfüllt. Bei einigen Metriken, z.B. I2-01M, werden mehrere Tests durchgeführt, aber nur für einen Test wird ein Score vergeben. In der folgenden Auflistung werden die F-UJI-Metriken aus Tabelle 2 und die Testergebnisse der drei Repositorien beschrieben.

- **F3-01M:** Diese Metrik legt besonderen Wert auf die Verweise auf Daten in den

Metadaten („point to data“), insbesondere auf die Praktik des Signpostings¹⁹. Dadurch wird sichergestellt, dass alle relevanten Ressourcen (z.B. Links zu Metadaten, Daten, Landing Pages und ergänzende Dokumente), die mit einem Datensatz verknüpft sind, leicht auffindbar sind.

- Die Metadaten enthalten den Identifier der Datensätze, die sie beschreiben, was eine bessere Auffindbarkeit und Zugänglichkeit ermöglicht.
- Am Beispiel von DepositOnce lässt sich zeigen, welchen Einfluss vorhandene Signposting-Informationen auf die Bewertung von Datensätzen haben: 13 Datensätze hätten in diesem Fall höhere Scores erreicht. Dies hat direkten Einfluss auf mehrere Metriken: F3-01M, F4-01M, A1-02D, A1-03D, I3-01M, R1.3-02D und teilweise FsF-R1-01MD. Fehlende Informationen über die Dateien wirken sich daher auf die Scores in mehreren Prinzipien aus, darunter A1, R1 und R1.3.
- **Refubium**: Die Tests für die Metrik schlagen für alle Datensätze fehl – Signpostings wurden im Header der Quellseite nicht gefunden.
- **edoc-Server**: Alle Datensätze erfüllen die Testbedingungen.
- **DepositOnce**: Die überwiegende Anzahl der Datensätze erfüllt die Testbedingungen. 13 Datensätze enthalten keine Angaben zu Dateigröße, -typ oder -link im HTML-Header.
- **A1-01M**: Diese Metrik prüft, ob Zugangsbedingungen in den Metadaten in maschinenlesbarer Form definiert sind.
 - Alle drei Repositorien zeigen eine unzureichende Umsetzung der Metrik, die jedoch ein relativ fortgeschrittenes Niveau voraussetzt.
 - **Refubium**: Etwa 53 % der Datensätze erfüllen die Testbedingungen.
 - **edoc-Server**: Etwa 9 % der Datensätze erfüllen die Testbedingungen.
 - **DepositOnce**: Etwa 20 % der Datensätze erfüllen die Testbedingungen.
- **A1-03D**: Der auflösbare Link zu den Daten soll in den Metadaten enthalten sein, die dem Standard-Kommunikationsprotokoll entsprechen.
 - Dieser Test ist mit dem F3-Prinzip verknüpft. Wenn F3 erfüllt ist, sollte auch A1 erfüllt sein.
 - **Refubium**: Die Tests für die Metrik schlagen für alle Datensätze fehl.
 - **edoc-Server**: Alle Tests für die Metrik sind erfüllt.
 - **DepositOnce**: Alle Tests für die Metrik sind erfüllt. 13 Datensätze enthalten keine Identifier im HTML-Header im Signposting-Format.
- **I1-01M**: Metadaten sollen in einer formalen Sprache zur Wissenrepräsentation für die maschinelle Verarbeitung der Informationen bereitgestellt werden. Dabei sollen standardisierte Formate wie RDF, RDFS oder OWL (z.B. serialisiert in JSON-LD, Turtle, RDF/XML Formaten) verwendet werden.
 - Der Erfüllungsgrad von 50,0 % bei Metrik I1-01M lässt sich dadurch erklären, dass einer der Tests („structured metadata embedded in the landing

¹⁹Signposting: <https://signposting.org/conventions/>

page XHTML/HTML code“) von keinem der untersuchten Repositorien bestanden wird, ein anderer hingegen („graph data (RDF, JSON-LD) through negotiation, typed links or sparql endpoint“) von allen.

- **Refubium, edoc-Server & DepositOnce:** Alle Tests für die Metrik sind erfüllt. Daher sind die Erfüllungswerte von unter 100 % auf die Umsetzung bei den Repositorien und nicht auf die Testkriterien zurückzuführen.
- **I2-01M:** Die Nutzung kontrollierter Vokabulare wird über deklarierte *namespaces* geprüft.
 - Die Tests für die Metrik I2-01M scheinen auf den ersten Blick bei allen Repositorien vollständig unerfüllt. Bei genauer Betrachtung der Test-Debug-Nachricht sowie der F-UJI-Dokumentation wird jedoch deutlich, dass die Entwickler*innen sogenannte „*default namespaces*“ von der Validierung ausschließen – dazu gehören *schema.org* und *Dublin Core*. Ausschließlich die in der Dokumentation gelisteten *namespaces*²⁰ führen zu einem positiven Testergebnis.
 - **Refubium, edoc-Server & DepositOnce:** Alle Tests für die Metrik sind erfüllt. Daher sind die Erfüllungswerte unter 100 % auf die Umsetzung bei den Repositorien und nicht auf die Testkriterien zurückzuführen.
- **I3-01M:** Die Metrik prüft das Vorhandensein von Verweisen auf verwandte Ressourcen.
 - Mehrere Datensätze von Refubium und sämtliche Datensätze von edoc-Server und DepositOnce fallen bei den Test für die Metrik I3-01M durch. Eine detaillierte Analyse der Testergebnisse zeigt, dass diese Datensätze tatsächlich keine verwandten Ressourcen in ihren Metadaten ausweisen. Das negative Testergebnis ist in diesen Fällen darum formal gerechtfertigt, wirft jedoch die Frage auf, inwiefern die Bewertung einer Ressource beeinträchtigt werden sollte, wenn sie tatsächlich in keiner Relationen zu anderen Ressourcen steht.
 - **Refubium & DepositOnce:** Die Tests für die Metrik schlagen für alle Datensätze fehl.
 - **edoc-Server:** Etwa 37 % der Datensätze erfüllen die Tests für die Metrik.
- **R1-01MD:** Die Metrik prüft, ob der Inhalt der Datensätze ausreichend in Metadaten beschrieben ist.
 - Alle drei Repositorien haben in der Metrik R1-01MD lediglich die niedrigste Stufe von drei möglichen erreicht. Ähnlich wie bei der Metrik F3-01M für die *findability* sind bei R1-01MD für die *reusability* Angaben wie Dateigröße, Ressourcentyp und Verlinkungen zu den Daten in den Metadaten relevant. Mit Ausnahme des Ressourcentyps *dataset* fehlen diese Informationen jedoch in den Metadaten. Insgesamt werden für diese Metrik acht Tests durchgeführt, wobei nur vier Tests einen Score erhalten.

²⁰https://github.com/pangaea-data-publisher/fuji/blob/master/fuji_server/data/default_namespaces.txt

- **Refubium, edoc-Server & DepositOnce:** Alle Tests für die Metrik sind erfüllt. Daher sind die Erfüllungswerte unter 100 % auf die inhaltliche Beschreibung durch die Repositorien und nicht auf die Testkriterien zurückzuführen.
- **R1.2-01M:** Die Metrik prüft, ob die Metadaten Angaben zur Provenienz bzw. zu Nutzungsbedingungen enthalten.
 - Die automatisierte Erkennung von Nutzungsbedingungen bzw. Lizenzen für die Metrik R1.2-01M erfordert eine genauere Überprüfung der Ergebnisse. Zwar werden die entsprechenden Metadatenfelder in der Regel korrekt erkannt, jedoch werden die angegebenen Standard-Lizenzen (wie Creative Commons) häufig nicht als gültig validiert, obwohl sie aus menschlicher Sicht korrekt erscheinen. Ein Grund könnte dafür sein, dass F-UJI eine exakte Übereinstimmung mit den offiziellen SPDX-Kennungen und -URLs erwartet. Bereits geringfügige Abweichungen, etwa ein zusätzlicher oder fehlender Zeichensatz in der Lizenz-URL, führen dazu, dass der Datensatz keinen Punkt erhält.
 - Die Metadaten enthalten Angaben zur Datenerstellung oder -versionierung.
 - **Refubium, edoc-Server & DepositOnce:** Alle Tests für die Metrik sind erfüllt. Daher sind die Erfüllungswerte unter 100 % auf die inhaltliche Beschreibung durch die Repositorien und nicht auf die Testkriterien zurückzuführen.
- **R1.3-02D:** Die Metrik prüft, ob die Daten in von der Forschungscommunity empfohlenen Formaten vorliegen.
 - Im Falle von Refubium, ist die Metrik R1.3-02D – verknüpft mit der Metrik F3-01M – ebenfalls nicht erfüllt, da keine Identifier für die Daten in den Metadaten vorhanden sind. Dadurch kann das von der Forschungscommunity empfohlene Dateiformat nicht nachgewiesen werden. In den Fällen, in denen ein Identifier vorhanden war – bei edoc-Server und DepositOnce – konnte das Dateiformat nicht validiert werden, da keine Dateien zur Überprüfung zugänglich waren.
 - Insgesamt werden für diese Metrik acht Tests durchgeführt, wobei nur vier Tests einen Score erhalten.
 - **Refubium:** Die Tests für die Metrik schlagen für alle Datensätze fehl.
 - **edoc-Server:** Etwa 77 % der Datensätze erfüllen die Tests für die Metrik.
 - **DepositOnce:** Etwa 86 % der Datensätze erfüllen die Tests für die Metrik.

4 Diskussion und Fazit

In diesem Beitrag kann keine Aussage dazu getroffen werden, wie sich das Aufkommen publizierter Datensätze an den Einrichtungen insgesamt entwickelt und auf verschiedene Typen von Repositorien – wie etwa institutionelle und disziplinär ausgerichtete – verteilt. Festgehalten werden kann jedoch, dass von der Möglichkeit, Datensätze in den institutionellen Repositorien abzulegen, Gebrauch gemacht wird, und dass die

Metrik	Status	Refubium	edoc-Server	DepositOnce
FsF-F1-01D	pass	78	31	378
	fail	0	0	0
FsF-F1-02D	pass	78	31	378
	fail	0	0	0
FsF-F2-01M	pass	78	31	378
	fail	0	0	0
FsF-F3-01M	pass	0	31	365
	fail	78	0	13
FsF-F4-01M	pass	78	31	365
	fail	0	0	13
FsF-A1-01M	pass	42	3	77
	fail	36	28	301
FsF-A1-02M	pass	78	31	365
	fail	0	0	13
FsF-A1-03D	pass	0	31	365
	fail	78	0	13
FsF-I1-01M	pass	78	31	378
	fail	0	0	0
FsF-I2-01M	pass	78	31	378
	fail	0	0	0
FsF-I3-01M	pass	29	0	0
	fail	49	31	378
FsF-R1-01MD	pass	78	31	378
	fail	0	0	0
FsF-R1.1-01M	pass	78	31	378
	fail	0	0	0
FsF-R1.2-01M	pass	78	31	378
	fail	0	0	0
FsF-R1.3-01M	pass	78	31	378
	fail	0	0	0
FsF-R1.3-02D	pass	0	24	328
	fail	78	7	50

Tabelle 3: Anzahl der erfolgreichen und durchgefallenen Tests für die jeweilige Metrik

Anzahl der veröffentlichten Datensätze stetig wächst, auch wenn die Gesamtzahl der Datenpublikationen und die Zuwachsrate bei den betrachteten Repositorien variieren. Die Ausgestaltung von Arbeitsabläufen bei der Datenpublikation kann einen Effekt auf die Dokumentation, Aufbereitung und Lizenzierung von Forschungsdaten haben. In allen drei untersuchten Repositorien werden die Metadaten beispielsweise von den Autor*innen bei der Ablieferung der Datensätze mit eingereicht und anschließend von Mitarbeiter*innen der Repositorien geprüft. Diese Prüfung berücksichtigt vor allem formale Kriterien, darum hängt die Beschreibungstiefe wesentlich von den Autor*innen der Datensätze ab.

Im Sinne einer Öffnung des Forschungsprozesses sollen Daten offengelegt werden, damit sie überprüft und nachgenutzt werden können. Um dies zu ermöglichen, müssen Metadaten möglichst vollständig und korrekt sein. Dieser Beitrag zeigt zwei Ansätze zur Bewertung von Metadaten auf. Die Auswertung der Metadaten im DIM-Austauschformat zeigt, dass die drei untersuchten Repositorien überwiegend Dublin-Core-Elemente zur Beschreibung ihres Forschungsdatenbestands nutzen, diese jedoch auch um Elemente aus lokal entwickelten Metadataschemata ergänzen. Dublin Core ist einerseits weit verbreitet und gewährleistet die Interoperabilität mit anderen Diensten, andererseits ist es jedoch sehr generisch. Die Anreicherung um Elemente aus lokal entwickelten Schemata kann dem entgegenwirken und Beschreibungen stärker auf den Objekttyp „Forschungsdaten“ zuschneiden.

Auch die FAIR Data Principles können zur Bewertung von Metadaten hinzugezogen werden. Werkzeuge wie F-UJI übersetzen die FAIR Data Principles in Metriken und messen, inwiefern Datenpublikation mit diesen übereinstimmen. Die Auswertung mit F-UJI zeigt in allen Repositorien ähnliche Muster auf: In der Dimension Auffindbarkeit werden hohe Werte erreicht, während in der Dimension Interoperabilität deutlich niedrigere Werte zu beobachten sind. Eine genauere Betrachtung der Implementierung von Metriken in F-UJI verdeutlicht jedoch, dass Vergleiche zwischen den vier Dimensionen der FAIR Data Principles (Auffindbarkeit, Zugänglichkeit, Interoperabilität und Nachnutzbarkeit) nur eingeschränkt möglich sind. Beispielsweise variiert die Punktevergabe zwischen diesen Metriken und damit auch die Scores, die maximal erreicht werden können. Außerdem sind die Metriken unterschiedlich anspruchsvoll – einige setzen eine hohe technische Reife voraus (z.B. R1), während andere mit weniger Aufwand erfüllt werden können. Es ist also möglich, dass die Unterschiede zumindest teilweise auf die verwendeten Metriken zurückzuführen bzw. deren Interpretation durch FAIRsFAIR [3] sind. Neun Metriken wurden vollständig ohne Fehlermeldungen bei allen drei Repositorien bestanden (s. Tabelle 3). Darunter befinden sich vier Metriken – I1-01M, I2-01M, R1-01M, R1.2-01M – für die die Erfüllungswerte unter 50,0 % liegen. Diese niedrigen Werte sind auf eine unzureichende Umsetzung durch die Repositorien zurückzuführen.

Die in diesem Beitrag dargestellten Ansätze zur Bewertung von Metadaten können in einigen Punkten zu unterschiedlichen Ergebnissen führen. Beispielsweise besteht bei der Nutzung generischer Metadataschemata die Gefahr, dass Elemente von F-UJI bei der automatisierten Evaluierung nicht zuverlässig berücksichtigt werden. Sowohl die Dokumentation als auch die Debug-Nachrichten von F-UJI machen deutlich, dass die

Nutzung von Dublin Core das Erreichen hoher Scores erschwert – da das Schema als „default“ oder „common“ gilt, werden dafür keine Punkte in den Metriken I2-01M oder I3-01M vergeben. Die genannten Metriken verlangen explizit den Einsatz domänen-spezifischer Vokabulare. Nichtsdestotrotz werden die Kernelemente des Dublin-Core-Metadatenschemas (s. Abbildung 4 – u.a. „titel“, „creator“, „data identifier“, „publisher“, „publication date“ usw. – bei allen drei Repositorien erfolgreich validiert, insbesondere in der Metrik F2-01M.

Weitere Diskrepanzen zeigen sich bei der Bewertung von Rechteinformationen. Diese werden von F-UJI nicht immer zuverlässig erfasst. Dies betrifft insbesondere Datensätze, die zwar eine standardisierte Lizenz in den Metadaten angeben, jedoch nicht in dem von F-UJI erwarteten Format oder nicht mit anerkannten Standards wie denen von SPDX. Möglichst freie Standardlizenzen wie Creative Commons, wie sie für eine optimale Nachnutzbarkeit von Forschungsdaten empfohlen werden, werden nicht für alle Datenpublikationen vergeben. Die Gründe dafür können unterschiedlich sein. Um diese zu ermitteln, wäre eine detailliertere Untersuchung notwendig, die beispielsweise einbezieht, um welche Art von Daten es sich jeweils handelt, ob Urheberrechtsschutz oder beispielsweise Datenbankschutzrechte greifen und ob diese Datensätze oder Teile darin ggf. nicht offen lizenziert werden können. Auch eine eingehende Befassung mit der Motivation Forschender, bestimmte Lizenzen zu verwenden, wäre lohnenswert. Für die strategische Weiterentwicklung des Themas wäre auch von Interesse, ob Unsicherheiten oder Unkenntnis in Bezug auf rechtliche Fragen dazu führen, dass Forschende eher restriktive Lizenzen wählen.

Einige Faktoren werden von F-UJI in mehreren Metriken geprüft. So reduzieren fehlende Angaben zu Dateigrößen, Dateitypen und Identifiern in den Metadaten bei allen drei Repositorien die erreichten Scores für mehrere Metriken. In einigen Fällen werden Angaben von F-UJI zudem nicht erkannt, obwohl sie prinzipiell vorhanden wären. So werden vorhandene DOIs in den Metadaten von Refubium nicht erkannt, da sie nicht im erwarteten Format vorliegen. Dadurch schlagen alle Tests für F3-01M für dieses Repository fehl. Durch die strengen formalen Anforderungen der Tests in F-UJI kann es zu „false negative“-Ergebnissen wie diesen kommen.

Eine weitere Auffälligkeit bei F-UJI ist, dass Null-Scores nicht zwischen nicht-erfüllten und fehlgeschlagenen Tests unterscheiden. Diese Ergebnisse sollten daher mit Bedacht interpretiert werden.

Die beschriebenen Analysemethoden können einen Überblick über Metadaten von Forschungsdaten liefern und Ansatzpunkte für die Verbesserung der Metadatenqualität aufzeigen. Werkzeuge wie F-UJI bieten einen automatisierten Ansatz für die Bewertung von Metadaten. Die Ergebnisse werden in einem Format ausgeliefert, das über verschiedene Repositorien vergleichbar ist. Sie sollten dennoch mit Vorsicht eingesetzt werden und Nutzende sollten sich bewusst machen, was die Metriken messen. Die ausgegebenen Werte sollten trotz des einfachen Ausgabeformats kritisch reflektiert und im Kontext des individuellen Repositoriums verstanden werden.

5 Danksagung

Der Beitrag wurde in einer frühen Fassung Vertreter*innen der drei beschriebenen Repositorien vorgelegt. Wir bedanken uns herzlich für die Rückmeldungen, die zur weiteren Ausarbeitung der Untersuchung beigetragen haben. Unser Dank gilt außerdem den beiden Gutachter*innen für die wertvollen Hinweise und Anregungen, die in die zweite Fassung des Beitrags eingeflossen sind.

6 Datenverfügbarkeit

Die der Auswertung zugrundeliegenden Daten sind über GitHub verfügbar²¹.

²¹Datenpublikation auf GitHub: <https://github.com/anastasiabright/bua-metadata>

Literatur

- [1] Abgeordnetenhaus von Berlin. *Gesetz über die Hochschulen im Land Berlin (Berliner Hochschulgesetz - BerlHG) § 41 - Forschungsberichte*. Techn. Ber. 2021. URL: <https://gesetze.berlin.de/bsbe/document/jlr-HSchulGBE2011V27P41> (besucht am 26. 06. 2024).
- [2] Abgeordnetenhaus von Berlin. *Hochschulverträge für die Jahre 2024 bis 2028 gemäß § 2a Berliner Hochschulgesetz*. Techn. Ber. Drucksache 19/1250. 2023. URL: <https://www.parlament-berlin.de/adosservice/19/IIIPlen/vorgang/d19-1250.pdf> (besucht am 26. 06. 2024).
- [3] Anusuriya Devaraju u. a. *FAIRsFAIR Data Object Assessment Metrics*. 2022. doi: [10.5281/ZENODO.6461229](https://doi.org/10.5281/ZENODO.6461229).
- [4] Freie Universität Berlin. *Forschungsdaten-Policy der Freien Universität Berlin*. 2021. doi: [10.17169/REFUBIUM-30560](https://doi.org/10.17169/REFUBIUM-30560).
- [5] Tatjana Fritz u. a. *Harmonisiertes Metadatenschema für die DSpace-Repositorien der Berliner Universitäten*. 2022. doi: [10.17169/REFUBIUM-36972](https://doi.org/10.17169/REFUBIUM-36972).
- [6] Horizon Europe. *Horizon Europe (HORIZON) Euratom Research and Training Programme. General Model Grant Agreement*. Techn. Ber. 2024. URL: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/agr-contr/general-mga_horizon-euratom_en.pdf (besucht am 04. 04. 2025).
- [7] Humboldt-Universität zu Berlin. *Grundsätze zum Umgang mit Forschungsdaten an der Humboldt-Universität zu Berlin*. 2014. URL: <https://www.cms.hu-berlin.de/de/dl/dataman/infos/policy> (besucht am 26. 06. 2024).
- [8] Maxi Kindling. *Hochschulverträge sind unterzeichnet: Berliner Hochschulen und das Land Berlin wollen Offene Wissenschaft (Open Research) stärken*. Feb. 2024. doi: [10.59350/hycr5-62h70](https://doi.org/10.59350/hycr5-62h70).
- [9] Maxi Kindling und Dorothea Strecker. "Data Quality Assurance at Research Data Repositories". In: *Data Science Journal* 21 (2022), S. 18. doi: [10.5334/dsj-2022-018](https://doi.org/10.5334/dsj-2022-018).
- [10] Open-Access-Büro Berlin u. a. *Open-Access-Bericht Berlin*. 2024. doi: [10.5281/zenodo.13142722](https://doi.org/10.5281/zenodo.13142722).
- [11] Nico Riedel, Miriam Kip und Evgeny Bobrov. "ODDPub – a Text-Mining Algorithm to Detect Data Sharing in Biomedical Publications". en. In: *bioRxiv* (2020). Publisher: Cold Spring Harbor Laboratory Section: New Results, S. 2020.05.11.088021. doi: [10.1101/2020.05.11.088021](https://doi.org/10.1101/2020.05.11.088021).
- [12] Senat von Berlin. *Open-Access-Strategie für Berlin*. 2015. doi: [10.17169/REFUBIUM-26319](https://doi.org/10.17169/REFUBIUM-26319).
- [13] Technische Universität Berlin. *Forschungsdaten-Policy der Technischen Universität Berlin*. 2019. URL: <https://www.tu.berlin/arbeiten/wichtige-dokumente/richtlinien-leitlinien/forschungsdaten-policy> (besucht am 26. 06. 2024).

- [14] Mark D. Wilkinson u. a. "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific Data* 3.1 (2016), S. 160018. doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).