

Entwicklung effizienter und nachhaltiger Datenmanagementdienste

Giuditta Parolini und Falko Glöckler

Abstract

Effizienz und Nachhaltigkeit spielen eine zentrale Rolle bei der Weiterentwicklung von Forschungsdatenmanagementdiensten. Angesichts des kontinuierlichen Anstiegs der Forschungsdaten und ihrer Nutzung lassen sich diese Ziele mit manuellen Prozessen kaum noch erreichen. In diesem Beitrag wird ein Beispiel für eine Prozesstransformation diskutiert, die zur Umsetzung von Effizienz und Nachhaltigkeit bei der Veröffentlichung von Daten mit DOIs (Digital Object Identifiers) durchgeführt wurde. Es wird die Django-Webanwendung vorgestellt, die das Wissenschaftsdatenmanagement (WSDM) Team am Museum für Naturkunde Berlin entwickelt hat, um den DOI-Registrierungsprozess mithilfe der vom DOI-Anbieter DataCite zur Verfügung gestellten API zu automatisieren.

Efficiency and sustainability play a central role in the continuing development of research data management services. Given the steady increase in research data and their usage, these objectives can hardly be achieved continuing to develop manual processes. The paper will discuss an example of process transformation implemented to achieve efficiency and sustainability when publishing data with DOIs (Digital Object Identifiers). It will present the Django web application that the Science Data Management team at the Museum für Naturkunde Berlin has developed to automate the registration process using the API made available by the DOI provider DataCite.

Unterstützung der Forscherinnen und Forscher, andererseits erkennen Förderinstitutionen immer mehr den Wert wissenschaftlicher Datenpublikationen und stellen höhere Anforderungen an die Kuratierung, Verwaltung und Archivierung von Daten.³ Darüber hinaus entwickeln sich auch die rechtlichen Anforderungen an Forschungsdaten rasch weiter.⁴ Dieser Wandel wirft ernsthafte Fragen zu den Prozessen des Forschungsdatenmanagements auf, wie sie sich weiterentwickeln sollten, um den neuen Anforderungen zu entsprechen. Insbesondere scheinen Effizienz und Nachhaltigkeit kaum mit Prozessen vereinbar zu sein, bei denen der Hauptanteil auf menschlichen Eingriffen und nicht auf automatisierten Vorgängen beruht. Der bisher im Datenmanagement vorherrschende manuelle Ansatz muss einem verstärkten Einsatz technologischer Lösungen weichen, wenn Forschungseinrichtungen die aktuellen Herausforderungen im Datenbereich erfolgreich bewältigen wollen. Der Business Case für die Automatisierung ist bereits klar.⁵ Darüber hinaus sind für Forschungsdaten die Umsetzung von Grundsätzen, wie die FAIR-Prinzipien und Openness, zwei weitere entscheidende Anforderungen, die mit den vorherrschenden manuellen Verfahren kaum vereinbar sind.⁶

Das Museum für Naturkunde Berlin (MfN) erlebt täglich die Herausforderungen, die sich aus den neuen Datenpraktiken ergeben. Das Museum ist mit einem Großprojekt zur Digitalisierung seiner Sammlungen befasst und muss gleichzeitig die vielen Forscherinnen und Forscher, die derzeit im Museum arbeiten, beim Datenmanagement unterstützen. Die Unterstützung der Forscherinnen und Forscher beschränkt sich nicht auf die Erstellung von Datenmanagementplänen. Beratung und Unterstützung

1 Einführung

Effizienz und Nachhaltigkeit spielen eine zentrale Rolle bei der Weiterentwicklung von Forschungsdatenmanagementdiensten.¹ Wie von mehreren Akteuren im Forschungsdatenmanagement hervorgehoben, befinden wir uns bei der Entwicklung von Datenmanagementdiensten an einem Wendepunkt.² Einerseits erzeugt die rasche und anhaltende Zunahme der produzierten Forschungsdaten die Notwendigkeit einer stärkeren

1 Rat für Informationsinfrastrukturen: Datendienste nachhaltig gestalten. Ein Diskussionsimpuls zur Weiterentwicklung von Forschungsdateninfrastrukturen, Göttingen 2020, 6 S.

2 Neuroth, Heike/ Putnings, Markus/ Neumann, Janna: Praxishandbuch Forschungsdatenmanagement, De Gruyter 2021.

3 Jäckel, Denise/ Helbig, Kerstin/ Odebrecht, Carolin: „Desiderate zum Forschungsdatenmanagement 2013 und 2022“, in: Information – Wissenschaft & Praxis 73(5-6) (2022), S. 265-276. <https://doi.org/10.1515/iwp-2022-2239>. Cremer, Fabian/ Fräbendorf, Mathis/ Neumann, Janna/ Petersen, Mareike/ Pramann, Bianca / Schaffner, Sandra/ Skowronek, Thomas/ Toepfer, Ralf: „Institutionelle Workflows zum Forschungsdatenmanagement“. Bestandsaufnahme und Lösungsvorschläge aus der Leibniz-Gemeinschaft, in: Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern 3 (2021), S. 142-150. <https://doi.org/10.17192/bfdm.2021.3.8346>.

4 European Commission, Directorate-General for Research and Innovation, van Echoud, Mirelle: Study on the Open Data Directive, Data Governance and Data Act and their possible impact on research. Publications Office of the European Union 2022. <https://data.europa.eu/doi/10.2777/71619>.

5 Hobart, Mark: „The Business Case for Automating Data Management“, in: Journal of ICT Standardization 2020, S. 199-216. <https://doi.org/10.13052/jicts2245-800X.832>.

6 Wilkinson, Mark D., et al.: „The FAIR Guiding Principles for Scientific Data Management and Stewardship“, in: Scientific Data 3(1) (2016). <https://doi.org/10.1038/sdata.2016.18>; Murray-Rust, P.: „Open Data in Science“, in: Nature Precedings (2008). <https://doi.org/10.1038/npre.2008.1526.1>

werden in allen Phasen des Forschungsprozesses angeboten, einschließlich der Datenveröffentlichung. Am MfN ist das Wissenschaftsdatenmanagement (WSDM) Team für den Workflow der Datenveröffentlichungen verantwortlich. Der von der Einrichtung eingerichtete Prozess der Datenveröffentlichung folgt den FAIR-Prinzipien. DOIs (Digital Object Identifiers) werden registriert, um die Auffindbarkeit, Zugänglichkeit und Wiederverwendung von Daten zu verbessern. Aufgrund der erheblichen Zunahme der in den letzten Jahren eingegangenen DOI-Anträge war es notwendig, die Prozesse für die DOI-Registrierung zu überdenken, um die Ziele der Effizienz und der Nachhaltigkeit zu erreichen. Der Beitrag diskutiert die Webanwendung, die das WSDM-Team entwickelt hat, um den DOI-Registrierungsprozess beim Anbieter DataCite zu automatisieren. In dem Beitrag wird zunächst die Bedeutung von DOIs für die Veröffentlichung von Forschungsergebnissen beschrieben. Anschließend wird die am MfN entwickelte Webanwendung zur Automatisierung des DOI-Registrierungsprozesses vorgestellt. Schließlich wird auf die Herausforderungen und Chancen hingewiesen, die sich aus der Entwicklung effizienter und nachhaltiger Datendienste ergeben.

2 DOI-Registrierung und ihre Bedeutung

Das DOI-System wurde 1998 von der Internationalen DOI Foundation gegründet.⁷ DOIs sind alphanumerische Zeichenfolgen, die als Persistent Identifier (PID) funktionieren. Das heißt, sie identifizieren die Objekte, auf die sie sich beziehen, immer eindeutig, sie führen immer zum richtigen Ort, an dem das digitale Objekt verfügbar ist, und sie stellen immer die Metadaten zur Verfügung, die dem Objekt zugeordnet wurden, als der DOI registriert wurde. Das DOI-System ermöglicht den Aufbau automatisierter Dienste, die einen Austausch von Forschungsdaten zwischen Produzenten und Konsumenten ermöglichen. Ein klarer Fall ist das Abrufen veröffentlichter Artikel über die von den Verlagen vergebenen DOIs.

Während sich die Informationen über ein Objekt im Laufe der Zeit ändern können, ändert sich der DOI nicht, und es ist möglich, die aktuellste Version der beschreibenden Metadaten über die von der DOI-Registrierungsagentur bereitgestellte Infrastruktur abzurufen. Das Digital Object Identifier System ist eine ISO-Norm (die aktuellste Version ist ISO 26324:2022), die international geregelt ist, um die

Interoperabilität zwischen DOIs zu gewährleisten, die bei verschiedenen Organisationen registriert sind.⁸

Bibliotheken und Forschungsdatenmanagement-Teams von Universitäten und Forschungseinrichtungen beteiligen sich immer mehr an DOI-Diensten. Sie registrieren routinemäßig DOIs für die Forschungsergebnisse ihrer Einrichtungen. Gleichzeitig profitieren sie von Diensten wie der Verfügbarkeit von durchsuchbaren Metadaten, die von den DOI-Registrierungsagenturen zur Verfügung gestellt werden. In Deutschland ist eine zunehmende Beteiligung wissenschaftlicher Bibliotheken und Forschungsdatenmanagement-Teams am DOI-Service zu beobachten.⁹ In den letzten Jahren hatten diese Einrichtungen einen steilen Anstieg der Zahl der eingegangenen DOI-Anfragen zu verzeichnen.¹⁰ Dies ist auf eine Reihe von Gründen zurückzuführen. Einerseits sind sich die Forscherinnen und Forscher mehr und mehr des Wertes von DOIs für die Verbesserung der Zitierfähigkeit wissenschaftlicher Arbeiten bewusst. Andererseits erkennen Förderorganisationen wie die Deutsche Forschungsgemeinschaft (DFG) den Wert dauerhafter Identifikatoren wie DOIs bei der Veröffentlichung wissenschaftlicher Datensätze, um optimale Verfahren für die Verwaltung von Forschungsdaten zu gewährleisten.¹¹

Das MfN registriert DOIs bei der weltweit tätigen gemeinnützigen Organisation DataCite.¹² Das Museum nutzt die Dienste von DataCite über eine Kooperationsvereinbarung mit der Deutschen Zentralbibliothek für Medizin (ZB MED), die ein offizielles Mitglied von DataCite ist und ein Konsortium deutscher Institutionen vertritt, die sich mit der Forschung in den Lebenswissenschaften beschäftigen.

DataCite hat im Jahr 2020 über vier Millionen DOIs registriert – im Jahr 2011 waren es nur etwas mehr als einhundertfünfzigtausend – und der Wachstumstrend ist ungebrochen, wie die Statistiken zeigen.¹³ Auch die Zahl der vom MfN registrierten DOIs ist im Laufe der Jahre erheblich gestiegen. Die Registrierungen im Jahr 2022 haben sich im Vergleich zu den Gesamtregistrierungen des Vorjahres mehr als verdoppelt (Abb. 1).

Die Registrierung eines DOI bei DataCite für MfN-Forschungsergebnisse erfolgt in mehreren Schritten und ist größtenteils ein manueller Prozess. Der manuelle Prozess

7 <https://www.doi.org> [15. Mai 2023].

8 <https://www.iso.org/obp/ui/#iso:std:iso:26324:ed-2:v1:en> [15. Mai 2023].

9 <https://datacite.org> [15. Mai 2023]. DataCite bietet seine Registrierungsdienste für deutsche wissenschaftliche Einrichtungen kostenlos an.

10 Liu, Jia: „Digital Object Identifier (DOI) and DOI Services: An Overview“. In: Libri 71 (4) (2021) S. 349-60. <https://doi.org/10.1515/libri-2020-0018>.

11 DFG-Positionspapier: „Wissenschaftliches Publizieren als Grundlage und Gestaltungsfeld der Wissenschaftsbewertung. Herausforderungen und Handlungsfelder“ (2022) https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/publikationswesen/positionspapier_publikationswesen.pdf.

12 <https://datacite.org>.

13 <https://datacite.org/documents/DataCite%202020%20Annual%20Report.pdf>, <https://stats.datacite.org> [15. Mai 2023].

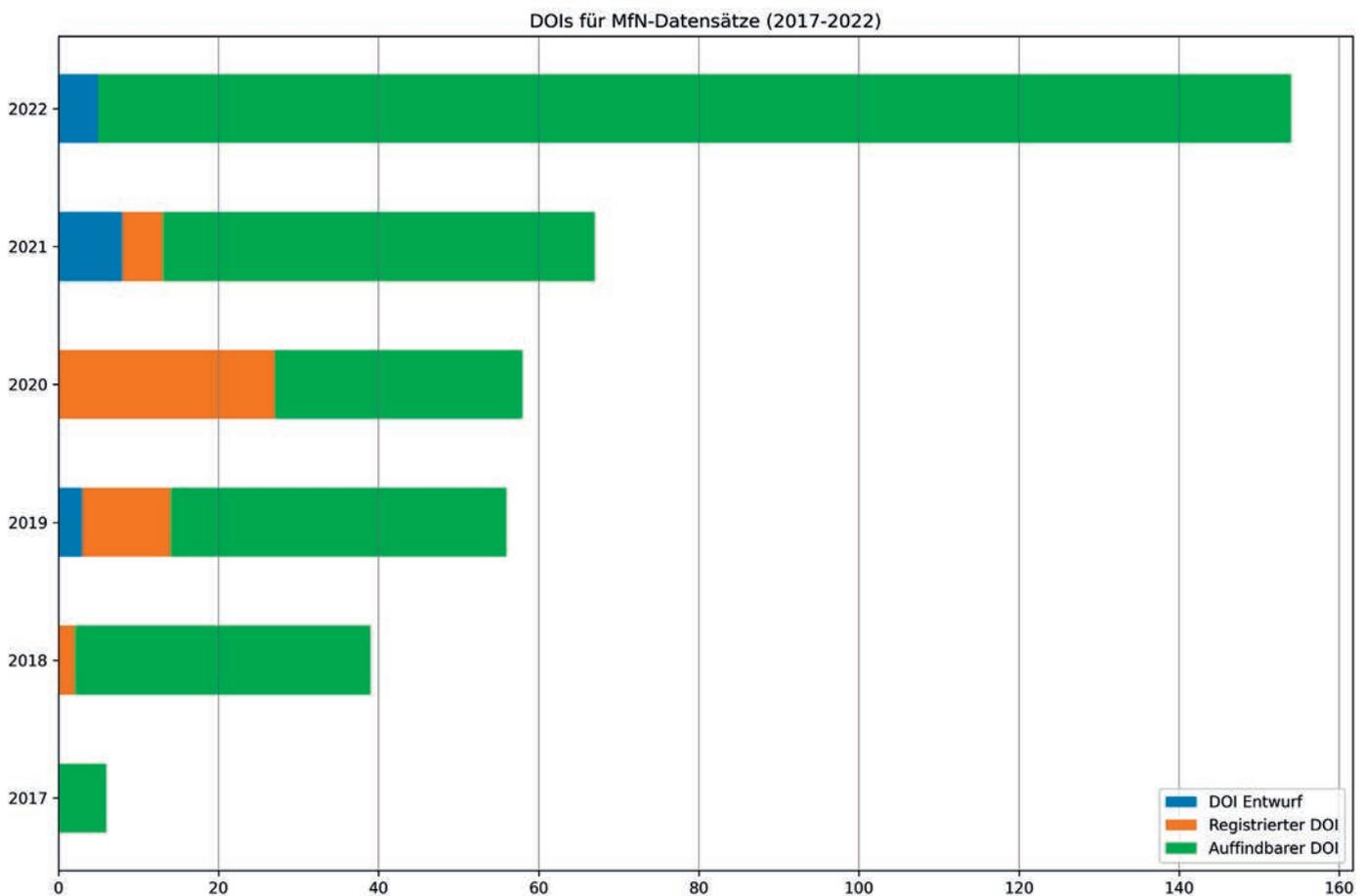


Abbildung 1: DOIs für MfN-Datensätze (2017–2022). Mit DataCite können DOIs in drei möglichen Status erstellt werden: Entwurf, registriert, auffindbar.

der DOI-Registrierung ist zeitaufwändig, da der DOI-Workflow mehrere IT-Systeme des MfN und die vom DOI-Anbieter angebotene Benutzeroberfläche einbezieht. Um die Zunahme der DOI-Anfragen zu bewältigen, hat das WSDM-Team des MfN beschlossen, den derzeitigen Arbeitsablauf zu überdenken und Lösungen zu suchen, die zumindest einige der Schritte im DOI-Erstellungsprozess automatisieren. Die Verfügbarkeit einer gut dokumentierten und zuverlässigen API (Application Programming Interface) für DataCite legte nahe, den Automatisierungsprozess des MfN-DOI-Workflows mit der Interaktion mit dem DataCite-Dienst zu beginnen. Die DOI-Erstellung sowie die Registrierung und Aktualisierung von Metadaten sind Aufgaben, die leicht über die DataCite-API automatisiert werden können und nicht über die DataCite-Benutzeroberfläche durchgeführt werden müssen.

3 DOI-Registrierung am MfN effizient und nachhaltig gestalten

3.1 DOI-Workflow

Am MfN beginnt der DOI-Registrierungsprozess mit der Erstellung eines DOI-Antrags durch ein Mitglied der Einrichtung (Abb. 2). Der Antrag wird durch Ausfüllen eines

speziellen Webformulars im internen Umfragetool gestellt. Das Formular geht beim WSDM-Team ein, das den Antrag manuell überprüft. Wenn der Antrag fehlerfrei ist und angenommen wird, erstellt der/die Datenmanager/Datenmanagerin einen DOI-Entwurf mit Hilfe der Weboberfläche DataCite Fabrica.¹⁴ Der DOI trägt das Präfix 10.7479, das ZB MED dem MfN als Konsortiumsmitglied zugeordnet hat.

Der DOI-Entwurf wird dann dem ursprünglich eingereichten Erhebungsformular hinzugefügt und die Formulardaten werden in die MfN-Metadaten-Datenbank, eine MySQL-Datenbank, übertragen. Der Datentransfer löst auch die Erstellung der Landing Page aus, auf die der DOI weiterleitet und auf der der Datensatz verfügbar sein wird. Um den Datensatz verfügbar zu machen, müssen die Dateien (z.B. CSV-Dateien, Mediendateien usw.) in das MfN-System zur Verwaltung digitaler Ressourcen (Digital Asset Management System) hochgeladen werden. Jede Ressource muss mit dem entsprechenden Datensatz in der MfN-Metadaten-Datenbank verknüpft werden. Sobald dieser Schritt abgeschlossen ist, wird der Datensatz auf der für den DOI erzeugten Landing Page dargestellt. Die Landing Page ermöglicht es auch die Metadaten im DataCite XML-Schema zu exportieren, so dass eine ex-

¹⁴ <https://doi.datacite.org> [15. Mai 2023].



Universität Marburg



TU München



Universitätsbibliothek
Salzburg



Berlin-Brandenburg
International School

zambelli

EINFACH MACHEN. AUS METALL.

Zambelli Bibliotheken Lernen und Wohlfühlen

Die Zambelli Bibliothekseinrichtungen begleiten wissenschaftliche und öffentliche Bibliotheken, die sich mit neuen Gegebenheiten auseinandersetzen und sich weiterentwickeln wollen. Wir helfen Ihnen Ihre Bibliothek so auszustatten, dass attraktive und funktionale Lernräume entstehen. Dabei können Sie sich auf in Sicherheit und Funktion bewährte Einrichtungs-lösungen verlassen.

Gemeinsam schaffen wir gestalterisch-kreative Raumkonzepte.

Nehmen Sie mit uns Kontakt auf!
regalsysteme@zambelli.com

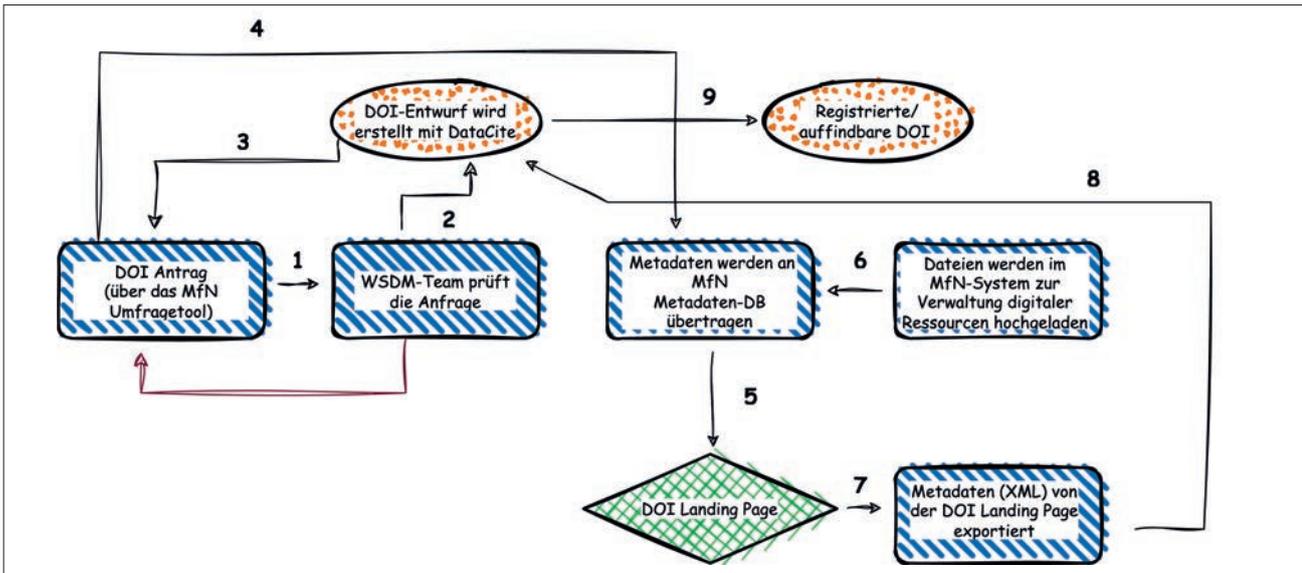
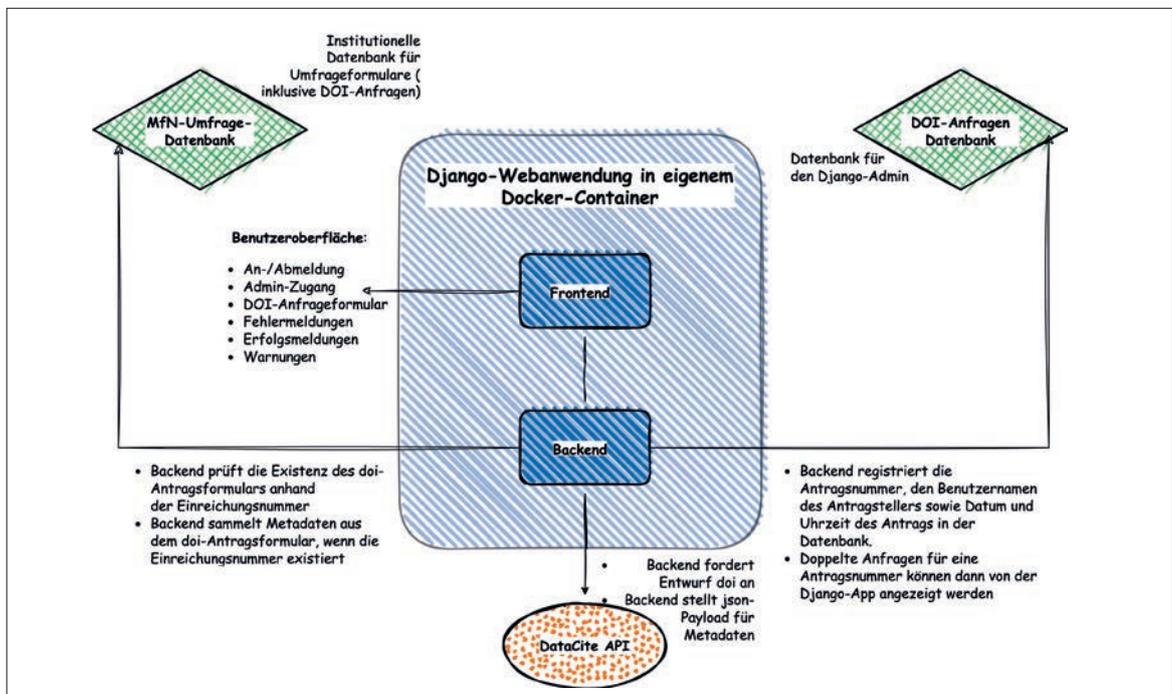


Abbildung 2: DOI-Arbeitsablauf im MfN

Abbildung 3:
Ablauf der
Webanwendung

portierte XML-Datei dazu verwendet wird, die Metadaten über die DataCite-Weboberfläche zum DOI-Entwurf hinzuzufügen. Sobald die Metadaten hinzugefügt wurden, kann der Status des DOI manuell auf „registriert“ oder „auffindbar“ durch den/die Datenmanager/Datenmanagerin geändert werden.

3.2 Webanwendung entwickelt, um die DOI-Registrierung zu automatisieren

Abb. 3 zeigt eine schematische Darstellung der Django-Webanwendung zur automatischen Registrierung von DOIs mit DataCite-API. Das Web-Framework Django

wurde ausgewählt, weil es auf Python basiert. Es ermöglicht aufgrund seiner zahlreichen bereits existierenden Funktionen eine schnelle und sichere Entwicklung und erfüllt alle Front-End-Anforderungen („Logging“, „Messaging“ usw.) sowie die Back-End-Anforderungen (Datenbankverbindung, Verfügbarkeit von Webformularen usw.) für die Webanwendung, die die DataCite-API aufruft.¹⁵ Darüber hinaus ermöglichte es eine einfache Integration mit Docker, welches am MfN die Technologie der Wahl für die isolierte Ausführung von Anwendungen in virtuellen Software-Containern ist.¹⁶ Die Django-Benutzeroberfläche ermöglicht die An- und Abmeldung, den

15 <https://www.djangoproject.com> [15. Mai 2023]. Mele, Antonio/ Bob Belderbos: Django 4 By Example: Build Powerful and Reliable Python Web Applications from Scratch (4. Edition), Birmingham, UK: Packt Publishing 2022.

16 Kofler, Michael/ Öggl, Bernd: Docker: das Praxisbuch für Entwickler und DevOps-Teams, Rheinwerk 2021.



Effiziente Medienerwerbung leicht gemacht

Alles Gute zum
25-jährigen Jubiläum
von **b.i.t.online!**



Profitieren Sie von Verlagsunabhängigkeit und Transparenz.

Werden Sie mit uns zum dynamischen Dienstleister für Ihre Studierenden und Lehrenden. Wir unterstützen Sie bei der Erwerbung, Verwaltung und Nutzung von Medien aus aller Welt durch:

- + **verlagsübergreifende Beratung** durch ein spezialisiertes Team
- + **hohe Transparenz** bei der Erwerbung von digitalen Medien
- + **individuelle Print-to-Digital-Analysen**
- + **komfortable Tools** wie das Schweizer Connect
- + **zuverlässige Lieferantendatenimporte**
- + **kostenlose Webinare** rund um ProQuest Ebook Central oder die Verlagsplattformen
- + **exklusive Partnerschaften** beispielsweise zu Kanopy



Fragen? Einfach E-Mail an:
academic@schweitzer-online.de
www.schweitzer-online.de

schweitzer
Fachinformationen

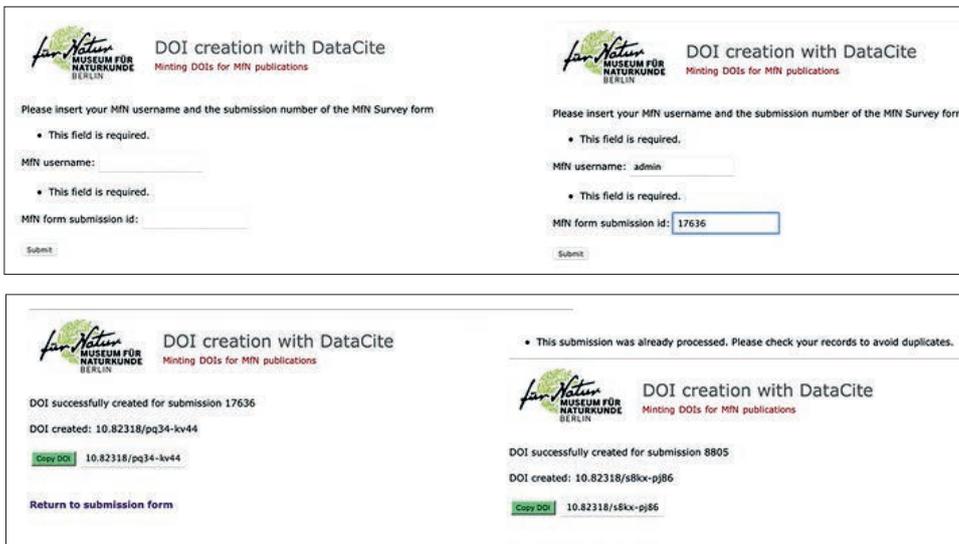


Abbildung 4a, oben: Benutzeroberfläche. Von links nach rechts: 1) Django-Formular zur Beantragung eines DOI; 2) ausgefülltes Django-Formular. Die Bilder wurden während der Code-Entwicklung aufgenommen.

Abbildung 4b: Benutzeroberfläche. Von links nach rechts: 1) Erfolgsmeldung für die DOI-Erstellung im Zusammenhang mit der Einreichung in Abbildung 4a (rechts); 2) Erfolgsmeldung für die DOI-Erstellung im Zusammenhang mit einer bereits in der Django-Admin-Datenbank vorhandenen Einreichungsnummer. Der Benutzer wird vor einem möglichen Duplikat gewarnt. Das DOI-Präfix bezieht sich auf das Test-Repository, das dem MfN von DataCite zur Verfügung gestellt wurde.

Zugriff auf die Django-Admin-Oberfläche und stellt den Benutzern das Webformular zum Ausfüllen bereit, um den DOI-Registrierungsprozess bei DataCite auszulösen. Über das Frontend wird der Benutzer über Fehler, Warnungen und erfolgreich abgeschlossene DOI-Registrierungen informiert. Das Backend der Webanwendung verwaltet alle Interaktionsprozesse mit DataCite und mit den MfN-Datenbanken, die für die DOI-Anforderung und das Funktionieren der Django-Anwendung notwendig sind. Es ist das Backend, das vor dem Auslösen des DOI-Registrierungsprozesses die MfN-Umfragedatenbank abfragt und überprüft, ob die Einreichungsnummer (ein interner ID-Code für das DOI-Antragsformular) korrekt ist. Wenn eine gültige Benutzeranfrage eingeht, sammelt das Backend die Metadaten für die DOI-Registrierung aus der MfN-Umfragedatenbank, registriert die erforderlichen Daten in der Datenbank, die mit dem Django-Admin verbunden ist, und ruft die API auf, um den DOI anzufordern und die Metadaten im JSON-Format bereitzustellen. Alle Aufrufe an die DataCite-API erfolgen über den oben erwähnten Python-Client. Der Client erleichtert die Arbeit mit der API erheblich, da er über alle erforderlichen Methoden zur Übertragung von Daten zum und vom DataCite-API verfügt. Darüber hinaus ermöglicht der Client den einfachen Wechsel zwischen der DataCite-Entwicklungs-API (zum Testen des Workflows) und der DataCite-Produktions-API.

3.3 Webanwendung Benutzeroberfläche

Das Frontend der Anwendung ermöglicht die Benutzerauthentifizierung und den Zugriff auf die Django-Admin-

Oberfläche, über die das WSDM-Team die bereits bearbeiteten Einreichungen überprüfen kann. Die Webanwendung ist nur für die Registrierung von DOIs bei DataCite vorgesehen, daher wird das Webformular, das den DOI-Registrierungsprozess auslöst, authentifizierten Benutzern sofort angezeigt. In dem Formular muss der/die Datenmanager/Datenmanagerin nur zwei Informationen angeben: den Benutzernamen und die Einreichungsnummer des DOI-Antragsformulars in der MfN-Datenbank (Abb. 4a). Die Einreichungsnummer ist die wichtigste Information für den gesamten Prozess. Mit Hilfe dieser Nummer ist es möglich, alle relevanten Metadaten aus dem Antragsformular zu sammeln und den Prozess zur Registrierung des DOI bei DataCite zu starten. Da die Einreichungsnummer für den Erfolg des Prozesses ent-

scheidend ist, wird ihr Vorhandensein in der MfN-Erhebungsdatenbank sofort geprüft, wenn das Formular eingereicht wird. Wird die Einreichungsnummer unter den DOI-Antragsformularen in der Datenbank nicht gefunden, wird eine Fehlermeldung erzeugt und dem Benutzer angezeigt. Sobald der Benutzer eine gültige Einreichungsnummer eingegeben hat, ruft die Webanwendung die DataCite-API auf, um einen DOI zu generieren und dem generierten DOI die Metadaten zuzuweisen, die aus dem in der MfN-Umfragedatenbank verfügbaren Antragsformular extrahiert wurden. Bei einer erfolgreichen DOI-Generierung wird dem Benutzer eine Erfolgsmeldung mit dem DOI und der Einreichungsnummer angezeigt (Abb. 4b(links)). Damit hat der/die Datenmanager/Datenmanagerin die volle Kontrolle über den Prozess. Darüber hinaus kann der/die Datenmanager/Datenmanagerin über eine Schaltfläche in der Benutzeroberfläche den DOI einfach mit einem Klick kopieren und in das ursprüngliche Erhebungsformular einfügen. Dies ist der letzte Schritt, bevor die Antragsdaten in die MfN-Metadatenbank übertragen werden können (siehe Abb. 2). Sollte der DOI für eine bereits vorhandene Einreichungsnummer generiert worden sein, erhält der/die Nutzer/Nutzerin eine Warnmeldung und kann auf mögliche Duplikate prüfen (Abb. 4b(rechts)). Auf der Erfolgsseite steht ein Link zur Verfügung, über den der/die Datenmanager/Datenmanagerin zum Einreichungsformular zurückkehren kann und den Prozess sofort erneut starten kann, um weitere DOIs zu erstellen. Mit Hilfe der entwickelten Django-Anwendung kann der Prozess von dem/der Datenmanager/Daten-

managerin durchgeführt werden, ohne die von DataCite bereitgestellte Weboberfläche zu nutzen. Damit entfallen zumindest einige Schritte in dem in Abb. 2 beschriebenen DOI-Workflow. Das Ergebnis ist nicht nur eine gewisse Zeitersparnis, sondern auch eine bessere Qualitätskontrolle, da die Metadaten bereits zum Zeitpunkt der Erstellung mit dem DOI verknüpft werden, wodurch das Risiko einer falschen Zuordnung von Metadaten und DOI ausgeschlossen wird.

4 Schlussbemerkungen: Herausforderungen und Chancen bei der Entwicklung effizienter und nachhaltiger Services

Forschungsdaten werden erst dann zu einer langfristigen Ressource für die Wissenschaft und die Gesellschaft, wenn sie von dem Labor, dem Feld oder der Gesellschaft, in dem sie erzeugt oder gesammelt wurden, zu den interessierten Nutzern gelangen können. An diesem Übergang sind mehrere Datenmanagementaktivitäten beteiligt, wie z.B. Datenkuratierung, Datenarchivierung und Datenpublikation. Für die Durchführung jeder einzelnen dieser Aktivitäten sind definierte Prozessabläufe erforderlich. Unabhängig davon, ob es sich um einen durchgängigen, in der Regel automatisierten Datenfluss (Pipeline) oder um Workflows, d.h. hauptsächlich um manuelle Statusänderungen der Daten, oder um eine Kombination aus beidem handelt: die Prozesse sind entscheidend für die Kernaufgabe des Forschungsdatenmanagements. Diese Prozesse müssen nachhaltig und effizient sein, um mit der Zunahme wissenschaftlicher Daten und der steigenden Nachfrage nach Daten Schritt halten zu können, trotz der potentiell begrenzten finanziellen und personalbezogenen Ressourcen des Forschungsdatenmanagements. Einerseits erfordern Nachhaltigkeit und Effizienz Investitionen in Technologien und Automatisierung, da die Zunahme der Datenoperationen nicht mit manuellen Verfahren erreicht werden kann und die neuen Systeme in bereits bestehende Dateninfrastrukturen integriert werden müssen. Andererseits erfordern sie auch eine enge Zusammenarbeit mit allen beteiligten Akteuren, damit der entwickelte Automatisierungsprozess auch wirklich ihren Bedürfnissen entspricht. Beim Entwurf des Codes für den Aufruf der DataCite-API wurden diese beiden Aspekte berücksichtigt. Die Webanwendung musste sich leicht in die bereits bestehende IT-Infrastruktur des MfN integrieren lassen. Gleichzeitig musste die Benutzeroberfläche für den/die Datenmanager/Datenmanagerin intuitiv und einfach zu benutzen sein. Die Ziele Nachhaltigkeit und Effizienz im Forschungsdatenmanagement erfordern nicht unbedingt große

Investitionen oder eine komplette Umstrukturierung der Datenprozesse. So profitierte die am MfN entwickelte Webanwendung zur Registrierung von DOIs von der am Museum bereits eingesetzten Microservices-Architektur und dem Einsatz von Frameworks wie Docker und Django, die über vorcodierte Funktionalitäten verfügen.¹⁷ Gleichzeitig bedeutete die Automatisierung des Prozesses nicht nur eine Zeitersparnis, sondern auch eine bessere Qualitätskontrolle. Wir sind daher der Meinung, dass sich die anfängliche Investition von Zeit und Ressourcen in die Entwicklung effizienter und nachhaltiger Prozesse im Forschungsdatenmanagement durch die daraus resultierenden Verbesserungen mehr als bezahlt macht und sehen die Chancen weitaus größer als die Herausforderungen, die der Entwicklungsprozess mit sich brachte. |



Giuditta Parolini

arbeitet als Datenwissenschaftlerin für das Wissenschaftsdatenmanagement-Team am Museum für Naturkunde Berlin. Sie hat einen Abschluss in Physik von der Universität Pavia und einen Dokortitel in Science, Technology, and Humanities von der Universität Bologna. Bevor sie zum MfN kam, arbeitete sie als Datenwissenschaftlerin für Nachhaltigkeit in einem deutschen Medienunternehmen und forschte im akademischen Bereich zu den Auswirkungen von Wetter und Klima auf die Landwirtschaft.

giuditta.parolini@mfn.berlin



Falko Glöckler

studierte Landschaftsökologie und Naturschutz an der Universität Greifswald. Er spezialisierte sich auf die Programmierung von wissenschaftsbegleitenden Websystemen und auf das Datenmanagement im Bereich der Ökologie und Biodiversitätsforschung. Ab dem Jahr 2011 arbeitete er als wissenschaftlicher Programmierer im Museum für Naturkunde Berlin – Leibniz-Institut für Evolutions- und Biodiversitätsforschung vor allem an Systemen und Datenbanken zur Begleitung von Digitalisierungs- und Datenveröffentlichungsprozessen. Dabei spielten die Themen der Datenqualitätssicherung und -standardisierung eine wesentliche Rolle. Der Hauptfokus seiner Forschungstätigkeit liegt im Feld der angewandten Biodiversitätsinformatik. Seit 2018 leitet er die Abteilung Wissenschaftsdatenmanagement am Museum für Naturkunde Berlin, die er maßgeblich mit aufgebaut hat.

Museum für Naturkunde

Leibniz Institute for Evolution and Biodiversity Science

Invalidenstr. 43, 10115 Berlin, Germany

falko.gloeckler@mfn.berlin

17 Newman, Sam: Vom Monolithen zu Microservices. Patterns, um bestehende Systeme Schritt für Schritt umzugestalten, O'Reilly 2020.