Langzeitverfügbarkeit von Forschungsdaten stellt eine große Herausforderung für Infrastruktureinrichtungen dar. Über den gesamten Entstehungsprozess der Daten kommt es immer wieder zu Grundsatzentscheidungen, deren Auswirkungen erst Jahre später erfasst werden können. Im Projekt LaVe (Langzeitverfügbarkeit von Forschungsdaten) wird seit 2022 an unterstützenden Materialien und Diensten gearbeitet, die darauf abzielen, einzelne Schritte auf dem komplexen Weg zur Langzeitverfügbarkeit zu verbessern. Als, im Vergleich zu anderen Initiativen, kleines Projekt müssen hier die Interventionspunkte so gewählt werden, dass mit geringem Ressourceneinsatz eine Verbesserung der Qualität der Forschungsdaten selbst oder der Workflows für die Langzeitverfügbarkeit erreicht werden kann. Die Auswahl der vor diesem Hintergrund gewählten Projektziele wird in diesem Aufsatz dargestellt und der aktuelle Arbeitsstand vorgestellt.

Ensuring the long-term availability of research data is a major challenge for infrastructure organizations. Throughout the entire data life cycle, basic decisions have to be made, the full impact of which can only be assessed years later. Since 2022, the LaVe (Langzeitverfügbarkeit von Forschungsdaten – Long-term availability of research data) project has focused on supporting materials and services aimed at improving individual steps in the complex process towards long-term availability. As a relatively small project compared to other initiatives, the individual interventions must be selected in such a way that any improvements in the quality of the research data itself or the workflows for long-term availability can be achieved with minimum use of resources. This article describes how the project goals were selected and presents the current status of the work.

ALEXANDER BERG-WEISS, JÜRGEN ROHRWILD, MARTIN SPENGER

Daten, die bleiben

Das LaVe-Projekt als Modell für nachhaltige Forschung

Einleitung

Bei der Langzeitverfügbarkeit (LZV) von Forschungsdaten gibt es zwei Dimensionen der Komplexität, die bei den meisten anderen Daten nicht vorhanden sind. Erstens sind Forschungsdaten an sich sehr heterogen. Sie können von strukturierten und kurzen textuellen Fragmenten, wie sie beispielsweise im Langzeitprojekt VerbaAlpina[1] entstehen, bis hin zu binären Daten in proprietären Formaten im Petabyte-Bereich wie in datenintensiven Vorhaben in der Physik alles umfassen. Zweitens sind die Use Cases für die aufbewahrten Daten hochgradig individuell. In manchen Bereichen sollen die Daten vor allem die Reproduzierbarkeit der Forschungsergebnisse ermöglichen, in anderen beispielsweise als Grundlage weiterer Forschung dienen. Der einfachste Fall ist hier sicherlich derjenige, bei dem die Forschungsdaten vor allem deshalb archiviert werden müssen, weil es die Förder- oder Publikationsrichtlinien vorgeben.

Diese zusätzliche Komplexität führt dazu, dass mit Archivierung und langfristiger Verfügbarkeit von Forschungsdaten sehr hohe Kosten einhergehen, wenn man sie bestmöglich realisieren will. Die Institutionen, an denen die Forschungsdaten entstehen, müssen daher nicht selten Kompromisse eingehen und sind gezwungen, die Qualität in Erschließung, Aufbereitung etc. zu reduzieren. Die notwendige Abwägung, wo und in welchem Umfang Einsparungen an der Qualität vorzunehmen sind, muss datensatzspezifisch erfolgen und benötigt sowohl tiefe Kenntnisse der Daten als auch eine Einschätzung der zukünftigen disziplinspezifischen Arbeitsmethoden, die in der Regel nicht hinreichend vorhersehbar sind. Strebt man beispielsweise bestmögliche Reproduzierbarkeit an, ist es wichtig zu wissen, mit welcher Technik der Datensatz im Detail erhoben wurde. Man müsste nun sicherstellen, dass diese Technik auch in zehn, fünfzehn oder mehr Jahren verfügbar ist. Um dies zu ermöglichen, müsste unter Umständen eine eigene Umgebung zur Emulation der Technik angelegt werden, was wiederum ständiger Wartung bedürfte. Schnell kann sich so der Personalbedarf für die Sicherstellung der Reproduzierbarkeit eines Forschungsdatums auf einen signifikanten Anteil einer Vollzeitstelle belaufen. Häufig ist dies weder finanzierbar noch sinnvoll, denn ob später überhaupt jemand anstrebt, diesen Datensatz vollständig zu reproduzieren, ist zum Zeitpunkt der Archivierung nicht bekannt und kann schwer prognostiziert werden. Man könnte die Frage der Reproduzierbarkeit auch gänzlich in die Zukunft delegieren, indem man die Ansicht vertritt, dass ein reines Aufbewahren der Bits ausreichend ist, da die Zukunft Techniken (»Stichwort KI«) hervor-

bringen könnte, die aus den Daten ableiten, wie diese reproduziert oder automatisch erschlossen werden können. Um zu bewerten, ob und wann dies eintreten könnte, wäre wieder Zukunftswissen notwendig.

Vor diesem Hintergrund ist auch das hier vorgestellte Projekt »Langzeitverfügbarkeit von Forschungsdaten« (LaVe) zu sehen, dessen Ziel es nicht ist, die höchstmögliche Qualität bei der langfristigen Verfügbarkeit von Daten zu erreichen. Vielmehr sollen Prozesse, Best Practices und Infrastrukturen erarbeitet werden, die bei einem beschränkten Einsatz von Ressourcen einen möglichst großen Beitrag für eine hohe Qualität leisten. LaVe ist ein Modellprojekt, das vom Bayerischen Staatsministerium für Wissenschaft und Kunst im Rahmen des Digitalen Campus Bayern[2] seit 2022 gefördert wird. Es baut auf einer längeren Kooperation der Universitätsbibliotheken der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) und der Ludwig-Maximilians-Universität München (LMU) auf. Im Folgenden werden der Projekthintergrund und das Umfeld, die Ziele und die diesen zugrundeliegende Motivation sowie erste Ergebnisse skizziert. Abschließend erfolgt ein Ausblick auf weitere Entwicklungen sowohl im Projekt selbst als auch im regionalen Kontext.

Projekthintergrund

Alle Forschungseinrichtungen stehen vor der Herausforderung, zumindest Teile ihrer Forschungsdaten langfristig bereitzustellen. Die grundsätzliche Frage ist immer, welche Stelle sich dem vielschichtigen Thema LZV annehmen soll oder muss. Unter Umständen müssen keine eigenen dedizierten Dienste aufgebaut werden, wenn institutionsübergreifende Infrastruktureinrichtungen ihre Dienste und Expertise auch Dritten zur Verfügung stellen. Eine Übersicht über eventuelle Vorarbeiten und Vernetzungsmöglichkeiten ist daher nicht nur zur Einordnung eines Projektes wie LaVe hilfreich. Im Folgenden werden die nationalen und regionalen Bestrebungen aufgezeigt.

Nationale Bestrebungen

Mit der Nationalen Forschungsdateninfrastruktur (NFDI) ist seit 2020 ein zentraler Player auf nationaler Ebene aktiv. Die NFDI hat die Zielsetzung, wissenschaftsgetrieben den Umgang mit Forschungsdaten in den Fachdisziplinen zu verbessern, wozu auch das Thema LZV zählt. Die Verankerung des Themas findet in der NFDI auf verschiedenen Ebenen statt, nämlich der Konsortialebene, der Sektionsebene und im Rahmen der gemeinsamen Initiative Base4NFDI.

Die 26 NFDI-Konsortien sind auf bestimmte Disziplinen und Forschungsgegenstände ausgerichtet. Die eigenständigen Konsortien treiben primär den Aufbau der NFDI voran. Hier werden konkrete, auf die Bedarfe der Fächer ausgerichtete Services und Unterstützungswerkzeuge für das Datenmanagement erarbeitet. So ist beispielsweise NFDI4Objects auf Forschung zu und mit materiellen Artefakten der Menschheitsgeschichte ausgerichtet – eine Forschungsrichtung, der langfristige Bewahrung von Forschungsobjekten und Ergebnissen inhärent ist.

Querschnittsthemen, wie etwa Datenkompetenzvermittlung, rechtliche Aspekte oder gemeinsame Infrastrukturen, die konsortialübergreifend bearbeitet werden, sind in sogenannten Sektionen des NFDI-Vereins organisiert. Sektionen bieten darüber hinaus auch Interessierten, die nicht direkt in einem Konsortium aktiv sind, die Möglichkeit, zur NFDI beizutragen.

In Base4NFDI werden sogenannte Basisdienste erarbeitet, die in praktisch allen Konsortien von Bedeutung sein werden oder die als Grundlage für darauf aufbauende spezifische Angebote der Konsortien dienen. Hierunter fallen beispielsweise Datenmanagementpläne (DMP4NFDI), Identitäts- und Zugangsmanagement (IAM4NFDI), Wissensgraphen und Terminologien (TS-4NFDI), aber auch Dienstleistungen zu Persistenten Identifikatoren (PID4NFDI).[3]

Ebenfalls können durch die NFDI die nationalen Lösungen zum Forschungsdatenmanagement auch für einen Austausch mit internationalen Netzwerken vorbereitet werden, wozu vor allem die European Open Science Cloud (EOSC) oder die weltweit agierende Research Data Alliance (RDA) zählen. Der Austausch der nationalen und internationalen Initiativen ist vor allem für die Etablierung von Standards und Best Practices ein zentrales Element.

Regionale Bestrebungen

Für das Forschungsdatenmanagement (FDM) spielen jedoch auch die Entwicklungen auf regionaler Ebene eine zentrale Rolle, wie sie in Deutschland beispielsweise durch die Landesinitiativen im Bereich FDM vorgenommen werden.[4] Die Situation im Bereich FDM ist auf bayerischer Ebene stark durch einen Projektcharakter geprägt, da es im Freistaat noch keine Landesinitiativenstruktur gibt. Hervorzuheben sind hier die vom Bayerischen Staatsministerium für Wissenschaft und Kunst geförderten Projekte »eHumanities – interdisziplinär«[5] sowie »Digitale Langzeitverfügbarkeit für Wissenschaft und Kultur in Bayern«[6] (im Folgenden kurz »Digitale LZV«). Das LaVe-Projekt geht aus dem eHumanities-Projekt sowie thematischen Vernetzungen mit dem Projekt »Digitale LZV« hervor. LaVe knüpft an diese Vorarbeiten an, um die teilweise konzeptionellen und grundlegenden Infrastrukturdienste um eine Langzeitarchivierungskomponente zu erweitern.

Das bereits abgeschlossene Projekt »eHumanites – interdisziplinär« befasste sich ausgehend von Bedarfen aus Forschungsprojekten der digitalen Geisteswissenschaften in sechs Arbeitspaketen, die entlang des gesamten Forschungsprozesses angesiedelt wurden, mit dem Datenmanagement. Schwerpunkte waren hierbei Metadatenan-

reicherung von Forschungsdaten, Bereitstellung und Entwicklung fachlicher Datenmanagementpläne, Erstellung von Schulungs- und E-Learning-Material, Publikation und Discovery von Forschungsdaten, Workflows zum FDM, Integration von Diensten in eine zentrale Service-Plattform sowie Öffentlichkeitsarbeit.[7] Als besonders erfolgreich stellten sich der DataCite Best Practice Guide[8] und der darauf basierende Metadatengenerator[9] heraus, die auch international nachgenutzt werden.

Das aktuell in der zweiten Förderphase laufende Projekt »Digitale LZV« befasst sich direkt mit der langfristigen Archivierung und Bereitstellung von Daten. Hierfür werden sowohl Workflows als auch technische Lösungen für unterschiedliche Objektklassen erarbeitet. Die konkrete Umsetzung umfasst diverse Use Cases – von Forschungsdaten und Digitalisaten bis hin zu Open-Access-Zeitschriften samt zugehörigem Publikationsprozess. Dies wird durch im Projekt entworfene Mustervereinbarungen und -verträge für die Datenübergabe, Policies und Checklisten zur Vorbereitung der Daten flankiert.[10]

Da »eHumanities – interdisziplinär« und »Digitale LZV« parallel gefördert wurden, war es möglich, sich während der Planungen für LaVe gezielt auszutauschen. So konnten ein einheitliches Vorgehen sowie die Vorbereitung von Kooperationsthemen vereinbart werden, die Ergebnisse aus den bestehenden Projekten miteinander verbinden. LaVe zielt damit als Projekt sowohl auf eine Erweiterung der zuvor bearbeiteten Themen als auch auf eine Vertiefung von technischen Komponenten ab, die zuvor nur modellhaft erarbeitet wurden.

Von Vernetzung profitieren auch LZV-Initiativen und -Projekte mit ursprünglich regionalem Fokus. Ein zentraler Schritt war der Zusammenschluss der LZV-Bestrebungen der Bundesländer in KoopLZV, der »im Juni 2022 als trilateraler Austausch zwischen den Initiativen in Bayern, Hessen und Nordrhein-Westfalen«[11] begann und seitdem um vier weitere Mitglieder erweitert wurde. »Ziel der Zusammenarbeit ist es, Erfahrungen zu sammeln, Synergien zu nutzen und sich gegenseitig bei der Entwicklung effizienter Abläufe zu unterstützen.«[12]

Auch wenn es eine Vielzahl von Akteuren im Forschungsdatenmanagement gibt, die oft bestrebt sind, durch gegenseitige Vernetzung und Nachnutzung von Ergebnissen und Materialien Synergieeffekte zu schaffen, bleibt die Landschaft trotzdem heterogen. Doppelstrukturen und unterschiedliche Bewertung von Bedarfen sind deshalb nicht zu vermeiden. Dies kann aber auch daran liegen, dass es für einige Fragestellungen keine offensichtlichen Antworten gibt und so mehrere Wege erprobt werden müssen. Hinzu kommen die einleitend beschriebene Komplexität und Diversität der Forschungsdaten, die durch unterschiedliche, berechtigte Anforderungen der Fachdisziplinen noch erhöht werden.

Trotz oder vielleicht gerade wegen der Dynamik im transinstitutionellen Umfeld sind Services an der eigenen Institution unabdingbar. Dies hat drei Gründe: Erstens sind die Dienstwege an der eigenen Institution für die Forschenden kürzer. Zweitens sind gerade die überregionalen Dienstleistungen standardisiert. Dieser Standard muss bekannt sein, und man muss die Fähigkeit besitzen, den Standard entsprechend umzusetzen. Beides ist bei Forschungsdaten komplex und zeitraubend. Drittens stellt sich bei der langfristigen Verfügbarkeit oftmals die Frage, wie lange es einen genutzten Service geben wird und wer ihn betreibt. Insbesondere in Projektkontexten konnten in den letzten Jahren einige Vorhaben unter anderem aufgrund von Ressourceneinsparungen beziehungsweise Personalengpässen nicht fortgesetzt werden. Bei externen Dienstleistern ist deshalb schwer einzuschätzen, ob ein zuverlässiger und langfristiger Betrieb gewährleistet ist.

Motivation, Vorgehen und erste Ergebnisse

Langfristige Verfügbarkeit von Daten und Dokumenten, auf denen Forschungsergebnisse basieren oder die Ergebnisse eines Forschungsprojektes sind, lässt sich infrastrukturseitig nur effektiv durch unterstützende Maßnahmen während des gesamten Forschungsprozesses fördern. Einzelne Dienste, wie etwa die Bereitstellung langfristig betreuter, redundanter Speichersysteme, können immer nur einen Beitrag zu besserer LZV leisten, diese aber nicht allein garantieren.

Im Projekt LaVe wurden daher drei Teilbereiche fokussiert, die jeweils eine Komponente nachhaltiger LZV darstellen. Diese sind:

1. Best Practices und Awareness für Daten und deren Organisation in der aktiven Projektphase,
2. niederschwellige, stabile Referenzierung von digitalen Objekten mittels lokaler PID-Dienste,
3. Anbindung lokaler Systeme und Repositorien an genuine Langzeitarchive.

Jeder Themenbereich ist für sich genommen immer noch zu groß, um im Rahmen eines kleinen Projektes sinnvoll bearbeitet zu werden. Daher wurde entsprechend der Vorerfahrung der Projektpartner und mit Blick auf Use Cases und Anforderungen der Forschenden jeweils ein Aspekt identifiziert, in dem mit vertretbarem Aufwand eine merkliche Verbesserung erreicht werden kann.

Die Motivation, das Vorgehen im Projekt und erste Ergebnisse werden im Folgenden dargestellt.

Best Practices und Awareness für Daten und deren Organisation in der aktiven Projektphase

Der im Vorgängerprojekt »eHumanities – interdisziplinär« entwickelte DataCite Best Practice Guide[13] und die darauf aufbauenden Workflows[14] zu Archivierung und Publikation von Datensätzen wurden bewusst mit verschiedenen Beispielen aus der Praxis angereichert. In LaVe sind daher analog Empfehlungen zu Datei- und Datenformaten für die Langzeitverfügbarkeit entwickelt worden.

Dabei muss klar sein, dass es bereits viele ausgezeichnete, von Expert*innen kuratierte Empfehlungen zu Datenformaten für die Langzeitverfügbarkeit gibt. Beispiele hierfür sind:

- der Katalog archivischer Dateiformate der Koordinierungsstelle für dauerhafte Archivierung elektronischer Unterlagen,[15]
- die Empfehlungen »Archivtaugliche Dateiformate« der ETH Zürich,[16]
- empfohlene Dateiformate für Forschungsdaten zur Langzeitarchivierung von OSTDATA[17]
- oder die »Best Practices for 3D Data Preservation.[18]

Eine weitere Handreichung zu diesem Aspekt bietet also keinen echten Mehrwert. Daher sind, ähnlich dem Vorgehen bei der Erstellung des DataCite Best Practice Guides, Empfehlungen anhand von Use Cases erarbeitet worden, die dabei denkbare Vorgehensweisen, Probleme und Fragestellungen für Forschende aufzeigen.

Für die Auswahl der Use Cases bot sich die im Abschnitt »Regionale Bestrebungen« angesprochene enge Kooperation mit dem Projekt »Digitale LZV« an. Das Projekt konnte bereits umfangreiche Erfahrung mit der Übernahme von Daten aus Forschung und aus Digitalisierungsprojekten in Langzeitarchive aufbauen. Unter anderem wurden die Übernahme und Langzeitverfügbarmachung heterogener Daten etwa aus elektronischen Laborbüchern und Projektdatenbanken erprobt. Die Kooperation erlaubt nicht nur einen Rückgriff auf einen breiten Fundus von Fällen, sondern ermöglicht es, die Ergebnisse direkt in die bereits über die Webseite *lzv-bayern.de* bereitgestellten Materialien zur LZV einzubauen.

Vorangestellt stand die Grundsatzentscheidung,die Empfehlungen und Hilfestellungen auf erwartete Nachnutzungsszenarien und Möglichkeiten der Zielgruppe, der Datenproduzenten an Hochschulen, auszurichten. Dies folgt der eingangs beschriebenen Philosophie, keine optimalen Lösungen anzustreben, sondern mit leistbarem Aufwand eine merkliche Verbesserung der Datenqualität zu erreichen. Leistbar bezieht sich sowohl auf den für die Umsetzung der Empfehlungen notwendigen Support durch Infrastruktureinrichtungen als auch auf den Aufwand und die Zeit, die Forschende selbst investieren müssen. Pragmatischen Lösungen, die den Erhalt der für das von den Forschenden anvisierte Nachnutzungsszenario nötigen Informationen gewährleisten, werden daher im Projektkontext der Vorzug gegeben.

Die Erarbeitung entsprechend kontextabhängiger Empfehlungen ist jedoch aufgrund der Vielzahl denkbarer Szenarien sehr aufwendig. Um diese Schwierigkeit abzumildern, wurden die Hilfestellungen an in Beratungen häufig auftretende Fälle angelehnt. Die Fallbeispiele wurden dann anhand eines festen Musters beschrieben, das grob einem Pattern-Konzept[19] folgt.[20]

Die Use Cases sollen idealerweise immer enthalten:

- eine knappe Beschreibung des Falls,
- Kontextinformationen, beispielsweise Rahmenbedingungen, angestrebtes Nachnutzungsszenario, Datentyp oder Fachrichtung,
- ein empfohlenes Vorgehen mit knapper Angabe von Gründen,
- mit dem Vorgehen einhergehende Konsequenzen und Kosten. Hier sollten sowohl Aufwand, finanzielle Kosten und Personalressourcen zur Umsetzung der Lösung genannt werden als auch potenzielle Risken anderer Handlungsweisen, insbesondere dem Verzicht auf Maßnahmen,
- weiterführende Hinweise, beispielsweise Verweise auf Empfehlungen zu den einzelnen Datenformaten oder zu Anleitungen zur Datenkonvertierung.

Die Bereitstellung der Materialien erfolgt als Markdown-Dateien über das GitHub-Repositorium der Universitätsbibliothek der LMU München[21] und als PDF über Zenodo[22]. Die Use Cases werden zusätzlich in die Webseite lzv-bayern.de[23] eingebunden.

Ein Nachteil der kompakten Fallbeispiele ist, dass die einzelnen Hilfestellungen zwar in einfachen Situationen mit klaren Datentypen angewendet werden können, aber für komplexe, individuelle Bedarfe zu oberflächlich sind. Ein Beispiel hierfür sind (Projekt-)Webseiten, die ein breites Spektrum an Datentypen und regelmäßig auch interaktive Elemente enthalten. Sie stellen eine besondere Herausforderung in der FDM-Beratung für die Langzeitverfügbarkeit dar, da hier häufig erst mit Projektende Unterstützung gesucht wird. Daher wurde in Zusammenarbeit mit der Max Planck Digital Library (MPDL) eine ausführliche Orientierungshilfe[24] zur Webseiten-Archivierung erarbeitet. Diese richtet sich an Beratende, aber auch an Administratoren der Webseiten und Projektverantwortliche und stellt in gewissem Sinn einen Gegenentwurf zum Ansatz der Fallbeispiele dar.

Die Handreichung betrachtet detailreich einzelne Aspekte einer Webseite und der darüber bereitgestellten Daten. Von grundlegenden Formaten über technische Überlegungen bis hin zu Barrierefreiheit und Datenschutz werden anhand von Leitfragen die Interventionspunkte aufgezeigt, an denen rechtzeitig grundlegende Weichen für die spätere Archivierung gestellt werden müssen. Hieraus werden schließlich Checklisten für die Konzeption und Archivierung von Webseiten abgeleitet. Die Orientierungshilfe ist auch via GitHub und über *lzv-bayern.de* verfügbar.[25]

Lokale PID-Dienste

Langfristig stabile Links sind eine Grundvoraussetzung für den dauerhaften Zugang zu digitalen Objekten. Einfache http-URLs leiden an den Hauptproblemen »Broken links / Link rot« (die URL führt nicht mehr

zum digitalen Objekt) und »Content drift« (der Inhalt, zu dem die URL aufgelöst wird, hat sich geändert).[26]

Persistente Identifikatoren (PID) entkoppeln Identifikator und Speicheradresse digitaler Objekte, indem eine Zwischenebene eingeführt wird. Der PID selbst ist lediglich eine eindeutige alphanumerische Zeichenfolge. Die Zwischenebene stellt das administrative und technische Kernstück dar. Es ist die Verantwortung des Betreibers dieser Ebene, dass ein PID dauerhaft zuverlässig aufgelöst werden kann. Unter anderem dies unterscheidet PID von dem konkurrierenden Ansatz der Cool URI,[27] der deutlich stärker auf technischen Maßnahmen fußt. Das wohl bekannteste PID-System ist der Digital Object Identifier (DOI).

Wissenschaftliche Bibliotheken sind als auf Dauer angelegte, vertrauenswürdige Institutionen für den Betrieb der Zwischenebene, den PID-Resolvern, prädestiniert. Gleichzeitig werden im Bereich Forschungsdaten Use Cases an Infrastruktureinrichtungen herangetragen, die sich nicht problemlos mit DOI bedienen lassen, beispielsweise:

- Projekte mit einer sehr großen Anzahl von Datensätzen infolge granularer Erschließung,
- Datensätze, die nicht statisch, aber auch nicht sinnvoll versionierbar sind,
- Datensätze, die nicht dauerhaft aufgehoben werden sollen.

Ein Ziel des Projektes war daher die prototypische Implementierung eines lokalen PID-Resolvers, der leicht angepasst und somit an anderen Einrichtungen eingesetzt werden kann. Dabei bezieht sich »lokal« lediglich auf den Einsatzbereich, PIDs für die eigenen Einrichtungen zu erstellen und den Dienst vor Ort mit minimalen Anpassungen aufsetzen zu können. Die Identifikatoren sind weltweit eindeutig und der Dienst vermittelt zuverlässig zwischen PID und der aktuell zugehörigen URL.

Neben den oben genannten Use Cases bieten frühzeitig vergebene, stabile IDs auch generell Vorteile für Workflows zwischen und innerhalb von Infrastruktureinrichtungen. Eindeutige IDs begleiten einen Datensatz unter Umständen vom Moment der Erzeugung, bis hin zur Übergabe an beispielsweise eine Bibliothek, über Transformationen, Mapping-Schritte, Metadatenanreicherung, Veröffentlichung in einem lokalen Repositorium bis schließlich zum Einbringen in ein genuines Langzeitarchiv.[28] Diese IDs zu PIDs aufzuwerten ist im Sinne der FAIRness[29] konsequent und bietet so einen Mehrwert ohne großen Aufwand. Mit einer Vielzahl der momentan verfügbaren und entstehenden von externen Anbietern betriebenen PID-Lösungen ist dies jedoch nicht ohne Weiteres möglich.

Schließlich hat ein lokaler Dienst auch Hoheit über die darin verwalteten (Meta-)Daten. Damit ist weniger die gegebene Möglichkeit gemeint, die Pflichtinformationen, die für die PID-Vergabe nötig sind, eigenständig festzulegen. Vielmehr liegt die Verantwortung für die langfristige Verwaltung der PIDs bei der heimischen Einrichtung der Forschenden, die – solange die Forschungseinrichtung als solche existiert – für die notwendige langfristige Pflege bürgt. Bibliotheken haben hier einen *Track Record*, mit dem ein nicht zu unterschätzender Vertrauensvorschuss einhergeht, der anderen Anbietern in dem immer noch wachsenden PID-Ökosystem so nicht eingeräumt wird. Dies trägt dazu bei, die Akzeptanz einer lokalen PID-Lösung zu fördern – gerade in Forschungsbereichen, die traditionell ein großes Interesse an generationenübergreifender Verfügbarkeit des Forschungsoutputs haben.

Ausgehend von diesen grundsätzlichen Überlegungen wurden im LaVe-Projekt zunächst Minimalanforderungen an einen lokalen PID-Resolver festgelegt. Hier konnte auf eine im Rahmen des Studiengangs Digitales Datenmanagements (DDM) erstellte Masterarbeit von Vanessa Gabriel (*Persistent-Identifier-Systeme für Forschungsdaten – Eine analytische Vergleichsstudie mit Empfehlungen zur Implementierung an wissenschaftlichen Einrichtungen, 2022*) zurückgegriffen werden. Darin wurden technische und administrative Anforderungen an das System formuliert.

Auf technischer Ebene muss das System leichtgewichtig, einfach aufzusetzen, leicht anzupassen und sicher sein. Dies legt eine Umsetzung in einem Micro-Service-Ansatz nahe, in dem einzelne Komponenten unabhängig voneinander entwickelt und die Bauteile dann anschließend zu einer Gesamt-Applikation »PID-Resolver« zusammengesetzt werden. Dies erlaubt es, einzelne Elemente je nach Bedarf nachzunutzen und / oder auszutauschen. Insbesondere wird das Backend, in dem PID, Metadaten und Identitäten / Rechte verwaltet werden, vom Frontend, über das Nutzende einzelne PIDs auflösen und zu den zugeordneten digitalen Ressourcen geleiten werden, getrennt. So können unterschiedliche Einrichtungen Frontend und Backend separat betreiben oder alternativ lediglich das Backend nutzen, wenn beispielsweise die REST API des Backends direkt für die Kommunikation genutzt werden soll.

Das Backend setzt auf eine Spring-Boot-Applikation, die Anfragen an eine PostgreSQL-Datenbank durchführt und mittels einem nginx-basierten Reverse Proxy angesprochen wird. Das Identitätsmanagement wird von Keycloak übernommen. Für jede Komponente wird, dem Modularitätsprinzip folgend, ein eigener Docker Container erzeugt. Das Frontend basiert auf React und wird ebenfalls als eigener Container angelegt. Die Struktur des Frontends kann leicht über Anpassungen in den JSX-Dateien geändert werden. Abbildung 1 zeigt das grobe technische Schema nochmals grafisch. Aus Sicht des Backends ist das Frontend lediglich ein spezieller API-Nutzer. Das Frontend ermöglicht so Unbekannten ohne eigene API-Privilegien, eingeschränkte Anfragen an das Backend zu richten.

Die API selbst stellt nur grundlegende Funktionalitäten zur Verfügung: die Abfrage der zu einem PID gehörigen Metadaten, das Auslesen aller PIDs, das Anlegen neuer PIDs und das Anpassen der Metadaten zu einem PID. Auf administrativer Ebene waren die primär zu treffenden Entscheidungen die minimal erforderlichen Metadaten, die einen PID begleiten müssen, die Form des PID selbst und die Interaktionsoptionen, die Nutzer*innen via dem Frontend zur Verfügung gestellt werden.

Als minimal notwendige Pflichtfelder in der Datenbank wurden identifiziert:

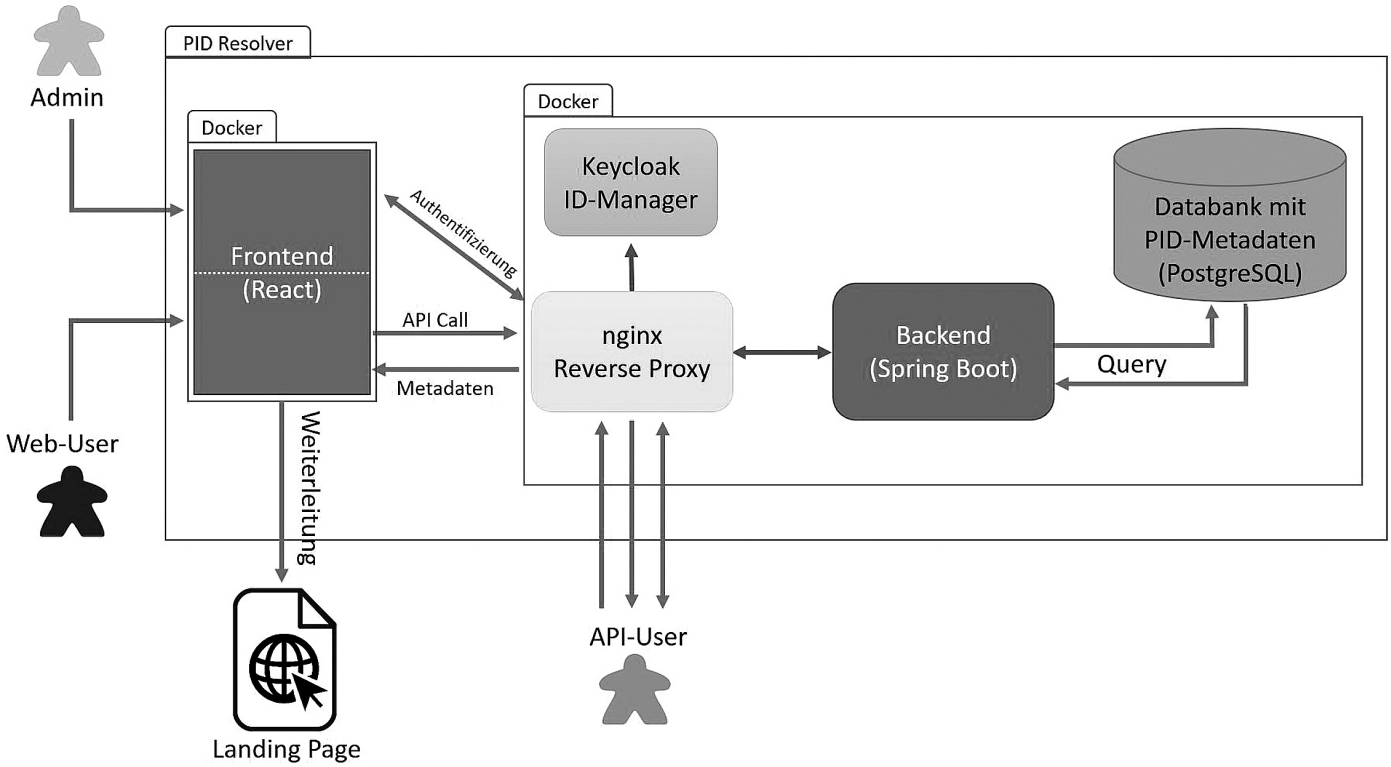
- der PID selbst,
- das Datum der Erzeugung des Datenbankeintrags,
- der Löschstatus der dahinterliegenden Ressource,
- der eindeutige Name des Systems, das die Ressource beinhaltet (Zielsystem),
- die interne ID der Ressource im Zielsystem,
- die URL der Landing Page, auf die der Resolver verweisen soll.

Optionale Metadatenfelder sind darüber hinaus die Angabe der Anzahl der Weiterleitungen auf eine Ressource, der Löschgrund und das Datum des letzten Metadaten-Updates. Die modulare Struktur des Systems erlaubt grundsätzlich beliebige Erweiterungen der erfassten Metadatenfelder, dies erfordert jedoch entsprechende Anpassungen in der API.

Der PID selbst sollte kurz und unmissverständlich sein. Insbesondere muss er von Nutzer*innen aus Print-Publikationen einfach abgetippt werden können. Daher wird eine alphanumerische Zeichenfolge der Form xxx-xxx – wobei x aus dem Zeichenbereich [a-z0-9] stammt – mit der URL des Resolvers als Präfix kombiniert. Ist beispielsweise die URL des Resolvers pid.lmu.de, dann ist pid.lmu.de/z45-p9r ein denkbarer PID. Mit dem gewählten Alphabet stehen etwa zwei Milliarden PIDs zur Verfügung, in denen kaum Homoglyphen[30] drohen.

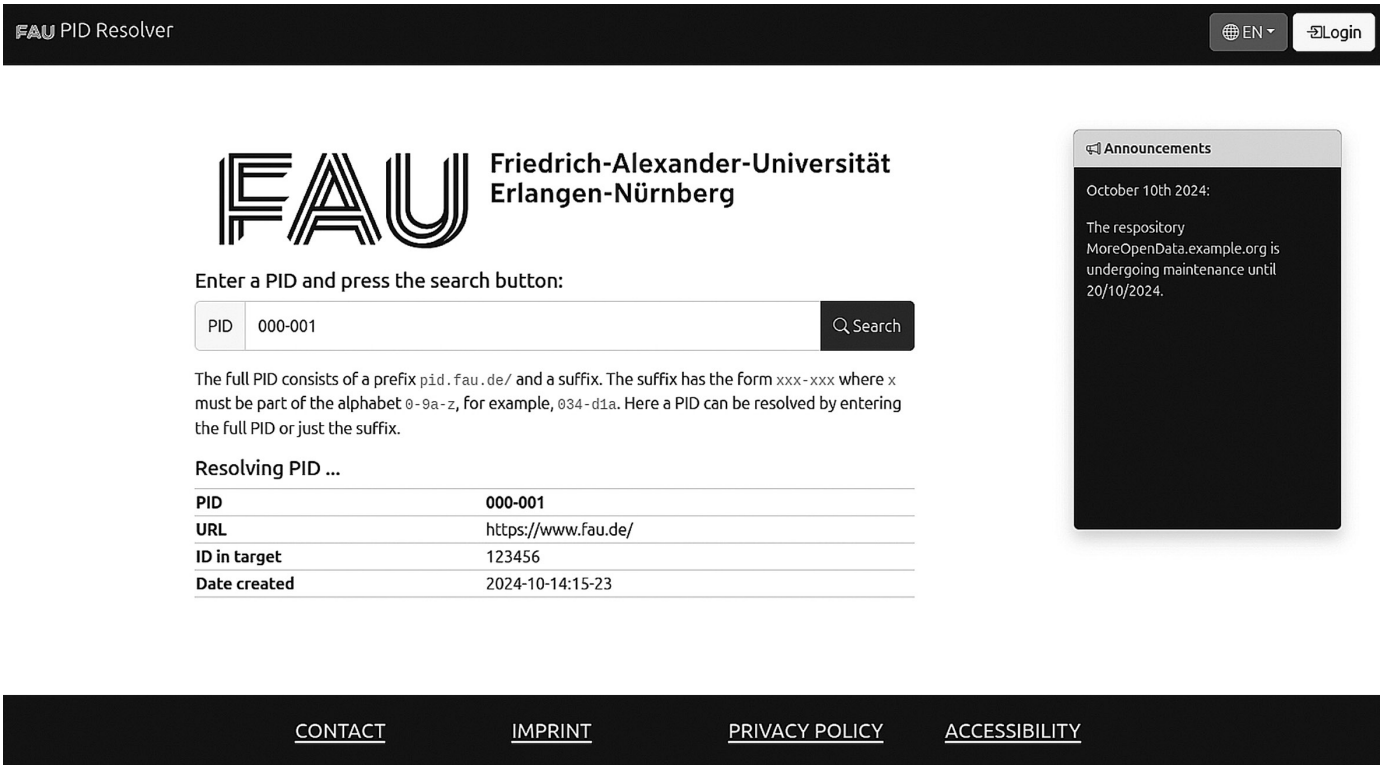
Der komplette PID ist damit selbst eine URL, die direkt vom Frontend des Resolvers verarbeitet werden kann. In diesem Fall werden Nutzer*innen direkt vom Frontend zur Landing Page weitergeleitet. Alternativ kann der PID auf der Hauptseite des Resolvers in ein Webformular eingegeben werden. Abbildung 2 zeigt die Hauptseite des Resolvers in einem Testsystem, nachdem ein PID abgefragt wurde.

Neue PID-Einträge können in der Datenbank auf zwei Wegen vorgenommen werden. Entweder wird ein API-Call eines autorisierten Systems – beispielsweise ei-



1 Darstellung des Zusammenspiels der einzelnen Komponenten des PID -Resolvers. Externe User können über die Weboberfläche des Frontends oder direkt über die API des Backends mit dem System kommunizieren. Für autorisierte User verfügt das Frontend über eine Administrationsfunktion, über die manuell PIDs angelegt und Metadaten aktualisiert werden können. (Architekturüberlegungen: Jürgen Rohrwild (FAU), Dominik Schmid (LMU))

Grafik: Jürgen Rohrwild (FAU)



2 Homepage des Frontends auf einem Testsystem. Es wird gerade der PID pid.fau.de/000-001 anhand des Suffixes 000-001 aufgelöst. Basale Metadaten werden angezeigt, bevor auf die hinlegte URL weitergeleitet wird

nes Repositoriums oder ein elektronischen Laborbuchs – direkt an das Backend gerichtet. In diesem Fall werden dem Zielsystem in der Antwort die Metadaten des PID zur Verfügung gestellt. Alternativ kann ein PID manuell über einen Admin-Zugang im Frontend angelegt werden. Momentan wird ein Prototyp des Resolvers exemplarisch an ein erstes Zielsystem, »Open Data LMU – Physics« angeschlossen. Gleichzeitig wird an einer Verbesserung der Sicherheit des Systems gearbeitet, wozu die Keycloak-Anbindung und die Verschlüsselung der Kommunikation nochmals neu implementiert werden. Mittelfristig müssen noch weitere Funktionalitäten bereitgestellt werden, die Administrator*innen den Betrieb des Resolvers erleichtern. Darunter fallen automatisierte Prüfroutinen, die überprüfen, ob die URLs der Ressourcen noch verfügbar sind, sowie statistische Auswertungen.

Anbindung lokaler Systeme und Repositorien an genuine Langzeitarchive

Das Einbringen von Daten in ein auf genuine Langzeitverfügbarkeit ausgelegtes Archiv findet in aller Regel erst gegen Ende einer Kette von Vorarbeiten statt – die Auswahl des Datenformats und das frühzeitige Verknüpfen mit eindeutigen IDs sind nur einige denkbare Schritte davon. Dabei kann es sinnvoll sein, ein einzelnes Archiv zu nutzen, um möglichst einheitliche Workflows einzusetzen. Die Verwendung unterschiedlicher Archive bietet sich unter Umständen an, wenn für bestimmte Datentypen lokal passende Lösungen vorhanden sind, während andere Ressourcentypen besser in auf diese Ressourcentypen spezialisierte externe Systeme abgelegt werden.

In LaVe sollen beide Ansätze exemplarisch bedient werden. Die Universitätsbibliothek der FAU verfolgt die Idee, verschieden Datentypen (Digitalisate, Dokumente, Forschungsdaten, etc.) in eigenen Workflows in Zielsysteme einzubringen – entweder direkt oder über einen Zwischenschritt in einem lokalen Repositorium.

Dies bietet gleichzeitig die Möglichkeit, auf eine vom Projekt »Digitale LZV« entwickelte Software-Lösung zurückzugreifen: Der FDOrganizer[31] bietet eine Zwischenebene, die genau für diese Situation entwickelt wurde. Ein leichtgewichtiger Vermittler zwischen dezentralen Datenquellen und einer (über-)regionalen LZV-Lösung. Verschiedene Datensätze können lokal gebündelt, bei Bedarf mit Metadaten angereichert und dann an den FDOrganizer weitergeleitet werden. Entsprechende in LaVe erarbeitete Workflows und Schnittstellen sind dann breit nachnutzbar und das Vorgehen bietet direkt Synergieeffekte zwischen den LZV-Projekten.

Für die Anbindung einzelner Systeme an die Vermittlerplattform ist ein jeweils maßgeschneidertes Plugin notwendig, das die Authentifizierung im lokalen System, die Extraktion und das Mapping aller Metadaten sowie schließlich den Import der zu archivierenden Ressour-

cen steuert. Hier wird im Rahmen von LaVe exemplarisch ein Python-Plugin für die Anbindung von DSpace 7 entwickelt. Als Vorlage kann das von FDOrganizer zur Verfügung gestellte Plugin für EasyDB dienen.[32] Grundlegende Mappings wurden hierfür bereits erstellt. Die sichere Authentifizierung und der so mögliche Zugriff auf nicht publizierte Ressourcen muss jedoch noch weiter ausgearbeitet werden. Dies wird frühestens mit Ende der offiziellen Projektlaufzeit von LaVe abgeschlossen sein.

Die Universitätsbibliothek der LMU verfolgt den zweiten Ansatz, bei dem nicht nach Datentyp (z. B. Forschungsdatum, Digitalisat etc.) unterschieden wird, sondern jeder Datentyp denselben Prozess und dieselbe Infrastruktur durchläuft. Jedes digitale Objekt wird in einem System (Fedora Repositories) gespeichert, welches sie für die langfristige Verfügbarkeit vorbereitet. Die unterschiedlichen Anforderungen, die von den Datentypen ausgehen, werden dabei ausschließlich mittels Metadaten abgedeckt. Dies ermöglicht es, den Fokus auf die Erschließung der Objekte und nicht der Objekt-Verwaltung zu legen. Um dies umzusetzen, war zunächst die Migration der eingesetzten Software auf eine Version, die das Oxford Common File Layout (OCFL) unterstützt, notwendig.

Den Datensätzen im System ist ein übergreifendes Metadatenschema gemein. Dies ermöglicht die einfache Bereitstellung der Daten für einen Ingest in das Langzeitarchiv Rosetta, da mittels OAI-PMH Schnittstelle das für Rosetta benötigte Datenformat leicht und für alle Datensätze erstellt werden kann.

Ein weiteres Ergebnis ist die Erarbeitung verschiedener Kostenmodelle für die langfristige Verfügbarkeit von Forschungsdaten. Hier bietet die Universitätsbibliothek der LMU ein kostenfreies Modell an, das sowohl durch die zur Verfügung gestellte Speicherkapazität (500 GB) als auch der aufgewendeten Erschließung- / Personalressourcen begrenzt ist. Werden seitens der Datengeber weitere Anforderungen gestellt, so müssen diese die Kosten tragen. Im Idealfall haben die Datengeber entsprechende Summen oder Personalstellen in ihre (Drittmittel-)Anträge aufgenommen. Dass dieser Idealfall immer öfter eintritt, ist zum einen auf die wachsende Nutzung der LMU-Instanz des Datenmanagementplantools RDMO[33] zurückzuführen als auch auf die Beratungsleistung der FDM-Beratungsstelle der Universitätsbibliothek der LMU.

Weitere Themen

Im Rahmen von LaVe wurden anhand der drei Arbeitspakete wichtige Fragen geklärt und erste Lösungsansätze entwickelt. Darüber hinaus umfasst der Umgang mit Forschungsdaten aber noch weitere Komponenten, die im Projekt identifiziert wurden und langfristig ebenfalls geeignete Lösungen erfordern. Die einleitend beschriebene Problematik im Zusammenhang mit der Reproduzierbarkeit der Forschungsergebnisse ist oft an eine entsprechende Software oder Umgebung gebunden, die zusammen mit den Forschungsdaten aufbewahrt werden müsste.

Dies stellt jedoch die Infrastruktureinrichtungen vor weitere Probleme, da neben der Archivierung von Forschungsdaten auch die Aufbewahrung von Software mitgedacht werden muss. Die Methoden hierfür werden gerade erst erarbeitet und die Verbreitung von Empfehlungen – beispielsweise in Form von Forschungssoftware-Policies – ist noch sehr gering. Während der Projektlaufzeit von LaVe gab es einige Neuerungen, die auch von den Projektbeteiligten getestet werden konnten. Darunter war eine Erweiterung des Datenmanagementplan-Tools RDMO um einen Fragenkatalog[34], der bei der Erstellung eines Softwaremanagementplans (SMP) unterstützt. Dieser SMP ist eine ideale Ergänzung, um sowohl die Forschungsdatenkomponente als auch die verwendete Software in Kontext zu setzen.

Ausblick

Während einige Projektziele erst zum Projektende abgeschlossen sein werden, verbessern andere Ergebnisse des Projekts schon jetzt die langfristige Verfügbarkeit von Forschungsdaten. Neben den Best-Practice-Guides trifft dies vor allem auf den PID-Resolver zu. Dieser wird nicht nur von den beiden am Antrag beteiligten Einrichtungen verwendet, sondern ist auch so konzipiert, dass er nach Projektende leicht von Dritten nachgenutzt werden kann.

Während des Projekts stellt sich immer wieder die Frage, wie die Projektergebnisse verstetigt werden können. Da bei LaVe die Langzeitverfügbarkeit im Mittelpunkt steht, wurde der Verstetigung der Projektergebnisse von Seiten der Projektbeteiligten eine hohe Wichtigkeit beigemessen, denn nicht nur die Technik eines PID-Resolvers muss aktuell gehalten werden, sondern auch die Best-Practice-Guides müssen immer wieder überarbeitet werden, damit sie stets die aktuellen Best Practices abbilden. Geplant ist deshalb, dass diese Dienste in einen HITS Forschungsdaten[35] integriert und weitergeführt werden. Dies ist den Projektbeteiligten zumindest insofern gelungen, dass die HITS-Antragsstellenden (FAU und Technische Hochschule Nürnberg Georg Simon Ohm) die Dienste als eine zentrale Komponente in den Antrag aufgenommen haben.[36]

Für das Projekt wurden auch konkrete Kooperationen eingegangen und die Vernetzung auf unterschiedlichen Ebenen vorangetrieben. Die Universitätsbibliothek der LMU ist dem nestor-Netzwerk[37] beigetreten und in den entsprechenden Arbeitsgruppen aktiv. Darüber hinaus sind Mitarbeitende der beiden Einrichtungen in der Arbeitsgruppe »Long-term Archival (LTA)« der NFDI-Sektion Common Infrastructures aktiv. Durch die Mitarbeit in den entsprechenden Gremien wird deutlich, dass die Arbeiten und gemeinsamen Vorgehensweise zur Langzeitverfügbarkeit von Forschungsdaten längst nicht

abgeschlossen sind, es aber auf überregionaler Ebene Ansätze gibt, diese zu vereinheitlichen und gemeinsame Strategien zu erarbeiten.

Abschließend hat aus Sicht der Verfasser das Projekt LaVe dazu beigetragen, die langfristige Verfügbarkeit von Forschungsdaten an den beiden beteiligten Institutionen und darüber hinaus zu verbessern. Hiermit wurden wichtige Schritte getan, die aber insgesamt nur die ersten Kilometer im Marathonlauf »Langfristige Verfügbarkeit von Forschungsdaten« sind.

Anmerkungen

1 KREFELD, Thomas und Stephan LÜCKE. Verba Alpina. 2016. Ludwig-Maximilians-Universität München. DOI: 10.5282/VERBA-ALPINA.

2 Das Innovationsprogramm Digitaler Campus Bayern fördert verbesserte digitale Infrastrukturen, IT-Grundausbildung und spezifische IT-Studiengänge. Siehe hierzu auch *Hochschule: Digitaler Campus* [Zugriff am 24.10.2024]. Verfügbar unter: https://www.stmwk.bayern.de/studenten/digitalisierung/hochschule-digitaler-campus.html

3 Siehe hierzu auch BERNARD, Lars und andere, *Base4NFDI – Basic Services for NFDI.* Zenodo, 2023. Verfügbar unter: https://doi.org/10.5281/zenodo.10245518, sowie die Homepage Base4NFDI [Zugriff am 24.10.2024]. Verfügbar unter https://base4nfdi.de/

4 Einen ersten Überblick zu den FDM-Landesinitiativen bietet forschungsdaten.info [Zugriff am 31.10.2024]. Verfügbar unter: https://forschungsdaten.info/fdm-im-deutschsprachigen-raum/deutschland/fdm-landesinitiativen-und-regionale-netzwerke/

5 eHumanities wurde in zwei Projektphasen von 2018 bis 2023 vom Bayerischen Staatsministerium für Wissenschaft und Kunst gefördert. Das Projekt war eine Kooperation der Universitätsbibliotheken von der FAU Erlangen-Nürnberg und der LMU München sowie der IT-Gruppe Geisteswissenschaften (ITG) der LMU München.

6 »Digitale Langzeitverfügbarkeit im Bibliotheksverbund Bayern« wird in zwei Projektphasen seit 2019 vom Bayerischen Staatsministerium für Wissenschaft und Kunst gefördert. Die Projektpartner sind die Universitätsbibliothek der Universität Bayreuth und der Universität Regensburg, die Bayerische Staatsbibliothek und das Leibniz-Rechenzentrum.

7 Details zu den Arbeitspaketen können der Projektwebseite entnommen werden: eHumanities – interdisziplinär [Zugriff am 24.10.2024]. Verfügbar unter: https://www.fdm-bayern.org/projekte/ehumanities-interdisziplinaer/

8 Für die aktuelle, dritte Version der Handreichung siehe: BAYER, Christiane und andere, DataCite Best Practice Guide, 2024. Zenodo. DOI: 10.5281/zenodo.7099881

9 GitHub UB-LMU [Zugriff am 24.10.2024]. Verfügbar unter: https://github.com/UB-LMU/datacite-metadata-generator

10 Umfassende Information finden sich auf der Projektwebseite, lzv-bayern.de [Zugriff am 24.10.2024]. Verfügbar unter https://lzv-bayern.de/, sowie in BRUGBAUER, Ralf, Klaus CEYNOWA und André SCHÜLLER-ZWIERLEIN, *Stets einen Schritt voraus: Die Schaffung einer kooperativen Infrastruktur für digitale Langzeitverfügbarkeit in Bayern*. Zeitschrift für Bibliothekswesen und Bibliographie. 2022, 69 (3), 131-141.

11 LZV Bayern [Zugriff am 24.10.2024]. Verfügbar unter: https://lzv-bayern.de/organisation/netzwerk

12 LZV Bayern [Zugriff am 24.10.2024]. Verfügbar unter: https://lzv-bayern.de/organisation/netzwerk

13 BAYER, Christiane und andere, *DataCite Best Practice Guide (Version 3.0)*. 2024. Zenodo. Verfügbar unter: https://doi.org/10.5281/zenodo.7099881

14 SPENGER, Martin, Christiane BAYER, Sonja KÜMMET und Julian SCHULZ. Workflows für das Forschungsdatenmanagement (FDM) in den Digital Humanities an der LMU München, 2021. Zenodo. DOI: 10.5281/zenodo.5031603

15 KOST, Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen, *Katalog archivischer Dateiformate* [Zugriff am 18.10.2024]. Verfügbar unter: https://kost-ceco.ch/cms/kad_main_de.html

16 ETH-Bibliothek, Fachstelle Forschungsdatenmanagement und Datenerhalt (FDD), *Archivtaugliche Dateiformate* [Zugriff am 18.10.2024]. Verfügbar unter: https://unlimited.ethz.ch/display/DD/Archivtaugliche+Dateiformate

17 OSTDATA, *Empfohlene Dateiformate für Forschungsdaten zur Langzeitarchivierung* [Zugriff am 18.10.2024]. Verfügbar unter: https://www.osmikon.de/fileadmin/uploadMountall/user_upload/2_Dateien_nach_Rubriken_geordnet/4_Forschungsdaten/empfohlene_Dateiformate_fuer_Forschungsdaten_OstData_v1.pdf

18 GOLUBIEWSKI-DAVIS, Kristina und andere, *Best Practices for 3D Data Preservation.* In MOORE, Jennifer und Adam ROUNTREY und Hannah Scates KETTLER. 2022. *3D Data Creation to Curation: Community Standards for 3D Data Preservation.* Association of College & Research Libraries.

19 ALEXANDER C., S. ISHIKAWA und M. SILVERSTEIN, et al. *A Pattern Language*. 1977. Oxford University Press, New York.

20 HENZEN, C. und andere. *A Community-driven Collection and Template for DMP Guidance facilitating Data Management across Disciplines and Funding*, Zenodo, 2022. DOI: 10.5281/zenodo.6903404.

21 GitHub – UB-LM [Zugriff am 24.10.2024]. Verfügbar unter: https://github.com/UB-LMU/Dateiformate_BestPracticeGuide

22 FRECH, Andreas und andere, *Dateiformate für die Langzeitverfügbarkeit – Best Practice Guide für Dateiformate*. 2023. Zenodo. Verfügbar unter: https://doi.org/10.5281/zenodo.10021391

23 LZV Bayern [Zugriff am 24.10.2024]. Verfügbar unter: https://lzv-bayern.de/langzeitverf%C3%BCgbarkeit/formats

24 FRECH, Andreas und Yves Vincent GROSSMANN, *Das Internet vergisst doch – Handreichung für die Archivierung von wissenschaftlichen Webseiten*, Zenodo. 2024. DOI: 10.5281/zenodo.10556375.

25 GitHub – UB-LMU [Zugriff am 24.10.2024]. Verfügbar unter: https://github.com/UB-LMU/Webseitenarchivierung_BestPracticeGuide ; LZV Bayern [Zugriff am 24.10.2024]. Verfügbar unter: https://lzv-bayern.de/langzeitverf%C3%BCgbarkeit/website-archivierung

26 WIMALARTNE, Sara und Martin FENNER. (2018). *D2.1 PID Resolution Services Best Practices* (Version Version1). Zenodo. DOI: 10.5281/zenodo.1324300

27 BAZZANELLA, Barbara, Stefano BORTOLI und Paolo BOURQUET. Can Persistent Identifiers Be Cool? *International Journal of Digital Curation.* 2013, 8. 14-28. DOI: 10.2218/ijdc.v8i1.246

28 Mit »genuin« werden in diesem Text Archive bezeichnet, die Dienste jenseits der reinen langfristig garantierten Speicherung der Daten übernehmen, also etwa Migration, Emulation oder zumindest aktive Kuration betreiben.

29 WILKINSON, Mark D. und andere. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

30 Ein häufiges Problem in verbreiteten Schrifttypen ist die Verwechslung von O (Großbuchstabe zu o) mit 0 (die Ziffer) oder I (Großbuchstabe zu i) mit l (Kleinbuchstabe zu L).

31 SIMON, Martin, Robert GÜNTHER und Alexandra ULLRICH, *Vom Labor ins Archiv in 20 Minuten – Ein niederschwelliges Tool zur Langzeitarchivierung von Forschungsdaten, 112*. Bibliocon. Verfügbar unter: https://nbn-resolving.org/urn:nbn:de:0290-opus4-187699

32 GitLab FDOrganizer [Zugriff am 24.10.2024]. Verfügbar unter: https://gitlab.lzv-bayern.de/forschungsdaten/fdorganizer/-/tree/master/

33 Research Data Management Organizer, die RDMO-Webseite [Zugriff am 24.10.2024]. Verfügbar unter: https://rdmorganiser.github.io/

34 Der SMP ist publiziert auf dem rdmorganizer-GitHub [Zugriff am 31.10.2024]. Verfügbar unter: https://github.com/rdm-organiser/rdmo-catalog/tree/master-rdmo1.x/rdmorganiser/questions. Details finden sich in FRANKE, Michael und Yves Vincent GROSSMANN, *Sustainable and FAIR Software in Research – A New RDMO Catalogue for Software Management Plans,* 2023. MPG.PuRe. Verfügbar unter: https://hdl.handle.net/21.11116/0000-000D-A277-6

35 HITS steht für Hochschulübergreifenden IT-Services, die vom Digitalverbund Bayern für die Hochschulen des Landes aufgebaut werden. Siehe hierzu auch die Webseite HITS – Digitalverbund Bayern im Hochschulbereich [Zugriff am 31.10.2024]. Verfügbar unter: https://digitalverbund.bayern/hits/

36 Zum Zeitpunkt der Artikelverfassung ist noch nicht bekannt, ob der HITS vom Land Bayern gefördert wird.

37 Siehe nestor-Portal langzeitarchivierung.de [Zugriff am 31.10.2024]. Verfügbar unter: https://www.langzeitarchivierung.de/Webs/nestor/DE/nestor/Mitglieder/IMU.html

Verfasser

Alexander Berg-Weiß,
Leitung Abteilung Digitale Dienste,
Universitätsbibliothek,
Ludwig-Maximilians-Universität,
Geschwister-Scholl-Platz 1,
80539 München,
Telefon +49 89 2180-5800,
alexander.weiss@ub.uni-muenchen.de,
ORCID: 0000-0002-7435-8676
Foto: Nadine Bollendorf

Dr. Jürgen Rohrwild, Leiter der Technisch-naturwissenschaftliche Zweigbibliothek, Referat Forschungsdatenmanagement, Fachreferent Physik, Biologie, Technisch-naturwissenschaftliche Zweigbibliothek, Universitätsbibliothek, Friedrich-Alexander-Universität Erlangen-Nürnberg,
Erwin-Rommel-Straße 60,
91058 Erlangen,
juergen.rohrwild@fau.de,
ORCID: 0000-0002-1167-0339
Foto: Stacey Warter

Dr. Martin Spenger, Referent für Forschungsdatenmanagement Teamleitung FDM-Beratung,
Universitätsbibliothek, Ludwig-Maximilians-Universität,
Geschwister-Scholl-Platz 1,
80539 München,
martin.spenger@ub.uni-muenchen.de,
ORCID: 0000-0002-8841-5985
Foto: Nadine Bollendorf