

Die hochkomplexe Aufgabe, digitale Daten langzeitverfügbar zu halten, ist nicht von einer Institution allein zu bewältigen. Gleichzeitig bedarf die Archivierung wissenschaftlicher Daten des engen Dialogs mit der jeweiligen Wissenschaft. Das vom bayerischen Staatsministerium für Wissenschaft und Kunst geförderte Projekt »Digitale Langzeitverfügbarkeit für Wissenschaft und Kultur in Bayern«, in dem die Universitäten Regensburg und Bayreuth mit der Verbundzentrale des Bibliotheksverbunds Bayern und der Bayerischen Staatsbibliothek sowie weiteren Bibliotheken und Hochschulen zusammenarbeiten, nimmt diesen kooperativen Gedanken als zentrales infrastrukturelles Element auf. Ziel des Projekts ist es, das wissenschaftliche Gedächtnis Bayerns durch ein gemeinschaftliches, systematisch organisiertes Langzeitverfügbarkeits-Netzwerk dauerhaft aktiv und entwicklungsfähig zu halten. Im Projekt wird, aufbauend auf der Rosetta-Installation der Bayerischen Staatsbibliothek / Verbundzentrale des Bibliotheksverbunds Bayern, eine bayernweit vernetzte, kombiniert zentral-dezentrale Infrastruktur (ZDI) im Rahmen des Kooperativen Leistungsverbands geschaffen.

It is not possible for one institution to cope by itself with the highly complex task of ensuring the long-term availability of digital data. At the same time, the archiving of scientific data necessitates close dialogue with the respective branches of science. The »Long-term availability of digital data for science and culture in Bavaria« project is funded by the Bayerisches Staatsministerium für Wissenschaft und Kunst (Bavarian State Ministry of Science and the Arts). Based on cooperation as a central infrastructural element, the universities of Regensburg and Bayreuth are working together with the Verbundzentrale des Bibliotheksverbunds Bayern, the Bayerische Staatsbibliothek and other libraries and universities. The aim of the project is for the collaborative, systematically organised long-term availability network to ensure that Bavaria's scientific memory remains permanently active and capable of development. Within the framework of the contractual service partnership, the project will create a combined centralised/decentralised infrastructure (ZDI) which is networked throughout Bavaria. This will build on the Rosetta installation of the Bayerische Staatsbibliothek / Verbundzentrale des Bibliotheksverbunds Bayern.

RALF BRUGBAUER, KLAUS CEYNOWA, ANDRÉ SCHÜLLER-ZWIERLEIN

# Stets einen Schritt voraus: Die Schaffung einer kooperativen Infrastruktur für digitale Langzeitverfügbarkeit in Bayern

»You are never done with digital preservation.«  
(Trevor Owens, 2017)

**W**ie bewahrt man etwas auf, das sich ständig ändern muss, um mit einem dynamischen Umfeld Schritt zu halten? Schritt halten ist das Motto der digitalen Langzeitarchivierung. Will man digitale Daten aufbewahren, müssen entweder die Daten in regelmäßigen Abständen bewegt und verändert werden (Migration), oder man passt die Umgebung der Daten so an, dass sie einen älteren Zustand imitiert (und dies immer wieder neu; Emulation), oder, schließlich, man bewahrt jede Alterungsstufe der Datenumgebung selbst museal auf (und behandelt diese ihrerseits laufend konservatorisch; Computermuseum). Von den drei Ansätzen hat sich lediglich die Migration als massentauglich und damit für Millionen von Dateien einsetzbar erwiesen. Doch auch diese birgt ihre eigenen Kom-

plexitäten. Denn wer legt fest, wann eine Datei auf ein neues Format migriert werden muss? Wer sorgt dafür, dass diese Entscheidung rechtzeitig erfolgt – also vor der den digitalen Daten permanent drohenden letalen Formatobsoleszenz? Wer garantiert, dass Entscheidungen, die für Millionen von Dateien getroffen werden, auch bei jeder einzelnen Datei zu brauchbaren Ergebnissen führen? Welche Eigenschaften muss, wenn migriert wird, die neue Datei (eine Datei im neuen Format ist eine neue Datei, ein stets neues »Original«) aufweisen, damit sie alle notwendigen Funktionen der alten erfüllt?<sup>1</sup> Wer definiert dies?<sup>2</sup> Wer kontrolliert, ob nach der Migration alle entsprechenden Funktionen und Eigenschaften jedes einzelnen Objekts vorhanden sind?<sup>3</sup> Was, wenn sich dieses Verständnis der notwendigen Funktionen und Eigenschaften über die Jahre hinweg ändert, etwa wenn sich die Bedürfnisse der Wissenschaft ändern? Der Weg zurück zur Ausgangsversion ist mit Höchstschwierigkeiten versehen und oft genug

nicht mehr begehbar – versucht man die Ausgangsversion dennoch mit zu archivieren? Bewahrt man zudem über die Jahre hinweg jede einzelne Version auf, um gegebenenfalls Fehler zurückverfolgen zu können? Wie überliefert man Teile der für die Nutzung notwendigen »Knowledge Base«<sup>4</sup> mit, also etwa technische Informationen, Informationen zu Darstellung und Inhalt – oder gar die nötige Software (alles ihrerseits Daten, die wiederum Metadaten erfordern) etc. *ad infinitum*? Die Komplexität dieser Fragestellungen deutet auf die strukturelle Schwachstelle der digitalen Langzeitarchivierung hin: Ihre Crux ist es, dass neben Software und Serverkapazitäten permanent ein hochspezialisiertes Team vorhanden sein muss, das sich genau diesen Fragen und vielen weiteren stellt – über Jahre, Jahrzehnte und, ja, man muss es aussprechen, Jahrhunderte. Das digitale Zeitalter zwingt uns, die diachrone Zugänglichkeit von Daten – also ihre Benutzbarkeit über die Zeit hinweg – als *Prozess* aufzufassen.<sup>5</sup> Die stets drohende Obsoleszenz erfordert *permanentes* und *proaktives* Management der Daten (*Preservation Planning*)<sup>6</sup> und die vorausschauende Schaffung flexibler Infrastrukturen, um dem Datenverlust stets einen Schritt voraus zu sein.

Geht es um wissenschaftliche Daten, wird dieser Prozess noch einmal schwieriger: Da nur Wissenschaftler\*innen viele der oben genannten Fragen entscheiden können, kann die Archivierung der Daten nicht ausschließlich durch eine zentrale spezialisierte Instanz geschehen, die zwar über entsprechende technische und organisatorische Kompetenz verfügt, nicht aber über die notwendigen wissenschaftlichen Kenntnisse. Mit einer einmaligen Ablieferung der Daten ist somit keine dauerhafte Archivierung bzw. zweckgerechte Zugänglichhaltung wissenschaftlicher Daten möglich. Gleichzeitig ist diese als technische und organisatorische Aufgabe so umfangreich, dass nicht viele Hochschulen in der Lage sind, hierfür eigene Infrastrukturen aufzubauen. Daher ist zur Archivierung wissenschaftlicher Daten zwingend ein kooperatives System mit zentralen und dezentralen Elementen notwendig, das die Vorteile zentraler Infrastrukturen und spezialisierter Kompetenzen mit dem laufenden direkten Kontakt zur lokalen Wissenschaft kombiniert, die traditionelle Trennung zwischen »Producer role« und »Archive role« aufhebt und die »Designated Community« dauerhaft mit einbindet.<sup>7</sup> Das vom bayerischen Staatsministerium für Wissenschaft und Kunst nun bereits in der zweiten Projektphase geförderte Projekt »Digitale Langzeitverfügbarkeit für Wissenschaft und Kultur in Bayern«, in dem die Universitäten Regensburg und Bayreuth mit der Verbundzentrale des Bibliotheksverbunds Bayern und der Bayerischen Staatsbibliothek sowie weiteren Bibliotheken und Hochschulen zusammenarbeiten, nimmt diesen kooperativen Gedanken als zentrales infrastrukturelles Element auf und setzt ihn Schritt für Schritt um.

## Projektkontext

Die Digitalisierung hat die Dynamik und Leistungsfähigkeit der Wissenschaft in den letzten Jahrzehnten enorm erhöht. Wissenschaftsrelevante Information liegt in zunehmendem Maße digital vor. Zur Digitalisierung gehört jedoch die Nachhaltigkeit in gleichem Maße wie die Innovation. Denn für eine sich dynamisch fortentwickelnde Wissenschaft reicht die geordnete Bereitstellung oder einmalige Ablage von Materialien nicht aus – sie ist vielmehr auf die *Langzeitverfügbarkeit*, also die dauerhaft und jederzeit mögliche Zugänglichkeit und vielfältige Nachnutzbarkeit wissenschaftsrelevanter Information angewiesen. Dies bedeutet, dass wissenschaftsrelevante Informationen über Jahrzehnte hinweg mit hohem Aufwand stets aktiv nutzbar und zugriffsbereit gehalten werden müssen. Diese langfristige Perspektive abzusichern, stellt eine eigene technische und organisatorische Herausforderung dar.

Die Langzeitverfügbarkeit der von wissenschaftlichen Bibliotheken gesammelten Objekte und Informationen wird von der Wissenschaft ebenso wie von der breiten Bevölkerung erwartet und für selbstverständlich genommen, ohne dass die Komplexität dieser Aufgabe realisiert wird. Während dieser Anspruch im analogen Bereich – trotz begrenzter, oft schwindender Magazinkapazitäten und verbreiteter Sanierungsstaus – noch weitgehend einlösbar ist, sind die notwendigen Infrastrukturen und Prozesse im digitalen Bereich bislang nur in Ansätzen vorhanden. Die Schaffung und Erhaltung digitaler Langzeitverfügbarkeit ist deutlich komplexer und personalintensiver als die Verfügbarhaltung gedruckter Materialien. Mit dem Projekt »Digitale Langzeitverfügbarkeit für Wissenschaft und Kultur in Bayern« hat der Freistaat Bayern begonnen, eine zukunftsorientierte, kooperative Infrastruktur zu schaffen, die in möglichst effizienter Weise die Langzeitverfügbarkeit wissenschaftlicher Information sichert und damit die Grundlage für innovative Wissenschaft erhält.

Mit der Investition in ein abgestimmtes, kooperatives Konzept erkennt der Freistaat Bayern die These an, dass Langzeitverfügbarkeit genauso zur Digitalisierung gehört wie Innovation. Gleichzeitig stellt die für die Langzeitverfügbarkeit der wissenschaftlichen und kulturellen Datenbestände notwendige intensive dauerhafte Kooperation einen Paradoxfall für die Umsetzung des in der bayerischen Hochschulgesetzgebung beschriebenen »Kooperativen Leistungsverbundes« zwischen der Bayerischen Staatsbibliothek und den Bibliotheken der bayerischen Hochschulen dar. Die *De-facto-Etablierung* funktionierender kooperativer Langzeitverfügbarkeitssysteme ist deutschlandweit von höchster Bedeutung.<sup>8</sup> Mit seinem systematisch kooperativen Konzept der Langzeitverfügbarkeit schafft der Freistaat Bayern auf diesem jungen Gebiet eine paradigmatische Struktur.

## Projektziele

Ziele und Struktur des Projekts leiten sich aus einigen grundlegenden Erkenntnissen ab:

- Die Schaffung und Erhaltung digitaler Langzeitverfügbarkeit ist ein eigener, spezialisierter Bereich, der über die kurz- und mittelfristige Bereitstellung von Publikationen und Daten in Repositorien hinausgeht.
- Digitale Langzeitverfügbarkeit wird nicht einmalig geschaffen, sondern kann nur in einem fortlaufenden Prozess erhalten werden.
- Für die digitale Langzeitverfügbarkeit wissenschaftlicher und kultureller Information gibt es keine objektklassen- und bedarfsübergreifende Patentlösung.
- Digitale Langzeitverfügbarkeit ist nicht durch bloßen Softwareeinsatz umsetzbar.
- Die aufwendigen und personalintensiven Arbeiten zur Erhaltung digitaler Langzeitverfügbarkeit sind nicht von einer einzelnen Bibliothek zu leisten, sondern nur kooperativ realisierbar.
- Die Schaffung und Erhaltung digitaler Langzeitverfügbarkeit involviert neben technischen Aspekten und Infrastrukturen ebenso organisatorische, personelle, rechtliche und wissenschaftsspezifische Fragen.

Die Konzeption einer zentral-dezentralen Infrastruktur (ZDI) für die Langzeitverfügbarkeit von verschiedenen Objektklassen setzt auf diesen Erkenntnissen auf: Das wissenschaftliche Gedächtnis Bayerns kann nur durch ein gemeinschaftliches, systematisch organisiertes Langzeitverfügbarkeitsnetzwerk dauerhaft aktiv und entwicklungsfähig bleiben. Für diese große Zukunftsaufgabe waren die bayerischen Infrastruktureinrichtungen bisher jedoch nicht ausreichend gerüstet: Mit der Rosetta-Installation der Bayerischen Staatsbibliothek / Verbundzentrale des Bibliotheksverbunds Bayern (2009/2014) konnte zwar eine wertvolle softwaretechnische Grundlage für ein Langzeitverfügbarkeitssystem geschaffen werden. Die Schaffung einer bayernweit vernetzten Infrastruktur stellt jedoch deutlich höhere Anforderungen: Jenseits des bloßen Software-Einsatzes (und hier sind über Rosetta hinaus eine ganze Reihe von weiteren Tools erforderlich)<sup>9</sup> kann digitale Langzeitverfügbarkeit nur durch ein integriertes System von Technologie, Personaleinsatz sowie kommunikativen und organisatorischen Regelungen im Sinne des Open Archival Information System (OAIS)<sup>10</sup> gewährleistet werden, das den internationalen Gold-Standard bei der Langzeitverfügbarkeit digitaler Informationen darstellt: »An OAIS is an Archive [...] consisting of an organization [...] of people and systems that has accepted the responsibility to preserve information and make it available for a *Designated Community*«.<sup>11</sup>

Insbesondere ist zur Etablierung eines bayernweiten Systems, das die Bedürfnisse der Wissenschaftler\*innen der bayerischen Universitäten und Hochschulen berücksichtigt, eine über die teilnehmenden Institutionen

verteilte Organisationsstruktur notwendig: Arbeitsintensive und wissenschaftsnahe Tätigkeiten wie Datenakquise, Beratungsdienstleistungen und Kommunikationsprozesse, Qualitätsmanagement und Aufbereitung der Daten, Einhaltung der Grundsätze zur SIP-Bildung (Submission Information Packages, also Dateien, die in ein OAIS eingebracht werden sollen),<sup>12</sup> Festlegung der Verfügbarkeitsziele, Abklärung rechtlicher Aspekte, die Vervollständigung von für die Archivierung notwendigen Daten, die Etablierung der Nutzung offener und standardisierter Formate u. v. m., die Bestandteil des Langzeitverfügbarkeitssystems/-prozesses sind, können nur dezentral in der jeweiligen Institution erbracht werden. Dezentrale Datenkurator\*innen sind unentbehrlich, damit die zentrale Infrastruktur effizient arbeiten kann: Ein intensiver<sup>13</sup> und methodischer, lokal basierter Pre-Ingest ist die Basis einer effektiven Langzeitarchivierung. Denn nur für qualitätsgesicherte und standardkonforme Daten mit festgelegten Verfügbarkeitszielen, ausreichender Dokumentation und detaillierten, zielgruppengerechten Metadaten (Representation Information, Preservation Description Information u. a.),<sup>14</sup> die technische Interpretierbarkeit und erzeugerunabhängige Verständlichkeit<sup>15</sup> ermöglichen, kann Langzeitverfügbarkeit garantiert werden. Gleichzeitig ist die Etablierung von Preservation Policies an den Hochschulen dringend erforderlich, um landesweit gemeinsame Ziele und Praktiken dauerhaft zu etablieren.

Der Perspektiv-Bericht, der 2017 im Auftrag des IT-Beirats von einer mit Vertreter\*innen aus bayerischen Bibliotheken und Rechenzentren besetzten Arbeitsgemeinschaft erstellt wurde, empfahl dementsprechend »die Einrichtung einer kombiniert zentral-dezentralen Infrastruktur [...] im Rahmen des Kooperativen Leistungsverbands. Diese Infrastruktur ist mit klar definiertem Personaleinsatz, fester Aufgabenteilung sowie schriftlich fixierten Geschäftsgängen und Dialogprozessen zu betreiben.«<sup>16</sup> In dieser stringent gemäß dem OAIS-Standard organisierten und dokumentierten Infrastruktur sind zentrale und dezentrale Leistungen notwendige Elemente eines bayernweit arbeitenden Gesamtsystems, das als gemanagtes »Sicherheitsnetz« hinter aktuellen Bereitstellungssystemen und Forschungsdateninfrastrukturen stehen sollte.

Ziel des Projekts ist es, das wissenschaftliche Gedächtnis Bayerns durch ein gemeinschaftliches, systematisch organisiertes Langzeitverfügbarkeits-Netzwerk dauerhaft aktiv und entwicklungsfähig zu halten. Im Projekt soll, aufbauend auf der Rosetta-Installation der Bayerischen Staatsbibliothek / Verbundzentrale des Bibliotheksverbunds Bayern, eine bayernweit vernetzte, kombiniert zentral-dezentrale Infrastruktur (ZDI) im Rahmen des Kooperativen Leistungsverbands geschaffen werden.

Nachdem in einer ersten Projektphase in kooperativer Arbeit technische Modelllösungen, Prozesse und

Workflows, Entwürfe für Preservation Policies, eine Mustervereinbarung sowie ein Entwurf für ein Kostenmodell erarbeitet werden konnten, soll die zweite Projektphase die Konsolidierung, Umsetzung und Nachnutzung der erarbeiteten Lösungen leisten, um ein nachhaltiges bayernweites System der Langzeitverfügbarkeit auf Basis von Rosetta zu etablieren.

### Die erste Projektphase (2019–2021)

Für den Aufbau der ZDI im Rahmen des Bibliotheksverbunds Bayern ist ein Zeitraum von insgesamt fünf Arbeitsjahren veranschlagt. Die erste Projektphase (2019–2021) war angelegt auf die Etablierung erster dezentral-zentraler Workflows als Kernelemente eines zukünftigen Gesamtsystems. In dieser Projektphase waren – aufbauend auf der Bedarfsermittlung der Arbeitsgemeinschaft Langzeitverfügbarkeit aus dem Jahr 2017, die besonders hohen Handlungsbedarf bei den Objektklassen Open-Access-Publikationen, Digitalisate und Forschungsdaten festgestellt hatte<sup>17</sup> – konkrete Vorarbeiten für die Ausgestaltung der ZDI zu leisten, insbesondere die Ausarbeitung der Kommunikationsprozesse, die Gestaltung der organisatorischen und technischen Geschäftsgänge, die Ausarbeitung von Detailkonzepten und nachnutzbaren Lösungen für die drei genannten Objektklassen, der testweise Ingest entsprechender Datenmengen, die Ausarbeitung von Preservation Policies, einer Mustervereinbarung sowie eines Geschäfts- bzw. Kostenmodells. Hinzu kam die Erarbeitung erster Grundlagen für die Integration relevanter Inhalte in die universitäre und bibliothekarische Ausbildung mit Blick auf die dauerhafte Gewinnung geeigneten Nachwuchses.

Das Hauptaugenmerk galt neben dem Ausbau der zentralen Infrastruktur in der Verbundzentrale (München) der Erarbeitung beispielhaft nachnutzbarer Lösungen durch zwei teilnehmende Hochschulen (Regensburg, Bayreuth) für die genannten, von den bayerischen Hochschulen am höchsten priorisierten Objektklassen: Open-Access-Publikationen und Digitalisate (Regensburg) sowie Forschungsdaten (Bayreuth). Entsprechend wurden die drei beantragten Personalstellen auf die Standorte verteilt. Die teilnehmenden Hochschulen bleiben über die Projektzeit hinaus spezialisierte Ansprechpartner für die Umsetzung von Lösungen der jeweiligen Objektklasse an weiteren bayerischen Hochschulen. Die Arbeitsgemeinschaft Langzeitverfügbarkeit des Bibliotheksverbunds Bayern und der bayerischen Hochschulrechenzentren hatte die Aufgabe, die Umsetzung des Gesamtsystems zu begleiten sowie Synergien mit anderen bayerischen Technologieprojekten auszuloten (z. B. der Forschungsdateninfrastruktur).

### Regensburg – Digitalisate

Im Rahmen des Projekts erarbeitet und erprobt die Universitätsbibliothek Regensburg gemeinsam mit der

Verbundzentrale Arbeitsabläufe für die Archivierung von Digitalisaten im Langzeitverfügbarkeitssystem Rosetta. Als Portal für die Präsentation der Digitalisate wird bisher der Bayerische Multimediasever, betrieben von der Verbundzentrale auf Basis der DigiTool-Software von Ex Libris, verwendet. Die in jahrelanger Praxis bewährten Abläufe mit DigiTool werden im Projekt auf die Archivierung und Präsentation der Digitalisate in Rosetta, ebenfalls von Ex Libris, übertragen, das als »light archive« dient. Als Testobjekte wurden Digitalisate aus dem bavarikon-Projekt »Die Rechnungsbücher des Regensburger St. Katharinenospitals 1354 bis 1934« gewählt. Das gesamte Projekt umfasst 1.113 Bände. Der Speicherbedarf nach Abschluss der Digitalisierung wird auf ca. 25 Terabyte geschätzt.

Entgegen dem üblichen Verfahren werden die Digitalisate nicht direkt in die digitale Plattform bavarikon eingespielt, sondern zuerst im Langzeitverfügbarkeitssystem Rosetta gespeichert. Die Übertragung nach bavarikon erfolgt dann in einem zweiten Schritt aus Rosetta heraus. Die Verbundzentrale entwickelt hierfür derzeit eine Schnittstelle von Rosetta zu bavarikon, die auf einer bewährten Schnittstelle von DigiTool nach bavarikon aufsetzt. Die beschreibenden Metadaten der Rechnungsbücher aus dem Archivsystem FAUST werden zukünftig ebenfalls über Rosetta nach bavarikon transportiert. So wird sichergestellt, dass alle notwendigen Daten, die für die reichhaltige Präsentation in bavarikon benötigt werden, im Archivsystem verwaltet und für spätere Anwendungen auch gespeichert werden können.

An der Universitätsbibliothek Regensburg wurden in der ersten Projektphase die für die Aufbereitung der Digitalisate und Metadaten notwendigen Programme fertiggestellt. Die mit bavarikon abgestimmten bibliografischen Metadaten liegen ebenso vor wie testweise erstellte Strukturmetadaten. Die Digitalisate wurden erfolgreich in eine Testinstallation von Rosetta eingespielt. Da viele Bibliotheken, die ihre Digitalisate auf dem Bayerischen Multimediasever veröffentlichen, bereits verwandte Programme und Abläufe nutzen, die von der Universitätsbibliothek Regensburg zusammen mit der Verbundzentrale entwickelt worden sind, sind gute Voraussetzungen für eine breite Nachnutzung gegeben; auch bereits existente Projekte können dann von DigiTool nach Rosetta überführt werden. Die in Arbeit befindliche Schnittstelle zu bavarikon schließlich wird Rosetta für weitere Bibliotheken attraktiv machen, da keine individuellen Lösungen mehr erarbeitet werden müssen.

### Regensburg – Open-Access-Zeitschriften

Im Rahmen des Projekts werden beim Projektpartner in Regensburg genuine digitale Open-Access-Publikationen für den Ingest in Rosetta zur Verfügung gestellt. Als Publikationssystem wurde die Software Open Journal Systems (OJS)<sup>18</sup> vom Public Knowledge Project



(PKP) gewählt, dessen Haupteinsatzgebiet das Publizieren von wissenschaftlich begutachteten Zeitschriftenartikeln ist. Dieses System wurde u. a. wegen seiner großen Verbreitung im Bibliotheksverbund Bayern (BVB) gewählt. Anhand der an der Universität publizierten Zeitschrift *Movement and Nutrition in Health and Disease* wurde gemeinsam mit der Verbundzentrale ein Verfahren zur Meldung von neuen Publikationen und zur Bereitstellung der Artikel inklusive Manuskriptversionen und ergänzender Daten sowie Metadaten erarbeitet. Als primäre Schnittstelle zwischen den Systemen Rosetta und OJS wurde die OAI-Schnittstelle gewählt, insbesondere da a) es sich um eine standardisierte Schnittstelle handelt, die in OJS und gängiger Repositoriensoftware standardmäßig implementiert ist; b) sie dem Service Provider (Rosetta) ermöglicht, neue Publikationen vom Content Provider (OJS) zu harvesten; c) sie erweiterbar ist bezüglich des verwendeten Metadatenschemas; d) sie ein etabliertes Verfahren darstellt, das z. B. bei der Pflichtablieferung von digitalen Werken an die Deutsche Nationalbibliothek genutzt wird. Über die OAI-Schnittstelle werden die Metadaten der Publikation im Format Dublin Core zur Verfügung gestellt. Für den Artikelvolltext wird der Link auf die frei verfügbare Datei angegeben, über den die Datei selbst in einem zweiten Schritt im Zuge des Ingests transferiert werden kann. Als dauerhafter Identifier für den Eintrag wird der DOI genutzt. Die durchgeführten Ingests waren erfolgreich, die Artikel der Zeitschrift können mit diesem Verfahren zukünftig archiviert werden. Die Nachnutzung der Archivierung von Zeitschriftenartikeln, die mit dem System OJS publiziert werden, ist somit für weitere bayerische Bibliotheken bereits möglich. Die notwendige Schnittstelle ist standardmäßig in OJS implementiert.

Bei der Publikation von Zeitschriftenartikeln werden in OJS jedoch neben den Druckfahnen auch weitere Dokumente und Texte – ausgehend vom eingereichten Manuskript über redaktionelle Entscheidungen und Gutachten von externen Reviewern bis zu den Druckfahnen – gespeichert. Gerade angesichts der aktuellen Diskussion um pseudowissenschaftliche Veröffentlichungen ist es wünschenswert, zusätzlich zum Artikel den gesamten Publikationsprozess inklusive des Qualitätsprüfungsprozesses und der unterschiedlichen Versionen zu archivieren. Auch dies wird im Projekt umgesetzt: Da sich die bereits in OJS angebotenen Exportformate für eine dauerhafte Erstellung des langzeitarchivierungsfähigen Formats als nicht nutzbar erwiesen haben, wurde ein neues Programm entwickelt, welches das Exportformat erstellt. Das Exportformat enthält alle Informationen zum Publikationsprozess sowie Links zu den (nicht publizierten) Vorversionen des Artikels. Da die Begutachtungsinformation im Normalfall nicht öffentlich zugänglich ist, muss das oben erwähnte Schema für die Bereitstellung über die OAI-Schnittstelle erweitert

werden. Mittels Network File System (NFS) werden diese Daten in einem definierten XML-Format für den Ingest zur Verfügung gestellt. Die NFS-Infrastruktur bietet die Möglichkeit, geschützte Inhalte sicher zu übertragen und unberechtigten Zugriff durch Dritte auf die Daten zu vermeiden. Da das Exportformat direkt aus der Datenbank erstellt wird, müssen bei zukünftigen Änderungen der OJS-Version lediglich die Datenbank auf Änderungen überprüft und gegebenenfalls das Exporttool angepasst werden. Dies wird im Rahmen von regelmäßigen Softwareupdates an der UB Regensburg erfolgen. Das Programm wird unter einer freien Lizenz zur Verfügung gestellt. Ziel ist es, die Hauptversionen ab Version 3.3.0 zu unterstützen. Für die Konfiguration muss lediglich der Pfad der Ablage der XML-Datei in der NFS-Infrastruktur angegeben werden. Alle weiteren Informationen erhält das Programm direkt aus der OJS-Installation. OJS-Betreiber können diese Lösung so mit geringem Aufwand nachnutzen.

Das erarbeitete Ablieferungsverfahren lässt sich zudem auf weitere Repositorien jenseits von OJS übertragen. Es kann z. B. auch für *enhanced publications* verwendet werden oder für interne Dokumente, die nicht öffentlich zur Verfügung stehen. Mit Blick auf *enhanced publications* wurde ein weiteres Publikationsrepository, das »Virtual Laboratory« (betrieben von der Universität Regensburg, dem Max Planck Institute for the History of Science, Berlin, und der Fakultät Medien, Bauhaus-Universität Weimar),<sup>19</sup> für die Ablieferung mittels OAI-Schnittstelle angepasst. Hierzu wurde das System auf die zukunftsweisende, am CERN entwickelte und weit verbreitete Software Invenio<sup>20</sup> umgestellt und die OAI-Schnittstelle konfiguriert. Diese Lösung ist für alle Invenio-Nutzer\*innen nachnutzbar. Erweiterungen zur Bereitstellung der verknüpften Dateien wurden als Open Source veröffentlicht. Die publizierten Artikel des »Virtual Laboratory« werden zukünftig in der ZDI archiviert. Die Archivierung interner Dokumente ist u. a. für Open-Access-Publikationen aus dem Repository einer Hochschule interessant. Neben den frei verfügbaren Volltexten können hier zusätzliche Dokumente zur Rechteübertragung oder zugehörige Rechnungen vorliegen, welche nicht öffentlich zugänglich sind, aber für eine Langzeitarchivierung relevant sind. Hier kann das Verfahren analog angewendet werden. Diese Erweiterung wird derzeit in dem DFG-geförderten Projekt *openCost: automatisierte, standardisierte Lieferung und offene Bereitstellung von Publikationskosten und Verlagsvereinbarungen* (2021–2024; Projektpartner: Universitätsbibliothek Regensburg, Universitätsbibliothek Bielefeld, DESY Hamburg) ausgearbeitet.

### Bayreuth – Forschungsdaten

Die Aufgabe der Bayreuther Projektpartner ist es, zusammen mit Datenlieferant\*innen Mechanismen zur Archivierung von Forschungsdaten aus allen Bereichen

der Universität zu konzipieren. Um eine möglichst hohe Abdeckung sowie eine einfache Nachnutzbarkeit zu ermöglichen, wurde ein pragmatischer Ansatz gewählt. Die Infrastruktur an der Universität Bayreuth bietet ein breites, sehr heterogenes Spektrum an lokalen Forschungsdaten-Repositories. In der Praxis verwenden Forschende zudem oft eigene oder fachspezifische Datenbanken, die teilweise lizenzpflichtig sind und/der auf cloudbasierten Services von Fremdanbietern aufsetzen. Die breite Streuung der Daten und Speicherungsformen erzeugt den Bedarf nach einer zentralen Verwaltungsschnittstelle, die eine Anbindung an verschiedenste Datenbanken ermöglicht, den Pre-Ingest aus heterogenen Quellen systematisiert und als niedrigschwelliges Web-Frontend für die Datenlieferant\*innen dient. Dieses zentrale Pre-Ingest-Tool wird im Projekt unter dem Namen *FDOrganizer* entwickelt und besteht aus drei Modulen:

Der *Repository-Manager* bildet die Schnittstelle zu verschiedenen Forschungsdaten-Repositories. Er ermöglicht den Forschenden, die Inhalte aus den Repositorien in das zentrale Verwaltungssystem zu exportieren und sie gegebenenfalls durch weitere Inhalte zu ergänzen. Technische wie inhaltliche Metadaten werden automatisiert aus den Repositorien extrahiert und in einem einheitlichen Format abgespeichert. Der *Metadata-Manager* unterstützt die Generierung von Metadaten nach unterschiedlichen Schemata. Ziel ist es, dass dieses Modul auch außerhalb des (Pre-)Ingest-Prozesses nachgenutzt werden kann. Hierfür wird den Forschenden die Möglichkeit zum Verwalten, Konvertieren und Exportieren von Metadaten in verschiedenen Formaten und Schemata gegeben. Mit Blick auf die Nachnutzbarkeit ist das Modul so konzipiert, dass weitere Schemata oder Formatierungen einfach hinzugefügt werden können. Der *Ingest-Manager* schließlich unterstützt die Forschenden bei der Zusammenstellung der Ingests, die sich aus den Daten aus dem Repository-Manager und den Metadaten aus dem Metadata-Manager zusammensetzen. Daten aus verschiedenen Repositorien können dabei von den Nutzer\*innen in Pakete zusammengefasst und mit Metadaten angereichert werden. Der Ingest-Prozess wird anschließend von den Nutzer\*innen angestoßen, wobei automatisiert das Ingest-Paket generiert wird. Der Prozess sieht vor, dass jeder Ingest kuratorisch überprüft wird (Review), um die Qualität der abgelegten Daten und insbesondere der Metadaten sicherzustellen. Im Anschluss wird der Ingest automatisiert durch das System bei Rosetta eingespielt. Alle Module des *FDOrganizer* werden unter ausschließlicher Verwendung von Open-Source-Software entwickelt. Durch die Abdeckung der meistgenutzten Forschungsdaten-Repositories durch den Repository-Manager wird eine breite Nachnutzung in bayerischen Institutionen ermöglicht.

In der ersten Projektphase wurde die Softwarearchitektur in Grundzügen entwickelt und mit einem ersten

System getestet: Labfolder ist ein digitales Laborbuch, das als SaaS (Software as a Service) genutzt wird. Diese Einschränkung spiegelt typische Herausforderungen bei der Archivierung von Forschungsdaten wider, etwa den komplexen Export von Daten oder die fehlende Datensouveränität bei extern gehosteten beziehungsweise kommerziellen Systemen. Die Schnittstelle wurde über die bereitgestellte REST-API von Labfolder umgesetzt, die Daten werden über diese auf den *FDOrganizer* übertragen und lokal gespeichert. Außerdem wurde die Möglichkeit geschaffen, lokale Dateien vom Rechner der Forschenden in den *FDOrganizer* hochzuladen. Weiterhin wurde der erste Entwurf für den Metadata-Manager entwickelt, inkl. JSON-Datenstrukturen zur Abbildung von Metadaten-Schemata und zur Formatierung von Metadaten in Standardformaten wie z. B. DataCite-XML. Der Metadata-Manager ermöglicht es Forschenden, Metadaten nach verschiedenen Schemata zu erzeugen. Durch die Umsetzung als Data-Driven-Application ist die Erweiterung um zusätzliche Schemata oder Formatierungen für den Export leicht möglich. Der Ingest-Manager schließlich ist in einem ersten Prototyp umgesetzt. Es ist möglich, Datenpakete hierarchisch strukturiert zusammenzustellen sowie mit Metadaten aus dem Metadata-Manager anzureichern und daraus einen Ingest zu erstellen. Nach dem Submit durch die Datenlieferant\*innen wird ein Review-Prozess für das Ingest-Paket angestoßen. Nach erfolgreichem Review und Freigabe durch eine Datenkuratorin oder einen Datenkurator wird automatisiert ein mit Rosetta kompatibles Ingest-Paket generiert und per Hotfolder im NFS auf Rosetta übertragen. Damit ist ein grundlegender Prototyp für eine breit nachnutzbare Pre-Ingest-Lösung im Bereich Forschungsdaten geschaffen, die in der zweiten Projektphase ausgearbeitet und an weitere Systeme angepasst werden soll. Durch einen methodischen Pre-Ingest aus heterogenen Datenquellen wird so die landesweite Nutzung der kooperativen Infrastruktur auf Basis von Rosetta erleichtert, ohne andere Nutzungsmodelle auszuschließen.

#### **Verbundzentrale/Bayerische Staatsbibliothek**

Die Projektpartner in Regensburg und Bayreuth wurden bei der Konzeption und Implementierung der Workflows durch die Verbundzentrale des Bibliotheksverbunds Bayern bedarfsgerecht unterstützt. In Kooperation mit den Projektpartnern wurde durch die Verbundzentrale ein Kostenmodell bzw. Umlagemodell für bestimmte Kostenanteile erarbeitet. Vorarbeiten sind ebenfalls im Bereich Datenschutz erfolgt; der zwischenzeitlich erreichte Stand bei der Umsetzung der Datenschutzgrundverordnung im Bereich der Langzeitarchivierung im Verhältnis der Bayerischen Staatsbibliothek zum Leibniz-Rechenzentrum bildet eine gute Ausgangslage, um im weiteren Projektverlauf zunächst exemplarisch für die Projektpartner die für ihre Rolle

in der angestrebten zentral-dezentralen Infrastruktur erforderlichen datenschutztechnischen Anpassungen und Ergänzungen vorzunehmen. Die Verbundzentrale übernahm im Rahmen des Projekts den Austausch mit anderen konsortialen Rosetta-Anwendern im deutschsprachigen Raum sowie den Austausch im Rahmen der Strategischen Allianz zwischen dem Bibliotheksverbund Bayern und dem Kooperativen Bibliotheksverbund Berlin-Brandenburg. Im Rahmen der derzeit deutschlandweit aufgebauten Nationalen Forschungsdateninfrastruktur (NFDI) ist die Bayerische Staatsbibliothek Co-Applicant im aktuell beantragten Konsortium für die historisch arbeitenden Geisteswissenschaften NFDI4Memory; zudem agiert sie als Participant im Konsortium NFDI4Culture, das sich mit Forschungsdaten zu materiellen und immateriellen Kulturgütern befasst. Auf diese Weise ist die Kommunikation und Kooperation an der Schnittstelle zwischen lokalen und regionalen Aktivitäten einerseits und dem nationalen Großvorhaben NFDI andererseits institutionell sichergestellt.

#### **Arbeitsgemeinschaft Langzeitverfügbarkeit und Vernetzung**

Die Arbeitsgemeinschaft Langzeitverfügbarkeit, die das Projekt in regelmäßigen Sitzungen begleitet, hat die Ausarbeitung einer Rahmen-Preservation-Policy für den Bibliotheksverbund Bayern, eines Musters für institutionelle Preservation Policies sowie einer Mustervereinbarung für die Langzeitarchivierung durch die Projektleitung in Regensburg unterstützt. Die Dokumente liegen im Entwurf vor und werden zeitnah finalisiert und mit den notwendigen Gremien abgestimmt. Es liegt eine erste Version einer Systemlandkarte inkl. einer allgemeinverständlichen Darstellung des OAIS vor, die zwischen Projektgruppe und Arbeitsgemeinschaft abgestimmt ist und die der Vermittlung von Kenntnissen zur Langzeitverfügbarkeit bei Schulungen dienen soll. Auf dieser Grundlage wird die Arbeitsgemeinschaft in Kooperation mit den Projektpartnern zukünftig an bayernweit nutzbaren Schulungskonzepten arbeiten. Um die laufende Kommunikation mit dem bayerischen Forschungsdatenprojekt »eHumanities – interdisziplinär« zu intensivieren, fanden mehrere Vernetzungstreffen beider Projektgruppen sowie von Vertreter\*innen der Arbeitsgemeinschaft Langzeitverfügbarkeit statt. Mögliche Nachnutzungsszenarien und Kooperationen wurden mit Blick auf die zweite Projektphase identifiziert.

#### **Meilensteine der ersten Projektphase (2019–2021)**

Die wesentlichen Meilensteine der ersten Projektphase sind damit erreicht: die Festlegung der im Projekt prototypisch zu bearbeitenden Workflows mit konkreter Anknüpfung an die beteiligten dezentralen Institutionen; die technische Umsetzung der Workflows im zentralen Testsystem; der produktive Ingest von Daten;

die Erarbeitung von Mustervereinbarungen zur digitalen Langzeitverfügbarkeit und die Erarbeitung eines Geschäfts- bzw. Kostenmodells. Die Entwicklung von Schulungsmaterialien und -konzepten wurde begonnen und wird in der zweiten Projektphase weiter ausgebaut.

#### **Die zweite Projektphase (2022–2025)**

Die zweite Projektphase steht insbesondere unter den Gesichtspunkten der *Konsolidierung* (Arbeitspaket 1) und der *Nachnutzung* (Arbeitspaket 2). Im Arbeitspaket 1 *Konsolidierung* geht es zunächst um die Etablierung einer stabilen Governance, einschließlich rechtlicher und datenschutztechnischer Regelungen, als Grundlage für die dauerhafte Nachnutzung der erarbeiteten Lösungen in bayerischen Bibliotheken.<sup>21</sup> Hierzu sind Nutzungsvereinbarungen zwischen den beteiligten Bibliotheken abzuschließen sowie ggf. datenschutztechnische Verarbeitungen mit dem Leibniz-Rechenzentrum anzupassen. Auch die durch die Hochschulen selbst zu schaffenden Voraussetzungen sind hierbei herauszuarbeiten, um die Nachhaltigkeit und Nutzung gerade im Bereich Forschungsdaten zu erhöhen. Dabei soll insbesondere die Auswahl von Daten mit Blick auf strukturelle, rechtliche und finanzielle Anforderungen sowie der Umgang mit verwaisten Daten berücksichtigt werden. Hierbei wird eine enge Zusammenarbeit mit dem Komplementärprojekt »Langzeitverfügbarkeit von Forschungsdaten« (LaVe) (s.u.) unter Federführung der Universitätsbibliothek Erlangen-Nürnberg angestrebt. Darüber hinaus soll die in der ersten Projektphase entworfene Rahmen-Preservation-Policy vom Verbundrat verabschiedet werden, die teilnehmenden Bibliotheken sollen entsprechend institutionelle Policies veröffentlichen. Dem Verbundrat wird zudem das entworfene Kostenmodell vorgelegt, um die Leistungen für die teilnehmenden Bibliotheken auf eine stabile Grundlage stellen zu können. Die weitere Etablierung von Rosetta als bayernweiter Komponente soll zudem auch im Kreis der Chief Information Officers der bayerischen Hochschulen vorangetrieben und die ZDI in die bayerische IT-Strategie integriert werden. Schließlich ist entsprechende Öffentlichkeitsarbeit durchzuführen. In diesem Arbeitspaket sollen darüber hinaus die erarbeiteten Geschäftsgänge und Lösungen mit Blick auf Nachnutzung ausführlich dokumentiert und im Open Access öffentlich verfügbar gemacht werden. Hierzu gehört neben der Prozessdokumentation auch die Erstellung und Veröffentlichung von Vorlagen, Formatempfehlungen (im Bereich Forschungsdaten gemeinsam mit dem Forschungsdaten-Modellprojekt) und Metadatenstandards. Auf dieser Basis wird im Rahmen der Projektlaufzeit eine offizielle Zertifizierung der Infrastruktur bzw. der Geschäftsgänge angestrebt, z. B. nach dem international anerkannten Core Trust Seal.<sup>22</sup>

Im Arbeitspaket 2 *Nachnutzung* werden die erarbeiteten Lösungen für die drei Objektklassen an baye-

rischen Partnerbibliotheken ausgerollt. Im Bereich *Open-Access-Publikationen* soll die erzielte Lösung für die automatisierte Einspielung von OJS-Publikationen in Rosetta an drei bayerischen Partnerbibliotheken (Universitätsbibliotheken Bayreuth, Bamberg, Passau) umgesetzt werden. Hierbei sollen auch die in der ersten Projektphase erarbeiteten Techniken für die Archivierung von Publikationsworkflows und weiteren zur Publikation gehörigen Daten ausgearbeitet und in die Nachnutzung gebracht werden. Im Bereich Digitalisate soll die gefundene Lösung an drei bayerischen Partnerbibliotheken (Universitätsbibliotheken Würzburg und Erlangen-Nürnberg sowie Bayerische Staatsbibliothek München) umgesetzt werden. Zudem sollen die bestehenden Workflows in enger Zusammenarbeit mit der Verbundzentrale durch die Erfahrungen der neuen Projektpartner erweitert werden. Für die Präsentation der digitalen Objekte soll der IIIF-Standard unterstützt werden, den die Bayerische Staatsbibliothek bereits für die Ausgabe und Darstellung ihrer digitalen Sammlungen einsetzt. IIIF steht für International Image Interoperability Framework und ist ein auf Offenheit und Austausch angelegter technischer Rahmen, der von einem weltweiten Zusammenschluss von Kultureinrichtungen getragen wird. Eine IIIF-Schnittstelle öffnet die im Langzeitverfügbarkeitssystem gespeicherten Digitalisate für eine breite, international ausgerichtete Nachnutzung.

Im Bereich Forschungsdaten ist eine Ausarbeitung der bisherigen Lösungen für die Nachnutzung erforderlich: Um den FDOrganizer – aktuell noch auf einem prototypischen Entwicklungsstand – hin zu einem niedrigschwelligen Instrument zu entwickeln, das unkompliziert und breit in der Routine einsetzbar ist, sind weitere Features und Optimierungen erforderlich. Beispielsweise ist im Sinne des Datenschutzes bei der Anbindung weiterer Projektpartner sicherzustellen, dass durch ein Benutzer\*innen-Management eine klare Mandantentrennung im System umgesetzt wird. Weiterhin ist eine Optimierung der Oberfläche des FDOrganizers nach modernen UX/UI (User-Experience, User-Interface)-Standards erforderlich. Um die Anbindung weiterer Projektpartner vorzubereiten, ist es schließlich auch notwendig, den Ingest-Manager zu erweitern, um die je nach Institution unterschiedlichen Anforderungen an den Workflow, die Genehmigungsprozesse und das Benutzer\*innen- bzw. Berechtigungs-Management abbilden zu können. Zudem ist eine Anpassung des Systems in konkreten Kooperationen geplant: So sollen in der zweiten Projektphase in enger Kooperation mit der Universität Regensburg Daten des Leibniz-Instituts für Ost- und Südosteuropaforschung (IOS) in Regensburg aus dem dortigen Forschungsdaten-Repository in die ZDI exportiert werden. Darüber hinaus wurde eine Kooperation mit der Universitätsbibliothek Augsburg vereinbart, die derzeit eine zentrale Forschungsdaten-

infrastruktur aufbaut; sie wird als Vermittler zwischen den Forschenden und der ZDI fungieren, insbesondere in Form von Schulungs- und Beratungsangeboten, und die Umsetzung der notwendigen Prozesse begleiten. Es ist vorgesehen, dass die hier gewonnenen Erkenntnisse und Erfahrungen eine Wissensbasis schaffen, die wiederum für ähnliche Projekte an bayerischen Institutionen (nach-)genutzt werden kann.

Ebenfalls soll die Kooperation mit dem bayerischen Modellprojekt »eHumanities – interdisziplinär« der Friedrich-Alexander-Universität Erlangen-Nürnberg und der Ludwig-Maximilians-Universität München sowie dem Projekt »Langzeitverfügbarkeit von Forschungsdaten« (LaVe) ausgedehnt und Synergien verstärkt werden. So wird die ZDI bei der Umsetzung der Archivierungsschicht jenseits des Zehn-Jahres-Lifecycles unterstützend mitwirken. Weitere gemeinsame Entwicklungspläne beinhalten unter anderem einen gemeinsam erarbeiteten Best-Practice-Guide für Dateiformate. Außerdem soll das geplante DSpace-Repository der Universitätsbibliothek Erlangen-Nürnberg über den FDOrganizer an Rosetta angebunden werden. »Langzeitverfügbarkeit von Forschungsdaten« (LaVe) ist als direktes Komplementärprojekt zum Projekt »Digitale Langzeitverfügbarkeit für Wissenschaft und Kultur in Bayern« konzipiert. Es setzt auf bestehenden Vorarbeiten der ZDI und des Projekts »eHumanities – interdisziplinär« auf und ergänzt sie um Aspekte der Datenkuratierung und Metadatenqualität sowie im Bereich des institutionellen PID-Resolving. Schließlich wird das Projekt »Digitale Langzeitverfügbarkeit für Wissenschaft und Kultur in Bayern« angesichts der wachsenden Bedeutung der anwendungsorientierten Forschung an den Hochschulen für Angewandte Wissenschaften/Technischen Hochschulen auch die stärkere Vernetzung (z. B. durch geeignete Schulungsangebote) über den universitären Bereich hinaus vorantreiben.

Die *Projektplanung* ist wiederum auf drei Jahre ausgelegt: Da in der ersten Projektphase die Stelle in der Verbundzentrale nicht besetzt werden konnte, wurden die Stellen für die zweite Projektphase nochmals für drei Jahre beantragt, um das anvisierte Gesamtprojekt bewältigen zu können. Grundlage der ZDI ist die Kooperation zwischen einem zentralen Infrastrukturangebot mit hocheffizienten Workflows auf der einen Seite und dezentralen Strukturen auf der anderen Seite, die die Schnittstellen für die Standardisierung heterogener Anforderungen darstellen sowie die lokale Datenaufbereitung und die notwendigen Kommunikations- und Rückkopplungsprozesse mit den Datenlieferanten leisten. Diese Architektur wird auch in den Stellenbesetzungen der zweiten Projektphase umgesetzt: Eine Stelle für das Projektmanagement und die Durchführung der Zertifizierung wird bei der Projektleitung in Regensburg besetzt. Die Aufgaben der Stelleninhaberin oder des Stelleninhabers beinhalten insbesondere die in Arbeits-



paket 1 beschriebenen Tätigkeiten, u. a. Durchführung des Zertifizierungsprozesses. Eine Stelle für die zentrale Infrastruktur wird in der Verbundzentrale besetzt. Aufgabe hierbei ist die standardmäßige Umsetzung und ggf. Anpassung der Anbindung der Workflows der beteiligten Institutionen an Rosetta. Schließlich werden zwei Stellen für die Betreuung dezentraler Einrichtungen und die Weiterentwicklung der ZDI im hochkomplexen und heterogenen Bereich Forschungsdaten jeweils an den Universitäten Bayreuth und Regensburg besetzt. Die Aufgaben der Stelleninhaber\*innen sind die Weiterentwicklung des FDOrganizers im Bereich Forschungsdaten, die Anbindung der Projektpartner an diesen sowie die Mitarbeit in der Kooperation mit dem bayerischen Forschungsdaten-Modellprojekt. Dafür ist es erforderlich, die individuellen Anforderungen (z. B. Genehmigungsprozesse der Institutionen, Besonderheiten von Daten- oder personellen Strukturen, etc.) der Institutionen und der anzubindenden Repositorien zu erfassen sowie die Integration in die ZDI systematisch zu planen und auszuführen. Weiterhin werden durch die Stelleninhaber\*innen notwendige Optimierungen und Erweiterungen am FDOrganizer konzipiert und durchgeführt, um die notwendigen Funktionalitäten sowie hohe Nutzerfreundlichkeit für eine breite Akzeptanz des FDOrganizers zu erreichen. Die Ansiedlung einer Stelle in Regensburg soll die intensive Betreuung der Arbeiten für das IOS sowie die enge Anbindung an die Projektleitung sichern.

Das *Projektziel* ist klar artikuliert: Nach der zweiten Projektphase wird in Bezug auf die zwei Objektklassen Digitalisate und Open-Access-Publikationen eine stabile, national konkurrenzfähige Langzeitverfügbarkeitslösung geschaffen sein, die bereits in verschiedenen bayerischen Hochschulen genutzt wird und leicht von weiteren Hochschulen übernommen werden kann. In Bezug auf die sehr heterogene Objektklasse Forschungsdaten werden generische Lösungen bereitgestellt sein, insbesondere im Bereich Metadaten, die die Anwendung für spezifische Bedarfsfälle der Hochschulen ermöglichen. Durch die weitere Integration von Repositorien mit einer bayernweiten Verbreitung wird der Grundstein für ein attraktives Angebot für die Langzeitarchivierung von Forschungsdaten auch für andere Institutionen gelegt. Die ZDI verfügt nach Projektabschluss über eine stabile Governance auf der Basis guter Dokumentation und eindeutiger rechtlicher Regelungen und Vereinbarungen. Die Kostenverteilung ist konsensual beschlossen und nachhaltig realistisch. Die zweite Projektphase wird wie bislang von der Arbeitsgemeinschaft Langzeitverfügbarkeit des Bibliotheksverbunds Bayern begleitet, die sich u. a. verstärkt der Öffentlichkeitsarbeit für die Langzeitverfügbarkeit sowie der Erörterung grundlegender Fragen zur Langzeitverfügbarkeit widmen wird (etwa wie Informationen zu Großereignissen – wie Pandemien – zukünftig optimal archiviert werden können).

## Ausblick

Das Projekt ist für die Schaffung und Etablierung der ZDI in Bayern gut aufgestellt – zum Projektende wird eine zuverlässige Langzeitverfügbarkeits-Infrastruktur für die bayerischen Hochschulen und wissenschaftlichen Bibliotheken geschaffen sein. Die im Projektverlauf etablierte zentral-dezentrale Langzeitarchivierungsarchitektur stellt zugleich das unentbehrliche Scharnier zwischen lokalen Aktivitäten des Forschungsdatenmanagements und dem deutschlandweiten Großvorhaben einer Nationalen Forschungsdateninfrastruktur (NFDI) dar. Voraussetzung für das *langfristige* Funktionieren der ZDI ist jedoch die Schaffung zusätzlicher dauerhafter Stellen für Data Stewards bzw. Digital Preservation Specialists in jeder beteiligten Einrichtung, um den wissenschaftlichen Output systematisch in die ZDI zu bringen; dies ist im Rahmen eines systematischen Forschungsdatenmanagements an den bayerischen Universitäten zu sehen – durch flächendeckende Information über standardisierte Verfahren und offene Formate, durch einheitliche Policies und die laufende Vermittlung der verfügbaren Services sowie durch ständige enge Rückkopplung mit den Bedürfnissen der Wissenschaft<sup>23</sup> wird die Grundlage für eine laufende Verfügbarhaltung gelegt. Hierbei können lokal inhaltliche Schwerpunkte gebildet werden (Kompetenzzentren). Im Rahmen der in der bayerischen IT-Strategie anvisierten *hochschulübergreifenden IT-Dienstleistungen* ist zu prüfen, wie diese Stellen geschaffen werden können. Nur so kann die ZDI mit der dynamischen Entwicklung der Wissenschaftslandschaft und ihrer Produktions- und Publikationsmechanismen nicht nur dauerhaft Schritt halten, sondern stets einen Schritt voraus sein.

## Anmerkungen

- 1 »Da bei den übernommenen Repräsentationen im Laufe der Zeit nicht alle Eigenschaften erhalten werden können, muss eine Auswahl getroffen werden. Diese Teilmenge aller Eigenschaften sind die signifikanten Eigenschaften. [...] Die signifikanten Eigenschaften sollen durch den gesamten Archivierungsprozess hindurch erhalten werden. An ihnen kann bis zu einem bestimmten Grad die Authentizität der aus migrierten oder emulierten Repräsentationen hervorgegangenen Performances festgestellt werden.« (Nestor (2012): *Leitfaden zur digitalen Bestandserhaltung*. Version 2.0. Frankfurt/Main: nestor, 2012. Verfügbar unter: [http://files.d-nb.de/nestor/materialien/nestor\\_mat\\_15\\_2.pdf](http://files.d-nb.de/nestor/materialien/nestor_mat_15_2.pdf), 10).
- 2 »Es wird erforderlich sein, die Bemühungen auf die Kernfunktionen (sogenannte »significant properties«) digitaler Objekte zu konzentrieren, vordringlich auf das, was ihren wesentlichen Informationsgehalt ausmacht.« (Nestor: *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Version 2.3. Göttingen: nestor, 2010. Verfügbar unter: [http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch\\_23.pdf](http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch_23.pdf), Kap. 1:4-5).

- 3 »Sobald sich eine Performance aufgrund einer Migration oder eines Emulators ändert, müssen die signifikanten Eigenschaften der alten Performance mit jenen der neuen Performance verglichen werden. [...] Signifikante Eigenschaften erlauben somit die Bewertung einer Performance [...]. Die Auswahl signifikanter Eigenschaften kann verändert werden (diese Änderungen sollten dokumentiert werden). Neu definierte Nutzungsziele können die Signifikanz gewisser Eigenschaften obsolet machen. Wenn der Erhaltungsprozess weiter gefasst wird, können aber auch Eigenschaften im Laufe des Erhaltungsprozesses neu entstehen. Das Langzeitarchiv muss entscheiden, ob dabei ein neues Informationsobjekt entsteht oder ob dieses als Nachfolgeobjekt des ›Ausgangsobjekts‹ gesehen werden kann.« (Nestor (2012): *Leitfaden zur digitalen Bestandserhaltung*. Version 2.0. Frankfurt/Main: nestor, 2012. Verfügbar unter: [http://files.d-nb.de/nestor/materialien/nestor\\_mat\\_15\\_2.pdf](http://files.d-nb.de/nestor/materialien/nestor_mat_15_2.pdf), 22).
- 4 ISO-Norm 14721:2012(E), 30.
- 5 HOLLMANN, Michael und André SCHÜLLER-ZWIERLEIN, Hrsg., 2014. *Diachrone Zugänglichkeit als Prozess: Kulturelle Überlieferung in systematischer Sicht*. Berlin: de Gruyter. – »Nothing has been preserved, there are only things being preserved. Preservation is the result of ongoing work of people and commitments of resources. The work is never finished.« (Trevor Owens: *The Theory and Craft of Digital Preservation*. Manuscript. 2017 (<https://osf.io/preprints/lissa/5cpjt>), 7), »Digital preservation is about making the best use of your resources to mitigate the most pressing preservation threats and risks. You are never done with digital preservation. It is not something that can be accomplished or finished. Digital preservation is a continual process of understanding the risks you face for losing content or losing the ability to render and interact with it and making use of whatever resources you have to mitigate those risks.« (Ebd., 8).
- 6 Wichtige Elemente dieses sogenannten »Preservation Planning« sind der »Community Watch« und der »Technology Watch« (vgl. (Nestor (2012): *Leitfaden zur digitalen Bestandserhaltung*. Version 2.0. Frankfurt/Main: nestor, 2012. Verfügbar unter: [http://files.d-nb.de/nestor/materialien/nestor\\_mat\\_15\\_2.pdf](http://files.d-nb.de/nestor/materialien/nestor_mat_15_2.pdf), 29–30).
- 7 ISO-Norm 14721:2012(E), 22, 28.
- 8 Vergleichbare kooperative Ansätze stellen das Projekt »Langzeitverfügbarkeit digitaler Inhalte an hessischen Hochschulen« (LaVaH) im HeBIS-Verbund (<https://www.lavah.de/>) sowie die Kooperation der drei Zentralen Fachbibliotheken für Lebenswissenschaften (ZB MED), für Technik und Naturwissenschaften (TIB) und Wirtschaftswissenschaften (ZBW) im Bereich der Langzeitarchivierung dar (<https://www.tib.eu/de/die-tib/policies/preservation-policy-der-drei-zentralen-fachbibliotheken>).
- 9 »Working to ensure long-term access to digital information is not a problem for a single tool to solve.« (Trevor Owens: *The Theory and Craft of Digital Preservation*. Manuscript. 2017 (<https://osf.io/preprints/lissa/5cpjt>), 5), »Software cannot preserve anything. Software cannot be a repository in itself. A repository is the sum of financial resources, hardware, staff time, and ongoing implementation of policies and planning to ensure long-term access to content.« (Ebd., 7).
- 10 ISO-Norm 14721:2012(E).
- 11 ISO-Norm 14721:2012(E), 12.
- 12 Vgl. Nestor: *Grundsätze zur SIP-Bildung. Handreichung der nestor-Arbeitsgruppe SIP-Konkretisierung*. Nestor, 2020. Verfügbar unter: <http://nbn-resolving.de/urn:nbn:de:0008-2020071606>
- 13 »Considerable iteration may be required to agree on the right information to be submitted, and to get it into forms acceptable to the OAIS.« (ISO-Norm 14721:2012(E), 40).
- 14 ISO-Norm 14721:2012(E), 30–34.
- 15 »The OAIS shall [e]nsure that the information to be preserved is *Independently Understandable* to the Designated Community. In particular, the Designated Community should be able to understand the information without needing special resources such as the assistance of the experts who produced the information.« (ISO-Norm 14721:2012(E), 39).
- 16 Bibliotheksverbund Bayern: *Perspektiven der digitalen Langzeitarchivierung in Bayern. Ein Bericht der Arbeitsgruppe Langzeitarchivierung*. Regensburg 2017, 3.
- 17 Bibliotheksverbund Bayern: *Perspektiven der digitalen Langzeitarchivierung in Bayern. Ein Bericht der Arbeitsgruppe Langzeitarchivierung*. Regensburg 2017, 6. Die aktuelle Untersuchung von Barrueco und Termens bestätigt den Mangel an systematischer Diskussion und Planung zur Langzeitarchivierung von Open-Access-Publikationen in institutionellen Repositorien und damit den hohen Handlungsbedarf; vgl. José Manuel Barrueco; Miquel Termens: »Digital Preservation in Institutional Repositories: a Systematic Literature Review«. *Digital Library Perspectives* (2021), <https://doi.org/10.1108/DLP-02-2021-0011>. Die Studie von Laakso et al. zeigt, dass die Gefahr des gänzlichen Verlusts von Open-Access-Publikationen real ist (Mikael Laakso; Lisa Matthias; Najko Jahn: »Open Is Not Forever: a Study of Vanished Open Access Journals«. *Journal of the Association for Information Science and Technology* 72 (2021), 1099–1112, <https://doi.org/10.1002/asi.24460>).
- 18 Siehe <https://pkp.sfu.ca/ojs/>
- 19 [http://vlp.uni-regensburg.de/index\\_html](http://vlp.uni-regensburg.de/index_html)
- 20 Siehe <https://invenio-software.org/>
- 21 »It is essential for an OAIS to have documented policies and procedures [...]. [...] The Producer and Consumer communities should be provided with submission and dissemination standards, policies, and procedures to support the preservation objectives of the OAIS.« (ISO-Norm 14721:2012(E), 43).
- 22 <https://www.coretrustseal.org/>
- 23 »The Designated Community should be monitored to be sure the Content Information is still understandable to them.« (ISO-Norm 14721:2012(E), 44).

## Verfasser



Ralf Bruggbauer, Direktor der Universitätsbibliothek Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Telefon +49 0921 55-3400, [Ralf.Bruggbauer@uni-bayreuth.de](mailto:Ralf.Bruggbauer@uni-bayreuth.de)

Foto: privat



Dr. Klaus Ceynowa, Generaldirektor, Bayerische Staatsbibliothek, Ludwigstraße 16, 80539 München, Telefon +49 89 28638-2201, [ceynowa@bsb-muenchen.de](mailto:ceynowa@bsb-muenchen.de)

Foto: BSB / H. R. Schulz



Dr. André Schüller-Zwierlein, Ltd. Bibliotheksdirektor, Universitätsbibliothek Regensburg, Universitätsstraße 31, 93053 Regensburg, Telefon +49 941 943-3900, [Andre.Schueller-Zwierlein@bibliothek.uni-regensburg.de](mailto:Andre.Schueller-Zwierlein@bibliothek.uni-regensburg.de)

Foto: UB Regensburg