

Infrastrukturen für die digitale Langzeitarchivierung in Archiven, Bibliotheken und Museen schaffen und nachhaltig finanzieren!

Statement von DINI, DBV-Sektion-4, nestor und dem DFG-geförderten Projekt „Nationales Hosting elektronischer Ressourcen“ (September 2015)

Executive Summary

Wie für analoge (gedruckte) Materialien sind auch für digitale Datenbestände und Ressourcen (Audios, Bilder, Bücher, Digitalisate, Forschungsdaten, Texte, Videos, WEB-Inhalte, Zeitschriften etc.) nachhaltige Infrastrukturen zu ihrer langfristigen Archivierung und Verfügbarkeit dringend erforderlich und zentrale Voraussetzung für Forschung, Lehre und Studium aller Fachdisziplinen. An vielen deutschen Archiven, Bibliotheken und Museen bestehen bereits Konzepte zur digitalen Langzeitarchivierung und Langzeitverfügbarkeit, die vielfach auch schon lange spartengerecht und erfolgreich umgesetzt werden. Allerdings sind diese Lösungen oft isoliert mit unterschiedlicher Intensität auf die jeweils langfristig vorzuhaltenden Informationstypen ausgerichtet. Eine übergreifende nationale Infrastruktur, die modular auszugestalten wäre, existiert noch nicht. Angesichts dessen besteht Handlungsbedarf, der in folgender Weise aufgegriffen werden soll. Beim DFG-Projekt „Nat-Hosting“ fokussiert die unmittelbare Sicherung des dauerhaften Zugangs auf lizenzierte Informationsressourcen (E-Books, E-Journals). Für die langfristige Sicherung weiterer forschungs- und lehrrelevanter Materialien (Bildmaterial, Digitalisate, Forschungsdaten, WEB-Inhalte etc.) sind bereits Lösungen produktiv, für die das Kompetenznetzwerk nestor ein Informations- und Austauschforum ist. Doch insgesamt fehlen abgestimmte und übergreifende Infrastrukturen zur langfristigen Verfügbarkeit aller relevanten Informationstypen und Materialien. Unabhängig vom jeweils gewählten technischen Ansatz verbinden sich alle auf Dauer ausgerichteten Lösungen mit einer Reihe organisatorischer, rechtlicher und technischer Vorkehrungen, die zu realisieren und nachhaltig zu finanzieren sind. Die Deutsche Initiative für Netzwerkinformation (DINI), das Kompetenznetzwerk Langzeitarchivierung nestor, die Partner des DFG-geförderten Projekts „Nationales Hosting“ und die Sektion 4 des Deutschen Bibliotheksverbands (DBV) „Wissenschaftlichen Bibliotheken“ sehen die unmittelbare Notwendigkeit, das Thema der digitalen Langzeitarchivierung und Langzeitverfügbarkeit sowie deren Realisierung aufzugreifen, die Hochschulleitungen und alle Träger der betroffenen Einrichtungen für die Notwendigkeit eines Aufbaus der dafür erforderlichen verteilten Infrastrukturen zu sensibilisieren und die damit zusammenhängenden finanziellen und organisatorischen Rahmenbedingungen zu thematisieren.

1. Welche Motivationen haben wir?

Im Zuge der digitalen Transformation entstehen wissenschaftliche Ergebnisse digital, sie werden digital aufbereitet, diskutiert, verbreitet und publiziert. Dieser Wandel verändert die Aufgaben und Rollen im Bereich der Informationsversorgung und –bewahrung mit zum Teil gravierenden Auswirkungen auf einzelne Arbeitsfelder.

So hat sich beispielsweise das herkömmliche Ownership-Modell als zentraler Ansatz der Bestandsentwicklung wissenschaftlicher Bibliotheken zum Access-Modell auf der

Grundlage von Nutzungslizenzen entwickelt. Für Informationsressourcen wie E-Books und E-Journals werden deshalb Nutzungslizenzen von den Verlagen erworben. Auch wenn zur Sicherung der langfristigen Zugänglichkeit digitaler Fachinformation in den Lizenzverträgen von Bibliotheken und Verlagen zunehmend Archivrechte festgeschrieben sind, werden diese zumeist lediglich auf Servern der Verlage in Anspruch genommen. Damit ist eine langfristige Verfügbarkeit über Gedächtnis- und Serviceeinrichtungen, wie Bibliotheken es sind, nicht sichergestellt. Denn zum einen haben Verlage nicht die Aufgabe oder den Auftrag, ihre Produkte langfristig zu archivieren oder zur Verfügung zu stellen, noch haben sie ein genuines Interesse an Langzeitarchivierung und Langzeitverfügbarkeit als Grundlage ihrer wirtschaftlichen Interessen. Zum anderen unterliegen Verlage als Wirtschaftsunternehmen den Einflüssen der Marktentwicklung, die eine langfristige Verfügbarkeit digitaler Bücher und Zeitschriften nicht sicher garantieren kann.

Hier sind also spezifische Maßnahmen erforderlich, sollen diese wichtigen Inhalte, für deren Lizenzierung erhebliche Investitionen getätigt werden, abgesichert werden. In ähnlicher Weise gilt dies auch für viele weitere Materialarten, seien es Forschungs- und Quelldaten (mit ganz anderen Fragen rund um ihre häufig verteilte Entstehung, ihre Analyse, ihre Kontextualisierung, den Zugang auf sie), Publikationen auf Open-Access-Repositoryn, aber auch Artefakte der wissenschaftlichen Diskussion in Blogs oder Wikis. Alle diese verschiedenen Ausprägungen der Arbeit von Wissenschaft und Forschung stellen Aspekte der wissenschaftlichen digitalen Information dar, deren Verfügbarkeit auf Dauer abgesichert werden muss. Um langfristig auf Inhalte digitaler Information zugreifen zu können, was ein Desiderat nahezu aller Wissenschaftsdisziplinen ist, kommen Hochschulen, Forschungseinrichtungen und ebenso Einrichtungen des Kulturerbes deshalb in die Situation, eigene Infrastrukturen für die langfristige Archivierung und Verfügbarkeit digitaler Inhalte nachhaltig aufzubauen und zu betreiben. Dabei geht es nicht nur um das „physische“ Vorhalten der entsprechenden Daten, sondern auch um die langfristige Absicherung ihrer Verfügbarkeit und Nutzbarkeit – über Dateiformatgenerationen und veränderten technische Bedingungen hinweg.

An einer Reihe von deutschen Archiven, Bibliotheken und Museen bestehen bereits Ansätze und Lösungen zur digitalen Langzeitarchivierung und Langzeitverfügbarkeit. Diese Lösungen sind oft spartenspezifisch und mit unterschiedlicher Intensität auf die jeweils langfristig vorzuhaltenden Informationstypen (Audios, Bilder, Digitalisate, Forschungsdaten, Texte, Videos, WEB-Inhalte etc.) ausgerichtet. Ihnen allen gemein ist die langfristige Perspektive, d.h. die regelmäßige Prüfung der Nutzbarkeit durch die Aktualisierung von Formaten oder die Bereitstellung von Emulationsumgebungen. In den vergangenen Jahren ist es auf Basis des Kompetenznetzwerks nestor gelungen, eine durchlässige Informations- und Austauschkultur zwischen den verschiedenen Akteuren und Plattformen aufzubauen, die insbesondere auch technische Werkzeuge kooperativ zur Nutzung bringt; eine systematisch koordinierte, übergreifende und ineinander verwobene nationale Infrastruktur für einen ortsunabhängigen und permanenten Zugriff auf langfristig vorgehaltene Inhalte existiert jedoch noch nicht. An anderen Stellen entstehen für ähnliche Aufgaben Doppelstrukturen, die unterschiedlichen konzeptionellen Ansätzen folgen und zueinander nicht kompatibel sind.

Eine von der AG Nationale Hosting-Strategie der Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen in Auftrag gege-

bene Studie von Charles Beagrie Ltd. aus dem Jahr 2009¹ mündete in 2012 in eine Ausschreibung der Deutschen Forschungsgemeinschaft. Das daraus hervorgegangene Projekt zum nationalen Hosting elektronischer Ressourcen (NatHosting)² verfolgt das Ziel, im Sinne der nachhaltigen und störungsfreien Verfügbarkeit digitaler Informationsressourcen für Wissenschaft, Forschung und Lehre, eine Strategie des doppelten Bodens für lizenzpflichtige Materialien von Drittanbietern zu entwickeln. Hier liegt die Dringlichkeit insbesondere in der raschen Implementierung eines solchen Konzepts, das in wesentlichen Zügen bereits beschrieben ist und Anfang 2016 der DFG vorgelegt werden wird. Vereinbarte Archivrechte werden damit konkret umsetzbar sein und stellen das Material für die langfristige Absicherung des Informationszugriffs durch entsprechende Langzeitarchivierungsmaßnahmen bereit. Auch die Aktivitäten des „DEAL“-Projekts, mit dem auf Initiative der Hochschulrektorenkonferenz (HRK) nationale Lizenzierungen von Zeitschriftenportfolios großer Verlage angestrebt werden, unterstreicht die Notwendigkeit des Aufbaus flächendeckend nutzbare Infrastrukturen für langfristige Archivierung und Verfügbarkeit von E-Books und E-Journals.

2. Welche Ansätze wollen wir verfolgen?

Das hier vorgelegte Statement der Partner adressiert einerseits die unmittelbare Sicherung des Zugangs auf lizenzierte Informationsressourcen (E-Books, E-Journals). Andererseits müssen diese Materialien sowie weitere für Forschung und Wissenschaft relevante Datenbestände und Ressourcen (Audios, Bilder, Digitalisate, Forschungsdaten, Videos, WEB-Inhalte etc.) ebenfalls langfristig gesichert werden, um den Zugriff für künftige Generationen zu gewährleisten. Beide Ansätze greifen direkt ineinander.

a) Projekt NatHosting

Das Konzept des Projekts NatHosting definiert folgende Eckpunkte einer Lösungsstrategie, die des Aufbaus, insbesondere aber auch der dauerhaften Finanzierung bedarf:

- Teilnahme eines nationalen Konsortiums bei dem non-profit-Dienstleister Portico³ zur Absicherung der dort archivierten E-Journals überwiegend größerer Verlage. Ein entsprechendes Angebot liegt vor.
- Aufbau und Betrieb eines nationalen Private LOCKSS Network (PLN) zur sukzessiven Absicherung der relevanten nicht durch Portico abgedeckten Inhalte – mit Schwerpunkt auf tendenziell eher gefährdete „long-tail“-Verlage. Einzurichten ist ein sog. „managed PLN“ mit einigen wenigen verteilten Knotenpunkten, das angesichts der Dimension des Vorhabens über existierende LOCKSS-Ansätze in Deutschland signifikant hinausgehen muss.
- Aufbau und Betrieb einer übergreifenden rechnergestützten Rechteverwaltung
- Einrichtung einer Hosting-Agentur mit Koordinationsaufgaben für das nationale Hosting

Beide genannten Systeme bieten für Inhalte, die für das nationale Hosting relevant sind (E-Journals, E-Books) bereits mit unterschiedlichen Konzepten realisierte

¹ http://www.allianzinitiative.de/fileadmin/user_upload/redakteur/hosting_studie_d.pdf

² <http://www.nathosting.de/display/ND/Home>

³ <http://www.portico.org>

Grundfunktionalitäten der Langzeitarchivierung, sei es als ggf. notwendige Vorab-Migration (bei Portico) oder als „Migration-on-Access“ (bei LOCKSS). Hier ist künftig zu prüfen, ob das vorgesehene PLN durch weitere Werkzeuge zur Sicherung der Langzeitverfügbarkeit der Inhalte ergänzt werden muss.

b) Infrastrukturen zur Sicherung der Langzeitverfügbarkeit

Für den Aufbau nationaler Infrastrukturen für Langzeitarchivierung und Langzeitverfügbarkeit dieser und weiterer Materialien bieten sich unterschiedliche technische Lösungen für die aufzubewahrenden Informationstypen (Audios, Bilder, Digitalisate, Forschungsdaten, Texte, Videos, WEB-Inhalte etc.) der verschiedenen Sparten (Archive, Bibliotheken, Museen etc.) an – s. dazu die im nestor-Wiki verfügbare Übersicht zu nationalen und internationalen LZA-Aktivitäten⁴. Tatsächlich sind bereits eine Reihe von kommerziellen und nicht kommerziellen Lösungen im Einsatz, überwiegend für statische Materialien wie Texte, Bilder oder Audiofiles. Dazu gehören die laufenden Rosetta-Instanzen (BSB München, SLUB Dresden, Goporis/TIB Hannover), das DIAS-System der Deutschen Nationalbibliothek, das DIMAG (Landesarchiv Baden-Württemberg), das Digitale Archiv Nordrhein-Westfalen u. a. Unabhängig von dem jeweils gewählten technischen Ansatz verbinden sich für einen dauerhaften Betrieb alle Lösungen mit einer Reihe organisatorischer, rechtlicher und technischer Vorkehrungen, die zu realisieren und dauerhaft zu finanzieren sind. Das nestor-Positionspapier⁵ adressiert Handlungsbedarf auf folgenden Feldern:

- **Ressourcenausstattung:** Projektfinanzierungen sind für den Anschlag von Maßnahmen dringend erforderlich, genügen aber zur Finanzierung der daraus erwachsenden Daueraufgaben nicht. Hier fehlt es an einheitlichen Maßstäben und Vorgaben.
- **Rechtliche Fragen:** Bestehende Regelungen zu Urheber-, Lizenz- und Persönlichkeitsrechten müssen im Hinblick auf digitale Langzeitarchivierung angepasst werden⁶. Hier existieren bereits konkrete Vorschläge, die transparent in verbindliche Anwendungsstandards gebracht und durchgängig angewandt werden müssen.
- **Verantwortung und Zuständigkeit:** Verantwortungs- und Zuständigkeitsbereiche müssen adäquat ausgestaltet werden. Dabei ist der Selbstorganisation der Wissenschaft ausdrücklich Rechnung zu tragen. Dabei handelt es sich um eine kontinuierliche Abstimmungsaufgabe. Um auf neue Anforderungen, Angebote und Rahmenbedingungen reagieren zu können, bedarf es der laufenden Dokumentation und Koordination der Infrastrukturangebote.
- **Forschung:** Für die langfristige Verfügbarkeit komplexer Informations- und Medientypen wie beispielsweise audiovisuelle Medien, Datenbanken und WEB-Seiten stehen weiterhin Forschungs- und Entwicklungsaufgaben an.
- **Nationale Strategie:** Die zu entwickelnde Informationsinfrastruktur muss die identifizierten Handlungsfelder der digitalen Langzeitarchivierung möglichst umfassend abdecken. Der Prozess zur Entwicklung eines entsprechenden Netzwerks, muss nachhaltig unterstützt und entsprechend ausgedehnt werden. Dafür steht nestor mit der Kompetenz seiner Partner.

⁴ <https://wiki.dnb.de/pages/viewpage.action?pageId=94679349>

⁵ <http://files.dnb.de/nestor/berichte/nestor-Positionspapier.pdf>

⁶ http://files.dnb.de/nestor/berichte/nestor-Stellungnahme_AG-Recht.pdf

Insgesamt weisen die Ansätze und Planungen von nestor in die richtige Richtung: Nur ein gesamtheitlicher Ansatz, der viele unterschiedliche Akteure und Themen vereint, ist in der Lage, eine ausgewogene und abgestimmte Infrastruktur zu initiieren und aufrecht zu erhalten. Die aus den unterschiedlichsten Bereichen stammenden Partner im nestor-Kompetenznetzwerk sind bereit, diesen Schritt in Kooperation mit anderen Trägern der Infrastruktur zu gehen.

Das Projekt NatHosting strebt nach Vorlage des Konzepts bei der DFG eine rasche Implementierung mittels Anschubfinanzierung an, um für die Absicherung lizenzpflichtiger Ressourcen ohne weiteren Zeitverzug Abhilfe schaffen zu können. Für den dauerhaften Betrieb bedarf es aber eines projektunabhängigen und tragfähigen Finanzierungskonzepts. Hier kommen beide Ansätze (NatHosting, Infrastrukturen zur Sicherung der Langzeitverfügbarkeit) zusammen.

Auch perspektivisch ist die Einbindung von NatHosting in einen von nestor angeordneten größeren Kooperationszusammenhang anzustreben, um potentielle Synergien zu erschließen und um bei Bedarf die Verfügbarkeit geeigneter Werkzeuge zur Einleitung von Erhaltungsmaßnahmen sicherzustellen. Auf diese Weise können die Aktivitäten von NatHosting, die auf den Aufbau eines abgesicherten Speicherraums als schnell verfügbarer Ersatz bei Störungen ausgerichtet sind, ggf. um Infrastrukturkomponenten ergänzt werden, die in der Lage sind, notwendige Langzeitarchivierungsmaßnahmen durchzuführen, soweit sie durch die Plattform nicht bereits geleistet werden können. Insgesamt geht es darum, unter Beteiligung der relevanten Kompetenzeinrichtungen tragfähige Strukturen für die langfristige Sicherung der Informationsinfrastruktur zu schaffen.

3. Worin besteht unser Anliegen?

Angesichts der skizzierten Gesamtsituation besteht dringender Handlungsbedarf, um den Aufbau von umfassenden Infrastrukturen zur Langzeitarchivierung und Langzeitverfügbarkeit finanziell, organisatorisch, rechtlich und technisch voranzutreiben und zu realisieren. Dafür ist vor allem im Hinblick auf die Nachhaltigkeitsverpflichtung der zu ergreifenden Maßnahmen ein hohes Maß an politischer Rückendeckung und substantieller Unterstützung der Infrastruktureinrichtungen der deutschen Hochschulen sowie deutscher Gedächtnisinstitutionen außerhalb des Hochschulkontextes erforderlich.

Die Deutsche Initiative für Netzwerkinformation (DINI), das Kompetenznetzwerk Langzeitarchivierung nestor, die Partner des DFG-geförderten Projekts „Nationales Hosting“⁷ und die Sektion 4 des Deutschen Bibliotheksverbands (DBV) „Wissenschaftlichen Bibliotheken“ sehen die Notwendigkeit, mit dem hier vorliegenden Statement die Hochschulleitungen und alle Träger der betroffenen Einrichtungen auf die Notwendigkeit eines Aufbaus der erforderlichen Infrastrukturen und der damit verbundenen, finanziellen und organisatorischen Implikationen hinzuweisen und dafür zu sensibilisieren.

⁷ Bayerische Staatsbibliothek, FIZ Karlsruhe, KIT-Bibliothek mit Steinbuch Computing Center, HU Berlin, UB Erlangen-Nürnberg, UB Frankfurt.