

Dauerhafter Zugriff auf digitale Publikationen – das DFG-Projekt NatHosting

Hildegard Schäffler, Bayerische Staatsbibliothek

Michael Seadle, Institut für Bibliotheks- und Informationswissenschaft (IBI), Humboldt-Universität zu Berlin

Karl-Heinz Weber, FIZ Karlsruhe, Leibniz-Institut für Informationsinfrastruktur

Zusammenfassung:

Das Arbeiten mit digitalen Veröffentlichungen und Daten in Forschung, Lehre und Studium erfordert eine kontinuierliche und möglichst störungsfreie Verfügbarkeit dieser Inhalte. Bei Publikationen, die von wissenschaftlichen Fachverlagen lizenzpflichtig vertrieben werden, können verschiedene Typen von Störungsfällen wie temporäre oder längerfristige Ausfälle von Verlagsservern, der Transfer von Zeitschriftentiteln zu anderen Verlagen, das komplette Ausscheiden von Verlagen aus dem Markt oder die Abbestellung der elektronischen Ressourcen durch die lizenznehmende Einrichtung zu Zugriffsunterbrechungen führen. Das DFG-Projekt „Nationales Hosting elektronischer Ressourcen“ (NatHosting) untersucht, wie unter Berücksichtigung bereits existierender Lösungsansätze eine nationale Strategie zur Absicherung des dauerhaften Zugriffs auf digitale Publikationen entwickelt werden kann. Das vom Projekt erarbeitete Lösungskonzept umfasst eine zweigleisige Strategie der Teilnahme eines nationalen Konsortiums bei Portico bei gleichzeitigem Aufbau eines „Private LOCKSS Networks“ für die durch Portico nicht abgedeckten Inhalte. Weitere Komponenten sind der Aufbau eines rechnergestützten Rechtemanagements und die Einrichtung einer „Hosting-Agentur“ mit Koordinationsaufgaben für das nationale Hosting.

Summary:

In order to work with digital publications and data in a reliable and sustainable way, researchers need to be sure that access to and immediate availability of the content is provided continuously and free of faults. An interruption of access to licensed content can occur for various reasons such as temporary or long-term breakdown of publishers' servers, the transfer of journals to another publisher, the withdrawal of publishers from the market or the termination of electronic resource subscriptions by the licensee. Against the background of existing solutions, the project NatHosting (National Hosting of Electronic Resources) aims at formulating a national strategy to solve the issue of ensuring perpetual access to licensed content. The project proposes a twin-track approach with both Portico membership of a national consortium as well as the implementation of a Private LOCKSS Network for content not covered by Portico. Other components of the concept would include computer-aided rights management as well as the establishment of a hosting agency for coordination tasks.

Zitierfähiger Link (DOI): <http://dx.doi.org/10.5282/o-bib/2015H4S279-284>

Autorenidentifikation: Schäffler, Hildegard: GND 118181408

Seadle, Michael: GND 1034760815

Schlagwörter: Dauerhafter Zugriff; E-Ressourcen

1. Hintergrund

Hintergrund des Projektes „NatHosting – Nationales Hosting elektronischer Ressourcen“ ist die Herausforderung, dass Wissenschaftlerinnen und Wissenschaftler nur dann verlässlich und nachhaltig mit digitalen Veröffentlichungen und Daten arbeiten können, wenn sichergestellt ist, dass der Zugriff auf diese Inhalte kontinuierlich gewährleistet ist. Bei digitalen Publikationen, die von wissenschaftlichen Verlagen lizenzpflichtig vertrieben werden, können verschiedene Typen von Störungsfällen wie temporäre oder längerfristige Ausfälle von Verlagsservern, der Transfer von Zeitschriftentiteln zu anderen Verlagen oder auch das komplette Ausscheiden von Verlagen aus dem Markt zu Zugriffsunterbrechungen führen. Ferner möchte eine Bibliothek auch bei Abbestellungen von digitalen Zeitschriften ihren Nutzern den Zugriff auf die einmal abonnierten Jahrgänge ermöglichen (Post-Cancellation Access, PCA). Für derartige Störungen des Normalbetriebs ist die Entwicklung von Lösungen erforderlich. Diese sollten sich möglichst in eine deutschlandweit ausgerichtete Gesamtstrategie einbinden lassen, so dass sich nicht jede Bibliothek selbst um die dauerhafte Zugangssicherung kümmern muss, sondern sich an einer gemeinsamen Nutzung von Diensten beteiligen kann.

Aufgabe des Projektes ist es, eine nationale Hosting-Strategie zu erarbeiten, die den dauerhaften Zugriff auf lizenzierte elektronische Ressourcen (vor allem E-Journals und E-Books) sicherstellt, und die vorgeschlagene Lösung in ihrer lizenzrechtlichen, technischen, organisatorischen und finanziellen Dimension zu beschreiben. Mit „Hosting“ ist hierbei gemeint, lizenzierte Inhalte technisch unabhängig vom Verlag zu speichern und unter definierten Bedingungen verfügbar zu machen. Es setzt nicht die Gewährung des primären Zugangs auf diese Inhalte voraus. Es wird vielmehr die Schaffung eines „Doppelten Bodens“ angestrebt, ein Hintergrunddienst, der den langfristigen Zugriff auf lizenzierte Ressourcen sicherstellt.

Das Thema „Nationales Hosting“ wird bereits seit Beginn der Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen im Jahre 2008 diskutiert. Eine in diesem Zusammenhang in Auftrag gegebene Studie,¹ durchgeführt von Charles Beagrie Ltd., lieferte einen umfassenden Überblick über die Problematik und vereinzelt bestehende Lösungsansätze. In einem Folgeschritt soll nun das aktuelle, von der Deutschen Forschungsgemeinschaft (DFG) ausgeschriebene Projekt ein rasch umsetzbares Lösungskonzept zur Absicherung der in Deutschland lizenzierten digitalen Publikationen entwickeln. Die Partner im Projekt NatHosting (Laufzeit 02/2014 – 01/2016) sind die Bayerische Staatsbibliothek, FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur, die Universitätsbibliothek Johann Christian Senckenberg Frankfurt/M., das Karlsruher Institut für Technologie (KIT) – Bibliothek und Steinbuch Centre for Computing (SCC), die Humboldt-Universität zu Berlin – Institut für Bibliotheks- und Informationswissenschaft sowie die Universitätsbibliothek Erlangen-Nürnberg.

1 Charles Beagrie Ltd. (Hg.): Dauerhaften Zugriff sicherstellen. Auf dem Weg zu einer nationalen Strategie zu Perpetual Access und Hosting elektronischer Ressourcen in Deutschland. Februar 2010. www.allianzinitiative.de/fileadmin/user_upload/redakteur/hosting_studie_d.pdf (13.10.2015).

2. Bestehende Dienste

An bereits bestehenden Diensten und technischen Systemen für die Gewährung eines dauerhaften Zugriffs auf Verlagspublikationen sind vor allem Portico² und LOCKSS³ zu nennen, die jeweils einen unterschiedlichen Ansatz verfolgen. Portico ist ein in Princeton (NJ) ansässiger Dienst, der – wie auch JSTOR – zur Non-Profit-Institution ITHAKA (NY) gehört. Portico archiviert seit 2005 im Auftrag von Verlagen deren digitale Publikationen zentral in einheitlichem Format und ermöglicht teilnehmenden Bibliotheken in definierten Fällen – wozu ganz überwiegend auch Post-Cancellation Access (PCA) gehört – den Zugang dazu. Neben dem zentralen Produktionssystem existieren mehrere geographisch verteilte Backup-Systeme. Portico finanziert sich über Beiträge der Verlage und der teilnehmenden Bibliotheken (aktuell mehr als 920).

LOCKSS (Lots of Copies Keep Stuff Safe) ist dagegen eine von der Stanford University frei verfügbare Software zum Harvesten von Inhalten von Verlags-Webseiten und Speichern dieser Inhalte in einem Netzwerk verteilter Server, sog. LOCKSS-Boxen. Das System wird seit 2004 vom LOCKSS-Team in Stanford (CA) z.B. für den Betrieb des Global LOCKSS Networks (GLN) eingesetzt, über das inzwischen viele Verlage ihre (vorwiegend Zeitschriften-)Inhalte regelmäßig harvesten und im GLN speichern lassen. Sobald für einen Nutzer ein Dokument auf dem Verlagsserver nicht verfügbar ist, wird er auf die entsprechende Version in der LOCKSS-Box geführt. Jede im GLN teilnehmende Bibliothek muss dabei aber eine eigene LOCKSS-Box betreiben. Die gleiche Software lässt sich jedoch auch für sog. Private LOCKSS Networks (PLN) einsetzen, mit denen Einrichtungen in eigener Verantwortung Inhalte in einem Netzwerk redundant speichern und damit gegen Zugriffsverlust absichern können. Das bekannteste PLN ist CLOCKSS⁴ (Controlled LOCKSS), ein Netzwerk bestehend aus derzeit 12 LOCKSS-Knoten, das E-Journals vieler, auch großer Verlage harvestet und ihre fortdauernde Verfügbarkeit bei kompletter Beendigung des entsprechenden Verlagsangebotes sicherstellt. Damit gehört Post-Cancellation Access oder Zugang bei vorübergehenden Störungen nicht zu den Diensten von CLOCKSS. CLOCKSS finanziert sich ebenfalls über Verlage und Bibliotheksbeiträge, das GLN dagegen ausschließlich über die Beiträge der in der LOCKSS Alliance vertretenen Bibliotheken.

Beide Dienste bzw. Systeme, Portico und LOCKSS, wurden im NatHosting-Projekt intensiv im Hinblick auf ihre technischen, organisatorischen und finanziellen Eigenschaften bzw. Anforderungen sowie auf Einsatzmöglichkeiten innerhalb einer zu entwickelnden Hostinglösung hin untersucht. Bezüglich der Inhalte lässt sich feststellen, dass Portico gerade bei größeren Verlagen (z.B. > 50 Zeitschriften) eine gute Abdeckung hat. LOCKSS ist dagegen mit dem GLN bei kleineren, also den „Long-Tail“-Verlagen recht erfolgreich.

Auf vereinzelt in Deutschland schon bestehende Hosting-Aktivitäten soll hier nicht eingegangen werden, da sie jeweils nur einen vergleichsweise geringen Teil der insgesamt abzusichernden Lizenzen, die im Projekt detailliert quantifiziert wurden, umfassen.

2 www.portico.org

3 www.lockss.org

4 www.clockss.org

3. Lösungskonzept

Aufgrund der spezifischen Eigenschaften von Portico und LOCKSS schlägt das Projekt als Lösung eine zweigleisige Strategie vor:

1. Teilnahme eines nationalen Konsortiums bei Portico zur Absicherung der dort archivierten E-Journals und E-Books
2. 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.

Die Nutzung dieser beiden Komponenten ist zu ergänzen durch

3. Aufbau und Betrieb einer rechnergestützten Rechteverwaltung, die an das PLN (und perspektivisch auch an Portico) zu koppeln ist, sowie
4. Einrichtung einer „Hosting-Agentur“ mit Koordinationsaufgaben für das nationale Hosting.

Mit einer Teilnahme bei Portico kann ein signifikanter Anteil lizenzierter Zeitschriften für folgende Fälle sehr rasch abgesichert werden (Portico-Notation):

- Verlag hat seinen Dienst eingestellt
- ein Zeitschriftentitel wird nicht länger angeboten
- ältere Bestände einer Zeitschrift sind nicht mehr verfügbar (für mehr als 90 Tage)
- ein technischer Ausfall (z.B. aufgrund einer Katastrophe) des Verlagsservers (für mehr als 90 Tage)
- Kündigung einer Zeitschrift (spätestens 30 Tage nach Anforderung bei Verlag, wenn der Verlag grundsätzlich Portico als PCA-Lösung vorsieht).

Es bestehen – neben einer individuellen Teilnahme einzelner Bibliotheken bei Portico – grundsätzlich zwei Teilnahmeoptionen: der Abschluss einer Art „National-Lizenz“ bei Portico für praktisch alle potentiell interessierten deutschen Bibliotheken zu einem zu vereinbarenden jährlichen Pauschalbeitrag oder ein „Opt-in“-Modell, wobei ein Konsortium aus Bibliotheken, die diesem explizit beitreten, eine Teilnahme bei Portico abschließt. Hierbei wird der jährliche Konsortialbeitrag von der genauen Zusammensetzung des Konsortiums abhängen. Das Projekt ist derzeit noch dabei, die jeweiligen Konditionen gemeinsam mit Portico zu prüfen.

Parallel zu dieser Portico-Teilnahme soll ein nationales „Private LOCKSS Network“ (PLN) zur sukzessiven Absicherung der relevanten, *nicht* durch Portico abgedeckten Inhalte aufgebaut werden. Das dazu notwendige Einverständnis der betreffenden Verlage ist zuvor von diesen einzuholen. Auch müssen die Verlage bestimmte technische Vorkehrungen treffen, damit die Web-Inhalte von einem LOCKSS-Netzwerk geharvestet werden können.

Um – anders als bei GLN – darauf verzichten zu können, dass jede teilnehmende Bibliothek eine eigene LOCKSS-Box betreiben muss, soll der Betrieb des Netzwerks über eine kleine Zahl von LOCKSS-Knoten erfolgen. Aus technischen und theoretischen Überlegungen heraus sollte ein PLN aus mindestens 5–6 Knoten mit identischen Inhalten bestehen, da hierdurch eine durch die

LOCKSS-Software automatisch durchgeführte, kontinuierliche Überprüfung und (bei Bedarf) Wiederherstellung der Integrität der Daten gewährleistet ist. Dies macht jedoch den Aufbau einer rechnergestützten Rechteverwaltung erforderlich, die an das PLN (und perspektivisch auch an Portico) zu koppeln ist. Man spricht dann von einem „Managed PLN“. Die Konzeption dieser Rechteverwaltung erfolgt in enger Abstimmung mit einem weiteren DFG-geförderten Projekt, das sich den Aufbau eines überregionalen „Electronic Resource Management System“ (ERMS) zum Ziel gesetzt hat. Das Rechtemanagement verfügt letztlich über die Information, welche Einrichtung zu welcher Publikation über welchen Zeitraum Zugriffsrechte hat bzw. hatte (Letzteres relevant für PCA). Die Daten aller relevanten Lizenzen (National-, Allianz-, weitere Konsortial- und lokale Lizenzen) werden dazu zentral erfasst bzw. dem zentralen System zugeliefert. Um im Anwendungsfall rechtliche Auseinandersetzungen zu vermeiden, müssen die Verlage selbstverständlich von den im System hinterlegten Rechten Kenntnis haben.

In Großbritannien gibt es ebenfalls ein vergleichbares Projekt (SafeNet),⁵ das ein solches „managed PLN“ zur Archivierung von E-Journals zum Ziel hat. Es wird von EDINA⁶ und Jisc⁷ durchgeführt; ein Hosting-Dienst soll 2016 in Betrieb gehen. Das Projekt NatHosting steht mit SafeNet in engem Kontakt und Erfahrungsaustausch.

Die mit dem Aufbau und Betrieb eines solchen Gesamtsystems (Portico-Teilnahme + Betrieb eines PLN) erforderlichen Koordinations- und Managementaufgaben sollten von einer sog. Hosting-Agentur erledigt werden, die neu zu gründen bzw. die an einer geeigneten bestehenden Einrichtung anzusiedeln ist. Zu den Aufgaben dieser Agentur sollten die Koordination der Contentauswahl und die Contentpriorisierung für das PLN gehören, die Verhandlungen mit Verlagen zur Teilnahme am PLN für die Lizenzen der teilnehmenden Bibliotheken, die Koordination des inhaltlichen Aufbaus der „Entitlement Registry“ in Kooperation mit den Bibliotheken, Verhandlungen/Vertragsabschlüsse mit Portico und ggf. der Stanford University (LOCKSS) in Bezug auf technische Dienstleistungen, Beauftragung eines technischen Betreibers des PLN inklusive Rechtemanagement bzw. der Knotenpunkte, an denen die Daten gehalten werden, Abstimmung/Verhandlungen mit möglichen finanziellen Förderern sowie die Information und Kommunikation mit teilnehmenden und interessierten Bibliotheken.

4. Finanzierung

Der Aufbau und der Betrieb des angestrebten Dienstes zur nachhaltigen Absicherung des Zugriffs auf einmal lizenzierte digitale Publikationen, der allen Wissenschaftseinrichtungen zugutekommen soll, muss freilich auch finanziert werden. Das Projektkonsortium hat Kostensondierungen bei Portico vorgenommen und auch die absehbaren Aufwände für das genannte PLN, das Rechtemanagement und die Hosting-Agentur grob abgeschätzt. Für die Aufbauphase des PLN mit Rechtemanagement wird zunächst eine Anschubfinanzierung benötigt. Da ein wie hier beschriebener Dienst im Zuge

5 <http://safenet.blogs.edina.ac.uk/>

6 <http://edina.ac.uk/>

7 <https://www.jisc.ac.uk/>

der fortschreitenden Digitalisierung und des damit einhergehenden Umstiegs vieler Bibliotheken auf rein elektronische Publikationen prinzipiell für alle Universitäten und außeruniversitären Forschungseinrichtungen eine wichtige Infrastruktur darstellt, hält das Projekt-Konsortium eine strukturelle Finanzierung des Dauerbetriebs durch Bund und Länder für gerechtfertigt. Sollte eine solche Finanzierung nicht oder nicht ausreichend zu erreichen sein, ist stattdessen bzw. ergänzend eine finanzielle Beteiligung der von der Lösung profitierenden wissenschaftlichen Bibliotheken notwendig. Angesichts der Gesamtzahl der Einrichtungen, bei denen potentieller Bedarf angenommen werden kann, sollten die resultierenden Beiträge für eine einzelne Institution verkraftbar sein. Es geht schließlich um die Absicherung von Investitionen, die in der Summe ein Vielfaches höher sind als die Kosten des dazu notwendigen Infrastrukturbetriebes.

Literaturverzeichnis

- Charles Beagrie Ltd. (Hg.): Dauerhaften Zugriff sicherstellen. Auf dem Weg zu einer nationalen Strategie zu Perpetual Access und Hosting elektronischer Ressourcen in Deutschland. Februar 2010. www.allianzinitiative.de/fileadmin/user_upload/redakteur/hosting_studie_d.pdf (13.10.2015).