

## Hosting – ein doppelter Boden für den sicheren Zugang zu wissenschaftlichen Informationsressourcen

Textbasierte wissenschaftliche Informationsressourcen (z. B. E-Journals und E-Books) sind zunehmend ausschließlich in elektronischer Form vorhanden, umfangreiche Bestände nur in Papierform vorliegender Materialien werden digitalisiert und zugänglich gemacht. Der Begriff »Hosting« meint in diesem Zusammenhang das Bereithalten dieser Ressourcen für den unmittelbaren, sofortigen Zugriff, wie er z. B. über Verlagsserver realisiert wird. Die Lizenzen, die mit den Verlagen verhandelt werden, ermöglichen den Nutzern für einen bestimmten Zeitraum den Zugriff auf die jeweils lizenzierten Bestände. Der Zugriff auf diese Bestände kann aus verschiedenen Gründen gestört bzw. beendet werden – mit schwerwiegenden Folgen für Wissenschaft und Forschung. Damit dieser Fall nicht eintritt, bedarf es der Einrichtung eines »doppelten Bodens« zur Absicherung des geregelten Zugriffs auf die erworbenen Ressourcen. Dieser Beitrag befasst sich mit den sich daraus ergebenden Entwicklungen und Herausforderungen, die nicht nur technischer, sondern auch organisatorischer, politischer und rechtlicher Natur sind.

Text-based scholarly information resources are being made available increasingly only in electronic formats (e.g., eJournals and eBooks). Extensive collections of existing materials in paper format are being digitalized and re-published. In this connection the term »hosting« is used to indicate immediate access, such as through a publisher's server. The licenses which have been negotiated with the publisher allow the user access to specific collections for a specific period of time. The fact that access may be interrupted or ended due to a variety of reasons has serious consequences for research and teaching. To prevent the situation ever occurring, there is need for a »double floor« to secure orderly access to resources once acquired. This article deals with the resulting developments and challenges, which are not only technical, but also organizational, political and legal in nature.

Der Begriff »Hosting« wurde im wissenschaftlichen Umfeld in der Vergangenheit fast ausschließlich im Zusammenhang mit dem Angebot von fachspezifischen Datenbanken verwendet. Durch die vielfältigen Entwicklungen des Internets, die zu immer mehr digital verfügbaren Informationen unterschiedlichster Ausprägung führen, gewinnt auch das Hosting im wissenschaftlichen Bereich zusätzliche Bedeutung. Textbasierte wissenschaftliche Informationsressourcen (Zeitschriften, Bücher, Berichte, Dokumentationen etc.) sind zunehmend ausschließlich in elektronischer Form vorhanden, umfangreiche Bestände nur in Papierform vorliegender Materialien werden digitalisiert und zugänglich gemacht. Diese Entwicklungen machen den Aufbau und die Bereitstellung leistungsfähiger Infrastrukturen zur Gewährleistung eines dauerhaften Zugriffs auf diese digitalen Informationsressourcen und zur Sicherung der damit verbundenen Investitionen dringend erforderlich. Vermehrt müssen auch rein digitale Informationsobjekte (z. B. Forschungsdaten) auf entsprechenden Systemen zugreifbar und nachhaltig verfügbar gehalten werden. Für das traditionell ge-

prägte Zusammenspiel zwischen Anbietern und Nutzern digitaler Inhalte ergeben sich neue Anforderungen sowie neu zu gestaltende Prozesse, wodurch neue Herausforderungen auf die Informationseinrichtungen zukommen.

Dieser Beitrag konzentriert sich auf die Entwicklungen und Herausforderungen der Informationsversorgung, die im Zusammenhang mit der nachhaltigen Bereitstellung digitaler wissenschaftlicher Zeitschriften und Bücher (E-Journals und E-Books) wichtig sind, denn hier stellen sich dringend zu lösende Fragen. Wenn in dem Beitrag von Hosting gesprochen wird, dann ist damit das Bereithalten von Inhalten für den unmittelbaren, sofortigen Zugriff (Access) gemeint. Eng damit verbunden ist das Thema der Langzeitarchivierung (LZA): Dieser Begriff wird für die Aktivitäten verwendet, die den verlässlichen Zugriff auf digitale Bestände so lange wie möglich und unabhängig von technischen Problemen und / oder technologischen Veränderungen gewährleisten. Hosting dient also dem direkt verfügbaren, schnellen Zugriff, LZA dem langfristigen Erhalt der digitalen Ressourcen. Auf die komplexeren Zusammenhänge der LZA kann hier nur am Rande eingegangen werden. Entsprechende Prozesse und Systeme sind derzeit in der Entwicklung und sollten sich in Hosting-Plattformen integrieren.<sup>1</sup>

### DIE DAUERHAFTHE ABSICHERUNG DES ZUGANGS ALS AKTUELLE HERAUSFORDERUNG

In der Entwicklung der wissenschaftlichen Kommunikation der letzten 15 Jahre haben sich Wissenschaftler, Verlage und Bibliotheken zweifellos immer stärker auf die elektronischen Medien ausgerichtet. Die generellen Möglichkeiten, das Verhalten aller Akteure und auch die Geschäftsmodelle haben sich gravierend verändert. Was jedoch nach wie vor fehlt, ist eine Absicherung der Zugriffsrisiken, wie sie bei gedruckten Medien durch den Erwerb und die Speicherung durch die Bibliotheken gegeben war. In einem Strategiepapier von 2005 heißt es dazu: »For electronic journals, the academy has as yet no functional equivalent in long-term maintenance and control over the scholarly record that ›owning a copy‹ provided for printed journals.«<sup>2</sup> Die weithin erkannte und anerkannte Herausforderung besteht also im Aufbau einer zuverlässigen und verlagsunabhängigen Infrastruktur zur Absi-



Leni Helmes



Ralf Schimmer

Zugriffsrisiken

cherung des Zugriffs auf die (kommerziellen) elektronischen Informationsressourcen und damit auch der oft sehr erheblichen Investitionen in diese Ressourcen.

Wissenschaftlerinnen und Wissenschaftler als Nutzer elektronischer Informationsressourcen erwarten stabile und dauerhafte Zugriffsmöglichkeiten über performante, einfach zu bedienende und leicht auffindbare Rechercheinstrumente. In der Regel organisieren Informationseinrichtungen wie die Bibliotheken an den Hochschulen und außeruniversitären Forschungseinrichtungen die Informationsversorgung und stellen die entsprechenden Zugangssysteme zur Verfügung. Szenarien wie Serverausfälle, Veränderungen in den Geschäftsmodellen der Verlage, Einstellungen von Zeitschriften oder auch Ereignisse wie Brände, Erdbeben oder sonstige Zerstörungen, die in vordigitalen Zeiten durch die zum Teil auch kooperativ abgestimmte Magazinierungspraxis und Schwerpunktbildung von Bibliotheken aufgefangen wurden, dürfen auch im elektronischen Kontext nicht zu einer Versorgungsunterbrechung führen. Zur Absicherung des Zugriffs und der getätigten Investitionen in die Informationsressourcen sind im Hintergrund nachhaltige Infrastrukturen aufzubauen, die mittel- und kurzfristig bei Systemausfällen und sonstigen Zugangsverlusten einspringen und langfristig dauerhaft die Versorgung mit Inhalten und den darauf aufsetzenden Services als Grundlage der täglichen Arbeit weiter gewährleisten.

Zur Sicherung der permanenten und nachhaltigen Zugänglichkeit der digitalen Ressourcen sind nicht nur technische, sondern auch vielfältige organisatorische, politische und rechtliche Herausforderungen zu bewältigen. Es ist daher sinnvoll, in Hinblick auf die verfügbaren Handlungsoptionen zwischen lizenzpflichtigen und lizenzfreien Ressourcen (bzw. Ressourcen in fremder und in eigener Verantwortung) zu unterscheiden:

- Der Zugriff auf lizenzpflichtige E-Journals und E-Books erfolgt überwiegend über Server der Verlage, die als Eigentümer der Daten mit eigenen Interfaces und Retrieval-Werkzeugen ihre Inhalte zugänglich machen und unter Kontrolle halten – das Hosting liegt in fremder Verantwortung.

- Zugriffe auf retrodigitalisiertes, lizenzfreies Material erfolgen über die Server der digitalisierenden Institutionen – das Hosting liegt in eigener Verantwortung.

Die Lizenzen, die mit den Verlagen verhandelt werden, ermöglichen den Nutzern für einen bestimmten Zeitraum den Zugriff auf die jeweils lizenzierten Bestände bei den Verlagen. Dieses Vorgehen unterscheidet sich wesentlich von dem in der reinen »Papierwelt«: Hier werden gedruckte Zeitschriften von den Bibliotheken

erworben, sie erhalten eine entsprechende physische Kopie, um deren Zugänglichkeit und langfristige Verfügbarkeit sie sich ab dem Erwerb in eigener Verantwortung kümmern. Einmal erworbene gedruckte Materialien hat man also weiterhin vor Ort verfügbar – für ausschließlich digitale Materialien gilt dies nicht ohne Weiteres, selbst wenn man sich entsprechende Rechte gesichert hat.

Der Zugriff auf die Verlagsserver kann aus verschiedenen Gründen gestört bzw. beendet werden, beispielsweise:

- Kündigung von Verträgen oder Vertragsbestandteilen. Hier stellt sich dann die Frage, wie es mit dem weiteren Zugriff auf die bis dahin lizenzierten Bestände aussieht (»post-cancellation access«).
- Der Verlag bietet bestimmte Inhalte (einzelne Titel oder Jahrgänge) nicht mehr auf seinem Server an, etwa aufgrund eines Wechsels der betroffenen Titel zu anderen Anbietern. Dabei können leicht auch Inhalte verloren gehen, für die man glaubte, »post-cancellation access« gesichert zu haben.<sup>3</sup>
- Der Verlag / Anbieter stellt den Betrieb ein.
- Es liegt ein technischer (zumeist temporärer) Totalausfall des Serverbetriebs beim Anbieter vor.
- Das Verlagshosting entspricht qualitativ nicht den Ansprüchen der Nutzer bzw. der Verlag bietet (auf Dauer) kein eigenes Hosting an.

Es ist zwar immer häufiger möglich, für die lizenzierten digitalen Inhalte auch dauerhafte Zugangsrechte eingeräumt zu bekommen – eben »post-cancellation access« oder auch »perpetual access«. Diese dauerhaften Zugangsrechte werden typischerweise über die Verlagsserver realisiert (Zugangsmöglichkeiten auf zurückliegende Jahrgänge, Archive). Viele Verlage und Fachgesellschaften – darunter so wichtige Anbieter wie das Institute of Electrical and Electronics Engineers (IEEE) oder die Electrochemical Society (ECS) – verweigern jedoch die Ausübung dauerhafter Zugriffsrechte auf ihren Servern und liefern statt dessen die bezahlten Jahrgänge der Zeitschriften nur auf DVD aus. Kaum eine Einrichtung dürfte damit jedoch etwas anfangen können, um diese Inhalte dann auch tatsächlich ihren Nutzern effizient zugänglich zu machen. Die Ausbedingung von dezidierten Hosting-Rechten in den Lizenzverträgen ist zwar eine Option, verharrt aber eher im Bereich des Theoretischen, denn längst nicht alle Verlage sind dazu bereit oder in der Lage, die dafür notwendige Distributionsinfrastruktur aufzusetzen.<sup>4</sup> Problematisch ist allerdings auch hier, dass nur wenige Einrichtungen finanziell und technisch in der Lage sind, ein adäquates System zur Realisierung eines eigenverantwortlichen Hosting aufzubauen und zu betreiben.

In Deutschland gibt es zwar wichtige Projekte zur Langzeitarchivierung und auch eine Reihe von lokalen oder regionalen Hosting-Aktivitäten. Eine nationale Strategie im Sinne einer klaren Zielsetzung, eines gemeinsamen Vorgehens, der Entwicklung und Einhaltung gewisser Standards und Richtlinien sowie des Aufbaus und Betriebs einer nachhaltigen zentralen oder verteilten Infrastruktur in Bezug auf die Speicherung von Verlagsdaten steht allerdings noch aus. In diese Lücke stößt die Allianz der deutschen Wissenschaftsorganisationen mit ihrer Arbeitsgruppe Hosting vor. Sie ist bestrebt, eine leistungsfähige Infrastruktur zum Speichern digitaler Volltexte im Sinne von Sicherung und prinzipieller Verfügbarhaltung für die aktuelle Nutzung vorzubereiten, um darüber den nachhaltigen Zugriff auf elektronische Veröffentlichungen und Daten abzusichern. Hosting im Sinne der angestrebten bundesweit abgestimmten Strategie beschreibt also die Absicherung des Zugriffs für den Fall, dass der Normalbetrieb nicht mehr funktioniert. Es geht darum, einen doppelten Boden einzuziehen, um den ununterbrochenen Zugriff auf die einmal erworbenen Inhalte sicherzustellen. Die dafür notwendigen neuen Strukturen sind nach Überzeugung der Allianz-Organisationen von ähnlich grundlegender Bedeutung wie der Aufbau elektronischer Bibliotheksdatenbanken und Fachinformationssystemen in früheren Jahren.

### **AKTIVITÄTEN DER ALLIANZ-AG HOSTING**

Die Aktivitäten der Allianz-Arbeitsgruppe Hosting waren von Anfang an konsequent auf einen einrichtungsübergreifenden, von Konsens und Abstimmung getragenen sowie potentiell allen deutschen Wissenschaftseinrichtungen offen stehenden Lösungsansatz hin ausgerichtet. Zielsetzung der Arbeit war es, eine »nationale Hostingstrategie« im Sinne eines bundesweit koordinierten Konzepts zu entwickeln, wie es parallel auch in den Arbeitsgruppen zu Lizenzierung und Open Access angestrebt wird. Dabei wurde sehr schnell klar, dass das Feld des Hosting sehr komplex und anspruchsvoll ist, weshalb man sich gleich zu Beginn auf die Beauftragung einer externen, gemeinsam finanzierten Studie verständigte, um die weiteren Überlegungen auf eine fundierte Grundlage stellen zu können. Besagte Studie wurde federführend von Neil Beagrie von der Firma Charles Beagrie Ltd. in Zusammenarbeit mit der Firma Globale Informationstechnik durchgeführt und im Dezember 2009 unter dem Titel »Dauerhaften Zugriff sicherstellen: Auf dem Weg zu einer nationalen Strategie zu Perpetual Access und Hosting elektronischer Ressourcen in Deutschland«<sup>5</sup> vorgelegt.

Die Studie wurde in jeder Phase von der Allianz-Arbeitsgruppe Hosting begleitet. Auf Basis der Zielvorgaben durch die Arbeitsgruppe umfasste die Studie folgende Schwerpunkte:

- Präzisierung von relevanten Begriffen,
- Beschreibung und Analyse existierender internationaler Ansätze und aktueller Initiativen in Deutschland (auf Basis von Interviews, Befragungen und Literaturstudium),
- Durchführung einer Gap-Analyse zum Vergleich von Ist-Zustand und Idealvorstellung in Bezug auf verschiedene Materialtypen,
- Definition von Anwendungsfällen zur Beschreibung des konkreten Handlungsbedarfs,
- Entwicklung von Lösungen und Szenarien für eine Strategie. Die skizzierten Lösungen werden jeweils auf ihre Eignung für die definierten Anwendungsfälle, Risiko, Aufwand und Innovationsbedarf hin überprüft.

In der Studie werden umfangreiche Materialien dokumentiert und vielschichtige Analysen zusammengetragen. Sie kulminiert jedoch nicht in einer eindimensionalen Zuspitzung, sondern präsentiert eine Reihe von konkreten Einzelvorschlägen und empfiehlt ausdrücklich die Betrachtung mehrerer Lösungsansätze auch in Kombination miteinander und im internationalen Kontext.

Die Ergebnisse und Empfehlungen der Studie wurden im März 2010 im Rahmen eines Workshops einem breiten Kreis von Experten aus Bibliotheken und Wissenschaftsorganisationen mit dem Ziel der Diskussion möglicher nächster Schritte vorgestellt. Bei der Definition der Anforderungen an die Studie hatte noch die Vorstellung bestanden, dass zwingend eine eigene Hosting-Struktur aufzubauen sei, wobei die Frage nach zentral versus dezentral im Vordergrund stand. Mit Vorliegen der Studie geht es aber nun vielmehr darum, welche Optionen der Absicherung des dauerhaften Zugriffs auf elektronische Ressourcen bereits bestehen und genutzt werden könnten – nicht zuletzt im internationalen Kontext – und welche Ansätze ggf. näher verfolgt werden sollten. Der Aufbau einer eigenen Hosting-Struktur ist hier nur ein Ansatz von mehreren möglichen, die unter Umständen miteinander zu kombinieren sind.

Im Rahmen des Workshops wurden drei der in der Studie diskutierten Ansätze ausgewählt, die eine nähere Untersuchung wert erscheinen ließen. LOCKSS, Portico und eigene Hosting-Lösungen wurden auf Grund der vorliegenden Informationen in einem ersten Schritt einer groben Bewertung unterzogen, eine detailliertere Analyse ist jedoch notwendig und stellt

**leistungsfähige  
Infrastruktur**

**doppelter Boden**

**koordiniertes Konzept**

den nächsten Schritt auf dem Weg zur Entwicklung einer Strategie des dauerhaften Zugriffs dar.

### **DREI KERNBAUSTEINE FÜR EIN GEMEINSAMES, ABGESTIMMTES KONZEPT IN DEUTSCHLAND**

**Kombination**

Als Kernbausteine für ein gemeinsames, abgestimmtes Konzept in Deutschland zeichnet sich spätestens seit diesem Workshop im März 2010 eine noch genauer zu definierende Kombination aus den möglichen Bestandteilen LOCKSS, Portico und dem Auf- bzw. Ausbau eigener Hosting-Lösungen ab. Seither arbeitet die Allianz-Arbeitsgruppe Hosting insbesondere an einer Vertiefung vor allem der Stränge LOCKSS und Portico, bei denen noch weitere Details in Erfahrung gebracht werden sollen.

#### **LOCKSS**

**Kopien auf unterschiedlichen Servern**

Das Konzept von LOCKSS (Lots of Copies Keep Stuff Safe), initiiert von und bis heute beheimatet an der Stanford University, folgt der Prämisse, dass der Erhalt digitaler Inhalte umso sicherer ist, je mehr Kopien auf unterschiedlichen Servern es von ihnen gibt.<sup>6</sup> Die LOCKSS-Software ist Open Source, sie kann allerdings nur dann zur Archivierung der Inhalte von Verlagen genutzt werden, wenn diese durch ein entsprechendes Lizenzabkommen LOCKSS explizit unterstützen. Teilnehmende Bibliotheken können die im Internet veröffentlichten E-Journals- (oder auch E-Books)-Dateien der Verlage einsammeln (»crawl«) und sie auf ihren lokalen Servern (LOCKSS-Boxen) speichern. Zusätzlich zu den eigentlichen Inhalten wird in LOCKSS das »look and feel« der Präsentationsform beim Verlag gespeichert, was eine originalgetreue Darstellung der Inhalte ermöglicht. Der Zugriff zu diesen lokal gespeicherten Inhalten wird – sofern die erforderliche Lizenzierung vorliegt – unmittelbar ermöglicht, wenn der Zugriff auf die Verlagsserver (aus irgendeinem Grund, z. B. auch bei kurzfristigen Netzwerkproblemen) gestört ist. Bisher haben rund 450 Verlage der Speicherung ihrer Inhalte im weltweiten LOCKSS-Netzwerk, das derzeit aus ungefähr 200 teilnehmenden Bibliotheken besteht, zugestimmt. Im Mittel ist der gleiche Inhalt ungefähr 40 Mal in diesem verteilten LOCKSS-Netzwerk gespeichert. Sobald neue Inhalte in das LOCKSS-System aufgenommen wurden, überprüfen die LOCKSS-Boxen untereinander regelmäßig ihre gespeicherten Daten. Sind Inhalte in einer LOCKSS-Box beschädigt oder unvollständig, kann über andere Boxen eine Reparatur erfolgen.

**»trigger events«**

Die vielfältige Replikation von Daten auf verteilten Servern führt bei einigen Verlagen zu einem gewissen Misstrauen, da sie befürchten, möglicherweise

die Kontrolle über die Nutzung ihrer Inhalte zu verlieren. Dies hält vor allem größere Verlage davon ab, ihre Zustimmung zur Teilnahme an LOCKSS zu geben. Mit einem »privaten« LOCKSS-Netzwerk (PLN)<sup>7</sup>, das unter Verwendung der LOCKSS-Technologie ein nicht öffentliches, sondern überschaubares, begrenztes verteiltes System aufbaut, könnte dieses Problem umgangen werden. Organisationen, die ein PLN betreiben, besitzen und kontrollieren die Inhalte so, wie im öffentlichen LOCKSS-Netzwerk. So könnte die Allianz der deutschen Wissenschaftsorganisationen eigene Vereinbarungen mit Verlagen über das Hosting ihrer lizenzierten Bestände in einem nationalen PLN treffen. Neben den zu klärenden Fragen hinsichtlich der Organisation und Finanzierung zum Aufbau und Betrieb eines solchen nationalen PLN gibt es noch eine Reihe weiterer offener Fragen, z. B. zur eingesetzten Software, zu Schnittstellen und zur Zusammenarbeit mit anderen Diensten, zu rechtlichen Aspekten sowie zu LZA-Diensten.

#### **Portico**

Portico ist im Gegensatz zu LOCKSS ein zentraler Dienst zum Hosting und zur (nach eigenen Aussagen) Langzeitarchivierung von überwiegend lizenzpflichtigen Inhalten (E-Journals und zunehmend ausdrücklich auch E-Books).<sup>8</sup> Die kooperierenden Verlage liefern die Daten und Dokumente an Portico, wo in einem proprietären technischen System sowohl die Source Files als auch die Presentation Files gespeichert und die Metadaten im NLM-Format standardisiert werden. Eine Prüfung auf Vollständigkeit der gelieferten Objekte ist dabei nicht vorgesehen. Zur Replikation gibt es derzeit drei Spiegelserver in den USA und einen in Europa, doch verfügen diese nur über die gespeicherten Inhalte und verwalten keine Zugriffsberechtigungen, so dass diese Server nur der Absicherung, nicht aber der Nutzung dienen.

Technische Interoperabilität ist nicht das Ziel des Ansatzes von Portico, sondern ein kontrollierter Zugang unter bestimmten, klar und eng definierten Bedingungen, den »trigger events« und dem »post-cancellation access«. Ein »trigger event« setzt dann ein, wenn Inhalte einer Zeitschrift nicht mehr vom Verlag zugänglich gemacht werden, etwa wenn der Verlag eine bestimmte Zeitschrift insgesamt oder in bestimmten Jahrgängen nicht mehr vertreibt, wenn er seine Geschäfte komplett einstellt oder wenn seine Vertriebsplattform durch einen Katastrophenfall für einen längeren Zeitraum nicht zur Verfügung steht. Bisher verzeichnet Portico genau vier solcher »trigger events«, in denen die Inhalte dann über Portico den teilnehmenden Einrichtungen zugänglich gemacht

wurden. Nach Angaben von Portico besteht bei 85 % der geladenen Inhalte für die teilnehmenden Einrichtungen derzeit die Option auf »post-cancellation access«, und es gehen auch immer mehr Verlage dazu über, die Ausübung von »post-cancellation access« über Portico in ihren Standardlizenzverträgen zu erwähnen. Aktuell besteht aber noch sehr wenig Erfahrung mit der konkreten Ausübung dieser Rechte. Bisher verfügt Portico weder über eine proaktive Rechteverwaltung noch über zügige Aktivierungsverfahren oder eine komfortable Zugangsoberfläche. Auch Hochverfügbarkeit ist bei Portico nicht gewährleistet. Bei den bisherigen Aktivierungen war ein Zeitverzug von etwa 120 Tagen zu beobachten. Außerdem müssen die Rechte immer erst vom Verlag bestätigt werden, bevor Portico den Zugang freigibt.

Derzeit kooperieren 121 Verlage mit Portico, und es gibt bereits über 700 teilnehmende Bibliotheken, davon ungefähr die Hälfte außerhalb der USA. Aus Deutschland ist noch kein Teilnehmer verzeichnet, auch wenn es vereinzelt bereits Kontaktaufnahmen und Gespräche gab. Die Kosten für eine Teilnahme an Portico richten sich nach den Erwerbsetats, eine Rahmenvereinbarung für deutsche Einrichtungen wäre verhandelbar. Die Absicherung über Portico besteht nur so lange wie eine Einrichtung Teilnehmer bleibt und die entsprechenden Jahresbeiträge erbringt. Der zentrale Ansatz und die einmalige Speicherung der Daten an einem Ort sind für die großen Verlage vertrauenswürdiger als der LOCKSS-Ansatz und erleichtern ihnen die Entscheidung zur Teilnahme. Der zentrale Ansatz bereitet andererseits den nutzenden Bibliotheken und Organisationen einige Probleme, da sie nur sehr wenig bzw. gar keine Kontrolle über das System haben, das letztendlich eine Versicherung ihrer lizenzierten Inhalte und damit ihrer Investitionen darstellt. Portico ist zwar verlagsunabhängig, doch dies bedeutet nicht, dass die Kontrolle über die Inhalte durch eine teilnehmende Bibliothek ausgeübt wird.

### — Eigene Hosting-Lösungen

Für den dritten Baustein, den Auf- bzw. Ausbau eigener Hosting-Lösungen, gibt es eine Reihe von organisatorischen Beispielen, an deren Erfahrungen sich anknüpfen lässt. Alle bisherigen Ansätze in Deutschland sind von ihrem Datenvolumen sowie von der inhaltlichen Abdeckung und Reichweite ihrer Operationen<sup>9</sup> her jedoch erheblich kleiner als maßgebliche Vorbilder im internationalen Kontext wie z. B. das an der Universität von Toronto beheimatete Scholars Portal<sup>10</sup> für die 21 Universitäten der Provinz Ontario, das Electronic Journals Center<sup>11</sup> des gesamten Verbundsystems von Ohio oder das e-Depot<sup>12</sup> der Königlichen Biblio-

thek in den Niederlanden. All diese Ansätze und Erfahrungen sind sehr umfassend in der bereits erwähnten Hosting-Studie beschrieben und können hier nicht im Einzelnen vorgestellt werden. Die Allianz-Arbeitsgruppe Hosting hat sich aber intensiv mit allen Initiativen befasst und dabei diejenigen identifiziert, bei denen ein Erfahrungsaustausch und die Erörterung möglicher Kooperationsbeziehungen besonders sinnvoll erscheinen. Wie in der Studie schon herausgearbeitet wurde, ist gerade im europäischen Kontext das niederländische e-Depot ein viel versprechendes Vorbild für den technischen und organisatorischen Aufbau in Deutschland.

Die konkrete Planung und Ausbaustufe einer dedizierten eigenen Hosting-Infrastruktur in Deutschland wird maßgeblich davon abhängen, welche Anteile in der Absicherung durch LOCKSS (als PLN) und/oder Portico abgedeckt werden können und welche konkreten Inhalte dann noch für die gemeinsame, ggf. verteilte Datenspeicherung im Rahmen eines abgestimmten Gesamtkonzepts vorgesehen sind. Auch die Berechnung des Ressourcenbedarfs hängt davon ab, welche konkrete Anforderung an die Hosting- und LZA-Services gestellt werden, wie heterogen die zu speichernden Materialien sind, um welche Größenordnung es sich handelt, welche Leistungen erbracht werden (Umfang der Qualitätssicherungsmaßnahmen, Prüfung auf Vollständigkeit der gespeicherten Objekte etc.) und welche Organisationsform angestrebt wird.

### AKTUELLER STAND UND WEITERES VORGEHEN

Die Allianz-Arbeitsgruppe Hosting verfolgt diese drei Bausteine weiter und strebt unter Berücksichtigung auch der weitergehenden Empfehlungen und Analysen der Hosting-Studie nach wie vor ein implementierungsfähiges Gesamtkonzept für einen »doppelten Boden« in Deutschland an. Dabei wird zunehmend klar, dass die zahlreichen offenen Fragen nicht allein in Arbeitsgruppen neben dem laufenden Alltagsgeschäft geklärt werden können. Deshalb wurde diskutiert, wie man personell, organisatorisch und operativ einen Schritt weiter und von der kontinuierlichen Beobachtung und Evaluation in eine erste Konkretisierungsphase kommen kann. Als geeigneter Weg dahin erscheint in der derzeitigen Phase die Aufsetzung eines Projekts mit praktischen Arbeitszusammenhängen und entsprechender personeller Ausstattung, in dem die bisher zusammengetragenen Erkenntnisse und Eckpunkte zu einer umsetzbaren nationalen Strategie weiterentwickelt werden. Offene Fragen sollen in Form von Tests und Experimenten geklärt und ne-

»post-cancellation access«

Kooperationsbeziehungen

Gesamtkonzept für einen »doppelten Boden«

ben technischen Fragestellungen vor allem auch organisatorische und rechtliche Anforderungen und Umsetzungsmöglichkeiten abgesteckt und geprüft werden. Die für ein solches Projekt notwendigen Vorarbeiten und Verfahrenswege wurden bereits identifiziert und angegangen. Eine konkrete Ausschreibung, auf die interessierte Einrichtungen in Deutschland reagieren können, wird für Ende 2011 angestrebt.

Unabhängig davon, wie die Empfehlungen der Allianz-Arbeitsgruppe Hosting für ein Gesamtkonzept am Ende aussehen werden, ist klar, dass vernetzte, nachhaltige Organisationsstrukturen in öffentlicher Hand benötigt werden, um die langfristige Verfügbarkeit elektronischer Informationsressourcen zu gewährleisten. Der initiale Aufbau des organisatorischen Rahmens und der Infrastruktur ist mit Mitteln aus projektbezogener Förderung sicherlich möglich. Für den laufenden und nachhaltigen Betrieb werden jedoch weitergehende Beauftragungen und eine dauerhafte Finanzierung erforderlich sein, womit Fragen der institutionellen Grundförderung berührt werden. Auch ist eine Beteiligung der teilnehmenden Einrichtungen an den Kosten der Sicherungsinfrastruktur denkbar – als Mitglieder oder unter der Maßgabe, dass auch bei den elektronischen Informationsressourcen genau wie bei den gedruckten Medien nicht nur Kosten für die Beschaffung anfallen, sondern auch für die dauerhafte Verfügbarhaltung. Für die Steuerung derartiger bundesweit abgestimmter und ineinander greifender Organisationsstrukturen müssten geeignete Gremien und Beiräte geschaffen werden. Ein Organisationsvorbild könnte das Deutsche Forschungsnetz (DFN) sein, das vom DFN-Verein getragen wird, einer zentralen Einrichtung der Wissenschaft in Deutschland. Das Deutsche Forschungsnetz, so heißt es auf der Homepage, »ist das von der Wissenschaft selbst organisierte Kommunikationsnetz für Wissenschaft und Forschung in Deutschland. Es verbindet Hochschulen und Forschungseinrichtungen miteinander und ist nahtlos in den europäischen und weltweiten Verbund der Forschungs- und Wissenschaftsnetze integriert.«<sup>13</sup> Die Anforderungen an eine leistungsfähige wie zeitgemäße Hostinginfrastruktur sind vergleichbar: Auch hier würde die Wissenschaft ein Datennetz zur Zugriffssicherung selbst organisieren und daran die Hochschulen und Forschungseinrichtungen sowie internationale Kooperationspartner anbinden.

Wie wichtig das Thema Hosting ist, zeigt sich auch daran, dass es zu den acht Themenfeldern des von der »Kommission Zukunft der Informationsinfrastruktur« (KII) erarbeiteten »Gesamtkonzepts für die Infrastruktur in Deutschland« zählt. Im Auftrag der Gemeinsamen Wissenschaftskonferenz von Bund und Ländern

(GWK) erarbeitete die KII unter Federführung der Leibniz-Gemeinschaft bis Frühjahr 2011 ein Gesamtkonzept für die Informationsinfrastruktur in Deutschland.<sup>14</sup> Die KII verfolgte das Ziel, die optimierte »Landschaft« der Informationsinfrastruktur in Deutschland darzustellen, die dafür notwendigen Rahmenbedingungen (Strukturen, Prozesse, Akteure und Koordination) zu beschreiben, Synergien und Möglichkeiten der Arbeitsteilung, Kooperationen und Ressourcenfreisetzung aufzuzeigen und darauf aufbauend Handlungsempfehlungen an die GWK zu formulieren. Das Thema Hosting und Langzeitarchivierung ist dabei eines von acht identifizierten Handlungsfeldern, zu denen Empfehlungen ausgesprochen werden. Die Allianz-Arbeitsgruppe Hosting war am KII-Prozess mit beteiligt, und der Erkenntnisstand der Arbeitsgruppe ist in den Gesamtbericht mit eingeflossen. Die Forderung zum Aufbau nachhaltiger Hosting-Strukturen und deren Finanzierung schlägt sich auch in diesem Bericht nieder.

Sobald sich ein konkretes Hosting-Gesamtkonzept und die Schaffung eines Organisationszusammenhangs abzeichnen, wird dies auch Auswirkungen und Rückkopplungseffekte auf andere Themenfelder und Arbeitsgruppen haben. Vor allem für die Allianz-Arbeitsgruppe zu Lizenzen werden sich sehr konkrete neue Anforderungen an die Verhandlungen und Ausgestaltung der Verträge mit den Verlagen ergeben, sobald Strategie und Zielvorstellung definiert sind. Die neuen Anforderungen werden schnell Eingang finden in die »Grundsätze für den Erwerb DFG-geförderter überregionaler Lizenzen«<sup>15</sup> und in den Mustervertrag für die Allianz-Lizenzen. Aber auch die Wissenschaftsorganisationen und die in der GASCO organisierten Regionalkonsortien in Deutschland werden in ihren Geschäftsbeziehungen mit den Verlagen die Anforderungen aufgreifen und auf die entsprechenden Zielvereinbarungen hinwirken. Auch außerhalb Deutschlands wird sich mit möglichen Kooperationspartnern wie dem niederländischen e-Depot oder globalen Initiativen wie LOCKSS und Portico sehr viel konkreter sprechen lassen, wenn das Gesamtgefüge und seine Einzelteile zumindest in Grundzügen geklärt sind.

Mit einem bundesweit abgestimmten Gesamtkonzept eines »doppelten Bodens« für den sicheren Zugang zu elektronischen Informationsressourcen wird eine wichtige Lücke geschlossen. Die Absicherung des Zugriffs auf elektronische wissenschaftliche Inhalte für den Fall, dass der Normalbetrieb nicht mehr funktioniert, ist ein funktionales Äquivalent im elektronischen Zeitalter zu den bisherigen Magazinbeständen an gedruckten Medien in den Bibliotheken und somit eine zwingend notwendige Sicherungsmaßnahme für den Wissenschaftsstandort Deutschland.

**Organisationsstrukturen**

**Finanzierung**

**Kommission Zukunft der Informationsinfrastruktur**

<sup>1</sup> Vgl. hierzu auch den Beitrag zum Thema Langzeitarchivierung in diesem Heft, S. 184–196.

<sup>2</sup> Vgl. Donald J. Waters, Urgent Action Needed to Preserve Scholarly Electronic Journals, 15. Oktober 2005, zugänglich unter [www.arl.org/bm~doc/ejournalpreservation\\_final.pdf](http://www.arl.org/bm~doc/ejournalpreservation_final.pdf)

<sup>3</sup> Dies ist gerade auf längere Sicht ein bedrohliches und kaum zu kontrollierendes Szenario. Ein aktueller Fall im Frühjahr 2011 bezieht sich auf die Zeitschrift »Entomologia experimentalis et applicata«, die bereits ab 2003 von Springer zu Blackwell transferiert wurde. Bisher waren die zurückliegenden Jahrgänge 1958–2002 noch auf Springer-Link im Rahmen der DFG-finanzierten Nationallizenzen bzw. über Zugriffsrechte aus institutionellen Lizenzen verfügbar; nun jedoch verlangt der neue Rechteinhaber Wiley-Blackwell die Entfernung dieser Inhalte vom Springer-Server. Auf der *Wiley Online Library* sind die Inhalte dieser Zeitschrift jedoch erst ab dem letzten Heft des Jahrgangs 1996 verfügbar. Die alten Jahrgänge sind derzeit nirgendwo online erhältlich, und es ist unklar, wann Wiley-Blackwell die früheren Jahrgänge wieder zugänglich macht und ob dann überhaupt die Rechte aus Verträgen mit dem früheren Rechteinhaber anerkannt werden. Der »Transfer Code of Practice« ([www.uksg.org/Transfer/Code](http://www.uksg.org/Transfer/Code)) adressiert derartige Probleme und versucht über ein gewisses Maß an Selbstverpflichtung verlässliche neue Standards zu etablieren. Harte Praxistests stehen jedoch noch aus, aber es ist eher unwahrscheinlich, dass dieser »Code of Practice« zu einem manifesten, d. h. ggf. auch einklagbaren Rechtsanspruch führen wird. Wie dieser aktuelle Fall zeigt, besteht ohne die Absicherung durch eine entsprechende Infrastruktur für alle wissenschaftlichen Einrichtungen das Risiko, bereits erworbene und vermeintlich sogar über dauerhafte Nutzungsrechte abgesicherte Zugriffe zu verlieren und diese Inhalte dann vielleicht sogar ein zweites Mal erwerben zu müssen.

<sup>4</sup> Bei den DFG-geförderten Nationallizenzen ist die physische Auslieferung der Daten zum Zwecke des Hosting eine *conditio sine qua non*. In allen Verträgen im Rahmen der Nationallizenzen ist das Hosting daher gesichert. Scheinbar unüberwindbare Hürden sind somit doch durchaus überwindbar.

<sup>5</sup> Siehe [www.allianzinitiative.de/fileadmin/hosting\\_studie\\_d.pdf](http://www.allianzinitiative.de/fileadmin/hosting_studie_d.pdf)

<sup>6</sup> LOCKSS, so der programmatische Anspruch, »is an international community initiative that provides libraries with digital preservation tools and support so that they can easily and inexpensively collect and preserve their own copies of authorized e-content.« Vgl. <http://lockss.stanford.edu/lockss/Home> – zuletzt besucht am 27.02.2011.

<sup>7</sup> Es handelt sich hierbei um ein nicht öffentliches, sondern überschaubares, begrenztes System auf Basis der LOCKSS-Technologie. Die darin gespeicherten Inhalte sind nur einer geschlossenen Community zugänglich.

<sup>8</sup> Hinter Portico steht der Anspruch: »Portico provides libraries and publishers with a reliable, cost-effective solution to one of the most critical challenges facing the scholarly community today – ensuring that the electronic resources you rely on everyday will be acces-

sible to future researchers, scholars, and students.« Vgl. [www.portico.org/digital-preservation/services/](http://www.portico.org/digital-preservation/services/) – zuletzt besucht am 27.02.2011.

<sup>9</sup> Die derzeit aktiv handelnden Einrichtungen in Deutschland haben in der Regel keinen überregionalen oder gar flächendeckenden Auftrag und etablieren daher auch keine organisatorische Infrastruktur zur Mitversorgung von Interessenten außerhalb ihres lokalen organisatorischen Zuständigkeitsbereichs.

<sup>10</sup> »Scholars Portal Journals is a digital repository of over 15 million scholarly articles drawn from journals covering every academic discipline.« Vgl. <http://spotdocs.scholarsportal.info/display/sp/Journals> – zuletzt besucht am 27.02.2011.

<sup>11</sup> »The Electronic Journal Center contains more than 9,000 scholarly journal titles from 101 publishers across a wide range of disciplines.« Vgl. [www.ohiolink.edu/about/what-is-ol.html](http://www.ohiolink.edu/about/what-is-ol.html) – zuletzt besucht am 27.02.2011.

<sup>12</sup> »The e-Depot is a digital archiving environment that ensures long-term access to digital objects.« Vgl. [www.kb.nl/hrd/dd/index-en.html](http://www.kb.nl/hrd/dd/index-en.html) – zuletzt besucht am 27.02.2011.

<sup>13</sup> Siehe [www.dfn.de/](http://www.dfn.de/)

<sup>14</sup> Siehe [www.wgl.de/?nid=infrastr&nidap=&print=o](http://www.wgl.de/?nid=infrastr&nidap=&print=o)

<sup>15</sup> Siehe [www.dfg.de/download/programm/wissenschaftliche\\_literaturversorgung\\_informationsysteme/antragstellung/12\\_18/12\\_18.pdf](http://www.dfg.de/download/programm/wissenschaftliche_literaturversorgung_informationsysteme/antragstellung/12_18/12_18.pdf)

## DIE VERFASSER

**Dr. Leni Helmes** ist Bereichsleiterin »Entwicklung und Angewandte Forschung«, FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Tel.: 07247 – 808-408, Mail: [Leni.Helmes@fiz-karlsruhe.de](mailto:Leni.Helmes@fiz-karlsruhe.de)

**Dr. Ralf Schimmer** ist Bereichsleiter »Wissenschaftliche Informationsversorgung«, Max Planck Digital Library, Amalienstr. 33, 80799 München, Tel.: 089 – 38602-261, Mail: [Schimmer@mpdl.mpg.de](mailto:Schimmer@mpdl.mpg.de)