

Bundesverdienstkreuz für Professor Christoph-Hubert Schütte



*(v.l.) Wissenschaftsminister
Prof. Dr. Peter Frankenberg,
Prof. Christoph-Hubert Schütte,
Prof. Dr. Horst Hippler, Präsident des KIT*

■ Im Rahmen einer Feierstunde im Festsaal des Studentenhauses in Karlsruhe wurde am 2. August dem ehemaligen Direktor der KIT-Bibliothek Prof. Christoph-Hubert Schütte für seine Verdienste um die Entwicklung des Bibliothekswesens das Bundesverdienstkreuz am Bande verliehen. Ausgehändigt wurde der Orden durch Prof. Dr. Peter Frankenberg, Minister für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg.

Die Auszeichnung erhielt Christoph-Hubert Schütte für seine herausragende Lebensleistung. Er hat den Wandel der Universitäts- und späteren KIT-Bibliothek von einer traditionellen Universitätsbibliothek zu einem leistungsfähigen Informationszentrum ermöglicht und das Bibliothekswesen weltweit nachhaltig geprägt. Unter seiner Leitung ist die Universitätsbibliothek zum „Pionier für

Internet-Dienstleistungen“ geworden: als eine der ersten Bibliotheken in Deutschland hat die KIT-Bibliothek einen Online-Katalog in Betrieb genommen, der Karlsruher Virtuelle Katalog (KVK) ist seit Jahren eine feste Größe in der Informationslandschaft. Als weitere Höhepunkte im Wirken Christoph-Hubert Schüttes nannte der Minister in seiner Laudatio den Umbau der traditionellen Magazinbibliothek zur ersten vollautomatischen 24-Stunden-Bibliothek. Die Mitarbeit in zahlreichen Gremien auf Landes- und Bundesebene und die Tatsache, dass er sich daneben stets um die Lehre bemühte, insbesondere bei der Ausbildung von Bibliothekaren und Studierenden, waren weitere herausragende Leistungen. Nicht zuletzt gab Schütte den Anstoß zur Gründung des BSZ – Bibliotheksservicezentrum für Baden-Württemberg und Sachsen. Zudem wurde auch sein Engagement bei B.I.T.online als einer der ersten elektronisch verfügbaren bibliothekarischen Fachzeitschriften würdigend hervorgehoben.

Herzlichen Glückwunsch!
Michael Mönlich

Einführung eines einheitlichen Systems für die Universitätsbibliographie

Open Access Server und Hochschulschriften an der Universität Regensburg

Gernot Deinzer und Rafael Ball

■ Bislang gab es an der Universität Regensburg zwei unterschiedliche Systeme zum elektronischen Veröffentlichen von Forschungsergebnissen. Aufgrund von dadurch entstandenen Unsicherheiten seitens der Bibliothekskunden und einem organisatorischen Mehraufwand erfolgte im Oktober 2009 die Umstellung auf ein einheitliches System, das internationale Standards erfüllt. Dadurch ergab sich eine deutliche Serviceverbesserung für die Wissenschaftler der Universität Regensburg.

Ausgangslage an der Universität

Mitarbeiter der Universität konnten bereits seit 1999 Hochschulschriften auf einem Server der Universitätsbibliothek online veröffentlichen. Technisch realisiert war dieser Vorgänger des jetzigen Systems als reiner Volltextserver, der vor allem für Dissertationen und die Schriftenreihe „Regensburger Diskussionsbeiträge zur Wirtschaftswissenschaft“ sehr gut angenommen wurde. Dieser Server wurde mit der Software OPUS¹ betrieben.

Im Jahr 2006 startete die Universitätsbibliothek Regensburg mit einem institutionellen Repositorium (Publikationsserver) als zusätzliche Dienstleistung. Auf diesem Server konnten Angehörige der Universität ihre Publikationen auflisten und verwalten. Möglich war dabei auch das Einbringen von Volltexten, die im Sinne von Open Access frei im Internet zur Verfügung gestellt wurden. Dazu wurde die Software EPrints² eingesetzt.

Durch den Betrieb unterschiedlicher Systeme ergaben sich bei einigen Kunden der Universitätsbibliothek Unsicherheiten. So konnten Dokumente mit freier Zugänglichkeit in beiden oder nur in einem der beiden Systeme gefunden werden. Dem Kunden war der Unterschied zwischen den Systemen häufig nicht bewusst. Zusätzlich sah sich die Universitätsbibliothek vor dem Problem, zwei Server mit unterschied-

licher Software zu betreiben und Mitarbeiter in zwei unterschiedliche Systeme einzuarbeiten. Deshalb lag die Entscheidung der Universitätsbibliothek nahe, die unterschiedlichen Systeme in eine einheitliche Server- und Softwarelösung zu migrieren.

Anforderungen

Das neue System sollte als institutionelles Repositorium dienen. Dabei sollte es für die Wissenschaftler³ der Universität Regensburg möglich sein, alle ihre Veröffentlichungen in diesem System nachzuweisen. Hier bestand der Konflikt, dass Wissenschaftler einerseits eine komplette Liste ihrer Publikationen einbringen möchten, sie aber andererseits nicht das Recht besitzen, alle ihre Veröffentlichungen frei zugänglich auf einem Server der Universitätsbibliothek zusätzlich zur Verlagspublikation zu veröffentlichen. Ein reiner Open Access Server stellte damit keine ideale Lösung dar. Erforderlich war die Möglichkeit, sowohl rein bibliographische Daten als auch Volltexte in das System einzubringen. Zusätzlich wünschenswert war dabei die Möglichkeit, eine sogenannte Embargoperiode für die Freisichtung von Volltexten zu setzen. Diese werden somit erst nach Erreichen eines definierten Datums frei zugänglich gemacht, da bei manchen Verlagen die Selbstarchivierung erst nach dem Ende einer Sperrfrist möglich ist. Durch das Einrichten von Embargoperioden können Wissenschaftler der Universität bereits beim bibliographischen Eintrag den Volltext auf den Server stellen und brauchen nach dem Ablauf der Sperrfrist keine weiteren Aktionen durchzuführen.

Idealerweise sollten die Dokumente sowohl als Primärveröffentlichung der Universität, als auch parallel zu einer Verlagspublikation öffentlich zugänglich gemacht werden können. Dabei war vor allem bei den Primärpublikationen darauf zu achten, dass die dauerhafte Zitierfähigkeit gewährleistet ist. Zu

diesem Zweck war die Vergabe einer URN (Uniform Resource Name)⁴ zwingend erforderlich.

Eine weitere Anforderung war das Erstellen von Publikationslisten. Wissenschaftler wollen für unterschiedliche Einsatzmöglichkeiten Publikationslisten zusammenstellen und diese nachnutzen. So sollte es möglich sein, eine Liste nach bestimmten Kriterien zu generieren und entweder in ein Literaturverwaltungsprogramm oder auf verschiedene Internetseiten (z. B. der Fakultät, der Arbeitsgruppe, eines Forschungsprojekts) einzubringen. Wichtig war hierbei, dass jedem Mitarbeiter der Universität eine eindeutige Identifikationsnummer zugeordnet wird. Ziel ist hier, dass Mitarbeiter der Universität ihre Veröffentlichungen an einer zentralen Stelle online pflegen können und redundante Mehrfacharbeit vermieden wird. In der Zukunft soll das System auch als vollständige Universitätsbibliographie dienen.

Darüber hinaus sollte der Publikationsserver keine Stand-alone-Lösung darstellen, sondern mit weiteren, vor allem fachlichen, Repositorien vernetzt werden. Dabei war es der Universitätsbibliothek Regensburg wichtig, dass der Wissenschaftler die freie Wahl hat, ob er seine Einträge in ein fachliches Repositorium oder in das institutionelle Repositorium einbringen möchte. Das institutionelle Repositorium musste also sowohl Import- als auch Exportfunktionen besitzen.

Auch für weitere Austauschfunktionen musste das System eine standardisierte OAI-Schnittstelle⁵ aufweisen. Diese sollte einfach erweiterbar sein, damit spezielle Dienste, wie die Vergabe einer URN, möglich sind.

Entscheidung für ein geeignetes System

Bei der Auswahl eines Systems, das diesem Anforderungsprofil entspricht, beschränk-

1 Siehe <http://opusdev.bsz-bw.de/trac>

2 Siehe <http://www.eprints.org/>

3 Wenn im Folgenden die verallgemeinernde männliche Form für die Bezeichnung von Personengruppen verwendet wird, sind sowohl weibliche als auch männliche Personen gemeint.

4 Siehe <http://www.persistent-identifier.de/>

5 Siehe <http://www.openarchives.org/>.



Abbildung 1: Der Rektor der Universität Regensburg Prof. Dr. Thomas Strothotte weiht zusammen mit dem Direktor der Universitätsbibliothek Regensburg Dr. Rafael Ball den Publikationsserver der Universität Regensburg offiziell ein (Universitätsbibliothek Regensburg, 21. Oktober 2009).

Um die Eingabe von bibliographischen Daten für den Wissenschaftler auf ein absolutes Mindestmaß zu beschränken, sind bestimmte Import- und Exportmöglichkeiten unbedingt notwendig. Zuvorderst muss eine Import- und Exportfunktion für alle gängigen Literaturverwaltungsprogramme vorhanden sein, da der Wissenschaftler hier selbst bestimmen soll, ob er die Einträge in seinem Literaturverwaltungsprogramm oder im institutionellen Repositorium durchführen will. Bei der Vernetzung mit fachlichen Repositorien ist auf die jeweilige Arbeitsweise eines Wissenschaftlers, bzw. einer akademischen Disziplin, einzugehen. So besteht beispielsweise für Wirtschaftswissenschaftler der Universität Regensburg die Möglichkeit, Diskussionspapiere auf dem Publikationsserver zu veröffentlichen. Damit diese aber in der gesamten fachspezifischen Community gesehen werden, importiert man diese automatisiert in RePec¹³. Anders hingegen verhält es sich im Fach Physik: Ein Großteil der Physiker bringt Veröffentlichungen in arXiv¹⁴ ein. Hier soll es möglich sein, dass Einträge in arXiv einfach mittels der arXiv-ID in das institutionelle Repositorium eingepflegt werden. Weitere fachliche Repositorien sollen auf analogem Wege eingebunden werden. Bei der Software EPrints sind solche Import- und Exportmöglichkeiten bereits vorhanden, bzw. können leicht technisch realisiert werden. Oftmals werden solche Zusatzprogramme von der internationalen Eprints Community bereits bereitgestellt. Beispielsweise wird eine Softwarelösung für den Import aus arXiv als Download angeboten. Herr Schallehn von der Universitätsbibliothek München hat uns ein File im ReDIF¹⁵-Format für den Export von bibliographischen Daten und Volltexten an RePec zur Verfügung gestellt, welches nur noch angepasst werden musste. Der Export von Publikationslisten für die Einbindung auf Internetseiten war von großer Bedeutung. Hierbei ist der Aspekt der Vergabe von Identifikationsnummern entscheidend, da über diese auf einfache Weise Literaturlisten erstellt werden können. Ein

ten wir uns auf die beiden Softwarelösungen OPUS und EPrints, da diese bereits an der Universitätsbibliothek Regensburg im Einsatz waren und somit sehr viel Erfahrung mit diesen Softwarepaketen vorhanden war. Diese sind zusammen mit DSpace⁶ die beiden weltweit am meisten eingesetzten Systeme. Laut dem Registry of Open Access Repositories⁷ werden 649 Repositorien mit DSpace, 336 mit EPrints und 40 mit OPUS betrieben. Bei OpenDOAR⁸ sind 592 DSpace-, 280 EPrints- und 54 OPUS-Installationen registriert. Anhand der genannten Anforderungen wurden die beiden bisher bestehenden Systeme analysiert und das für unsere Zwecke geeignetere System bestimmt. Die Universitätsbibliothek Regensburg hat sich entschieden, die Software EPrints für den neuen Service einzusetzen. Die Gründe hierfür waren:

Das System sollte eine einfache Benutzerverwaltung mit sich bringen. Dabei ist es wünschenswert, dass dieses mit dem bereits bestehenden System der Universität Regensburg verbunden ist. Eine Anbindung mittels Shibboleth⁹ oder Lightweight Directory Access Protocol (LDAP)¹⁰ ist hier notwendig. Sowohl Shibboleth als auch LDAP lassen sich mit EPrints verwirklichen. Die Rechenzentrumsnummer soll dabei als eindeutige ID, sowohl als Benutzerkennung als auch zur eindeutigen Bezeichnung für Autoren und Herausgeber dienen. Wissenschaftler, die nicht der Universität Regensburg angehören, können ihre E-Mail-Adresse angeben, die somit zur eindeutigen Identifikation dienen kann. Damit diese auch von

Jedem ohne große Suche eingegeben werden kann, wird bei der Eingabe von Autoren bzw. Herausgebern bereits nach Einträgen von Personen mit dem Namensbeginn gesucht und es werden Vorschläge angezeigt, welche durch einen einfachen Klick übernommen werden können (Technik: AJAX¹¹).

Auch die Struktur der Universität muss flexibel darstellbar sein. Gerade bei Änderungen der Fakultätsstruktur, neuen Lehrstühlen und Arbeitsgruppen oder der Versetzung eines Professors in den Ruhestand, soll die neue Struktur richtig abgebildet werden. Für die Mitarbeiter der Universität sollen keine Änderungen nötig und die Strukturveränderung für die Bibliothek mit einem einmaligen, geringen Aufwand umzusetzen sein. Durch die Verwendung von EPrints ist dies komfortabel möglich.

Wichtig ist die möglichst einfache Eingabe aller relevanten Daten für den Benutzer. Neben der Autovervollständigung der Personen gibt es auch Vorschläge für die Titel von Zeitschriften, Projekten, Workshops, Schriftenreihen usw. Im Falle der Zeitschriften wird dabei eine bestehende Datenbank zum Vorschlag von Zeitschriftentiteln verwendet. Es werden nicht nur der Zeitschriftentitel ergänzt, sondern zusätzlich auch die ISSN und der Verlag eingefügt. Gerade diese Daten werden erfahrungsgemäß selten eingegeben. Zusätzlich werden noch Farben analog zur SHERPA/RoMEO-Liste¹² angezeigt, die den Wissenschaftler auf einen Blick über die rechtlichen Standardbedingungen informieren.

6 Siehe <http://www.dspace.org/>

7 Siehe <http://roar.eprints.org/>, Stand 27.07.2010

8 Siehe <http://www.openoar.org/>, Stand 27.07.2010.

9 Siehe <http://shibboleth.internet2.edu/>.

10 Siehe beispielsweise <http://www.openldap.org/>.

11 AJAX steht für Asynchronous JavaScript and XML. Für eine Einführung siehe <http://www.w3schools.com/ajax/default.asp>

12 Die Liste kann unter <http://www.sherpa.ac.uk/romeo/> eingesehen werden. Eine deutsche Version ist zugänglich unter <http://open-access.net/de/allgemeines/rechtsfragen/sherparomeo/>

13 Siehe <http://repec.org/>

14 Siehe <http://arxiv.org/>

15 Siehe <http://ideas.repec.org/p/rpc/rdfdoc/redif.html>

bank des Hochschulschriftenservers ausliest und in ein für EPrints importfähiges xml-File schrieb. Dieses wurde dann in EPrints eingelesen. Wichtig dabei war, dass alle Daten aus dem Hochschulschriftenserver im neuen System gespeichert werden.

Schließlich mussten noch Hilfeseiten zum Onlinepublizieren, dem neuen System entsprechend angepasst und eingepflegt werden.

Einstellen des alten Services

Um einen reibungslosen Übergang des Hochschulschriftenservers in den Publikationsserver zu gewährleisten, wurde im ersten Schritt die Möglichkeit zum Veröffentlichen abgeschaltet. In einem zweiten Schritt wurden Weiterleitungen von allen Einträgen im Hochschulschriftenserver mittels der URN zum Publikationsserver eingerichtet. Somit blieb gewährleistet, dass alle Zitate und Links auf die alten Seiten weiterhin erreichbar sind.

Im Rahmen der Open Access Woche 2009 wurde der neue Publikationsserver vom Rektor der Universität Regensburg, Herrn Prof. Dr. Thomas Strothotte, offiziell eingeweiht.

Zusammenfassung

Mit der Wahl der Software EPrints hat sich die Universitätsbibliothek entschlossen, ein international bewährtes System einzusetzen.

Durch die Zusammenlegung des Hochschulschriftenservers mit den bisherigen institutionellen Repositorien wurde eine benutzerfreundliche Veröffentlichungsplattform geschaffen.

Dank

Wir bedanken uns bei der Universitätsbibliothek München, Herrn Volker Schallehn, für die Bereitstellung der einzelnen Programme.

■ AUTOREN

DR. RAFAEL BALL

Direktor der Universitätsbibliothek Regensburg
 Universitätsbibliothek
 Universitätsstraße 31
 93053 Regensburg
 rafael.ball@bibliothek.uni-regensburg.de



DR. GERNOT DEINZER

Fachreferent für
 Mathematik, Physik und
 Informatik
 Open Access Beauftragter
 der Universität Regensburg
 Universitätsbibliothek
 Regensburg
 93042 Regensburg
 gernot.deinzer
 @bibliothek.uni-regensburg.de



STAR-Kundentreffen 2010
Münchner Künstlerhaus
vom 13. bis 14. Oktober 2010

Integrierte Knowledge Center Lösungen

Bibliotheken, Archive, Dokumentations- und Informationszentralen, Museen und Landtage werden mit Anforderungen konfrontiert, die sich schnell verändern und stetig wachsen. Die **Cuadra STAR** Information Management Suite hat sich in 25 Jahren immer neu definiert, um dieser rasanten Entwicklung von Technologie und Anwendererwartung stets gerecht werden zu können.

- ▶ Archivmanagement
- ▶ Bibliotheksverwaltung
- ▶ Bild- und Medienarchiv
- ▶ Dokumentenmanagement
- ▶ eGovernment
- ▶ Integrierter Document Delivery Service
- ▶ Literaturverwaltung
- ▶ Museumsmanagement
- ▶ Normenverwaltung
- ▶ Parlamentsdokumentation
- ▶ Patentinformationsverwaltung
- ▶ Thesaurusmanagement
- ▶ Wissensmanagement
- ▶ Zeitschriftenverwaltung

Wir bieten Ihnen für Ihre individuellen Anforderungsprofile übersichtliche und anwenderfreundliche Lösungen!

GLOMAS Deutschland GmbH
 Germaniastraße 42
 80805 München

Fax 089 36 11 066
 Tel. 089 3 68 19 90

sales@glomas.de
www.glomas.com

LuKII (LOCKSS-und-KOPAL-Infrastruktur-und-Interoperabilität)

Kathrin Grzeschik, Niels Fromm und Pamela Aust

Einführung

■ LuKII (LOCKSS-und-KOPAL-Infrastruktur-und-Interoperabilität) ist ein DFG-gefördertes Projekt, das die digitale Langzeitarchivierung zur Aufgabe hat. LuKII will jedoch nicht nur eine neue Software aus zwei bestehenden Softwarekomponenten bzw. Systemen kreieren, sondern zusätzlich eine organisatorisch kompetente und betriebsfähige Infrastruktur für Langzeitarchivierung auf die Beine stellen. Diese Infrastruktur beinhaltet nicht nur ein Netzwerk an teilnehmenden Institutionen, sondern auch ein Kompetenzzentrum für den technischen Support.

Digitale Langzeitarchivierung ist essentiell, denn Forschung geschieht digital und wird auch in den meisten Bereichen digital zur Verfügung gestellt. Langzeitarchivierung ist daher für alle Disziplinen und Forschungsbereiche entscheidend.

In den letzten 10 Jahren hat es vermehrt gute Forschung [z.B. Rosenthal, Robertson, Lipkis, Reich und Morabito (2005) oder Altenhöner (2005)] im Bereich digitaler Langzeitarchivierung gegeben. Es gibt jedoch bis jetzt keine Lösung auf dem Markt, die den durch ein Modell dargestellten Ablauf einer ‚sicheren‘ Langzeitarchivierung, komplett abbildet. Das OAIS Reference Modell (siehe Abbildung 1) beschreibt Langzeitarchivierungssysteme. Das Modell verdeutlicht, wie die Komponenten eines Langzeitarchivierungssystems aussehen könnten, um so u.a. eine gemeinsame Diskussions- und Entwicklungsbasis zu schaffen. Bis jetzt gibt es jedoch kein System, das dem OAIS-Modell in seiner ausführlichen Weise entspricht, kostengünstig ist, die Authentizität und Integrität der Dokumente wahrt und eine Benutzung der gespeicherten Daten ermöglicht. LuKII hat genau das zum Ziel.

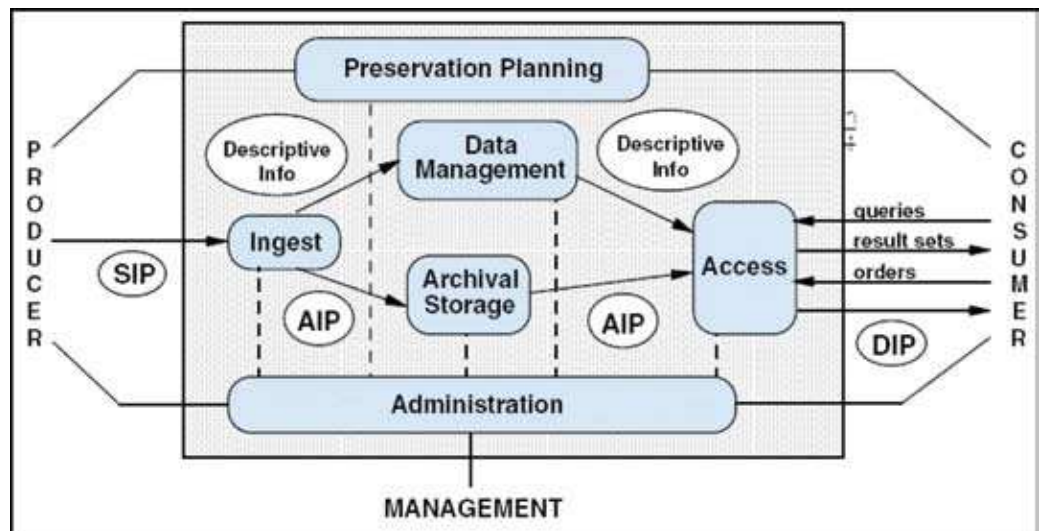


Abbildung 1: OAIS-Modell¹

LuKII

Die Ziele von LuKII sind viergeteilt. Zum Einen soll der Aufbau einer kosteneffektiven LOCKSS-Community in Deutschland erfolgen. Zeitgleich erfolgt der Aufbau eines Kompetenzzentrums zur Unterstützung von LOCKSS-Anwendungen in Deutschland. Ein weiteres Ziel ist es, die Interoperabilität zwischen LOCKSS und kopal zu ermöglichen. Beide Systeme bilden bis dato einige Teilbereiche des OAIS-Modells mehr oder weniger gut ab, eine erfolgreiche Zusammenführung könnte die LZA nach dem OAIS-Modell kompetent und komplett gewährleisten. Des Weiteren ist die Zusammenarbeit mit CLOCKSS (Closed LOCKSS) geplant. LuKII profitiert durch die Teilnahme am CLOCKSS PLN (Private LOCKSS Network) an dem Wissen und den schon vorhandenen Erfahrungen der in CLOCKSS beteiligten Partner und kann diese in das Projekt LuKII einfließen lassen.

LOCKSS

LOCKSS (Lots of Copies Keep Stuff Safe) ist ein Langzeitarchivierungssystem der Stan-

ford University, Kalifornien, das auf der Idee des distributed computing beruht. Der Name ist Programm und die Grundidee der Langzeitarchivierung ist das verteilte Speichern von mindestens 7 Kopien einer Masterkopie (z.B. ein Zeitschriftenartikel) als Bitstrom, auf 7 verschiedenen Servern, die weltweit verteilt sein können oder alle in einem Raum stehen.

Diese Anzahl von Servern, die über ein TCP/IP-Netzwerk miteinander verbunden sind und miteinander kommunizieren, bezeichnet man auch als LOCKSS-Netzwerk. Um ein Dokument von einem Dokumentenserver zu archivieren, würden sieben Server des LOCKSS-Netzwerkes das Dokument in ihr lokales Speichersystem herunterladen. Auf jedem dieser sieben LOCKSS-Server würde für dieses Dokument eine Prüfsumme (oder Hashwert) berechnet.

Nun folgt der entscheidende und technologisch interessante Schritt. Es kommt im LOCKSS-Netzwerk zu einer Abstimmung über die Integrität des Dokumentes, alle einzelnen Server vergleichen ihre Version des Dokumentes und die zugehörige Prüfsumme. Erst wenn eine Mehrheit der Server eine Übereinstimmung feststellt, eine Mehrheit also die gleiche Prüfsumme berechnet hat, wird davon ausgegangen, dass dieses Dokument eine identische Kopie vom ursprünglichen Dokumentenserver ist. Alle

¹ http://preserv.eprints.org/img/OAIS_full.gif

in der Abstimmung unterlegenen LOCKSS-Server kopieren sich dann das Dokument mit der richtigen Prüfsumme von einem der anderen Server. Dieses Dokument ist damit mit dieser Prüfsumme im LOCKSS-Netzwerk archiviert.

Das Abstimmungssystem ist auch die Grundlage für die Integritätssicherung der LZA in LOCKSS, so werden diese Abstimmungen kontinuierlich für alle archivierten Dokumente in dem LOCKSS-Netzwerk durchgeführt. Sollte irgendwann bei einem einzelnen LOCKSS-Server der Inhalt des Speichersystems korrupt oder völlig verloren sein, so kann der Inhalt oder Teile davon von den anderen Servern im Netz wiederhergestellt werden. Dass die Mehrheit der Server in einem LOCKSS-Netzwerk gleichzeitig ihre Daten verlieren oder verfälschen ist sehr unwahrscheinlich.

Dabei ist ein Sicherheitsmerkmal die räumliche Verteilung der einzelnen Server. Sind diese weltweit verteilt, so ist es sehr unwahrscheinlich, dass alle gleichzeitig von technischen Problemen oder Katastrophen bedroht werden. Gleichzeitig ist es auch sehr unwahrscheinlich, dass alle Server gleichzeitig von einem Angreifer übernommen oder sabotiert werden.

Die Kommunikation in einem LOCKSS-Netzwerk ist zwar über Zertifikate verschlüsselt und die Server selbst durch eine eigene Firewall geschützt, aber ein Angriff von außen oder auch die mutwilligen Beschädigungen von Dokumenten durch den Administrator eines LOCKSS-Servers selbst können nicht vollständig ausgeschlossen werden. Durch das Abstimmungssystem

kann allerdings das Gesamtsystem durch einen einzelnen Server nicht gestört werden, in LOCKSS werden Dokumente wirklich sicher archiviert.

LOCKSS bietet des Weiteren die Möglichkeit einer „on-the-fly Migration“. LOCKSS kann als proxy-Server im Netz, im Gegensatz zu den meisten anderen vorhandenen Langzeitarchivierungssystemen, automatisch Zugang zu den auf den Servern liegenden Dokumenten geben (siehe Abbildung 2).²

Da die Daten als Bitstrom vorliegen, kann auf Wunsch in jedes Format migriert werden. Das ist besonders hilfreich, wenn z.B. der Server eines Verlags gerade Probleme hat und das Originaldokument nicht an den Nutzer geliefert werden kann. Wenn LOCKSS als Proxy agiert, ist es für den Nutzer völlig unauffällig, woher er das Dokument erhält.

kopal

kopal (Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen) ist ein deutsches Langzeitarchivierungssystem, das mit Hilfe eines durch das BMBF (Bundesministerium für Bildung und Forschung) unterstützten Forschungsprojekts entstanden ist. kopal bildet zwar die drei Hauptbereiche des OAIS-Modells gut ab, (Ingest, Archiving, Dissemination), aber es ist z.T. proprietär und dahingehend unsicher und leider momentan auch sehr teuer. kopal besteht

2 https://docs.google.com/present/view?skipauth=true&id=dkvvcck_22gcnpsgfx

aus den zwei Teilen DIAS (Digital Information Archiving System) und koLibRI (kopal Library for Retrieval and Ingest). DIAS ist ein Produkt der Firma IBM. „Entsprechend des OAIS-Standards für Archivsysteme verwaltet es die gespeicherten Daten und stellt eine Umgebung für unterschiedlichste Preservation-Strategien zur Verfügung.“³ koLibRI ist eine aus Java-Elementen bestehende Software. Sie ist Open Source und „bewusst so angelegt, dass sie als Ganzes oder in Teilen auch in anderen Zusammenhängen nachnutzbar ist.“⁴ Diese Eigenschaften macht sich LuKII zu Nutzen. koLibRI ist an erster Stelle für den Ingestbereich des OAIS-Modells zuständig und packt mit Hilfe von JHOVE Metadaten in ein UOF (Universal Object Format).

Interoperabilität

Zwei Dinge sind für die Interoperabilität der beiden Systeme entscheidend.

- koLibRI muss in der Lage sein mit LOCKSS zu kommunizieren,
- LOCKSS muss in der Lage sein koLibRI zu verstehen und „lernen“, METS Metadaten anzureichern.

Eine LOCKSS-Box kann nur über das Internet „kontaktiert“ werden, d.h. Veränderungen des Inhalts einer LOCKSS-Box können immer nur über Veränderungen des Originals der Ursprungsquelle, geschehen. Wenn z.B. ein Verlag eine Publikation überarbeitet, wird auch in der LOCKSS-Box die neue Version gespeichert. Die Inhalte der LOCKSS-Box können jedoch nicht wirklich verändert werden, eine neue Masterversion bedeutet nicht, dass die alte Masterversion in der LOCKSS-Box gelöscht wird, sondern nur, dass eine neue Masterversion als Version 1.1. etc. des alten digitalen Dokuments erstellt wird. Diese Art der Versionskontrolle sichert sowohl die Authentizität als auch die Integrität der digitalen Objekte. Es ist immer nachvollziehbar wann, wer, welche Veränderungen gemacht hat. Auf diese Art und Weise wäre es z.B. auch möglich eine prophylaktische Migration durchzuführen (im Gegensatz zur on-the-fly-Migration). Hier muss ein neues SIP (Submission Information Package) erzeugt werden, das als neue Version eines bereits vorhandenen digitalen Dokuments gespeichert werden kann. Diese Migration kann (und sollte) mit Hilfe eines Proxy-Servers vollzogen werden, da man als LOCKSS-Box Betreiber nicht den Verlag, bzw. den Betreiber der Ursprungsquelle bitten kann, eine Migration der Ursprungsdatei durchzuführen. Der Proxy-

3 http://kopal.langzeitarchivierung.de/index_software.php.de

4 http://kopal.langzeitarchivierung.de/index_koLibRI.php.de

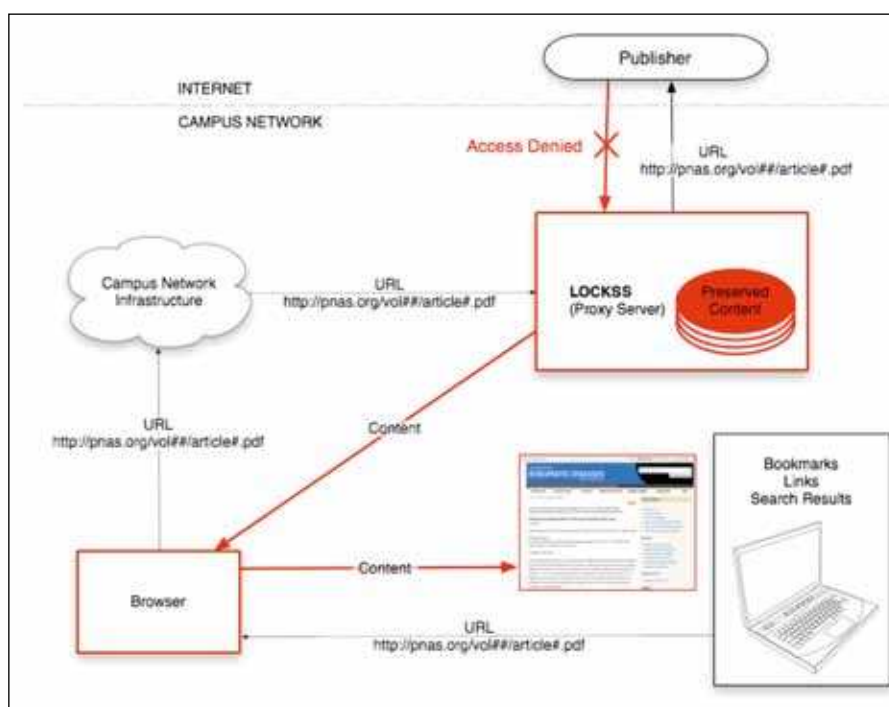


Abbildung 2: LOCKSS als Proxy-Server im Netz²

Server würde die LOCKSS-Box in dem Glauben lassen, dass das Original geändert wurde. Hier kommt koLibRI ins Spiel.

Zum Einen ist koLibRI in der Lage, als Proxy zu agieren und den Inhalt aus der LOCKSS-Box zu extrahieren, um so z.B. eine Migration durchzuführen. Zum Anderen kann koLibRI die extrahierten digitalen Originalobjekte mit Metadaten versehen, auch innerhalb der sieben LOCKSS-Server gibt es immer eine Masterkopie. LOCKSS benutzt momentan keine Metadaten, abgesehen von den fundamentalen technischen Metadaten. Da koLibRI jedoch die extrahierten digitalen Objekte mit Metadaten versehen kann, muss erstens koLibRI so geändert werden, dass es mit LOCKSS kommunizieren kann und zweitens muss LOCKSS so umprogrammiert werden, dass es standardmäßig Metadaten speichern kann.

Die Metadatenanreicherung soll mit Hilfe des METS (Metadata Encoding & Transmission Standard) Schema durchgeführt werden. Momentan packt koLibRI die Metadaten in ein UOF (Universal Object Format). „Ein Archivobjekt besteht beim UOF aus einer gepackten Datei, die eine beliebige Ordnerstruktur mit beliebig vielen Dateien enthält. Auf der Wurzelebene der Ordnerstruktur muss sich eine Datei namens „mets.xml“ befinden, die eine gültige XML-Datei gemäß des Schemas von METS 1.4 darstellt.“⁵ UOF entstand im Rahmen von kopal, basiert auf den Standards METS⁶ und LMER⁷ und definiert, wie die beschreibende XML-Datei aufgebaut ist, bzw. welche Teile von METS zwingend vorhanden sein müssen. Enthalten sind beschreibende und technische Metadaten zu allen im Objekt vorhandenen Dateien. Die technischen Metadaten bestehen aus der Ausgabe von JHOVE⁸, die als LMER-Element in der mets.xml vorhanden sind. Da so alle zur LZA benötigten Informationen zu dem Objekt enthalten sind, entspricht ein UOF-Objekt einem Information Package (IP) aus dem OAIS-Modell. In kopal wird für jedes zu archivierende Dokument ein UOF-Objekt generiert, das dann in DIAS abgelegt wird. Eine Frage, die sich im Zuge der LZA stellt, ist die Definition eines digitalen Dokuments. Was ist ein digitales Dokument? Ist ein digitales Dokument nur eine Datei, z.B. word, pdf, ps, etc oder nur eine URL, also genau eine Website, oder eine URL plus alle Verlinkungen? LOCKSS kann z.B. jeden beliebigen Teil einer Website als digitales Objekt erkennen und nach den oben beschriebenen Methoden speichern. Im Web gibt es kei-

ne bindenden Konzepte die definieren, was ein digitales Dokument ist. Um „ein“ digitales Objekt für die LZA zu erhalten, müssen manchmal viele Webseiten/URLs gerendert werden. Auch Pdfs können heutzutage aktive Inhalte enthalten, die den Rahmen eines digitalen Dokuments verschieben. Objekte sind nicht ‚nur‘ ein Bitstrom, sondern z.B. Cascading Style Sheets, Links und Bilder in Webseiten, etc. – also aktive Dokumente die aktiv gespeichert werden müssen. Diese unterschiedlichen Ansätze eines digitalen Objekts (aktiv vs. passiv/eingeraht) müssen deutlich sein und für LuKII und damit Deutschland neu definiert werden. Momentan erstellt koLibRI Archivpakete, die auf dem Gedanken der „einen Datei“ basieren, diese werden ebenso von koLibRI in DIAS eingespielt. Im Rahmen von LuKII müssen diese Archivpakete neu bestimmt werden.

LuKII hat daher ein mehrschichtiges Vorgehen beim Aufbau des Netzwerks geplant. Zum Einen wird zuerst ein traditionelles LOCKSS-Netzwerk ohne koLibRI aufgebaut, in dem zum Anderen verschiedene Testphasen vorgesehen sind, in denen die in den Institutionellen Repositorien vorhandenen Dokumente abgearbeitet werden. In einer zweiten Runde wird dann der LuKII-Prototyp, also das LOCKSS/koLibRI-Netzwerk mit den gleichen Schritten aufgebaut und getestet. Diese Vorgehensweise bringt unter anderem eine automatische Öffnung des Verständnisses eines digitalen Dokuments/Objekts in der heutigen Zeit. Außerdem können bei mehrschichtigen Testphasen eventuell auftretende Fehler schneller lokalisiert und behoben werden. Die beteiligten Partnerinstitutionen stehen bereits fest, ebenso ist der Aufbau des Kompetenzzentrums in vollem Gange. Nach den Sommermonaten erwartet LuKII die ersten Resultate, die unter anderem auf der Projektwebsite zu finden sein werden⁹.

Literatur

Altenhöner, Reinhard. Daten für die Zukunft – Das BMBF-Projekt Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen (kopal) und seine Hintergründe. 2005. Bibliothek: Forschung und Praxis, Volume 29, Issue 2, Pages 192 – 198

Rosenthal, David S. H. & Robertson, Thomas S. & Lipkis, Tom & Reich, Vicky & Morabito, Seth. A Fresh Look at the Reliability of Long-term Digital Storage. 2005. Proceedings of EuroSys Conference in Leuven (Belgien) <http://www.eecs.harvard.edu/~mema/publications/hotel.pdf>

Internet

DFG-Projekt: LuKII (LOCKSS-und-Kopal-Infrastruktur-und-Interoperabilität

<http://www.ibi.hu-berlin.de/forschung/digibib/forschung/projekte/LuKII>

(zuletzt aufgesucht am 21.08.2010)

http://preserv.eprints.org/img/OAIS_full.gif

http://preserv.eprints.org/img/OAIS_full.gif

(zuletzt aufgesucht am 27.07.2010)

JHOVE – JSTOR/Harvard Object Validation

Environment <http://hul.harvard.edu/jhove/>

(zuletzt aufgesucht am 20.08.2010)

kopal

http://kopal.langzeitarchivierung.de/downloads/kopal_Universelles_Objektformat.pdf

(zuletzt aufgesucht am 01.08.2010)

kopal – kooperativer Aufbau eines Langzeitarchivs digitaler Informationen – koLibRI: kopal

Library for Retrieval and Ingest http://kopal.langzeitarchivierung.de/index_koLibRI.php.de

(zuletzt aufgesucht am 01.08.2010)

kopal – kooperativer Aufbau eines Langzeitarchivs digitaler Informationen – Software

http://kopal.langzeitarchivierung.de/index_software.php.d

(zuletzt aufgesucht am 01.08.2010)

Providing Access

https://docs.google.com/present/view?skipauth=true&id=dskvck_22gcnpsgxf

(zuletzt aufgesucht am 28.07.2010)

AUTOREN

Kathrin Grzeschik

hat mit einem Master of Arts in Bibliotheks- und Informationswissenschaft am Institut für Bibliotheks- und Informationswissenschaft der Humboldt Universität zu Berlin abgeschlossen, bevor sie die Stelle als Projektkoordinatorin für LuKII (LOCKSS-und-KOPAL-Infrastruktur-und-Interoperabilität) übernommen hat. Außerdem ist sie Assistant Editor der Zeitschrift Library Hi Tech publiziert von Emerald Group Publishing Ltd.

Dipl. Inf. Niels Fromm

ist der stellvertretende Leiter der Arbeitsgruppe Elektronisches Publizieren, einer gemeinsamen Einrichtung von Universitätsbibliothek und Computer- und Medienservice der Humboldt-Universität zu Berlin. Davor betreute er die IT der Bibliothek des Deutschen Instituts für Normung und arbeitete als Entwickler für verschiedene Start-Up-Unternehmen.

Pamela Aust

hat einen Bachelor of Arts am Institut für Bibliotheks- und Informationswissenschaft der Humboldt Universität zu Berlin absolviert und ist jetzt Studentin im Masterstudiengang in Vollzeit. Seit Februar 2010 ist sie Studentische Mitarbeiterin im Projekt LuKII.

Humboldt Universität zu Berlin

Institut für Bibliotheks- und Informationswissenschaft

Unter den Linden 6

10099 Berlin

LuKII.IBI@hu-berlin.de

5 http://kopal.langzeitarchivierung.de/downloads/kopal_Universelles_Objektformat.pdf (p.3)

6 Ibid

7 Ibid

8 <http://hul.harvard.edu/jhove/>

9 <http://www.ibi.hu-berlin.de/forschung/digibib/forschung/projekte/LuKII>