

## PROJEKTE

# Das Projekt LuKII (LOCKSS-und-KOPAL-Infrastruktur-und-Interoperabilität) – eine Herangehensweise an Digitale Langzeitarchivierung für Deutschland



**Kathrin Grzeschik**  
Projektkoordination LuKII  
Humboldt-Universität zu Berlin  
Institut für Bibliotheks- und Informationswissenschaft  
Unter den Linden 6  
D-10099 Berlin  
E-Mail: Kathrin.Grzeschik@ibi.hu-berlin.de

LuKII (LOCKSS-und-KOPAL-Infrastruktur-und-Interoperabilität) ist ein von der Deutschen Forschungsgemeinschaft (DFG) gefördertes, gemeinsames Langzeitarchivierungsprojekt der Deutschen Nationalbibliothek (DNB) und der Humboldt-Universität zu Berlin (HU). Im Rahmen des Projekts werden zwei existierende Systeme, LOCKSS (Lots Of Copies Keep Stuff Safe) und kopal (Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen), bzw. Teile der Systeme, aufgegriffen, um eine funktionierende Infrastruktur für Langzeitarchivierung in Deutschland aufzubauen. Digitale Langzeitarchivierung betrifft jede wissenschaftliche Disziplin, jedoch gibt es bis dato kein effizientes Langzeitarchivierungssystem, das Bitstrom-Speicherung mit der Wahrung der Integrität der digitalen Objekte, sowie die Benutzbarkeit der selbigen gewährleisten kann.

Schlüsselwörter: LOCKSS; kopal; Langzeitarchivierung in Deutschland; Interoperabilität; formatunabhängige Speicherung

### **The Project LuKII (LOCKSS-and-KOPAL-Infrastructure-and-Interoperability) – An Approach to Digital Long-Term Archiving for Germany**

LuKII (LOCKSS-and-KOPAL-Infrastructure-and-Interoperability) is a German research project in the field of long-term preservation. It is funded by the German Research Foundation (DFG) and project partners are the German National Library (DNB) and the Berlin School for Library and Information Science (IBI) at Humboldt University, Berlin (HU). The project will take parts of two existing long-term preservation systems, LOCKSS (Lots Of Copies Keep Stuff Safe) and kopal (Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen) to build an efficient infrastructure for long-term preservation in Germany. Long-term preservation touches all scholarly fields, but currently no efficient system exists that uses bitstream-

preservation and warrants integrity and long-term usability of the digital objects.

Keywords: LOCKSS; kopal; long-term preservation in Germany; interoperability; format agnostic preservation

## 1 Einführung

LuKII (LOCKSS-und-KOPAL-Infrastruktur-und-Interoperabilität) ist ein von der Deutschen Forschungsgemeinschaft (DFG) gefördertes, gemeinsames Langzeitarchivierungsprojekt der Deutschen Nationalbibliothek (DNB) und der Humboldt-Universität zu Berlin (HU). Grundlage der Projektidee ist die Tatsache, dass wissenschaftliche Informationen heutzutage hauptsächlich digital erstellt, verarbeitet und veröffentlicht werden. Diese Realität macht deutlich, dass Langzeitarchivierung alle Bereiche der Wissenschaft gleichermaßen betrifft. Wissenschaftler erkennen auch, dass Langzeitarchivierung ein wichtiges Thema für die Zukunft ist, aber sie gehen im Allgemeinen davon aus, dass sich andere darum kümmern.

Es ist auch in der Tat so, dass sich Gedächtnisorganisationen seit knapp 10 bis 15 Jahren<sup>1</sup> mit der Problematik der Langzeitarchivierung auseinandersetzen, aber bis jetzt gibt es keine wirkliche Lösung. Mehrfache Sicherheitskopien helfen, adressieren aber nicht die Problematiken der Authentizität, Integrität und Benutzbarkeit digitaler Objekte. Physische Datenträger wie Magnetband sind anfällig für Zerfall oder magnetische Zerstörung.

“Today a much broader community must deal with and accept definitions of terms like trust, integrity, and authenticity, and they must trust implementation algorithms in programming languages that they can no more read than they can generally read ancient Greek.” (Seadle, 2006)

LOCKSS (Lots of Copies Keep Stuff Safe)<sup>2</sup> und kopal (Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen)<sup>3</sup> sind Ergebnisse jahrelanger Bemühungen im Bereich der digitalen Langzeitarchivierung. Im Rahmen von LuKII greifen wir diese Systeme, bzw. Teile der Systeme,

1 Ein Beispiel ist der Konferenzband der European Commission on Preservation and Access von 1997, mit dem Titel: „Choosing to Preserve: Towards a Cooperative Strategy for Long-Term Access to the Intellectual Heritage. Papers of the International Conference Organized by the European Commission on Preservation and Access and Die Deutsche Bibliothek, Leipzig/Frankfurt am Main, March 29-30, 1996.

2 <<http://lockss.org/lockss/Home>>.

3 <<http://kopal.langzeitarchivierung.de/index.php.de>>.

auf, um eine funktionierende Infrastruktur für Langzeitarchivierung in Deutschland aufzubauen. Im Vordergrund dieser Infrastruktur stehen Aspekte wie Bitstrom-Speicherung, die Integrität der digitalen Objekte sowie die Benutzbarkeit der selbigen zu gewährleisten. Tests mit Inhalten aus deutschen Institutionellen Repositorien sollen das entstandene System dann erproben. Ein Hauptaugenmerk des Projekts liegt auf der Herstellung der Interoperabilität zwischen den beiden System LOCKSS und kopal, um mögliche Synergieeffekte bestmöglich auszunutzen.

Im Wesentlichen verfolgt LuKII vier Zielstellungen:

- Aufbau einer kosteneffektiven LOCKSS-Community in Deutschland,
- Aufbau eines Kompetenzzentrums zur Unterstützung von LOCKSS-Anwendungen in Deutschland,
- Interoperabilität zwischen LOCKSS und kopal schaffen,
- Untersuchungen / Testreihen zur Interoperabilität von LOCKSS und kopal.

LuKII will also nicht nur aus zwei bestehenden Softwarekomponenten, bzw. Systemen, eine neue Software kreieren, sondern ebenso eine organisatorisch kompetente und betriebsfähige Infrastruktur für Langzeitarchivierung in Deutschland aufbauen. Diese Infrastruktur soll aus einem Netzwerk an Partnern, aber auch aus einem Kompetenzzentrum für technischen Support bestehen. Die erste Phase des Projekts beinhaltet vor allem das organisatorisch orientierte Unterfangen, den Aufbau eines LOCKSS-Netzwerks in Deutschland zu bewältigen.

Zuerst sind mögliche Kooperationspartner gesucht und gefunden worden, die sowohl die technischen als auch die finanziellen Voraussetzungen erfüllen, um die benötigten LOCKSS-Boxen einzurichten. Um Zusammenhänge des Projekts besser darstellen zu können, wird zunächst ein kurzer Überblick über LOCKSS und kopal gegeben.

## 2 LOCKSS

LOCKSS ist ein US-amerikanisches Open-Source-System und -Projekt zur Langzeitarchivierung digitaler Objekte, das von David Rosenthal an der Stanford University in Kalifornien entwickelt wurde. Ziel von LOCKSS war es, Bibliotheken die Möglichkeit zu geben, eine Kopie der digitalen Bestände, die sie erwerben, zu besitzen und nicht nur mit Lizenzen abgefertigt zu werden. LOCKSS ist mittlerweile in seinem elften Jahr und bietet nicht nur Bibliotheken, sondern auch anderen Institutionen wie Museen, Universitäten, Regierungsstellen (z. B. der US-amerikanischen Buchdruckerei / US Government Printing Office [GPO])<sup>4</sup>) eine kostengünstige, effektive Methode der Langzeitarchivierung.

Obwohl der Businessplan von LOCKSS die Speicherung von elektronischen Zeitschriften betont, ist LOCKSS in der Lage, jede Art von digitalen Objekten in Form des Bitstroms zu speichern. LOCKSS sammelt Inhalte, indem

es bestimmte Seiten im Internet ‚crawl‘ und die gefundenen digitalen Objekte in dem Format speichert, wie sie auf der jeweiligen Webseite zu finden waren. Natürlich wird vorab mit Hilfe von Plugins und einem ‚Manifest‘ auf der Webseite des Partners bestimmt, was gecrawlt werden darf.

Wie der Name ‚Lots Of Copies Keep Stuff Safe‘ bereits andeutet, basiert LOCKSS auf der Idee des distributed computing (verteilttes Rechnen). Verteilttes Rechnen beruht auf dem Prinzip der verteilten Bearbeitung oder Speicherung digitaler Objekte auf verschiedenen Rechnern. LOCKSS speichert digitale Objekte als Bitstrom auf mindestens sieben verschiedenen Festplatten (Rechnern). Diese Rechner können weltweit verteilt oder alle in einem Raum stehen, letzteres wird aus Sicherheitsgründen jedoch nicht empfohlen. Die räumliche Verteilung ist ein Sicherheitsmerkmal des LOCKSS-Systems. Sind Server weltweit verteilt, ist es unwahrscheinlich, dass alle gleichzeitig von technischen Problemen oder Naturkatastrophen betroffen sein werden. Ebenso unwahrscheinlich ist ein mutwilliger Angriff auf sieben Server gleichzeitig.

Die verteilt archivierten digitalen Objekte werden automatisch im Rahmen des Systems regelmäßig auf ihre Integrität überprüft. Wenn zum Beispiel die Originalversion eines digitalen Objekts (z. B. der Zeitschriftenartikel eines Verlags) Veränderungen erfährt, reagiert die LOCKSS-Software auf die Veränderung, indem eine neue Version dieses Objekts im Archival Unit (AU = Speichereinheit) archiviert wird. Diese neue Version gilt dann auch als neue Originalversion, da die Version, die im Internet vorliegt, immer die autoritative Version ist. Die alte Version wird jedoch nicht gelöscht, sondern auch weiterhin archiviert. Durch diese Versionenkontrolle wird auch die Authentizität der digitalen Objekte gewährleistet. Auch der Kontext der digitalen Objekte wird durch die gleichzeitige Speicherung der URL gewährleistet.

LOCKSS wurde in Deutschland zu Anfang oft als reines Speichersystem verkannt und daher nicht als Lösung zur Langzeitarchivierung digitaler Objekte gesehen. LOCKSS ist jedoch in andern Teilen der Welt für seine Vielfältigkeit bekannt und bietet ein robustes System, fähig der Wahrung von Integrität und Authentizität, sowie der Benutzbarkeit der gespeicherten Objekte. Das weltweite LOCKSS-Netzwerk hat zudem reale Angriffe getestet und abgewehrt.

„Military intelligence seeks to develop so-called ‚appreciations‘ of a potential adversary’s ‚capabilities‘ (what the adversary *could* do) and ‚intentions‘ (what the adversary is *expected* to attempt with the capabilities) as a basis for planning. Similarly, plans and techniques for defending distributed systems exposed to the Internet need an appreciation of the capabilities and intentions of the adversary they may encounter when deployed. Our assessment identified the following probable adversary capabilities:

- Unlimited Power
- Unlimited Identities
- Conspiracy
- Eavesdropping and Spoofing
- Exploiting Common Vulnerabilities

<sup>4</sup> LOCKSS News vom 14. Juni 2010, <<http://lockss.org/lockss/News>>.

– Uncovering Secrets” (Rosenthal, Maniatis, Roussopoulos, Guili, Baker, 2003)

Auch Formatmigrationen wurden erfolgreich getestet, indem veraltete Objekte durch neue Versionen in einem anderen Format ersetzt wurden. Dieser Test wurde, ebenso wie die Angriffsszenarien, weltweit im internationalen LOCKSS-Netzwerk durchgeführt. „Our implementation is capable of transparently presenting content collected in one Web format to readers in another Web format, with no changes needed to browsers. The reader need take no special action to cause this to happen, nor even be aware that it is happening.” (Rosenthal, Lipkis, Robertson, Morabito, 2005) LOCKSS war das erste Langzeitarchivierungssystem, das ein solches On-the-Fly-Migrationsverfahren<sup>5</sup> getestet und erfolgreich implementiert hat. Trotz dieses bereits getesteten Verfahrens ist LOCKSS neutral in Bezug auf Benutzbarkeits- und Lesbarkeitsverfahren. Gerade diese Neutralität machen wir uns im Projekt LuKII, im Zuge der Interoperabilitäts Herstellung zwischen LOCKSS und kopal, zu nutzen.

Ein weiterer positiver Aspekt von LOCKSS ist die finanzielle Unabhängigkeit von externen Geldgebern. LOCKSS finanziert sich zu 100 Prozent durch Beiträge von Mitgliedern der LOCKSS-Alliance.

„Voluntary dues from the LOCKSS Alliance members pay the salaries and operating costs for the six to eight members of the Stanford LOCKSS team. LOCKSS also pays Stanford an overhead fee for use of its facilities. Stanford does not fund LOCKSS, nor do any foundations.” (Seadle, 2010)

Zu Beginn hat LOCKSS natürlich auch Fördergelder erhalten, diese flossen seitens der National Science Foundation, SUN Microsystems Laboratories, Andrew W. Mellon Foundation, Library of Congress und des Joint Information Systems Committee (JISC).

Technischer Support für LOCKSS-Nutzer wird momentan hauptsächlich seitens des LOCKSS-Teams an der Stanford University und zu einem kleineren Anteil seitens des Digital Curation Centre in Schottland geleistet. LuKII, und damit Deutschland, soll zunächst im deutschsprachigen Raum, und dann auch im Rahmen der Europäischen Union, diese technische Unterstützungsfunktion (ich verweise auf das zweite Ziel von LuKII) leisten.

5 Es werden momentan zwei Arten der Migration unterschieden. Zum einen die prophylaktische Migration und zum anderen die On-the-Fly Migration. Eine prophylaktische Migration versucht vorab festzustellen, welche digitalen Objekte in einem LZA-System demnächst auf Grund von Formatalterung nicht mehr benutzbar sind. Daraufhin werden alle Objekte dieses Formattyps in ein neues, aktuelles Format migriert. Diese Art der Migration erfordert eine hohe Rechnerleistung, da meistens viele Dokumente auf einmal migriert werden sollen. Die digitalen Objekte sind während einer solchen Migration auch nicht benutzbar. On-the-fly-Migration betrifft immer nur ein digitales Objekt, das genau in dem Moment, von z. B. einem Bibliotheksbenutzer, angefragt wurde. Wenn sich herausstellt, dass der Rechner, an dem der Benutzer arbeitet, z.B. kein Tiff-Format mehr darstellen kann, dann würde dieses digitale Objekt im Tiff-Format in ein von dem Rechner lesefähiges Format (z. B. jpg) umgewandelt. Diese Migration benötigt nur wenig Rechnerleistung und passiert für den Nutzer völlig unbewusst.

Was oft für Missverständnisse in Bezug auf LOCKSS gesorgt hat, ist die allgemein gültige Vorstellung, dass ein Langzeitarchivierungssystem für digitale Objekte irgendwo extern sitzt, digitale Objekte dort gespeichert sind und diese nur bei Havarien wieder hervor geholt werden, um dann in das System der Gedächtnisorganisation wieder eingespeist zu werden. LOCKSS besitzt jedoch, im Gegensatz zu anderen auf dem Markt vorhandenen digitalen Langzeitarchivierungssystemen, auch eine Anzeigekomponente (in Form eines transparenten Proxy). Das bedeutet, die gespeicherten digitalen Objekte werden von LOCKSS selber, z. B. für einen Bibliotheksbenutzer wieder ausgegeben, wenn notwendig. So würde ein Bibliotheksbenutzer nicht bemerken, ob er die Kopie eines digitalen Dokuments vom Verlag vorliegen hat oder die Kopie aus einer LOCKSS-Box.

LOCKSS behauptet nicht, alle Aspekte der digitalen Langzeitarchivierung abzudecken, aber die Vorteile liegen klar in der Wirtschaftlichkeit des Systems, in der Robustheit, der Benutzbarkeit der digitalen Objekte und der Wahrung der Integrität und Authentizität der selbigen.

### 3 kopal

kopal ist ein deutsches Langzeitarchivierungssystem, das im Rahmen eines durch das BMBF (Bundesministerium für Bildung und Forschung) geförderten Projekts entstanden ist. Projektpartner waren die DNB (Die Deutsche Nationalbibliothek), die SUB Göttingen (Niedersächsische Staats- und Universitätsbibliothek Göttingen), das GWDG (Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen) und IBM Deutschland. Das Projekt wurde 2007 abgeschlossen, seitdem gibt es zwei aktive kopal-Nutzer, die Deutsche Nationalbibliothek und die SUB Göttingen. Ursprung des Projekts war die Besorgnis der Bibliotheken um die seit Beginn des Computerzeitalters immer wieder neu erscheinenden Dateiformate, die an bestimmte Rechnertypen und Betriebssysteme gebunden sind und somit eine langfristige Nutzbarkeit und Speicherung der digitalen Objekte erschweren. Ziel war es, eine technische und organisatorische Lösung zur Langzeitarchivierung digitaler Dokumente zu entwickeln, die sich an internationalen Standards und vorhandenen Bibliothekssystemen orientiert. Das Augenmerk auf Gedächtnisorganisationen zu haben, war von Anfang an beabsichtigt.

kopal besteht im Wesentlichen aus zwei Kernelementen, zum einen ist das die proprietäre Software DIAS (Digital Information Archiving System) der Firma IBM, und zum anderen ist es die selbstentwickelte Open-Source-Software koLibRI (kopal Library for Retrieval and Ingest). Kopal benutzt OAIS-kompatibles Vokabular<sup>6,7</sup>, so beinhaltet DIAS das Herzstück von kopal, die Elemente AIP (Archival Information Package(s)) also das Archival Storage als auch die Administration sowie das Preservation Planning einschließlich des Data Management mit Ingest und Access Bereich.

6 <<http://public.ccsds.org/publications/archive/650x0b1.pdf>> (Seite 38).

7 Weitere Informationen zum OAIS-Modell finden Sie hier: <[http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor\\_handbuch\\_artikel\\_62.pdf](http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_62.pdf)>.

koLibRI ist eine nach dem Bausteinprinzip entworfene Java-Bibliothek, die sogenannte SIPs (Submission Information Packages) und DIPs (Dissemination Information Packages) herstellt und diese sowohl in das Kernelement DIAS einspeisen als auch wiederauffinden und ausgeben kann. Die AUs werden in Form eines universellen Objektformats (UOF) erstellt, das im Rahmen von kopal spezifiziert wurde. Dieses universelle Objektformat definiert die Art und Weise der Speicherung der digitalen Objekte im System. Diese Objektspezifizierung benutzt ebenso die Begrifflichkeiten des OAIIS-Modells<sup>8</sup>.

## 4 koLibRI

koLibRI ist, wie bereits erwähnt, ebenso wie LOCKSS eine Open-Source-Software und bietet den Rahmen für die Integration eines Langzeitarchivs (im existierenden Fall DIAS) in die Infrastruktur einer Institution.

koLibRI wurde in der plattformunabhängigen Programmiersprache Java erstellt und besteht aus einzelnen Modulen. Es wurde bewusst so konzipiert, dass es als Ganzes oder in Teilen von anderen Systemen benutzt werden kann.

„Intention ist es, dass Gedächtnisorganisationen und andere an der Langzeitarchivierung elektronischer Materialien interessierte Einrichtungen diese Tools analysieren und evaluieren, um zu sehen, ob und inwieweit diese Tools potentiell zur eigenen Software-Architektur bzw. den eigenen Archivierungskonzepten passen.“ ([http://kopal.langzeitarchivierung.de/index\\_koLibRI.php](http://kopal.langzeitarchivierung.de/index_koLibRI.php))

Diese Eigenschaft macht sich LuKII im Rahmen des Projekts zunutze, da koLibRI nach momentaner Planung hauptsächlich für den Ingestbereich zuständig sein soll, um von Webseiten geharvestete digitale Objekte mit Hilfe von JHOVE Version 1.5 (JSTOR/Harvard Object Validation Environment<sup>9</sup>) mit derzeitig weiteren technischen Metadaten zu bereichern und in das UOF zu bringen. Die jeweiligen Module übernehmen im Rahmen von kopal momentan den Vorgang der Extraktion von Metadaten und die Validierung einer Datei.

## 5 Interoperabilität

Zwei Voraussetzungen müssen für die Interoperabilität der beiden Systeme geschaffen werden. Zum einen muss koLibRI in der Lage sein, mit LOCKSS zu kommunizieren, und zum anderen muss LOCKSS in der Lage sein, Kommunikation mit koLibRI zu verstehen und „lernen“, Metadaten in einem standardisierten Format (z. B. METS, WARC) anzureichern und als solche zu erfassen.

Um ersteres zu erfüllen, müssten die betroffenen Java-Module in koLibRI, die momentan mit DIAS kommunizieren, derart umprogrammiert werden, dass sie mit LOCKSS

interagieren können. Das betrifft vor allem den sogenannten Ingestbereich, also koLibRI-Module für den Ingest, zur Metadatenanreicherung, zur Erstellung des UOF und vor allem die SQL-Schnittstelle zur DIAS-Datenbank.

Veränderungen des Inhalts einer LOCKSS-Box finden immer nur über Veränderungen des Originals (der sogenannten Masterkopie auf der Webseite des host / Datenbankbetreibers, z. B. ein Verlag) statt<sup>10</sup> oder wenn fehlerhafte Dokumente entdeckt und repariert werden müssen.

Im ersten Fall kommen die Informationen von der Webseite des Informationsanbieters, im zweiten Fall kommen die Informationen von LOCKSS-Boxen, die die gleichen Inhalte speichern. Überprüfungen und Wiederherstellungen des Inhalts innerhalb eines LOCKSS-Systems erfolgen via automatisierten Anfragen und geschehen in regelmäßigen Intervallen.

Es ist immer nachvollziehbar, wann wer welche Veränderungen gemacht hat. Diese zwei typischen Vorgehensweisen von LOCKSS dienen der Integritätssicherung, erschweren jedoch die Anreicherung mit Metadaten, die koLibRI im Rahmen von kopal bereits leistet, und die LuKII im Rahmen des Projekts in LOCKSS integrieren will. So muss auch LOCKSS dementsprechend geändert werden. Eine Speichereinheit in LOCKSS enthält momentan nur elementare, technische Metadaten, also neben der URL des digitalen Objekts Informationen zum Zeitpunkt des Crawlens und den MIME-Type. JHOVE kann jedoch sehr viel umfangreichere technische Metadaten für gängige MIME-Types erzeugen, die momentan in LOCKSS nicht als Metadaten zu einem digitalen Objekt dazugehörend gespeichert würden. Diese Art von Informationen würde momentan nur als weiteres digitales Objekt unabhängig vom Bezugsobjekt gespeichert werden. Um diese Fähigkeit zu erreichen, gibt es zwei von LuKII entwickelte Szenarios, die u. a. WARC-files<sup>11</sup> verwenden. Das LOCKSS-Team in Stanford hat die LOCKSS-Software bereits bearbeitet, um WARC-file kompatibel zu sein. Es ist geplant, dass in der neusten, von Stanford freigegebenen Version von LOCKSS die Speichereinheiten bereits als WARC-file abgelegt werden können. Ebenso werden WARC-files anstatt eines eigenen Crawl durch LOCKSS zur Datengewinnung verwendet werden können, d. h. die LOCKSS-Box nutzt die Informationen aus dem WARC-file so, als hätte er die entsprechende Webseite selber geharvestet. Im Rahmen von LuKII muss LOCKSS dementsprechend weiter ausgearbeitet werden, dass die durch koLibRI zusätzlich erzeugten Metadaten mit in diese WARC-files integriert werden. Die Metadaten sind dann nicht mehr, wie zuvor, als eigenes digitales Objekt in einer eigenen Speichereinheit vorhanden, sondern Teil der besagten Speichereinheit. Seitens

8 Weiterführende Literatur zum Universellen Objektformat finden Sie hier: [http://kopal.langzeitarchivierung.de/downloads/kopal\\_Universelles\\_Objektformat.pdf](http://kopal.langzeitarchivierung.de/downloads/kopal_Universelles_Objektformat.pdf).

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

10 Wenn z. B. eine Publikation eines Verlages seitens des Verlags überarbeitet wird, erfährt die LOCKSS-Box durch regelmäßiges Crawl der Verlagswebseite von dieser Veränderung. Die neue Version wird dann als Version 1.1. im LOCKSS-System gespeichert, die alte Version wird nicht gelöscht, sondern wird weiter als Version 1.0 gespeichert.

11 „The WARC (Web ARChive) format specifies a method for combining multiple digital resources into an aggregate archival file together with related information.“ <http://www.digitalpreservation.gov/formats/fdd/fdd000236.shtml>.

der Deutschen Nationalbibliothek werden einzelne Module von koLibRI ebenso umprogrammiert, um WARC-files erstellen und lesen zu können.

Der Ablauf zur Erzeugung einer Speichereinheit wäre im Rahmen des ersten Szenarios dann wie folgt: Digitale Objekte aus dem Netz würden demnach wie üblich mit Hilfe des LOCKSS-Crawlers geharvestet und mit den einfachen technischen Metadaten in der LOCKSS-Box archiviert. Danach wird koLibRI die betreffenden digitalen Objekte aus der LOCKSS-Box exportieren, um wie bereits bewährt mit Hilfe von JHOVE komplexere, technische Metadaten zu dem digitalen Objekt hinzuzufügen. Anschließend wird koLibRI diese angereicherten digitalen Objekte erneut in die LOCKSS-Box einspeisen. Das resultierende digitale Objekt wäre dann nicht eine neue Speichereinheit, sondern dieselbe Speichereinheit, die zusätzliche Informationen enthält. In Schaubild 1 ist eine vereinfachte Darstellung des Vorgangs zu sehen.

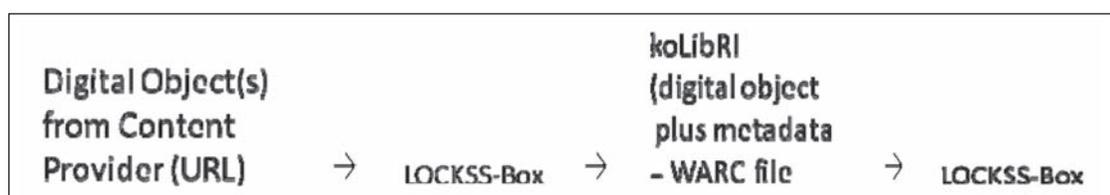


Schaubild 1: Metadatenanreicherung mit Hilfe von WARC-files Szenario 1

Im Rahmen des zweiten Szenarios würde koLibRI als Ingestkomponente agieren, um ebenfalls mit den so erhaltenen Daten WARC-files zu erzeugen, um diese dann wieder selbständig auf eine Webseite, sprich auf einen Server, der HTTP versteht, zu laden. Daraufhin würde die LOCKSS-Box diese neu erzeugten WARC-files als digitales Objekt von dem entsprechenden Server holen. LOCKSS würde dann nicht mehr das Web selber crawlen, sondern als reines Speichersystem agieren. In Schaubild 2 ist eine vereinfachte Darstellung dieses Vorgangs zu sehen.



Schaubild 2: Metadatenanreicherung mit Hilfe von WARC-files Szenario 2

Es ist angestrebt, beide Prototypen mit Open-Access-Inhalten aus deutschen Institutionellen Repositorien zu testen.

LuKII beschäftigt sich ferner mit der Integration von beschreibenden Metadaten in WARC files. Die beschreibenden Metadaten könnten über eine OAI-PMH-Schnittstelle aus den Institutionellen Repositorien extrahiert werden. Zumindest die durch DINI zertifizierten Server bieten bereits qualitativ hochwertige beschreibende Metadaten an.

## 6 Organisatorische Aspekte

Zu Beginn des Projektes wurden seitens der DNB und HU ausgewählte wissenschaftlich orientierte Gedächtnisorganisationen in Deutschland angefragt, ob sie Interesse haben, im Rahmen von LuKII mitzuwirken. Das Hauptaugenmerk bei der Auswahl lag auf der technischen Infrastruktur, die in jeder Einrichtung vorhanden sein sollte. Jede partizipierende Institution muss als Teil des Projekts eine LOCKSS-Box einrichten. Das Projekt LuKII kann durch die Natur der LOCKSS-Struktur nur mit Hilfe von Partnern realisiert werden.

Das erste Ziel ist, wie erwähnt, der Aufbau eines Netzwerks an Institutionen in Deutschland, die sich aktiv am Aufbau eines digitalen Langzeitarchivierungsnetzwerks beteiligen, denn ein Hauptziel ist, eine Infrastruktur für digitale Langzeitarchivierung in Deutschland zu entwi-

ckeln. Teilnehmende Institutionen sind die Bayerische Staatsbibliothek, das hbz (Hochschulbibliothekszenrum NRW) in Köln, das Karlsruher Institut für Technologie, die Niedersächsische Staats- und Universitätsbibliothek Göttingen, die Sächsische Landesbibliothek – Staats- und Universitätsbibliothek in Dresden, die Staatsbibliothek zu Berlin – Preussischer Kulturbesitz, die Universität Konstanz, die Universitätsbibliothek Stuttgart und die Universitäts- und Landesbibliothek Münster. Seit Dezember 2010 steht die Basis der Infrastruktur, und wir erwarten,

bis spätestens Ende März 2011 das Harvesting des ersten Dokumentenservers (edoc-Server der HU) abgeschlossen zu haben.

Teil einer funktionierenden Infrastruktur ist natürlich auch die Fähigkeit, ein permanentes Betreuungsangebot für technische und organisatorische Belange zu bieten. Dieser Support wird seitens der Humboldt-Universität zu Berlin geleistet werden. Der Computer- und Medienservice der HU

hat zunächst, im Rahmen von LuKII, das Kompetenzzentrum aufgebaut, um Aufgaben wie

- die Installation von LOCKSS,
- die Konfiguration von LOCKSS,
- die Konfiguration und Implementierung von LOCKSS-Plugins,
- die Konfiguration von ACCESS,
- und Troubleshooting

zu unterstützen

Zukunftsfähigkeit und Nachhaltigkeit waren bereits im Projektantrag wichtige Aspekte: „For LuKII sustainability comes in two forms“ (Projektantrag LuKII, 2009). Zum einen soll Nachhaltigkeit erreicht werden, indem gegen Ende des Projekts eine ausreichende Menge an LOCKSS-LuKII-Nutzern vorhanden ist, die nicht nur mit Hilfe des Kompetenzzentrums existieren können, sondern sich auch gegenseitig helfen können. PLNs weltweit verfolgen diese Struktur bis dato erfolgreich<sup>12</sup>. Dadurch, dass technisches Personal in deutschen Bibliotheken im Vergleich zu amerikanischen Bibliotheken zwar weniger vorhanden ist, aber länger an einem Arbeitsplatz bleibt, geht LuKII auch davon aus, dass das notwendige Wissen, ein LOCKSS-Netzwerk zu erhalten, in Deutschland länger stabil bleibt.

Zum anderen beruht die Zukunftsfähigkeit des Projekts, und damit der digitalen Langzeitarchivierung mit LOCKSS in Deutschland, auch auf der Entwicklung einer systematischen, finanziellen Unterstützung, die eventuell durch partizipierende Institutionen (ähnlich der LOCKSS Alliance) getragen wird.

LuKII ist des Weiteren in beratender Funktion an der National Hosting Debatte (Establishing a Federated Strategy on Perpetual Access and Hosting of Electronic Resources for Germany)<sup>13</sup> beteiligt. Wir sehen eine deutsche, durch LuKII etablierte LOCKSS-Gemeinschaft als gute Möglichkeit, die National-Hosting-Strategie zu unterstützen.

Aktuelle Informationen zum Projekt finden Sie auch auf unserer Homepage: <<http://www.ibi.hu-berlin.de/forschung/digibib/forschung/projekte/LuKII>>.

12 „Beispiele für PLNs, die Nachhaltigkeit u. a. mit Elementen der gleichen Strategie verfolgen, sind z.B. das MetaArchive (<<http://www.metaarchive.org/>>), COPPUL (<<http://www.coppul.ca/pln.html>>) oder auch PeDALS (<<http://pedalspreservation.org/Default.aspx>>).

13 <[http://www.allianzinitiative.de/en/core\\_activities/national\\_hosting\\_strategy/](http://www.allianzinitiative.de/en/core_activities/national_hosting_strategy/)>.

## Literaturverzeichnis

- Rosenthal, David; Maniatis, Petros; Roussopoulos, Mema; Guili, TJ; Baker, Mary: Notes On The Design Of An Internet Adversary. In: Adaptive and Resilient Computing Security Workshop, November, (2003), <<http://www.eecs.harvard.edu/~mema/publications/adversary.pdf>>.
- Rosenthal, David; Lipkis, Thomas; Robertson, Thomas S.; Morabito, Seth: Transparent Format Migration of Preserved Web Content. In: D-Lib Magazine. Bd. 11/1. Bath (2005), <<http://www.dlib.org/dlib/january05/rosenthal/01rosenthal.html>>.
- Seadle, Michael: A Social Model for Archiving Digital Serials: LOCKSS. In: Serials Review. Bd. 32/2. Elsevier (2006), S. 73–77.
- Seadle, Michael: Archiving in the networked world: LOCKSS and national hosting. In: Library Hi Tech. Bd. 28/4. Emerald Group Publishing, Bingley (2010) S. 710–717.

## Quellen im Internet

Öffentliche Version von OAIS als CCSDS Blue Book (Januar 2002), <<http://public.ccsds.org/publications/archive/650x0b1.pdf>> (zuletzt aufgerufen am 15.12.2010).  
Kopal – kooperativer Aufbau eines Langzeitarchivs digitaler Informationen / koLibRI: kopal, Library for Retrieval and Ingest, <[http://kopal.langzeitarchivierung.de/index\\_koLibRI.php.de](http://kopal.langzeitarchivierung.de/index_koLibRI.php.de)> (zuletzt aufgerufen am 28.12.2010).

WARC, Web ARChive File Format, <<http://www.digitalpreservation.gov/formats/fdd/fdd000236.shtml>> (zuletzt aufgerufen am 02.01.2011).

## Weitere Quellen

Antrag auf Förderung durch die DFG im Förderungsbereich Wissenschaftliche Literaturversorgungs- und Informationssysteme (LIS). Antragsteller: Niggemann, Elisabeth; Seadle, Michael; Schirnbacher, Peter: LOCKSS und KOPAL – Infrastruktur und Interoperabilität. Berlin/Frankfurt a.M. (2009) S. 1-34.