

Stefan Hein, Karlheinz Schmitt, Virginie Werb

## LuKII – LOCKSS-und-kopal- Infrastruktur-und-Interoperabilität



Das Problem digitaler Langzeitarchivierung ist bis heute ungelöst. Die weltweiten Bemühungen an sowohl praktikablen als auch wirtschaftlichen Lösungen für diese Aufgaben stellen heute eine Vielzahl an Lösungsansätzen mit teilweise sehr unterschiedlichen Schwerpunkten und Herangehensweisen zur Verfügung.

Alle bisherigen Erfahrungen und Bestrebungen auf dem Gebiet der Langzeitarchivierung führten allgemein zu der Erkenntnis, dass eine wirtschaftliche und praktisch umsetzbare Langzeitarchivierung am sinnvollsten kooperativ geleistet werden sollte. Diesem Grundgedanken folgt das Projekt LuKII.

Das Ziel dieses Projektes ist es, zwei existierende Langzeitarchivierungsansätze LOCKSS (Lots of Copies Keep Stuff Save<sup>1)</sup>) und kopal (kooperativer Aufbau eines Langzeitarchivs digitaler Informationen<sup>2)</sup>) miteinander zu kombinieren, um auf diese Weise die Vorteile beider Systeme miteinander zu einem Gesamtsystem zu vereinen. Hierdurch soll innerhalb Deutschlands ein Langzeitarchivierungsnetzwerk aufgebaut werden, welches zum einen die bewährten und kostengünstigen Fähigkeiten von LOCKSS zur Bewahrung der Bitstreamintegrität digitaler Objekte besitzt. Zum anderen sollen die nicht minder erprobten Fähigkeiten von kopal und der in diesem Projekt entwickelten Werkzeuge koLibRI (kopal Library for Retrieval and Ingest)<sup>3)</sup> zur Erhaltung der Nutzbarkeit (z. B. durch Migration) digitaler Objekte Verwendung finden.

Das LuKII-Projekt verfolgt drei Hauptziele:

- Die Etablierung eines LOCKSS-Netzwerks in Deutschland,
- die Konzeption und Implementierung der Interoperabilität von LOCKSS und kopal sowie
- die Durchführung von Tests der Interoperabilität durch Archivierung ausgewählter deutscher Institutsrepositorien.

LuKII ist ein von der Deutschen Forschungsgemeinschaft (DFG) gefördertes Projekt, das sowohl von der Humboldt-Universität zu Berlin (HUB) als auch von der Deutschen Nationalbibliothek (DNB) kooperativ durchgeführt wird. Die Gesamtprojektleitung wurde von der HUB übernommen. Insgesamt teilt sich das Projekt in vier Arbeitspakete und endet im Oktober 2011 nach einer Projektlaufzeit von zwei Jahren. Für das Erreichen des ersten Projektziels, dem Aufbau eines deutschlandweiten LOCKSS-Netzwerks, konnten auch die Universität Regensburg, die Bayerische Staatsbibliothek München, die Staatsbibliothek zu Berlin - Preußischer Kulturbesitz, das Forschungszentrum Jülich, die Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB) und das Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz) gewonnen werden. Die Institutionen beteiligen sich jeweils als ein Knoten des LOCKSS-Netzwerks und stellen durch die Anbindung ihrer Repositorien zudem eine ausgewogene und repräsentative Datenmenge zur Evaluierung von LOCKSS zur Verfügung.

Für die Umsetzung der Interoperabilität zwischen LOCKSS und kopal besteht darüber hinaus ein reger Austausch mit den LOCKSS-Entwicklern aus Stanford.

Neben den bereits genannten verbleibenden Zielen, die sich auf die Arbeitspakete zwei und drei aufteilen, verfolgt die HUB im letzten Arbeitspaket die Anbindung an CLOCKSS (Controlled LOCKSS). CLOCKSS stellt ein zugangskontrolliertes, auf LOCKSS basierendes Speichersystem dar, das weltweit wissenschaftliche Publikationen von Verlagen und Universitäten langzeitarchiviert.

### LOCKSS

Im Jahr 2004 wurde an der Stanford University (USA) eine erste produktive Version der LOCKSS-Software unter einer Open-Source-Lizenz veröffentlicht.

Vielzahl an Lösungsansätzen zur Langzeitarchivierung

Projektziel

Projektstruktur

Arbeitspakete und Projektteilnehmer

Anbindung an CLOCKSS

licht. Die Weiterentwicklung und Pflege des Systems wird zurzeit von Stanford und der LOCKSS Alliance, einer Organisation bei der Bibliotheken kostenpflichtig Mitglied werden und somit finanzielle Unterstützung leisten, betrieben. Die mittlerweile über 200 Bibliotheken umfassende Allianz, bildet insbesondere in Großbritannien in Zusammenarbeit mit dem Joint Information Systems Committee (JISC)<sup>4)</sup> einen Schwerpunkt im Bereich des technischen Supports.

Das LOCKSS-Konzept beruht auf einem bestehend einfachen Prinzip. Je mehr Kopien es von einem digitalen Objekt gibt, desto wahrscheinlicher ist dessen Erhalt für die Zukunft. Mithilfe von Peer-to-Peer-Verbindungen sind alle Teilnehmer des LOCKSS-Netzwerks miteinander verknüpft, sodass auf möglichst vielen (mind. sieben) Knoten des Netzwerks eine Kopie des digitalen Objekts auf einfachste Weise verteilt werden kann. Aufgrund regelmäßiger Vergleiche können Fehler im Bitstream der digitalen Objekte erkannt und korrigiert werden. Ein weiterer Vorteil dieses Systems besteht in den ausdrücklich geringen Hardwareanforderungen und damit sehr geringen Hardwarekosten für die beteiligten Institutionen.

## kopal

kopal wurde im Rahmen eines 3-jährigen BMBF-Projektes im Zeitraum von 2005 bis 2007 von der DNB und der SUB entwickelt.<sup>5)</sup> Auf der Grundlage des Open Archival Information System (OAIS) – ISO-Standards, dem anerkannten Referenzmodell für digitale Langzeitarchivierungssysteme, ist in Deutschland hierdurch erstmalig ein kooperativ betriebenes Langzeitarchivierungssystem entstanden. Im Gegensatz zu LOCKSS besitzt kopal ein erprobtes Preservation-Planning-Konzept, insbesondere in Bezug auf die Unterstützung für Formatmigrationen. Andererseits erfordert der Einstieg in das ursprüngliche kopal-System durch die unterliegende Hardware-Struktur sehr hohe Anfangsinvestitionen, die berücksichtigt werden müssen. Außerdem fallen durch das Präventivkonzept (Objekte werden dann migriert, wenn sie als gefährdet gelten müssen, nicht erst im Fall eines Zugriffs) laufend Aufwände zur Beobachtung und

Überwachung von Entwicklungen und zur Durchführung von Preservation Planning-Maßnahmen an.

## Arbeitsschritte

Im ersten Schritt ist der Aufbau eines LOCKSS-Netzwerks in Deutschland vorgesehen. Hierbei richten alle beteiligten Institutionen so genannte LOCKSS-Boxen ein, die als Knoten im LOCKSS-Netzwerk fungieren.

Da LOCKSS auf Open-Source-Software basiert, fallen keine Lizenzkosten an. Innerhalb einer Institution kann als LOCKSS-Knoten ein handelsüblicher Computer mit ausreichend dimensionierter Speicherkapazität verwendet werden, sodass die anfallenden Kosten für die Einrichtung eines LOCKSS-Knotens marginal bleiben. Bei der Auswahl der Hardware muss auf Kompatibilität zu anderen LOCKSS-Knoten keinerlei Rücksicht genommen werden. Aus diesem Grund ist es für eine beteiligte Institution auch jederzeit möglich die verwendete Speichertechnologie teilweise oder vollständig zu wechseln. Fehlendes Archivmaterial wird automatisch wieder von anderen Mitgliedern des LOCKSS-Netzwerkes ergänzt.

Im zweiten Schritt soll zuerst ein Konzept erarbeitet werden mit dem die Symbiose beider Langzeitarchivierungssysteme hergestellt werden kann. Ein denkbares Szenario wäre beispielsweise, das bisher in kopal verwendete Speichersystem (DIAS) durch das LOCKSS-Netzwerk zu ersetzen und die notwendigen Schnittstellenanpassungen vorzunehmen. Prototypisch wird das konzipierte Gesamtsystem im Anschluss an die Konzeptionsphase umgesetzt, sodass erste Erfahrungen bezüglich der Praxistauglichkeit gesammelt werden können.

Um nachhaltige Aussagen in Bezug auf die zu erprobende Praxistauglichkeit zu erhalten, wird im dritten Schritt eine repräsentative Auswahl an digitalen Medienwerken aus den Repositorien aller beteiligten Institutionen in das kombinierte Langzeitarchivierungssystem eingespielt und hinsichtlich verschiedener qualitativer Kriterien (z. B. Leistungsverhalten) und funktionaler Anforderungen (z. B. Zurückspielen der archivierten Objekte) getestet.

Aufbau eines LOCKSS-Netzwerks in Deutschland

Konzept zur Symbiose von LOCKSS und kopal

Beitritt der HUB  
zu CLOCKSS

Über die gesamte Projektlaufzeit erstrecken sich die Aufgaben des Arbeitspakets vier, das den Beitritt der HUB zu CLOCKSS umfasst.

## Aktueller Stand

Für eine gemeinsame Kommunikations- und Dokumentationsplattform wurde an der HUB ein Wiki eingerichtet, das allen beteiligten Partnern zur Verfügung steht. Zur Unterstützung wurde das offizielle Tutorial für die Installation einer LOCKSS-Box ins Deutsche übersetzt und im Wiki hinterlegt. Diese Bemühungen sind der erste Schritt zur Etablierung eines deutschen LOCKSS-Kompetenz-zentrums, das seinen Sitz an der HUB haben wird. Zu dessen Aufgaben gehört primär die Unterstützung bei der Installation und Konfiguration der LOCKSS-Software und der im Projekt entwickelten Softwarebausteine.

Alle Projektbeteiligten haben jeweils nach Schaffung der nötigen technischen Voraussetzung, die

Installation der LOCKSS-Box abgeschlossen. Jeder Knoten arbeitet auf Basis der aktuellsten LOCKSS-Software und bildet gemeinsam mit allen anderen Knoten das zugangsbeschränkte LuKII Privat LOCKSS-Netzwerk.

Aktuell werden die softwaretechnischen Anforderungen umgesetzt, um die auf verschiedenen Systemen basierenden Repositorien (z. B. OPUS, edoc-Server) mit LOCKSS zu verknüpfen. Nach Fertigstellung der dafür benötigten Software können die Inhalte der Repositorien in das LOCKSS-Netzwerk übertragen werden.

Parallel dazu befindet sich das Konzept für die Interoperabilität zwischen LOCKSS und kopal in Bearbeitung. Hierbei gilt es einen gemeinsamen Weg zu finden, wie die beiden Ansätze zu einem Gesamtsystem für die Langzeitarchivierung kombiniert werden können. Welche Metadaten durch koLibRI geliefert werden, wie diese abgespeichert und welches Archivpaketformat verwendet werden soll, sind nur einige der noch offenen Fragen, die beantwortet werden müssen.

Umsetzung soft-  
waretechnischer  
Anforderungen

Wiki als Kommu-  
nikationsplattform

## Anmerkungen

1 <<http://www.lockss.org/lockss/Home>>

2 Wollschläger, Thomas: Kopal goes live. In: Dialog mit Bibliotheken, 19 (2007) 2, S. 17 - 22.

3 <[http://kopal.langzeitarchivierung.de/index\\_koLibRI.php.de](http://kopal.langzeitarchivierung.de/index_koLibRI.php.de)>

4 <<http://www.jisc.ac.uk>>

5 <<http://kopal.langzeitarchivierung.de>>