

Thomas Wollschläger

[kopal goes live]

Ergebnisse und Fortführung von kopal an der Deutschen Nationalbibliothek

Das Ziel von kopal (Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen) war der Aufbau einer nachnutzbaren technischen und organisatorischen Infrastruktur zur Sicherung der Langzeitverfügbarkeit elektronischer Publikationen. Es handelte sich um ein Förderprojekt des Bundesministeriums für Bildung und Forschung (BMBF) über eine Laufzeit von drei Jahren. Das Projekt wurde am 30. Juni 2007 offiziell beendet. Über die Ausgangslage auf dem digitalen Publikationsmarkt, die Herausforderungen an die Langzeitarchivierung, die Ziele, den Projektaufbau und die Planung von kopal ist Anfang 2005 an dieser Stelle bereits berichtet worden¹⁾. Nunmehr ist es Zeit, auf den Projektverlauf zurückzublicken, eine Bilanz der Ergebnisse zu ziehen und den Weiterbetrieb des Archivs zu betrachten.

Den Kern des kopal-Archivsystems bildet das von IBM in Zusammenarbeit mit der Koninklijke Bibliotheek (KB), Den Haag, entwickelte Digital Information Archiving System (DIAS). Innerhalb des Projekts sind digitale Materialien fast aller Art der Deutschen Nationalbibliothek (DNB) und der Niedersächsischen Staats- und Universitätsbibliothek (SUB), Göttingen, in das Archiv eingestellt worden. Damit haben zwei sehr unterschiedliche Bibliotheken zusammengearbeitet, was sowohl den jeweiligen Auftrag (National- bzw. Universitätsbibliothek) betrifft, als auch Schwerpunkte bei den bisherigen Aktivitäten. Der technische Betrieb des Systems ist ausgelagert und bei der Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWGDG) angesiedelt. Der Partner IBM

Deutschland sorgte für eine professionelle Anpassung der Software und bot eine langfristig stabile Unterstützung. Im internationalen Rahmen erarbeiteten die Partner gemeinsam mit der KB Anforderungen an künftige Weiterentwicklungen von DIAS.

Das kopal-System erfüllt den Anspruch an ein universell nutzbares Archivsystem, indem zum einen die Sicherstellung der langfristigen Verfügbarkeit durch Migration und Emulation unterstützt wird. Zum anderen gibt es in kopal keinerlei Einschränkungen für die Art des Materials, welches in das Archiv eingespielt werden kann (Text, Bilder, Audio, Video), und für die möglichen Dateiformate. Obwohl das skalierbar angelegte kopal-System bisher eine begrenzte Gesamtkapazität hat, ist die Größe des einzelnen Archivobjekts nicht begrenzt. Jeder der Partner ist völlig frei in der Auswahl und Regelfestlegung beim Einspielen der von ihm gesammelten Objekte.

Die Nachnutzung von kopal durch weitere Institutionen, die eine Langzeitarchivierung benötigen, ist ausdrücklich erwünscht. Die kopal-Lösung ist von vornherein auf unterschiedliche Bedürfnisse ausgerichtet. Um die Nachnutzbarkeit zu gewährleisten, werden etablierte Standards genutzt. Der Transfer der Objekte in ein digitales Archiv über standardisierte Formate, Transportwege und Schnittstellen ist dabei ein wichtiges Erfordernis. Bereits das Kernsystem DIAS von IBM basiert auf bewährter Standardsoftware (wie der DB2-Datenbank, dem Content Manager und dem Tivoli Storage-Manager) und weist eine zukunftsfähige Trennung von Speicherkonzept und Datenverwaltung auf. Das Open Archival Information System-Referenz-

Künftige Weiterentwicklung von DIAS

Nachnutzung von kopal ist erwünscht

Rückblick Projektverlauf

Mandantenfähigkeit ist möglich

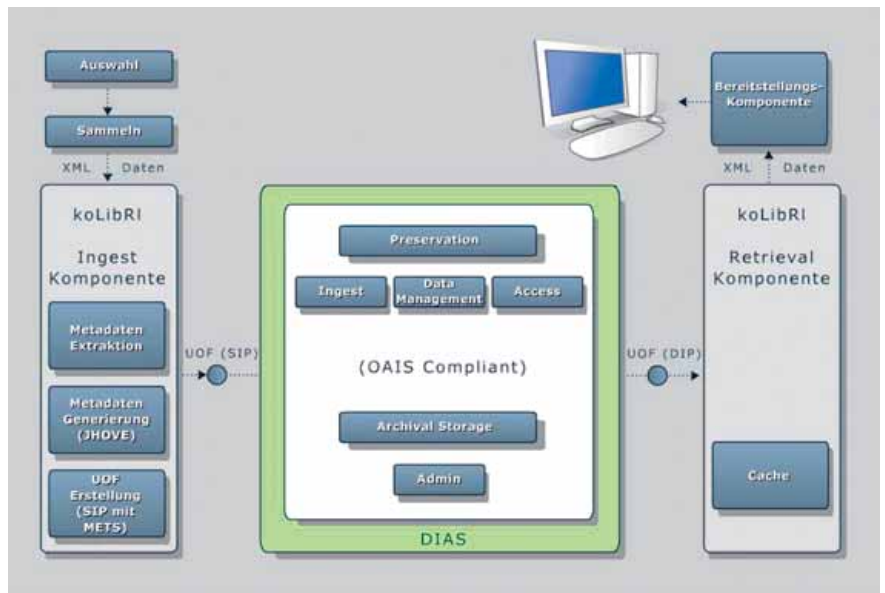
renzmodell (OAIS) für digitale Archivierung ist in DIAS konsequent implementiert. Die Erweiterbarkeit von DIAS für neue Nutzerinstitutionen und präzise definierte Import- und Exportschnittstellen ermöglichen es, eine Archivnutzung in verschiedenste Workflows unterschiedlichster Institutionen zu integrieren.

Das kopal-System erlaubt eine getrennte Nutzung durch verschiedene Institutionen (Mandantenfähigkeit). Vergleichbar getrennten Schließfächern, verfügt jede Institution über einen eigenen Bereich, in den nur sie Daten einstellen, verändern und abrufen kann.

Die Institutionen können per Fernzugriff auch über sichere Internetverbindungen auf das kopal-System (derzeit auf den Speicherstandort der GWDG in Göttingen) zugreifen. Upgrades des DIAS-Systems sind jedoch für alle Mandanten gemeinsam

gültig. Die Mandantenfähigkeit und der mögliche Fernzugriff sind die Kernvoraussetzungen für die Nutzbarkeit des kopal-Systems durch andere Gedächtnisorganisationen (Bibliotheken, Archive und Museen). Außerdem ist die DIAS-Software durch flexible Module erweitert worden. Hierfür haben die DNB und die SUB Göttingen auf DIAS abgestimmte Softwareprodukte erstellt, die als »kopal Library for Retrieval and Ingest« (koLibRI) unter einer Open Source Lizenz veröffentlicht worden sind. Bei diesen kopal-Tools

geht es hauptsächlich um den Bereich des Einspielens von Objekten in das DIAS-System sowie den Zugriff auf die archivierten Objekte. koLibRI ist eine generische freie Softwarebibliothek (basierend auf JAVA) zur Ein- und Anbindung unterschiedlicher Mandanten und der automatischen Erstellung von technischen Metadaten. koLibRI besitzt konfigurierbare Workflows. Das Projekt kopal hat dafür das »Universelle Objektformat«



Aufbau des kopal-Archivsystems.

vorgelegt, mit dem digitale Objekte zusammen mit Metadaten archiviert und zwischen Institutionen und Systemen ausgetauscht werden können.

Die Systementwicklung ist dabei so offen angelegt worden, dass eine Ausdehnung der kooperativen Nutzung auf weitere Nutzer aus dem Kreis aller Gedächtnisorganisationen möglich ist. Bereits seit März 2006 stand für Testzwecke in einer Beta-Version ein vorläufiger Release der neu erstellten koLibRI-Software zur Verfügung, der zu Beginn

Universelles Objektformat

des Jahres 2007 als Beta-2-Release aktualisiert worden ist. Mit dem Projektabschluss steht seit Juli 2007 ein nachnutzbares System und ein endgültiges Release der vollständig entwickelten koLibRI-Software (Version 1.0) bereit.

Nach einer Pilotphase zur Systemevaluierung und Entwicklung des Universellen Objektformats im Jahr 2004 folgte für kopal eine zweieinhalbjährige Entwicklungsphase. Ende 2005 wurden das bei der GWDG installierte System und die neu entwickelten Tools erfolgreich getestet. Im Jahr 2006 wurde das aus den Parametern des erfolgreich getesteten Referenzsystems abgeleitete eigentliche Produktivsystem von kopal aufgesetzt. Erklärtes Ziel war, zunächst anhand einer Vielzahl von Objekten in verschiedenen Formaten die praxisnahe Nutzung der kopal-Lösung aufzuzeigen. In einem nächsten Schritt waren die in kopal entwickelten Arbeitsabläufe prinzipiell an die Erfordernisse eines künftig in der Routine laufenden Einspielbetriebs in den Bibliotheken anzupassen. Zuletzt sind die Komponenten für Administration und Prozess-Monitoring im Archivsystem fertig gestellt sowie die Voraussetzungen geschaffen worden, um in kopal Migrations- und Emulationsprozesse durchzuführen, die eine langfristige Interpretierbarkeit der archivierten Dokumente sicherstellen.

Im August 2006 konnte die erste Stufe des Produktivbetriebs erfolgreich aufgenommen werden, wobei die Projektpartner DNB und SUB allein im ersten Durchgang über 40.000 zu archivierende Dokumente in das bei der GWDG gehostete System eingespielt haben. Weitere im Projektantrag dezidiert genannte einzuspielende Objektklassen bilden E-Jour-

nal-Artikel, Digitalisate und CD-ROMs bzw. DVDs. Letztere Objekte müssen vor dem Einspielen in nicht trägergebundene digitale Formen überführt werden; hierbei soll die Archivierung möglichst in Form von Images nach ISO 9660 erfolgen. In der DNB wurden hierzu von der Abteilung Benutzung exemplarische Mengen solcher Publikationen ausgewählt. Durch kopal erfolgten die Erstellung von Images einschließlich der Analysen auftretender Problemfälle und das Einspielen dieser Objekte in das Langzeitarchiv. Anhand dieser Objekte erfolgten auch wesentliche Anstöße für die Umgestaltung der Workflows in der Bibliothek.

Bisher existierten für die Verarbeitung von Onlinedokumenten mehr oder weniger halbautomatische Verfahren, außerdem gab es verschiedene Workflows für Netzpublikationen, Onlinedissertationen und weitere Materialien. Ziel ist die Schaffung eines automatischen, einheitlichen Verfahrens mit der Übergabe der Archivobjekte an kopal (Ingest) bzw. beim Access die Übergabe von kopal an Arbeitsplatzrechner oder das neu entstehende Bereitstellungssystem. Dabei sind zahlreiche Abteilungen in der DNB involviert: Das PICALTIS-Team, die Portal-Mitarbeiter, die Abteilung Erwerbung/Formalerschließung, die Nutzungsabteilung und nicht zuletzt die externen Ablieferer. Letztere benötigen für die Anbindung an die Prozesse in der DNB bzw. die Zulieferung in der Regel Unterstützung. Der Workflow von elektronischen Materialien auf Datenträgern (d. h., CD- bzw. DVD-Veröffentlichungen) bildet, wie bereits erwähnt, einen derjenigen, die aufgrund der Anforderungen seitens Archivsystem und künftiger Bereitstellung anzupassen sind. Nach der Er-

Ziel: Schaffung eines automatisierten Übergabeverfahrens

2006 kopal-Produktivsystem

Verschiedene Objektklassen

**Überarbeitung
des Workflows**

stellung der Images und Analysen wurde ein Änderungs- und Ergänzungsvorschlag für den Geschäftsgang dieser Materialien vorgelegt. Der Workflow wird derzeit überarbeitet. Ebenso wird der Workflow für genuin online vorliegende Netzpublikationen unter Einbeziehung der Anforderungen der Langzeitar Archivierung neu gestaltet und der Workflow an die Schnittstellen des Archivsystems angepasst. Für so genannte fortlaufende Publikationen (vor allem elektronische Zeitschriftenartikel) entsprechen die künftigen Archivobjekte oft nicht der aktuellen Abbildung im Onlinekatalog. Bibliografische Metadaten von Archivobjekten müssen künftig in ILTIS Pica abgebildet werden. Dazu müssen eine Festlegung von Erschließungsvarianten und ein Mapping von Archivobjekten auf Katalogobjekte erfolgen. Das URN-Management in der DNB wurde bereits erweitert. Da jedes Objekt zum Einspielen in das Archiv einen Persistent Identifier (PI) benötigt, erfolgt für bereits gesammelte Objekte ohne URN eine Retrovergabe der URN. Alle neuen Objekte müssen mit URN geliefert werden bzw. bei Eingang/Bearbeitung eine URN erhalten. Zusammenfassend kann zum Workflowbereich gesagt werden, dass wesentliche Voraussetzungen für die Einbindung des Archivs in die Geschäftsumgebung der Institution vorliegen bzw. gerade geschaffen werden: Das Produktionssystem wurde aufgesetzt und läuft, das produktive Einspielen von Material wurde erprobt und wird fortgesetzt, nötige Weiterentwicklungen (z. B. noch fehlende Module zur Auswertung von Dateiformaten) wurden und werden ermittelt und Änderungen in diversen Workflows wurden angestoßen. Derzeitige Aufgaben bestehen in der

**Zusammenfassung
Workflowbereich**

Übergabe des kopal-Systems vom Projektteam an eine ständige Arbeitseinheit sowie die Gestaltung des Managements von rechtzeitigem Einspielen aller Retroobjekte und des aktuellen und künftigen Zugangs (direkt nach kopal).

Da die sichere Archivierung von Internetressourcen nunmehr ein wichtiges Arbeitsfeld darstellt, hat die DNB bereits vor dem Inkrafttreten des neuen Gesetzes verschiedene Möglichkeiten für automatisiertes Harvesting und entsprechende Archivierungsmechanismen getestet. Um diese Daten archivieren zu können, ist eine rechtzeitige Erhöhung des Speichervolumens von kopal zur Aufnahme aller Retroobjekte und für den laufenden Zugang erforderlich. Angesichts großer Datenmengen und teilweise großer Einzelobjekte (z. B. digitalisierte DVDs) ist die Gewährleistung einer ausreichenden Performanz des Systems einschließlich einer entsprechenden Benutzerbetreuung (etwa für Wartezeiten, rechtliche Beschränkungen u. a.) im Blick zu behalten.

Rechtzeitig vor Projektende hat kopal außerdem Modelle für die künftige Nachnutzung der Projektergebnisse bzw. des kopal-Archivsystems vorgelegt. kopal bietet ein flexibel konfigurierbares und anpassbares System an. Die Komponenten von kopal können je nach Kundenanforderung vereinbart und ausgestaltet werden. Die Modalitäten hängen außerdem vom gewünschten Nutzungsmodell ab. kopal bietet hier prinzipiell drei unterschiedliche Nutzungsmodelle an²⁾:

- kopal-Teilnehmer: Eine Institution lässt ihre Daten kommissarisch durch einen kopal-Mandanten archivieren.
- kopal-Mandant: Eine Institution verwaltet

**Erhöhung des
Speichervolumens
von kopal****Modelle zur
Nachnutzung**

selbstständig einen eigenen Bereich (Schließfach) des kopal-Archivsystems, der Serverbetrieb selbst bleibt ausgelagert.

- **kopal-Eigenbetrieb:** Eine Institution betreibt unter Rückgriff auf Erfahrungen des kopal-Projekts ein eigenes vollständiges Archivsystem.

Abhängig vom individuellen Servicekonzept einer Institution sind für die jeweiligen Nutzungsmodelle eine Reihe von Kostenfaktoren wie Zahl und Komplexität der Workflows bei einer Kundeninstitution, Menge, Heterogenität und Komplexität der zu archivierenden Objekte und ihrer Metadaten oder die gewünschten Zugriffsmöglichkeiten und Schnittstellen zu berücksichtigen.

Die Vorstellung der verschiedenen im kopal-Projekt erarbeiteten Servicemodelle bildete

Servicemodelle

auch den Schwerpunkt des Workshops »kopal goes live«, der am 18. Juni 2007 in der DNB in Frankfurt am Main stattfand. Mehr als 80 Teilnehmer aus ganz Deutschland waren gekommen, um sich über Projektergebnisse, Nachnutzungsmöglichkeiten und technische Entwicklungen aus dem Projekt zu informieren. Zu Beginn des Workshops konnte Ute Schwens (DNB) berichten, dass

kopal-Workshop

die kopal-Weiterführung durch die bisherigen Projektpartner gesichert ist. Die Repräsentanten der vier Einrichtungen (DNB, SUB, IBM Deutschland und GWDG) hatten ein »Memorandum of Understanding«, einen Kooperationsvertrag, unterzeichnet, in dem sie sich verpflichteten, das gemeinsam entwickelte kopal-System weiterzuführen. Außerdem wurde vereinbart, durch Forschungsvorhaben und Einbringen von Ressourcen die kopal-Lösung weiterzuentwickeln sowie die im Projekt entstandene ko-

operative Struktur weiter zu pflegen und auszubauen.

Reinhard Altenhöner (DNB) stellte den Projektverlauf und die im Projekt erreichten Ergebnisse vor. Norbert Lossau und Frank Klaproth (beide SUB) stellten die Kooperations- und Nachnutzungsmodelle im Detail vor. Die möglichen Varianten der kopal-Nutzung sowie die relevanten Kostenfaktoren wurden eingehend erläutert. Zu diesem Vortrag wurden die meisten Fragen aus der Zuhörerschaft gestellt. Dies verdeutlicht das rege Interesse an diesem Thema.

Es folgte der technisch orientierte Teil des Workshops. Tobias Steinke (DNB) stellte den technischen Aufbau des kopal-Systems und seine Aufgaben vor. Dagmar Ullrich (GWDG) beschrieb die Aufgaben des Dienstleisters GWDG sowie die Hardwarestruktur, in welche das Archivsystem in Göttingen eingebunden ist. Jürgen P. Schwarz (IBM) erläuterte neben der Perspektive für kopal-DIAS innerhalb der IBM die Lizenzmodelle, welchen die verschiedenen Varianten der kopal-Nachnutzungsoptionen (jeweils für die IBM-Softwareprodukte) unterliegen. Zur Abrundung des Technikteils stellte Jens Ludwig (SUB) die von den Bibliotheken entwickelte Open-Source-Library koLibRI vor und erläuterte, inwieweit diese von nachnutzenden Institutionen an die individuellen Erfordernisse der Einrichtung angepasst werden kann. In der Abschlussdiskussion wurde deutlich, dass viele Teilnehmer in ihren Einrichtungen derzeit in die aktuelle Diskussion einbezogen sind, wie die digitale Langzeitarchivierung geleistet werden soll.

Welche Perspektiven ergeben sich aus dem kopal-Projekt für die DNB? Einige Eckpunkte

Projektverlauf und Nachnutzungsmodelle

Technischer Teil

**Perspektiven
für die DNB**

sind bereits benannt worden: Die abgeschlossene Kooperationsvereinbarung stellt den Weiterbetrieb von kopal sicher und gewährleistet auch die Pflege des bestehenden Systems. Dies bietet die Grundlage, nach und nach alle weiteren digitalen Materialien aus den Bibliotheksbeständen einzuspielen. Dazu müssen in den kommenden Monaten die entsprechenden Workflows angepasst bzw. weiterentwickelt werden. Gleichzeitig ergibt sich eine Außenperspektive für die DNB. Zur Wahrnehmung des gesetzlichen Auftrags wird die DNB mit anderen Bibliotheken auf dem Gebiet der digitalen Langzeitar-

chivierung zusammenarbeiten. Im Rahmen dieser Kooperation sind unter Berücksichtigung der erarbeiteten Service- und Nachnutzungsmodelle von kopal derzeit Planungen und Vorarbeiten im Gange, einer Reihe von weiteren Einrichtungen die Archivierung ihrer Materialien im bestehenden kopal-System zu ermöglichen. Hierbei werden die DNB und die SUB Göttingen Langzeitarchivierungsaufgaben übernehmen und entsprechende Services anbieten. Der Dauerbetrieb von kopal und die Anbindung weiterer Einrichtungen stellen an die DNB eine nicht geringe, aber spannende Herausforderung.

Anmerkungen**1**

Altenhöner, Reinhard: Das KOPAL-Projekt des Bundesministeriums für Bildung und Forschung. In: Dialog mit Bibliotheken, 18 (2005) 1, S. 21 – 34.

2

Ausführliches Papier zu den Servicemodellen unter
<http://kopal.langzeitarchivierung.de/downloads/kopal_Services_2007.pdf>