

**Gesamtprojektleitung**

**Deutsche Nationalbibliothek**  
Reinhard Altenhöner  
Adickesallee 1  
60322 Frankfurt am Main  
Tel. 069/15251700  
E-Mail: altenhoener@dbf.d-nb.de

**Projektleitungen der Partner**

**Deutsche Nationalbibliothek**  
Tobias Steinke  
Adickesallee 1  
60322 Frankfurt am Main  
Tel. 069/15251762  
E-Mail: steinke@dbf.d-nb.de

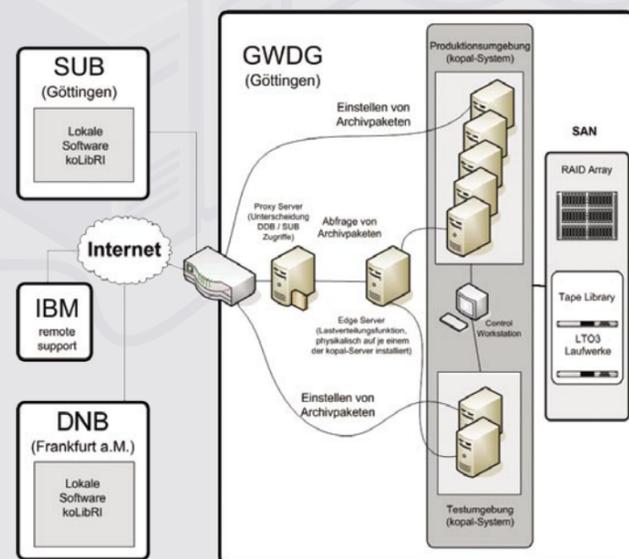
**Niedersächsische Staats- und  
Universitätsbibliothek**  
Frank Klaproth  
Platz der Göttinger Sieben 1  
37073 Göttingen  
Tel. 0551/39-5228  
E-Mail: klaproth@sub.uni-goettingen.de

**IBM Deutschland GmbH**  
Dr. Wilhelm Stoll  
Altrottstr. 31  
69190 Walldorf  
Tel. 06227/73-1075  
E-Mail: wstoll@de.ibm.com

**Gesellschaft für wissenschaftliche  
Datenverarbeitung mbH Göttingen**  
Dagmar Ullrich  
Am Fassberg  
37077 Göttingen  
Tel. 0551/201-1827  
E-Mail: dullric@gwdg.de

E-Mail: info@kopal.langzeitarchivierung.de

kopal.langzeitarchivierung.de

**Technische Infrastruktur**


Die Grafik zeigt die technische Infrastruktur des kopal-Projekts mit Fokus auf die Installation des Systems bei der Gesellschaft für wissenschaftliche Datenverarbeitung gmbH Göttingen (GWDG). Es handelt sich aktuell um sieben IBM p550 Server. Fünf dieser Server sind für den Einsatz im Produktivbetrieb vorgesehen, zwei dienen als Testumgebung. Das kopal-System ist mit einer Bandbreite von einem Gigabit an das Deutsche Wissenschaftsnetz (X-Win) und damit an das Internet angeschlossen. Ein Großteil der Zugriffe auf das System wird über einen vorgeschalteten Proxy-Server abgewickelt. Hier ordnet ein Web-server eingehende Anfragen den einzelnen Mandanten, derzeit DNB oder SUB, zu. Dann leitet er sie so an das System weiter, dass Zugriffe nur auf die jeweils institutseigenen

Daten möglich sind. Die Anfragen werden vom so genannten Edge-Server entgegengenommen, der sie so an die anderen Server verteilt, dass eine optimale Auslastung der Systemkapazitäten sichergestellt ist. Sowohl für die Speicherung von kopal-Archivpaketen als auch für das Backup des Systems werden zwei Bandroboter mit LTO3-Laufwerken eingesetzt. Diese beiden Scalar 10K Tape Libraries der Firma ADIC befinden sich an zwei räumlich getrennten Standorten in Göttingen, eine direkt bei der GWDG, die zweite im Bereich Humanmedizin der Georg-August-Universität Göttingen, Geschäftsbereich Informationstechnologie. Neben diesen Bandmedien nutzt kopal derzeit vorwiegend für Caching-Funktionen ein IBM DS4500 RAID-Array.

**Projektverlauf**
**Bisherige Meilensteine**

Zu Beginn des Projektes führte die Deutsche Nationalbibliothek (DNB) und die SUB Göttingen unter Mitwirkung der GWDG und der IBM verschiedene Tests zur Beurteilung des Digital Information Archiving System (DIAS) durch. Hierzu zählten unter anderem Belastungstests, Umgang mit extrem großen Datenpaketen und die Fehlerbehandlung. Vor dem Beginn der eigentlichen Projektphase stand die Definition des oben bereits vorgestellten Universellen Objektformats (UOF).

Die erste Entwicklungsphase umfasste die Beschaffung und Bereitstellung der erforderlichen Systemkomponenten durch die GWDG und die Anpassung der Kernkomponente **DIAS-Core** an die Anforderungen der Partner durch die IBM. Hierzu zählte insbesondere die Unterstützung des „Universellen Objektformats“. Außerdem wurden die systemseitigen Voraussetzungen für eine separate Nutzung seitens DNB und SUB geschaffen. Diese Mandantenfähigkeit stellt einen unabhängigen und nicht ortsgebundenen Zugriff beider Partner auf das System sicher.

**Weitere Entwicklung**

Im Zentrum der zweiten Entwicklungsphase steht die Entwicklung und Umsetzung eines tragfähigen Konzeptes zum Preservation Planning. Eine Grundlage für die Nutzbarkeit der digitalen Objekte über den technologischen Wandel hinweg ist die Erhaltung des Datenstroms. Darunter versteht man das sichere und verlustfreie langfristige

Sie ist die Grundvoraussetzung für die künftige kooperative Nutzung des Archivsystems.

Die Partner Deutsche Nationalbibliothek und SUB Göttingen implementieren auf der Basis des DIAS-Core zusätzliche Software zum automatisierten Einspeisen (Ingest) und Abfragen (Access) von digitalen Objekten, die so genannt „kopal Library for Retrieval and Ingest“ (koLibRI, siehe oben).

Ende November 2005 hat kopal eine erste Testphase des von IBM weiterentwickelten DIAS erfolgreich abgeschlossen. Im Fernzugriff auf das System wurde das Zusammenspiel von DIAS mit den kopal-Tools untersucht. Dokumentiert wurde außerdem das Verhalten des Systems bezüglich der Größe der Objekte sowie die Vollständigkeit und Korrektheit der Lieferungen. Schließlich wurden mögliche Störungen und Systemausfälle simuliert, die erfahrungsgemäß im praktischen Betrieb vorkommen.

Erhalten von digitalen Objekten. Die dafür nötigen Speicher- sowie Migrations- und Emulationsstrategien werden von den Projektpartnern entwickelt.

Die Deutsche Nationalbibliothek und die SUB Göttingen werden eine Auswahl ihrer vorhandenen Datenbestände in das System einstellen.

Anhand einer Vielzahl von Objekten in verschiedenen Formaten wird damit die praxisnahe Nutzung der **kopal-Solution** aufgezeigt. Auf diese Weise werden nachnutzbare

**Projektabschluss 2007**

Mit dem Projektabschluss im Juni 2007 wird ein nachnutzbares Archivsystem und ein erstes Release der vollständig entwickelten

**Service**

Eine sichere Speicherung und langfristige Verfügbarkeit digitaler Daten zu gewährleisten, umfasst die Integration der Archivierungslösung in bestehende Informationssysteme und Arbeitsabläufe. Hierbei ist die Nachnutzung durch weitere Kulturerbeinstitutionen sowie durch sonstige Institutionen und Firmen, die eine Langzeitarchivierung benötigen, ausdrücklich erwünscht.

Die **kopal-Solution** ist von vornherein auf unterschiedliche Bedürfnisse ausgerichtet. kopal verfolgt dabei das Ziel, künftig verschiedene

Lieferungsverfahren von digitalem Archivgut bereitgestellt. Ergänzend werden Absprachen mit nationalen und internationalen Zulieferern getroffen.

koLibRI-Software bereitgestellt. Hierauf bauen unterschiedliche kopal-Dienste und Services auf.

Nutzungsmodalitäten bereitzustellen: Es wird einerseits die Möglichkeit geboten werden, als Mandant mit einem eigenen „Schließfach“ das bestehende System mit abgesichertem Speicherplatz und eigener Verwaltung der Daten zu nutzen. Diese Lösung ist besonders geeignet für Institutionen mit geringerem Archivgutaufkommen. Andererseits wird die Möglichkeit einer Nachnutzung der kopal-Solution durch eine eigene Installation des **DIAS-Systems** bestehen. Dies wird in der Regel zusammen mit einer Nutzung der koLibRI-Software erfolgen.

**DIAS User Group**

Zum gegenseitigen Erfahrungsaustausch haben die kopal-Partner gemeinsam mit der Königlichen Bibliothek der Niederlande eine DIAS User Group ins Leben gerufen. Die beteiligten Institutionen halten mehrmals im Jahr Arbeitstreffen ab, um Informationen und Erfahrungen auszutauschen, notwendige Anpassungen zu diskutieren und gemeinsame Strategien für künftige Nutzanwendungen zu entwickeln.

Neue Partner und Interessenten, die sich für eine Nutzung von kopal entschieden haben, können unmittelbar in die DIAS User Group einbezogen werden. Sie profitieren von den vielseitigen Erfahrungen der anderen DIAS-Nutzer, gewinnen wertvolle Erkenntnisse für die eigenen Archivierungsaktivitäten und wirken aktiv an der Weiterentwicklung von DIAS bzw. kopal mit.

**Beirat kopal**

Dem kopal-Beirat gehören Experten aus Wirtschaft, Wissenschaft und öffentlicher Verwaltung an, die den Gesamtfortschritt und Ablauf des Projektes beratend begleiten. Die Zusammensetzung des Beirats gewährleistet, dass die Ziele von kopal gut unterstützt

und die strategische Zielsetzung erreicht werden kann. Der Beirat trifft sich in der Regel halbjährlich, um die erreichten Meilensteine zu evaluieren und Hinweise zu geben, wie die erreichten Projektfortschritte praxisbezogen kommuniziert werden können.

**Links**

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

**DC (Dublin Core)**  
<http://www.dublincore.org/>

**LMER (Langzeitarchivierungsmetadaten für elektronische Ressourcen)**  
<http://www.ddb.de/standards/lmer/>

**METS (Metadata Encoding & Transmission Standard)**  
<http://www.loc.gov/standards/mets/>

**ISO Archiving Standards**  
<http://ssdoo.gsfc.nasa.gov/nost/isoas/>

**Reference Model for an Open Archival Information System (OAIS)**  
<http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.o-B-1.pdf>

## Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen

Digitale Dokumente für die Nachwelt erhalten, dafür steht das Projekt **kopal**. Es verfolgt den Aufbau eines kooperativ entwickelten und betriebenen Archivsystems zur sicheren Langzeitverfügbarkeit digitaler Dokumente. Die Daten müssen dabei nicht nur physisch erhalten, sondern vor allem auch in Zukunft noch fehlerfrei interpretiert werden können.

### Workflow

Im Folgenden wird ein Überblick über den Arbeitsablauf bei der Nutzung des Archivsystems gegeben. Dieser vollzieht sich in folgenden Schritten:

**Auswahl:** Die jeweilige Institution wählt digitale Objekte aus, die langzeitarchiviert werden sollen. Sie wird dabei inhaltliche und formale Kriterien berücksichtigen wie Sammelschwerpunkte, besondere Digitalisate etc.

**Sammeln und Erstellen der Metadaten:** Um ein systematisches Ablegen und Wiederfinden von Objekten zu ermöglichen, werden diesen Zusatzinformationen wie bibliographische Daten hinzugefügt. Technische Metadaten sind notwendig, um die Objekte regelmäßig auffrischen und migrieren zu können. Die Metadaten werden zum Teil aus Informationssystemen geholt, die die Objekte bereits erschließen, zum Teil mittels spezieller Software aus den Objekten selbst generiert.

**Erstellung eines Einlieferungspaketes:** Die digitalen Objekte werden zusammen mit ihren Metadaten in einem speziellen Format, dem Universellen Objektformat (UOF, siehe unten) als Paket gebündelt.

**Einbringen in das Langzeitarchiv (Ingest):** Vor dem Einbringen in das Archiv überprüft die Software die Daten noch auf Vollständigkeit und

formale Korrektheit.

**Umwandlung in ein Archivierungspaket:** Die Metadaten werden in das Data Management überführt. Die Contentdaten und die zugehörige Metadatendatei wandern auf einen von DIAS verwalteten Massenspeicher.

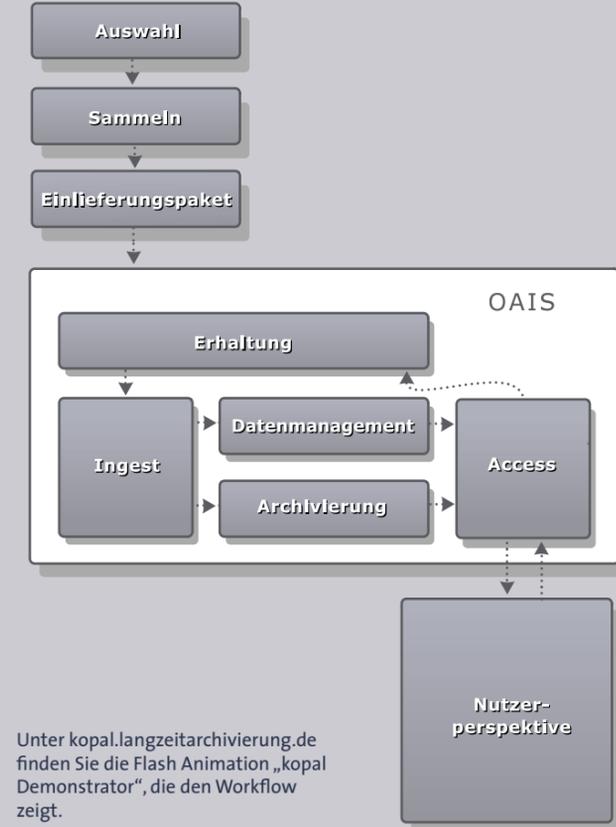
**Abfrage und Anforderung von Informationen (Access):** Über das Datenmanagement kann gezielt auf die Metadaten und damit auf die Archivpakete zugegriffen werden.

**Auslieferung:** Je nach Bedarf werden Metadaten und/oder das Archivmaterial selbst in einem entsprechenden Paketformat ausgeliefert.

**Verwendung der Daten:** Nutzer von digitalen Objekten greifen i.d.R. über ein Informationssystem auf die Daten zu. Dabei kann dem Nutzer angezeigt werden, dass er gerade auf langzeitarchivierte Daten zugreift. Zugleich kann ihm die Option eingeräumt werden, die Daten in einem bestimmten – nicht notwendig aktuellen – Format auszuwählen.

## Das Archivsystem

Das Archivsystem von **kopal** bietet eine technische und organisatorische Infrastruktur, mit der Gedächtnisorganisationen wie Archive, Bibliotheken und Museen ihre digitalen Sammlungen langfristig verfügbar halten können. Auch Verlage, Behörden und Firmen können künftig Mandanten dieses kooperativ betriebenen Archivsystems werden.



Unter [kopal.langzeitarchivierung.de](http://kopal.langzeitarchivierung.de) finden Sie die Flash Animation „kopal Demonstrator“, die den Workflow zeigt.

## Objektspezifikation

Damit die archivierten Daten rekonstruiert und gepflegt werden können, bedarf es strukturierter Archivpakete. Hierfür kommt bei **kopal** und seinem Archivsystem DIAS das „Universelle Objektformat“ (UOF) zum Einsatz. Es beschreibt eine Paketstruktur inklusive Metadaten und ist gleichermaßen als Archiv- und Austauschformat für Langzeitarchive geeignet. Das UOF kann beliebige Dateistrukturen aufnehmen und inhaltliche sowie technische Metadaten transportieren. Dabei gibt es keinerlei Beschränkungen auf bestimmte Dateiformate oder Medienkategorien. Somit lassen sich Objekte in beliebigen Dateiformaten wie TIFF, PDF, XML etc., aber auch ISO-Images von CD/DVD im Archiv speichern.

Für die Metadaten kommen die Standards METS (Metadata Encoding and Transmission Standard), LMER (Langzeitarchivierungsmetadaten für elektronische Ressourcen) und DC (Dublin Core) zur Anwendung. Darüber hinaus lassen sich weitere XML-Metadaten nutzen. Spezifische technische Metadaten extrahiert das kopal-System mittels des Software-Tool JHOVE, das ursprünglich für das Zeitschriftenarchiv JSTOR entwickelt wurde. Ein zentraler Teil der Metadaten im UOF dokumentiert die komplette Migrationshistorie. Dafür werden alle Konvertierungen aufgelistet, die zur Sicherstellung der Langzeitverfügbarkeit nötig waren. Diese Daten werden mit den technischen Informationen zu den einzelnen Dateien eines Objekts erhalten, damit künftige Nutzer auch in vielen Jahren noch auf den Inhalt der Objekte zugreifen können.

## Software

Das **kopal**-Archivsystem besteht aus dem DIAS-Core der IBM und den entwickelten **kopal**-Tools, welche die Partner SUB Göttingen und Deutsche Nationalbibliothek entwickelt haben. Der DIAS-Core beruht auf Standardsoftwarekomponenten der IBM, die hierfür weiterentwickelt wurden.

### DIAS-Core

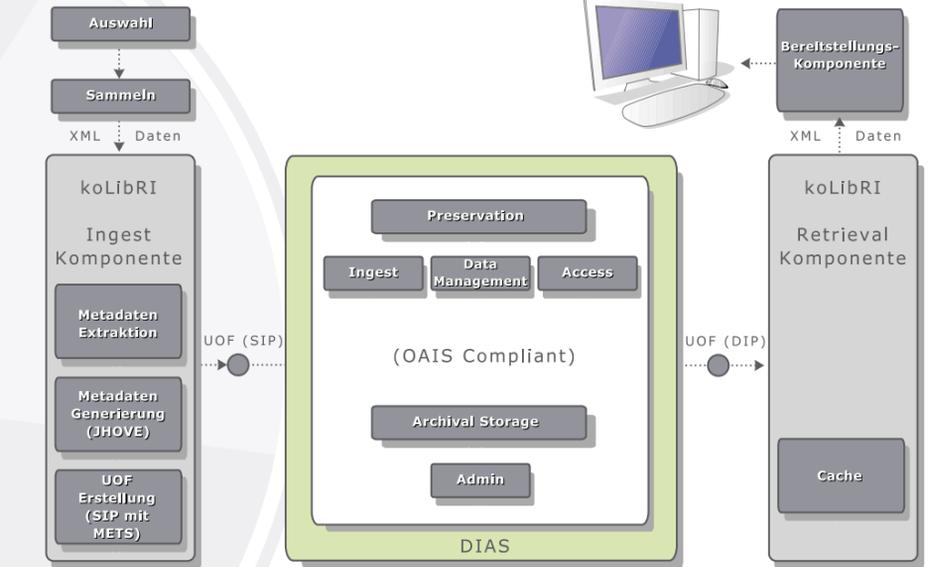


Den Kern des **kopal**-Archivsystems bildet das von IBM in Zusammenarbeit mit der Nationalbibliothek der Niederlande entwickelte **Digital Information Archiving System** (DIAS). Dem OAIS-Standard für Archivsysteme entsprechend, verwaltet DIAS-Core die gespeicherten

Daten und stellt eine Umgebung für unterschiedlichste Erhaltungsstrategien zur Verfügung. DIAS' Mandantenfähigkeit und präzise definierte Import- und Exportschnittstellen ermöglichen es, eine Archivnutzung in verschiedensten Workflows unterschiedlichster

Institutionen zu integrieren. Durch den Einsatz von vielfach bewährter IBM-Standardsoftware wie der DB2-Datenbank, dem Content Manager und dem Tivoli Storage Manager sind langfristige Stabilität, Performanz und Skalierbarkeit gewährleistet.

## koLibRI



Die Deutsche Nationalbibliothek und die Staats- und Universitätsbibliothek Göttingen entwickeln ein Software-Paket für die Benutzung der **kopal**-Solution: die „**kopal** Library for Retrieval and Ingest“ (**koLibRI**). Die **kopal**-Tools unterstützen das Einspielen von Objekten in das DIAS sowie den Zugriff auf die archivierten Objekte. Die Schnittstelle für den **Datenimport** (Ingest) erfüllt folgende Anforderungen:

1. Der Ingest kann automatisiert erfolgen.
2. Das universelle Objektformat (**kopal-UOF**) wird unterstützt.

3. Die flexible Schnittstelle ermöglicht eine Integration in verschiedenste Umgebungen und Informationssysteme.

4. Es werden internationale Standards angewandt.

5. Eine Nachnutzbarkeit durch Dritte ist gewährleistet.

6. Eine grafische Oberfläche ist in Entwicklung.

Die Anforderungen an den **Datenexport** (Access) unterscheiden sich bei der Deutschen Nationalbibliothek und der SUB Göttingen deutlich, so dass generische Module genutzt werden, welche die einzelnen Institutionen je nach Bedarf erweitern können und damit die

Nachnutzung für andere ermöglichen. Diese Software wird künftig auch die **Administration** des **kopal**-Systems unterstützen. Aufgrund der sehr unterschiedlichen und teils heterogenen Systemstrukturen der beiden Institutionen wird flexible Software benötigt. Das so genannte Workflow-Tool bietet eine gemeinsam nutzbare Infrastruktur für Module. Dieses Workflow-Tool kann als Asset-Builder zum Erstellen von Archivpaketen dienen. Weiterhin kann es als eine zentrale Vermittlungsinstanz zum DIAS-System, als Client-Loader, genutzt werden, indem es Archivpakete von mehreren Asset-Buildern sammelt und an DIAS übermittelt. Weitere Nutzungsmöglichkeiten lassen sich problemlos integrieren.