

Overall Project Manager

German National Library  
(Deutsche Nationalbibliothek)  
Reinhard Altenhoener  
Adickesallee 1  
60322 Frankfurt am Main, Germany  
49-69-152-51700  
r.altenhoener@d-nb.de

Partner Project Management

German National Library  
(Deutsche Nationalbibliothek)  
Tobias Steinke  
Adickesallee 1  
60322 Frankfurt am Main, Germany  
49-69-1525-1762  
t.steinke@d-nb.de

Goettingen State and University Library  
(Niedersächsische Staats- und  
Universitätsbibliothek)  
Frank Klaproth  
Platz der Goettinger Sieben 1  
37073 Goettingen, Germany

49-551-39-5228  
klaproth@sub.uni-goettingen.de

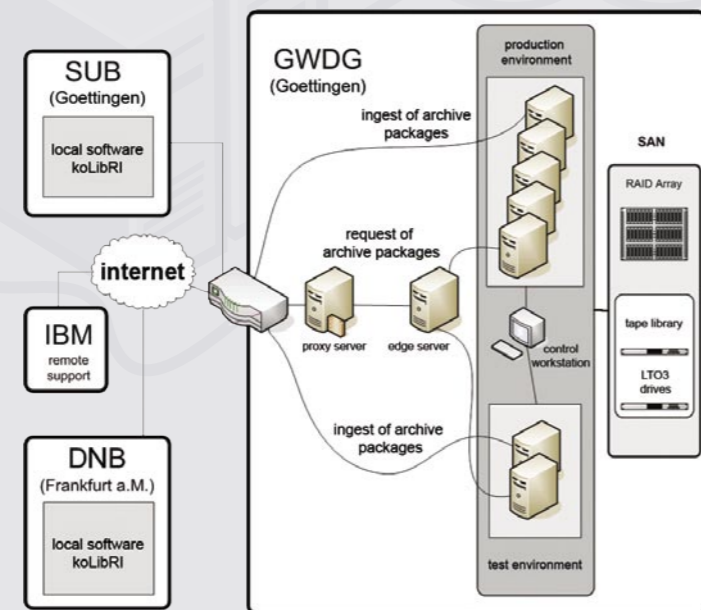
IBM Deutschland GmbH  
Dr. Wilhelm Stoll  
Altrottstr. 31  
69190 Walldorf, Germany  
49-6227-73-1075  
wstoll@de.ibm.com

Gesellschaft fuer wissenschaftliche  
Datenverarbeitung mbH Goettingen  
Dagmar Ullrich  
Am Fassberg 11  
37077 Goettingen, Germany  
49-551-201-1827  
dagmar.ullrich@gwdg.de

e-mail: info@kopal.langzeitarchivierung.de

kopal.langzeitarchivierung.de

Technical Infrastructure



The graphic shows the technical infrastructure of the kopal project with a focus on the installation of the system at the Gesellschaft fuer wissenschaftliche Datenverarbeitung mbH Goettingen (GWDG). It currently involves seven IBM p550 servers. Five of these form the production system; two serve as a test environment. The kopal system is connected to the Deutsche Wissenschaftsnetz (X-Win) and thus to the Internet with a bandwidth of 1 gigabit per second. Most accesses to the system are handled through an upstream proxy server. Here, a web server reassigned incoming requests to individual clients, currently the Deutsche Nationalbibliothek (German National Library) and the Niedersächsische Staats- und Universitätsbibliothek Goettingen (Goettingen State and University Library). Then, the requests are

forwarded on to the system, so that access is possible only to the institution's own data. The requests are received from a so-called Edge Server, which distributes them to the other servers in such a way that an optimal utilization of the system capacities is ensured. Two tape robots with LTO3 drives are deployed for the storage of kopal archive packages as well as for system backup. These two Scalar 10K Tape Libraries from the firm ADIC are located in two different places in Goettingen. Besides these tape media, kopal presently uses an IBM DS4500 RAID array primarily for caching functions.

Project Chronology

Milestones

At the beginning of the project, the German National Library and the Goettingen State and University Library, with the collaboration of the GWDG and IBM, conducted various tests in order to evaluate the Digital Information Archiving System (DIAS). Among the tests conducted were load tests, the handling of extremely large data packets, and error handling. The definition of the Universal Object Format (UOF), discussed above, was developed at the beginning of the project.

The first development phase involved the acquisition and installation of the required system components at the GWDG and the customization by IBM of the core components, DIAS-Core, according to the requirements of the partners. In particular, this included the support of the UOF. In addition, the system-side requirements of the German National Library and the Goettingen State and University Library for individual operation were realized. This multi-client capability ensures independent access to the system from any location by both partners and is the basic requirement for the future cooperative

Future Development

The core of the second development phase is the development and implementation of a sustainable plan for preservation planning. A basic requirement for the usability of digital objects in the face of technological change is the preservation of the data stream, i.e., the secure and loss-free long-term preservation of digital objects. The

operation of the archival system.

The partners, the German National Library and the Goettingen State and University Library, are implementing supplementary software on top of the DIAS-Core, the so-called "kopal Library for Retrieval and Ingest" (koLibRI, see above), in order to automate import (ingest) and retrieval (access) of digital objects.

At the end of November 2005, kopal successfully completed a first test phase of a version of DIAS that had been adapted by IBM to the project's needs. The interaction of DIAS with the kopal tools was examined via remote access. The behavior of the system with respect to the size of the objects as well as the completeness and correctness of the system responses were also documented. Finally, possible malfunctions and system failures, which experience has shown might occur in everyday operation, have been simulated. The system behavior has been tested accordingly.

required storage strategies as well as migration and emulation strategies are being developed by the project partners.

Starting in summer 2006, the German National Library and the Goettingen State and University Library will enter a selection of their existing data holdings into the system.

The practical operation of the kopal solution will be demonstrated in an environment with a number of objects in various formats. Thus, reusable procedures for delivery of digital archival objects

Project Completion 2007

With the completion of the project in June 2007, a reusable archival system and a first release of the fully developed koLibRI-Software

Service

Ensuring secure storage and long-term access to digital data involves the integration of the archiving solution into existing information systems and procedures. Here, the reuse by other cultural heritage institutions as well as by other institutions and firms that need long-term archiving is expressly desired.

From its inception, the kopal solution has been designed to serve a number of different needs. kopal pursues the goal of making various modes of operation available in the future. On the one hand, a

will be provided. In addition, agreements will be made with national and international suppliers.

will be made available. This will be the basis of several kopal services.

client will be given the opportunity to have its own „locker“ in the existing system, i.e., secure storage space with data under its own administrative control. This solution is especially appropriate for institutions having a small amount of material to be archived. On the other hand, the opportunity will also be available for an institution to reuse the kopal solution through its own installation of the DIAS system, normally in conjunction with the koLibRI-Software.

DIAS User Group

To facilitate the exchange of experiences, the kopal partners, along with the Koninklijke Bibliotheek (National Library of the Netherlands), have created a DIAS User Group. The participating institutions hold several workshops during the year in order to exchange information and experiences, discuss necessary customizations, and develop common strategies for future applications. New partners and others who have decided to

use kopal can immediately get involved in the DIAS User Group. They profit from the wide experience of other DIAS users, gain valuable insight for their own archiving activities, and take part in the further development of DIAS and kopal respectively.

kopal Advisory Board

The kopal Advisory Board includes experts from the business, academic, and public service sectors who oversee and advise the overall progress and development of the project. The composition of the Board ensures that the goals of kopal can be well supported and

that strategic objectives can be reached. The Board normally meets twice a year in order to evaluate the accomplishments and to give advice about how the project's progress can be communicated in a practical way.

Links

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

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

LMER (Long-term Preservation Metadata for Electronic Resources)  
<http://www.d-nb.de/eng/standards/lmer/lmer.htm>

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>

## Co-operative Development of a Long-Term Digital Information Archive

To preserve digital documents for future generations – that is what the **kopal** project stands for. It pursues the building of a cooperatively developed and operated archival system to ensure the long-term access to digital documents. In the process, the data must not only be physically preserved, but moreover, they must be interpretable without error in the future.

### Workflow

In the following, an overview of the procedures in the operation of the archival system will be given. The workflow consists of the following steps:

**Selection:** The institution selects digital objects to be long-term archived. In doing so, content and formal criteria, such as special collections, certain digitized documents, etc., are taken into account.

**Collection and production of metadata:** To enable the systematic storage and retrieval of objects, supplementary information such as bibliographic data is added. Technical metadata are required to be able to regularly refresh and migrate objects. The metadata are in part retrieved from information systems that may index the object in advance, and are in part generated from the object itself using special software.

**Production of a Submission Package:** The digital objects are bundled along with their metadata as a package in a special format, the Universal Object Format (UOF, see below).

**Import into the long-term archive (ingest):** The software verifies the data for completeness and formal correctness before ingesting it into the archive.

**Transformation into an archiving package:** The metadata are transferred to the data management system. The digital document and the corresponding metadata file are moved into mass storage administered by DIAS.

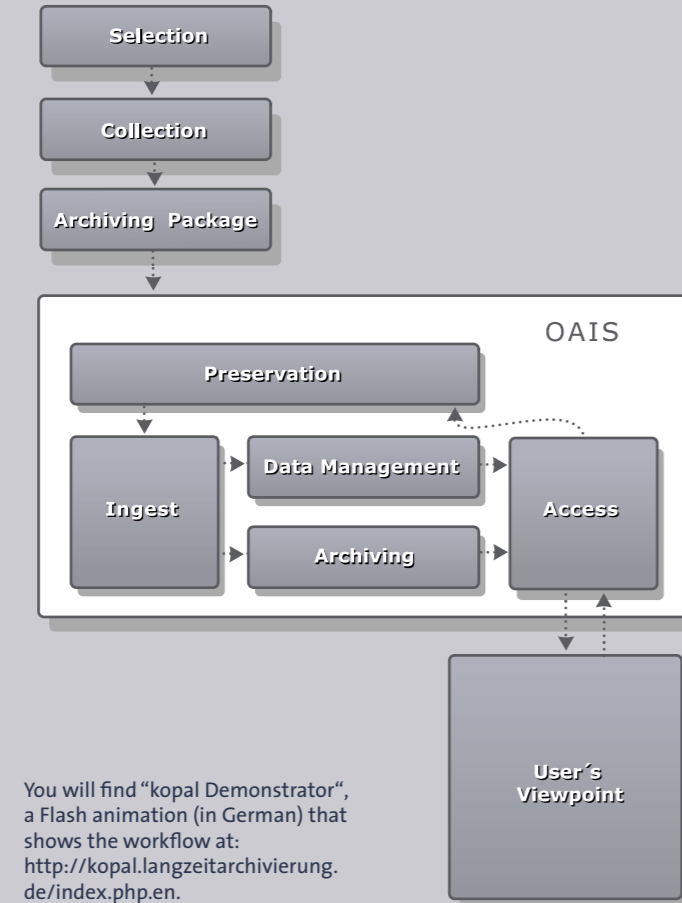
**Request and retrieval of information (access):** The metadata and thus the archive packages can be accessed via the data management system.

**Delivery:** As needed, the metadata and/or the archive material itself are delivered in a corresponding package format.

**Utilization of the data:** Users of digital objects usually access the data via an information system. Through such a system, the user can be notified that s/he has just accessed long-term archived data. At the same time, s/he may be given the option to choose data in a particular, not necessarily current, format.

## The Archiving System

**kopal's** archival system offers a technical and organizational infrastructure with which memory organizations such as archives, libraries, and museums can make their digital collections available over the long-term. Publishing houses, agencies, and firms can become clients of this cooperatively-operated archival system in the future.



## Object Specification

In order to maintain and reconstruct the archived data, a structured archive package is needed. For this purpose in **kopal** and its DIAS archival system, the Universal Object Format (UOF) is employed. The UOF describes a package structure including metadata and is suitable both as an archive format and as an exchange format for long-term archives. The UOF can accept as many data structures as needed and can handle descriptive as well as technical metadata. There are absolutely no restrictions on file format or media type. Thus, any file type such as TIFF, PDF, XML, etc., as well as ISO images from CD/DVD, can be stored in the archive.

The metadata comply with the standards METS (Metadata Encoding and Transmission Standard), LMER (Long-term preservation Metadata for Electronic Resources), and Dublin Core. Furthermore, other XML metadata can also be used. The kopal system extracts specific technical metadata using the software tool JHOVE, which was originally developed for the journal archive JSTOR.

A vital part of the metadata in the UOF documents is the complete history of the migration. All of the conversions that were necessary for long-term access are listed. These data are preserved with the technical information pertaining to the individual files of an object, so that users will still be able to access the content of an object even after many years.

## Software

The kopal archival system consists of IBM's DIAS-Core and of kopal tools, which the partners, the Goettingen State and University Library (Niedersächsische Staats- und Universitätsbibliothek Goettingen) and the German National Library (Deutsche Nationalbibliothek), have developed. The DIAS-Core is based on standard IBM software components that

have been further extended for kopal. The German National Library and the Goettingen State and University Library are developing powerful kopal tools which are precisely matched with DIAS-Core and which will be available in the future as the "kopal Library for Retrieval and Ingest" (koLibRI) under an Open Source license.

### DIAS-Core

Delivery&Capture



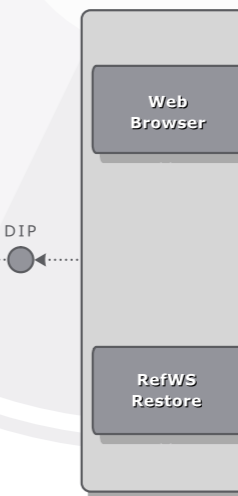
**The Digital Information Archiving System (DIAS)**, developed by IBM in cooperation with the Koninklijke Bibliotheek (National Library of the Netherlands), represents the core of the kopal archival system. Conforming to the OAIS standard for archival systems, DIAS-Core

Ingest/Access  
Data Management/Archival Storage  
Administration  
Preservation(planned)



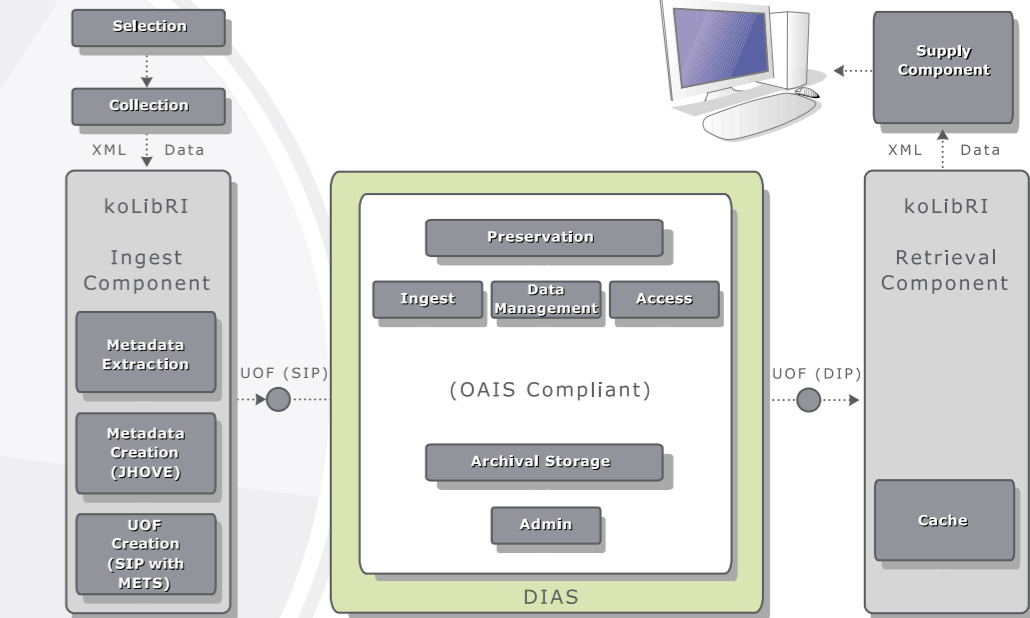
administers the stored data and makes available an environment for various preservation strategies. DIAS' multi-client capability and precisely defined import and export interfaces make it possible to integrate the operation of the archive into various workflows and

Packaging&Delivery



institutions. Through the adoption of well-established standard IBM software such as DB2 Database, Content Manager, and Tivoli Storage Manager, long-term stability, performance, and scalability are assured. The German National Library and the Goettingen State and Univer-

## koLibRI



sity Library are developing a software package for the use of the kopal solution: the "kopal Library for Retrieval and Ingest" (koLibRI). The **kopal tools** support the import of objects into DIAS as well as access to the archived objects. The interface for **data import (ingest)** fulfills the following requirements:

1. The ingest can be automated.
2. The Universal Object Format (kopal-UOF) is supported.
3. The flexible interface enables integration into various environments and information systems.

4. International standards are utilized.
5. Reuse by third parties is ensured.
6. A graphical interface is in development.

Since the requirements for **data export (access)** are clearly different for the German National Library and the Goettingen State and University Library, generic modules are used which an institution can extend as needed, thus making possible the reuse for others. In the future, this software will also support the **administration** of the kopal system. Because of the quite different and in part heterogeneous system structures of the two institutions, flexible software is required.

The so-called Workflow Tool offers a jointly usable infrastructure for modules. This Workflow Tool can serve as an asset builder for the production of archive packages. Furthermore, it can be used as a central relay to the DIAS system, as a client loader, in which it collects archive packages from many asset builders and delivers them to DIAS. Further possible uses can be integrated without problem.