GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

kopal
Data into the future

Technical concepts of kopal

Tobias Steinke, Deutsche Nationalbibliothek
June 11, 2007, Berlin

Overview

kopal
Data into the future

- Project kopal
  - Ideas
  - Organisation
  - Results
- Technical concepts
  - DIAS
  - koLibRI
- Models of reusability

Project kopal: Ideas

kopal
Data into the future

- Build-up a deposit for long-term preservation of electronic information
- Usable by different partners and institutions
- Support for long-term accessibility by migration and emulation
- No restrictions on categories (text, images, audio, video) and file formats (PDF, HTML, TIFF, etc.)
- Usage of existing standards like OAIS and METS
- Reliable core software enhanced by flexible open source modules

Project kopal: Organisation

kopal
Data into the future

- German National Library (Deutsche Nationalbibliothek)
  - Objects: Electronic theses, eJournals, CD-ROM's, scanned books, eBooks, music files, web pages
  - Role: Project leader, user, software development
- Lower Saxon State and University Library (SUB Göttingen)
  - Objects: Electronic theses, eJournals, scanned books, videos
  - Role: User, software development
- IBM Germany
  - Role: Core software development
- Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG)
  - Role: Service provider, hosting

Project kopal: Results

kopal
Data into the future

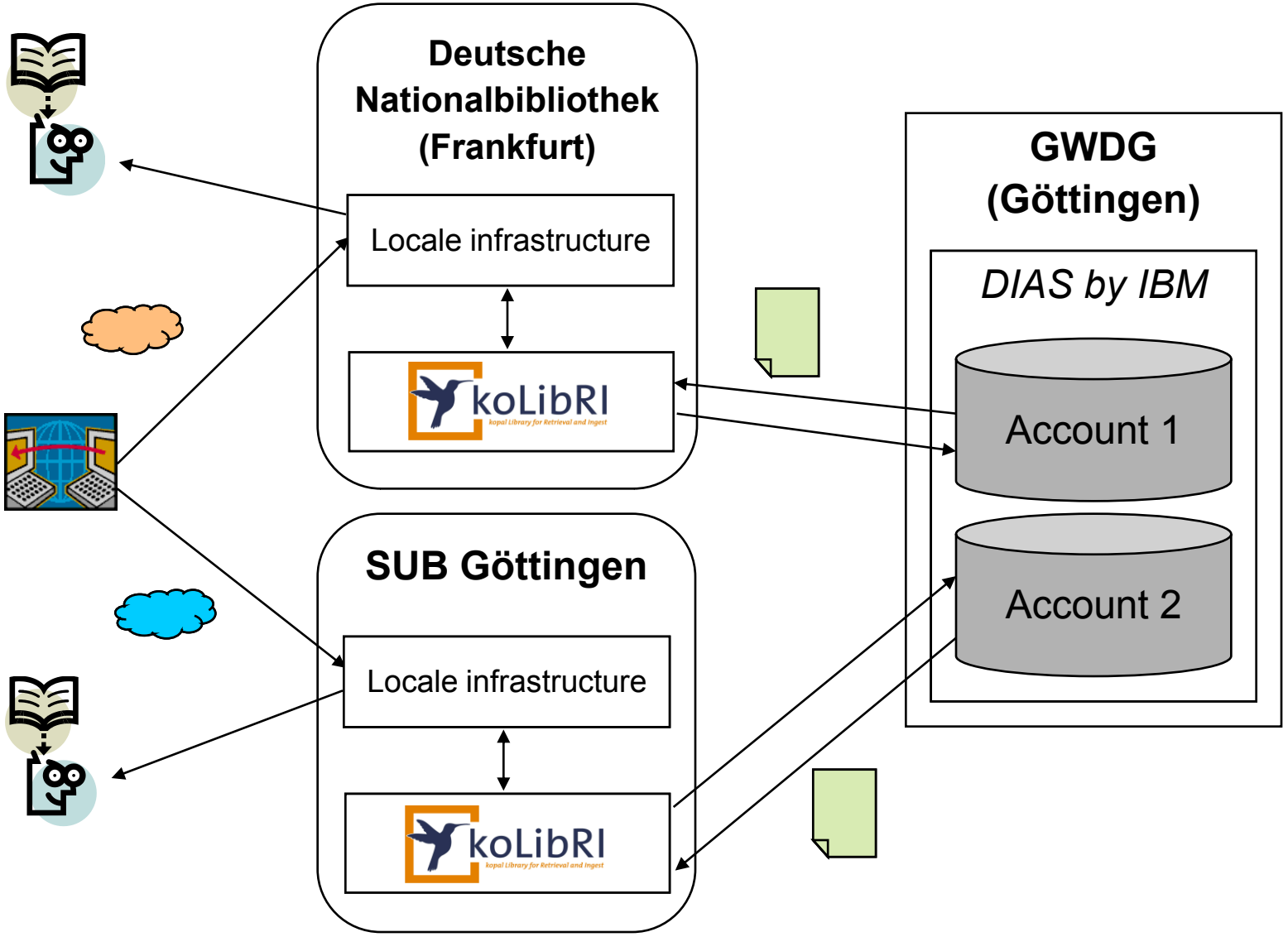
- 2004 – 2007
- Enhancements of DIAS by IBM
- Development of the open source software koLibRI
- Reusable specifications of SIP and DIP (Universal Object Format)
- Working installations of DIAS at GWDG and koLibRI at SUB Göttingen and Deutsche Nationalbibliothek

Technical concepts of kopal

kopal
Data into the future

- Global core (DIAS) and local software (koLibRI)
- Open interfaces: Universal Object Format
- Bitstream Preservation: Storage independence
- Preservation Planning: Migration Manager

Global core and local software



Core: DIAS

kopal
Data into the future

- IBM development for the National Library of the Netherlands
- Conforming to the OAIS reference model
- Based on IBM standard software (DB2, Content Manager, Tivoli Storage Manager, Websphere)
- Enhancements for kopal
  - Open interfaces (Universal Object Format)
  - Support for long-term preservation metadata and versioning
  - Multi-user support (accounts)
  - Query interface to Data Management

Locale software: koLibRI

kopal
Data into the future

- kopal Library for Retrieval and Ingest
- Open source development by Deutsche Nationalbibliothek and SUB Göttingen
- Flexible Java software library to connect the locale software infrastructure to DIAS
- Generates technical metadata using JHOVE
- Creates ingest packages (SIP) in Universal Object Format
- Encapsulates the communication with DIAS (security)
- Migration Manager: Workflow for format migration

Universal Object Format

- Generic description of packages for archiving and exchanging of digital objects
- Based on METS (special METS profile) and LMER (Long-term preservation Metadata for Electronic Resources)
- One packed file (e. g. ZIP or tar) per logical object, containing one METS file with descriptive metadata (e. g. Dublin Core) and technical metadata for long-term preservation (LMER)
- In kopal
  - Used as SIP and DIP (OAIS)
  - Dublin Core and LMER is stored in Data Management of DIAS
  - Other metadata in METS of SIP will be maintained and delivered in METS of DIP

Bitstream Preservation

- No “hard-wired” connection of Data Management and storage type
- Migration of storage media is possible without change in Data Management (Tivoli Storage Manager)
- Checksums to guarantee integrity
- No functionality in DIAS to change archival objects
- Normal hosting: Redundant storage, backup and refresh policies

Preservation Planning

kopal
Data into the future

- DIAS: Query interface to Data Management (technical/preservation metadata and optional Dublin Core)
- koLibRI: Migration Manager to perform queries, call conversion tools and re-ingest migrated objects
- Planned support for a File Format Registry: Database about file format characteristics, view paths, migration paths and emulation environments

Models of reusability

- kopal self-provider
  - Build-up a new system based on DIAS and koLibRI
  - Needed: Hardware, software, hosting service
- kopal user
  - New account in DIAS in Göttingen and customised local koLibRI installation
  - Needed: Software licence, local software integration
- kopal participant
  - Management of data by existing kopal user
  - Needed: Service agreement

Thank you!

kopal
Data into the future

http://kopal.langzeitarchivierung.de

- Download of koLibRI
- kopal demonstrator (Flash animation)
- Specifications of the Universal Object Format