

Attila Zabos - Deutsche Nationalbibliothek (DNB)

Architecture Considerations for Digital Long-Term Preservation at the German National Library (DNB)

Overview

- LTP challenges and cooperated efforts
- Current state at the DNB
- Current research projects
- Future orientation

Challenges for memory institutions

- High volume of digital data
 - Rapid increase of digital object production rate (born-digital object)
 - Automated scan process for books (mass digitisation)
- Ensure discovery and long-term access
- Technology dependency
 - Mainly based on rigid hardware/software solutions
- Authenticity and integrity of digital objects

LTP services

- Joint effort of national libraries in the EU
 - Objective: identification and specification of common LTP services for digital libraries
- Main focus on *ingest*, *retention* and *access*
- Introduce flexibility through a workflow-engine to orchestrate services
 - Enabling the implementation of institution specific business processes

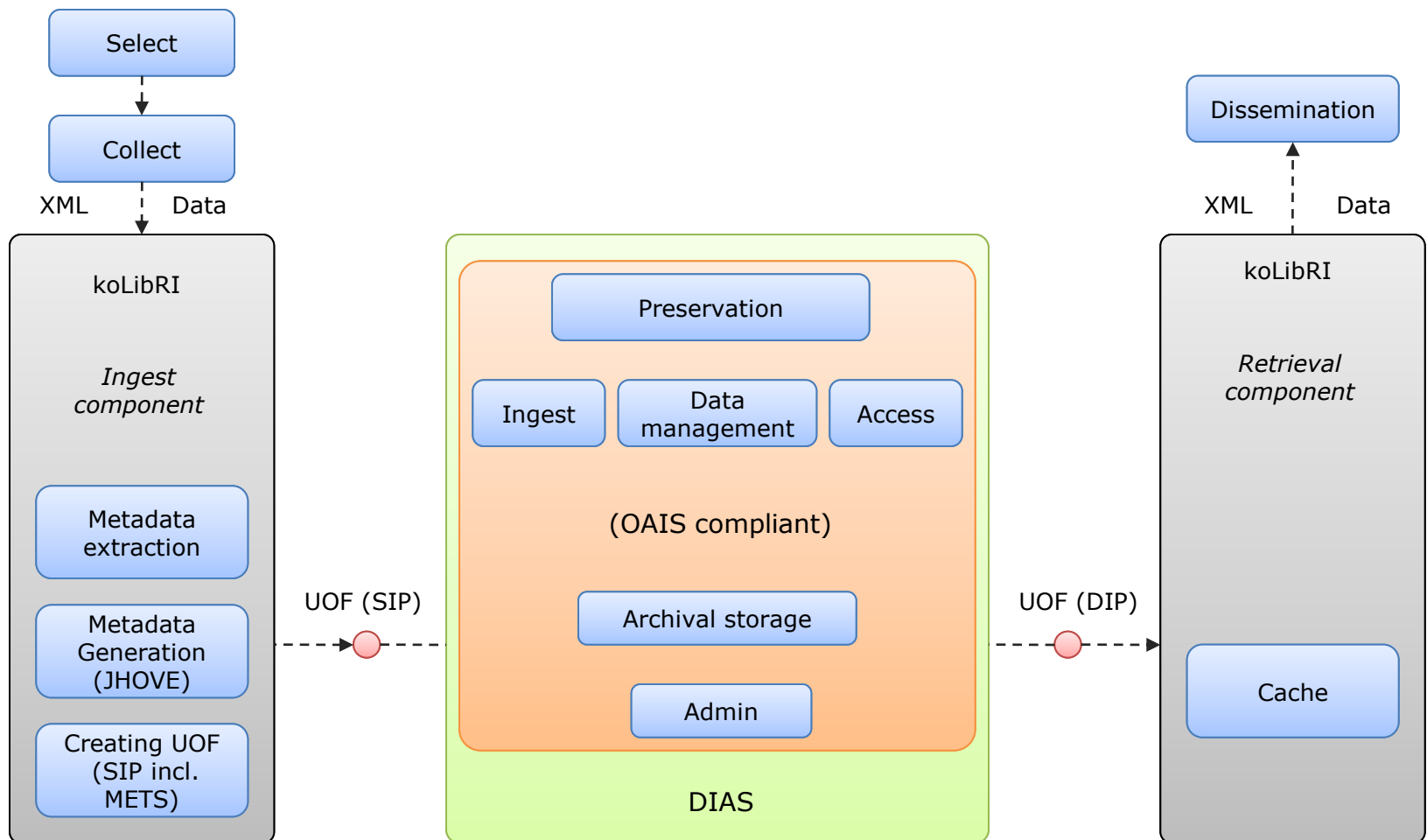
nestor

- Competence network for digital preservation
- DNB is member of standardisation committees
 - Foster the development of IT architectures
 - Influenced by DNB strategies for LTP

kopal

- First *high impact* LTP research project at the DNB
 - Digital archive implemented according to international standards for LTP
 - Evaluating the integration of a digital archive into existing infrastructure
- Ended in 2007
- End result: working prototype

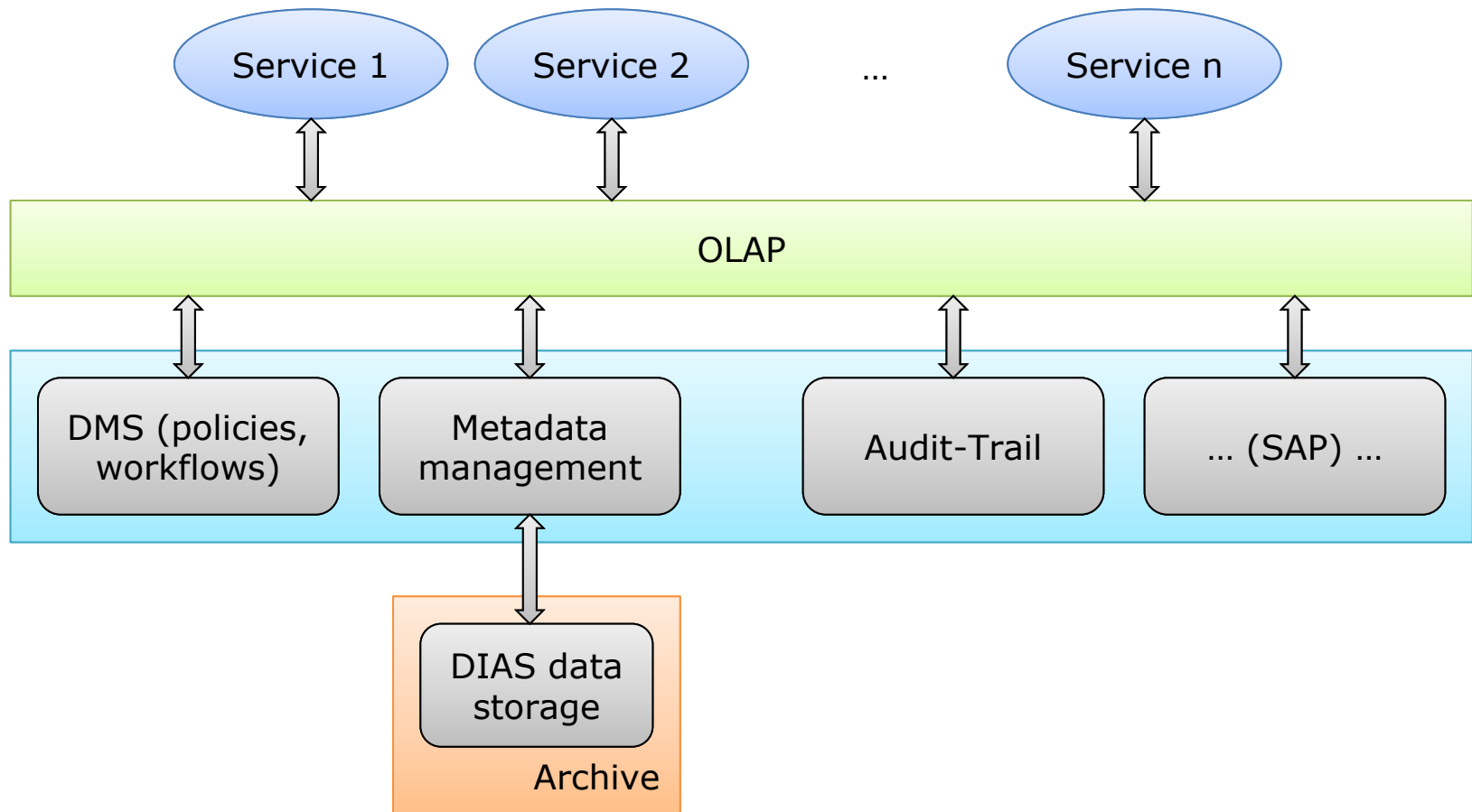
kopal (cont.)



DP4Lib

- Reuse and extend kopal functionality
 - Facilitate component integration into new environments (web services)
 - Maintain field-tested preservation tools (koLibRI)
- Reduce dependency between components
 - Redundant storage at different locations
 - Increase technology independence
- Establish LTP services for productive use

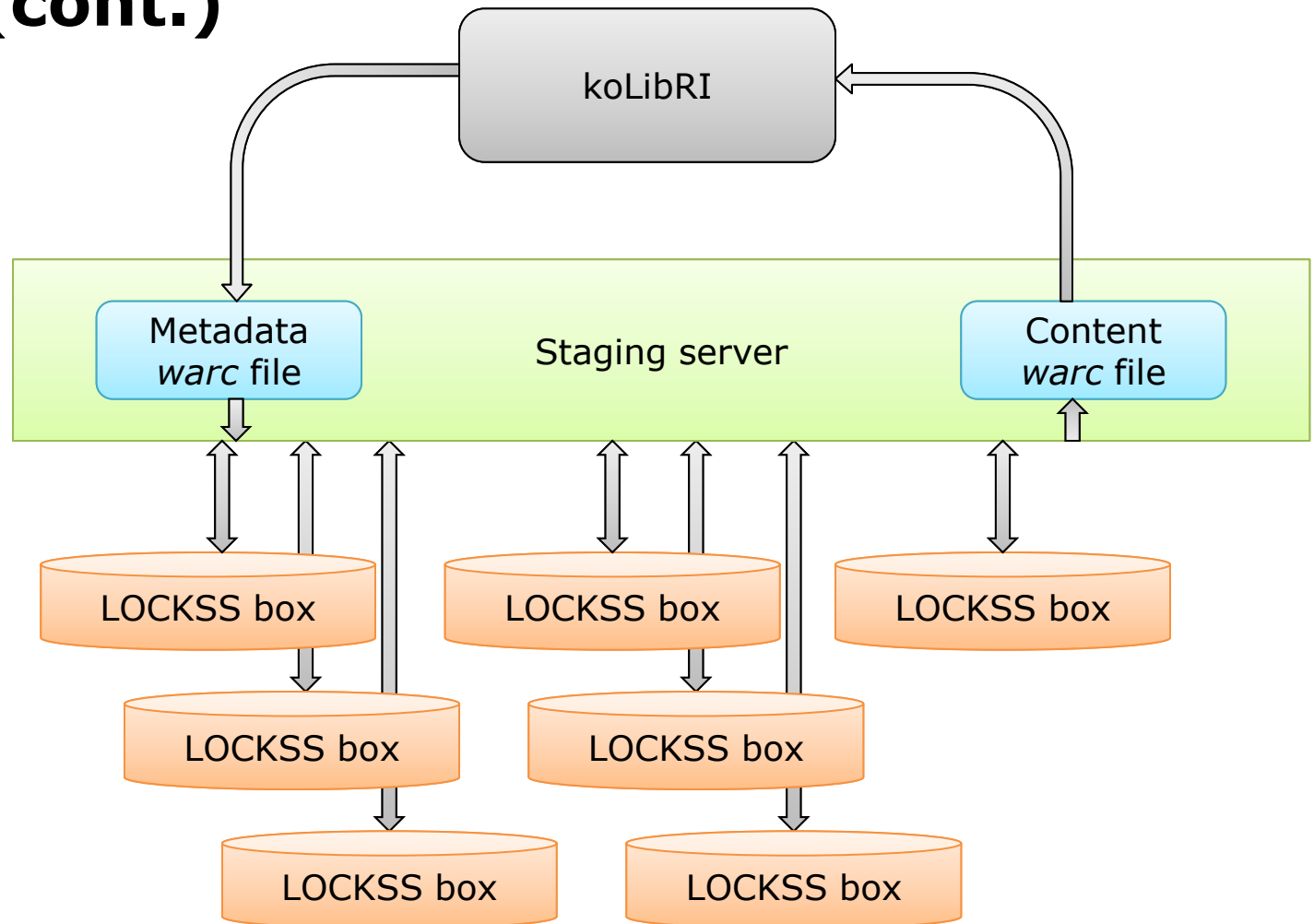
DP4Lib (cont.)



LuKII

- Reuse of established components
 - Cost-effective LOCKSS network
 - Integration of LOCKSS and koLibRI
- Construct and establish a LOCKSS based network in Germany
- Distributed storage, content spread over several nodes
 - Provide safety by repairing lost or corrupted data in a distributed network storage

LuKII (cont.)



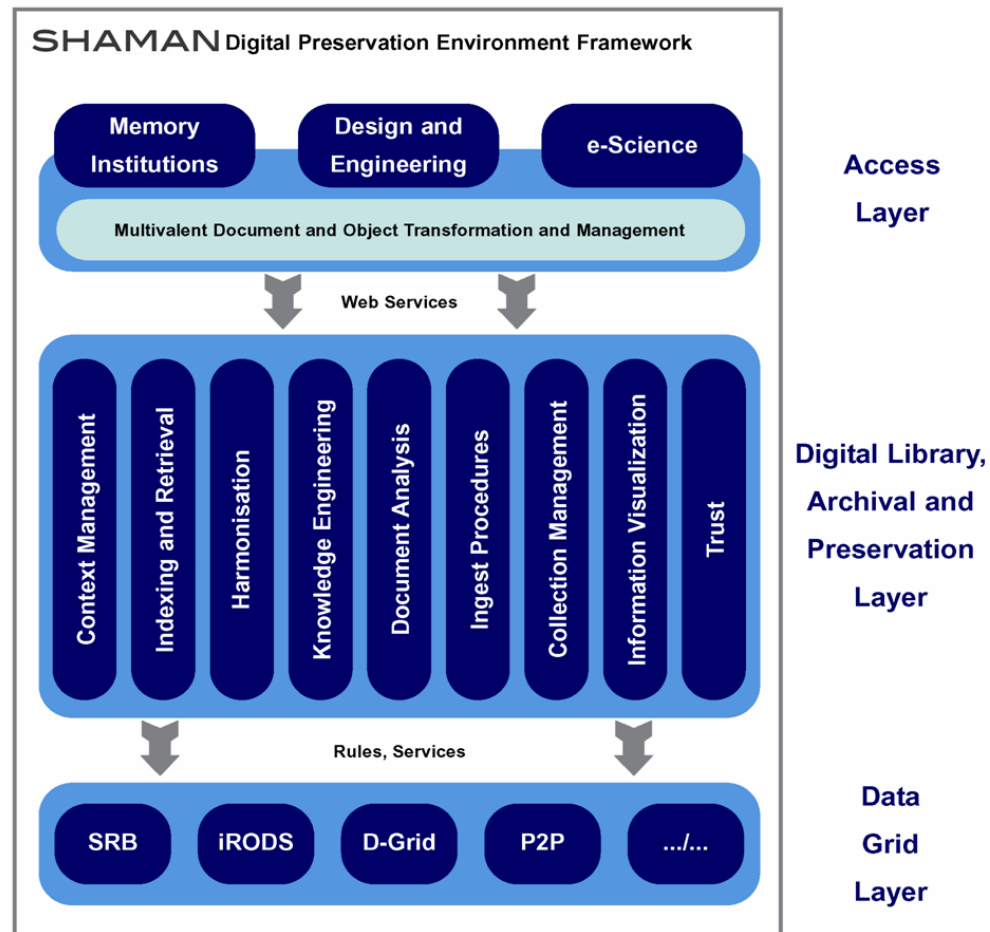
KEEP

- Portability of emulation environment
 - Separation of concerns
 - Capturing the bit-stream content of disk media
 - Rendering captured bit-stream data
 - Extendable frameworks to
 - Capture, store and access digital object
- Created disk images enhanced with metadata
- Layered virtual machine minimises adaptation to new systems

SHAMAN

- Layered framework
 - Technology independence
 - Infrastructure independence
- Distributed system
 - Data grid, managing high data volume
 - Deploying LTP services (using SOA)
- Preservation policies
 - Define strategies, satisfying LTP objectives
 - Origin of workflow specification

SHAMAN (cont.)



DNB strategies and activities

- Short term objectives
 - Build up LTP services to archive digital publications
 - Loosely coupled services enable the assignment of tasks to costumers
 - Integration of monitoring and reporting services
 - Stronger integration and scaling of LTP processes

- Mid and long term objectives
 - Ensure access
 - Shared and networked archiving infrastructure
 - Choosing appropriate architecture to facilitate audit and certification efforts
 - Integration of emulators and transfer tools into existing workflows
 - Preservation risk management

Summary

- Avoid monolithic solutions
- Common trend in research project
 - Distributed architecture
 - Development and providing LTP services
 - Minimise or eliminate technology dependencies
- Current approaches
 - Distributed systems
 - Service Oriented Architecture
 - Layered design, abstraction, virtual machines

Questions?

Contact: Attila Zabos

German National Library

(a.zabos@dnb.de)