

# Neue Entwicklungen zur digitalen Langzeitarchivierung (dLZA) in Bayern



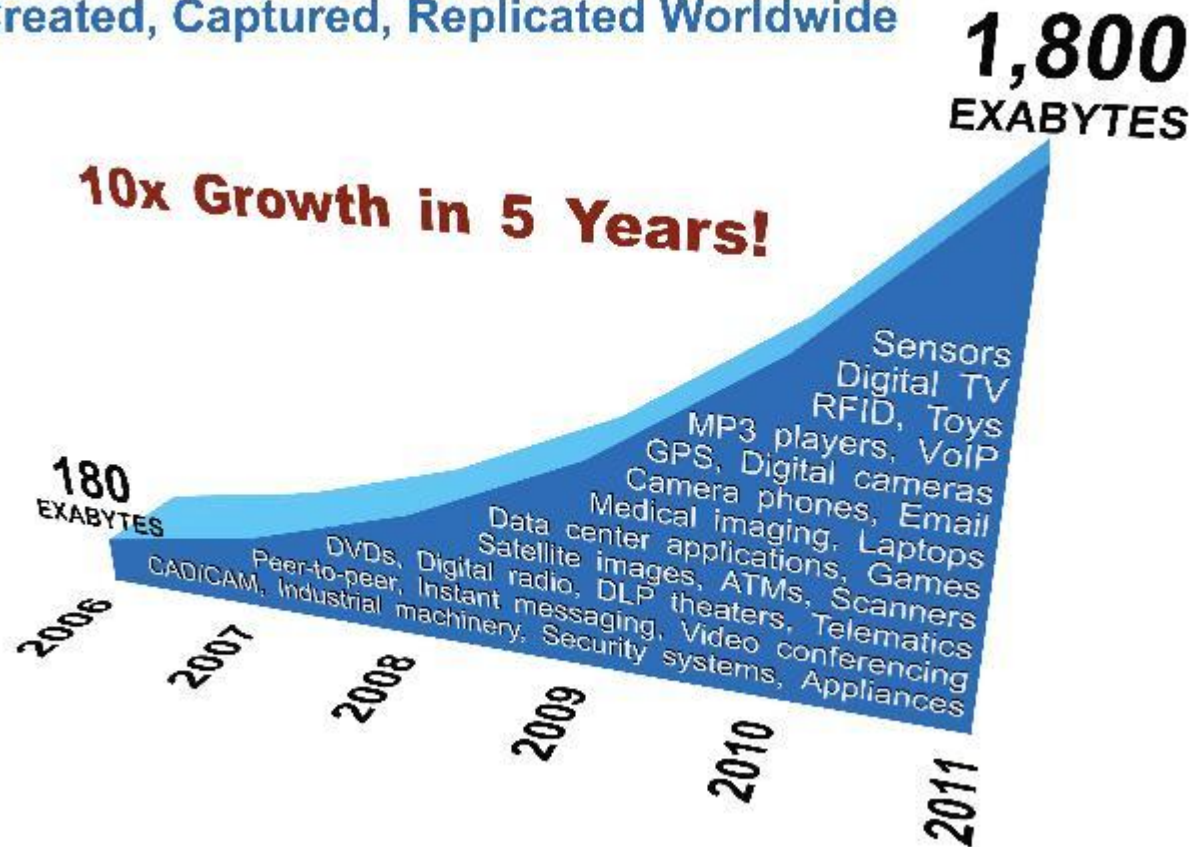
1. Digitale Langzeitarchivierung:  
Herausforderungen, Definition und  
Lösungsansätze
2. Von BABS zu Rosetta – dLZA an der BSB
3. Das Pilotprojekt: Rosetta Digital Preservation  
System in Bayern 2010-2011

---

# 1. Digitale Langzeitarchivierung: Herausforderungen, Definition und Lösungsansätze

# Rasanten Wachstum digitaler Daten

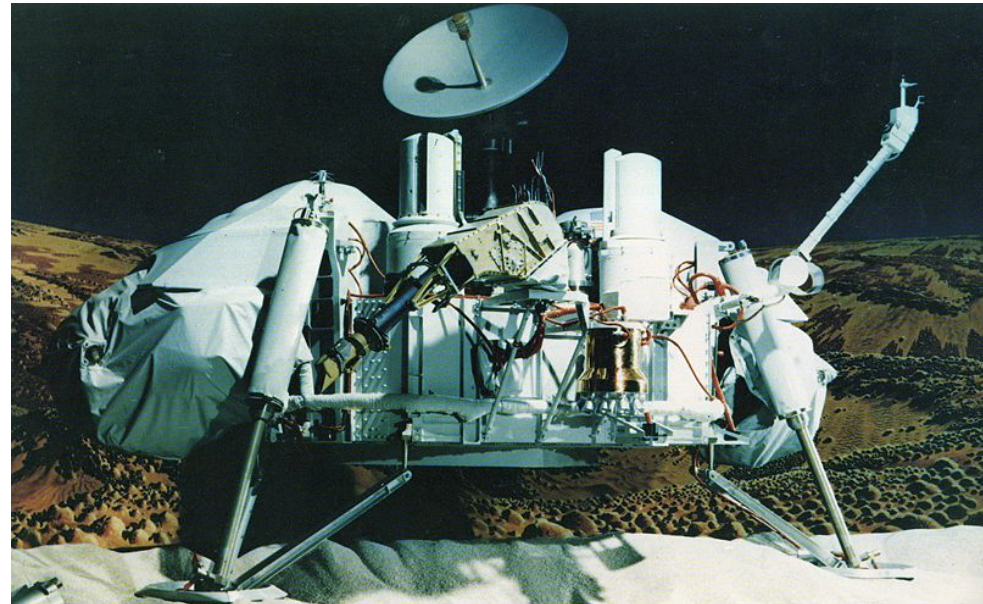
Digital Information  
Created, Captured, Replicated Worldwide



Source: IDC White Paper, "The Diverse and Exploding Digital Universe," sponsored by EMC, March 2008.

# Verlust der Lesbarkeit digitaler Daten

- 1976 Landung der Viking-Sonden auf dem Mars
  - Sammlung von Bodendaten auf Anzeichen von Leben
  
- 1999 Neurobiologie J.D. Miller plant Reanalyse der Viking-Daten
  
- NASA hat die Bänder **aber** in einem **1999 unbekanntem und nicht mehr lesbarem Dateiformat**



## Verlust des Zugriffs

### □ Primärquellen zum Viking-Datenverlust:

Nachrichtenagentur Reuters am 27. Juli 2001

- [http://www.reuters.com/news\\_article.jhtml?type=science&StoryID=137333](http://www.reuters.com/news_article.jhtml?type=science&StoryID=137333)
- <http://reuters.activebuddy.com/s?id=DS1DEKNG8BBN>



### PAGE NOT FOUND

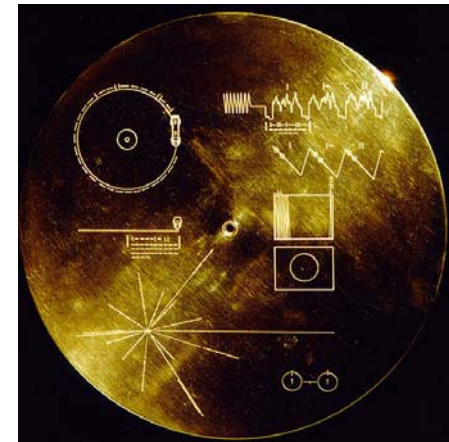
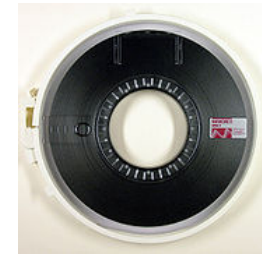
**Our apologies, the requested page was not found. Please double-check the URL for proper spelling and capitalization. If you're having trouble finding a page in Reuters, please chose from the options below:**

[Go to the Reuters home page.](#)

[Browse the Reuters site map](#)

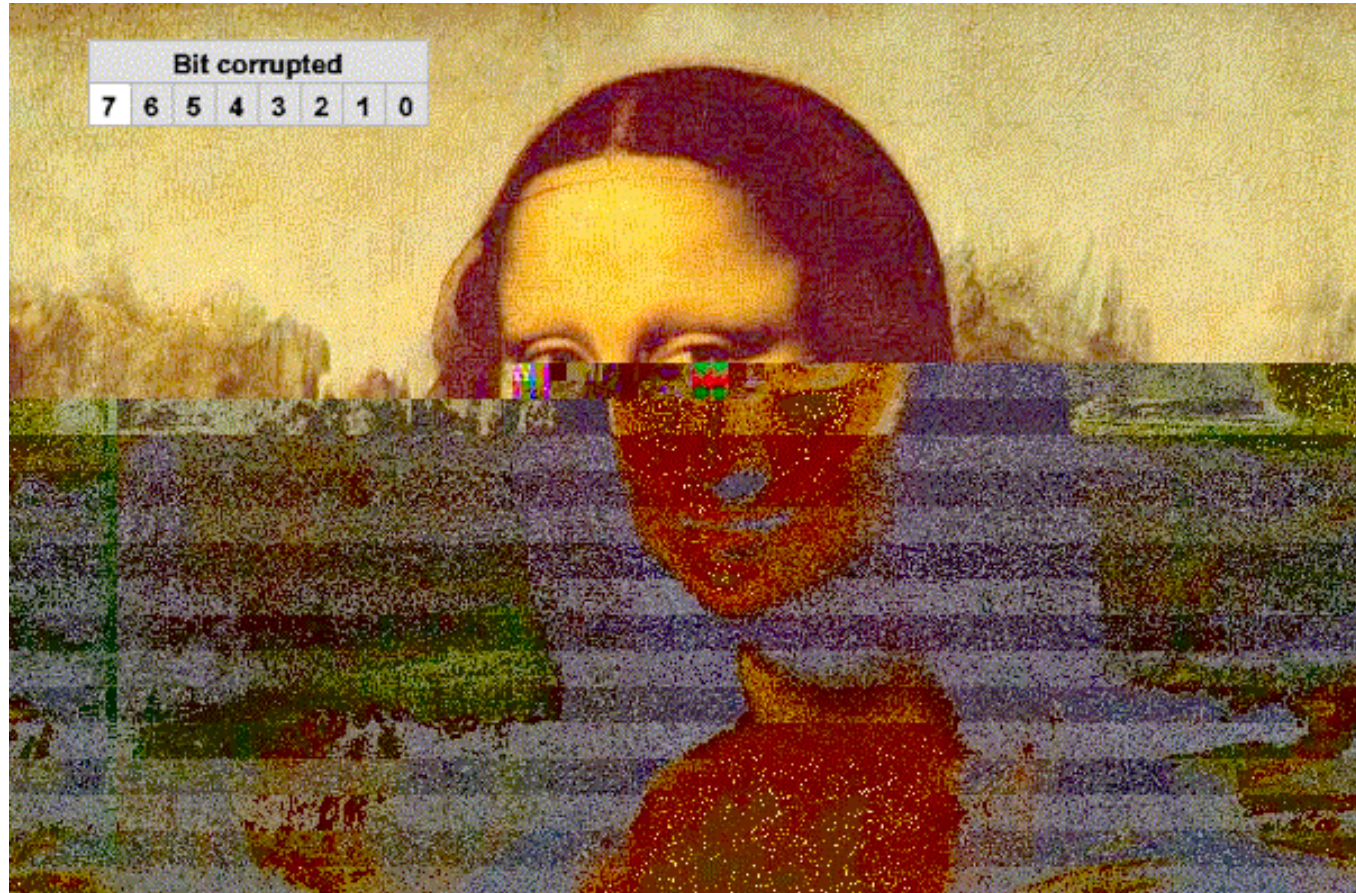
If you still can't find what you're looking for, you may wish to send a message to the [Help Desk](#)

# Verlust durch Alterung der Datenträger



Golden Record –  
Haltbarkeit: „500 Mio.  
Jahre“ (!)

# Verlust durch korrupte Bits oder Bytes





## Definition der dLZA

- Unter digitaler Langzeitarchivierung versteht man die **Sammlung**, die **langfristige Aufbewahrung** und die **Erhaltung der dauerhaften Verfügbarkeit** von Informationen in digitaler Form
- „Langzeit“ = keine Garantierklärung über fünf oder fünfzig Jahre, sondern die **verantwortliche Entwicklung von Strategien**, die den beständigen, vom Informationsmarkt verursachten Wandel bewältigen können

# Lösungsansätze

Erhaltung des Bitstream + Migration



Emulation

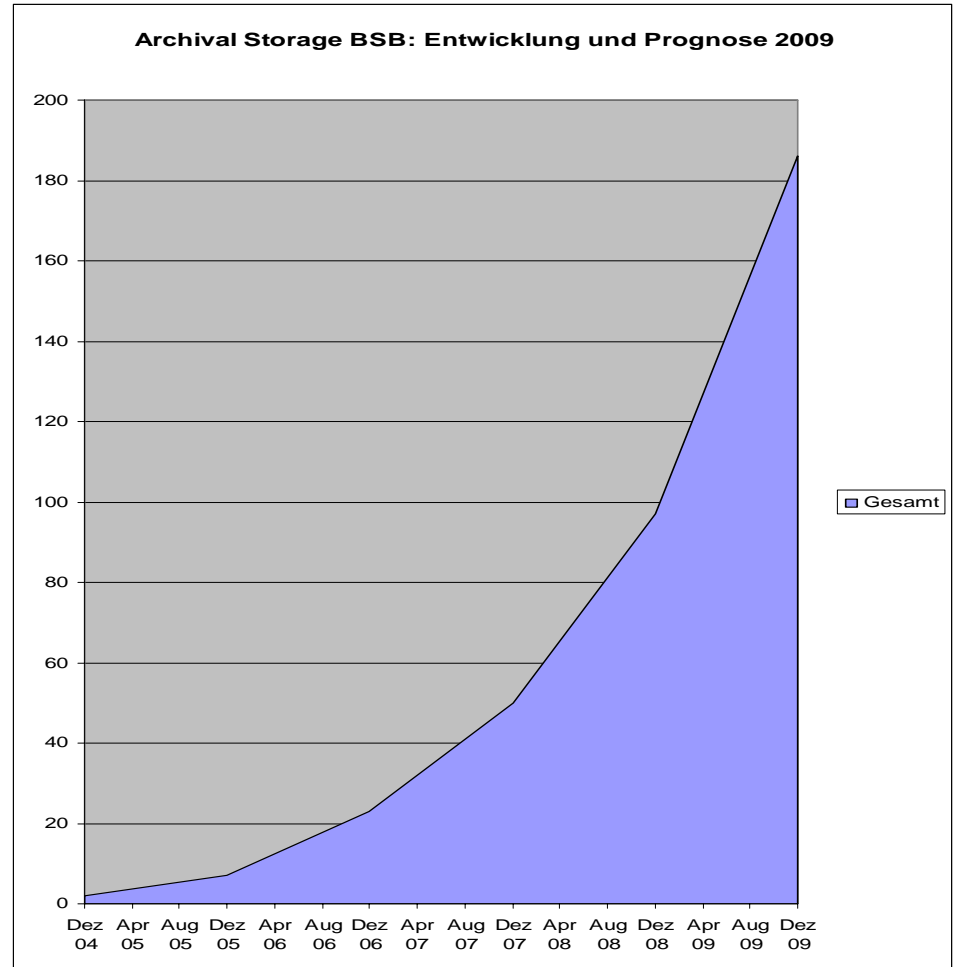
Hardware-Museum

---

## 2. Von BABS zu Rosetta - digitale Langzeitarchivierung an der BSB

# BSB-dLZA: Entwicklung der Datenmenge

- Nov. 2009 im Archivspeicher des LRZ gelagerte Dateien des MDZ („born digital“ und Retrodigitalisate m. Google):
  - 259 Mio.
  - Gesamtdatenvolumen: 171 Terabyte
- Mit Daten aus Kooperationsprojekten, z.B. Staatliche Bibliotheken, universitäre und außeruniversitäre Forschungsprojekte
- BSB drittgrößter Archivierungskunde des LRZ



# Kernaufgaben der dLZA an der BSB

---

## Drei überlappende Bereiche

- Schaffung der notwendigen **inhaltlich-konzeptionellen, organisatorischen und technischen** Voraussetzungen
- **operativer Betrieb** eines digitalen Langzeitarchivs
- **Forschung und Entwicklung**

# BSB: dLZA-Projekte seit 1999

- 1999
  - Organisatorisch-technisches Konzepts für die Langzeitarchivierung beigebundener digitaler Publikationen in Bibliotheken
- Seit 2003
  - nestor - Aufbau eines Expertennetzwerkes für dLZA in Deutschland
- 2005-2007
  - BABS I - Aufbau einer organisatorisch-technischen Infrastruktur
- 2009-2010
  - BABS II Selbstevaluierung und Verbesserung des BABS-Systems bezogen auf Vertrauenswürdigkeit und Skalierbarkeit (2009-2010)
- 2008-2009
  - Studie Organisations- und Geschäftsmodelle für die dLZA
- 2009
  - LOCKSS-Teilnahme - internationales Projekt in Kooperation mit der Library of Congress

der bundeswehr  
universität  münchen

**DFG**

nestor 

**BABS**  
Das Bibliothekarische  
Archivierungs- und  
Bereitstellungssystem

**DFG**

**lrz**

**BABS2**

**DFG**

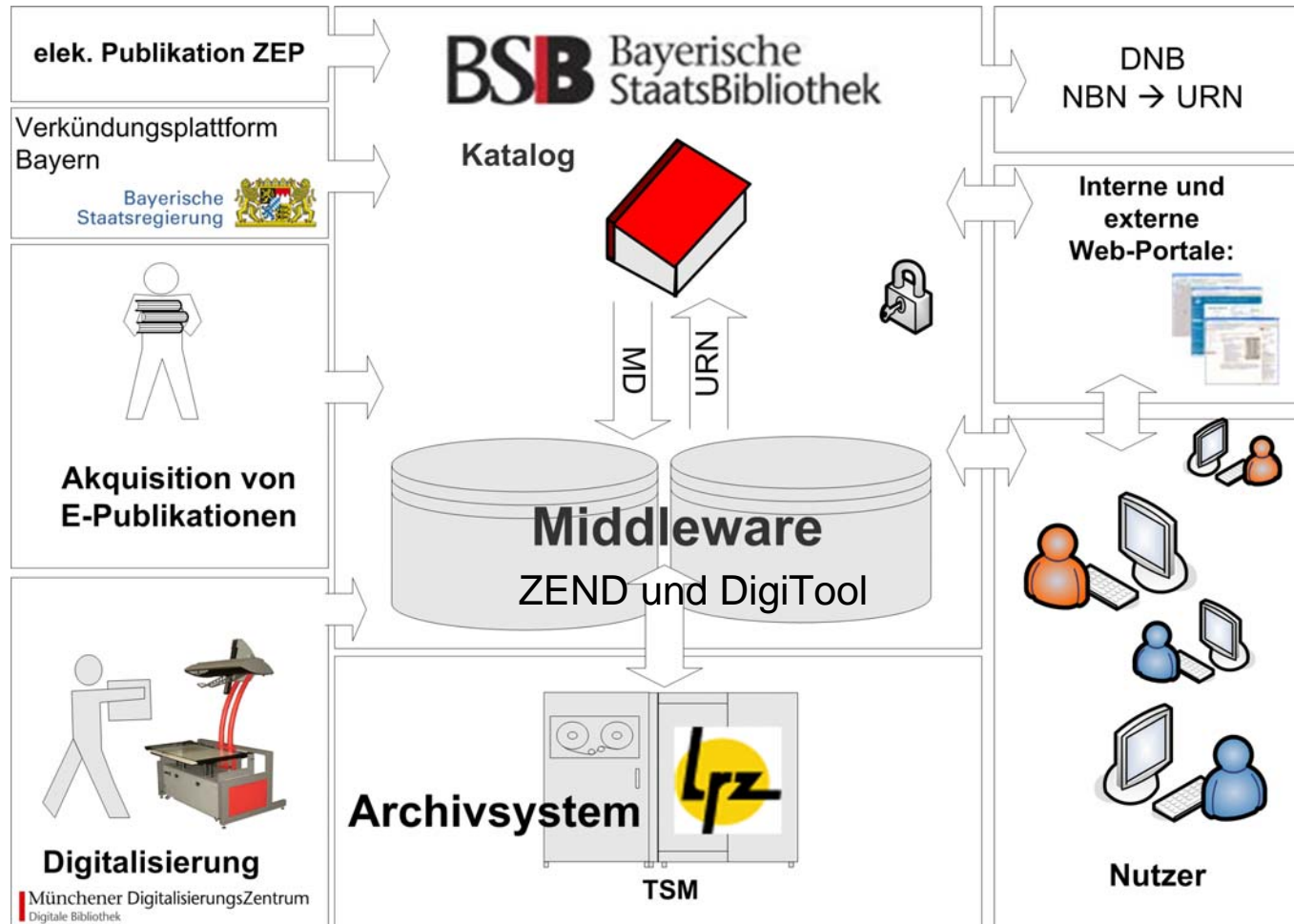
**lrz**

der bundeswehr  
universität  münchen

**DFG**

  
**LOCKSS**

# BABS-Architektur



## BABS-Technik

- Ziel: **Integrierter Ansatz** von der Datenübernahme ins Archiv bis zur Endnutzerbereitstellung
  
- Infrastruktur aus bestehenden technischen Modulen der BSB (MDZ u. BVB) und des LRZ
  - BSB-Module
    - **ZEND** (MDZ-Eigenentwicklung; Electronic Publishing Software) für Digitalisate und
    - **DigiTool 3.0** (ExLibris) für „born digital“-Dokumente
  - LRZ: **Backup- and Archival System**
  
- **Aber** es werden nicht alle Anforderungen an ein digitales Langzeitarchiv, z.B. OAIS erfüllt ...



# Open **A**rchival **I**nformation **S**ystem

---

## Referenzmodell

- Beschreibt ein digitales (oder physisches) Archiv als Organisation, in dem Menschen und Systeme zusammenwirken, um dessen Inhalte einer definierten Nutzerschaft verfügbar zu machen

## Gemeinsame Konzeption von NASA und ESA

## ISO-Standard 14721:2003 (*Space data and information transfer systems - Open archival information system - Reference model*)

# OAIS in Kürze

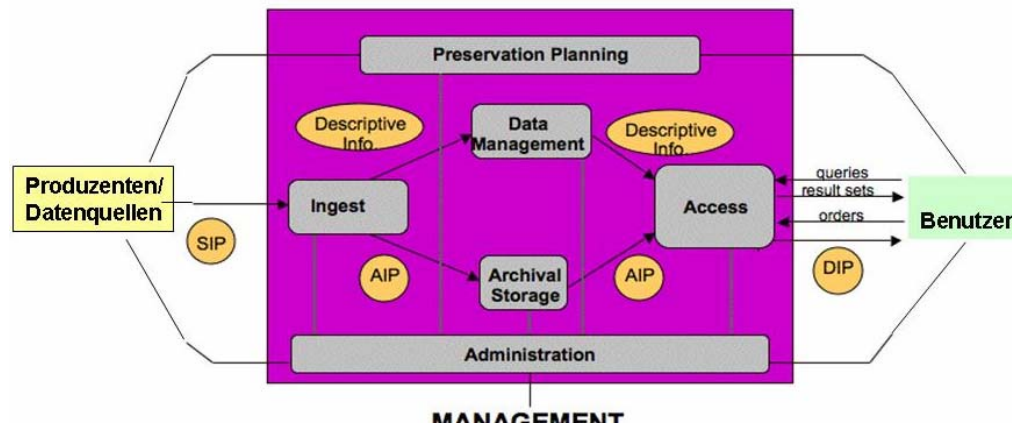
## □ OAIS definiert verschiedene Rollen

- Produzent, Benutzer, Management im OAIS-Archiv

## □ OAIS enthält

- Informationsmodelle (SIP, AIP, DIP)
- Funktionales Modell: Ingest, Archival Storage, Data Management, Administration, Preservation Planning, Access

SIP = Submission Information Package  
AIP = Archival Information Package  
DIP = Dissemination Information Package



# Beispiel: OAIIS-Preservation Planning

---

- Iterativer Prozess der Risikobewertung, z.B. von Dateiformaten
  1. Identifizieren
  2. Bewerten
  3. Plan erstellen
  4. Umsetzung
  5. Aktualisierung und Überprüfung

# Auswahl-Kriterien für das neue dLZA-System

---

u. a.

- Skalierbarkeit – mehrere Mio. Objekte
- Sicherheits- und Qualitätskonzept
- Rechteverwaltung
- Offene
  - Schnittstellen via API, SDK
  - Standards: OAIS, PREMIS, METS, DC, OAI, SRU ...
- Implementierung des OAIS-Referenz-Modells mit
  - Übernahme und Verwaltung digitaler Objekte
  - Planung und Durchführung von Erhaltungsmaßnahmen, Migration
  - Archivspeicherung und die Administration des Archivs
  - Bereitstellung der Informationen für die Nutzer

# dLZA-Systeme kommerzieller Anbieter

---

- IBM *DIAS*
- Tessella *Safety Deposit Box*
- ExLibris *Rosetta Digital Preservation System*
  - DigiTool 3.0 wird als *Rosetta Digital Asset Management System* ab 2011 Teil des *Rosetta Digital Preservation Systems*
  - DigiTool seit 2005 in der Verbundzentrale lizenziert

## Rosetta Digital Preservation System- Vorarbeiten und Entwicklung

- **2004** Start des National Digital Heritage Archive Program (NDHA) in Neuseeland
- **2005/2006** Detaillierte Spezifikation der Anforderungen mit der **Peer Review Group** und Analyse bestehender Systeme durch das NDHA
- Seit **2007** Entwicklungspartnerschaft der Nationalbibliothek von Neuseeland (NLNZ) und ExLibris
- **Nov. 2008** Version 1.0 startet
- **Jan. 2009** Start des "Charter Group Program" durch ExLibris
  - Ziel: Mitwirkung an der Entwicklung von Rosetta

**NDHA Peer Review Group**  
British Library  
Cornell University  
Library  
Getty Research  
Institute  
National Library of  
Finland  
Koninklijke Bibliotheek  
National Library Board  
of Singapore  
National Library of  
China  
Sun Microsystems  
University of Glasgow  
Yale University

---

## 4. Das Pilotprojekt: Rosetta Digital Preservation System in Bayern 2010-2011

# Rosetta in Bayern - Organisation

---

## □ Projektteam:

- BSB (Münchener Digitalisierungszentrum + Verbundzentrale)
- Leibniz-Rechenzentrum
- Ex Libris
- im Projektverlauf weitere Pilotpartner (UB Augsburg, Regensburg, Würzburg)

## □ Lenkungsausschuss

- LS für Informatik, UniBW
- IuK-Referat des Bay. Staatsministeriums für Wissenschaft, Forschung und Kunst
- KEP
- BSB



## Rosetta – Planungen für das Projekt (1)

---

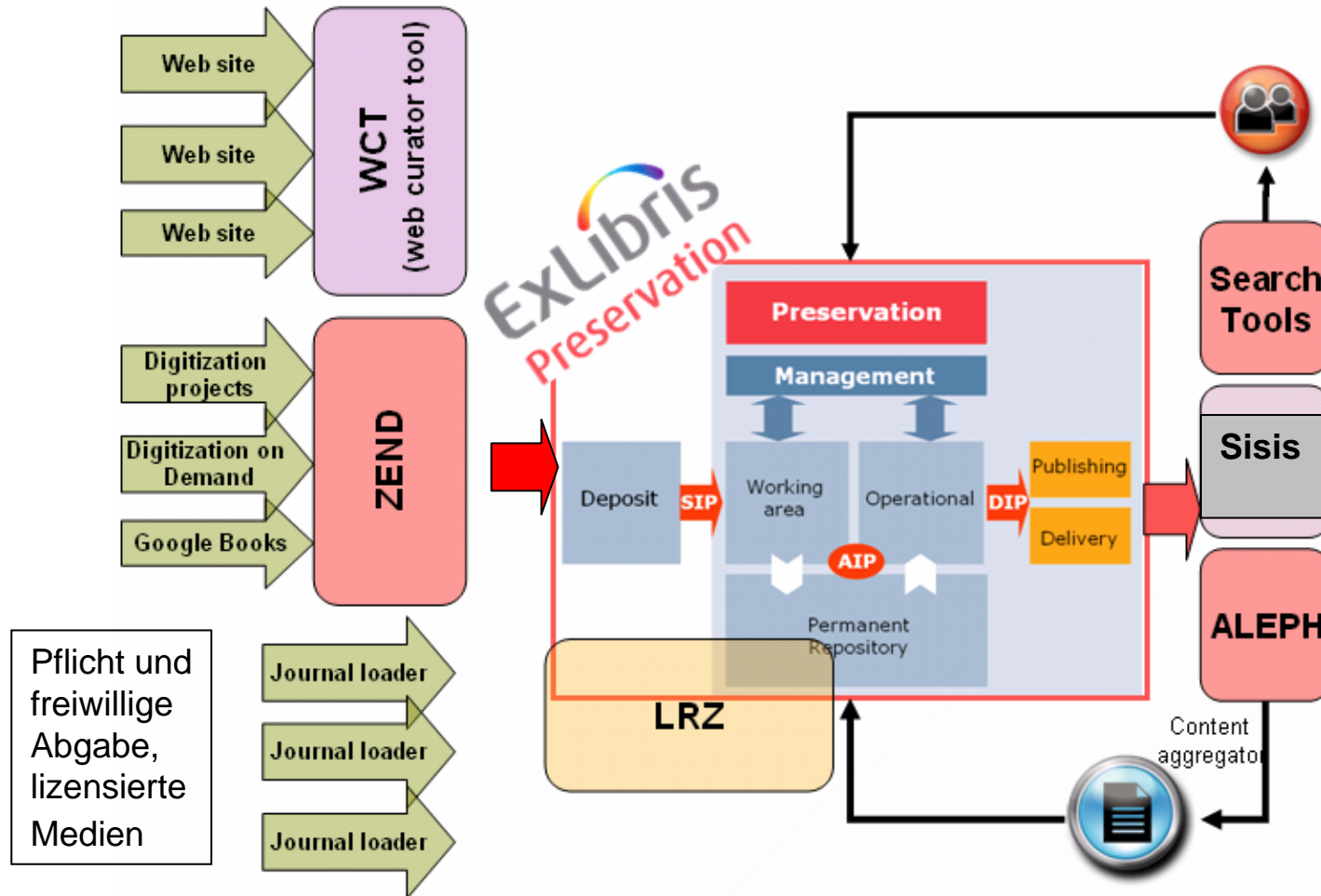
- Implementierung, Konfiguration, Test und gemeinsame Weiterentwicklung von Rosetta im Rahmen eines Charter-Group-Programms von Ex Libris
- Spezifikation, Aufbau, Test und Konsolidierung des Pilot-Systems durch BSB, LRZ und Ex Libris für 2-3 ausgewählte Workflows und Materialgruppen
- Integration in die bestehende technische Systemstruktur

## Rosetta – Planungen für das Projekt (2)

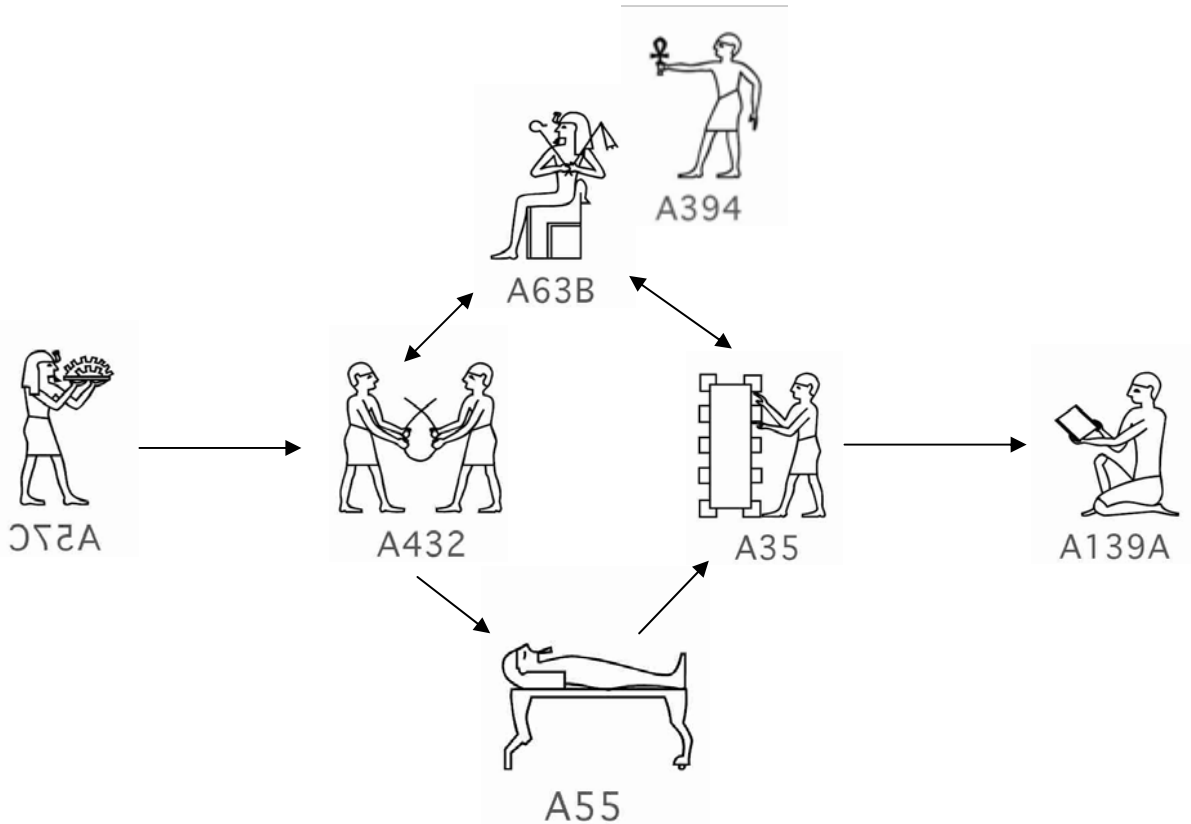
---

- Integration digitaler Bestände weiterer bayerischer Pilotbibliotheken
- Übergang in den Regelbetrieb
- Ausdehnung auf weitere Workflows der BSB
- Konsolidierung der Infrastruktur für digitalen Langzeitarchivierung auch in den Dimensionen Organisation, Personal, Finanzen, Recht

# Rosetta in Bayern - Einbindung in die Softwarearchitektur



# Rosetta-Module leicht gemacht



## Rosetta: Roadmap

---

- Dez. 2009: Beta-Release von Version 2.0
- Q1/2010: General release von Version 2.0
- 2011: Migration von DigiTool zu Rosetta Digital Asset Management Module (Rosetta DAM)
- 2010/2011: Beta-Release von Version 3.0

