

Langzeitarchivierung in der Bayerischen Staatsbibliothek

Dr. Thomas Wolf-Klostermann
Dresden, 24. März 2006



Ref. Digitale Bibliothek



Langzeitarchivierung in der Bayerischen Staatsbibliothek

Agenda

- Langzeitarchivierung elektronischer Dokumente im Überblick
- Digitalisierung und Langzeitarchivierung in der Praxis der BSB
- Aktuelles DFG-Projekt *Langzeitarchivierung von Netzpublikationen*



Die Digitale Bibliothek der BSB

- Gegründet 1997 als *Münchner Digitalisierungszentrum* mit Anschubfinanzierung der DFG
- 2003 als Referat *Digitale Bibliothek* in die Abteilung *Bestandsaufbau und Erschließung* der Bay. Staatsbibliothek integriert
- Sachgebiete:
 - Digitalisierung
 - Aufbau von Fachportalen
 - Langzeitarchivierung digitaler Medien
- Kontakt: mdz@bsb-muenchen.de

Elektronische Dokumente – Aufgaben der Bibliothek

- Akquisition / Eigenproduktion
 - durch Kauf / Lizenzierung
 - durch (gesetzlich geregelte) Pflichtabgabe
 - Harvesting von Internet-Ressourcen
 - Digitalisierung, E-Publishing
- Erschließung
- Bereitstellung zur Benutzung
- Langzeitarchivierung**

Bereitstellung elektronischer Dokumente in der BSB

Elektronische Zeitschriftenbibliothek - Microsoft Internet Explorer

ViFaOst - Geschichte - Microsoft Internet Explorer

Digitale Bibliothek - Münchener Digitalisierungszentrum - Microsoft Internet Explorer

Bayerische Landesbibliothek Online - Bayerische Landesbibliothek Online (BLO) - Microsoft Internet Explorer

Adresse <http://www.bayerische-landesbibliothek-online.de/>

Bayerische Landesbibliothek Online (BLO)

Startseite

Orte

Personen

Schnellsuche:

Ort

- Bayerische Bibliographie
- Bayerische Landesgeschichtliche Zeitschriftenschau
- Historisches Lexikon Bayerns
- WebWeiser
- Landtagsprotokolle
- Historische Karten
- Ortsblätter des 19. Jahrhunderts
- Historischer Atlas von Bayern
- Zeitschrift für bayerische Landesgeschichte
- Nürnberg Mittelungen
- Augsburger Projekte
- Zeitungen
- Franconica-Online
- Regensburger Portratsammlung
- Preysing-Tagebücher
- Ausstellungskataloge des Münchner Glaspalastes

Die "Bayerische Landesbibliothek Online" ist ein Informationsportal **bayerischer Bibliotheken**, das umfangreiche und vielfältige digitale Ressourcen zur Geschichte Bayerns zusammenfaßt.

Das Angebot wird ständig erweitert. Derzeit sind folgende Quellen abrufbar:

- Ortsdatenbank - historische Quellen, Daten, Fakten und Landkarten, gegliedert nach den bayerischen Ortsnamen
- Personendatenbank - Quellen, Daten und Fakten zu bayerischen Persönlichkeiten aus 11 Jahrhunderten
- Historisches Lexikon Bayerns - ein im Aufbau befindliches wissenschaftliches Online-Lexikon zur bayerischen Geschichte
- Bayerische Bibliographie - eine der größten Regionalbibliographien im deutschsprachigen Raum mit einem komfortablen Zugriff auf *gedruckte* Bibliotheksbestände
- Bayerische Landesgeschichtliche Zeitschriftenschau 1995 bis 2001 mit Artikeln über Bayern aus 300 historischen Zeitschriften
- WebWeiser Bayern - eine thematische Suchmaschine
- Protokolle des Bayerischen Landtags (16. - 18. Jh., 1918-33, 1946)
- Historische Karten
- Ortsblätter des 19. Jahrhunderts - Bayerns Städte und Dörfer auf 2400 historischen Kartenblättern des Landesamtes für Vermessung und Geoinformation
- Historischer Atlas von Bayern - die Besitz-, Herrschafts- und Verwaltungsstruktur Bayerns vom Mittelalter bis heute statistisch dokumentiert - Online-Version, mit derzeit 67 vergriffenen Bänden
- Zeitschrift für Bayerische Landesgeschichte Online-Version (verfügbare Bände 1-62 1928-1999)
- Nürnberg Mittelungen Online-Version der Mitteilungen des Vereins für Geschichte der Stadt Nürnberg (verfügbare Bände 1-88 1879-2001)
- Franconica-Online: Digitale Quellen zur Geschichte Frankens in der BLO
- Augsburger Projekte
- Regensburger Portratsammlung
- Die Tagebücher von Johann Maximilian Emanuel von Preysing
- Die Ausstellungskataloge der Kunstausstellungen im Münchner Glaspalast (1869-1931)
- Zeitungen - Historische Zeitungen aus verschiedenen Regionen Bayerns

Die Ausstellungskataloge des Münchner Glaspalastes (1869-1931)

Dieses Projekt wird gefördert vom Bayerischen Staatsministerium für Wissenschaft, Forschung und Kunst

Internet

Langzeitarchivierung?

Warum Langzeitarchivierung?

- Steigende Zahl digitaler Dokumente
 - ausschließlich digitale Dokumente (born digital)
 - in vielfältigen Formaten und Erscheinungsformen
- Mehrwert digitaler Dokumente
 - Bereitstellung
 - Produktion von Derivaten
- Sicherung der Wertschöpfung

- Physischer Verfall
- Technologische Obsolenz von Soft- und Hardware

Was ist Langzeitarchivierung?

- **Langzeitarchivierung** („*long-term digital preservation*“) = Aktivitäten, die dem Erhalt der Verfügbarkeit der Dokumente über die Lebensdauer der Trägermedien und die Grenzen des technologischen Wandels hinweg dienen
- **Langzeit...?** = eine unbegrenzt in die Zukunft reichende Zeitspanne mit tiefgreifenden technologischen Veränderungen und sich wandelndem Nutzungsverhalten

Kernaufgaben der Langzeitarchivierung

- Physische LZA („bitstream preservation“)
 - Systemsicherheit, Notfallvorkehrungen,
 - Refreshing, Kopieren auf neue Speichermedien...

- Logische LZA
 - Migration
 - Emulation
 - Erhalt der alten Hard-/Software

Langzeitarchivierung international Beispiele

- A. WebArchiving:
 - Nationalbibliothek Australien:
<http://pandora.nla.gov.au/index.html>
 - Nordic Web Archive <http://nwa.nb.no/>
 - Library of Congress: Minerva <http://www.loc.gov/minerva/>

- B. Archivierung von Verlagspublikationen
 - Niederländische Königliche Bibliothek in Kooperation mit IBM
http://www.kb.nl/kb/resources/frameset_kenniscentrum.html

Langzeitarchivierung in Deutschland

- Die Deutsche Bibliothek
 - Dissertationen und Netzpublikationen (Rahmenvertrag mit dem Börsenverein zur freiwilligen Abgabe von Netzpublikationen)
 - Kooperatives Projekt KOPAL (<http://kopal.langzeitarchivierung.de>)
- Baden-Württembergisches Online-Archiv (<http://www.boa-bw.de>)
- Rheinland-Pfälzischer Archivserver Edoweb (<http://www.rlb.de/edoweb.html>)
- Projekt NESTOR: Aufbau eines Kompetenznetzwerks für die Langzeitarchivierung und Langzeitverfügbarkeit in Deutschland (<http://www.langzeitarchivierung.de>)
 - Partner: Bibliotheken (Frankfurt, Göttingen, Berlin, München), Archive, Museen

Die Langzeitarchivierung in der Digitalen Bibliothek der BSB

- **1999-2001:** Erste strategische Konzepte, DFG-Projekt mit der Univ. der Bundeswehr, Bestandsaufnahme und Migration von Altdaten
- **Seit Mitte 2003:** Kompetenznetzwerk *nestor* zusammen mit nationalen Partnern (BMBF-Förderung)
- **Januar 2004:** Auflösung des CD-ROM-Archivs, Umstellung auf Bandspeicherung (Leibniz-Rechenzentrum München)
- **Seit Mitte 2004:** Zentrale Erfassungs- und Nachweisdatenbank für Digitalisate (ZEND), Workflow mit automatischer Archivierung
- **Aktuell:** Pilotprojekt Langzeitarchivierung von Netzpublikationen (DFG-Förderung) und *nestor*-Mitarbeit (Schwerpunkt *Vertrauenswürdige Archive*)

Digitalisierung und Langzeitarchivierung in der Praxis der BSB

Standard-Workflow

- 1. Vorbereitung
- 2. Digitalisierung
- 3. Strukturerrfassung und Nachbearbeitung
- 4. Bereitstellung
- 5. Archivierung

Unterstützt durch die Zentrale Erfassungs- und
Nachweisdatenbank (ZEND)



1. Vorbereitung

- Recherche/Erfassung im Katalog, Übernahme der bibliographischen Daten
- Vergabe der eindeutigen Dateinamen:
 - Bsp: **bsb00001119_00001.tif**
- Vergabe eines persistenten Identifikators:
 - National Bibliography Number
- Bsp: urn:nbn:de:bvb:12- **bsb00001119**
 - Vergabe erfolgt lokal, Verwaltung und Auflösung an der DDB:
[http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:12- **bsb00001119**](http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:12-bsb00001119</b)

2. Digitalisierung

- ❑ In der eigenen Digitalisierungsstraße oder durch Dienstleister
- ❑ Auflösung 400 bis 600ppi
- ❑ Formate bis A0 möglich
- ❑ Archivformat: grundsätzlich TIFF unkomprimiert



3. Strukturierung

- Standard: TEI/Ebind
- XML-Strukturdateien:
 - Header: bibliographische Daten,
 - Body: Verknüpfung von inhaltlicher Information und Imagedateien
- Qualitätskontrolle und Nachbesserungen
 - Integrität
 - Vollständigkeit

Beispiel *Coburger Zeitung*

The screenshot shows a digital archive interface for the Coburger Zeitung. At the top, the browser address bar displays the URL: <http://nbn-resolving.org/urn:nbn:de:bsz:bsb11-coburger-zeitung/>. The page header includes the title "Coburger Zeitung" with the years "1854-1955" and a navigation menu with "Übersicht Jahrgänge". Below this, it specifies "Coburger Zeitung, Coburg 1881, Bd. 21, 1881".

The main content area is a scanned page of the newspaper from Monday, January 3, 1881. The masthead reads "Coburger Zeitung" and "Regierungs-Blatt für das Herzogthum Coburg." It is published by "Verlag von C. B. Wimmer" at "Koburg, Brühl Nr. 35." The page contains several columns of text, including a section titled "Freies Reich." and a circular stamp from the "Coburger Landesbibliothek".

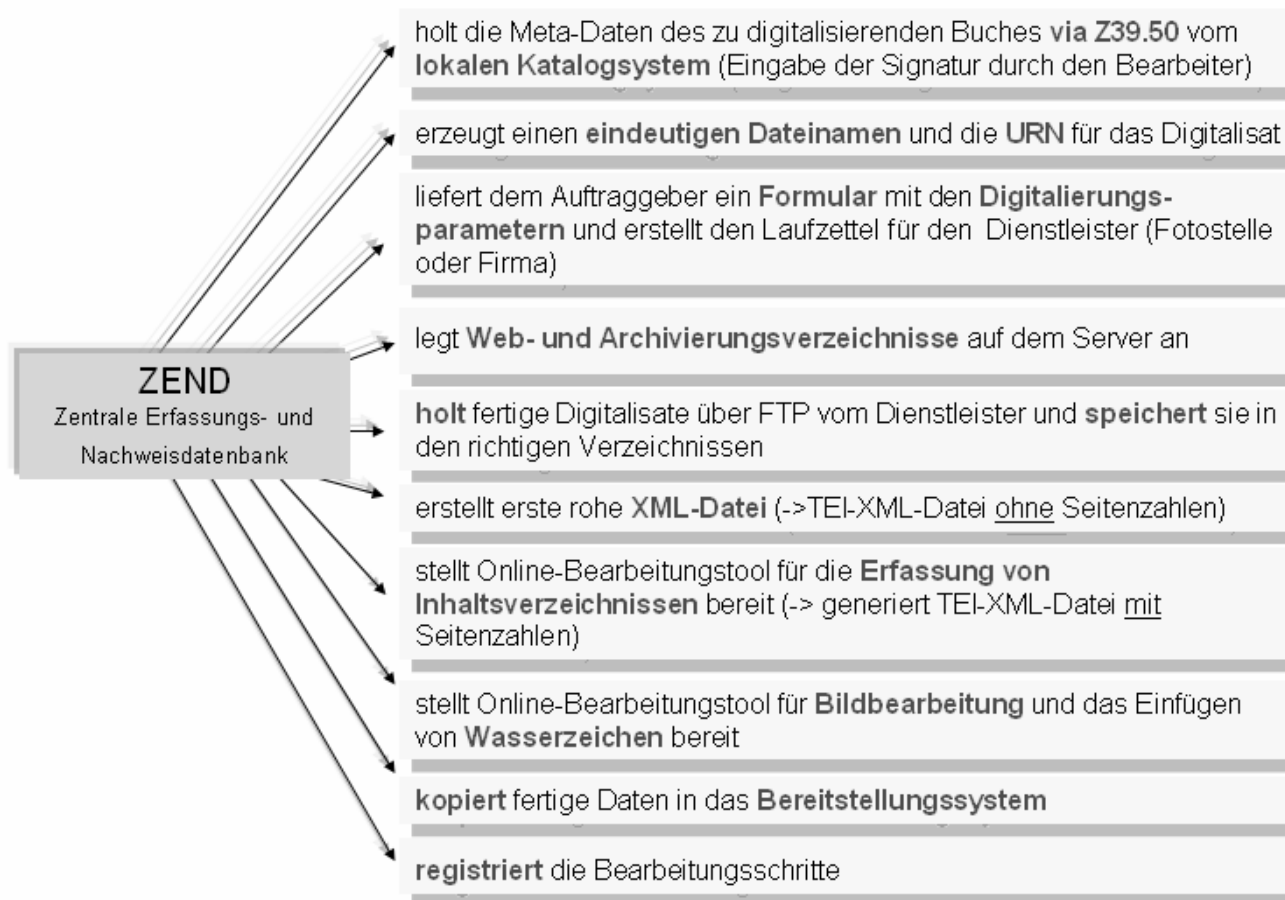
On the left side of the browser window, there is a table of contents (INHALT) listing the following issues:

- Nr. 1 vom 3. Januar
- Nr. 2 vom 4. Januar
- Nr. 3 vom 5. Januar
- Nr. 4 vom 6. Januar
- Nr. 5 vom 7. Januar
- Nr. 6 vom 8. Januar
- Nr. 7 vom 9. Januar
- Nr. 8 vom 11. Januar
- Nr. 9 vom 12. Januar
- Nr. 10 vom 13. Januar
- Nr. 11 vom 14. Januar
- Nr. 12 vom 15. Januar
- Nr. 13 vom 17. Januar
- Nr. 14 vom 18. Januar
- Nr. 15 vom 19. Januar
- Nr. 16 vom 20. Januar
- Nr. 17 vom 21. Januar
- Nr. 18 vom 22. Januar
- Nr. 19 vom 24. Januar

Beispiel *Coburger Zeitung*

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<TEI xmlns="http://www.tei-c.org/ns/1.0"
      xmlns:eb="http://sunsite.berkeley.edu/Ebind">
<teiHeader xml:id="bsb00001119" type="book">
  <fileDesc>
    <titleStmt>
      <title>Coburger Zeitung</title>
    </titleStmt>
    ...
  <sourceDesc>
    <imprint>
      <pubPlace>Coburg</pubPlace>
      <date>1881</date>
    </imprint>
  </body>
<div type="chapter">
<head>Nr. 1 vom 3. Januar</head>
  <eb:page>
    <eb:image entityref="bsb00001119_00001" seqno="1" nativeno="00" />
  </eb:page>
  <eb:page>
    <eb:image entityref="bsb00001119_00002" seqno="2" nativeno="00" />
  </eb:page>
```

3. Strukturfassung



4. Bereitstellung

- Automatische Erstellung von Präsentationsderivaten
 - GIF
 - JPG
 - PDF
- Automatische Bereitstellung im WWW
 - in den Digitalen Sammlungen
 - in den Fachportalen (ViFas)
 - in Themenportalen (z. B. Bayerische Landesbibliothek Online)
- Einbindung in die Meta-Suche der *Digitalen Sammlungen*
- Nachweis im Lokal- und Verbundsystem (OPAC)

5. Archivierung

- Archivierung nach Abschluß aller Arbeiten
- Qualitätskontrolle
- Übertragung der Daten zum Leibniz-Rechenzentrum
- Verwaltung, Speicherung auf Platten und Bändern in Robotersystemen
 - Client/Server-System
 - Ständiger Datentransfer
 - Rückholung per E-Mail möglich (Dokumentlieferung)
- File-System-Schnittstelle im Aufbau

Metadaten für die Langzeitarchivierung

- NBN als Persistent Identifier
- Bibliographische Daten: MAB und Dublin Core
- Strukturdaten: Bände, Kapitel, Volltext (XML)
- Technische Metadaten, z. B.
 - Mime-Typ, Größe, Checksumme, Farbprofile, ...
- Administrative Daten, z. B. Herkunft, Datumsangaben, History, Zuordnung
 - Entstehungsprozeß, Derivate

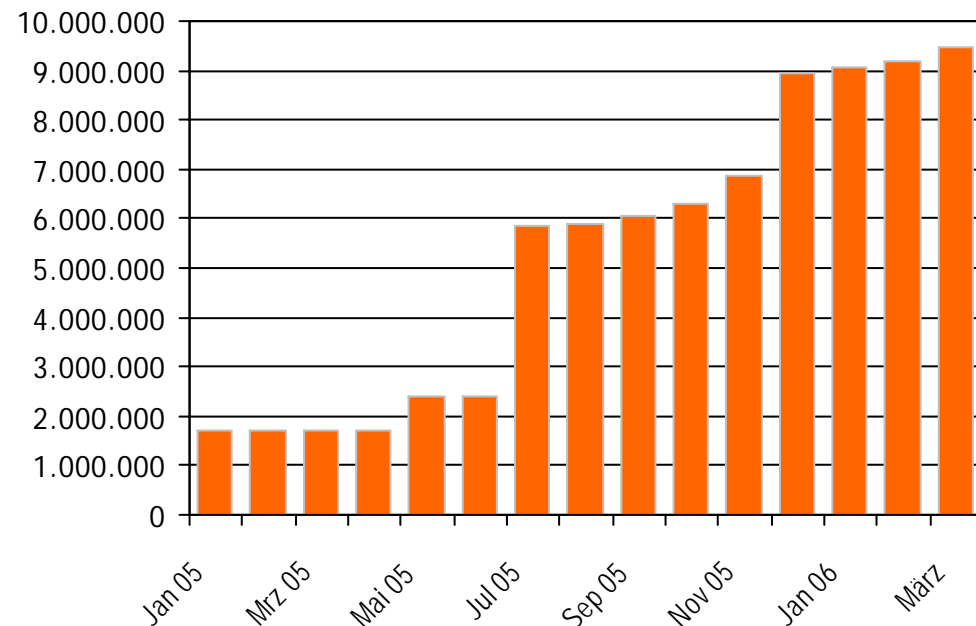
Lagerung der Daten



Datenbestand der Digitalen Bibliothek der BSB

im März 2006:

- mehr als 9 Mio. digitale Seiten
- = ~10.000 Einzelbände
- = ~120.000 Objekte
- = 8,9 TeraByte
- signifikante Steigerung seit Juni 2005 (Digitalisierungsstraße)
- Erwarteter Zuwachs: mindestens 10 TB p. A.



Pilotprojekt Langzeitarchivierung von Netzpublikationen

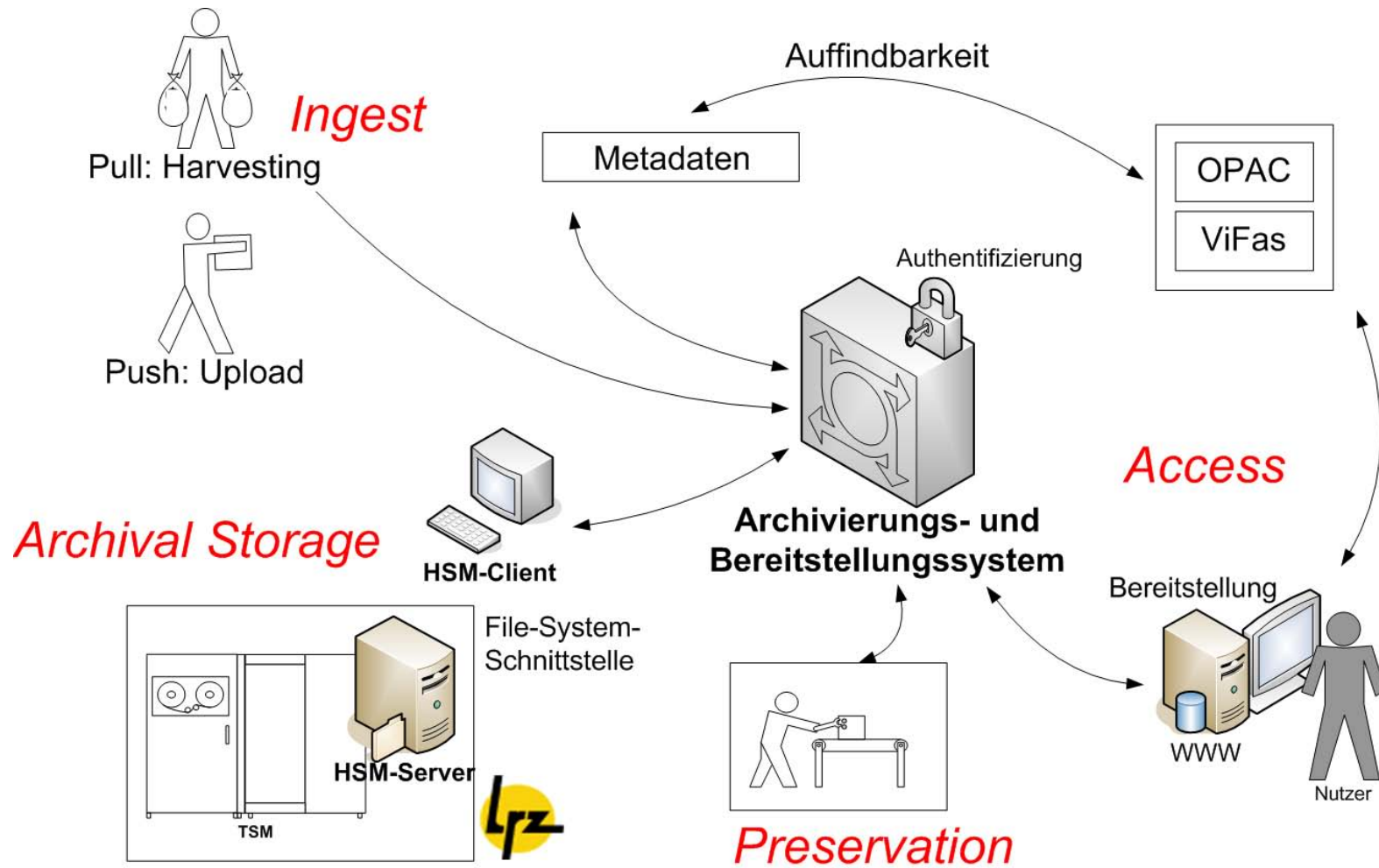
- Laufzeit: 2 Jahre (seit Juli 2005)
- Förderung durch die Deutsche Forschungsgemeinschaft
- Kooperation mit dem Leibniz-Rechenzentrum
- Ziele
 - Ausarbeitung von Kooperationsmodellen und Vereinbarungen für die Überlassung, Speicherung, Erschließung und Bereitstellung von elektronischer Information
 - Entwicklung von Workflows in Bibliothek und Rechenzentrum
 - Aufbau eines Bibliothekarischen Archivierungs- und Bereitstellungssystem (BABS)
 - Experimentelle Datensammlung und Erprobung von Methoden der Langzeitarchivierung mit einem heterogenen und komplexen Datenbestand

Pilotprojekt Langzeitarchivierung von Netzpublikationen

□ Material-Fokus

- Elektronische Veröffentlichungen verschiedener, vorwiegend regional angesiedelter, wissenschaftlicher Verlage
- Amtliche Veröffentlichungen von Landes- und Bundesbehörden
- Wissenschaftliche Informationsquellen in den Sondersammelgebieten der BSB (in Fachportalen erschlossen)
 - Geschichte
 - Osteuropa
 - Musikwissenschaften
 - Bavarica
- Retrodigitalisate der Digitalen Bibliothek

Das Bibliothekarische Archivierungs- und Bereitstellungssystem der BSB



Zu integrierende technische Infrastruktur

□ Bayerische Staatsbibliothek

- Zentrale Erfassungs- und Nachweisdatenbank (ZEND)
- Individuelle Präsentationslösungen (überwiegend Open Source)
z. B. MySQL, Cocoon
- DigiTool von Ex Libris
- Authentifizierungsmodule (HAN, Shibboleth, PDS)
- WinHTTRAck als Harvester

□ Leibniz-Rechenzentrum

- Tivoli Storage Manager von IBM mit Bandspeicher
- File-System-Schnittstelle über Client/Server-Lösung
- Datenleitung mit Gigabit-Geschwindigkeit

Langzeitarchivierung in der BSB Zusammenfassung

- Aufbau von Know-how durch Beschäftigung mit dem Thema Langzeitarchivierung seit 1999
- Leistungsfähige technische Infrastruktur, seit 2004 erfolgreich in Betrieb, wird derzeit erweitert
- Nationale Kooperation mit Partnerprojekten
- BSB als Partner für
 - Digitalisierung,
 - Hosting wissenschaftlicher Informationsangebote in kooperativen Projekten,
 - Langzeitarchivierung

Vielen Dank für Ihre Aufmerksamkeit

<http://www.bsb-muenchen.de>

<http://www.babs-muenchen.de>

Kontakt:

langzeitarchivierung@bsb-muenchen.de

