Klaus Kempf

Long Term Preservation: Needs and Activities at the Bavarian State Library (BSB)
Agenda

• BSB’s Institutional Profile

• Munich Digitization Center (MDZ)

• Current Responsibilities, Milestones, Activities and Organization of LTP

• Rosetta Project

• Future Prospects
BSB‘s institutional and collection profile

• Research library of worldwide importance
• Archive library of Bavaria
• Flagship of the Bavarian Library Network (BVB + more)
• Integrated Part of the „German virtual national library“ together with the Staatsbibliothek zu Berlin (Berlin State Library) and the Deutsche Nationalbibliothek (German National Library)
• Inventory: almost 10 million volumes with an annual increase of about 150,000 volumes, 52,500 current periodicals (print and electronic) + big collections of old manuscripts (no. 4 in the world), incunabula (no.1 in the world) and other valuable materials (maps, photographs, music collections)
Munich Digitization Center

- Founded as a national competence center for digitization and digital long-term preservation
- More than 100 projects since 1997, mostly third-party-funded
- Mass digitization since 2006
- Long-term preservation in cooperation with the Leibniz Supercomputing Center, Munich
- www.digital-collections.de
Munich Digitization Center

Main tasks:

• (Retro-)Digitization of printed material: (mass)-digitization projects, private public partnership with Google, digitization on demand, digitization for preservation reasons

• Data retrieval and e-publications: Bavarian Library Online, Virtual subject libraries

• Long-term preservation of the digital collections of BSB
Present Responsibilities in Long-Term Preservation (LTP)

- Digital objects from (mass-)digitization projects

- Born-digital objects from different sources:
  - Purchase/ licensing (negotiated by BSB at local, regional and national level)
  - Legal deposit (federal state level)
  - BSB‘s special subject collections
LTP at BSB: Three intersecting Spheres of Activity

- **Further development and consolidation of the required structures concerning:**
  - Profile and concepts
  - Organization
  - Technology

- **Continuous operation of a long-term preservation system**
  - BABS (Library Archiving and Access System)

- **Research and development**
Long-term Preservation Milestones

• **1999**: First project concerning removable media – testbed of a computer museum (Institute of Informatics, GFAF)

• **2004**: Start of the cooperation with the Leibniz Supercomputing Center (LRZ) Migration of CD-ROM-stored digitization data to the LRZ archiving system

• **2005**: Implementation of automatic archiving procedures during BABS1-project
  - Connecting the storage/archiving system of the LRZ and MDZ production line (ZEND software)
  - Daily online transfer of new data
  - Distributed storage of multiple copies

• **2006**: First complete migration of the archive hosted by LRZ during the relocation of LRZ from Munich City center to Munich Garching
Major Project Activities in long-term preservation (LTP)

**BABS I:**

**BABS II:**
Evaluation and improvement of the BABS-System in terms of trustworthiness and scalability (2009-2010)  

**nestor:**
Network of expertise for the long-term preservation in Germany (since 2003)

**Further studies:**
Organizational and business models for the long-term preservation of digitized materials et al.

**LOCKSS:**
Participation in a world-wide pilot project
How it is done: Architecture of BABS
Retro-Digitization: the costumized workflow tool ZEND

Munich Digitisation Centre
The Digital Library Department of the Bavarian State Library

Projects
Digitization On Demand
Preservation Purposes

Order
Image Capture
Delivery + Billing
Digital Documents
Outsourcing

Workflow Management
Title and Call Number
URN Generation
TOC Editor
Image Conversion
Indexing and Enrichment
Basic WWW Version
Internal Documentation
Long-term Preservation

The ZEND - Digital Asset Management System

Worldwide Users
Catalogue

URN

xEpicur / DNB

Long-term Preservation and Archival Storage
(at the Leibniz Computing Centre)

BSB Bayerische StaatsBibliothek
Information in erster Linie

"ALA 2010 Washington"
BSB and Google Books: Private Public Partnership – LTP
BSB and Google -books - workflow
LRZ: The Backbone of BSB‘s LTP strategy

• **Supercomputing Center of the Bavarian Academy of Sciences and Humanities**

• **Serviceprovider for**
  - general IT-services for the universities at Munich (more than 100,000 customers) and the Academy
  - powerful communications infrastructure called the Munich Scientific Network
  - backup and long-term preservation of large amounts of data
  - technical and scientific high performance Supercomputing Center for all german universities
BABS: Some Figures

- Total amount of long-term preserved data by end of May 2010:
  **217 Terabyte**
  **378 M. files**
  (with Google Books Project)

- Expected annual increase:
  with scanrobot production and Google: more than **100 TeraByte**
BABS2-Project: First Results

- Drafting of a Preservation Policy for BSB
- Inventory of in-house preservation needs
- Improved documentation of processes and workflows
- Developing organizational and cooperation models

- Concepts for securing authenticity and integrity of the digital objects
- Extented preservation metadata schema
- First experiences in Preservation Planning with PLANETS-Tool Plato
  -> From TIFF to JPEG 2000? (http://www.dlib.org/dlib/november09/kulovits/11kulovits.html)

- Evaluation based on the nestor catalogue of criteria for trusted digital repositories
- Intensive scalability testing of preservation infrastructure
Rosetta at the Bavarian State Library

A new Digital Preservation System at BSB for:

• Improving existing preservation workflows

• Implementing additional workflows

• Intensifying preservation planning activities
Why Rosetta?

- Comprehensive and specific functionality for all key areas of digital preservation:

- Flexible configuration, parameterisation and automation for different submission and ingest workflows / for internal and external stakeholders

- Digital Asset Management: checks for integrity / authenticity, format conversion, generation of metadata
Why Rosetta?

• Preservation planning and action: risk analysis and risk management, planning and implementation of preservation action (with V. 2.0)

• Support of standards (OAIS, METS, Dublin Core, PREMIS, OAI-PMH)

• Open Platform: APIs and software development kit support integration of additional software

• Easy exchange / export of Archival Information Package (AIP)
Rosetta: Integration into the existing software architecture

Legal and voluntary deposit; Licensed media
Possible Future Strategic Objectives

**General objectives:**

- BSB fulfills its duties as the archive library for Bavaria by preserving digital media from different sources (extension of Legal Deposit Act)
- BSB is willing to become a trusted partner in a cooperative solution for the preservation of the digital heritage of Germany (National Hosting Strategy)
- Further contribution to national and international research and development

**Next time objectives:**

- Digitization of the complete copyright-free holdings
- Further implementation of Preservation Planning and Action for different BSB holdings
- Offering digitization and long-term preservation services for third-party institutions (Bavarian regional libraries/ archives, university and research institutions etc.)
Future Prospects: Still a long way to go!

THANK YOU!

klaus.kempf@bsb-muenchen.de