Digitization at Bayerische Staatsbibliothek

markus.brantl@bsb-muenchen.de
Digitization of the whole BSB?

It’s not so easy: 8 million books, over 50,000 manuscripts ...

Perhaps in 10-15 years the most important collections!

Münchener Digitalisierungszentrum (MDZ)
ToC

- Who digitizes at BSB?

- The Munich Digitization Centre
  - Status of digitization activities
  - Profile
  - Production
  - Projects
  - Success

- Outlook
Who digitizes at BSB?

Since 1997
- Munich Digitization Centre - The library as electronic publisher
- Electronic document delivery department - „Subito“

Since 1998
- Picture Archive (potographic agency)

Since 2002
- Photographic Services
- The Users themselves
The Munich Digitization Centre - Introduction 1

- 1997 MDZ launched with funding by German Research Association (DFG)

- Today - part of the Digital Library Department at BSB, including
  - Subject gateways (portals)
  - Long term preservation of digital data
  - Overall staff of Digital Library: 4.5 persons

- Technical equipment: various servers, scanners etc.
The Munich Digitization Centre - Introduction 2

Tasks

- Grant acquisition
- Project management
- Production (scanning, creating metadata etc.)
- Evaluation of standards, tools and technics for image capture, text conversion, electronic document management and presentation of collections via WWW etc.
- Consultation and support for other library departments in house and outside
Current Profile

- Digital collection development
  - Along the lines of the DFG-supported special subject fields: History, Eastern Europe, Music
  - Content selection
    - Primary source material and information media for historians
    - Historic reference books
    - Selected items of cultural heritage

- Enhanced access
  - Added value through deep indexing

- Collaboration and partnership
  - Collaborative projects with scholars, academies (e.g. Regesta Imperii, emblem books)
Status of digitization activities

- 1999 Launch of the first digital collections

- 2003: ca. 4000 digital objects in 23 projects online - see http://mdz.bib-bvb.de

- Data online available:
  - ca. 2 GB text (XML) and
  - ca. 500,000 digitized images of book pages (manuscripts -> maps)
Production 1 – Stages in text indexing of a book → access

Each project defines new requirements in image capture, indexing and access!

- page images, ToC and extracted text (z.B. index)
  - search on extracted text and
  - view image

- Structured Full-text manually keyed text (99.98 % Correctness)

- Costs
Production 2 - Workflow

Enhancing Access

Metadata creation (database)
- Bibliographic record
- Structural description (ToC)
- Admin. Metadata

- Scanned images
- Technical metadata
- Text (range from extracted text to fulltext)

Delivery

Web-publishing Framework
- Document Management System
- RDBMS

OPAC
Excel

XML
Images

OPAC
Direct
access
by
local/
union
OPAC
WWW

Retrieval from storage, analysis, preparation, substitution of damaged books etc.
Production 3 - Image Capture

- Original or intermediate?
- Scanning from the original using
  - digital camera or planetary scanner
  - bookcradle
- Types of scanning requirements
  - 1 bit black/white, resolution 600 ppi – digital master: TIFF with ITU G 4 compression - for printed books 16th-21st century
  - 8 bit grayscale, resolution 300-400ppi – digital master: TIFF uncompressed - for text-only manuscripts, prints with engravings, photographs
  - 24 bit colour, resolution 300-400ppi – digital master: TIFF uncompressed with ICC-Profiles – for illuminated manuscripts, color illustrations, maps etc.
    - Color management system for color-true reproductions
Production 4 - Quality Assurance

- Set standards
  - MAB
  - TIFF
  - TEI, Ebind for text encoding
  - XML, ASCII
  - Character set: Unicode (e.g. use of formats)
- Control defined standards and quality at all production levels
  - Image capture
  - Text encoding
  - before final delivery to WWW
Production 5 – Access

Direct via Web interface of the DMS | local or library network catalogue
Projects 1 - Larger projects online

- Regesta Imperii (records of the documentary and historiographical sources of the Roman-German kings and emperors from 751-1519) (fulltext - images)
- Facsimile edition of imperial charters (751-1519) (extracted text – images))
- Protocols of the Reichstag (1867 - 1895) (extracted text - images)
- Protocols of the Bavarian parliament 1429-1669 and 1919-1933 (extracted text - images)
- Historical Atlas of Bavaria (extracted text – images, maps)
- Zeitschrift für Bayerische Landesgeschichte (extracted text, fulltext - images)
- Zedler-encyclopedia (extracted text - images)
- Allgemeine Deutsche Biographie – ADB (extracted text - images) and cumulated index with Neue Deutsche Biographie –NDB (extracted text)
- Emblem books of early modern times (current) (extracted text - images)
- Illustrations of incunabula printed in German-speaking countries (extracted text - images)
Projects 2 - examples

Our document server: http://mdz.bib-bvb.de

- Zedler - encyclopedia
- ADB/NDB – cumulated index of the most important German biographical dictionaries and images of ADB
- Incunabula illustrations
- Emblem books – zoom server technology
- The Hoffmann photographies
Projects 3 - in preparation

- dMGH – the digital edition of „Monumenta Germaniae Historica“ (in collaboration with the MGH)
  - New editions will be digitized with a „moving wall“ of 5 years

- Protocols of the German Reichstag 1919-1933

- 100(0) key documents of Russian-Soviet history (1917-1991) (in collaboration with the University of Erlangen and the Russian Academy of sciences) – borne digital no print publication
Ein Gemeinschaftsprojekt von

Münchener Digitalisierungszentrum (MDZ)
Success – usage of the collections in 2002

Grafik: MDZ 2002-7.614.730 Hits von ca. 83.993 hosts (+ MDZ2:816.915 Hits nicht beinhaltet)
Outlook

- From project to program: towards a stronger integration of the digitization in the library processes

- BSB as Content Provider for history, Eastern Europe, music

- Enlargement, cross-linking and expansion of existing digital collections

- New methods of automatisation in digitization and deep indexing → increaseamnt of productivity
Thank you for your attention!

markus.brantl@bsb-muenchen.de
Production - planetary scanners

Planetary (or book-) scanners: color, grayscale, bitonal
Production - Bookcradles
Production - Color management system – calibration