BSB Bayerische StaatsBibliothek

Information in erster Linie

Münchener DigitalisierungsZentrum

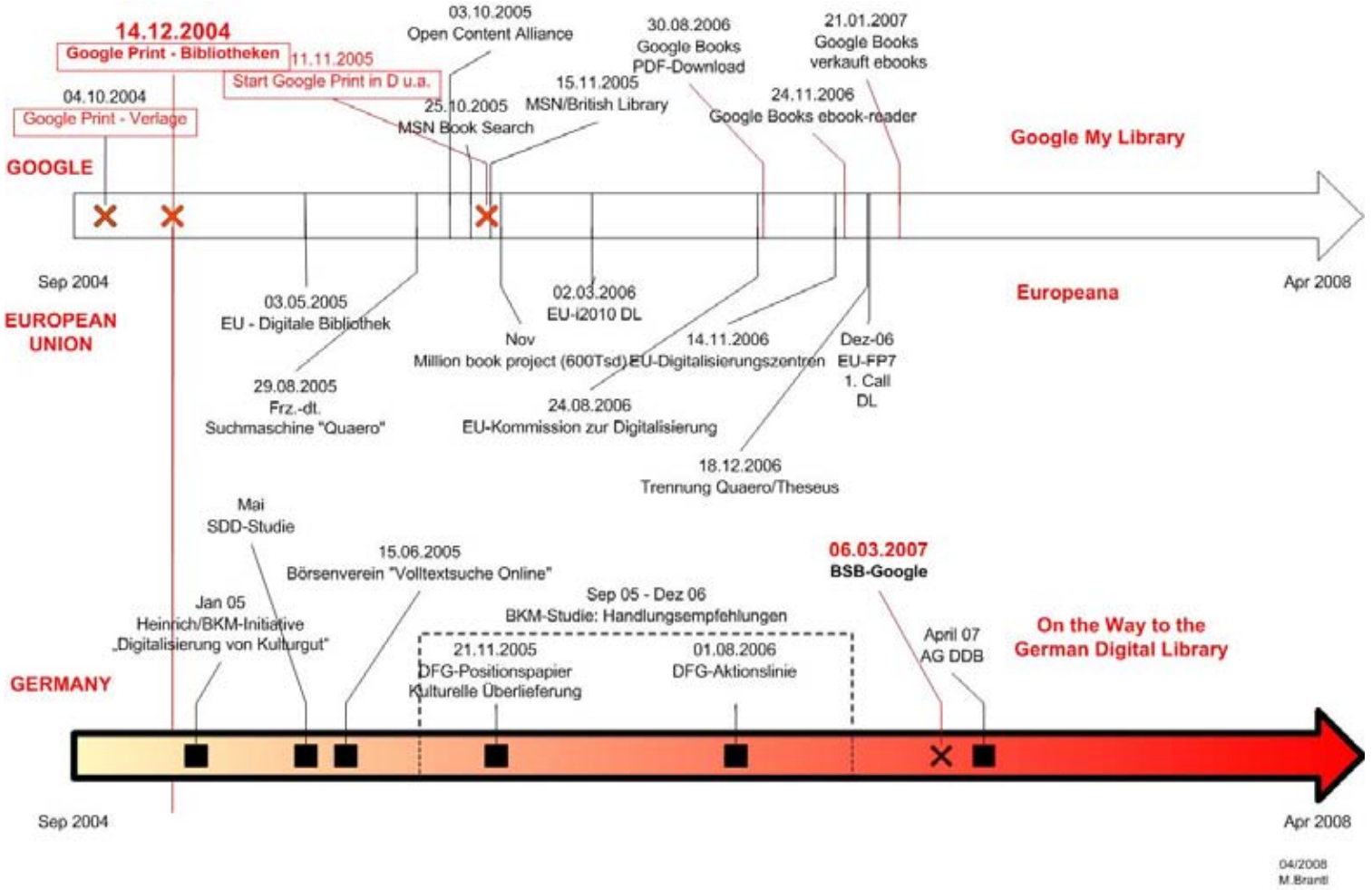
Digitale Bibliothek

Mass Digitisation and Long-term Preservation – Processes and Production at Munich Digitisation Centre

Dr. Markus Brantl

Copyright Bayerische Staatsbibliothek

Who invented Mass Digitisation?



Copyright Bayerische Staatsbibliothek

Agenda

1. The Bavarian State Library (BSB) and the Munich Digitisation Centre (MDZ)
2. BSB Digitisation Strategy and the Challenges for Mass Digitisation and Long-Term Preservation
3. Processes and Production
4. Projects, examples
5. Outlook

Copyright Bayerische Staatsbibliothek

Agenda

1. The Bavarian State Library (BSB) and the Munich Digitisation Centre (MDZ)
2. BSB Digitisation Strategy and the Challenges for Mass Digitisation and Long-Term Preservation
3. Processes and Production
4. Projects, Examples
5. Outlook

Copyright Bayerische Staatsbibliothek

The Free State of Bavaria and Munich

8 towns > 100.000 Inhabitants

Capital city Munich (1,5 million)

Population: 12,5 million

7 Counties

245.000 university students

Copyright Bayerische Staatsbibliothek

Bavarian State Library: Some Key Facts & Figures

- Founded in 1558 by duke Albrecht V.
- 680 employees
- 43 million euro/annual budget
- Inventory ~10 million volumes with an annual increase ~ 150.000 volumes
- ~1 million visitors in the reading room
- ...

450 Jahre BSB Bayerische StaatsBibliothek

Copyright Bayerische Staatsbibliothek

Major Functions of BSB

- International research library of worldwide importance
- Central regional and archival library of the Free State of Bavaria
- Head of the Bavarian Library Network (BVB) and Leader of the Bavarian Library Consortium
- One of three partner of the „German Virtual National Library“ (together with Staatsbibliothek zu Berlin and Die Deutsche Nationalbibliothek in Frankfurt/Leipzig)

Copyright Bayerische Staatsbibliothek

BSB as a Gateway to Cultural Heritage Resources

- Collections: Timeframe from 8th to 21th century
- Manuscripts and Rare Books Collection of world importance
  - 89,000 Manuscripts (number 4 worldwide)
  - 19,000 Incunabula (top position in the world)
  - 130,000 16th century prints (largest collection in Germany)
  - Illuminated Reichenau Manuscripts part of the UNESCO Memory of the World since 2004

Copyright Bayerische Staatsbibliothek

The Munich Digitisation Centre (MDZ)

- **1997** founded through the financing of the German Research Association as one of two national digitisation centres
- **2003** integrated as division in the department of Acquisition, Collection Development and Cataloguing

MDZ today

- **National competence center** for digitisation and long-term preservation
- **Central innovation und digital production** unit of BSB

Copyright Bayerische Staatsbibliothek

MDZ Tasks within the Framework of „The Digital Library“

- Research & Development
- Production
- Consulting (e.g. Workflows)
- Service Provider (e.g. scanning technology, Open-Source-Software for digitisation)

in

- (Retro-)Digitisation
- Long-term Preservation
- Subject Specific Portals

Copyright Bayerische Staatsbibliothek

MDZ – Some Facts

- Digitisation expertise in materials from 8th to 21th century

- ~32 M. files successfully processed

= ~ 10 M. pages – 27,000 books

= ~ 60 Terabyte of digital data stored and long-term preserved in cooperation with Leibniz Supercomputing Centre (Munich-Garching)

= the largest operating long-term preservation archive in Germany

- Complete production line from capture until long-term preservation based on own-developed Software ZEND

- Comprehensive Hardwareequipment, e.g. for the processing of the Google Books Search image copies

Copyright Bayerische Staatsbibliothek

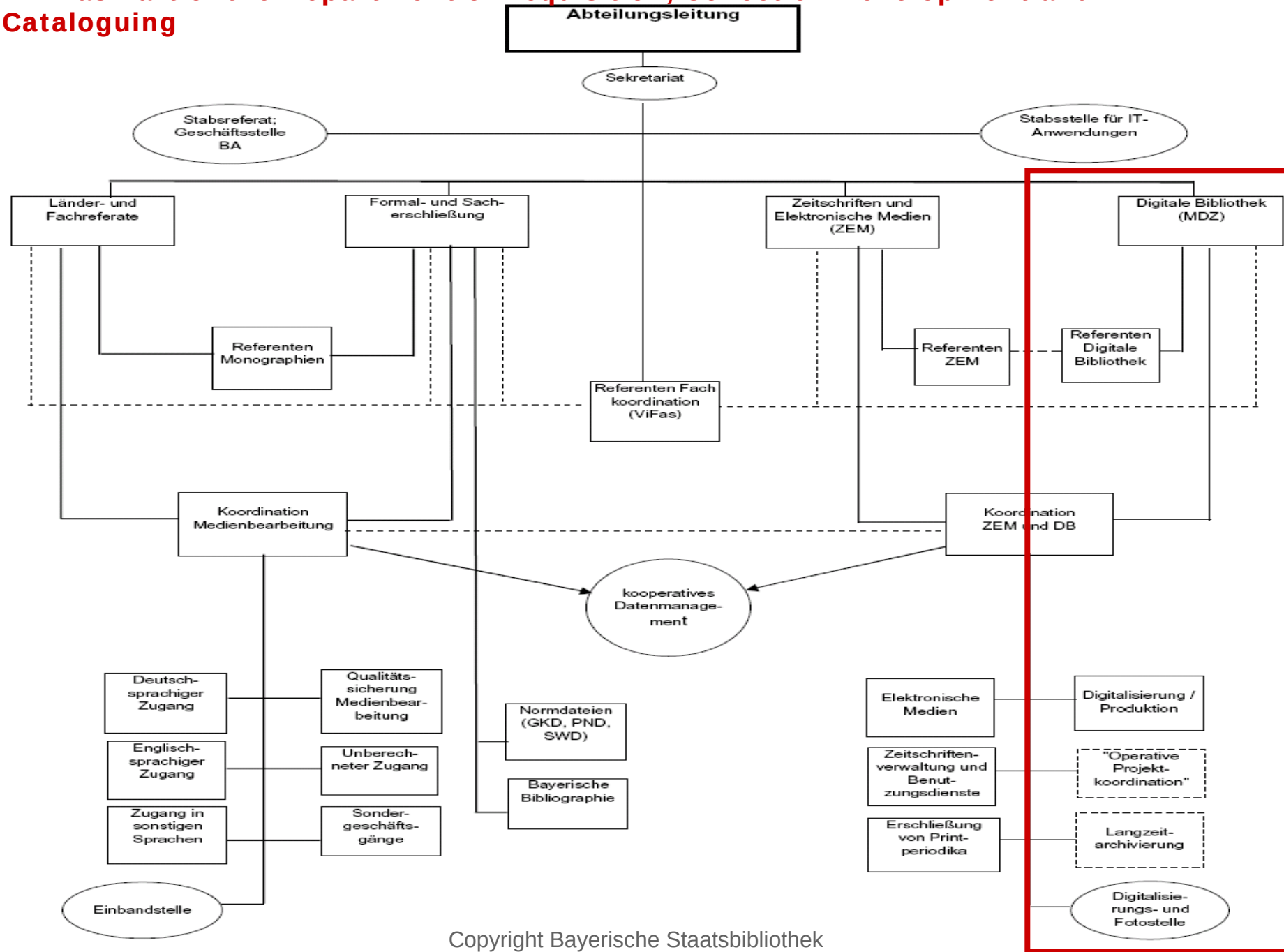
Examples of running R&D Projects at MDZ

- Long-term Preservation
  - Development of organisational and business models for the long-term preservation of digital objects in Germany (German Research Foundation)
  - BABS II – next step in deployment of practical long-term preservation (German Research Foundation)
  - ...

- Digitisation
  - **IMPr**oving **Ac**cess to **T**ext – Innovation in OCR
  - Development in new book scanning devices for fragile books
    - Automated book scanners for 16th century books – development partnership with theTreventus company
    - Book cradles
  - 3D-Animation of books via WWW
  - ...

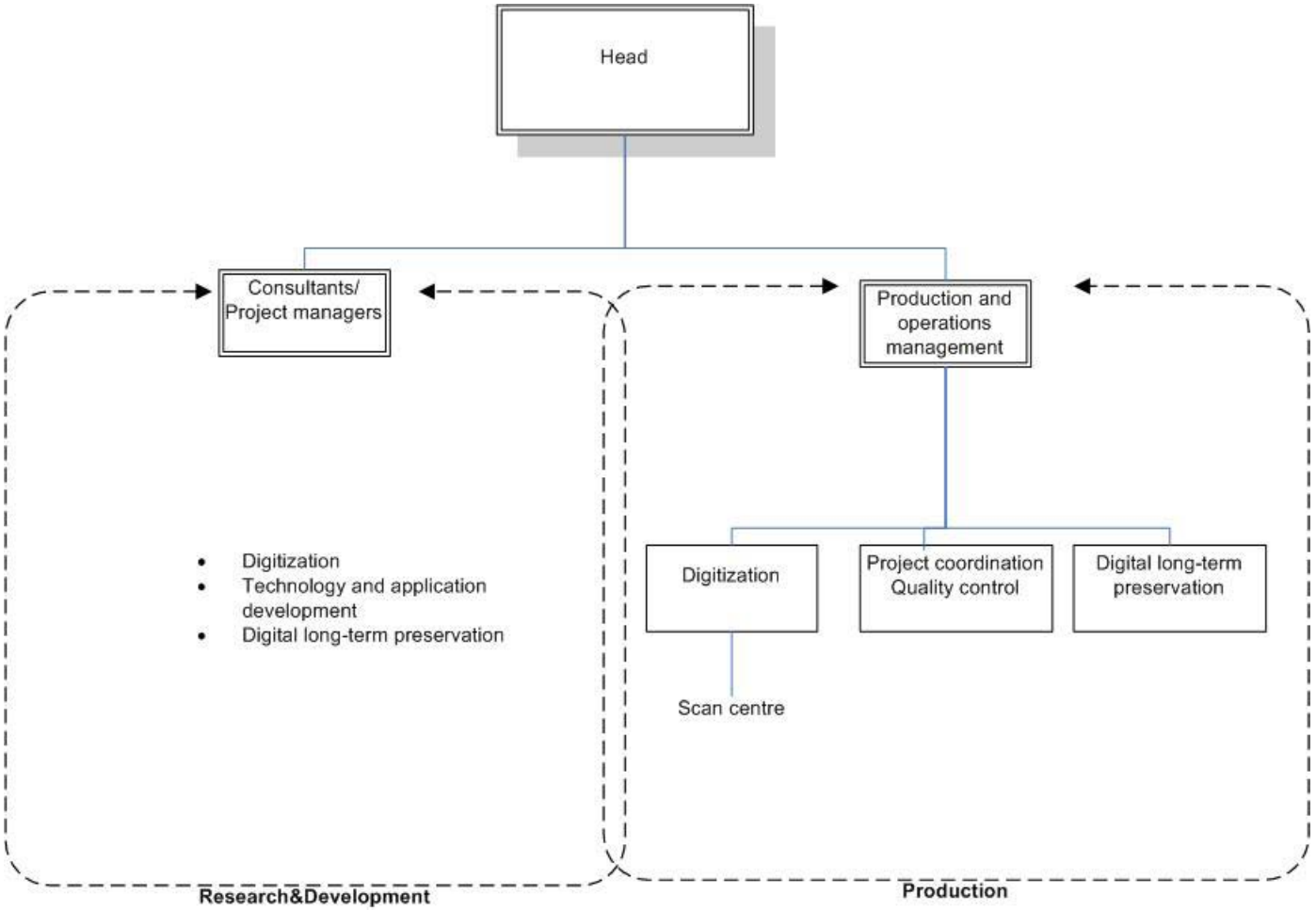
Copyright Bayerische Staatsbibliothek

MDZ as Part of the Department of Acquisition, Collection Development and Cataloguing



Copyright Bayerische Staatsbibliothek

MDZ Today's Organisation Unit in Detail



Copyright Bayerische Staatsbibliothek

MDZ Staff & Financing

- 5 fulltime and 1 part-time established posts
- 70 full- and part-time temporary employees (44,5 FTE) financed by third-party grants from:
  - European Union
  - German Research Association (DFG)
  - The Free State of Bavaria/Ministery of Science

Copyright Bayerische Staatsbibliothek

www.munich-digitisation-centre.de

http://www.digitale-sammlungen.de/index.html?c=startseite&l=en

Daily Status

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

BSB Bayerische Staatsbibliothek

- Main Page
- Digital Collections
- Digitisation
- Long-term Preservation
- Projects
- Service
- Disclaimer

↗ Deutsch

Scan Robot Days
June 18-20 2008

Welcome to the Munich Digitisation Centre

The Munich Digitisation Centre (MDZ) handles the digitisation and online publication of the cultural heritage preserved by the Bavarian State Library and by other institutions. It provides one of the largest and fastest growing digital collections in Germany, now comprising 27,786 titles available online. Access is free of charge!

The digitisation policy reflects the traditional special collection fields of the library: History, Classical Antiquity, Eastern Europe, Musicology. It comprises manuscripts, early prints, modern books, maps and photographic collections as well as journals and newspapers.

The MDZ frames the *Digital Library department* of the Bavarian State Library. It provides the ↗Digital Collections and is also responsible for the creation of subject portals and for the ↗long-term preservation of all the library's digital holdings.

W3C XHTML 1.0 W3C CSS

Astronomic figures (Call Nr. Cod.icon. 181, Augsburg 1555) from the digitisation project CodIcon Online - From the Digital Collections of the Bavarian State Library

- Digital Collections

Currently **27,786 Titles** online [23.06.2008]

↗ Brief Overview
↗ Extended Overview

Search

RSS

MDZ-News

The BSB elected 'Library of the year 2008'!
The BSB was elected 'Library of the year 2008' today by the ↗The German Library Association. Among other facts the statement of the jury pointed out the developement of the Munich Digitisation Centre to a service and technology provider for libraries as well as the engagement of the library in long-term preservation together with the Leibniz Supercomputing Centre. [18.06.2008]

Update: The programme of the presentation session of the Scan Robot Days on June 19 2008 is complete.
↗ to the programme
[13.06.2008]

↗ Other News
↗ MDZ-News via RSS

Copyright Bayerische Staatsbibliothek

Agenda

1. The Bavarian State Library (BSB) and the Munich Digitisation Centre (MDZ)
2. **BSB Digitisation Strategy and the Challenges for Mass Digitisation and Long-Term Preservation**
3. Processes and production
4. Projects - examples
5. Outlook

Copyright Bayerische Staatsbibliothek

Digitisation Strategy

Objective of retrodigital collection development at BSB:

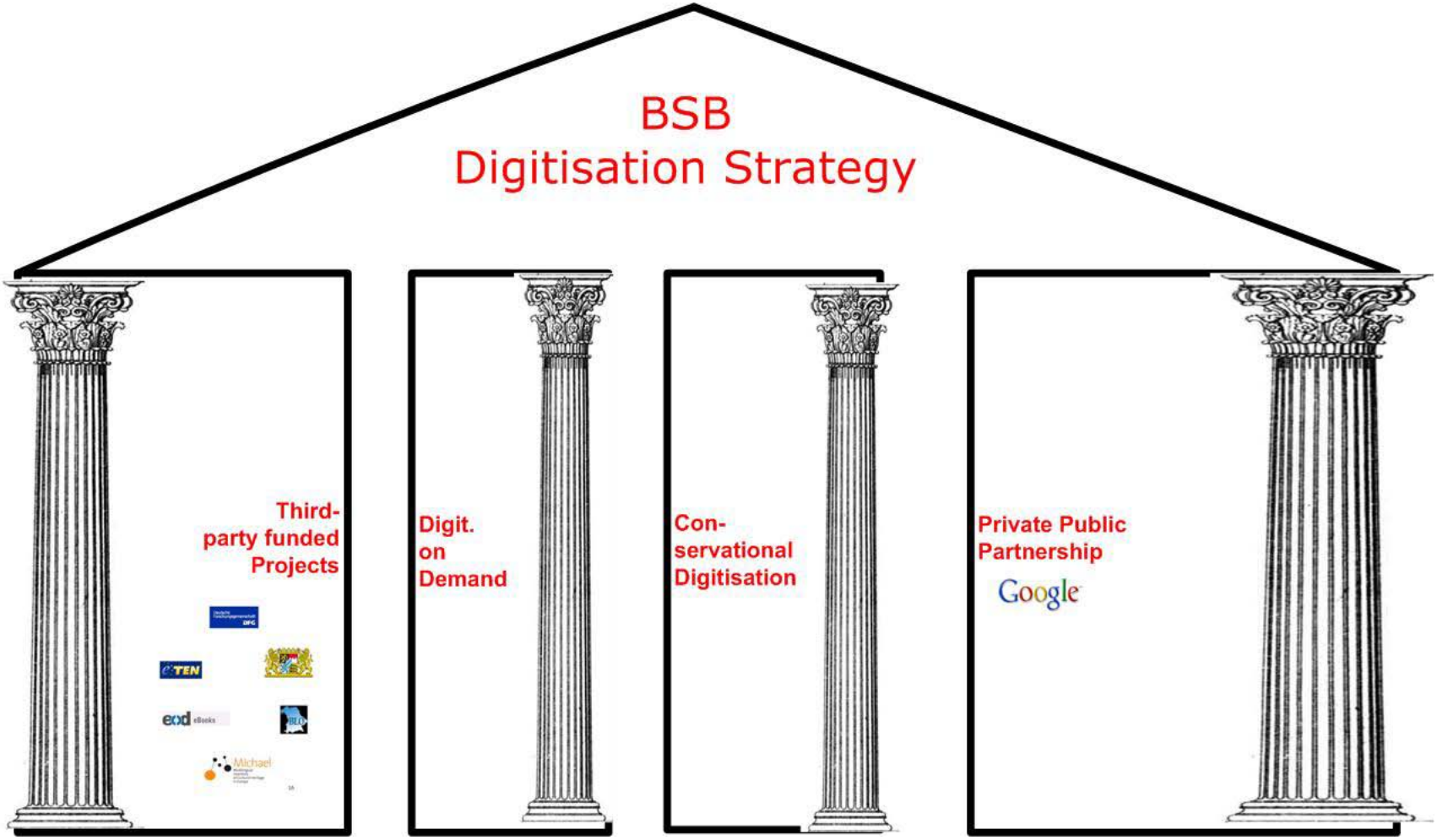
to digitize and make accessibly

– free of charge –

all copyright-free library holdings

Copyright Bayerische Staatsbibliothek

The Four Columns of



Copyright Bayerische Staatsbibliothek

Digitisation Strategy of the BSB

Third-party funds, Dig. On Demand, Conserv. Dig.	8th-16th century Manuscripts, incunabula, special collections
Public-Private Partnership with(Google)	17-19th century
Third-party funds	20-21st century (new programmes in the context of German Digital Library approach)

Copyright Bayerische Staatsbibliothek

Münchener DigitalisierungsZentrum
Digitale Bibliothek

What is Mass Digitisation?

Production of
more than a million
pages?
volumes ?

„Definition“ today (in Germany)

production of more than a million pages within a limited time

and tomorrow ...?

BSB Bayerische StaatsBibliothek
Information in erster Linie

Copyright Bayerische Staatsbibliothek

What have we done before Mass Digitisation?

“Boutique”-digitisation

- Since 1997
- Small – medium sized projects
- Explorative projects – learning and experimenting
- Selective picking of special objects under marketing aspects
- Deeper indexing of objects
- Manual work

Mass digitisation

- Started in 2006 with pilot
- Large projetcs with 1 million pages
- Using technology and automation whenever possible
- No selection
- Dynamic indexing – starting at low level; deeper indexing when technically possible (e.g. Blackletter script)

Copyright Bayerische Staatsbibliothek

Mass Digitisation: Plans for Quantities?

German Research Foundation (DFG)
target for digitisation of 16th.-18th. books

~1 M. titles =

~250 M. pages

Duration: Next 5-? years

Copyright Bayerische Staatsbibliothek

Münchener DigitalisierungsZentrum
Digitale Bibliothek

Current Mass Digitisation Projects at MDZ

Project	Volumes	Date	Pages
VD16digital -1 Deutsche Forschungsgemeinschaft DFG	~4.300	2006 -2008	~ 1 M.
VD16digital -2 Deutsche Forschungsgemeinschaft DFG	~37.000	2007-2009	~ 7.5 M.
Inkunabeln Deutsche Forschungsgemeinschaft DFG	~ 9.000	2008-2012	~ 1,6 M.
Private Public Partnership mit Google	> 1.2 M.	2008-	> 300 M.

3 ScanRobots in the MDZ-Scancentre

Linux-Cluster of the MDZ for Processing the Google book copy

Copyright Bayerische Staatsbibliothek

Mass Digitisation: Main Challenges

- Logistics: book pulling, transport, book-loan charging, quality management, return transport, book putting …
- Integration of metadata
- Storage and preservation – approx. 300,000 images per day
- Automated processing, e.g. automated book scanning devices
- …

Copyright Bayerische Staatsbibliothek

Mass Digitisation and Long-term Preservation

- MDZ data production
  - 1997-2004 – over 3.000 CD-ROMs
  - 2004 migrated from CD-ROMs to tapes at Leibniz-Computing Centre
  - 2005 Start of the newly built Scan-Centre ➔ fast growing archive

- New challenges: ScanRobots and Googles output – more and more Terabytes

Copyright Bayerische Staatsbibliothek

Agenda

1. The Bavarian State Library (BSB) and the Munich Digitisation Centre (MDZ)
2. BSB Digitisation Strategy and the Challenges for Mass Digitisation and Long-Term Preservation
3. **Processes and Production**
4. Projects
5. Outlook

a) Processes and Production Cycle

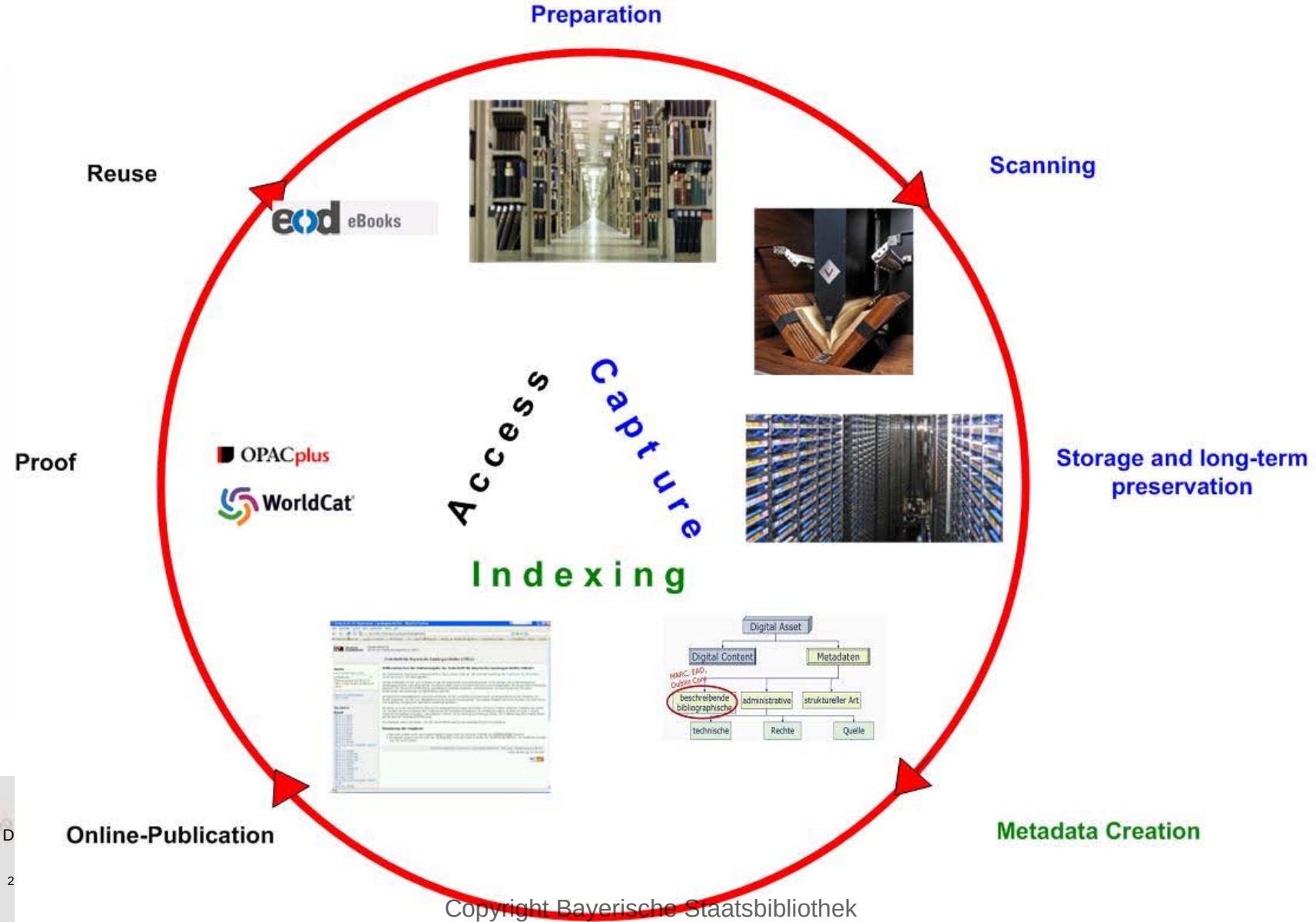
- Preparation
- Handling of Originals
- Scanning
- Storage and Long-term Preservation
- Metadata Creation
- Online Publication and Data Management
- Proof
- Reuse

b) ZEND Production Software

c) Example of a Production Workflow with ZEND

Copyright Bayerische Staatsbibliothek

Processes and Production Cycle



Copyright Bayerische Staatsbibliothek

Preparation

- Selection of all relevant titles for digitisation project from the collections
- Block the selected items for other forms of use (lending order e.g.)
- Printing of order-forms for collecting the relevant items from the shelves
- Checking the scanability of the selected items under conservational aspects
- Register the reason for non-scanability for every respective item
  - Correction of the catalogue if necessary
- Printing scanning order forms with Digital Identifier
  - Check if the books for scanning fit to the metadata of the selected title
- Delivery of the books of one charge together with scanning order forms to the Scan Centre

... Scanning ...

- Take the books back from the Scan Centre and check the completeness and actual condition of the books
- Putting the books in the shelves
- Unblocking lending orders

...

Copyright Bayerische Staatsbibliothek

Agenda

1. The Bavarian State Library (BSB) and the Munich Digitisation Centre (MDZ)
2. BSB Digitisation Strategy and the Challenges for Mass Digitisation and Long-Term Preservation
3. **Processes and Production**
4. Projects
5. Outlook

a) Production Cycle and Subprocesses
   - Preparation
   - **Handling of Originals**
   - Scanning
   - Storage and Long-term Preservation
   - Metadata Creation
   - Online Publication and Data Management
   - Proof
   - Reuse

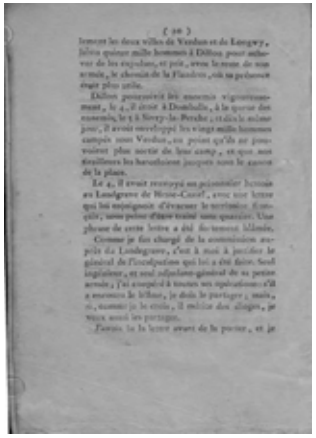
b) ZEND Production Software

c) Example of a Production Workflow with ZEND

Copyright Bayerische Staatsbibliothek

Handling of Different Material Types

- BSBs case: Wide range of materials from 8th-21th century
  - Manuscripts
  - Incunabulas
  - Early and rare prints
  - Historical Maps
  - Photographic material (slights; glass plates etc.)
  - Monographs
  - Newspapers
  - Magazines
  - ...



Copyright Bayerische Staatsbibliothek

Handling: Central Risks

- Damages by
  - Manually treatment of the originals
  - Room climate, e.g. during transport and reproduction
  - Light

Copyright Bayerische Staatsbibliothek

Handling: Damages by Opening Angle of 180°

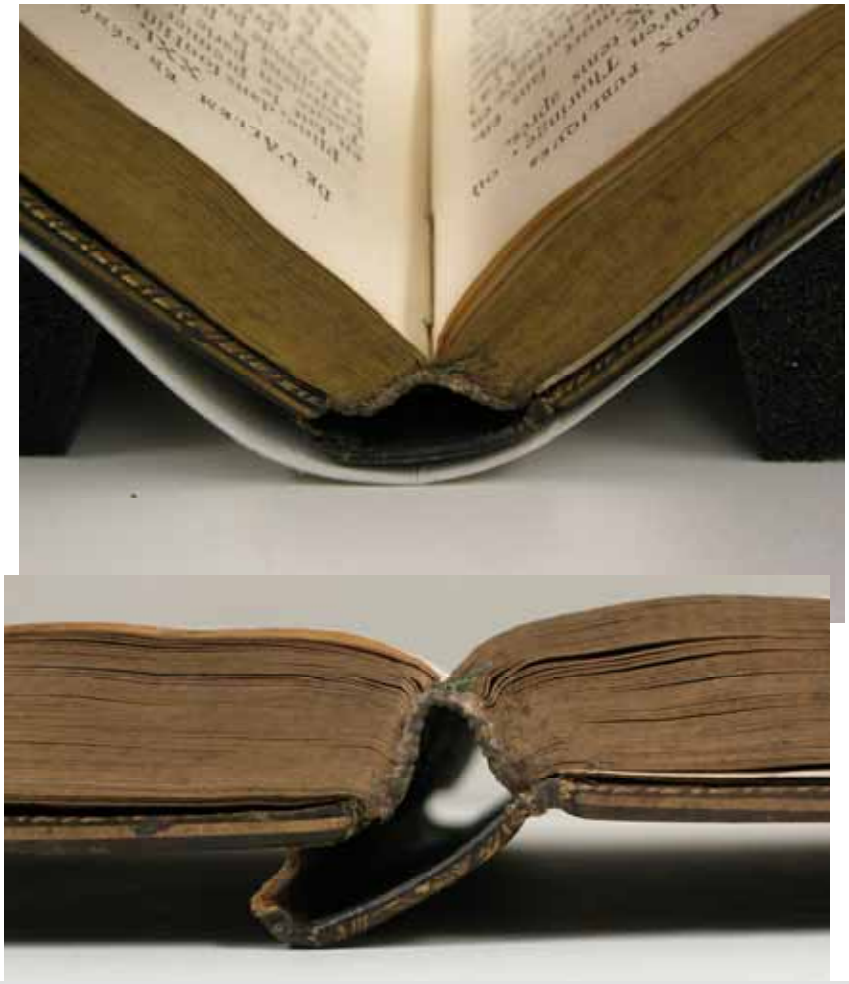
120° opening angle

- spine streched, but fold still okay

The same book:

180° opening angle

- fold broken



Copyright Bayerische Staatsbibliothek

Books Handling during Reproduction

BSBs Goal: Reduction of central risk factors of damage

- No fixation during re-production under glass-plate
  - ➔ **Deformation** of books body
- Close cooperation with the Institute for Book and Manuscript Restoration at BSB
  - Cooperation in the selecting process of scanning devices
  - Obligartory use of
  - Cool ligthing
  - Of different bookcradles for gentle book handling
  - ➔ Disadvantage: Higher set-up time

Copyright Bayerische Staatsbibliothek

Agenda

1. The Bavarian State Library (BSB) and the Munich Digitisation Centre (MDZ)
2. BSB Digitisation Strategy and the Challenges for Mass Digitisation and Long-Term Preservation
3. **Processes and Production**
4. Projects
5. Outlook

a) Processes and Production Cycle
   - Preparation
   - Handling of Originals
   - **Scanning**
   - Storage and Long-term Preservation
   - Metadata Creation
   - Online Publication and Data Management
   - Proof
   - Reuse

b) ZEND Production Software

c) Example of a Production Workflow with ZEND

Copyright Bayerische Staatsbibliothek

The MDZ Scan-Centre

- Especially for books and special collection materials before 1800
  (Modern books in outsourcing)
- 2005 developed from reorganisation of the anlaogue reproduction services

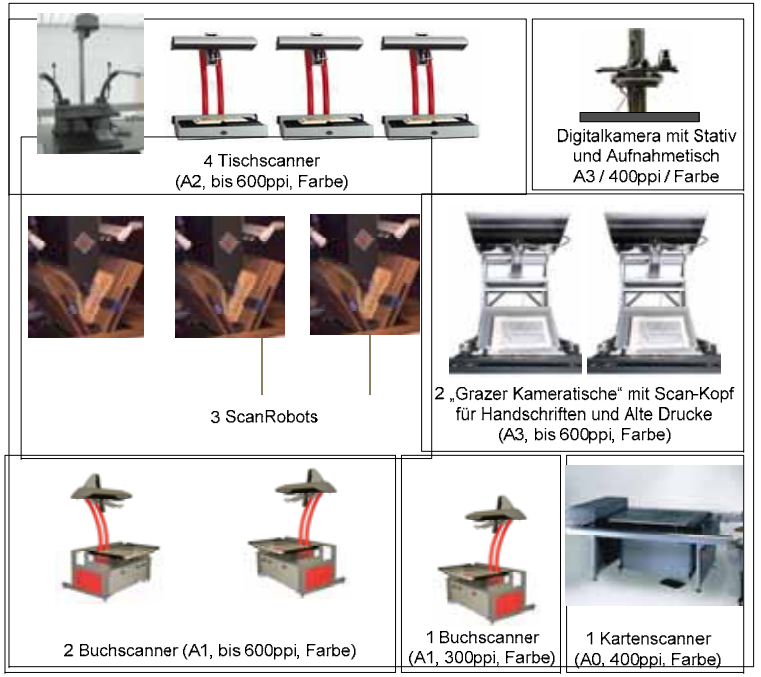
until2004

Since 2005

Copyright Bayerische Staatsbibliothek

Scan Facilities

- 14 bookscanners
  - 3 automated bookscanners, „ScanRobots“
  - 2 special bookscanners for rare manuscripts and treasure holdings - „Grazer camera table“
  - Formats up to ~0,85 * 1,18 meters (DIN A0)
  - High resolution up to 600 ppi



Copyright Bayerische Staatsbibliothek

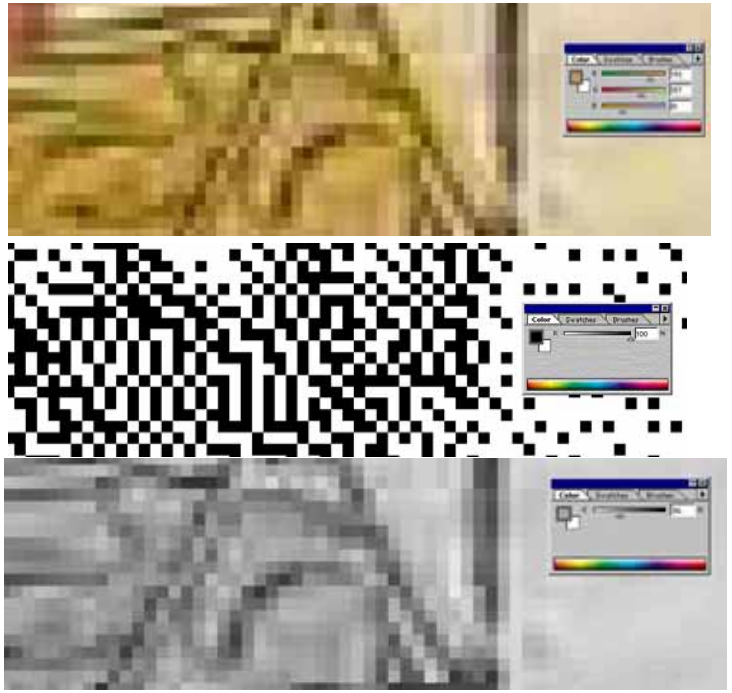
Image Scanning: Our Philosophy

- „Do it right the first time“
  - scanning at the best possible quality
  - Reuse mulitple
    - Facsimile, reprint
    - WWW
    - Digitisation on Demand
    - ...
- Images
  - Digital master files: TIFF
  - Derivates for WWW-Presentation: JPEG and PDF
- Extensive automation of workflows

Copyright Bayerische Staatsbibliothek

Image Quality and Scanning Parameters

Detail pictures (same image enlarged) In 24 Bit, 1 Bit, 8 Bit



High resolution digitisation in relation to the original size of

- Manuscripts, early printed books, historical Maps, special materials
  - colour (24 Bit)
  - 400 up to 600 ppi
  - Digital Master: TIFF uncompressed
  - Use of
    - Colour Management Software (ICC-Profiles)
    - Size – and Colour Targets
- Modern prints of 19th and 20th centuries in dependency of the original
  - Text only: black and white (1Bit)
  - Text/images: 1 Bit and grayscale (8 Bit) and/or colour
  - 300 up to 600 ppi (600 ppi only 1 Bit and TIFF-G4 compressed)

➔ **Image storage size between 100 Kilobyte up to 800 Megabyte per image**

Copyright Bayerische Staatsbibliothek

Scanning Output: Examples

Copyright Bayerische Staatsbibliothek

Agenda

1. The Bavarian State Library (BSB) and the Munich Digitisation Centre (MDZ)
2. BSB Digitisation Strategy and the Challenges for Mass Digitisation and Long-Term Preservation
3. **Processes and Production**
4. Projects
5. Outlook

a) Production Cycle and Subprocesses
   - Preparation
   - Handling of Originals
   - Scanning
   - **Storage and Long-term Preservation**
   - Metadata Creation
   - Online Publication and Data Management
   - Proof
   - Reuse

b) ZEND Production Software

c) Example of a Production Workflow with ZEND

Copyright Bayerische Staatsbibliothek

Digital Long-Term Preservation

- Still in experimental phase, no allround solution
- Different approaches
  - Migration
  - Emulation
  - Computer museum
- In cooperation on national and international level
  - nestor - the german Network of expertise in Digital long-term preservation
  - Institute of Informatics - University of the German Federal Armed Forces
  - European Union
  - USA
- MDZ focuses on
  - Digital long-term perservation in practice with the Leibniz Computing Centre
  - Trusted Repositories Certification for long-term archives

BSB Bayerische StaatsBibliothek
Information in erster Linie

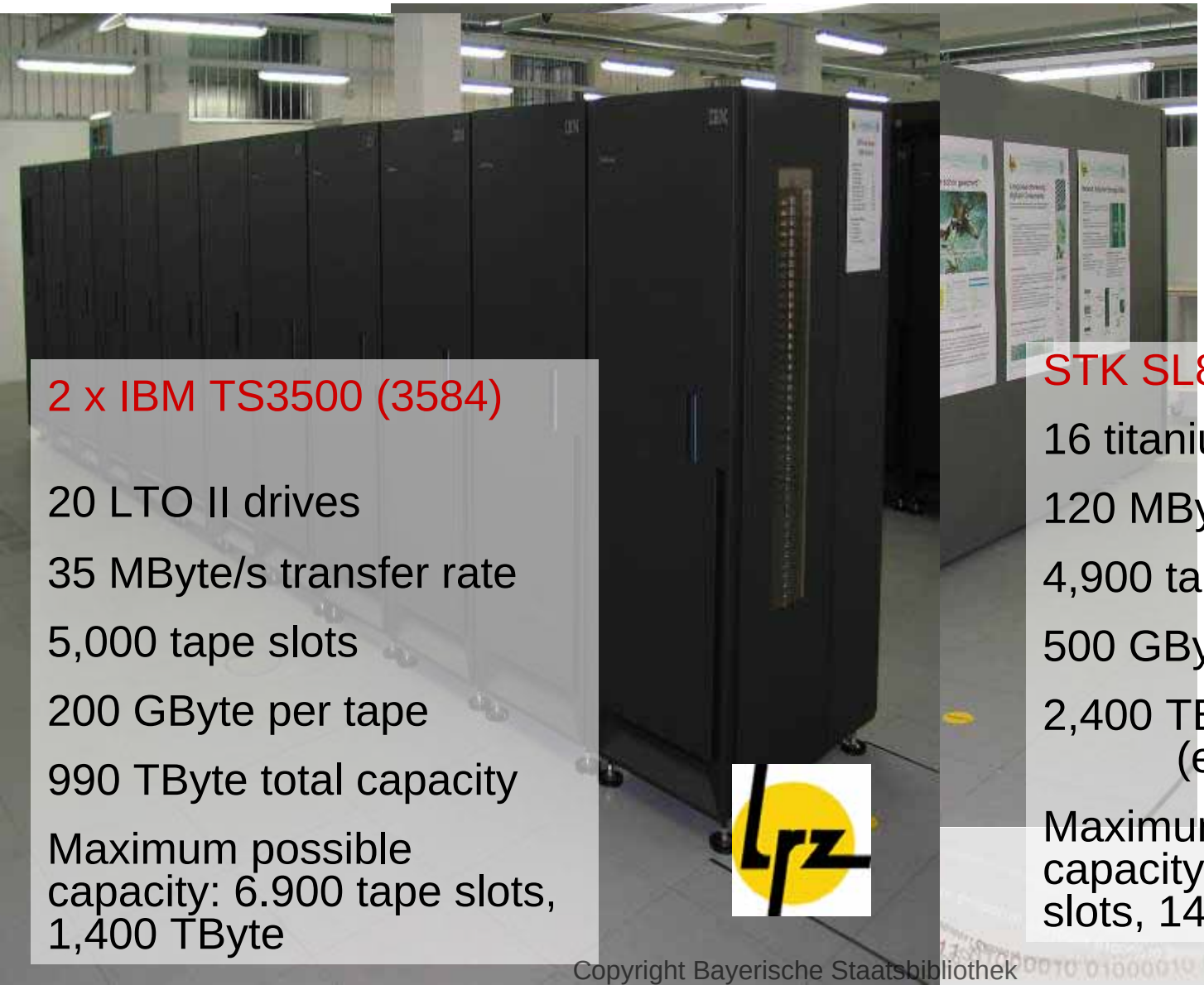
Copyright Bayerische Staatsbibliothek

MDZ:Milestones in long-term preservation

- 1999: First project concerning removable media – testbed of a computer museum (Institute of Informatics, GFAF)
- 2004: 1. Migration of CD-ROM stored digitisation data to a large-scale archiving system of the Leibniz Supercomputing Centre (LRZ) in Munich (IBM Tivoli Storage Manager)
- 2005: automatical storage procedures are implemented, connecting the storage/archiving system of the LRZ and MDZ-production infrastructure (ZEND-Software)
  - Daily online transfer of new data
  - Secure storage of multiple copies
  - Long-term preservation
  - Rapid restore functions

Copyright Bayerische Staatsbibliothek

Long-term Preservation: The Archival System of the LRZ



2 x IBM TS3500 (3584)

20 LTO II drives

35 MByte/s transfer rate

5,000 tape slots

200 GByte per tape

990 TByte total capacity

Maximum possible capacity: 6.900 tape slots, 1,400 TByte

STK SL8500

16 titanium drives

120 MByte/s transferrate

4,900 tape slots

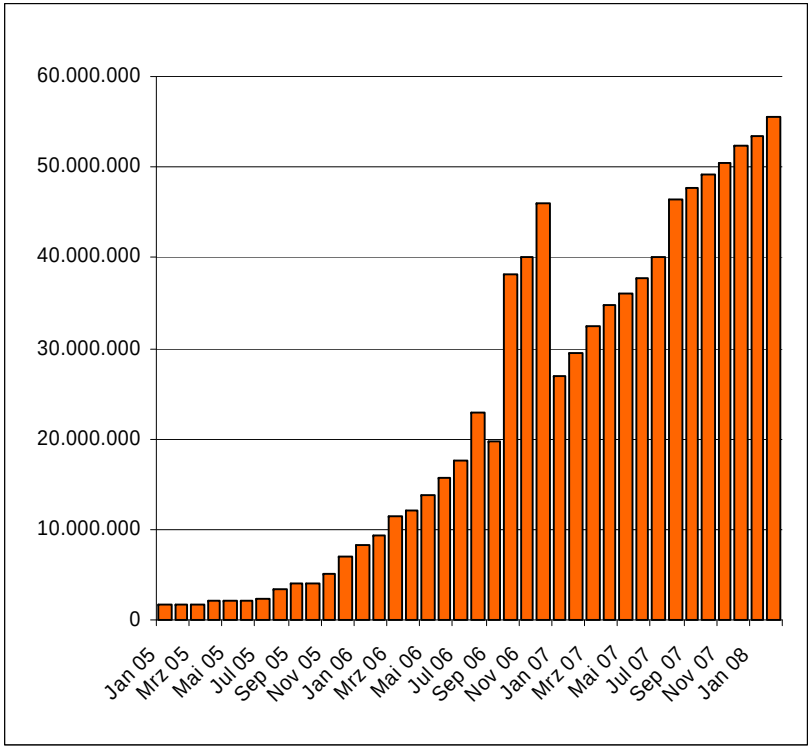
500 GByte per tape

2,400 TByte total capacity (eq. 3,8 M CDs)

Maximum possible capacity: 300.000 tape slots, 146,000 TByte

Copyright Bayerische Staatsbibliothek

MDZ: Increase of Long-Term Preserved Data



- Expected annual increase: at last 100-150 Terabyte

Copyright Bayerische Staatsbibliothek

Copyright Bayerische Staatsbibliothek

Agenda

1. The Bavarian State Library (BSB) and the Munich Digitisation Centre (MDZ)
2. BSB Digitisation Strategy and the Challenges for Mass Digitisation and Long-Term Preservation
3. **Processes and Production**
4. Projects
5. Outlook

a) Production Cycle and Subprocesses
   - Preparation
   - Handling of Originals
   - Scanning
   - Storage and Long-term Preservation
   - **Metadata Creation**
   - Online Publication and Data Management
   - Proof
   - Reuse

b) ZEND Production Software

c) Example of a Production Workflow with ZEND

Copyright Bayerische Staatsbibliothek

Indexing – Metadata Creation

- Metadata will needed for
  - Search and Retrieval
  - Navigation/Browsing

Creation of

1. Bibliographical (descriptive) Metadata = from OPAC
2. Structural metadata = stage model for indexing a book
   a) Only Images
   b) Image and text of the table of contents
   c) Images, text of the table of contents and indices
   d) Images and „hidden text“ (with errors)
   e) Complete, layout-proof , nearly error-free text

3. Technical MD = How the data were created?
4. Administrative MD = When the data were changed? By whom?

Copyright Bayerische Staatsbibliothek

Standards for Document formats and Character Set

- XML as data exchange format with
  - Metadata Encoding Transmission Standard (METS)
  - Text Encoding Initiative (TEI)
- Character set
  - Unicode, UTF-8 encoding

Copyright Bayerische Staatsbibliothek

Agenda

1. The Bavarian State Library (BSB) and the Munich Digitisation Centre (MDZ)
2. BSB Digitisation Strategy and the Challenges for Mass Digitisation and Long-Term Preservation
3. **Processes and Production**
4. Projects
5. Outlook

a) Production Cycle and Subprocesses
   - Preparation
   - Handling of Originals
   - Scanning
   - Storage and Long-term Preservation
   - Metadata Creation
   - **Online Publication and Data Management**
   - Proof
   - Reuse

b) ZEND Production Software

c) Example of a Production Workflow with ZEND

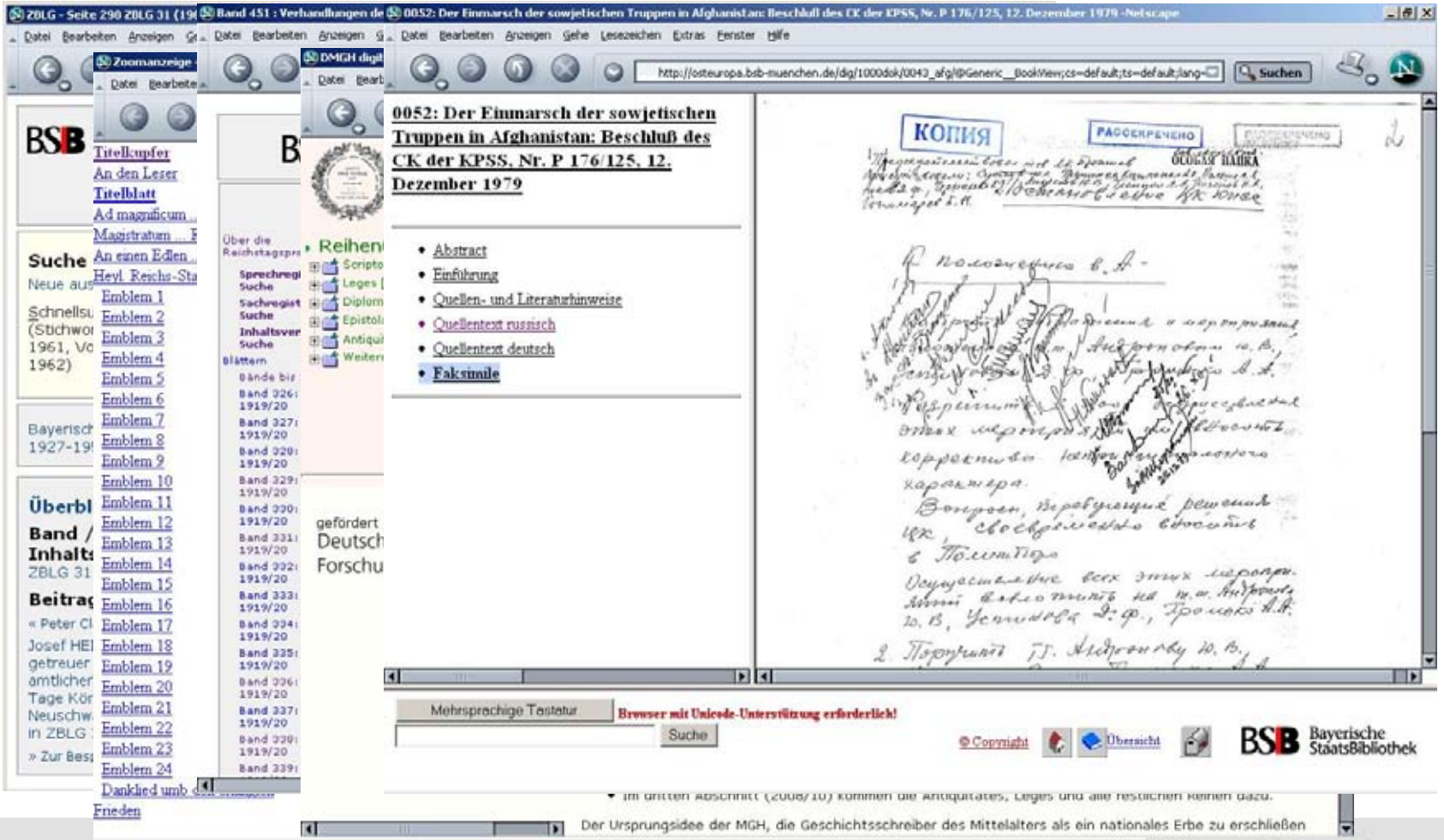
Copyright Bayerische Staatsbibliothek

Building the WWW-Interfaces

- In dependence of projects requirements and the depth of indexing for
  - Search and Retrieval
  - Navigation/Browsing
- Usability
- Accessibility

Copyright Bayerische Staatsbibliothek

Examples of Different Online Publications



Copyright Bayerische Staatsbibliothek

Agenda

1. The Bavarian State Library (BSB) and the Munich Digitisation Centre (MDZ)
2. BSB Digitisation Strategy and the Challenges for Mass Digitisation and Long-Term Preservation
3. **Processes and Production**
4. Projects
5. Outlook

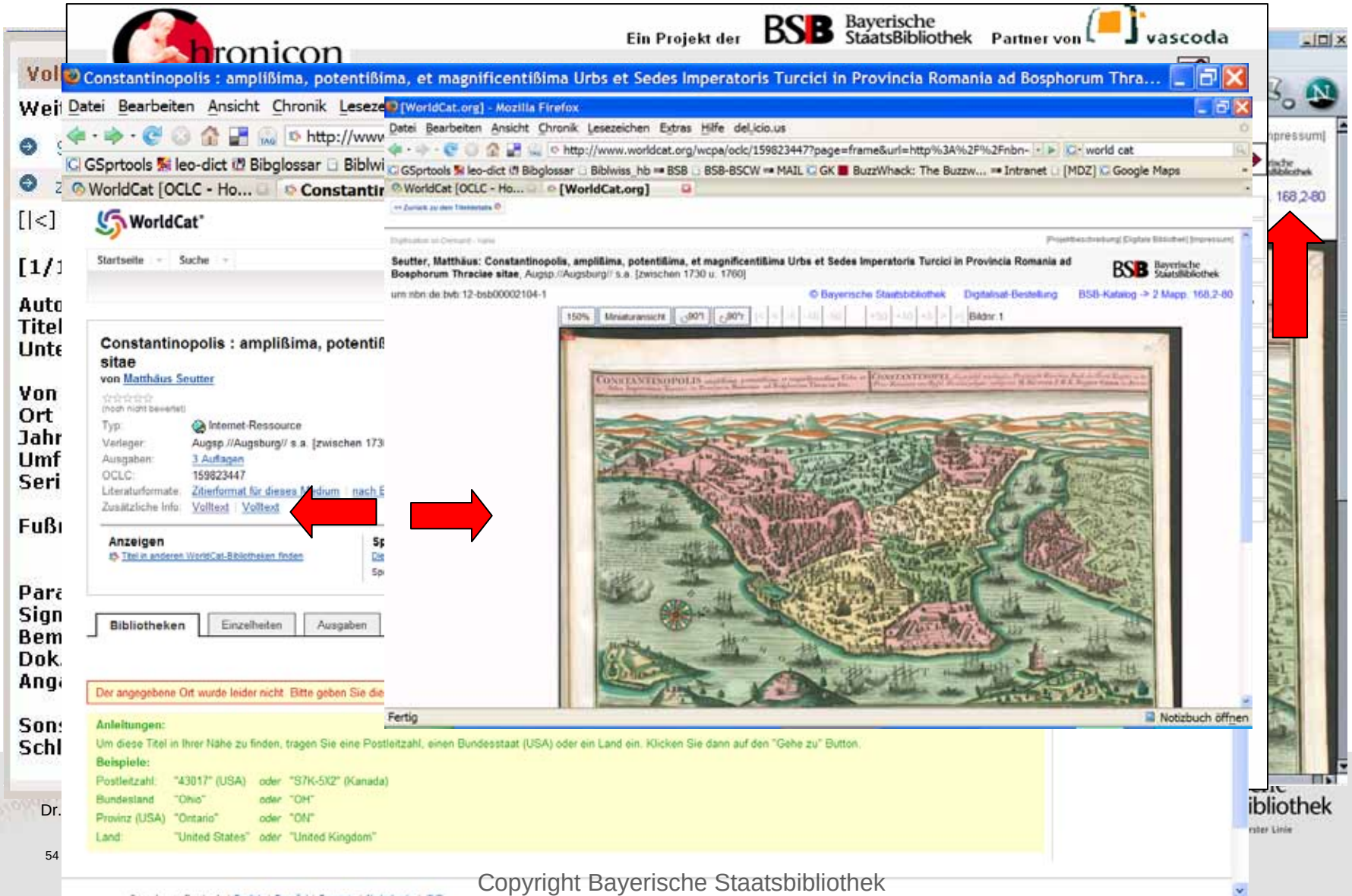
a) Production Cycle and Subprocesses
   - Preparation
   - Handling of Originals
   - Scanning
   - Storage and Long-term Preservation
   - Metadata Creation
   - Online Publication and Data Management
   - **Proof**
   - Reuse

b) ZEND Production Software

c) Example of a Production Workflow with ZEND

Copyright Bayerische Staatsbibliothek

Proofs of Digitisted Items in Catalogues, Subject Gateways ...



Copyright Bayerische Staatsbibliothek

Agenda

1. The Bavarian State Library (BSB) and the Munich Digitisation Centre (MDZ)
2. BSB Digitisation Strategy and the Challenges for Mass Digitisation and Long-Term Preservation
3. **Processes and Production**
4. Projects
5. Outlook

a) Production Cycle and Subprocesses
   - Preparation
   - Handling of Originals
   - Scanning
   - Storage and Long-term Preservation
   - Metadata Creation
   - Online Publication and Data Management
   - Proof
   - **Reuse**

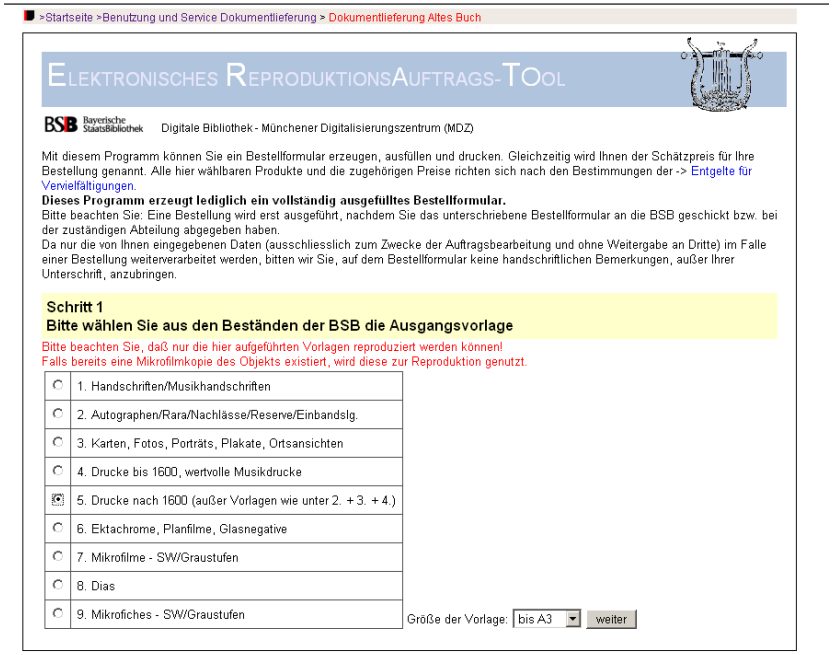
b) ZEND Production Software

c) Example of a Production Workflow with ZEND

Copyright Bayerische Staatsbibliothek

Reuse

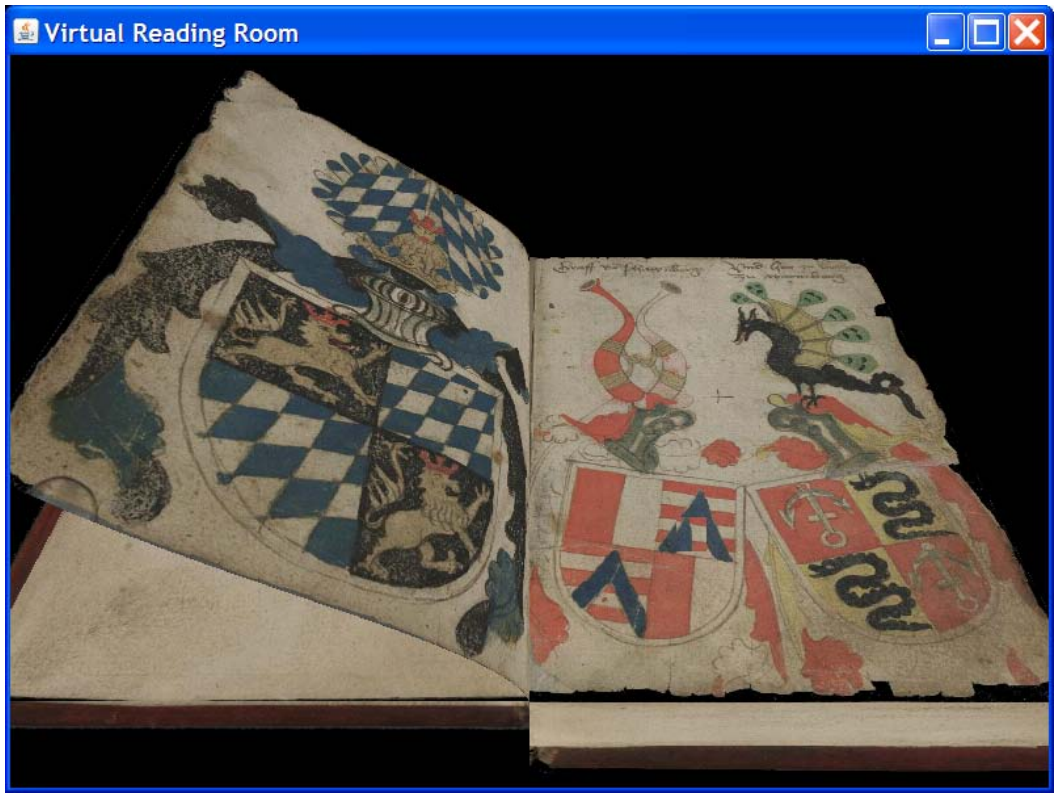
- Digital Master files at high resolution
  - Cross media publishing
    - Facsimilie and Reprint
    - Exhibiton catalogues
    - ...
  - Document delivery
  - Deeper indexing at any time
  - 3D



Copyright Bayerische Staatsbibliothek

Reuse: Example 3D-Animation with an Book of Arms

- Testbed in collaboration with book 2net



Copyright Bayerische Staatsbibliothek

Last but not least: Quality Control

- Everywhere in production at defined points
- But: 100% accuracy from „Boutique“-digitisation times is unrealistic for mass digitistion (only spot checks)
- New approaches, e.g.
  - checking page number sequence with OCR during scanning for completness
  - Involving the users in QC by offering a comment form with each image
- Problem: Each reprocessing increases costs

Copyright Bayerische Staatsbibliothek

Agenda

1. The Bavarian State Library (BSB) and the Munich Digitisation Centre (MDZ)
2. BSB Digitisation Strategy and the Challenges for Mass Digitisation and Long-Term Preservation
3. **Processes and Production**

---

4. Projects
5. Outlook

a) Production Cycle and Subprocesses
   - Preparation
   - Handling of Originals
   - Scanning
   - Storage and Long-term Preservation
   - Metadata Creation
   - Online Publication and Data Management
   - Proof
   - Reuse

b) ZEND Production Software

c) Example of a Production Workflow with ZEND

Copyright Bayerische Staatsbibliothek

ZEND Production Software

ZEND= Zentrale **E**rfassungs- und **N**achweis**D**atenbank
[central acquisition and proof database]

- ZEND developed by MDZ since 2003

- Electronic Publishing System with modules for
  - Document-Management
  - Web-Content-Management
  - Workflow-Management
  - Long term preservation of digital media

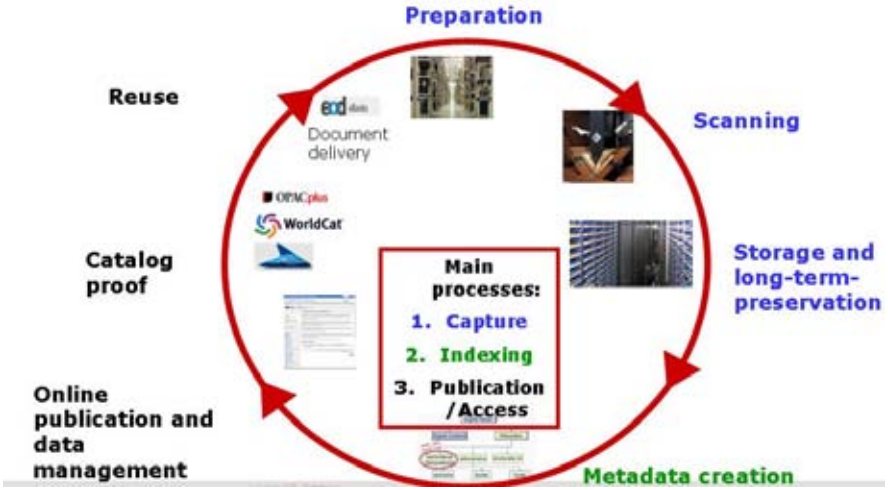
- Based on Open-Source-Software
  - LinuxApacheMySQLPHP (bzw. Perl)
  - Cocoon– XML-Publishing Framework von Apache
  - ...

- Flessible (for different media types)
- Scaleable and expandable (e.g. Cross Server Architecture)

Copyright Bayerische Staatsbibliothek

ZEND-Goals

- Mapping of the entire production cycle and its processes in a modular system
- Different service providers (scanning, text capture) can supply unlimited data to ZEND
- Workflow-control
- Every item of the BSB, which will be digitised, follows only the ZEND-workflow
- **Time and cost reduction** through extensive automation



Copyright Bayerische Staatsbibliothek

ZEND-Main Features

- Web-based user interface and administration
- Specific job control
- Collaborative working
- Import of scanned books processed by different, unlimited service providers
- Automatic URN allocation with linkresolving via German National Library
- OAI Data Provider
- Bibl. metadata-Import data via interface Z.39.50
- XML support for all common standards (for example, TEI, METS)
- Rights management: inhouse digital reading room or WWW publication
- Long-term archiving
- Search and Retrieval
- OCR server (works only for latin print type)
- PDF on Demand
- ...

Copyright Bayerische Staatsbibliothek

ZEND – Automated Processes

Full-automatic processes

- Collection of image data from different service providers
- Image conversion
- Fast supply of an simple working turning-pages version of the book (XML)
- Regular bibliographical data updates (ZEND from the OPAC)
- OCR
- Long-term archiving of master files at LRZ and
- subsequent deleting of master-fiels on the MDZ-servers...

Semi-automatic processes

- Loading of large amounts of bibliographic metadata from catalogue
- Re-import lists of the URNs in the catalogue
- Import of prepared Excel-sheets for the creation of ToCs in ZEND
- Allocation of page numbers to digitised books (for navigation)
- Transfer of digital master back from the long-term archive
- ...

Copyright Bayerische Staatsbibliothek

ZEND-Layers

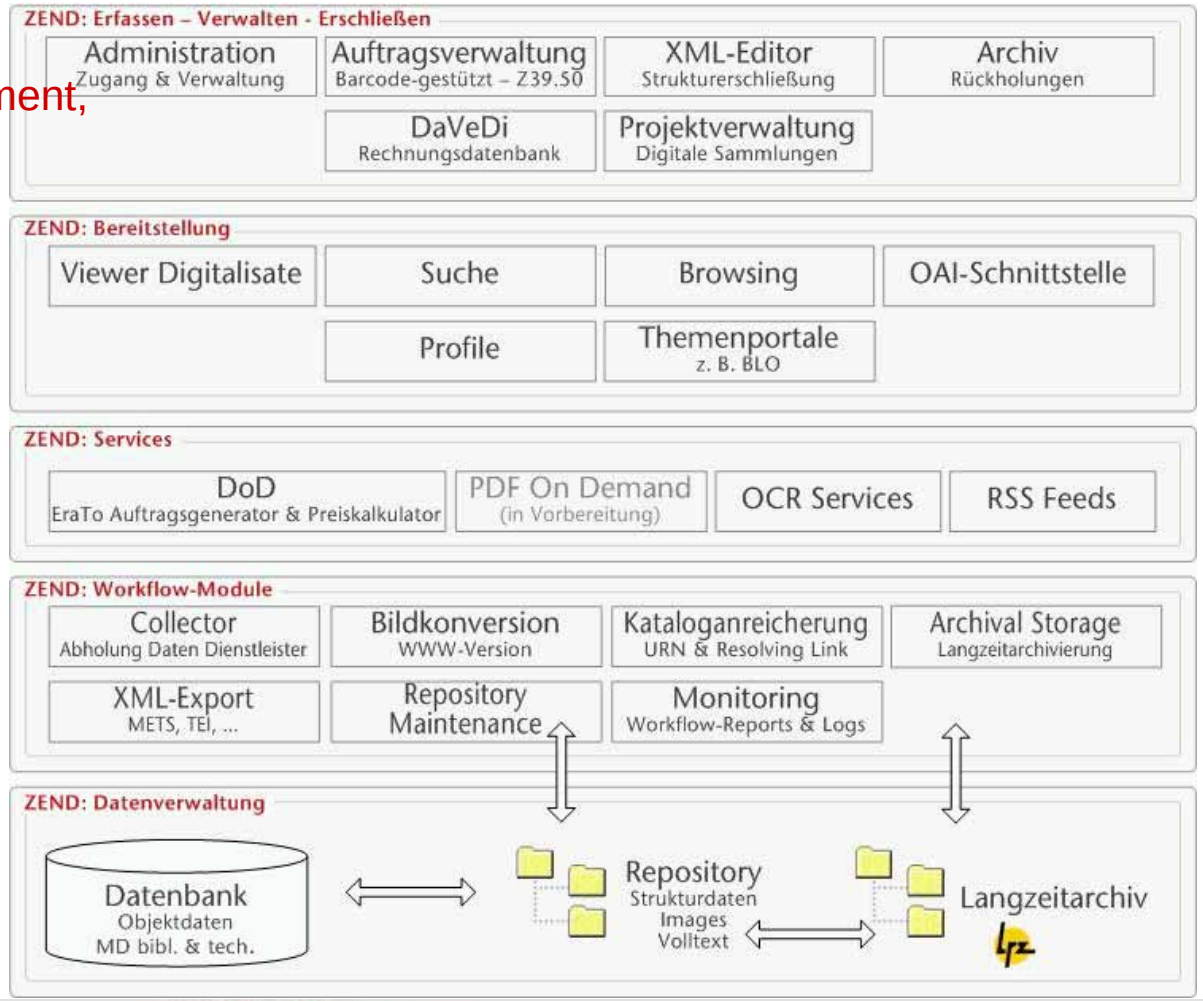
Capture, Content-Management, Indexing

Publication

Services

Workflow-Management

Data-Management



Copyright Bayerische Staatsbibliothek

Agenda

1. The Bavarian State Library (BSB) and the Munich Digitisation Centre (MDZ)
2. BSB Digitisation Strategy and the Challenges for Mass Digitisation and Long-Term Preservation
3. **Processes and Production**
4. Projects
5. Outlook

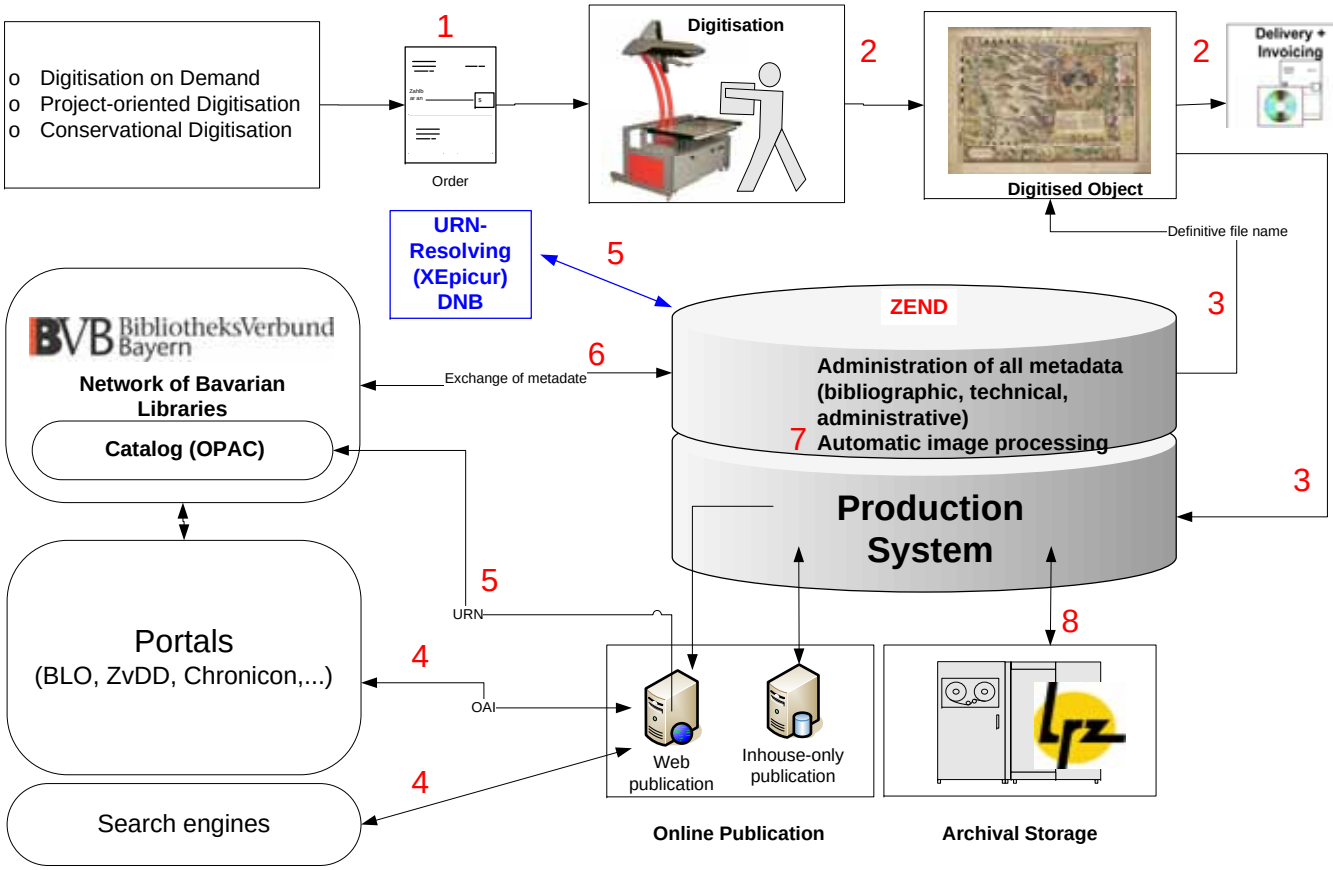
a) Production Cycle and Subprocesses
   - Preparation
   - Handling of Originals
   - Scanning
   - Storage and Long-term Preservation
   - Metadata Creation
   - Online Publication and Data Management
   - Proof
   - Reuse

b) ZEND Production Software

c) Example of a Production Workflow with ZEND

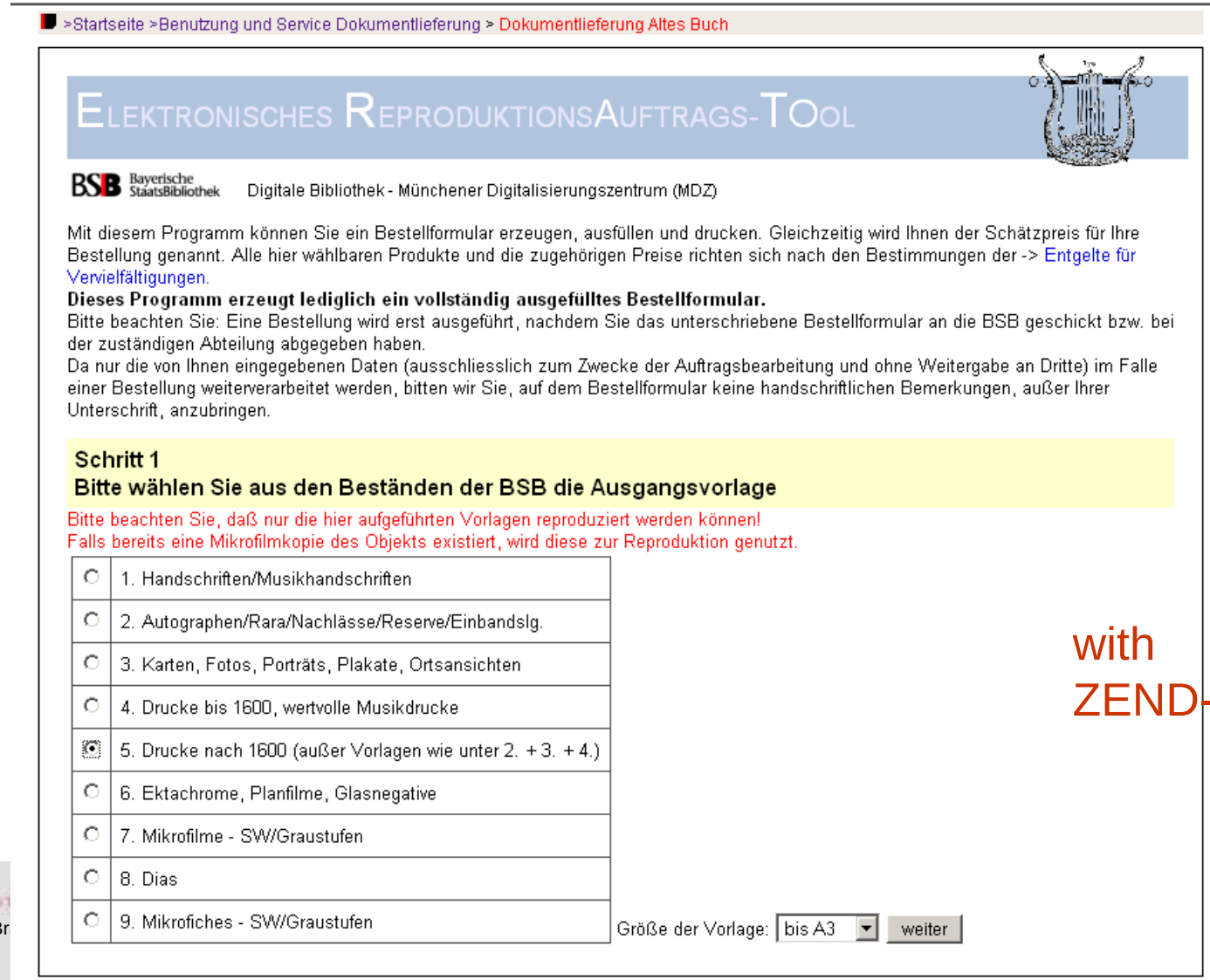
Copyright Bayerische Staatsbibliothek

Overview: ZEND-Production



Copyright Bayerische Staatsbibliothek

ZEND: Capture/Management (1)

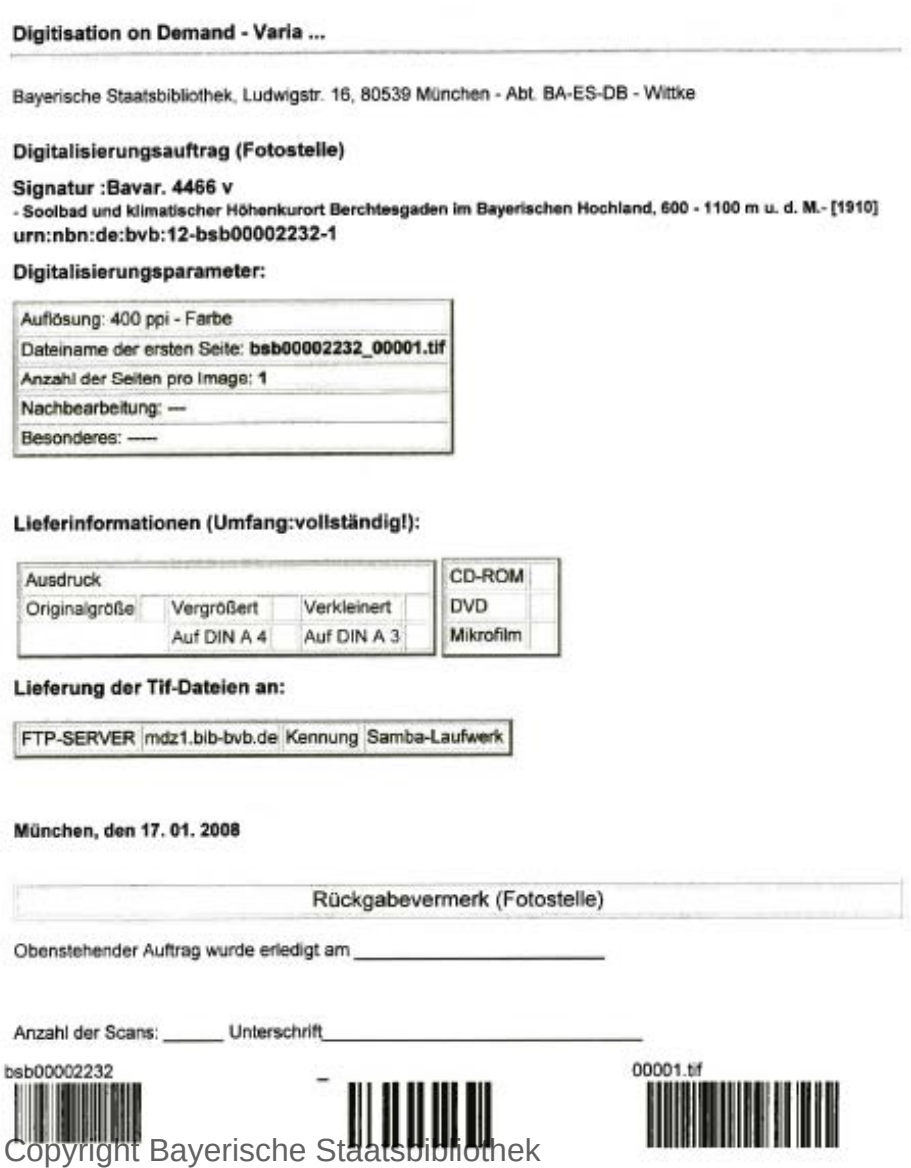


with
ZEND-Services

Copyright Bayerische Staatsbibliothek

ZEND: Capture/Management (2)

Result of order form: Process slip with barcode



Digitisation on Demand - Varia ...

Bayerische Staatsbibliothek, Ludwigstr. 16, 80539 München - Abt. BA-ES-DB - Wittke

Digitalisierungsauftrag (Fotostelle)

Signatur :Bavar. 4466 v
- Soolbad und klimatischer Höhenkurort Berchtesgaden im Bayerischen Hochland, 600 - 1100 m u. d. M.- [1910]
urn:nbn:de:bvb:12-bsb00002232-1

Digitalisierungsparameter:

Auflösung: 400 ppi - Farbe
Dateiname der ersten Seite: **bsb00002232_00001.tif**
Anzahl der Seiten pro Image: 1
Nachbearbeitung: ---
Besonderes: -----

Lieferinformationen (Umfang:vollständig!):

Ausdruck			CD-ROM
Originalgröße	Vergrößert	Verkleinert	DVD
	Auf DIN A 4	Auf DIN A 3	Mikrofilm

Lieferung der Tif-Dateien an:

FTP-SERVER	mdz1.bib-bvb.de	Kennung	Samba-Laufwerk

München, den 17. 01. 2008

Rückgabevermerk (Fotostelle)

Obenstehender Auftrag wurde erledigt am _____

Anzahl der Scans: _____ Unterschrift_____

bsb00002232 _ 00001.tif

Copyright Bayerische Staatsbibliothek

ZEND: Capture/Creation of File Name and URN (3)

- Automated assignment with **creation of the process slip**
  - Of the definitive file name

    Example: **bsb00001119_00001.tif**

    and at the same time

  - Assignment of a persistent identifier by use of the National Bibliography Number

    Example : urn:nbn:de:bvb:12- **bsb00001119**

  - Assignment is done locally, but the administration of the URN and the central link resolving is part of the duties by the DNB: http://nbn-resolving.de/urn/resolver.pl?urn= urn:nbn:de:bvb:12- **bsb00001119**

Copyright Bayerische Staatsbibliothek

ZEND: Capture/Management–3

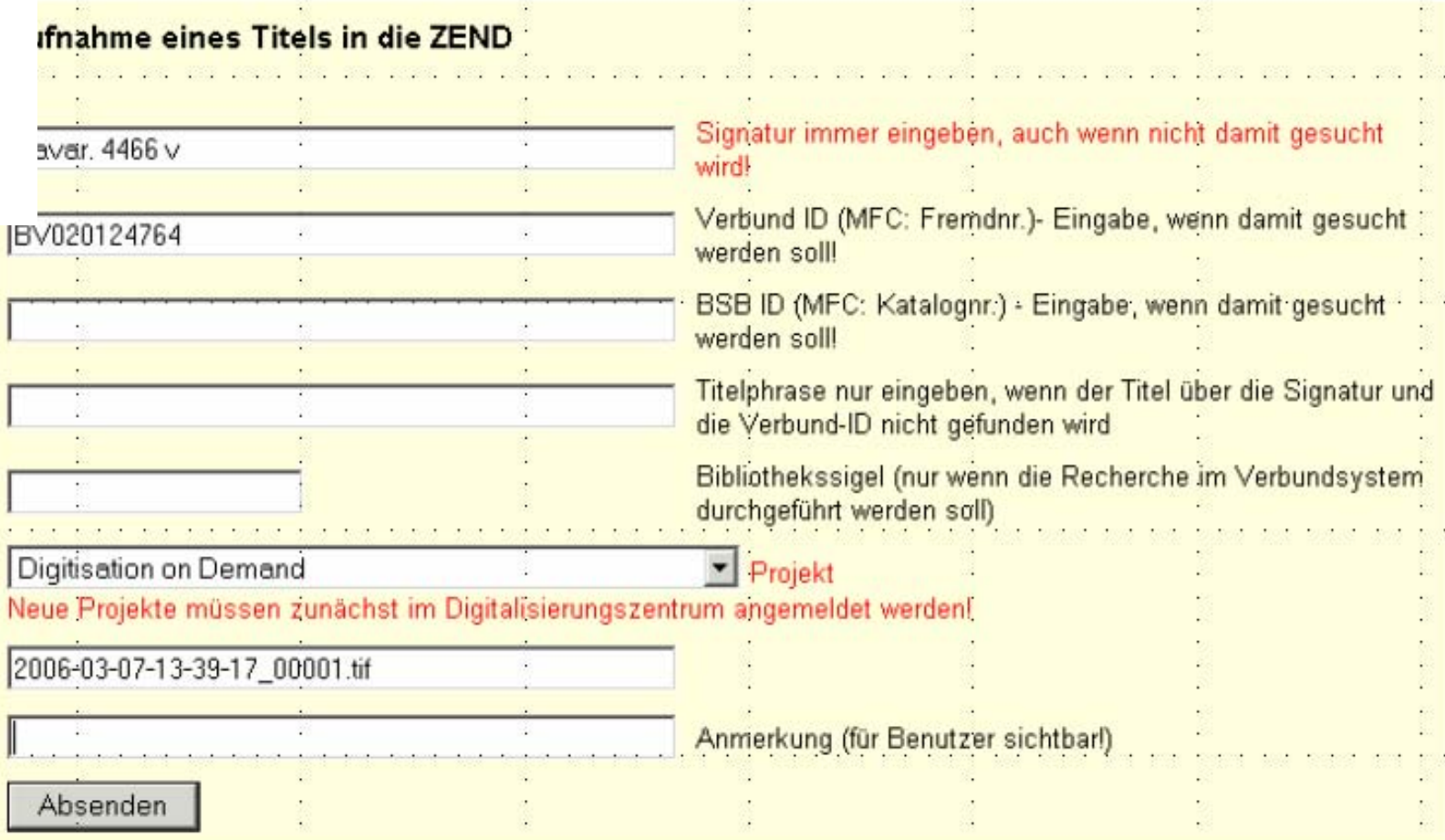
Books in process list: status

6-38_00001.tif	4 74.1458-1 [Hbh/Kp 146-1]/Capodilista: De viris illustribus/ BV018036378	H	Fotostelle			Titelaufnahme in ZEND	X		
4-55_00001.tif	Cod.icon. 394 a	H	Fotostelle			Titelaufnahme in ZEND	P	X	
2006-08-02-11-41-25_00001.tif	4 Mus.pr. 23189#Beibd.2	M	Fotostelle			Titelaufnahme in ZEND		X	
2006-06-29-15-27-46_00001.tif	Cod.icon. 181 FERTIG	H	Fotostelle			Titelaufnahme in ZEND	P	X	
2006-06-12-09-36-50_00001.tif	Res/4 H.ref. 804,2; Ajn neuwe disputatz; BV 001407073	H	Fotostelle			Titelaufnahme in ZEND	X		
2006-05-16-16-17-10_00001.tif	Cod.icon. 173	H	Fotostelle			Titelaufnahme in ZEND	P	X	
2006-05-16-16-11-33_00001.tif	Cod.icon. 141 FERTIG	H	Fotostelle			Titelaufnahme in ZEND	P	X	
2006-05-02-12-14-17_00001.tif	Cod.icon. 399 a FERTIG	H	Fotostelle			Titelaufnahme in ZEND	P	X	
2006-04-05-14-15-48_00001.tif	Ana 350.12.A. Auerbach, Erich FERTIG	H	Fotostelle			Titelaufnahme in ZEND	P	X	
2006-03-07-13-39-17_00001.tif	Bavar. 4466 w/Berchtesgaden/BV020124764	BA-ES-DB	Fotostelle			Titelaufnahme in ZEND		X	
2006-02-14-12-09-00_00001.tif	Clm 2936 FERTIG	H	Fotostelle			Titelaufnahme in ZEND	X		

Copyright Bayerische Staatsbibliothek

ZEND: Indexing (4)

Import of bibl. Metadata



Copyright Bayerische Staatsbibliothek

ZEND: Indexing (5)

Import of bibl. Metadata via Z39.50

Wittke / BA-ES-DB / D(2006-03-07-13-39-17_00001.tif)

1 Treffer

00567nM2.01000024 h
001 8278021
002 20051227000000.0
003 20080110000000.0
030 zz5d||rz|||37
040 BV020124764
050 ||||||||||||||
051 ||||z||
070 BVB
077 a12c1
078 a99996455006bBavar. 4466 vcd0e0f0
331 Soolbad und klimatischer Höhenkurort Berchtesgaden im Bayerischen Hochland, 600 - 1100 m u. d. M.
335 [Herausgegeben vom] Fremdenverkehrsverein Berchtesgaden-Land. [Umschlagtitel.]
410 (Berchtesgaden)
412 L. Vonderthann u. Sohn
425 [ca.1910]
425a1910
433 1 Taf. 18 S.
512 m
655 http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:12-bsb00002232-1

Signatur: -Verbund ID: BV020124764-BSB ID: 8278021Jahr: [ca.1910]
Ist das der richtige Datensatz? Ja, weiter / Neue Suche / Abbruch

Copyright Bayerische Staatsbibliothek

ZEND: After Scanning and Import of the Digital Master Files of the production System (6)

- Creation of an index file for the online publication of the images
- Automatic production of image formats for the web presentation (JPG, PDF etc.)
- Creation of the browsing structure for the object (ToC-Editor)
- OCR processing of the images (optional)
- Storage of all the data in the Leibniz Supercomputing Centre for long-term preservation

Bayerische StaatsBibliothek
Information in erster Linie

Copyright Bayerische Staatsbibliothek

ZEND: Indexing after Import (7)

Soolbad und klimatischer Höhenkurort Berchtesgaden im Bayerischen Hochland, 600 - 1100 m u. d. M. [Herausgegeben vom] Fremdenverkehrsverein Berchtesgaden-Land. [Umschlagtitel.]
(Berchtesgaden) : 1910
Verbund ID: BV020124764
BSB ID: 8278021
Signatur: Bavar. 4466 v
Bereitgestellt am: 2006-03-29
Vorläufiger Dateiname: 2006-03-07-13-39-17_00001.tif
Endgültiger Dateiname: bsb00002232_00001.tif
URN: urn:nbn:de:bvb:12-bsb00002232-1
Resolving URL:
http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:12-bsb00002232-1
0000000000 Digitisation on Demand - Varia

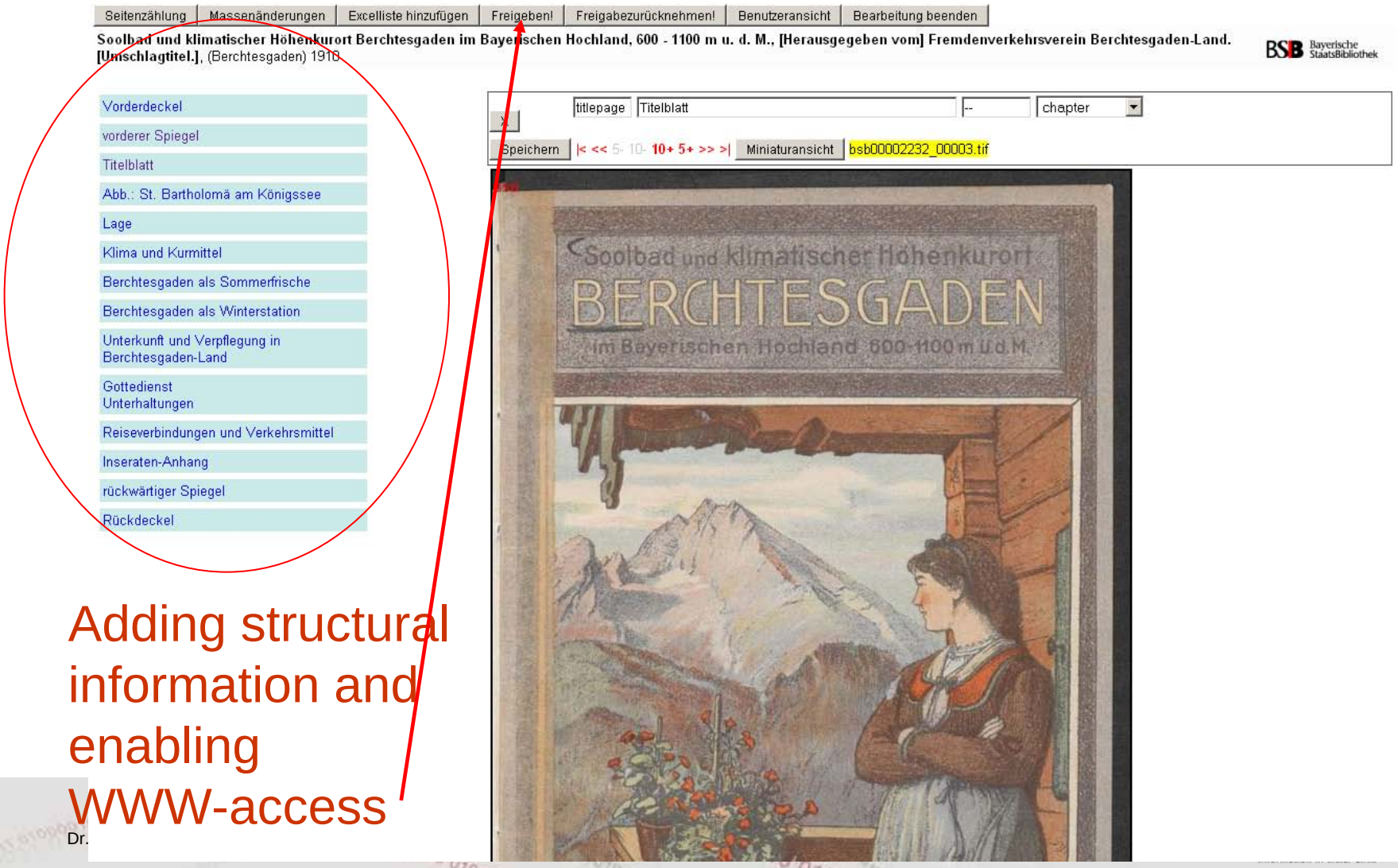
Datensatz ändern

http://mdz1.bib-bvb.de/db/bsb00002232/images/ Korrekturen Archiv X

URN – National bibliography number

Copyright Bayerische Staatsbibliothek

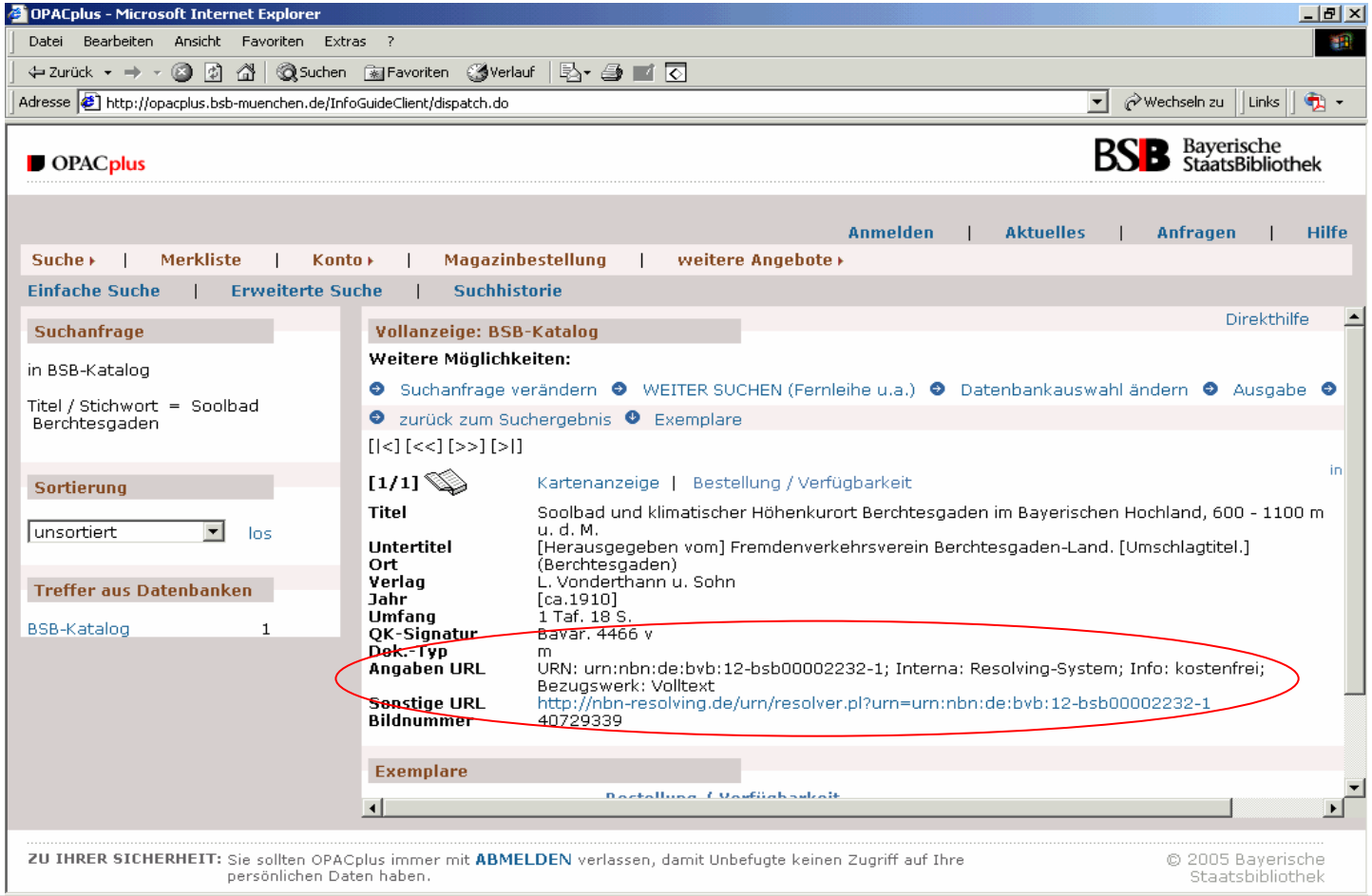
ZEND: Indexing - ToC Editor (8)



Adding structural information and enabling WWW-access

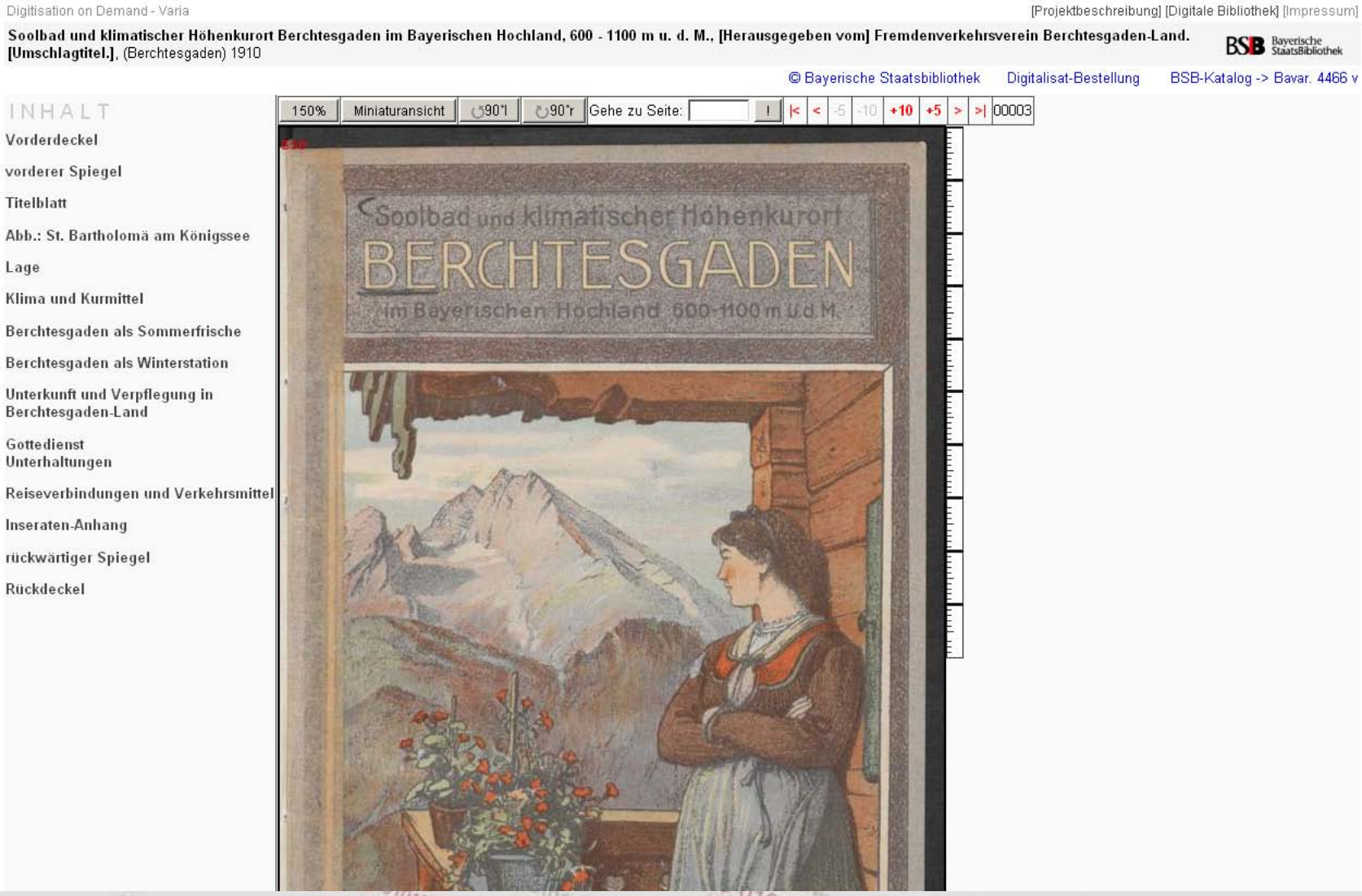
Copyright Bayerische Staatsbibliothek

Immediate availability of the link inside the local catalogue (OPAC) and the Bavarian Union Catalogue



Copyright Bayerische Staatsbibliothek

ZEND: Final Online Publication (9)



Copyright Bayerische Staatsbibliothek

Agenda

1. The Bavarian State Library (BSB) and the Munich Digitisation Centre (MDZ)
2. BSB Digitisation Strategy and the Challenges for Mass Digitisation and Long-Term Preservation
3. Processes and Production
4. **Projects**
5. Outlook

Copyright Bayerische Staatsbibliothek

Münchener DigitalisierungsZentrum
Digitale Bibliothek

Current Mass-digitisation Projects at MDZ

Project	Volumes	Project period	Pages
VD16digital-1 Deutsche Forschungsgemeinschaft DFG	**~4.300**	**2006 - 2008**	**~ 1 M.**
VD16digital-2 Deutsche Forschungsgemeinschaft DFG	**~37.000**	**2007-2009**	**~ 7.5 M.**
Inkunabeln Deutsche Forschungsgemeinschaft DFG	~ 9.000	2008-2012	~ 1,6 M.
Private Public Partnership mit Google	**> 1.2 M.**	**2008-**	**> 300 M.**

3 ScanRobots in the MDZ-Scancentre

Linux-Cluster of the MDZ for Processing the Google book copy

Copyright Bayerische Staatsbibliothek

„VD16-1 and VD16-2 digital“ -Projects

- Goal: Online Publication of all 16th century books which are unique in the inventory of BSB

Copyright Bayerische Staatsbibliothek

1501-1517(„VD16-1“)

- Pilot project for mass digitization
- Duration: 2006-2008
- Manual scanning with book scanners (A2)
- Digital Master with 300-400ppi, 24 bit, TIFF with ICC profiles, uncompressed
- 4,300 ~ ~ 20 terabytes of digital master data
- Deeper Indexing
- Processing from Capture to long-term preservation with ZEND

Deutsche Forschungsgemeinschaft DFG

Münchener DigitalisierungsZentrum
Digitale Bibliothek

1518-1600(„VD16-2“)

- Cooperation partner for long-term storage and archiving Leibniz Supercomputing Centre
- Duration: 2007-2009
- Scanning with automatic book scanners
- Digital Master with 300ppi, otherwise like V16-1
- 37000 titles = ~ 175 terabytes of digital master
- ZEND processing

Deutsche Forschungsgemeinschaft DFG

Copyright Bayerische Staatsbibliothek

Münchener DigitalisierungsZentrum
Digitale Bibliothek

Scanning Difficulties in the „VD16-1“-Project

- Conservational requirements:
  Only 70% of books could be opened up to 90 °
- Scanning with „normal bookscanners“
- Need for book cradle support: only one-sided scanning possible (1 Scan / click = 1 page)
- 3 process steps
  1. Scanning all left pages
  2. Scanning all right pages
  3. Assembling left and right pages

Due lack of pagination in 16th century books led to higher error rate

➡ **Reduction of throughput**

BSB Bayerische StaatsBibliothek
Information in erster Linie

Copyright Bayerische Staatsbibliothek

Münchener DigitalisierungsZentrum
Digitale Bibliothek

Scanning Throughput in the „VD16-1“-Project

- Average rate **3 books per Scanner** / 8 h-working day
- Increasing set-up time due conservational requirements
- Set-up time =
  - Adjustment of the book before the first scan
  - Scanning of the color/sharpness targets
  - Adjusting of sharpness during the scan-process (only for scanning without glass-plate)
  - Integration time – writing the data from the CCD/firmware to harddisk
  - Shut-down of the book
- Set-up time is changing from object to object

BSB Bayerische StaatsBibliothek
Information in erster Linie

Copyright Bayerische Staatsbibliothek

Münchener DigitalisierungsZentrum
Digitale Bibliothek

The Way to the ScanRobot

- Objective: optimisation of throughput for the scanning of early and rare prints with limited opening angle
- 2006 - Market evalution for automatic book scanners (hardware and software) and tender
- ➔Development partnership with TREVENTUS MECHATRONICS
- Milestones
  - July 2007: 2 prototypes in the BSB
  - January 2008: start of production
  - March 2008: third robot
  - April 2008: Start of 4-shift operation from 07.00 a.m. to 11.00 p.m.

BSB Bayerische StaatsBibliothek
Information in erster Linie

Copyright Bayerische Staatsbibliothek

Münchener DigitalisierungsZentrum
Digitale Bibliothek

„VD16-2“: ScanRobot-Throughput

- Status now:
  10 books per ScanRobot / 8 hour working day
  Total weekly production: 300 books ~ 75,000 pages ~ 2,2 Terabyte

  Interim result: The objective of a significant increase in production is achieved

  Next target - optimisation of the book cradle : 500 + x titles per week

BSB Bayerische StaatsBibliothek
Information in erster Linie

Copyright Bayerische Staatsbibliothek

ScanRobot in Action

- Scanning 16th centuries books with automated book scanners

Copyright Bayerische Staatsbibliothek

Digitisation in the GBS-programme

By Google and at the expenses of Google

Google Digital Copy

Integration in Google‘s services

Library Digital Copy

Integration in the BSB services

Copyright Bayerische Staatsbibliothek

Google™ Mass Digitisation

- No „Selective Picking“: no prioritisation of textcorpora, shelf numbers, material types etc.
- Logistics determines procedures and tempo of digitisation (Trucks, Carts, Operating Conditions, Shifts etc. etc.)
- Selection focusses strictly on conservation purposes, formats or Copyright-issues
- Continuous optimisation of Google‘s Scanning technologies („If it doesn‘t work today, it‘ll do tomorrow“)
- In principle: If Google cannot digitise for technical or conservational reasons, the MDZ will do

Bayerische StaatsBibliothek
Information in erster Linie

Copyright Bayerische Staatsbibliothek

Google™ Some Workflow-Facets

- Organisation of a workflow with many departments for the digitization of up to 1,000 books and more on the day
- Guaranteeing the quality of the workflow and its outcome: Track every book in every step of the process
- Collection and documentation of process data (Non-scanable marks, comments…)
- The digitisation process shouldn‘t disturb the normal library processes

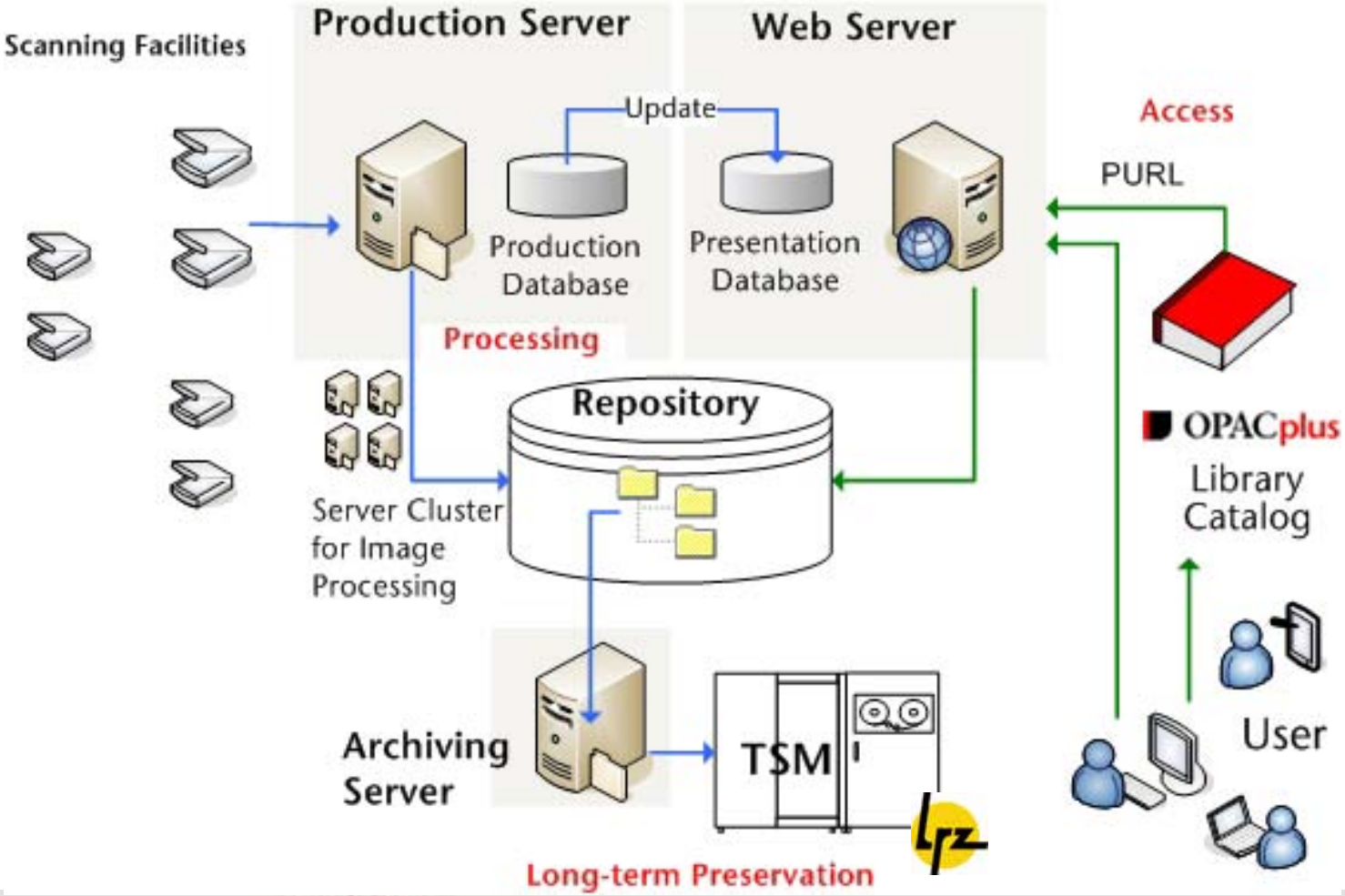
Copyright Bayerische Staatsbibliothek

Google™:The Role of MDZ

- MDZ as part of a BSB-wide (all departments involved) project organisation
- Responsible for the processing of the „**D**igital **L**ibrary **C**opy“ from data import to online-Publication
- Processing software: a special application of ZEND= „Google-ZEND“
- Storage und long-term preservation in cooperation with Leibniz Supercomputing Centre

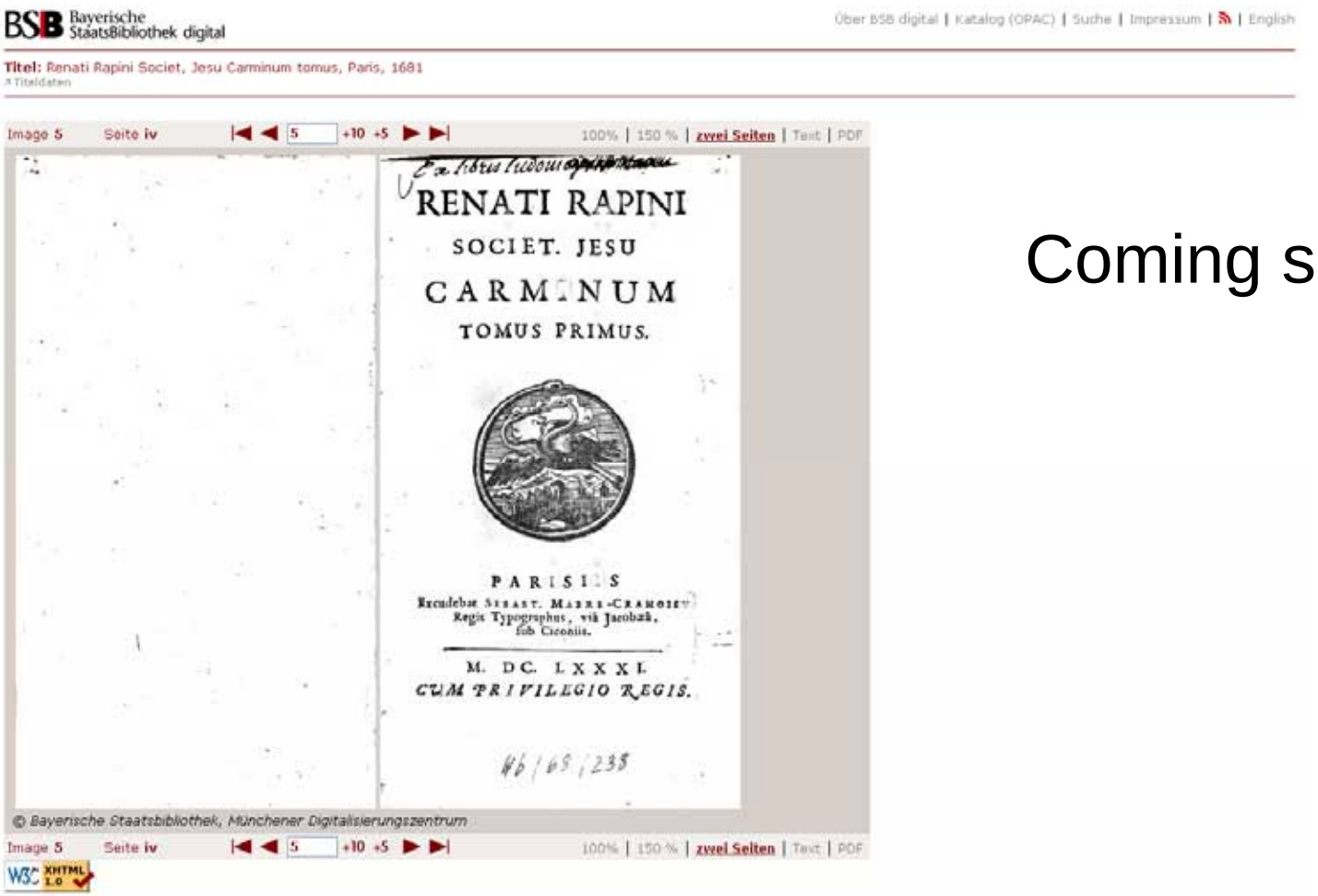
Copyright Bayerische Staatsbibliothek

Google-ZEND in a Nutshell



Copyright Bayerische Staatsbibliothek

DLC – Preview of Online Publication by MDZ



Coming soon …

Copyright Bayerische Staatsbibliothek

Agenda

1. The Bavarian State Library (BSB) and the Munich Digitisation Centre (MDZ)
2. BSB Digitisation Strategy and the Challenges for Mass Digitisation and Long-Term Preservation
3. Processes and Production
4. Projects
5. **Outlook**

Copyright Bayerische Staatsbibliothek

Outlook

- MDZ is just in the initial phase of mass digitisation
- Long-term preserved data can be processed in the future with new technologies
- Needs for further development and automation in
  - OCR
  - Indexing – Text Mining
  - Quality Control
- More online available digital content allows deeper linking between different repositories

There are still a lot of challenges …

Copyright Bayerische Staatsbibliothek

Thank you for your attention!

Contact:
markus.brantl@bsb-muenchen.de

BSB Bayerische StaatsBibliothek
Information in erster Linie

Copyright Bayerische Staatsbibliothek