Bringing it on the screen

The Mass Digitisation Projects of the Bavarian State Library

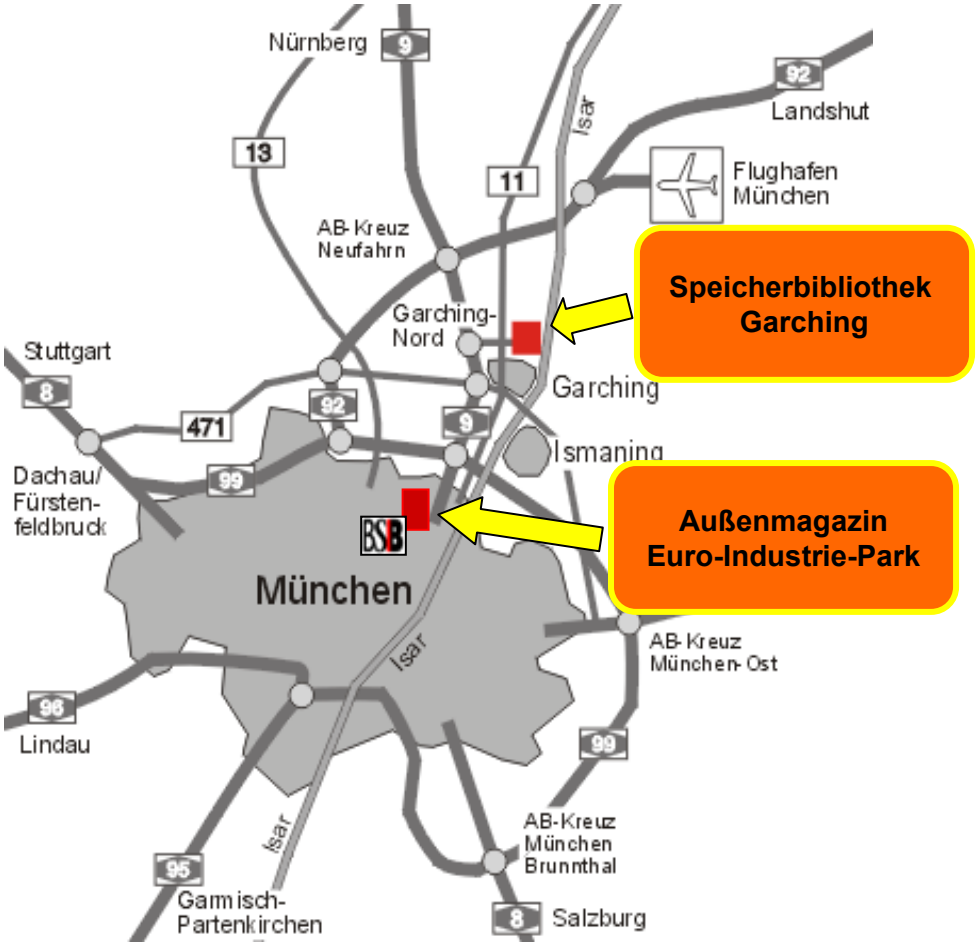
Dr. Thomas Wolf-Klostermann

INFORUM 2008
PRAHA

The Bavarian State Library at a glance

450 Jahre BSB Bayerische StaatsBibliothek

BSB – Depository site and Auxiliary stacks



Depository site (first wing – opened 1988)

Depository site (second wing – opened 2005)

BSB Bayerische StaatsBibliothek

Bavarian State Library – facts & figures

- 680 employees
- 43 million € budget p. a.
- ~ 10 million volumes
  - 89,000 manuscripts (n. 4 in the world)
  - 20,000 incunabula (n. 1 in the word)
  - 130,000 early prints (biggest collection in Germany)
- ~ 150,000 incoming volumes p. a.
- more than 40,000 current periodical titles
- 12,000 current e-journals (licenced), 600 databases
- 450,000 e-books

The Munich Digitisation Centre

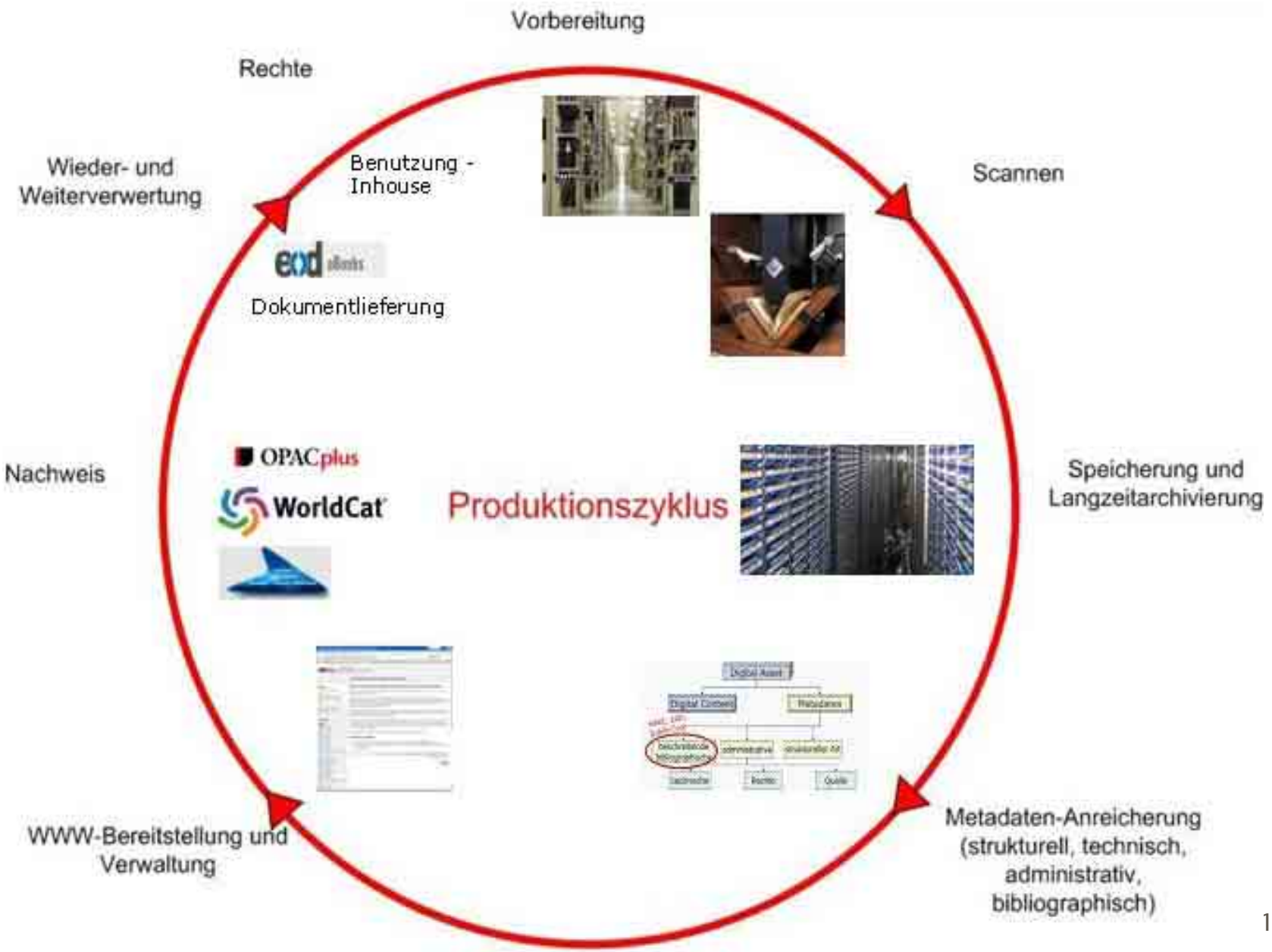
- Founded 1997
- 2003 Digital Library Department
- 2004/2005 Reorganisation
- 2007 Several new projects
- Staff: 5,5 regular employees and 29+ by third-party funds (DFG, FRG, EU)
- Main tasks
  - Digitisation
  - Subject portals/electronic publishing
  - Long-term preservation

Digital Documents at the Bavarian State Library

- Purchase / Licensing
  - Electronic journals, e-books, databases (consortia deals; pay-per-use agreements)
  - Large digital text corpora and back files of journals (national licensing program)
- Legal deposit (Federal State level)
  - Government publications, official documents
  - Output of Bavarian publishers
- Systematic Harvesting of internet resources
  - „Virtual subject gateways“
- Retrospective digitisation

Mass digitisation step by step

- 1. Organizing it
- 2. Image Processing and WWW publication
- 3. Indexing and access
- 4. Storage and Preservation


Vorbereitung
Rechte
Benutzung - Inhouse
Scannen
Wieder- und Weiterverwertung
Dokumentlieferung
Nachweis
OPACplus
WorldCat
Produktionszyklus
Speicherung und Langzeitarchivierung
WWW-Bereitstellung und Verwaltung
Metadaten-Anreicherung (strukturell, technisch, administrativ, bibliographisch)

Digitisation at the BSB: Four Reasons

1. Projects
2. Conservation
3. Digitisation On Demand
   - Scientific purposes
   - commercial purposes
4. Public-private partnership
   - Google
   - ….?

Digitisation strategy of the BSB…

- Manuscripts, incunabula, 16th century: third-party funds (e.g. DFG)
- 17-19th century: public-private partnership (Google)
- 20-21st century: funding programs of the DFG for "National Licenses" (Nationallizenzen) as well as for the BSB's special collection fields
- Special collections (maps, music supplies etc.): expecting new funding programs of the German Digital Library („Deutsche Digitale Bibliothek“)
- Allocation of bibliographic metadata via OCLC Worldcat → implementing the main search engines (Google, Yahoo etc.)

Mass Digitisation

Actual Definition

Production of more than one Million pages in a short period of time

at a constant (high!)quality konstant (following the Practical Guidelines for the Cultural Heritage Funding Programme of the DFG)

... and tomorrow?

Boutique vs. Mass Digitisation

"Boutique"-Digitisation

- Small amount of books
- Learning and experimenting (explorative projects)
- Selektive Auswahl besonderer Objekte besonders unter Marketing-Aspekten
- Indexing
- Extended manual effort

Mass Digitisation

- More than 1 M pages
- Mass production through automation
- No individual selection
- Only basic short indexing (allows deep indexing projects later on)

Recent projects in mass digitisation at the BSB

	Titles	Time period	Pages total
VD16-1 digital DFG	~4.000	2006 - 2008	~ 1 M
VD16-2 digital DFG	~37.000	2007-2009	~ 7,5 M
Partnership with Google	> 1 Mio.	2008 - ?	> 250 M

Hardware of the Munich Digitisation Centre

- 11 conventional bookscanners of different types and sizes (up to A0)
- 3 automated bookscanner
- 57 server
- 60 TB = 26,5 M files stored at the Leibniz Supercomputing Centre by March 2008

„Industrial“ Mass digitization

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

Mass Digitisation: Main Challenges

- Logistics: book picking, transport, return, allocation of dig. content, quality management
- Scanning
- Integration of metadata
- Storage and Preservation
- …

Digitisation in Practice
Some Impressions

Die Löwen 2007/2008

Google Buchsuche BETA

– Partner in the Library Program Google BookSearch

Bavarian State Library

Harvard University

The National Library of Catalonia

Ghent University Library

Princeton University

Cornell University Library

University of Wisconsin - Madison

University Complutense of Madrid

University of Michigan

Stanford University

University Library of Lausanne

University of California

University of Texas at Austin

Committee on Institutional Cooperation (CIC)

Oxford University

Keio University Library

The New York Public Library

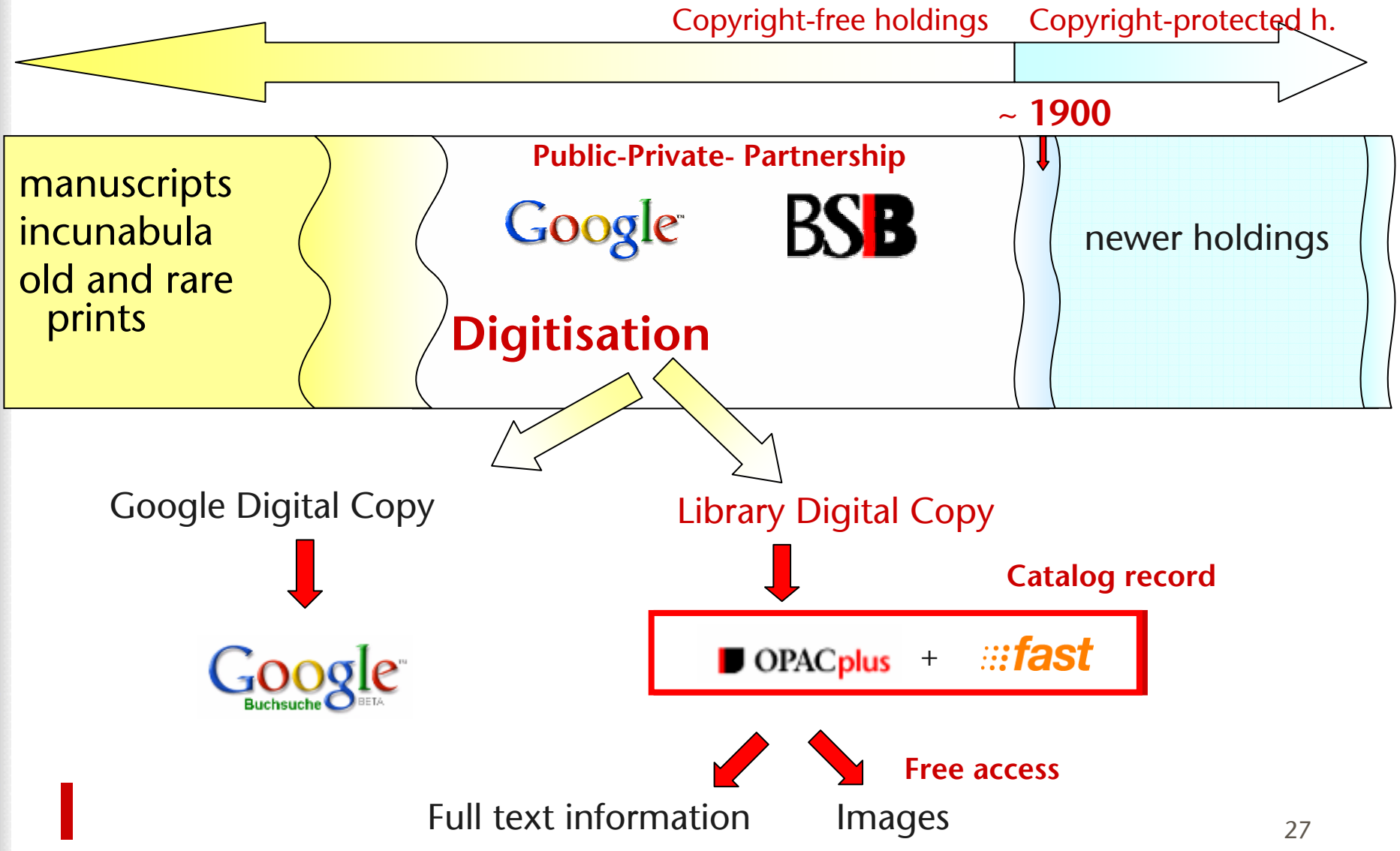
University of Virginia

Columbia University Libraries

„Mutual Non-Disclosure Agreement“

- no numbers
- no names
- no places
- no processes

The Cooperation with Google

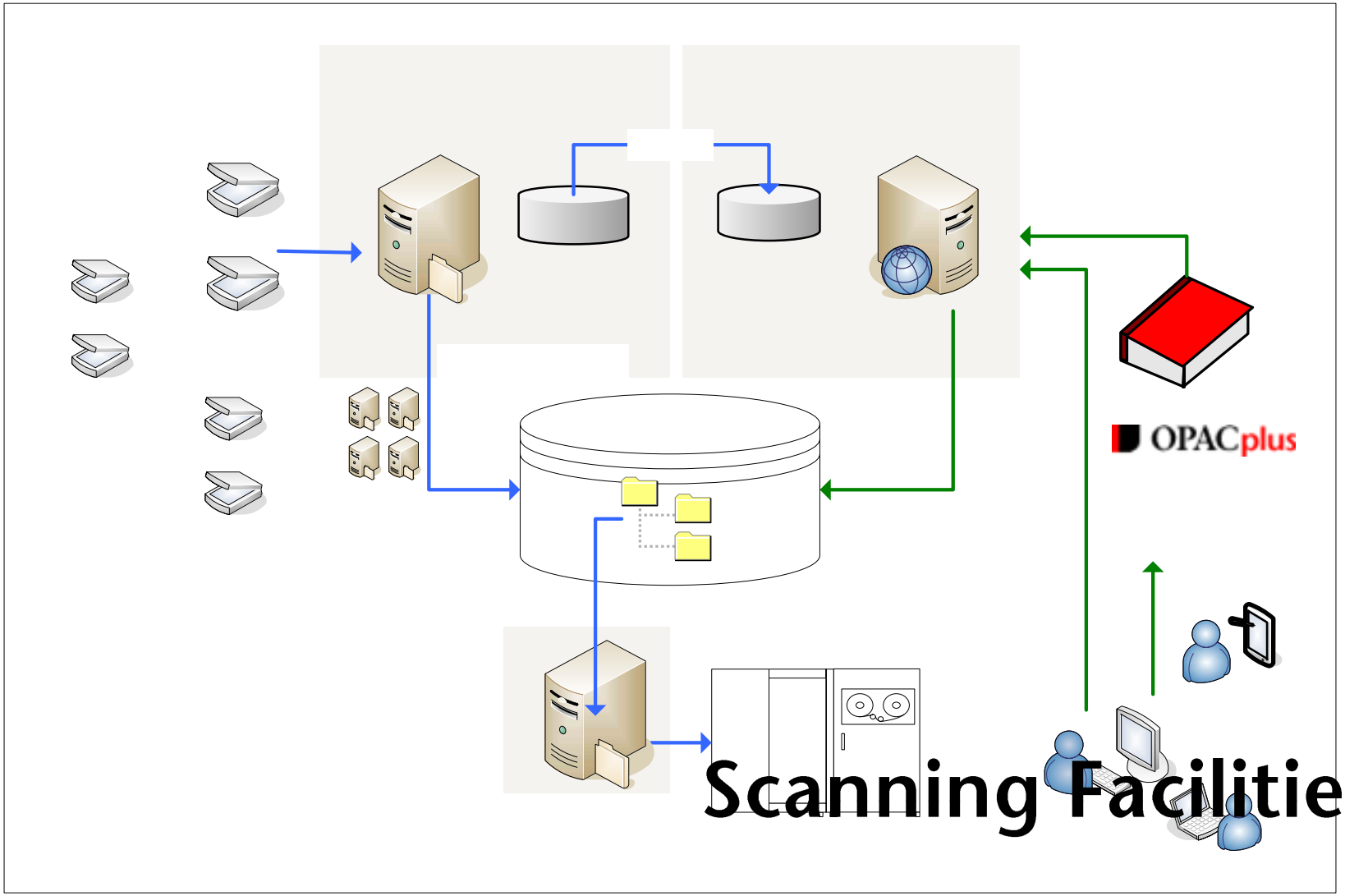


Cooperating with Google – The Workflow

- Involving participants from different departments of the library
- Designing workflows for the digitisation of up to 1,000 books a day
- Documentation and quality control at every single step
- Information
  - Information about books (not) being scanned

- Avoud disturbance of the day-to-day business of the library
  - Lending should remain possible
  - Picking the books just in time (not more books than needed)

Technical Infrastructure for Mass Digitisation Projects

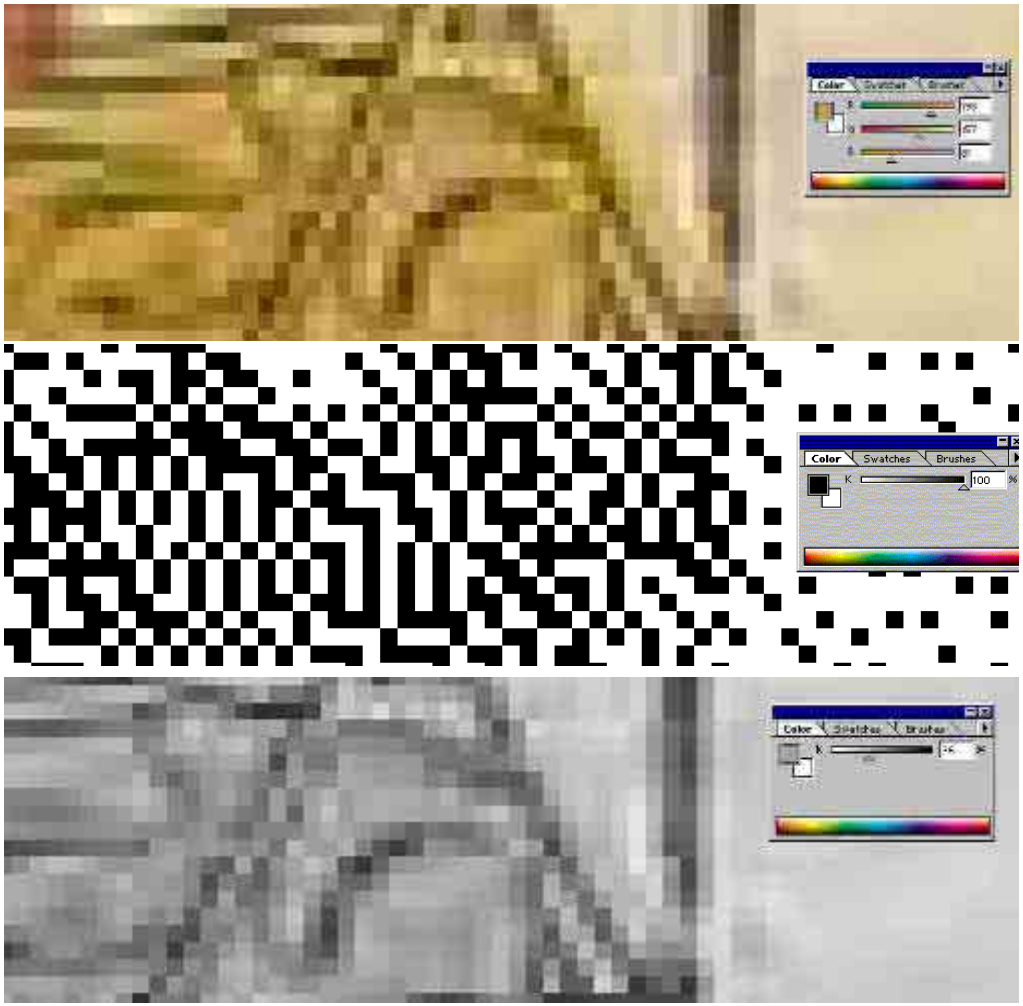


Mass digitisation step by step

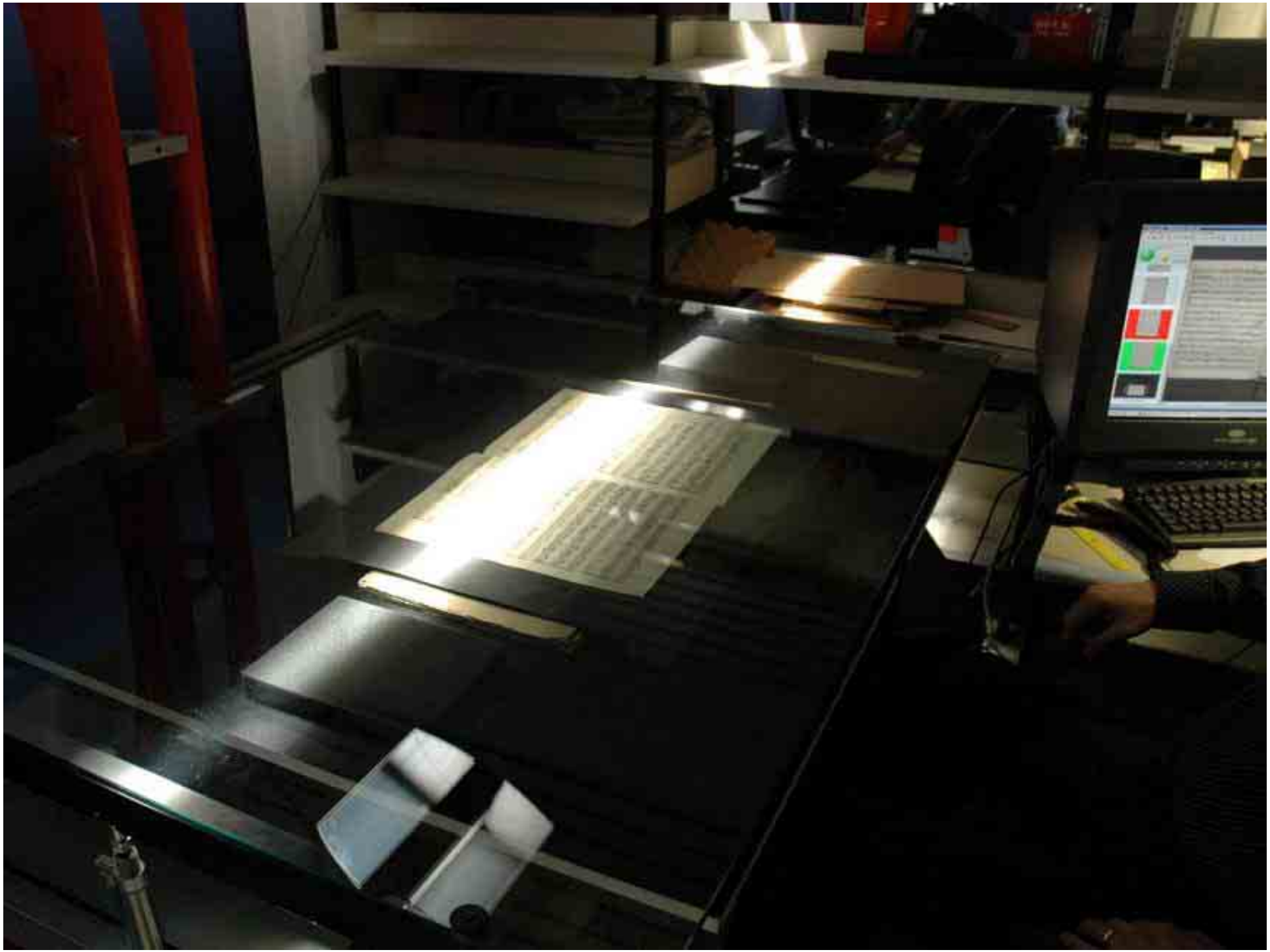
- 1. Organizing it
- 2. Image Processing and WWW publication
- 3. Indexing and access
- 4. Storage and Preservation

Digitisation Quality

colour, b/w, grayscale



Milberzhofn
S. Georg
Hadern



Mass digitisation step by step

- 1. Organizing it
- 2. Image Processing and WWW publication
- 3. Indexing and access
- 4. Storage and Preservation

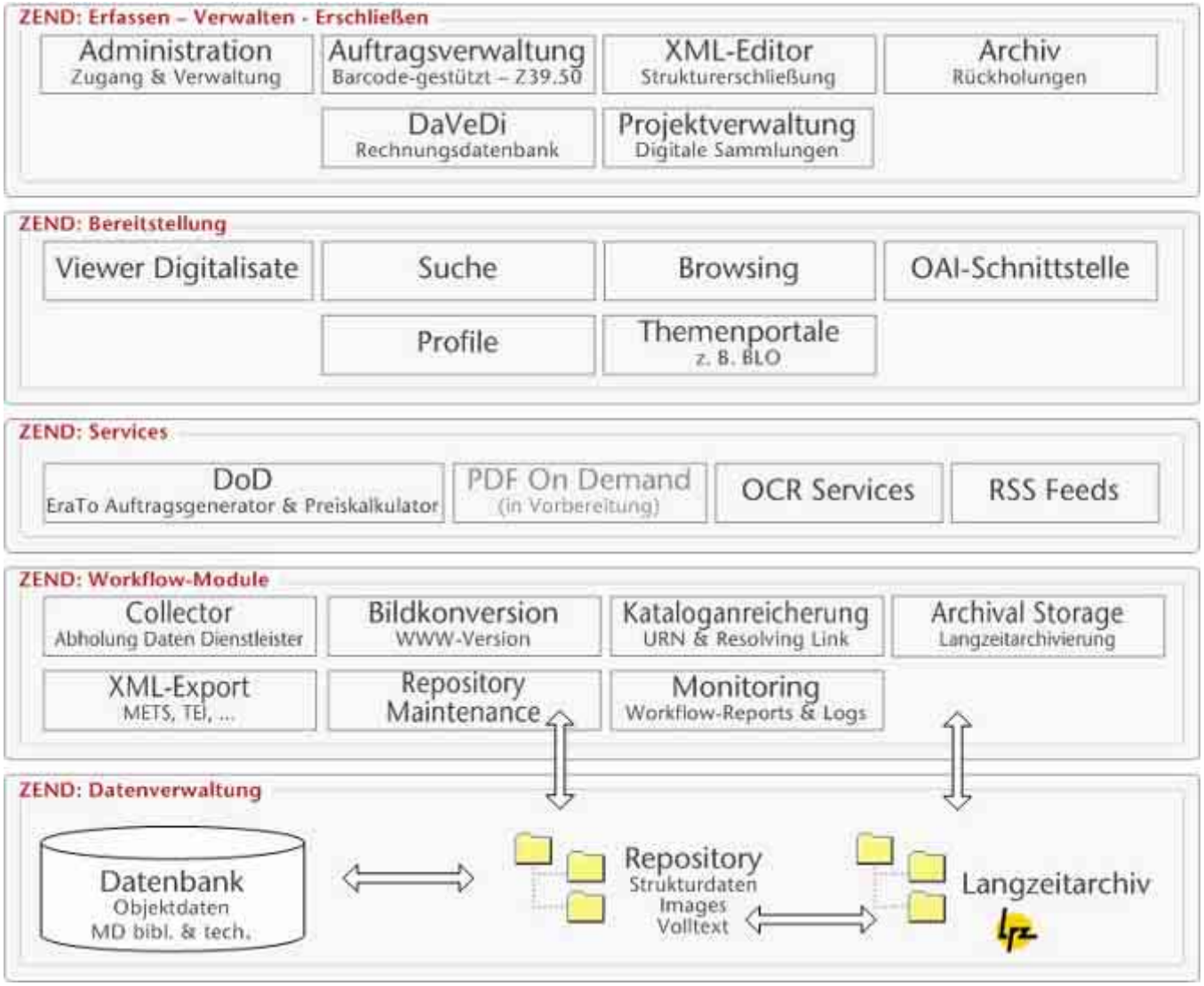
The ZEND

- Using the *Zentrale Erfassungs- und Nachweisdatenbank für Digitalisierungsprojekte* (ZEND)
  - Self-made by use of open source software
  - Backbone of all digitisation processes
  - Scaleable
  - Automated tasks
  - Interconnected with heterogeneous external resources (Library Catalogue, search engines, storage system, …)

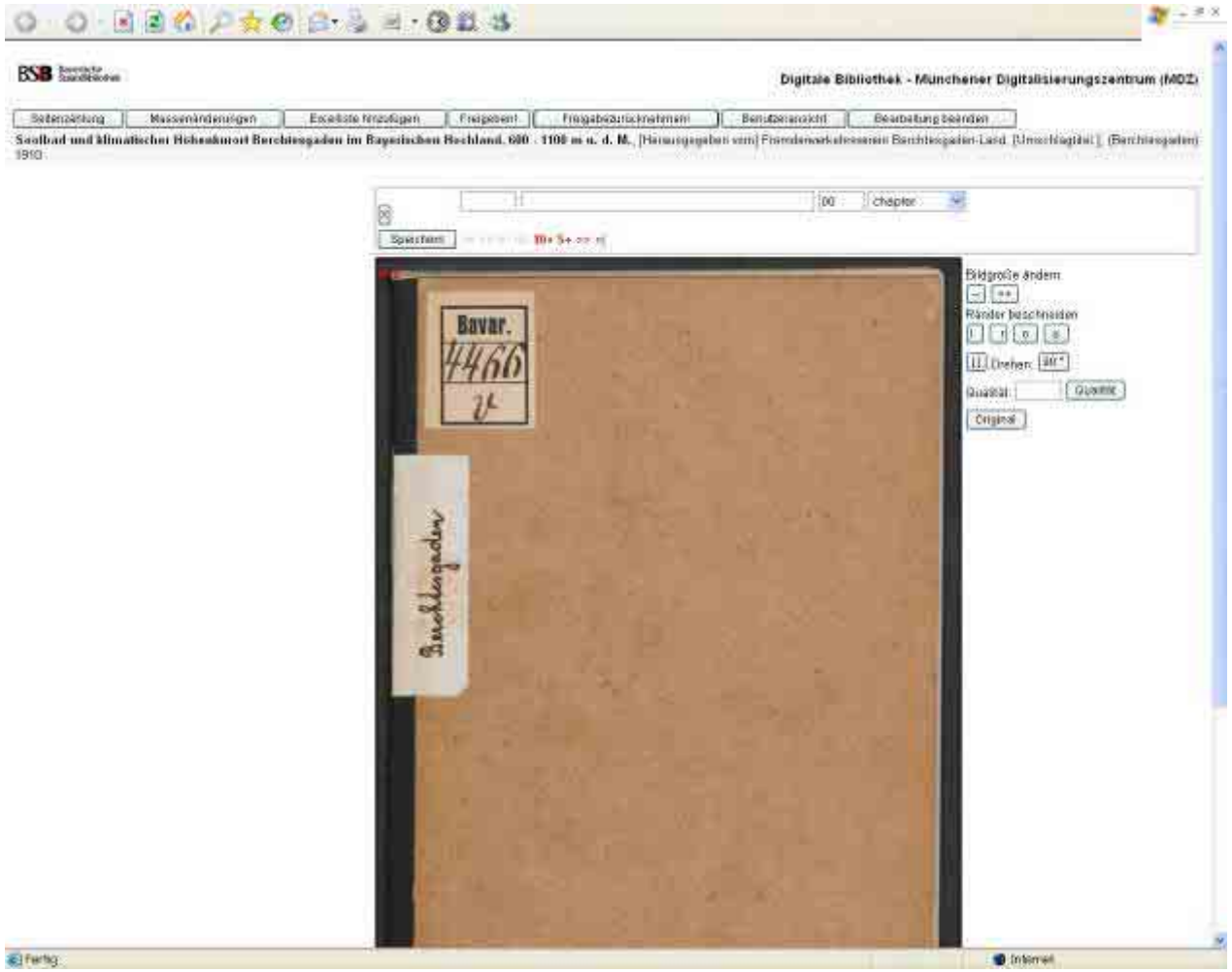
The ZEND: Basic Features

- Web-based interface for use and administration
- Job management / cooperative workings
- Allocation of an URN with link resolving via "Deutsche Nationalbibliothek"
- OAI-DataProvider
- Data import via Z.39.50
- XML-support for all usual standards (e.g. TEI, METS)
- Inhouse provision: "Digital reading room"
- Archival storage and long-term preservation in cooperation with the Leibniz Supercomputing Centre
- Search and retrieval
- OCR-Server
- PDF on Demand

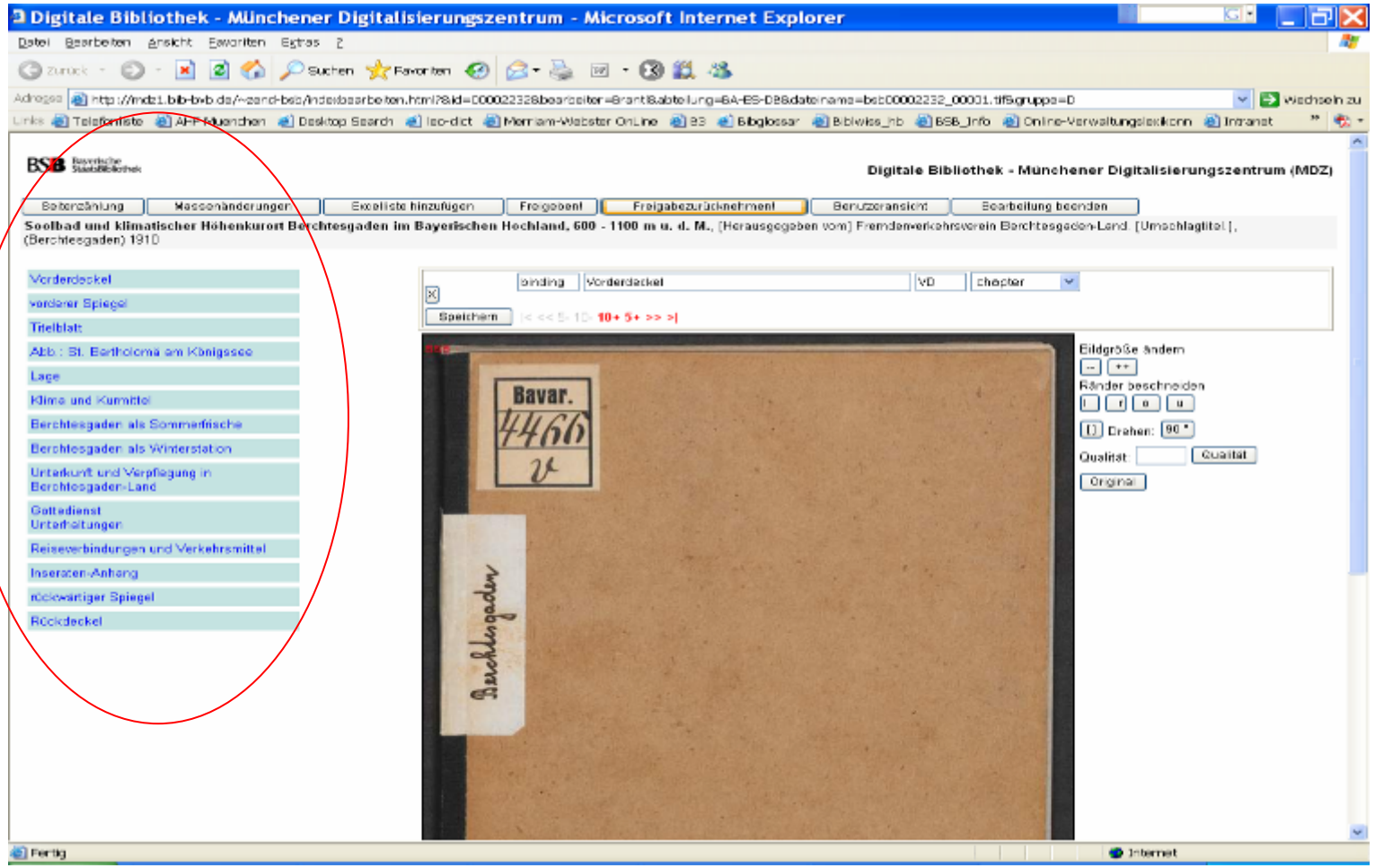
The ZEND: Basic Features



ToC(Table of Contents)-Editor



Adding structural information using the ToC Editor



Adding structural information using the ToC Editor

Seitenzählung | Massenänderungen | Excelliste hinzufügen | Freigeben! | Freigabezurücknehmen! | Benutzeransicht | Bearbeitung beenden

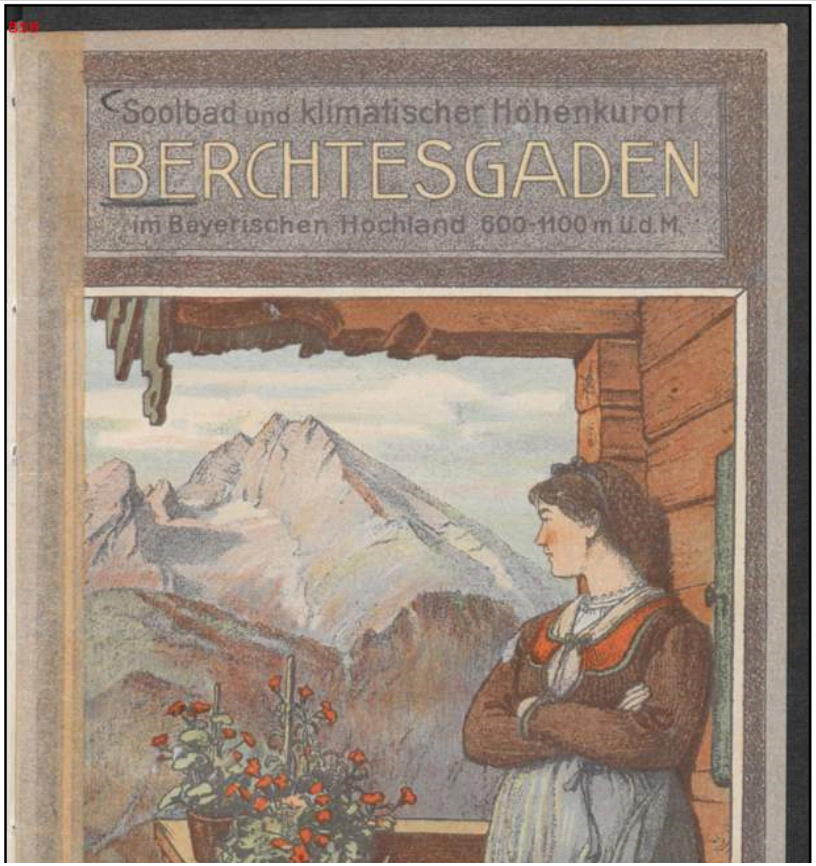
Soolbad und klimatischer Höhenkurort Berchtesgaden im Bayerischen Hochland, 600 - 1100 m u. d. M., [Herausgegeben vom] Fremdenverkehrsverein Berchtesgaden-Land. [Umschlagtitel.], (Berchtesgaden) 1910

BSB Bayerische StaatsBibliothek

- Vorderdeckel
- vorderer Spiegel
- Titelblatt
- Abb.: St. Bartholomä am Königssee
- Lage
- Klima und Kurmittel
- Berchtesgaden als Sommerfrische
- Berchtesgaden als Winterstation
- Unterkunft und Verpflegung in Berchtesgaden-Land
- Gottedienst Unterhaltungen
- Reiseverbindungen und Verkehrsmittel
- Inseraten-Anhang
- rückwärtiger Spiegel
- Rückdeckel

X | titlepage | Titelblatt | -- | chapter

Speichern | |< << 5- 10- 10+ 5+ >> >| | Miniaturansicht | bsb00002232_00003.tif



File name and URN

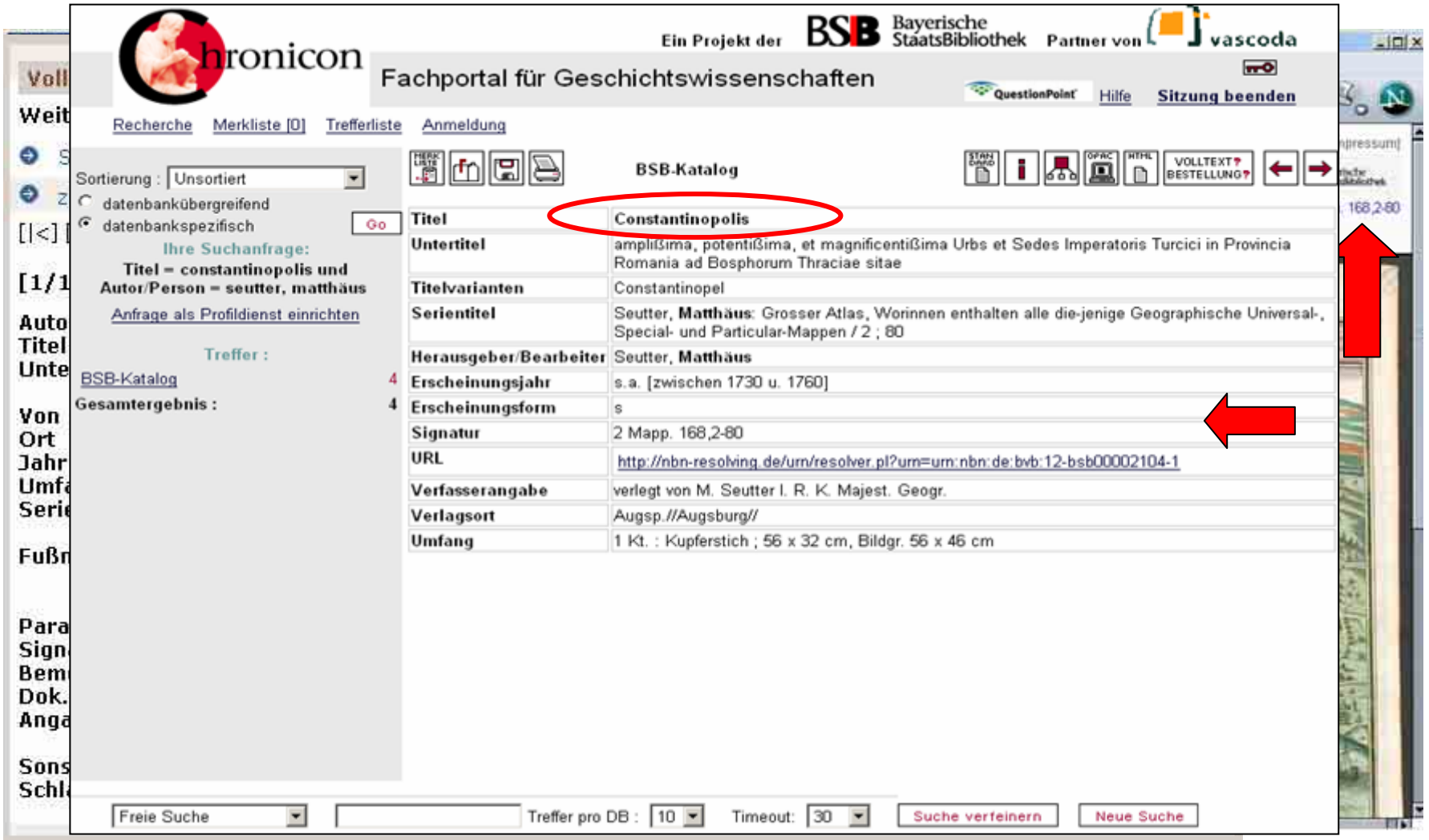
- Automated assignment
  - …of the unique identifier forming the internal BSB-file name

    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 is part of the duties of the DNB:
    - `http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:12-`**`bsb00001119`**

Access by a wide range of linkings



Mass digitisation step by step

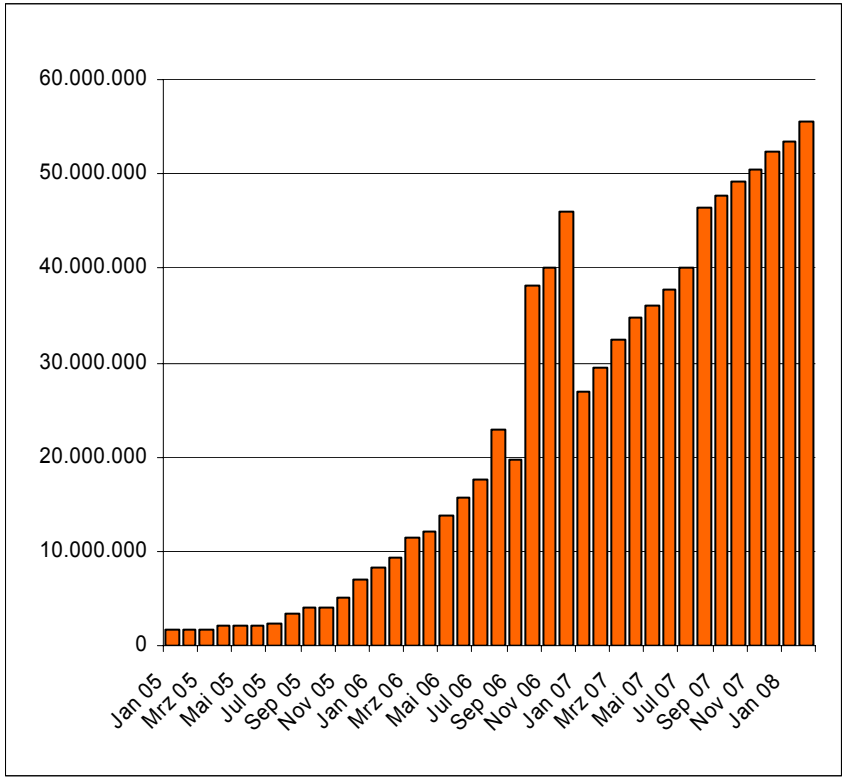
- 1. Organizing it
- 2. Image Processing and WWW publication
- 3. Indexing and access
- 4. Storage and Preservation

Long-term preservation at the MDZ

- Several projects since 1999
- Partner in the nestor project (main focus: trusted repositories)
- Intensive cooperation with the Software Technology Institute of the University of the German Federal Armed Forces
- Since 2004: Cooperation with the Leibniz Supercomputing Centre
  - Migration of the archive of CD-ROMs (2004)
  - Development of the ZEND Workflow with automated storage and retriaval functions
  - Partner in third-party projects (DFG)

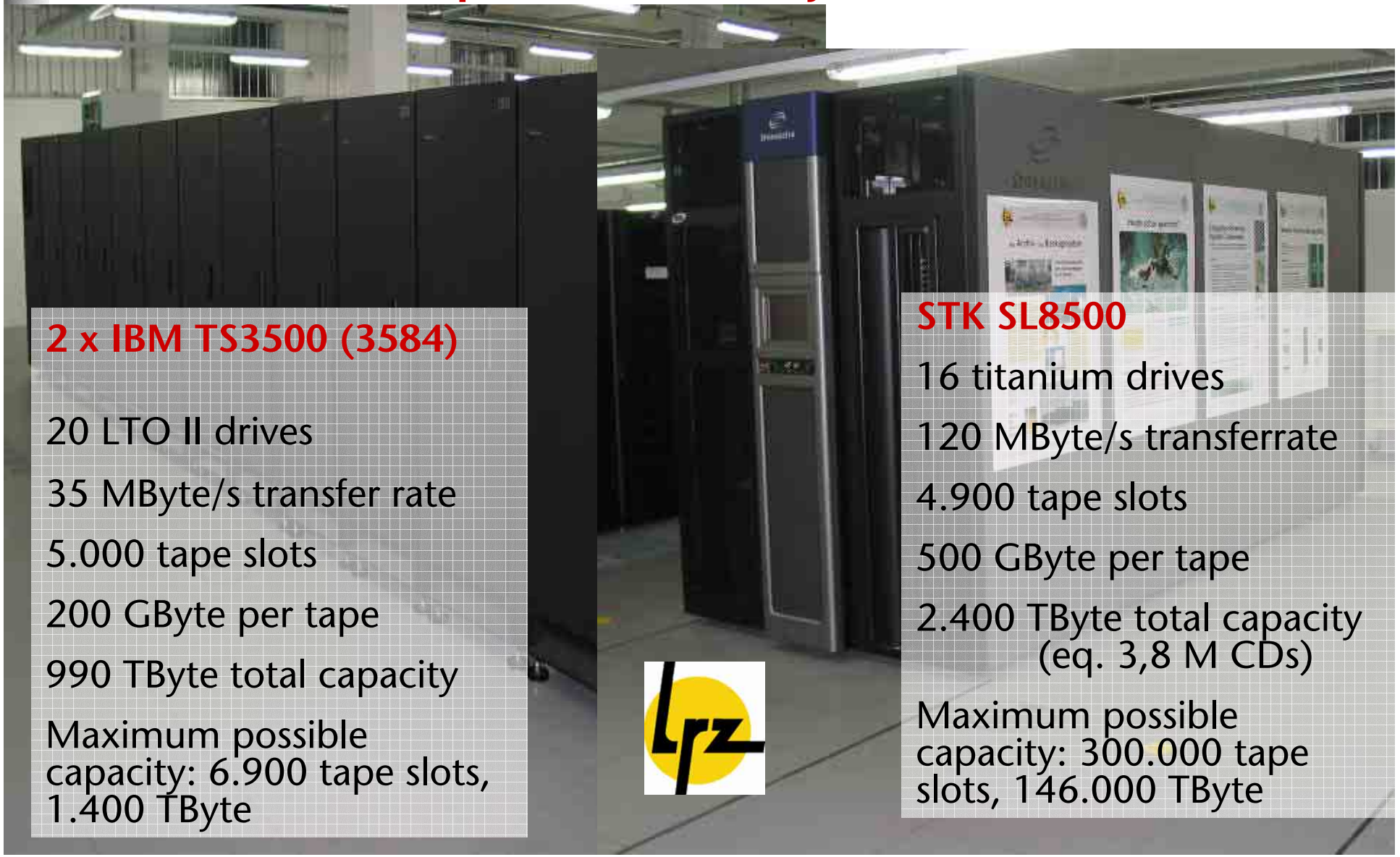
Some figures of the digital production…

- April 2008: 34.8 million files (56 TeraBytes data)
  - = more than 25.000 volumes
  - = more than 10 million pages
- Expected annual increase: at least 15 Terabyte - with scanrobot production approx. 100 Terabyte.



But with Google …?

The Backup and 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

BSB Bayerische StaatsBibliothek
Thank you!
Any Questions?
thomas.wolf-klostermann@bsb-muenchen.de

Das Münchener Digitalisierungszentrum [MDZ] - Windows Internet Explorer

http://www.digital-collections.de

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

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 25,969 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.

Swedish troops entering Munich in 1632, during Thirty Years' War. Detail from a copperplate engraving by Matthäus Merian (Hbks/E 24-2#564/565, Frankfurt 1637)

Digital Collections

Currently **25,969 Titles** online [15.05.2008]

↗ Brief Overview
↗ Extended Overview

Find collection

RSS

MDZ-News

Scan Robot Days at the Bavarian State Library

The Munich Digitization Centre and the Institute of Book and Manuscript Conservation of the Bavarian State Library will host a meeting of the leading manufacturers of robotic book scanning devices: June 18-20, 2008 the four leading manufacturer of automatic scanning technology will present their latest developments in scan robot technology and provide an opportunity to test the equipment. See ↗ http://www.digital-collections.de/robots for details. [07.05.2008]

Enhanced language support

Titles and content descriptions for almost all our ↗ digital collections are now available in English too. [19.03.2008]

Internet 100%

Münchener DigitalisierungsZentrum
Digitale Bibliothek

450 Jahre BSB Bayerische StaatsBibliothek

Scan Robot Days 2008

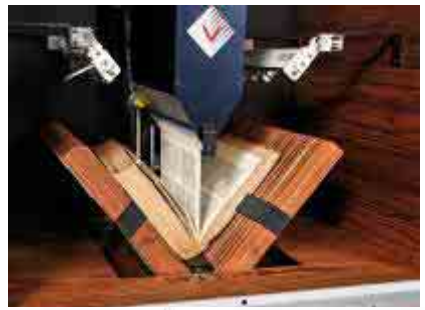
QIDENUS TECHNOLOGIES

4DigitalBooks™

A Summit Meeting of robot scanning expertise

Exhibition
Live demonstration
Presentations

June 18 – 20 2008
Bavarian State Library Munich



TREVENTUS MECHATRONICS

KIRTAS Technologies, Inc

http://www.digital-collections.de/robots