Digitization Workflow
of the
Bavarian State Library

Gabriele Messmer
Bavarian State Library
Munich, Germany
Digitization process at a glance

Preparation

Scanning

Reuse

Catalog proof

Online publication and data management

Main processes:
1. Capture
2. Indexing
3. Publication/Access

Metadata creation

Storage and long-term preservation

Document delivery
ERaTO – a tool

❖ to create
❖ fill in and
❖ print an order form
❖ that informs the patron about the estimated price
Order Form for Digitization on Demand (DoD)

Elektronisches Reproduktionsauftrags-Tool

Digitale Bibliothek - Münchener Digitalisierungszentrum (MDZ)

With this software you can create, fill in and print an order form. At the same time, you will be informed about the estimated price. All available products and their prices comply with the charges for photographic work. This software merely creates a completely filled order form.

Please note: An order is only then valid after you sign it and will be considered after you send it to the Bavarian State Library or submit it at the responsible library branch. With the exception of your signature, please don’t add any other handwritten notes or edit the form, since only those information, which you have entered, will be processed. The customer data will solely be used for the purpose of order processing and won’t be referred to a third party.

Step 1
Please select the document type from the stock of the Bavarian State Library

Please note: Only the below-mentioned document types can be reproduced!
If a microfilm copy of the original document exists, then this will be used for reproduction.

| 1. manuscripts/music manuscripts |
| 2. autographs/literary remains/book cover collection/shelf number beginning with Rar. or Res/ |
| 3. maps, photos, portraits, bills, vedute |
| 4. precious score prints and prints till the year 1600 |
| 5. prints after the year 1600 and does not belong to 2., 3. or 4. |
| 6. ektachrome, sheet film, reverse image on glass |
| 7. microfilms - black-and-white/grayscales |
| 8. slides |
| 9. microfiches - black-and-white/grayscales |

Document size (mandatory input)
- up to A3 (DIN standard)
- larger than A3 (DIN standard)

In case of doubt on document size, please refer to OPAC (full details/page number).
Step 2  
Please select one of the following reproduction products

<table>
<thead>
<tr>
<th>Digital images</th>
<th>1,50 Euro / page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Images (JPG) - colour - resolution 150 dpi - (for private/scientific use)</td>
<td></td>
</tr>
<tr>
<td>Images (TIFF) - grayscale - resolution 400 dpi (for prepress reproduction)*</td>
<td>4,00 Euro / page</td>
</tr>
<tr>
<td>Images (TIFF) - colour - resolution 400 dpi (for prepress reproduction)*</td>
<td>6,00 Euro / page</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hard copies</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A4 - black-and-white - standard paper</td>
<td>1,50 Euro / page**</td>
</tr>
<tr>
<td>A3 - black-and-white - standard paper</td>
<td>1,50 Euro / page**</td>
</tr>
</tbody>
</table>

If you desire high quality prints, please order the digital copy here and contact a digital service provider of your choice.

Please note, that an image size for prepress reproduction can vary from 7 MB to 350 MB.

** Please consider the chapter 2.1. in the -> Charges for photographic work.
Overview about the estimated cost - calculated for 20 pages
Price without quantity discount: 30.00 Euro - Price after 10% quantity discount: 27.00 Euro

Basic price: 27.00 Euro

Additional charges:

<table>
<thead>
<tr>
<th>Costs for dispatching</th>
<th>Postal charges</th>
</tr>
</thead>
<tbody>
<tr>
<td>USB stick</td>
<td>Mailing within Germany</td>
</tr>
<tr>
<td>5,00 Euro</td>
<td>3,00 Euro</td>
</tr>
<tr>
<td>DVD-ROM / piece</td>
<td>Mailing outside Germany</td>
</tr>
<tr>
<td>7,00 Euro</td>
<td>5,00 Euro</td>
</tr>
<tr>
<td>Internet-Download / per 500 MB</td>
<td>Airmail up to 500 g (=ca, 100 pages A4)</td>
</tr>
<tr>
<td>10,00 Euro</td>
<td>10,00 Euro</td>
</tr>
</tbody>
</table>

Extra efforts:
In case of special requests, extensive work or difficulties due to the special condition of the original document (e.g. aperture enlargements, pictures from the cut edge, heterogeneous assets, strict conservational guidelines) additional costs have to be paid. The hourly rate amounts to 50,00 Euro. In such cases, you will be informed in advance.

Express delivery charge:
A basic price of 30 Euro has to be payed for each order. The mentioned prices will double.
Due to operational reasons not all orders can be processed as express within 3 days.

Step 4: Order
ZEND = Zentrale Erfassungs- und NachweisDatenbank
[digital asset management system]

- Mapping of the entire production processes to a modular system
- Different service providers (scanning, text capture) can supply unlimited data to ZEND
- Workflow control
- Every object of the BSB, which will be digitized, follows the ZEND workflow
- Time and cost reduction through extensive automation
ZEND at a glance

1. Order
2. Digitised Object
3. Definitive file name
4. OAI
5. URN-Resolving (XEpicur)
6. ZEND
7. Administration of all metadata (bibliographic, technical, administrative)
8. OCR

- Digitisation on Demand
- Project-oriented Digitisation
- Conservation Purposes

Network of Bavarian Libraries
Catalog (OPAC)

Portals (BLO, ZvDD, Chronicon,...)

Search engines

Online Publication
Inhouse-only publication

Archival Storage
ZEND modules

Modules of the ZEND Workflow Tool

ZEND Administration
- ERATo
  Order Tool
- Order Management
- DaVeDi
  Workflow Database

ZEND Enduser Interface
- Viewer
- Search/Browse
  Title / Full Text
- RSS Feeds
- Subject Portals
- PDF-Download
  Order & Billing

ZEND Workflow Modules
- Image Conversion
- Catalog Enrichment
  Resolving Link
- Interfaces
  OAI, Z39.50
- Metsificator
  METS Export
- URN Generator
  Resolving Link
- Monitoring
  Access & Error Log
- OCR Processing
- Archival Storage
- Data Harvester
- Data Management
  via TSM-Client
- Repository Maintenance
- Retrieval from Archive

ZEND Data
- Database
  Object Information
  Metadata
- Repository
  Structural Information
  Full Text
- Archival Storage
  (TSM)
Scanning process – step by step

1. Transport of the original objects to the Scanning Center
2. Conservational check
3. Preparation for the scanning process
4. Image capture
5. Quality control
6. Indexing and OCR
7. Storage and digital long-term preservation
8. Publication on the web / Access
ZEND – login screen
Digitization order in ZEND

<table>
<thead>
<tr>
<th>Aug-16-2011 17:51:25</th>
</tr>
</thead>
</table>

### Digitalisierungsauftrag

- **Dienstleister:**
- **Falls nicht im Haus Name/Adresse:**
- **Signatur/Kurztitel/Verbund ID:**
- **Farbtiefe, Illumination:**
- **Seiten pro Image:**
- **Bildnachbearbeitung:** keine bes. Vorgabe
- **Besondere Hinweise für die Digitalisierung:**

---

Gabriele Messmer: Digitization Workflow - © Bayerische Staatsbibliothek 2011
## Digitization order in ZEND

<table>
<thead>
<tr>
<th>Objektbeschreibung</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Objektart:</strong></td>
</tr>
<tr>
<td><strong>Schrift:</strong></td>
</tr>
<tr>
<td><strong>Sprache(n):</strong></td>
</tr>
<tr>
<td>bulgarisch</td>
</tr>
<tr>
<td>deutsch (default)</td>
</tr>
<tr>
<td>deutsch [N. Rechtschreibung]</td>
</tr>
<tr>
<td>englisch</td>
</tr>
<tr>
<td><strong>Struktur:</strong></td>
</tr>
<tr>
<td>Inhaltsverzeichnis/Kap.überschriften</td>
</tr>
<tr>
<td>Register</td>
</tr>
<tr>
<td>Lemmata</td>
</tr>
<tr>
<td>erschließenswerte Bilder</td>
</tr>
<tr>
<td><strong>gew. Erschließungstiefe:</strong></td>
</tr>
<tr>
<td>OCR</td>
</tr>
<tr>
<td>OCR-Korrektur1</td>
</tr>
<tr>
<td>OCR-Korrektur2</td>
</tr>
<tr>
<td>Volltext-Dienstleister</td>
</tr>
<tr>
<td>Inhaltsverzeichnis/Kap.überschriften</td>
</tr>
<tr>
<td><strong>Phys. Form der Vorlage:</strong></td>
</tr>
<tr>
<td>Originaldokument</td>
</tr>
</tbody>
</table>
Import of the bibliographic metadata into ZEND via Z39.50 from the local catalogue system
Data import via Z39.50 into ZEND

Gabriele Messmer: Digitization Workflow - © Bayerische Staatsbibliothek 2011
... and at the same time generation of the definitive file name and allocation of an URN
Escursus: File name and URN

Creation of the definitive ZEND identification number …

➢ Example: bsb000019118_00001.tif

… and at the same time creation of the Unified Resource Name (URN)

➢ Example: urn:nbn:de:bvb:12-bsb00019118-6

Assignment is done locally in BSB, the resolving service is hosted by the German National Library

Resolving link:

http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:12-bsb00019118-6
Digitisation on Demand - Varia...

Bayerische Staatsbibliothek, Ludwigstr. 16, 80538 München - Abt. BA-ES-IDB - Witke

Digitalisierungsauftrag (Fotostelle)

Signatur : Bavar. 4406 v
- Sohlbad und klimatischer Höhenkurort Berchtesgaden im Bayerischen Hochland, 600 - 1100 m ü. d. M. - [1890]
urn:nbn:de:bvb:12-bvb00002232-1

Digitalisierungsparameter:

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

Lieferinformationen (Umfang: vollständig!):

| Ausdruck |
| Originalgröße | Vergrößert | Verkleinert |
| DVD |
| 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 __________________

---

ZEND – Process slip with barcode
Digitization parameters

Focus of in-house digitization
- Manuscripts
- Music manuscripts and printed music
- Old books (6th to 16th centuries)
- Rare and expensive books

Production parameters
- 24 bit
- High resolution
  (300 - 600 ppi in relationship to the originals’ size)
- ICC profiles
- TIFF uncompressed
Scanning speed depends on:

- Desired reproduction quality
- Preservation requirements
- Age and state of the original (difficult fixing)
- Format/binding of the original
Conservational requirements at BSB

Institute of Book and Manuscript Conservation (IBR)

The scanning devices have to follow the book requirements and not vice versa!

General requirements: lighting of the object and room climate

Specific requirement: use the scanning device most appropriate to the original work
Scanning devices of Munich Digitization Center

- 20 scanners (state of technology 2006-2010)
  - among them
    - 4 automated scanning devices
    - 1 thermographic scanner for watermarks
    - Hasselblad digital camera
Munich Scanning Center
Camera Work Table from Graz/Austria
Scanning by robots – the VD16 Project

Challenge: to scan old books with scan-robots

- Project cofinanced by the German Research Foundation
- Start: December 2007
- Equipment 2011: 4 scanrobots with each 800 pages/h (old books)

ScanRobot: EU „ICT“ Grand Prize Winner at the Cebit 2007
Scanning process

- Choice of most appropriate scanner
- Installing of a scan-job on the scanner PC
- Image capture
- Finalizing the scan job
- Moving the digitized images to a file called „Abholfach“
Quality control and approval

- Quality control
- Entering structural data
- Approval
- Return of the original work to the ordering library department or to the stocks
Automatic processing after scanning is finalized

- OCR processing of the images (optional)
- Creation of an index file for the web presentation 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)
- Storage of all the data in the Leibniz Supercomputing Centre for long-term preservation
ZEND: XML-ToC(Table of Contents) editor

- Quality control and correction (image size & orientation, etc.)
- Flagging the structural elements of the document (i.e. frontispiece, chapter titles, images etc.) - allows navigation inside the digital object
- Subsequent "activation" and instant availability on the WWW.
ZEND: ToC editor
Paulus <de Venetiis>: Summulae logicae, Venedig, 1500.10.16 [BSB-Ink P-111 - GW M30356]
Finished version for Web presentation
Immediate availability inside the local catalogue (OPAC)
Archival storage of the data

Long-term Preservation

The Digital Library of the Bavarian State Library has been engaged in long-term preservation activities since 1999. In cooperation with the Leibniz Computing Center, it has worked on a comprehensive organisational and technical infrastructure for long-term archiving since 2004. In the context of a two-year pilot project, a comprehensive new infrastructure was developed. It allows an efficient and largely automated ingest and archiving of the data as well as access to the data via existing online catalogues and subject gateways. Fortunately the times of the "archive" consisting of an ordinary wooden locker with several hundred CDs and DVDs are history since 2004.

Together with the Leibniz Computing Center, the Digital Library runs one largest electronic archives in Germany. At present (October 2007), the archive contains an overall amount of 45 terabytes of data (more than 22 million pages). Redundant storage and regular migration measures guarantee the long-term preservation of the data.
Digitization workflow – archiving report

Archivreport: bsb00047184

| Normal File--> | 851 | /sub/db/public_html/0004/bsb00047184/images/bsb00047184_kopf.xml [Sent] |
| Normal File--> | 9,667 | /sub/db/public_html/0004/bsb00047184/images/bsb00047184_stru.xml1312187602 [Sent] |

Archive processing of '/sub/db/public_html/0004/bsb00047184/images/*' finished without failure.

| Total number of objects inspected: | 220 |
| Total number of objects archived: | 220 |
| Total number of objects updated: | 0 |
| Total number of objects rebound: | 0 |
| Total number of objects deleted: | 0 |
| Total number of objects expired: | 0 |
| Total number of objects failed: | 0 |
| Total number of bytes inspected: | 2.88 GB |
| Total number of bytes transferred: | 2.88 GB |
| Data transfer time: | 12.55 sec |
| Network data transfer rate: | 240,540.29 KB/sec |
| Aggregate data transfer rate: | 37,467.78 KB/sec |
| Objects compressed by: | 0% |
| Total data reduction ratio: | 0.00% |
| Elapsed processing time: | 00:01:20 |

| Normal File--> | 162 | /sub/db/public_html/UU04/bsb00047184/images/index.html [Sent] |
| Normal File--> | 1,397,698 | /sub/db/public_html/0004/bsb00047184/images/bsb00047184_sig.tif [Sent] |
| Normal File--> | 225,032 | /sub/db/public_html/0004/bsb00047184/images/profil.icc [Sent] |
| Normal File--> | 34,519,510 | /sub/db/public_html/0004/bsb00047184/images/bsb00047184_00001.tif [Sent] |
| Normal File--> | 35,446,846 | /sub/db/public_html/0004/bsb00047184/images/bsb00047184_00002.tif [Sent] |
| Normal File--> | 28,428,094 | /sub/db/public_html/0004/bsb00047184/images/bsb00047184_00003.tif [Sent] |
Overview: Document Processing Steps

1. Scanning and Document Import:
   - Scanning: Images are scanned by an operator on the Scanning Station and then sent to Recognition Server.
   - Document Import: The Server Manager imports files from the designated input source (i.e. shared folder, FTP folder, or mailbox folder) and places them in a processing queue.

2. Recognition: The system evenly distributes files among the available Processing Stations for recognition. Workflow processing on all available Processing Stations is administered and monitored by the Server Manager. Both components function as Windows® services, can be installed separately in the network or on the same computer.

3. Verification (optional): In case of low quality document input or when precise accuracy is required, an optional Verification Station can be set up for manual proofing by an operator. The Server Manager routes documents to the Verification Stations according to workflow settings.

4. Document Separation: After recognition and verification, documents can be sorted using blank page, barcode, or fixed page number separators. Scripts may also be used to create custom separation rules.

5. Classification and Indexing (optional): Organisations may choose to index manually on the Indexing Station, or index automatically using customised scripts which enable intelligent export and delivery using document properties and metadata.

6. Export: Following recognition and verification, the Server Manager delivers the output document to its destination, which can be a network folder, a SharePoint® library, an e-mail address or applications which use the Recognition Server API.

Cited from:
http://www.abbyy.de/recognition_server/brochure_en/
## Objektbeschreibung

<table>
<thead>
<tr>
<th>Objektart:</th>
<th>------------------</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schrift:</td>
<td>---</td>
</tr>
</tbody>
</table>

### Sprache(n):
- bulgarisch
- deutsch (default)
- deutsch (N. Rechtschreibung)
- englisch

### Struktur:
- Inhaltsverzeichnis/Kap.überschriften
- Register
- Lemmata
- erschließenswerte Bilder

### gew. Erschließungstiefe:
- OCR
- OCR-Korrektur1
- OCR-Korrektur2
- Volltext-Dienstleister
- Inhaltsverzeichnis/Kap.überschriften

### Phys. Form der Vorlage:
- Originaldokument
Historiographic metafiction in modern American and Canadian Literature

Autor: Engler, Bernd
Verlagsort: Paderborn
München u.a.
Verlag: Schöningh
Erscheinungsjahr: 1994
Signatur: 94.46161

Challenge OCR output
Fulltext search in books digitized by Google
Contact

messmer@bsb-muenchen.de