

Klaus Kempf

L'organizzazione e conservazione del digitale in biblioteca: l'esperienza bavarese



The Bavarian State Library

Some facts & figures

- 780 employees
- ~ 10,1 million volumes
- ~ 130,000 incoming volumes p.a.
- more than 63,000 current periodical titles
- 24,147 current e-journals (licensed), 680 databases
- 1,2 Mio licensed e-books
- 1,08 digitised books



Long-term Preservation Milestones

- **1999** First Long-term preservation project
- **2004** Switch from CD to tape storage (Leibniz Supercomputing Centre)
- **2005** Implementation of the *BABS* Library Archiving and Access System
- **2006** First successful hardware migration
- **2007** Start of the Google-Project
- **2009** Rosetta-Partnership with ExLibris
- **2012** Rosetta goes live at BSB



Present Preservation Responsibilities

Digital Objects from the (Mass-)Digitisation Programme

- ~1.08 mio. volumes, >300 mio. Pages

Born-digital materials

- Electronic journals (->Hosting strategy)
- Databases
- E-books
- Legal Deposit of Public authorities
(legal obligation since 01/2009)
- Deposit of regional publishers
(voluntary)
- Open Access: BSB's special collections fields



Lots of Digitisation Activities

Maps **16th Century Books** Abwehrblätter

Newspapers Biographic 100(0) Key documents on German

History Literary & Scientific Remains **Latin**

Manuscripts Incunabula **Digi20** German

Manuscripts Exhibition Catalogs Monumenta Germaniae

Historica **Reichstag Session Protocols**

100(0) Key Documents on Russian History **Encyclopaedia**

Google Music Supplies **Digitisation**

on Demand Block Books Incunabula Journals on

Bavarian History **Bavarica**



Lots of Legal Deposit Activities

eBooks Biographic Databases Online
Dissertations Sound Storage Media **Scientific**
Databases Digitised Maps Subject Portals Bavarian
Official Governmental Gazettes Electronic Newspapers
Scientific Journals Photographic
Collections **Web Resources** Literary
& Scientific Remains **Digitised**
Books



Long-term Preservation: Three intersecting Spheres of Activity

Routine operation of the existing infrastructure

- BABS Library Archiving and Access System

Further development and consolidation of the required structures

- Profile and concepts, workflows
- Organisation
- Technology
- Rosetta Project (to be mentioned later on)
 - Technical support + library-related business

Research and development

- Evaluating future technologies through innovative projects



LRZ: The backbone of BSB's LTP strategy

The Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities

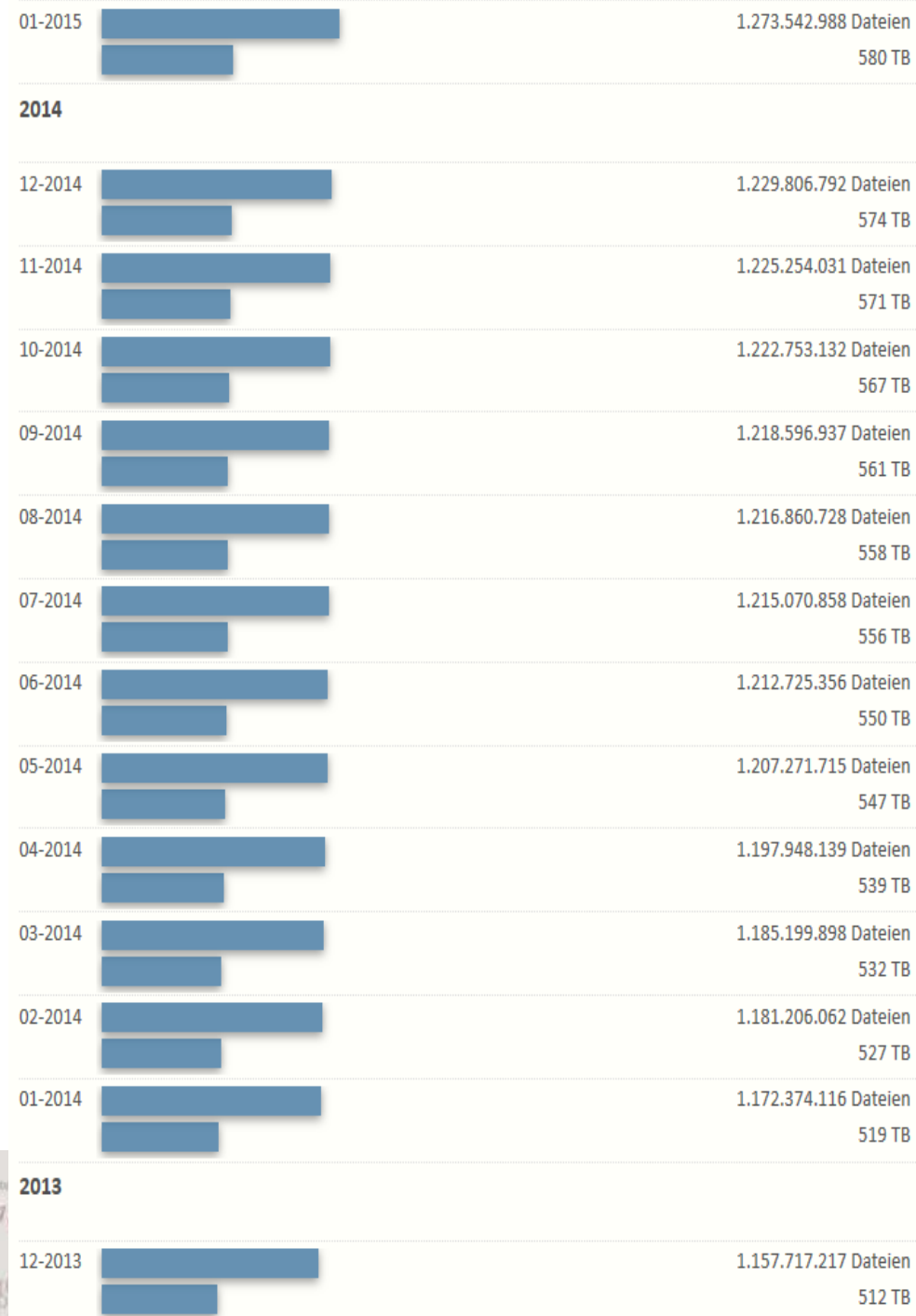
- IT-services for the university and research institutions in Munich
- Part of Munich Scientific Network
- Backup and long-term preservation
- High Performance Supercomputing Centre for all German universities



Status Quo 01/2015

→ Growing electronic archive

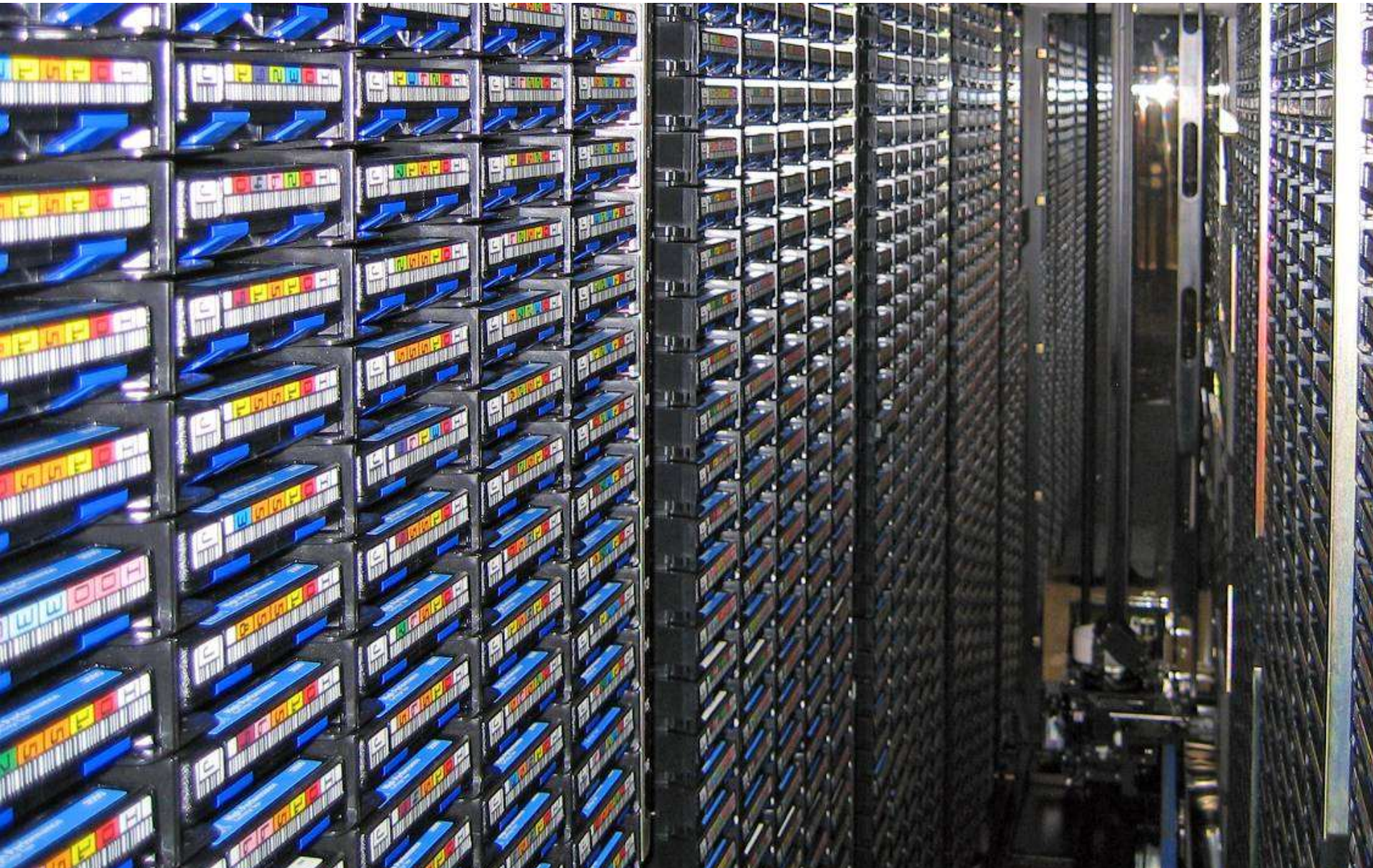
- 1,27 Billion Files
- 580 TB
- + ca.3- 8 TB/month



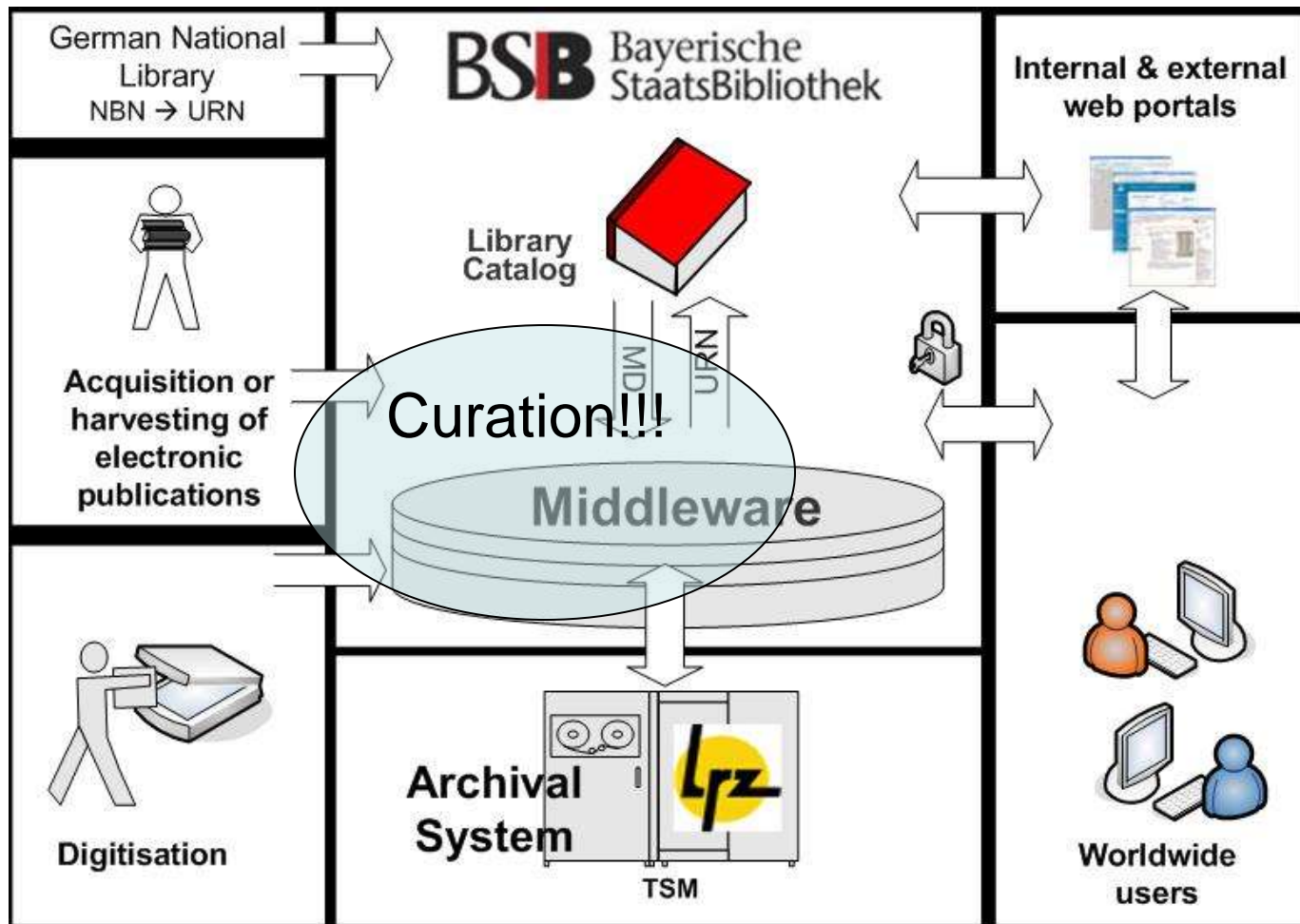
Shelfie 1: In the age of print



Shelfie 2: In the digital age



How it is done: Existing Architecture (BABS)



BSB's Preservation Philosophy for curated data

- Increasing importance of selection in the digital age
- Well defined workflows and processes
- Legal basis (law vs. license/contract) for archiving has to be clear
- Adoption of international standards and specifications
 - Open, widespread und well documented formats for all digital objects and metadata
- Permanent access for all users (not a dark archive)
- Combine expertise through cooperation
 - Strategic partnerships and projects

What a curated AIP needs

AIP= Archival Information Package

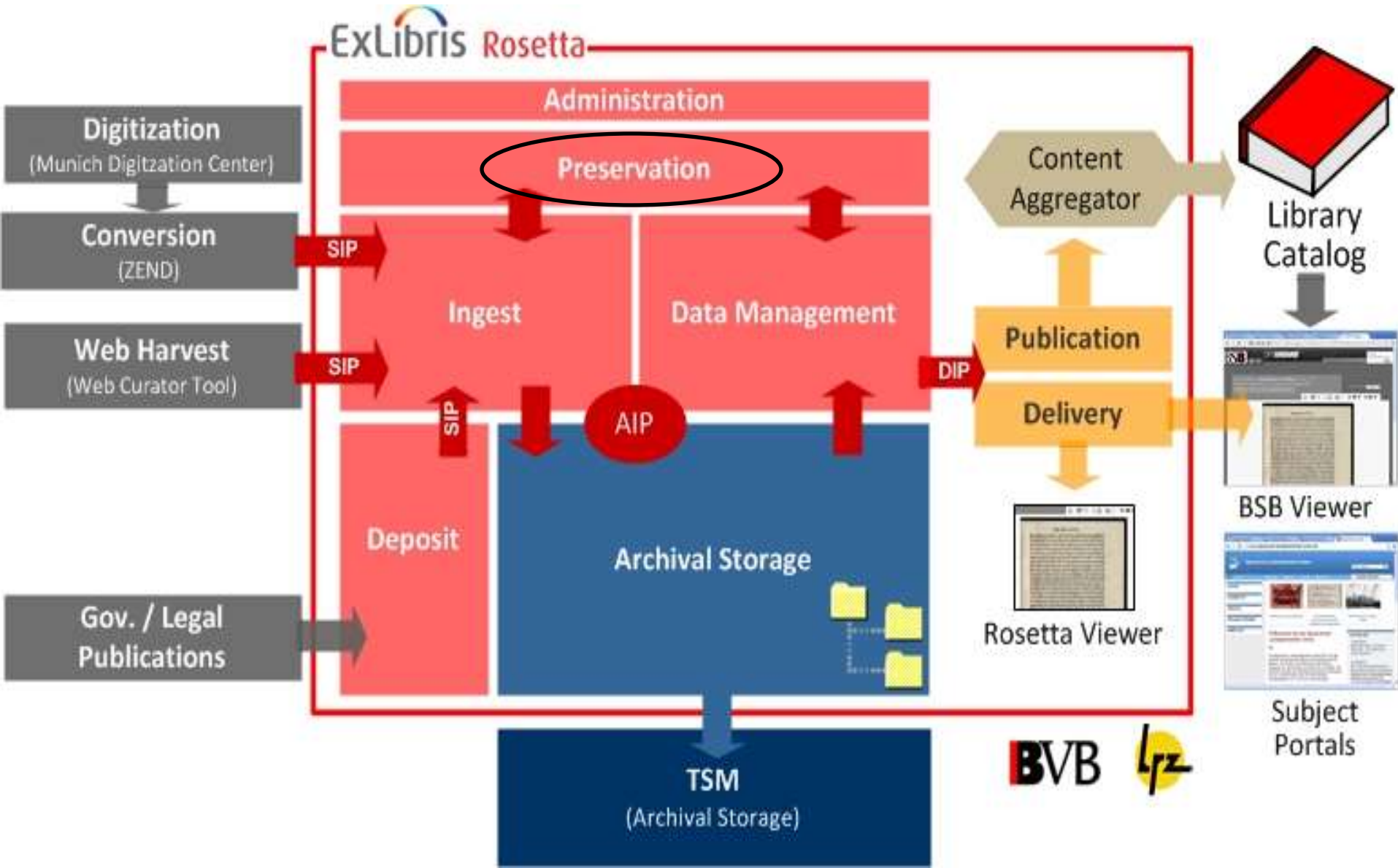
- Content (files)
- Bibliographic metadata
- Metadata needed for Preservation
 - Persistent identifiers
 - Technical information
 - Structural metadata
 - Provenance
 - Documentation of all changes
 - Documentation of legal status and access rights



Example: Curation Decisions for retrodigitization

- Do it once, do it right!
- Images in stable and widely used formats (tif, jp2, jpeg)
- XML for structural metadata
- Use of persistent identifiers for every object: URNs
- All Images and structural metadata are saved (redundant storage)
- Quality Control and Checks for Integrity (checks)
 - Format Validation with JHOVE
- Preparation and occasional testing of migration scenarios

What's new: Rosetta at BSB: System architecture for unified preservation



Rosetta's Preservation Module

The screenshot displays the ExLibris Rosetta Management web application. The browser window title is "Preservation". The page header includes the ExLibris Rosetta logo and navigation links: Home, Producers, Submissions, Data Management, and Preservation (highlighted). The user is identified as "WK Thomas" with links for Help and Logout.

The main content area is titled "Home > Preservation" and contains several sections:

- Risk Analysis** (circled): View the Format Risk Report, search the Repository
 - View Format Risk Reports
 - Manage Preservation Sets
 - Manage Exploratory Sets
- Preservation Plans** (circled): Manage Preservation Plans
 - List of Preservation Plans
 - Evaluate Test Results
 - Handle Technical Issues
- Preservation Executions** (circled): Execute and monitor Preservation executions
 - List of Signed-off Plans
 - Handle Technical Issues
- Advanced Preservation Activities** (circled): Manage the Format, Application, Risk and Classification Libraries and run reports
 - Manage Global Libraries
 - Manage Local Libraries
 - Run Reports
 - Configure Settings

The right sidebar contains:

- Quick Links**
 - Technical Issues - Loading
 - Arranger Work Area
 - User List
 - Technical Issues - Validation
 - Technical Issues - Deposit
 - Assessor Work Area
- Search Rosetta**: Find:
- Top Three Formats at Risk**

Risk	Format Name	Format ID	Files
old_version_of_format	Tagged Image File Format	fmt/7	1457
NON_ARCHIVE_PDF_18	Portable Document Format	fmt/18	39
PDF fmt/18	Portable Document Format	fmt/18	39

[More...](#)
- What else can I do?**
 - Deposit material
 - Administer the system

Rosetta: Still Challenges...

- Scalability
 - Original goal: 1000 Books a Day (for large scale migration of digitised objects) could not yet be reached, but BSB's daily production could be processed
- Improvement of viewers
 - Hierarchical collections (e.g. e-journals)
 - Multimedia objects
- Better integrated support for research data
- Documentation of the system

Long-term Preservation at the BSB

Overall Objectives

- To be a trusted partner in a cooperative solution for the preservation of the digital heritage of Germany
- To preserve digital media from different sources
- To keep everything accessible and available
(= No dark archive)



Speichersystem



Grazie per l'attenzione!
Kontakt:klaus.kempf@bsb-muenchen.de