

Rosetta DPS

Implementierung und Einsatz an der Bayerischen Staatsbibliothek

Matthias Groß
Bayerische Staatsbibliothek München
gross@bsb-muenchen.de

Langzeitarchivierung - Materialien

Digitalisate Biographische Datenbanken Online-

Dissertationen Tonträger **Wissenschaftliche**

Datenbanken Fachportale Amtsblätter Elektronische Zeitungen

Wissenschaftliche Zeitschriften Digitalisierte

Fotosammlungen **Internet-Quellen** Nachlässe

eBooks

Langzeitarchivierung: Fokus

- lokale/ regionale/ nationale Verantwortung
- Materialfokus: breites Spektrum
 - Eigene Produktion: Digitalisate, ZEP
 - Kooperationsprojekte
 - Bayerische AD, Bundes-AD (Erlass zur Abgabe amtlicher Veröffentlichungen an Bibliotheken)
 - Bayerische Pflichtablieferung (Gesetz in Novellierung)
 - Lizenzierte Materialien (lokal, regional, national)
 - SSG: Wissenschaftlich relevante Open-Access-Publikationen & Websites

Digitale Langzeitarchivierung an der BSB: Historisches

Münchener Digitalisierungszentrum/Digitale Bibliothek

- 1999: Erstes DFG-Projekt zur digitalen Langzeitarchivierung von Bibliotheksmaterialien
- 2004: Erste Migration des CD-Archivs (3000 CDs des MDZ) auf das Magnetbandarchiv des LRZ: Start der Kooperation
- 2005: Implementierung des Bibliothekarischen Archivierungs- und Bereitstellungssystems *BABS* mit dem LRZ
- 2006: Zweite erfolgreiche Datenmigration des MDZ-Tapearchivs (Hardware-Migration)
- 2007: Start des Google-Projekts: rasanter Datenzuwachs
- Einsatz von Scan-Robotik

Langzeitarchivierung: Größere Aktivitäten

- BABS – Bibliothekarische Archivierungs- und Bereitstellungssystem
 - Konzept und Setup (2005-2007)
 - Selbst-Evaluierung & Skalierbarkeit (2009-2010)
- Mitwirkung bei Nestor: Informationsnetzwerk zur Langzeitarchivierung (seit 2003)
- LOCKSS: Mitwirkung an einem weltweiten Pilotprojekt mit der LoC & weiteren Partnern



Das Rückgrat der BSB für Produktion & digitale Langzeitarchivierung

ist das **Leibniz-Rechenzentrum** der Bayerischen Akademie der Wissenschaften



- IT-Services für die Forschungsstätten in München
- Betreiber des Münchener Wissenschaftsnetzes
- Supercomputing-Zentrum für alle deutschen Universitäten

Infrastruktur am LRZ

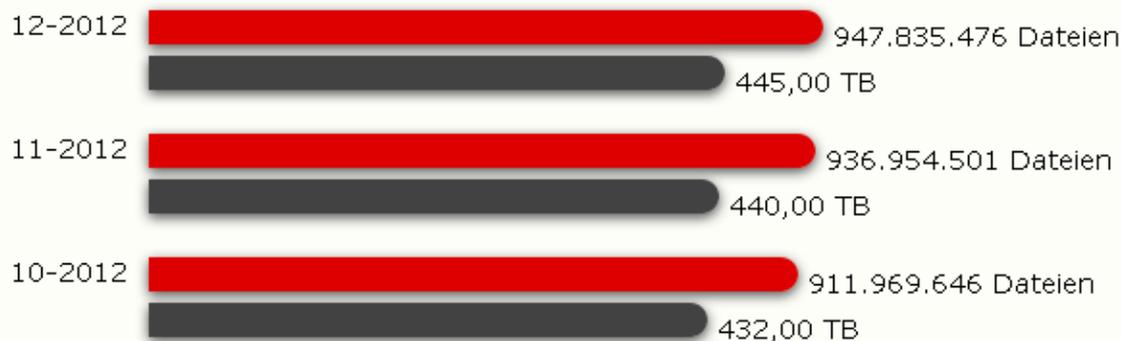
- NAS - Online-Storage für Produktion & WWW-Bereitstellung
 - Von dort in 2012: 794.900 PDF-Downloads ganzer Werke (= 76 TB Downloadvolumen)
- Archivspeicher: 476 Terabyte mit über **1 Milliarde Dateien** (Ende März 2013)

Entwicklung des BSB/MDZ-Archivspeichers am LRZ

2013



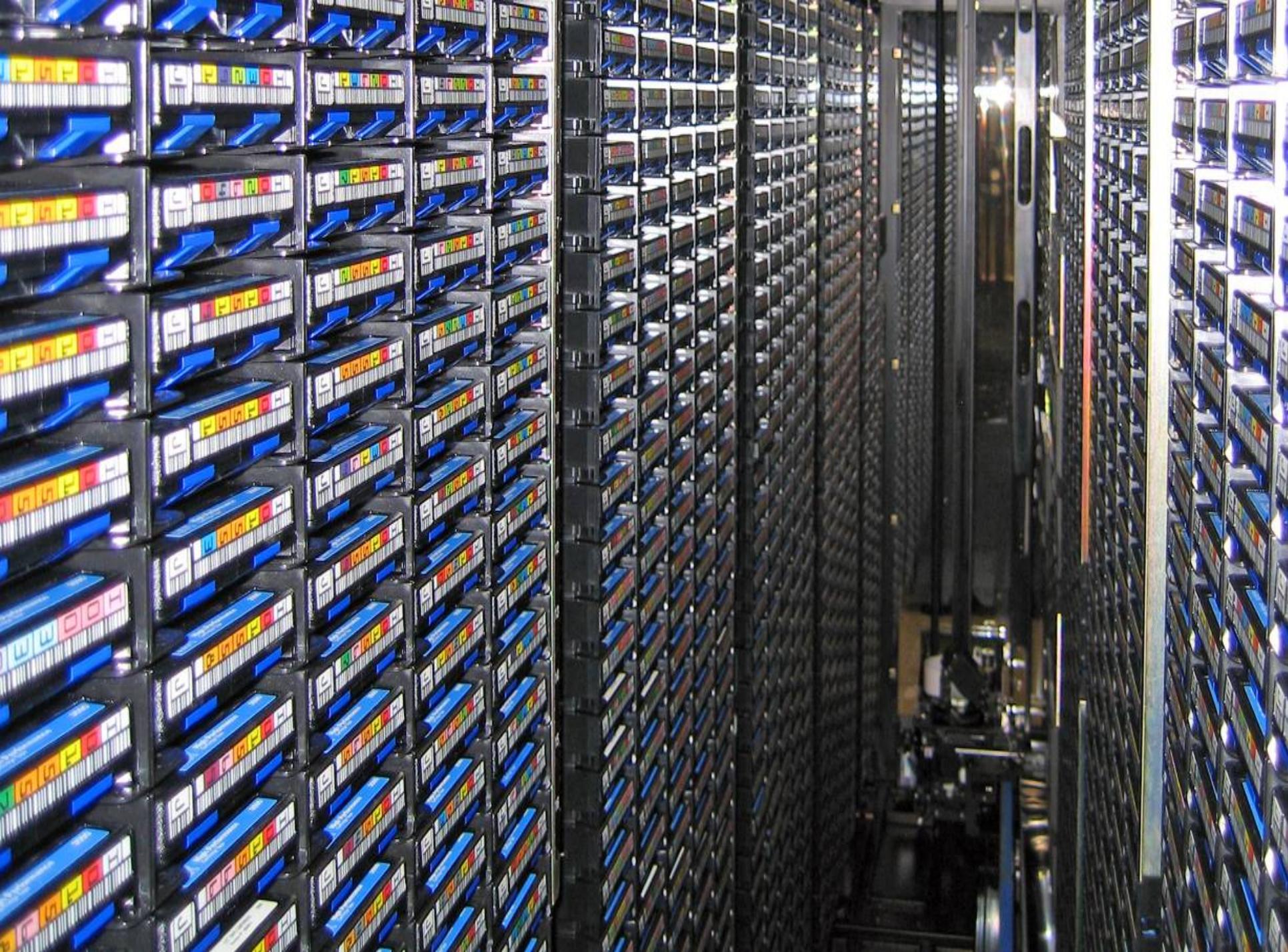
2012



Eine Milliarde Dateien im Archiv

Seit März 2013 umfasst das Langzeitarchiv der BSB exakt 1.034.200.065 Dateien. Die archivierte Datenmenge beträgt nun 476 TB [mehr](#)

[08.04.2013]



Herausforderung Datenvolumen

- **Durchsatz**
 - Datenpuffer zur Zwischenspeicherung
 - Schnelle Datenleitungen; Firewall
- **Performanz**
 - Skalierbarkeit (Multi-Server-Umgebung)
 - Redundanz & Ausfallsicherheit
 - Backup
- **Verfügbarkeit** (Retrieval)
 - Strukturierung der Datenablage (→ Archiv-“Knoten“)
 - Stetige Zugänglichkeit (Archiv-Rückholung alltäglich)

Rosetta an der BSB: Warum?

Wichtigste Kriterien:

- Volle OAIS-Compliance
 - Preservation Planning and Action
 - Kein reines „dark archive“: Direkter Nutzerzugang möglich
 - Rechteverwaltung
- Open-Platform-Strategie
 - Programmierschnittstellen: API, SDK
 - Standards: OAIS, PREMIS, METS, DC, OAI, SRU ...

Projektphasen, Team

- Pilotphase:
 - BSB (Projektleitung)  Bayerische Staatsbibliothek
Information in erster Linie
 - Münchener Digitalisierungszentrum 
 - Virtuelle Bibliothek Bayern (BVB-Verbundzentrale) 
 - Ex Libris 
 - Leibniz-Rechenzentrum 
- Ausbauphase: Tests mit
 - UB Augsburg
 - UB Regensburg
 - UB Würzburg

Rosetta an der BSB: Pilotphase (2010-2012)

- Installation des Systems
- Schulung der Mitarbeiter
- Implementierung von drei Pilot-Workflows
- Integration in die existierende Infrastruktur: Schnittstellen zu Digitalisierungsstraße/ZEND, Webarchivierung, Katalog, Speichersystem)
- Probebetrieb
 - Konnektivität, Datenverarbeitung
 - Multiserver-Umgebung
 - Test der Mandantenfähigkeit von Rosetta
 - Skalierungs- & Lasttests
- Beteiligung an der *Rosetta Advisory Group*

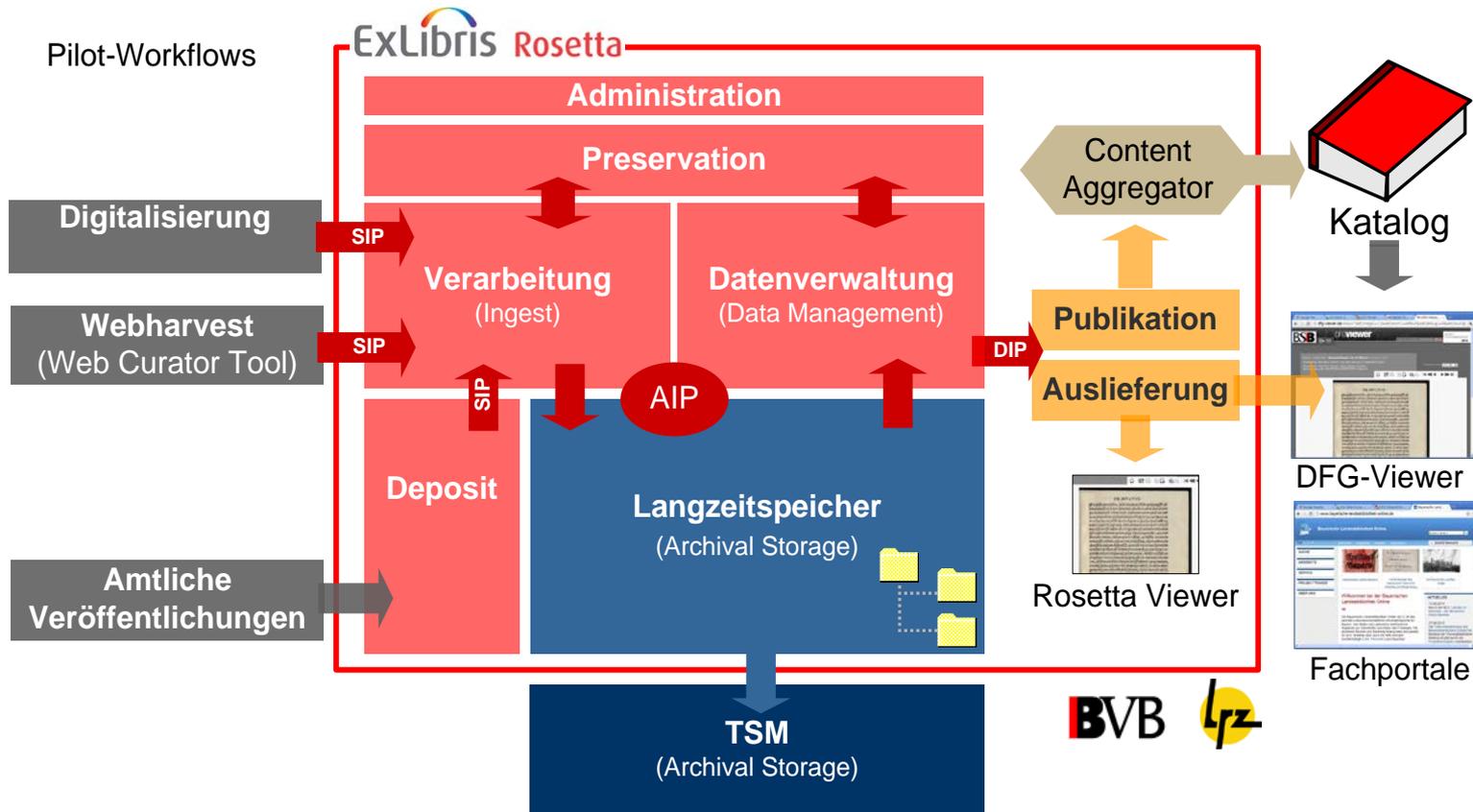
Rosetta an der BSB: Livegang

- Sommer 2012: Migration auf Rosetta Version 3
- seit Ende September 2012: Echtbetrieb für die drei Pilotworkflows
- bis Januar 2013: Aufbau einer Mehrserver-Testumgebung
- seit Februar 2013: Tests mit UB-Pilotworkflows

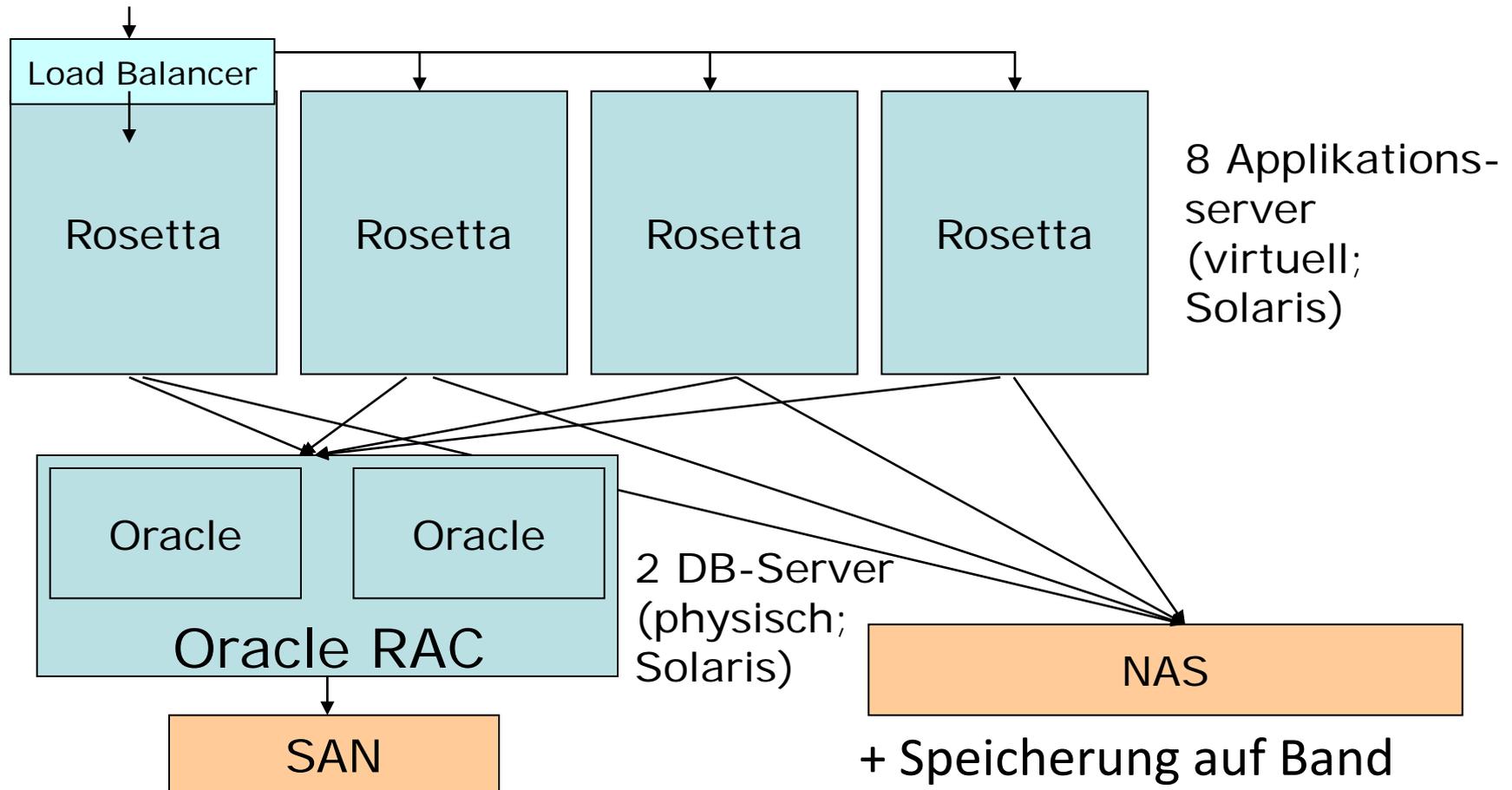
Pilot-Workflows

- Digitalisate
- Born-Digitals
 - aus elektronischer Pflichtablieferung (Amtsdruckschriften) & Sammlung
 - zunächst nur Monographien
- Web-Archive
 - aus Sondersammelgebieten der BSB

Rosetta: System-Architektur an der BSB



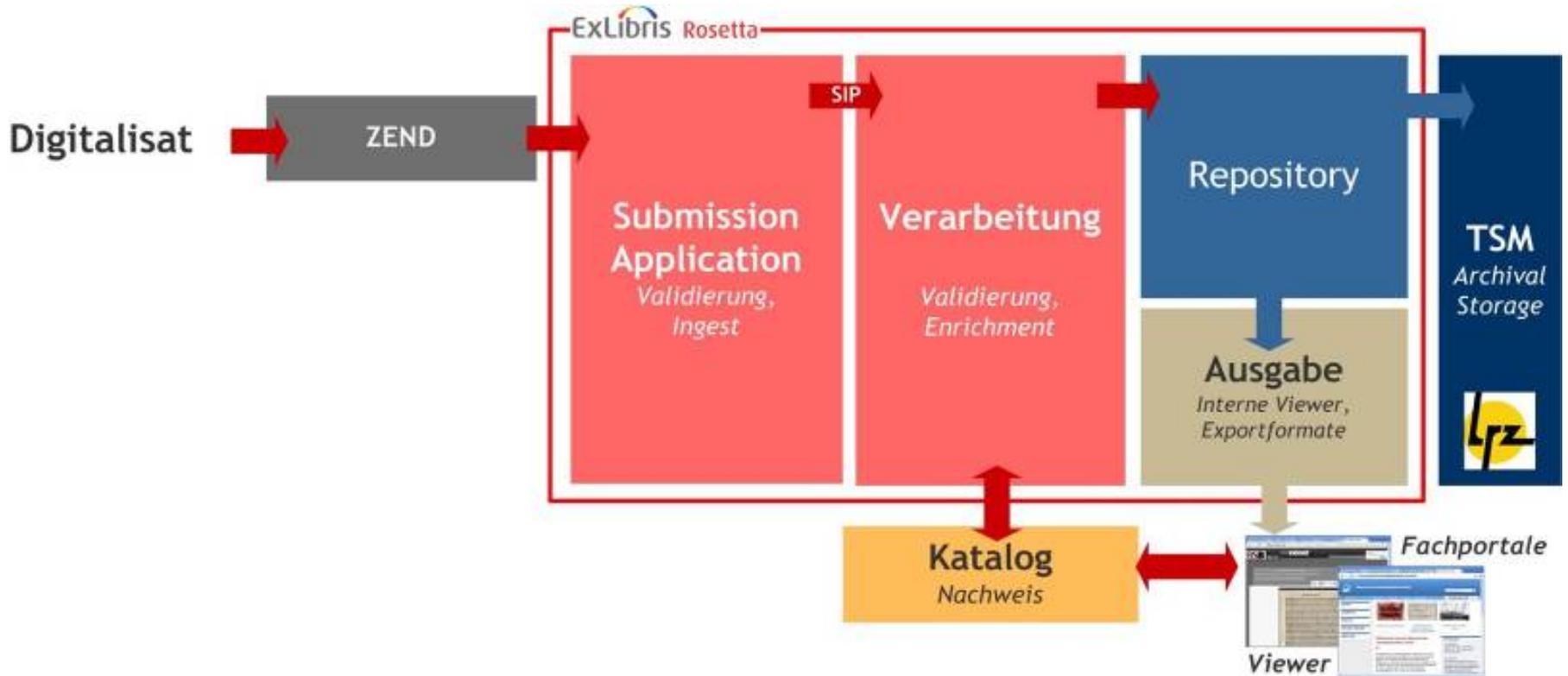
Rosetta: Serverarchitektur



Workflow #1: Digitalisate

- Erste Kollektion, die mit Rosetta bereitgestellt wird: Griechische Handschriften (Cod.graec.)
 - laufendes DFG-gefördertes Digitalisierungsprojekt
 - Masterformat TIF, Präsentationsformat JPG
 - METS-Strukturdaten (in der Regel nur einfache Blätterlösung)
 - Ingest über Submission Application
 - Verknüpfung zu Handschriften-Kurzaufnahmen in Aleph über Verbund-ID
 - Bereitstellung aus Rosetta über externe Viewer, u.a. DFG-Viewer

Workflow #1: Digitalisate



Workflow #2: Elektronische Pflichtablieferung

Erfassen wesentlicher Metadaten

BSB Bayerische Staatsbibliothek Deposit Producer: Bayerisches Landesamt für Umwelt | User: Jutta Wunder | Help | Logout

Home

Descriptive Information 1 2 3 4 5

Codierte Angaben zum Datensatz in Aleph (Zielfeld: MAB 030_\$_)

* Sprachencode gemäß ISO 639-2 (Zielfeld: MAB 037b_\$a)

Materialcodes zum Datenträger (Zielfeld: MAB 050_\$_)

Veröffentlichungsspezifische Angaben zu begrenzten Werken (Zielfeld: MAB 051_\$_)

Ansetzungsform der 1. Person für Haupteintragung nach (Zielfeld: MAB 100_\$a)

Ansetzungsform der 1. Person für Nebeneintragung (Zielfeld: MAB 100b_\$a)

Ansetzungsform der 1. Körperschaft für Nebeneintragung (Zielfeld: MAB 200b_\$a)

* Hauptsachtitel in Vorlageform (Zielfeld: MAB 331_\$a)

Zusatz zum HST (Zielfeld: MAB 335_\$a)

Verfasserangabe (Zielfeld: MAB 359_\$a)

Ausgabebezeichnung (Zielfeld: MAB 403_\$a)

* Ort 1, Verlag (Zielfeld: MAB 410_\$a)

Name 1, Verlag (Zielfeld: MAB 412_\$a)

* Erscheinungsjahr in Vorlageform (Zielfeld: MAB 425_\$a)

* Umfangsangabe (Zielfeld: MAB 433_\$a)

Illustrationsangabe (Zielfeld: MAB 434_\$a)

ISBN formal richtig (Zielfeld: MAB 540a_\$a)

* Erg. Bemerkung Computerdatei, Bestandsschutz (Zielfeld: MAB 659_\$a kostenfrei und \$b Langzeitarchivierung durch die Bayerische Staatsbibliothek)

Selektionskennzeichen für Ebooks (MAB 078e_\$a)

Bibliotheksspezifisches Selektionskennzeichen (Zielfeld: MAB 078q_\$a)

[< Back](#) [Cancel](#) [Next >](#)

Workflow #2: Elektronische Pflichtablieferung

Nutzungsrechte

BSB Bayerische Staatsbibliothek Deposit Producer: Bayerisches Landesamt für Umwelt | User: Jutta Wunder | Help | Logout

Home

 Assertion of Copyright 1 2 3 4 5

Access Rights

 No restrictions

Agreement

Als alleiniger Besitzer der Nutzungsrechte bzw. Bevollmächtigter/gesetzlicher Vertreter des Besitzers/der Besitzer der Nutzungsrechte für die elektronische Publikation räume ich hiermit der Bayerischen Staatsbibliothek das unbefristete Recht ein, eine Kopie dieser Publikation in ihrem in Kooperation mit dem Leibniz-Rechenzentrum betriebenen elektronischen Langzeitarchiv zu speichern und für einen unbefristeten Zeitraum zu archivieren.

Alle dafür erforderlichen Rechte der Vervielfältigung, elektronischen Speicherung und Bearbeitung räume ich ein. Insbesondere ist gestattet, die Daten in andere Datenformate zu konvertieren oder umzubenennen, falls der Archivierungsprozess es erfordert. Nicht gestattet sind inhaltliche Veränderungen an den Daten. Die Archivierung erfolgt auf Kosten der Bayerischen Staatsbibliothek. Diese ist berechtigt, nicht aber verpflichtet, die Langzeitarchivierung durchzuführen.

Die Bayerische Staatsbibliothek wird die Archivkopie katalogisieren. Die Bayerische Staatsbibliothek ist berechtigt, ihren Bibliotheksnutzern die Archivkopie ohne Benutzungseinschränkungen frei im Netz über die Bayerische Staatsbibliothek zur Verfügung zu stellen.

Ich versichere hiermit, dass dieser Archivierungs- und Nutzungsbewilligung keine Rechte Dritter entgegenstehen. Für den Fall, dass Dritte behaupten, Inhaber entgegenstehender Rechte zu sein, werde ich etwaige Schadensersatzansprüche abwehren oder befriedigen und die Bayerische Staatsbibliothek insofern freistellen.

I accept the terms

[< Back](#) [Cancel](#) [Next >](#)

© Ex Libris Ltd., 2012 Terms of Use

Workflow #2: Elektronische Pflichtablieferung

BSB Bayerische Staatsbibliothek Deposit Producer: Bayerisches Landesamt für Umwelt | User: Jutta Vunder | Help | Logout

Home

Summary of Deposit (1) (2) (3) (4) (5)

Deposit Activity ID: 1521
Material Type: Material_Flow_Monograph

Descriptive Information

<p>Codierte Angaben zum Datensatz in Aleph (Zielfeld: MAB 030_ \$_) az1dcr 7</p> <p>Materialcodes zum Datenträger (Zielfeld: MAB 050_ \$_) g </p> <p>Ansetzungsform der 1. Person für Haupteintragung nach (Zielfeld: MAB 100_ \$a) -</p> <p>Ansetzungsform der 1. Körperschaft für Nebeneintragung (Zielfeld: MAB 200b \$a) Bayerisches Landesamt für Umwelt</p> <p>Zusatz zum HST (Zielfeld: MAB 335_ \$a) das Bayerische Landesamt für Umwelt</p> <p>Ausgabebezeichnung (Zielfeld: MAB 403_ \$a) aktualis. Neuauf.</p> <p>Name 1, Verlag (Zielfeld: MAB 412_ \$a) LIU</p> <p>Umfangangabe (Zielfeld: MAB 433_ \$a) 28 S.</p> <p>ISBN formal richtig (Zielfeld: MAB 540a \$a) -</p> <p>Selektionskennzeichen für Ebooks (MAB 078e \$a) ebook</p>	<p>Sprachencode gemäß ISO 639-2per (Zielfeld: MAB 037b \$a)</p> <p>Veröffentlichungsspezifische Angaben zu begrenzten Werken (Zielfeld: MAB 051_ \$_) m </p> <p>Ansetzungsform der 1. Person für Nebeneintragung (Zielfeld: MAB 100b \$a) -</p> <p>Hauptsachtitel in Vorlageform (Zielfeld: MAB 331_ \$a) Wasser, Boden, Luft, Natur</p> <p>Verfasserangabe (Zielfeld: MAB - 359_ \$a)</p> <p>Ort 1, Verlag (Zielfeld: MAB 410_ Augsburg \$a)</p> <p>Erscheinungsjahr in Vorlageform (Zielfeld: MAB 425_ \$a) 2012</p> <p>Illustrationsangabe (Zielfeld: MAB 434_ \$a) III, Kt.</p> <p>Erg. Bemerkung Computerdatei, Langzeitarchivierung durch die Bayerische Bestandschutz (Zielfeld: MAB 659_ \$a kostenfrei und \$b Langzeitarchivierung durch die Bayerische Staatsbibliothek)</p> <p>Bibliotheksspezifisches Selektionskennzeichen (Zielfeld: MAB 078q \$a) BSBROSUZAD</p>
---	--

Access rights

Group: No restrictions - No restrictions

Selected files

File Name	Label	Note	Source	Status	Size	Date
1 ifu_all_00038_2012.pdf				VALID	5,572.93	

< Back Cancel Save as Draft Submit Deposit

© Ex Libris Ltd., 2012 Terms of Use

Workflow #2: Elektronische Pflichtablieferung

Verknüpfung mit Aleph (Variante: Titelaufnahme bereits vorhanden)

BSB Bayerische Staatsbibliothek Management

User: Jutta Wunder | Help | Logout

Home Producers Submissions Data Management Preservation Quick Launch Site Map

Home > Search External Repositories

Search Database: BVB01 RefreshIndexes

Find: BV040575340

Search Category: dc.identifier

GO

1 - 1 of 1 Records

	ID	Title	Creator	Created on
1	BV040575340	Messstellen Grundwasserstand in Bayern staatliches Grundnetz	Bayern Landesamt fü...	2009 2009

Cancel Link

© Ex Libris Ltd., 2012 Terms of Use

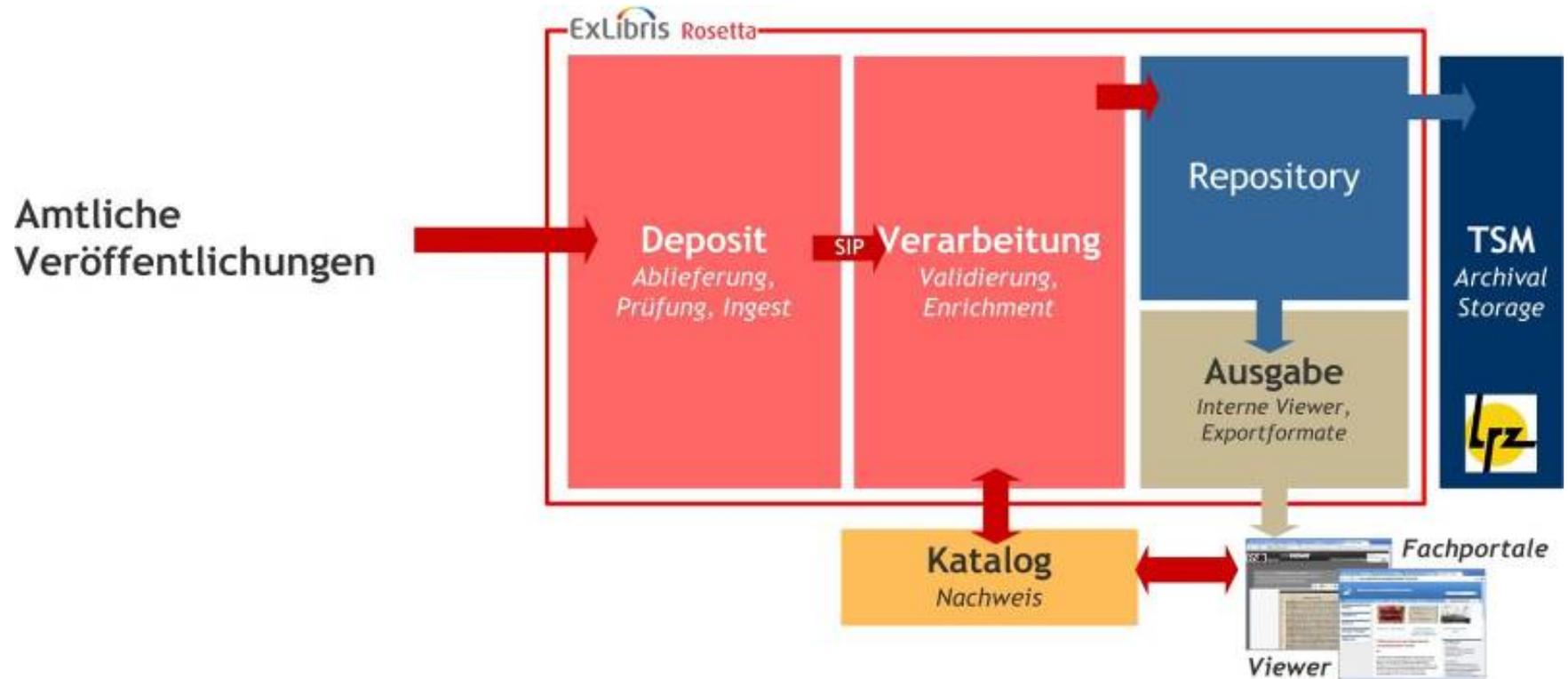
Workflow #2: Elektronische Pflichtablieferung

- „Hochkatalogisieren“

The screenshot shows the ALPHABIS cataloging software interface. The title bar indicates 'Katalogisierung ALPH - Version 20.01' and the user is '85801226'. The main window displays a record for 'MH Systemnr. 18902924 Wölfel, Roland (DE-588)109210366 (Praxistipps zur Standortbelegung), 2006'. The left sidebar shows a tree view with 'BVB01-18902924' selected. The main area displays a list of fields and their values:

Field	Value
Satzkennung	LDR -- 00000rM2.01200024-----h
ID Datensatz	001 a BV036092343
Dat. Erf.	002 a 20100323
Dat. Kor.	003 a 20110801
Überreg. ID	025 o a 6440B236B
Regionale ID	026 a BVBBV036092343
Datensatzcodes	030 a a 1dazz 37
Sprachencode	037 b a ger
Materialcodes	050 - 11111g 1111
Codes Begr.W.	051 - m
Codes eLRes.	058 - cr 1111111111
ID Bearb. Inst.	070 a DE-12
ID Bearb. Inst.	070 b a DE-12
Anwendercodes	076 a Online-Ressource
Anwendercodes	078 e a ebook
Anwendercodes	078 a a by
Stufe	080 a 05
1. Person	100 b p Wölfel, Roland 9 (DE-588)109210360
1.Körperschaft	200 b k Bayern b Staatsministerium für Wirtschaft, Infrastruktur, Verkehr und Technologie 9 (DE-588)10073845-X
HST VF	331 a Praxistipps zur Standortbelegung
AMB	334 a Elektronische Ressource
Zusatz zum HST	335 a 10 erfolgreiche Beispiele zur Stärkung von Innenstädten, Ortszentren und Mittel Flächenmanagement
Verfasserangabe	359 a [Hrsg.: CIMA Stadtmarketing GmbH, Red.: Roland Wölfel ...]
Ort 1. Verlag	410 a München
Ersch.jahr	425 a 2006
Ersch.jahr	425 a a 2006
Umfang	433 a 39 S.
Illustrationen	434 a Ill., graph. Darst., Kt.
Spez. Material	652 a a Online-Ressource
Erg. Bemerkung	659 a kostenfrei b Langzeitarchivierung durch die Bayerische Staatsbibliothek
Notation	700 w a 601500*by

Workflow #2: Elektronische Pflichtablieferung



Workflow #3: Webarchive

Web Curator Tool zum „Einsammeln“

The screenshot displays the Web Curator Tool interface in a web browser. The browser's address bar shows the URL `129.187.255.103:8000/webcurator/home.html`. The browser's tab bar includes several open tabs, with "WEB CURATOR TOOL" being the active one. The interface features a navigation menu at the top right with links for "Home", "Queue", "Harvested", "Help", and "Logout", along with the text "User: super is logged in".

The main content area is organized into two columns of panels, each with a title, an icon, and a list of items with associated actions:

- In Tray:** 14 tasks, 4345 harvest entries. Action: `open`.
- Harvest Authorisations:** 841 harvest authorisations. Actions: `open`, `add new`.
- Targets:** 12 Targets. Action: `open`.
- Target Instances:** 8 Scheduled instances, 2 ready for Quality review. Actions: `open`, `queue`, `harvested`.
- Groups:** 0 Target Groups. Action: `open`.
- Permission Request Templates:** Actions: `open`, `add new`.
- Reports:** Action: `open`.
- Harvester Configuration:** Actions: `general`, `bandwidth`, `profile`.
- Users, Roles, Agencies & Rejection Reasons:**
 - Users:** `open`, `add new`
 - Roles:** `open`, `add new`
 - Agencies:** `open`, `add new`
 - Rejection Reasons:** `open`, `add new`

Workflow #3: Webarchive

- Bereitstellung durch Rosetta mittels „Way-Back-Machine“

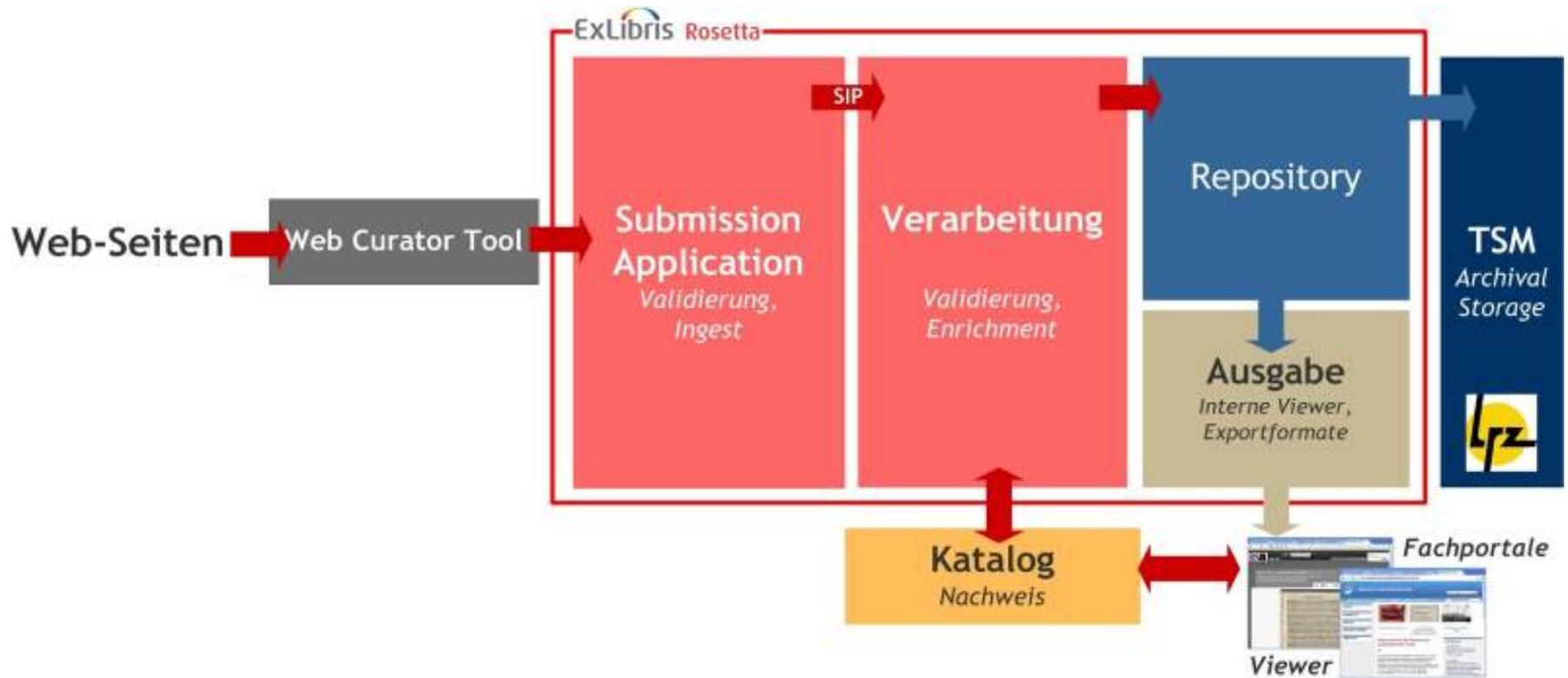
Internet Adresse: 2012 Suche Erweiterte Suche

Suche nach <http://stadtarchiv.mammingen.de/> Set Anchor Window: none 3 Treffer

Suchergebnis für 01.01.2010 - 01.01.2013		
2010	2011	2012
2 Seiten	1 Seite	0 Seiten
11.09.2010 +	05.04.2011 +	
01.10.2010 -		

[Hans](#) | [Hilfe](#)

Workflow #3: Webarchive



Ausbauphase: Fragebogen

Fragebogen für die Pilot-UBs Augsburg, Regensburg, Würzburg
auf der Basis der Projekterfahrungen in der BSB (Rücklauf
Anfang Februar 2012)

Schwerpunkt: für Tests geeignete Objekte/Workflows, Ist-
Situation bzgl. Datenhaltung und Nutzung, Erwartungshaltung
an Rosetta

Auswertung

- In allen 3 UBs kommen Digitalisate von Büchern vor
 - hier soll jeweils ein konkreter Workflow abgebildet werden, um sicherzustellen, dass die Bandbreite der lokalen Gegebenheiten abgedeckt werden kann
- In allen 3 UBs kommen (in Bezug auf die Pilotworkflows der BSB sowie den Fragebogenrücklauf) singuläre Workflows vor (Audio-, Video-, Bilddaten)
 - hier sollen durch eine sinnvolle Auswahl die Erfahrungen mit Rosetta ausgeweitet werden; hier ist ein Echteinsatz von Rosetta teils nur im Sinne eines „dark archive“ angestrebt

Beispiel: UB Regensburg, Historisches Werbefunkarchiv

ExLibris Rosetta AMBA ▾

Type	Usage	Type	Preservation Type	No. of Files	Technical Requirement
	VIEW	DERIVATIVE_COPY		1	View
	VIEW	MODIFIED_MASTER		2	View
	VIEW	PRESERVATION_MASTER 1			View

Audio: MP3 (für Delivery)

Scan: Dokumentation

Audio: WAV (Master)

1/1388	Peppi	131/7	181/59
1/1389	Peppi	131/9	414/59
1/1390	Shell	132/1	106/54
1/1391	Shell	132/4	511/59
1/1392	Bananenmark	133/1	121/54