

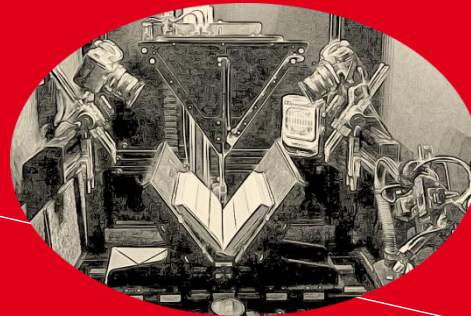


**SLUB**

Wir führen Wissen.

# Automatische Langzeitarchivierung

für die Digitalisierung mit Goobi



Andreas Romeyke, Dr. H. Berthold, C. Beisert, J. Sachse, Fotos mit freundlicher Genehmigung von **Jörg Sachse**  
Sächsische Landesbibliothek, Staats- und Universitätsbibliothek Dresden (SLUB)

**Juni 2014**



**SLUB**

Wir führen Wissen.

# **Europäischer Fonds für regionale Entwicklung – EFRE Konzeption und Realisierung einer effizienten Langzeitarchivierung**



Europäische Union

Europa fördert Sachsen.



Europäischer Fonds für  
regionale Entwicklung

# Agenda

- Was ist ein Langzeitarchiv?
- Digitales Langzeitarchiv SLUB
- Goobi
- Architektur und Datenfluss
- Ingest-Preprocessing
- Ingest
- Access-Postprocessing
- Produktiver Einsatz
- Unsere Erfahrungen

# Digitale Langzeitarchivierung

## Definition und Ziele

**DIN 31644:** „Organisation (bestehend aus Personen und technischen Systemen), die die Verantwortung für den Langzeiterhalt und die Langzeitverfügbarkeit von Information in digitaler Form sowie die Bereitstellung für eine bestimmte Zielgruppe übernommen hat.“



### Ziele

- Sicherung der langfristigen Verfügbarkeit und Nutzung von digitalen Objekten (50 Jahre +)
  - Berücksichtigung zukünftiger Nutzungsszenarien
- => Erhalt der Korrektheit (Bitstream Preservation) und  
=> Erhalt der Interpretierbarkeit und Nutzbarkeit (Content Preservation)

# Digitales Langzeitarchiv SLUB

## Ziele und Stand

Aufbau des Digitalen Langzeitarchivs der SLUB erfolgt im Rahmen eines Projektes (Mai 2012 bis Oktober 2014)

### Ziele

- Sichern der Langzeitverfügbarkeit der Digitalen Sammlungen der SLUB (Digitalisierung mit Goobi, Elektronische Publikationen, Digitale Sammlung der Deutschen Fotothek, Digitales Audio/Video-Material der Mediathek)
- Vorbereitung einer Dienstleistung für andere sächsische Institutionen

### Stand:

- Implementierung und Test von Workflows für die Digitalisierung mit Goobi und Elektronische Publikationen (Qucosa) erfolgreich
- Erweiterter Testbetrieb für Goobi-Workflow; Inbetriebnahme Q3/2014
- Anpassung und Inbetriebnahme des Qucosa-Workflows nach Umstellung des Repositories; Inbetriebnahme Q1/2015

# Digitales Langzeitarchiv SLUB

## Grundsätze

Verwendung als Dark Archive, in dem die Masterdaten verwaltet und archiviert werden – die Präsentationsdaten bleiben in einem separaten Repository, können aber aus den Masterdaten erzeugt werden

Automatisierung des Ingest, d.h. der Übernahme ins Langzeitarchiv, und des Access, d.h. des Zugriffs auf die Daten aus dem Langzeitarchiv (bis auf Fehlerfälle)

Prüfsummen werden bereits im Produktionsprozess (bei der Digitalisierung) bzw. der Annahme (bei Elektronischen Publikationen) erzeugt und bei der Übernahme ins Langzeitarchiv geprüft

Unterstützung einer definierten Menge von LZA-fähigen Datenformaten

Übernahme ins Langzeitarchiv nur für erfolgreich geprüfte Dokumente

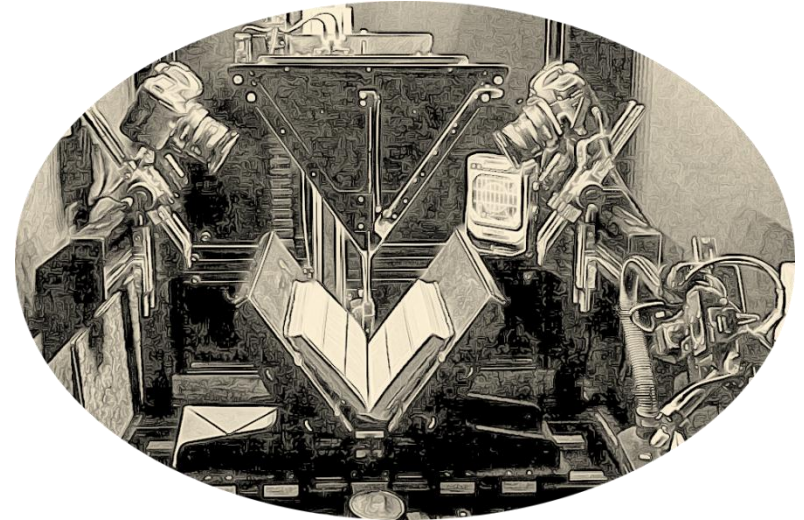
# Goobi

Goobi ist eine Software zur Unterstützung des Digitalisierungsworkflows

SLUB Digitalisierungszentrum erzeugt ca. 3 Mio. Scans im Jahr

Digitalisiert werden ausgewählte Printmaterialien

Digitale Dokumente sollen automatisiert nach beendeter Bearbeitung ins SLUB Langzeitarchiv übernommen werden



# Herausforderungen

## Allgemein

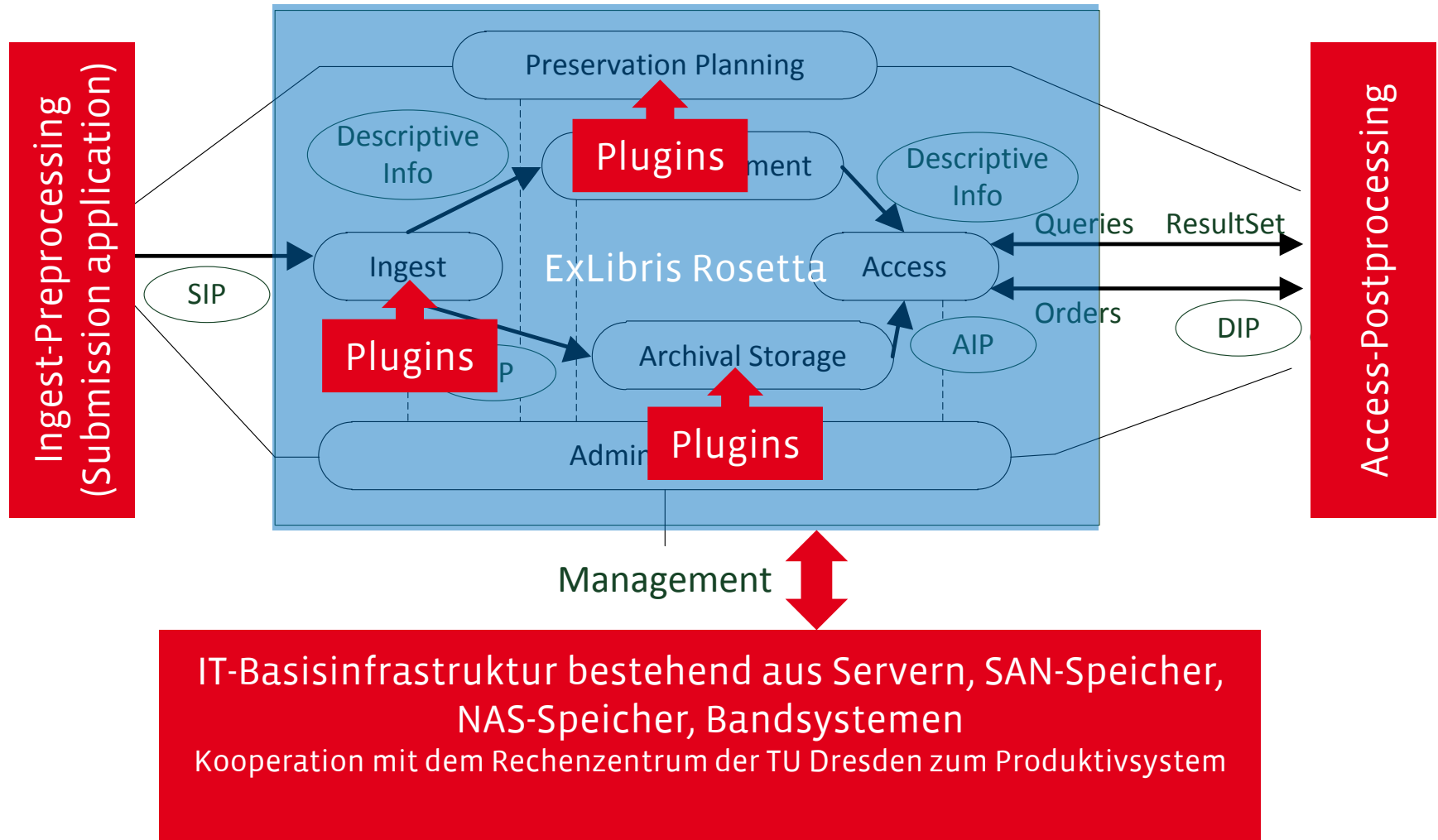
- Schneller Medien-, Format- und Systemwandel
- Begrenzte Haltbarkeit der Trägermedien
- Integrität der Daten nimmt durch gezielte Modifikation oder Systemfehler ab
- Hardwareausfälle, Softwarefehler, Unglücksfälle
- Veralten der Dateiformate – Software, die das Datenformat korrekt interpretieren wird nicht mehr entwickelt/gepflegt

## Speziell

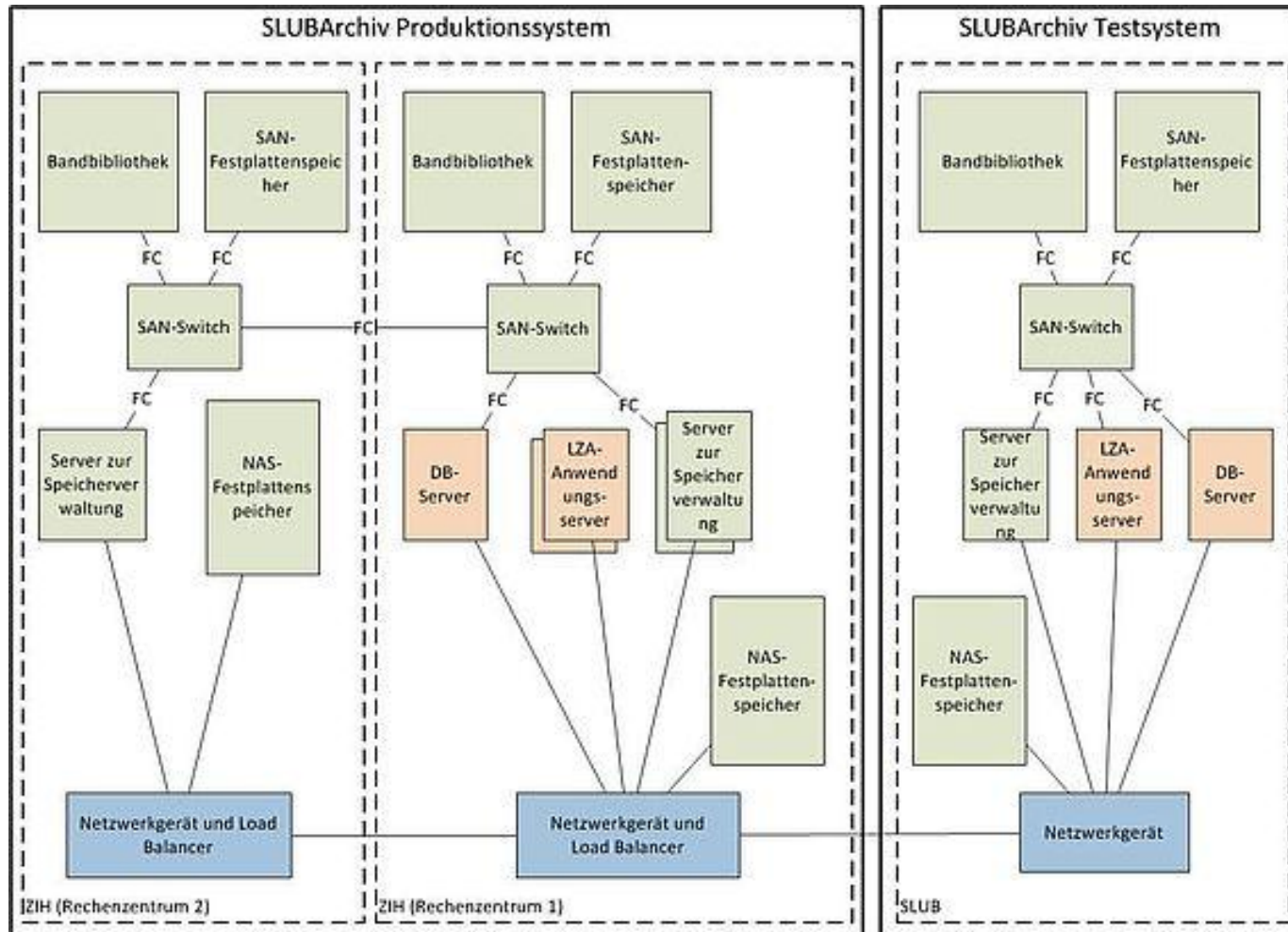
- Datendurchsatz zur Verarbeitung der täglich produzierten Daten (200 bis 500GB pro Tag + Migration)
- Prüfen der Korrektheit bei großen Datenmengen über Monitordaten
- Intellektuelle Prüfung beim Ingest (1% der SIPs)



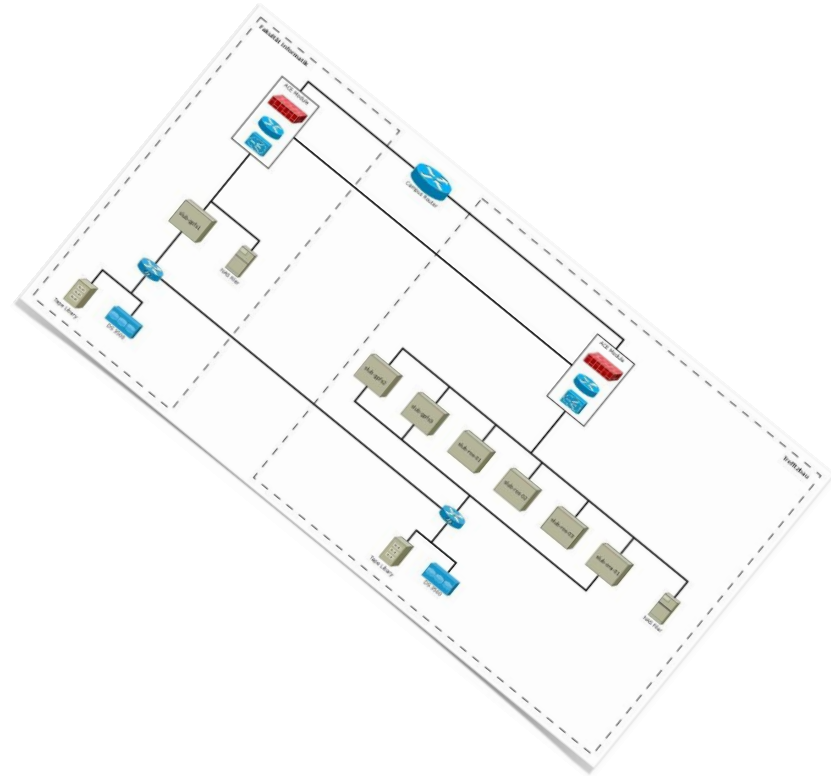
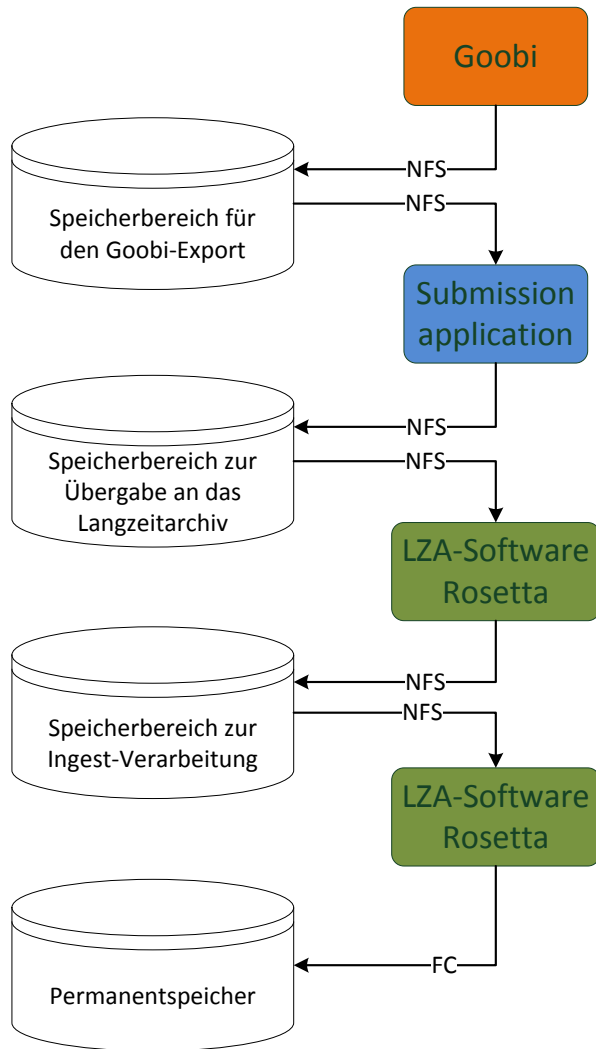
# Architektur



# IT-Basisinfrastruktur

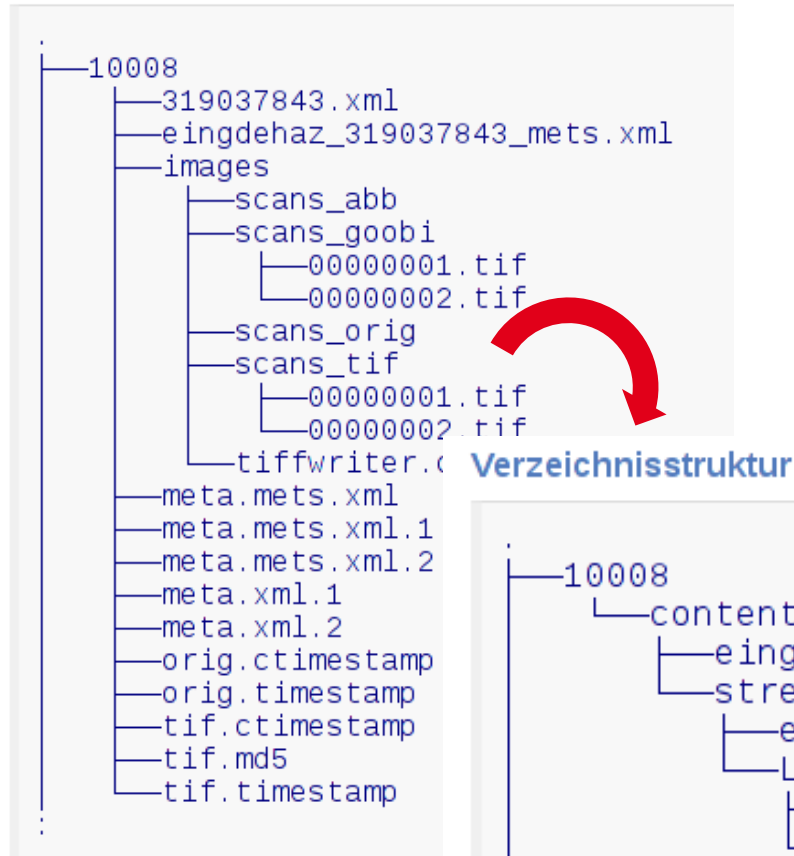


# Datenfluss

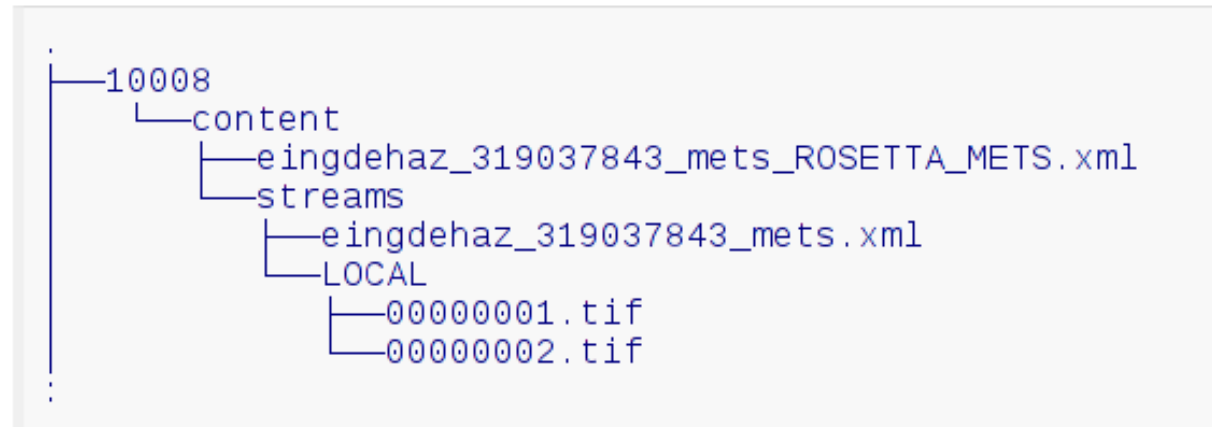


# Ausgangsdaten und Ingest-Preprocessing

## Verzeichnisstruktur Goobi (Detail eines Vorganges)



- Prüfen, ob beendete Goobi-Vorgänge vorhanden
- Prüfung der Vollständigkeit
- Prüfen der Integrität über Prüfsummen
- Transformation der Metadaten
- Übergabe ins Langzeitarchiv



# Ingest

Automatische Übernahme (muß für jeden Workflow konfiguriert und über Plugins/Programme angepasst werden)

- Prüfen der Vollständigkeit und Integrität der Dateien (Prüfsummen)
- Virusprüfung
- Identifikation des Datenformates (PRONOM-ID)
- Validierung = Prüfen der Korrektheit der Daten (Jhove)
- Extraktion von technischen Metadaten
- Erstellen eines Archivpaketes
- Speicherung im Archivbereich

# Access-Postprocessing

Export der Daten aus dem SLUB Langzeitarchiv

„Postprocessing application“ erstellt automatisch Goobi-Vorgang zur Wieder- oder Weiterverarbeitung in Goobi

# Produktiver Einsatz

Dafür müssen noch folgende Ziele erreicht werden:

- Erreichen des erforderlichen Durchsatzes für tägliche Produktion und Migration
  - Lasttests und Tuning
- Automatisierung der Änderungen (AIP Update) mit allen nötigen Parametern

# Unsere Erfahrungen

- Dark Archive und lose Kopplung an das Produktionssystem hat sich bewährt
- Automatisierung spart Ressourcen und minimiert „menschliche Fehler“
- Entwurf praxisnaher Testfälle und automatische Durchführung ist wichtig
- Ein Langzeitarchiv ist kein Archiv und kein Backup
- Größter Aufwand ist es, andere zu überzeugen in Jahrhunderten zu denken
- LZA-fähige Formate nutzen (siehe nächster Vortrag)
- Probleme mit Ausgangsdaten vor der Aufnahme ins Langzeitarchiv lösen - Option auf zukünftige Ressourcen ist problematisch
- Kooperieren, Vernetzen, Austauschen



# Kontakt/Nachfragen

Web: <http://www.slub-dresden.de/ueber-uns/digitale-langzeitarchivierung>

Email: [romeyke@slub-dresden.de](mailto:romeyke@slub-dresden.de)