



SLUB

Wir führen Wissen.

Ein Dienst für die Digitale Langzeitarchivierung

Dr. Henrike Berthold
Sächsische Landesbibliothek –
Staats- und Universitätsbibliothek Dresden (SLUB)

März 2016

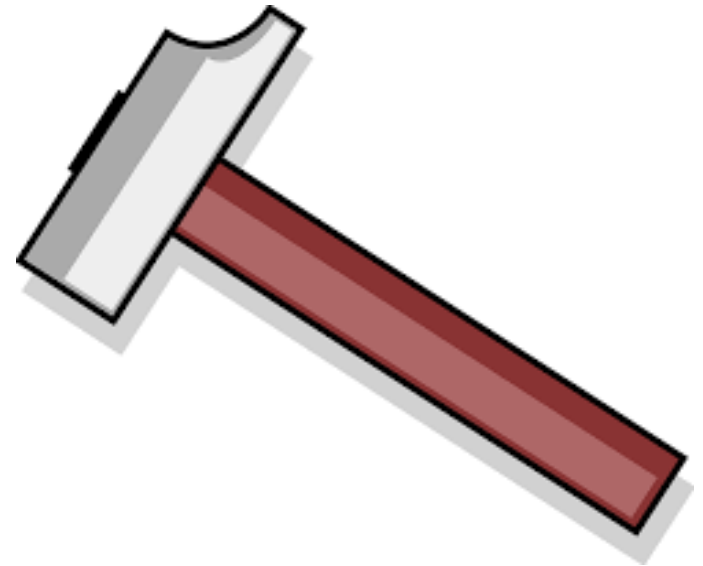
Ein Dienst für die Digitale Langzeitarchivierung

Agenda

- Überblick über Dienstleistungen
- Kurzvorstellung SLUBArchiv
- SLUB Dienstleistungen
- Ausblick

Überblick über Dienstleistungen

Dienstleistungen und Werkzeuge



Überblick über Dienstleistungen

Dienstleistungen



HathiTrust : Partnerschaft von akademischen Einrichtungen und Forschungseinrichtungen (aktuell ca. 110 Einrichtungen, meist aus USA und Kanada)

- für die Präsentation und die Langzeitarchivierung von digitalen Sammlungen

LOCKSS (Lots Of Copies Keep Stuff Safe): Allianz von Bibliotheken und Verlagen

- für eContent

Portico: Organisation mit teilnehmenden Bibliotheken und Verlagen

- für E-Journals und E-Books

AREDO: Kooperationsangebot der DNB zur digitalen Langzeitarchivierung

ArchivesDirect: gehostetes Archivemata auf der Basis von DuraCloud und Amazon S3/Glacier

SLUBArchiv (siehe Vortrag ...)

Weitere Dienstleistungsangebote in Deutschland in Vorbereitung

SLUBArchiv Stand

Produktiver Einsatz seit Januar 2015

Zertifizierung mit dem Data Seal of Approval im Juni 2015



Automatisierte Übergabe aus der Digitalisierung mit **Goobi** (TIFF6-Baseline-Images, ALTO XML und METS/MODS)

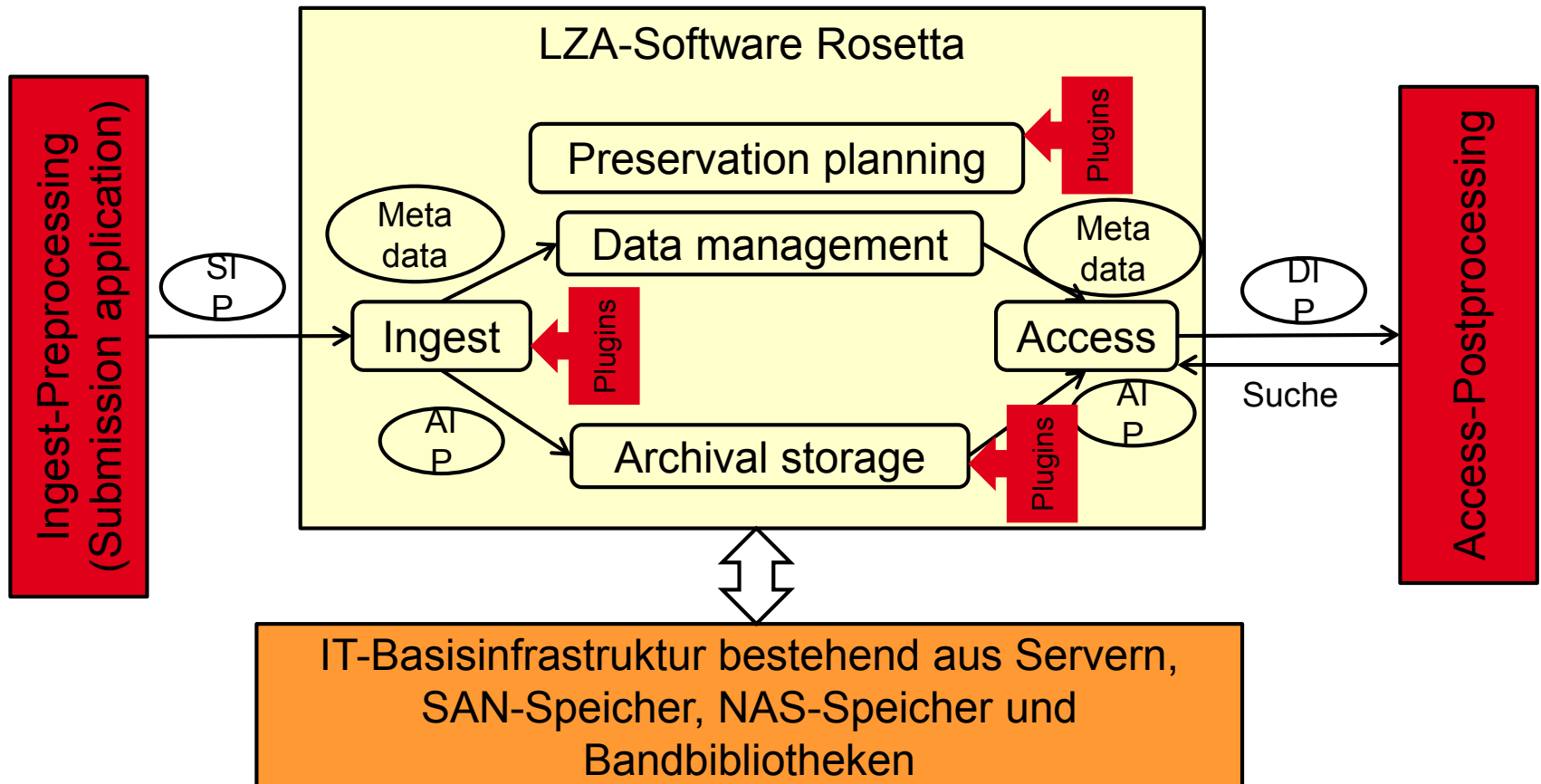
⇒ Vortrag zum Goobi-Workflow beim Bibliothekartag in Bremen

Weitere Workflows

- Elektronische Publikationen/Qucosa: PDF/A 1 oder 2
- Digitalisate von Fotos: TIFF6-Baseline-Images und LIDO
- Video: Matroska/FFV1 Version 3 und METS/MODS

Vernetzt: Partner in nector, als Rosetta-Anwender aktiv in der internationalen Nutzergruppe

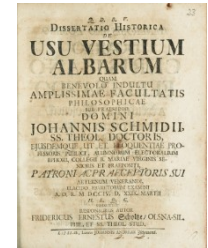
SLUBArchiv Architektur



Betrieb der IT-Basisinfrastruktur durch das Rechenzentrum der TU Dresden (ZIH)

SLUBArchiv

Betriebsaufgaben



Prüfung des Bitstreams

- Muster-basiertes Verfahren
- Stichproben-Verfahren

49	49	2A	00	08	00	00	00	10	00	FE	00	04	00	01	00	00	00	00	00	00	00	00	01	03	00	01	00	00	00	CE	06	00	00	01	01	03	00	01	00	00	00	C9	08
00	00	02	01	03	00	03	00	00	00	CE	00	00	00	03	01	03	00	01	00	00	00	01	00	00	06	01	03	00	02	00	00	00	0A	01	03	00	01	00	00	01	00		
00	00	01	00	00	00	11	01	04	00	01	00	00	00	A8	2A	04	00	12	01	03	00	01	00	00	00	01	00	00	00	15	01	03	00	01	00	00	00	03	00	00	16	01	
04	00	01	00	00	00	C9	08	00	00	17	01	04	00	01	00	00	00	3A	57	B3	00	1A	01	05	00	01	00	00	D4	00	00	1B	01	05	00	01	00	00	00	DC	00		
00	00	28	01	03	00	01	00	00	00	02	00	00	00	73	87	07	00	C4	29	04	00	E4	00	00	00	00	00	00	08	00	08	00	08	00	2C	01	00	00	01	00	00	00	
2C	01	00	00	01	00	00	00	04	29	C4	41	50	50	4C	02	40	00	00	73	63	6E	72	52	47	42	20	4C	61	62	20	07	DF	00	0A	00	0D	00	09	00	26	00	21	
61	63	73	70	4D	53	46	54	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	F6	D6	00	01	00	00	00	00	D3	2D		
4C	4F	47	4F	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00

Überwachung der Formate (Risikoanalyse), Recherche und Test von Tools für Formatmigrationen, Durchführung von Formatmigrationen

Monitoring: Überwachung der technischen Komponenten

Pflege

- Aktualisieren der LZA-Software & umfangreiche Prüfungen des Zusammenspiels mit den anderen Softwarekomponenten
- Aktualisieren der Systemhardware, Wartung der technischen Komponenten



Fehlerbearbeitung



SLUBArchiv

Entwicklungsaufgaben

Entwicklung der Unterstützung für neue Medienarten, Datenformate und Metadatenformate

- Formatidentifikation/Formatvalidierung und Metadatenanreicherung
- Recherche nach Open Source oder kommerziellen Werkzeugen
- Entwicklung der Werkzeuge bei Bedarf und Bereitstellung als Open Source

Weiterentwicklungen/Verbesserungen

- Planung der Erneuerung der IT-Infrastruktur und Datenmigration
- Zertifizierung mit dem nestor-Siegel
- weitere Automatisierung
- ...

SLUB Dienstleistungen

Motivation und Stand

Motivation

- Verantwortung als Landesbibliothek
- Aufbau eines digitalen Langzeitarchivs ist aufwändig; der Betrieb ist teuer
- Erwartung von Synergieeffekten durch „Standardisierung“ z.B. für Digitalisierung mit Goobi oder gemeinsam genutzte Datenformate

Stand

- Zunächst für Daten aus dem Digitalisierungssystem Goobi
- Erarbeitung von konkreten Spezifikationen/Anforderungen an die Datenformate und die Datenübergabe
- Gespräche mit interessierten Einrichtungen; Vorprojekt in Q2/2016

SLUB Dienstleistungen

Standardlösung

- Einrichten der Übernahme und Übergabe für eine bereits unterstützte Materialart bestehend aus Datenformat und Metadatenformat, gemäß einer Spezifikation
- Speicherung von 3 Kopien an zwei Standorten
- Information über die erfolgreiche Übernahme bzw. aufgetretene Fehler
- Pattern-basiertes Prüfverfahren (4mal im Jahr) für den Bit-Stream
- Stichprobenbasiertes Prüfverfahren (2mal im Jahr) für den Bit-Stream
- Überwachung der Formate, Recherche und Test von Tools für Formatmigrationen
- Reporting
- Langfristige Pflege/Betrieb und Erneuerung der IT-Infrastruktur (Hardware, Systemsoftware, LZA Software) und ggf. Migration der Daten

SLUB Dienstleistungen

Beispiele für Projekte

Prüfung des Datenmaterials und Erstellung einer Bewertung und einer Empfehlung für das Vorgehen zur Vorbereitung der digitalen Langzeitarchivierung

Unterstützung bei der Untersuchung und Behebung eines Fehlers

Preservation planning & action für spezifizierte Menge von Dokumenten

- Überprüfung der signifikanten Eigenschaften; Durchführung von Migrationen anhand einer Stichprobe der Daten; Erstellung einer Bewertung und Empfehlung; Durchführung der Preservation action

Einrichten einer spezifischen Übernahme und Übergabe

Einrichten der Unterstützung für ein vom Kunden gefordertes, nicht unterstütztes Datenformat

- Formaterkennung, Formatvalidierung, Metadatenanreicherung; Recherche nach open source oder kommerziellen Werkzeugen; Entwicklung der Werkzeuge bei Bedarf; Bereitstellung open source

Migration der Daten nach der Kündigung des Vertrages zu einem anderen System

SLUB Dienstleistungen

Eckpunkte Kostenmodell: Zusammensetzung

Kosten für den Betrieb des Langzeitarchivs

Kosten für den Betrieb und die fachgerechte Erneuerung der IT-Basisinfrastruktur (Hardware, Systemsoftware)

Kosten für die Nutzung der LZA-Software Rosetta (Lizenzgebühren)

Überblick über mögliche Kostenmodelle wurde Projekt [4C – Collaboration to Clarify the Costs of Curation](#) erarbeitet

Grundlage für das SLUB Kostenmodell ist das [LIFE-Modell](#), das im Rahmen des [Projektes Life Cycle Information for E-Literature](#) entstanden ist

umfassen Lebenszyklusabschnitte Ingest, Bit-Stream Preservation, Content Preservation und Access

SLUB Dienstleistungen

Kostenaufteilung

Nach Datenvolumen

- Pflege/Betrieb und Erneuerung der IT-Infrastruktur
- Muster-basiertes Prüfverfahren des Bitstreams
- Stichprobenbasiertes Prüfverfahren des Bitstreams

Nach Format

- Risikoanalyse und das Prüfen und Testen von Formatmigration

Nach Workflow

- Gemeinsam genutzte Verfahren für die Übernahme (z.B. Submission application), Übergabe, Reporting

Ausblick

Archivwürdigkeit und Archivfähigkeit

Erarbeitung von Kriterien zur Feststellung der **Archivwürdigkeit**



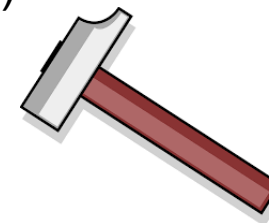
Sicherstellen der **Archivfähigkeit**

Integritätsüberprüfung durch das Erzeugen und Überprüfen von Prüfsummen
z.B. mit md5sum

Formatidentifikation mit standalone DROID/PRONOM

Formatvalidierung z.B. mit

- JHOVE
- checkit_tiff (baseline TIFF, Eigenentwicklung SLUB, [Open source](#))
- Callas pdfaPilot (PDF/A)



Datenkorrektur



SLUB

Wir führen Wissen.

Ein Dienst für die Digitale Langzeitarchivierung

Dr. Henrike Berthold
Sächsische Landesbibliothek –
Staats- und Universitätsbibliothek Dresden (SLUB)

März 2016