

Digitale Langzeitarchivierung beim KOBV



Wolfgang Peters-Kottig

CC BY 4.0

Die KOBV-Verbundzentrale ist ein Haus im Haus

...seit 1997 angesiedelt am
Konrad-Zuse-Zentrum für
Informationstechnik Berlin
(ZIB)

- Eigener Haushalt
- Eigene Aufsichtsgremien
- Eigenständiger Jahresbericht

Aktuell 20 Mitarbeiterinnen
und Mitarbeiter

Trägerschaft:

- Senatskanzlei Kulturelle Angelegenheiten Berlin
- Ministerium f. Wiss. Forsch. Kultur Brandenburg
- Bibliotheken!

Mitgliedsbibliotheken des KOBV

Hochschulbibliotheken: 23

- Alle Universitäten in Berlin und Brandenburg
- 16 Hochschulen/Fachhochschulen (2 private)

Spezial-, Forschungs-, Behördenbibliotheken: 63

- Potsdam Museum
- Bundeskanzleramt
- Leibniz-Institut für Gewässerökologie und Binnenfischerei
- Haus der Wannseekonferenz
- ...

Öffentliche Bibliotheken: ~ 239

- Verbund d. öffentl. Bibliotheken in Berlin VÖBB inkl. ZLB
- Verbund der öffentlichen Bibliotheken im Land Brandenburg VOEB inkl. SLB Potsdam

Tätigkeiten des Verbundes



Allgemeine Services

- Recherche in den Katalogen der Region: KOBV-Portal
- Technische Realisierung der Fernleihe
- Datenkonvertierungen
- Digitale Langzeitarchivierung



Betrieb von Anwendungen für Kooperationspartner („Hosting“)

- Discoverysystem ALBERT
- Dokumentenserver OPUS 4
- Verbundkatalog öffentlicher Bibliotheken in Brandenburg (VÖB)



Softwareentwicklung

- ALBERT
- OPUS 4



Bibliothekarische Projekte

- DeepGreen – Open Access Transformation

Digitale Langzeitarchivierung

Kernproblem: Digitale Information ist fragil



- ▶ Um digitale Information anzeigen/bearbeiten zu können muss ein ganzer Stapel voneinander abhängiger Technologie perfekt zusammenarbeiten:
 - Dateiformate (PDF, HTML, TIFF)
 - Anwendungssoftware (Textverarbeitung, Webbrowser,...)
 - Betriebssystem (Windows, Linux, Mac OS)
 - Computer (PC, Smartphone)
 - Speichermedien (Festplatte, USB, DVD...)
- ▶ Jede technologische „Schicht“ in diesem Stapel veraltet schnell
- ▶ Damit Informationen benutzbar sind müssen alle Schichten funktionieren

www.memory-of-mankind.com

Eternal 5D data storage could record the history of humankind

Published: 18 February 2016



Eternal 5D data storage

Scientists at the University of Southampton have made a major step forward in the development of digital data storage that is capable of surviving for billions of years.

Using nanostructured glass, scientists from the University's [Optoelectronics Research Centre](#) (ORC) have developed the recording and retrieval processes of five dimensional (5D) digital data by

RESEARCH ARTICLE | APPLIED PHYSICS

Long-term data storage in diamond

SHARE

Siddharth Dhomkar^{1,*}, Jacob Henshaw^{1,2,*}, Harishankar Jayakumar¹ and Carlos A. Meriles^{1,2,†}

¹*Department of Physics, City University of New York (CUNY)—City College of New York, New York, NY 10031, USA.*

²*CUNY—Graduate Center, New York, NY 10016, USA.*

✉[†]Corresponding author. Email: cmeriles[at]ccny.cuny.edu

✉^{*} These authors contributed equally to this work.

Science Advances 26 Oct 2016:

Vol. 2, no. 10, e1600911

DOI: 10.1126/sciadv.1600911



0110111000001010001011101000110110
1101011100100100011110101011010001
0110111000001010001011101000110110
1110101100010001010111001111101001
1011011101000101000101110100011111
0110111000001010001011101000110110
0110111000111010001011101000110110
1010111101001011110101010101101101
1011011101000101000101110100011011
1011011101110101000101110100011011
1011011101000101000101110100011111
0111110101010110111100100100101010

NATIONAL FILM UNIT FILMS
LISTED IN THE FILM CENSOR'S
REGISTER BUT NOT IN
N.F.U. CATALOGUES

Supplement to the
CHRONOLOGICAL ARRANGEMENT OF
FILMS LISTED IN
NATIONAL FILM UNIT AND
TOURIST & PUBLICITY DEPARTMENT
FILM CATALOGUES
1951-1990

Compiled by

„Ewige Speicherung“ ist
keine Digitale Langzeitarchivierung!

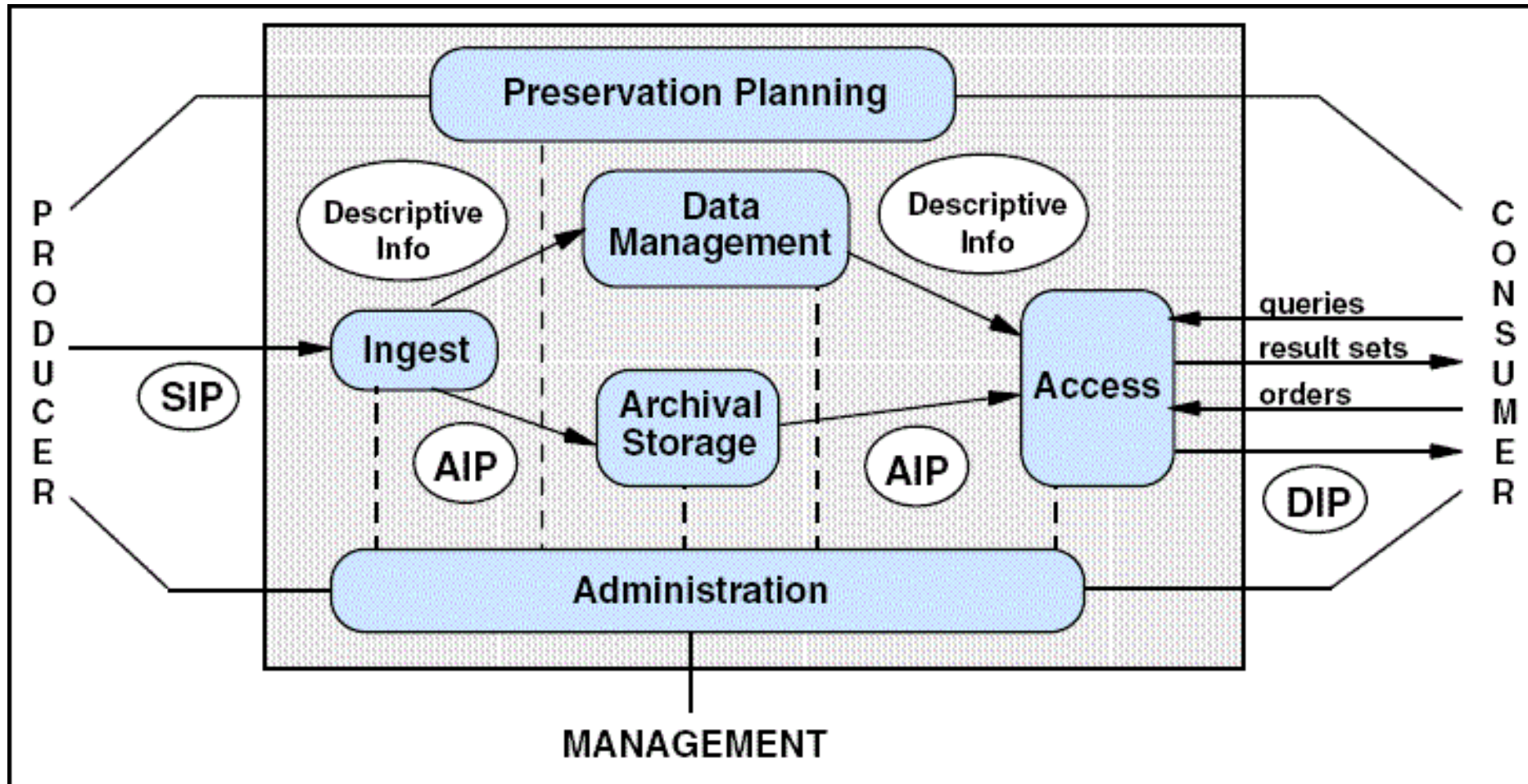
...sondern:

Sicherstellen, dass Daten nachnutzbar bleiben

- Dateiformate und Speichermedien laufend prüfen
- technologische Weiterentwicklung beobachten
- Bei Bedarf in neue Formate konvertieren (Migration)
- Nur unverschlüsselte Daten ohne Passwortschutz archivieren
- Keine proprietären Dateiformate archivieren
- Möglichst viele Metadaten mitspeichern

-61.78	138.10	3	-0.7	104	Oct-1991
-59.63	139.17	3	-0.2	32	Oct-1991
-58.13	139.83	3	1.5	38	Oct-1991
-55.90	140.85	3	1.2	32	Oct-1991
-54.12	141.67	3	1.8	39	Oct-1991
-54.12	141.67	3	1.8	130	Oct-1991
-53.85	140.67	3	3.2	70	Oct-1991
-53.53	141.98	3	1.4	53	Oct-1991
-53.13	142.25	3	2.5	39	Oct-1991
-51.85	142.58	3	3.9	50	Oct-1991
-51.45	142.58	3	4.3	64	Oct-1991
-51.25	142.90	3	7.0	126	Oct-1991
-50.67	143.25	3	5.6	57	Oct-1991

Entwicklung von Langzeitarchiven ist geprägt durch ISO-Norm: OAIS



Speichermedium am ZIB: Magnetband



- Magnetbandkassetten mit 20-30 Jahren Haltbarkeit (Tape-Libraries mit Robotersteuerung)
- > 100 Petabyte Speicherplatz im ZIB

Das Digitale Langzeitarchiv EWIG

- Entwickelt und betrieben vom Forschungs- und Kompetenzzentrum Digitalisierung Berlin (digiS) und dem KOBV
- Seit 2014 Gemeinsame Arbeitsgruppe „Digital Preservation“ von KOBV und digiS



Funktionen eines Langzeitarchivierungssystems



Ingest

- Prüfen der Vollständigkeit eines Pakets und der Integrität der Daten
- Virenprüfung
- Identifikation der Datenformate aller Dateien
- Prüfen der Korrektheit der Daten (Formatvalidierung)
- Extraktion von technischen Metadaten



Access

- Auslieferung von archivierten Inhalten / Präsentation



Administration

- Verwaltung des Archivs, z.B. Rechtemanagement

Funktionen eines Langzeitarchivierungssystems



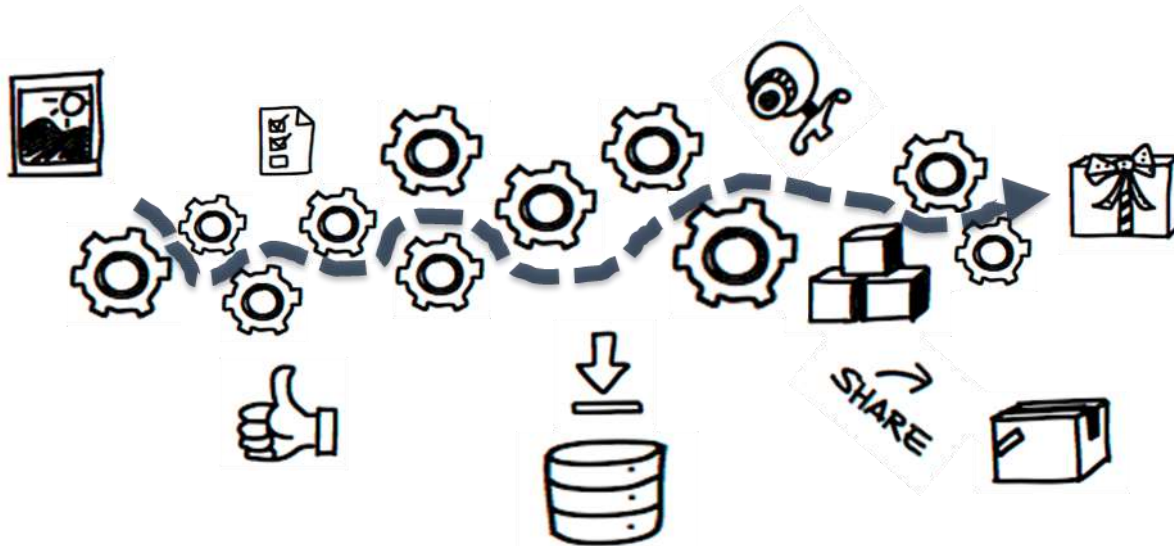
Preservation Planning

- Beobachten der technischen Entwicklungstrends
- Rechtzeitige Reaktion bei Technologiewandel



Data Management / Archival Storage

- Dokumentation von Vorgängen im System
- Redundante Speicherung der Daten



Das Digitale Langzeitarchiv EWIG

Einzelne Software-Komponenten, die über Skripte verknüpft sind

 archivematica®

Ingest, Preservation Planning



Fedora™



Reporting Dashboard

Datenmanagement, Access

iRODS

Archival Storage



...diverse Python-Skripte

Vor dem Start einer Datenübernahme

Kommunikation Datengeber <-> Langzeitarchiv

- Sind die Daten *einzigartig*?
- Sind die Daten mittelfristig *nicht mehr nutzbar*?
- Sind die *Kosten für die erneute Erstellung* hoch?
- Muss wirklich *alles* aufbewahrt werden?
- Prüfung der Archivwürdigkeit
→ „Garbage in – Garbage out“:
Langzeitarchive können Daten
nicht verbessern



Vor dem Start einer Datenübernahme

Vertrag: Rechtliche Grundlage mit
Versprechen der Langzeitverfügbarkeit

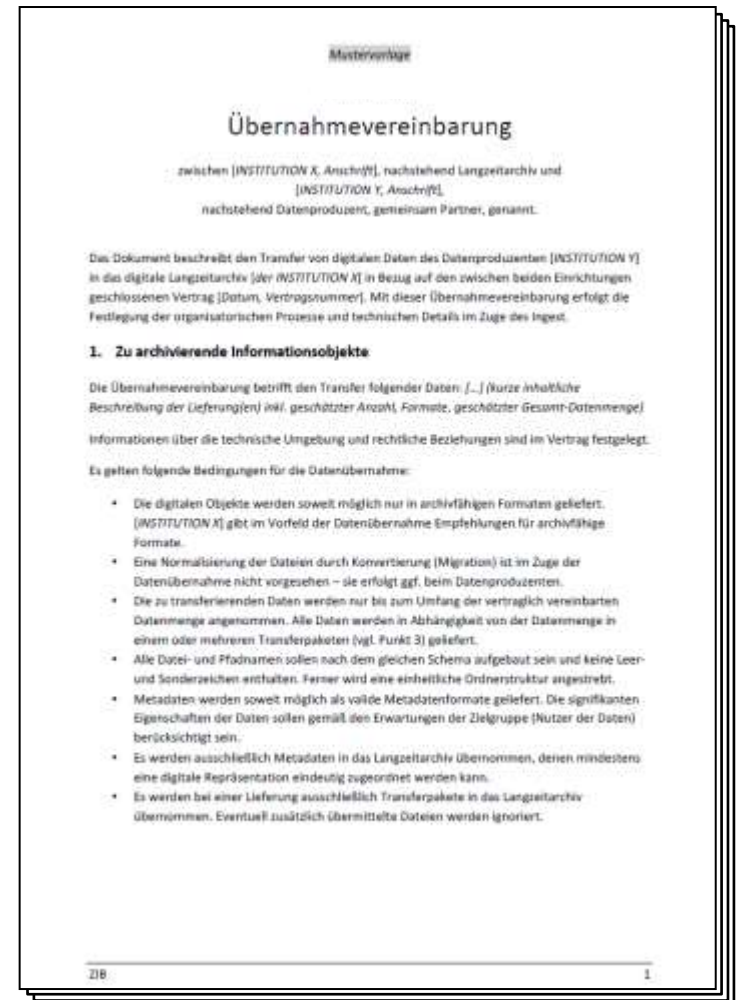


- Beschreibung der Inhalte
- Rollenverteilung Langzeitarchiv / Datengeber
- Technische Realisierung (Einlagerung, Speicherung, Bereitstellung)
- Ansprechpartner, handelnde Parteien
- Entgelte
- Rechteeinräumung, Haftung, Gewährleistung

Vor dem Start einer Datenübernahme

Übernahmevereinbarung: Konkrete Vereinbarungen zu

- Ordnerstruktur
- Transferwege (Festplatte, Internet)
- Beschreibung eines Transferprozesses
- „Packzettel“: notwendige Metadaten über die Datenlieferung



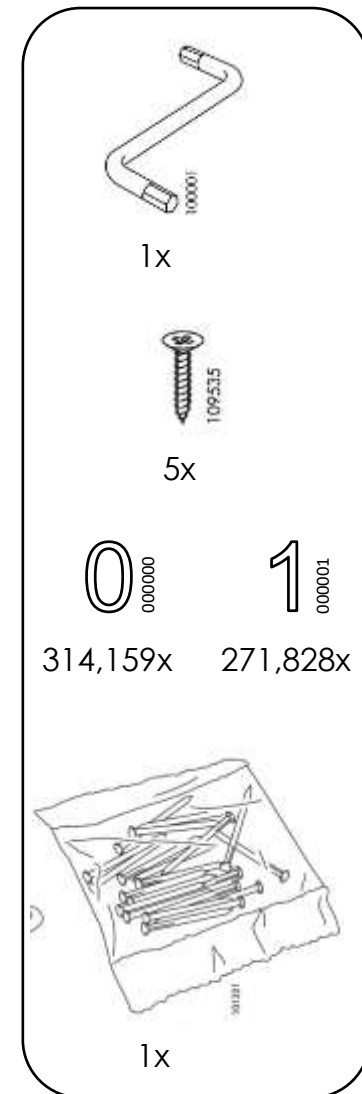
Inhalt eines Transferpakets

Administrative Informationen
(Submission Manifest)

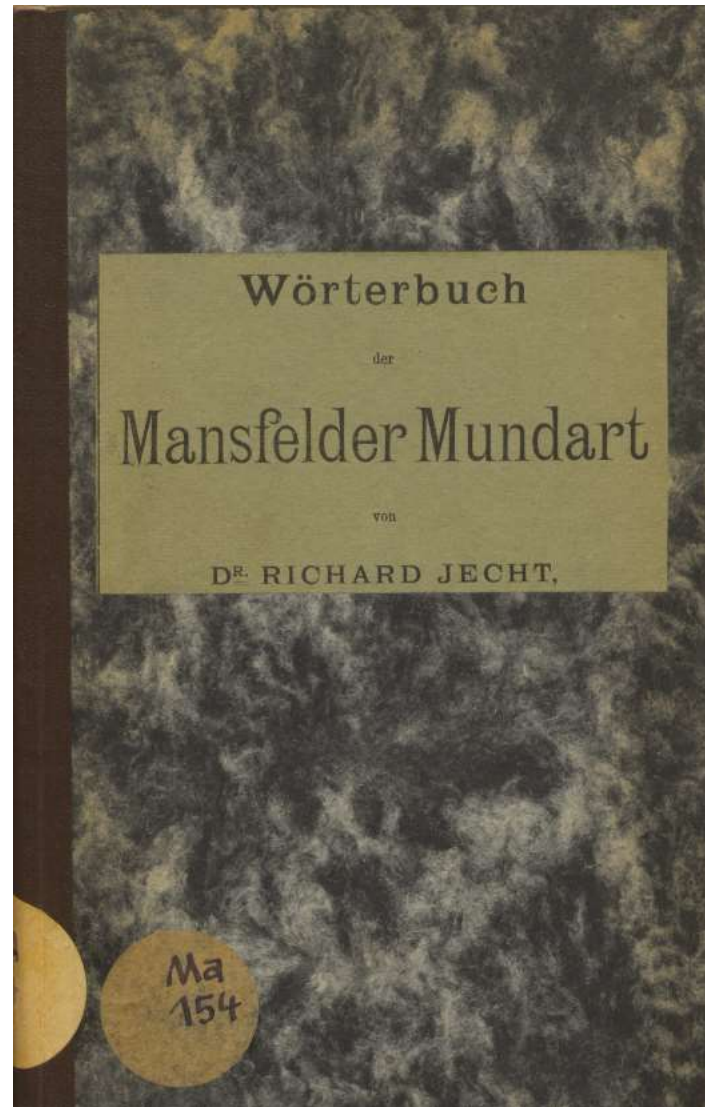
Deskriptive Metadaten
(MARC/MODS, DC)

Masterdateien
(z.B. *.tif)

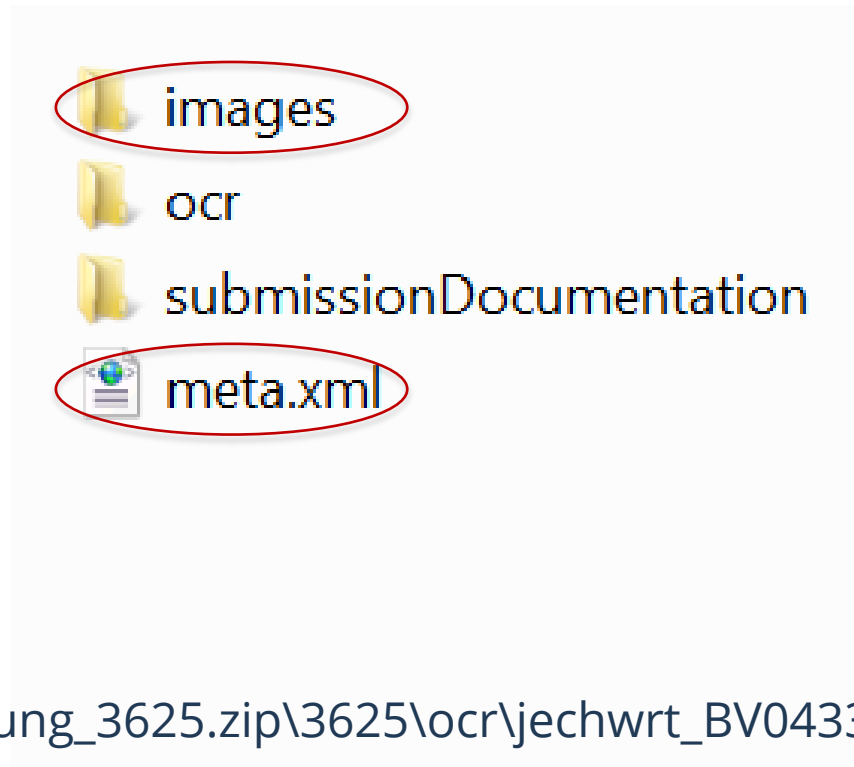
Zusätzliche beschreibende Informationen
(Submission Documentation, optional)



Beispiel Retrodigitalisierungsprojekt



Beispiel Retrodigitalisierungsprojekt



Beispiel_Lieferung_3625.zip\3625\ocr\jechwrt_BV043371129_pdf

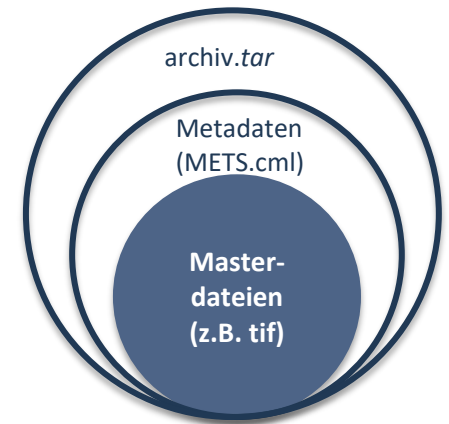
▼ Micro-service: Quarantine			
Job: Workflow decision - send transfer to quarantine [?]	Completed successfully		
Job: Find type to process as	Completed successfully		
Job: Move to workFlowDecisions-quarantineSIP directory	Completed successfully		
Job: Designate to process as a standard transfer	Completed successfully		
▼ Micro-service: Generate METS.xml document			
Job: Generate METS.xml document	Completed successfully		
▼ Micro-service: Verify transfer checksums			
Job: Verify metadata directory checksums	Completed successfully		
▼ Micro-service: Reformat metadata files			
Job: Process transfer JSON metadata	Completed successfully		
▶ Micro-service: Assign file UUIDs and checksums			
▼ Micro-service: Include default Transfer processingMCP.xml			
Job: Include default Transfer processingMCP.xml	Completed successfully		
▼ Micro-service: Rename with transfer UUID			
Job: Rename with transfer UUID	Completed successfully		
▼ Micro-service: Verify transfer compliance			
Job: Verify mets_structmap.xml compliance	Completed successfully		
Job: Verify transfer compliance	Completed successfully		
Job: Attempt restructure for compliance	Completed successfully		
Job: Remove unneeded files	Completed successfully		
Job: Remove hidden files and directories	Completed successfully		
Job: Set transfer type: Standard	Completed successfully		
Job: Move to processing directory	Completed successfully		
Job: Set file permissions	Completed successfully		
▼ Micro-service: Approve transfer			
Job: Approve standard transfer	Completed successfully		

Eigenschaften von EWIG

- ▶ „Flexible Datenübernahme“: Betreuung der Partner flexibel je nach Kompetenzen und Technik vor Ort = hoher Aufwand
- ▶ „Wir nehmen alles“: verschiedene Metadatenformate, alle Datenformate (mit Risikorückmeldung)
Service-Leistung ist dabei auch abhängig von der Qualität gelieferter Daten und Metadaten

Eigenschaften von EWIG

- ▶ Die Archivpakete enthalten alle Daten (*.tar-Dateien) und sind selbsterklärend („Self-contained“)
→ Wichtig für die Exit-Strategie
- ▶ Hardware wird innerhalb des ZIB betrieben
- ▶ „Dark Archive“: Online-Präsentation erfolgt nicht über EWIG
→ Viele Archive koppeln LZA und Online-Präsentation
- ▶ KOBV (und digiS) sind Partner, die im Hintergrund für die Bibliotheken agieren.



Der Service des KOBV

- Basis-Finanzierung durch Trägereinrichtungen: Land Berlin, Land Brandenburg, KOBV-Bibliotheken
- Im Rahmen der KOBV-Mitgliedschaft kostenloses Basis-Speichervolumen für Universitäts- und Hochschulbibliotheken, beide Landesbibliotheken
- Weitere KOBV-Mitglieder: Leistung im Rahmen von bilateralen Kooperationsvereinbarungen



Die Arbeitsgruppe „Digital Preservation“



- Fünf wissenschaftliche Mitarbeiter von digiS und KOBV, aber nicht 100% Arbeitszeit für LZA
→ umgerechnet ~3 Vollzeitstellen („FTE“)
- 2 Informatiker, 1 Mathematiker, 1 Umweltwissenschaftler:
 - Software-Entwicklung, Betrieb
- 1 Bibliothekar „Digital Preservation Manager“:
 - Ansprechpartner nach außen
 - Entwicklung Geschäftsmodell, Betriebsmodell
 - Diverse Texte schreiben: Vertragsentwürfe, Systemdokumentation, Infotexte, Projektanträge
 - Datenqualität prüfen, Metadatenmappings
 - Technische, rechtliche, organisatorische Entwicklungen nachvollziehen
 - Vertretung des ZIB in Gremien, auf Tagungen