

DIE ÖSTERREICHISCHE BIBLIOTHEKENVERBUND
UND SERVICE GMBH

obv sg

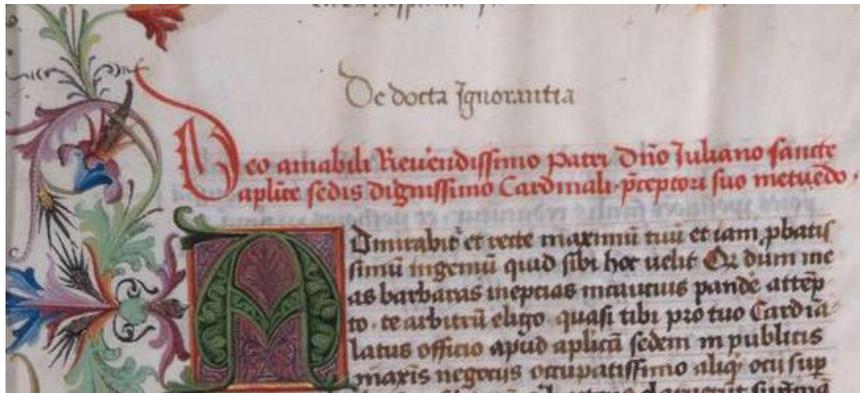
SAVE THE DATA! DIGITALE LANGZEITARCHIVIERUNG ALS DIENSTLEISTUNG FÜR UND VON BIBLIOTHEKEN

111. BIBLIOCON HANNOVER
26. MAI 2023
SILVIO WIESE

- Langzeitarchivierung Wozu?
- Herausforderungen bei der Langzeitarchivierung
- Die Implementierung unseres ersten Teilnehmers im konsortialen Dienst
 - Workflow
 - BagIt und SIP
 - Einzelne Schritte / Prüfungen beim Ingest
 - AIP – Was passiert mit den Daten im System?
 - DIP
- Unser Dienst in der Übersicht
- Ausblick

WARUM MACHEN WIR LANGZEITARCHIVIERUNG?

- Gesetzliche Forderungen (Bundesarchivgesetz (BArchG), Hochschulgesetze)
- Bewahren von kulturellem Erbe
- Schutz von Investitionen (Digitalisierungsprojekte)
- Wissenstransfer (Forschungsdaten)

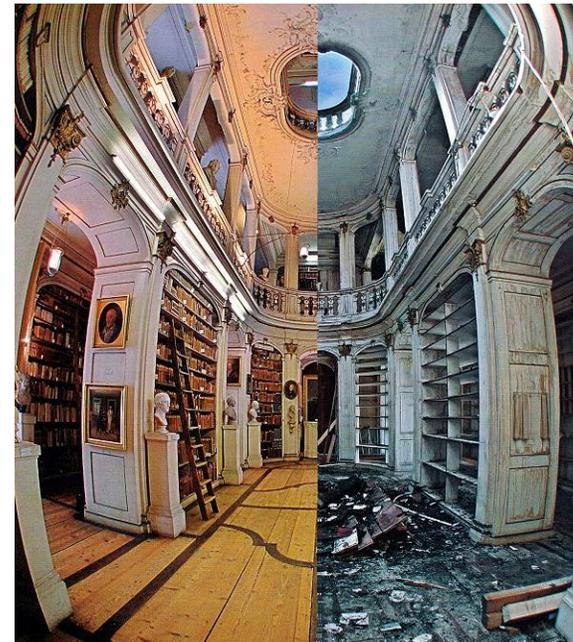
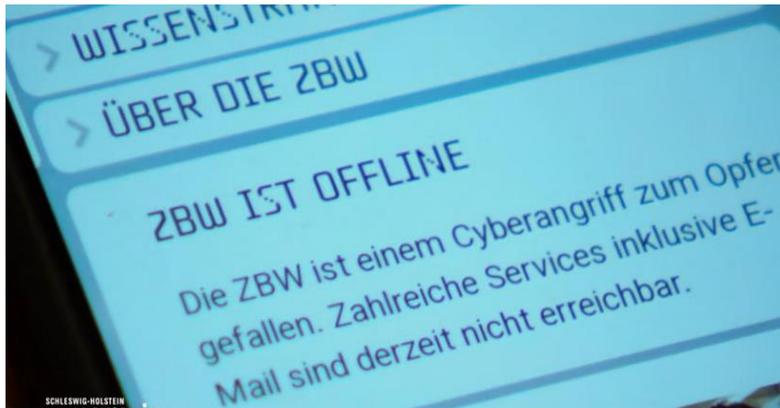


© St. Nikolaus-Hospital/Cusanusstift, Bernkastel-Kues



© Getty Images/iStockphoto

- Technische Probleme (produktionsbedingt/nutzungsbedingt)
- Technologische Veränderungen
- Menschliche Ursachen (innerhalb und außerhalb ihrer Einrichtung)
- Höhere Gewalt (Naturkatastrophen)
- Digitale Angriffe (Ransom Attacke)



© Stiftung Weimarer Klassik

Langzeitarchivierung ≠

- nur Dateitransfer zwischen zwei technischen Systemen
-> Langzeitarchiv als organisatorisches System

Langzeitarchivierung =

- Erfassung mit **Integrität** (Fixity)
- Langfristige Aufbewahrung mit **Authentizität**
- Erhaltung der dauerhaften **Verfügbarkeit**

„Langzeit“ bedeutet für die Bestandserhaltung digitaler Ressourcen nicht die Abgabe einer Garantieerklärung über fünf oder fünfzig Jahre, sondern die verantwortliche Entwicklung von Strategien, die den beständigen, vom Informationsmarkt verursachten Wandel bewältigen können.

Eine kleine Auswahl der vielfältigen Herausforderungen

- Herausforderungen vor der eigentlichen Archivierung
 - Was wird archiviert?
 - In welchen „Qualitätsstufen“ wird archiviert?
 - Anzahl und Ort der Kopien
 - Bevorzugung verlustfreier Formate  schnell sehr hoher Speicherbedarf
 - Workflowanpassungen in den Einrichtungen bzgl. des Datenmanagements
 - Erstellung des Submission Information Package (SIP)
 - Herausforderungen bei und nach Empfang der Daten
 - Vollständigkeit
 - Checksums
 - Formaterkennung (PRONOM, DROID)
 - Formatvalidierung (JHOVE)
 - Dokumentation von Ereignissen der Datenentstehung und -verarbeitung (PREMIS)
 - Formatmigration / Emulation
-  **All diese Dinge machen LZA kompliziert und kostenintensiv**

The most advanced digital preservation platform

Our mission is to safeguard the world's research and cultural heritage. Forever.

We created **LIBNOVA** in 2009 to provide next-generation digital preservation platforms that allow organizations to safeguard their content in a better way.

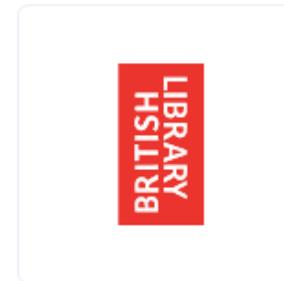
<https://www.libnova.com>

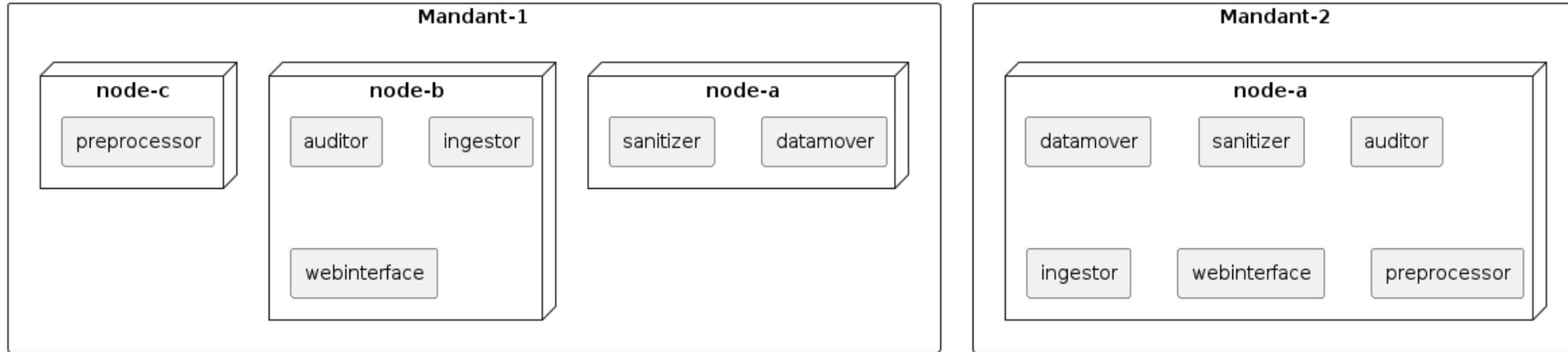
- **2018 – 2020** Projekt: Ausschreibung mit der ÖNB; Gewinner: LIBNOVA mit LIBSAFE
- **2020 – 2022** Projekt: Implementierung von Libsafe für die ÖNB
- **2022 – 2023** Projekt: Implementierung von Libsafe für die WUW

Already with us

Across the world, large and compact

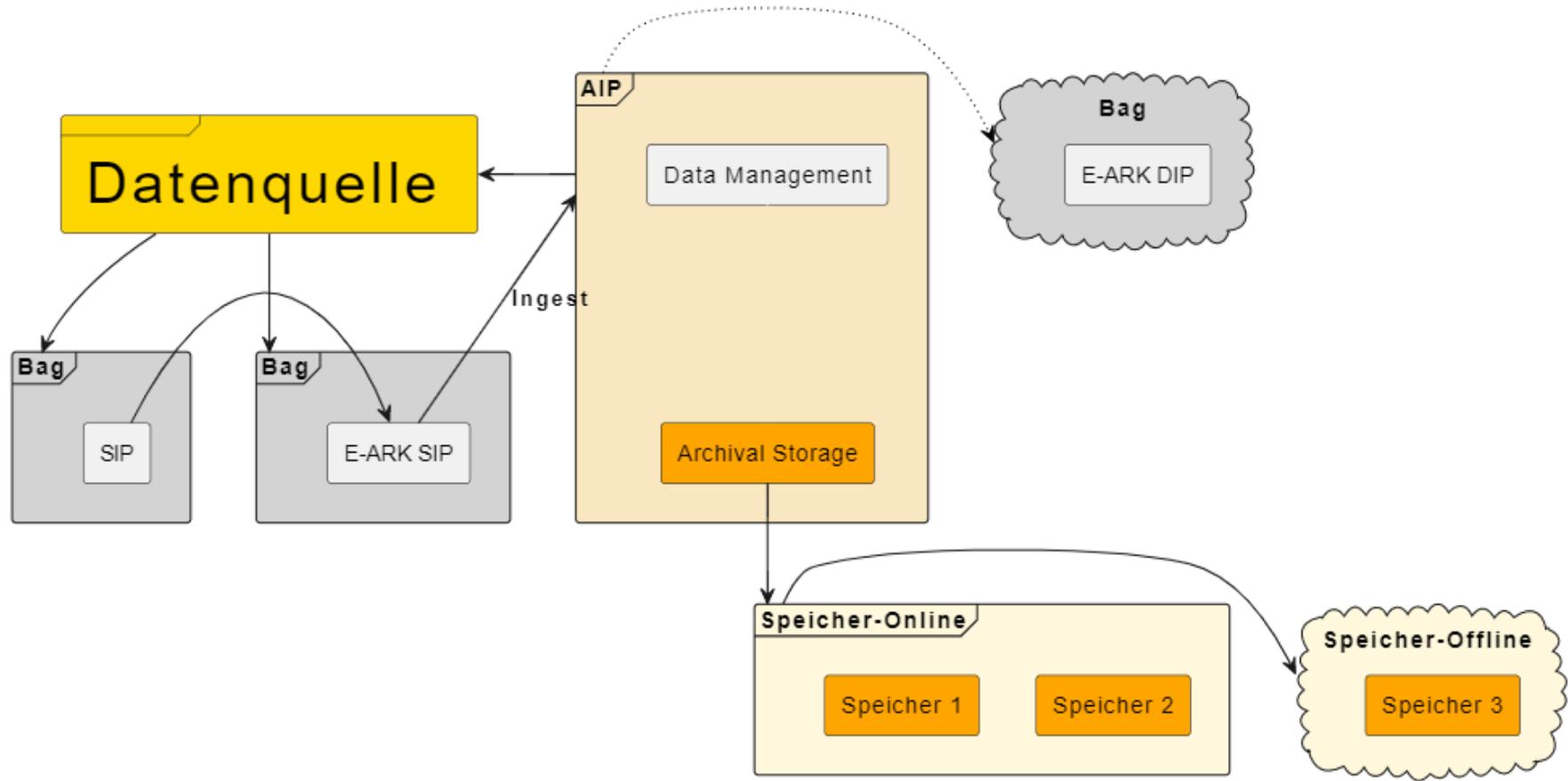
More than 50 institutions in 15 countries





- Serviceorientierte Architektur
- Horizontal verteilbares System
- Gute Skalierbarkeit
- Mandantenfähigkeit

IMPLEMENTIERUNG WORKFLOW AUFBAUEND AUF DEM „OPEN ARCHIVAL INFORMATION SYSTEM (OAIS)“

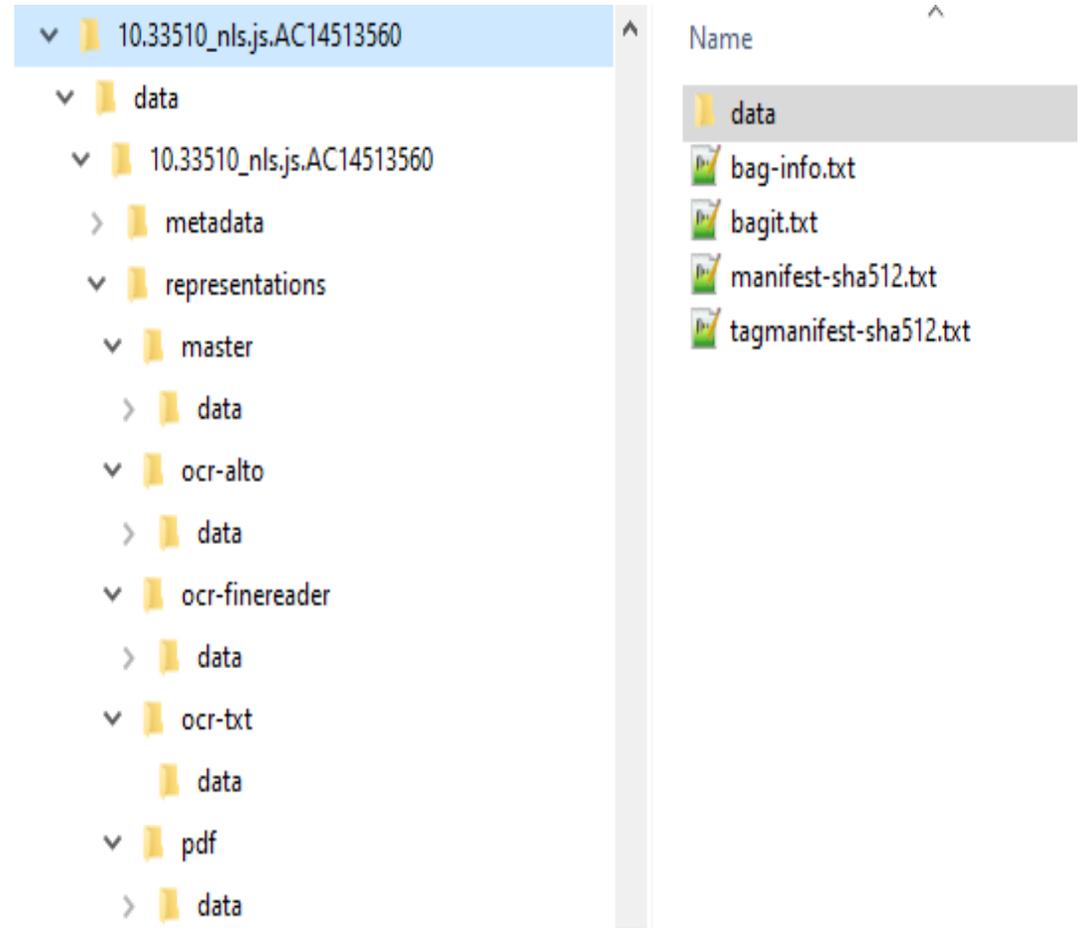


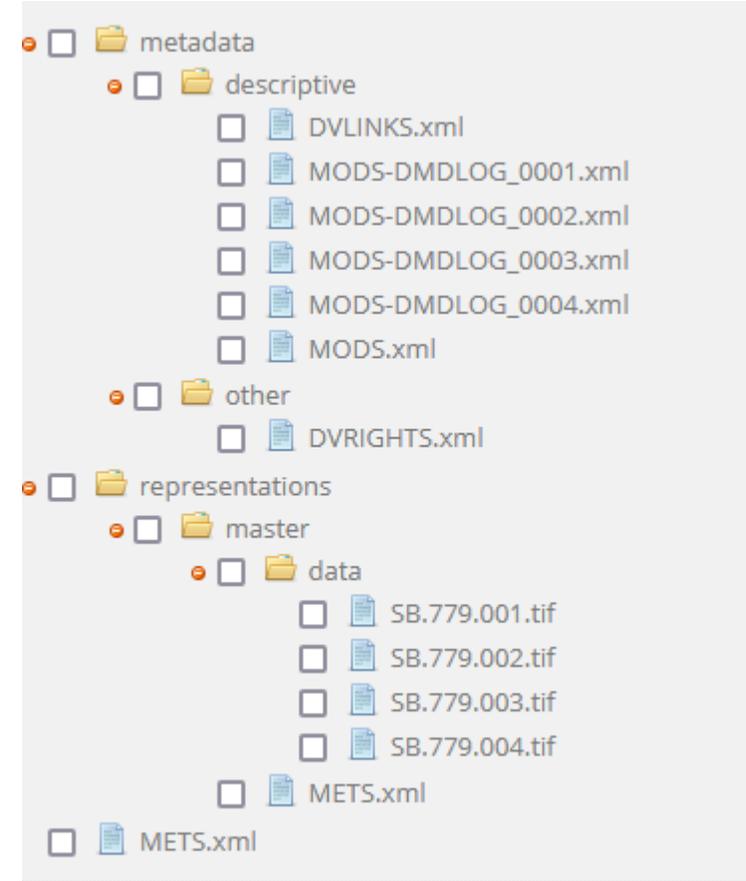
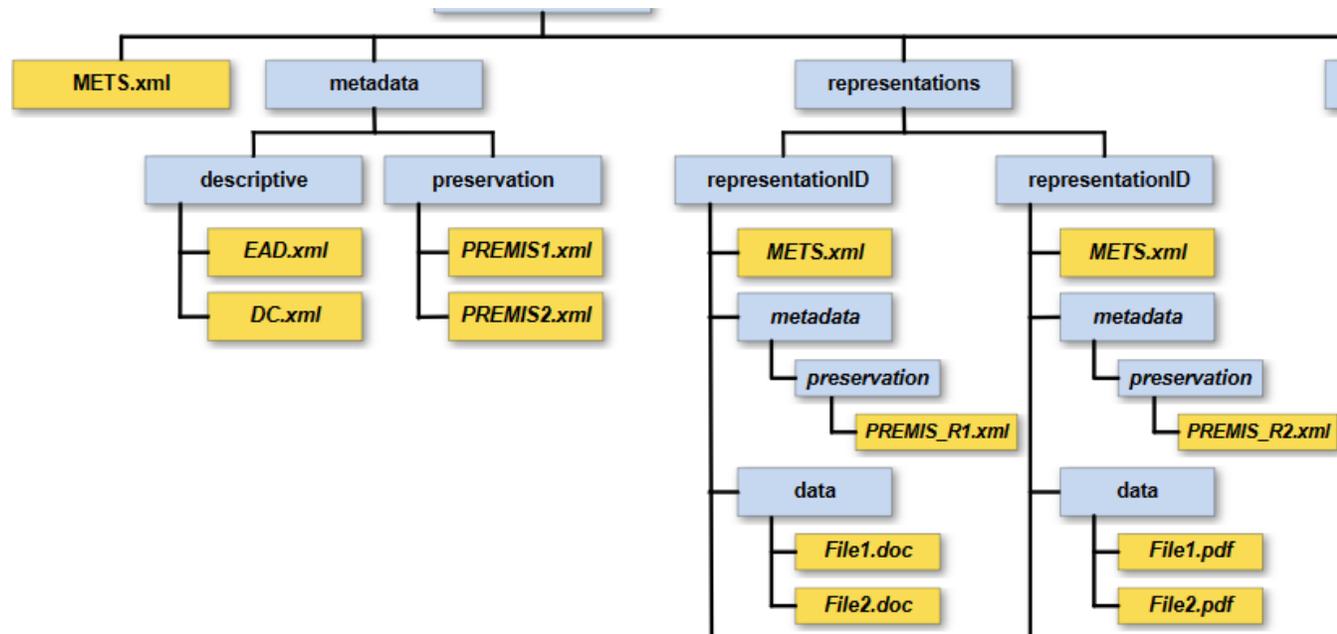
AUFBAU SIP

BAGIT

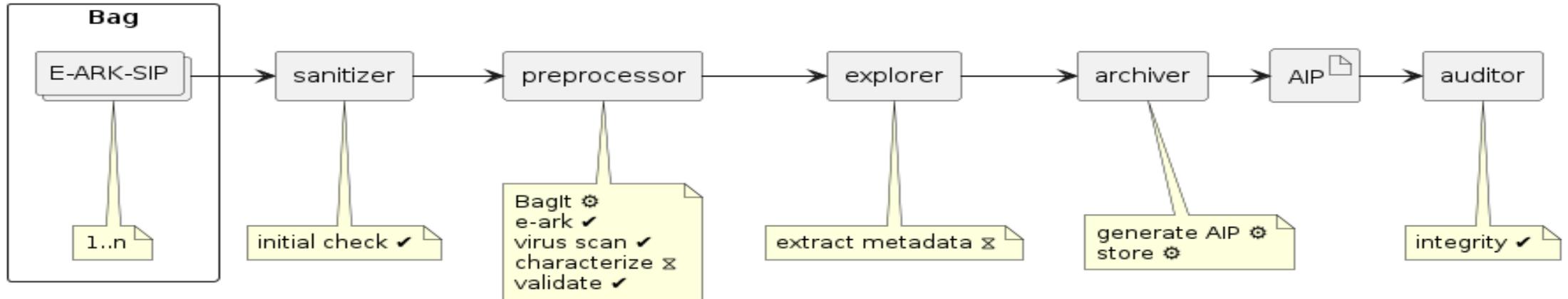
Examplebag/

- bagit.txt
- bag-info.txt
- manifest-sha512.txt
- tagmanifest-sha512.txt
- data/
 - [Struktur E-ARK]





<https://eakcsip.dilcis.eu>



- Preprocessor
 - 1. Virencheck
 - 2. BagIt Validierung und Extraktion (strukturelle Integrität, Prüfsummencheck)
 - 3. Umwandlung in E-ARK, falls vom Kunden nicht entsprechend angeliefert
 - 4. Formaterkennung (PRONOM, DROID) und Validierung (JHOVE)
 - 5. Erzeugung von PREMIS-Daten

AIP–Archival Information Package

- Automatisierte Audits, die Datenintegrität überwachen
- Kontinuierliche (Re-)Evaluierung bestehender Daten anhand neuerer Signaturen und Verfügbarkeit von Validatoren
- Vorgefertigte Reports zu Struktur der archivierten Daten (erkannte Formate, Validierungsstatus) und Risikoeinschätzung

Acrobat PDF 1.4 - Portable Document Format (fmt/18)	1135	2.46 GB	557	6 / 10
Acrobat PDF 1.5 - Portable Document Format (fmt/19)	348	1.20 GB	200	6 / 10
Acrobat PDF 1.6 - Portable Document Format (fmt/20)	367	1.09 GB	116	7 / 10
Acrobat PDF 1.7 - Portable Document Format (fmt/276)	186	253.58 MB	178	8 / 10
Acrobat PDF/A - Portable Document Format (fmt/95)	54	48.46 MB	35	9 / 10
Acrobat PDF/A - Portable Document Format (fmt/354)	213	355.42 MB	174	9 / 10
Acrobat PDF/A - Portable Document Format (fmt/477)	31	170.30 MB	31	9 / 10
MPEG 1/2 Audio Layer 3 (fmt/134)	75	3.38 GB	60	8 / 10
Microsoft Word Document (fmt/40)	1	2.52 MB	1	6 / 10

DIP–Dissemination Information Package

- **DIP** kann ausgespielt werden als:
 - Vollständiges AIP
 - AIP mit einzelner Repräsentation
 - Dateien des AIP + abgeleitete Formate
- Provenienz-/Eventdaten aus dem Archivsystem (PREMIS, inklusive aus Ursprungssystem übernommener Provenienzdaten) sind immer vorhanden

Die Österreichische Nationalbibliothek

- Implementierungsprojekt
 - 04.12.2020 – 08.06.2022
- Produktiv seit 08.06.2022
- Bisher im System
 - ca. 120 TB
 - 150.000 Objekte mit 5.000.000 Files
- Datenquellen:
 - Digitalisierungsprojekte
 - Born Digital Material
 - gecrawlte Webseiten
 - Bundesgesetzblätter, elektronische Dissertationen

Die Wirtschaftsuniversität zu Wien (WUW)

- Implementierungsprojekt
 - 28.06.2022 – 02.10.2023
- Ziele:
 - Einrichten einer eigenen Instanz der LZA-Plattform Libsafe
 - Festlegen auf Standards, welche für alle weiteren Teilnehmer gelten
 - Definieren von Workflows für die Anforderungen der WUW
 - Entwickeln von unterstützenden Tools und Integrationen mit Blick auf Wiederverwendbarkeit
- Datenquellen:
 - Goobi (E-ARK-Plug-in)
 - Archivdaten aus Acta Pro (geplant)

Fachliche Betreuung

- Analyse der zu archivierenden Daten gemeinsam mit den KundInnen
- Systemkonfiguration
- Laufende Anpassungen der Workflows falls notwendig
- Durchführung beauftragter statistischer Auswertungen
- Durchführen der Risikoanalyse und des Preservation Planning gemeinsam mit den KundInnen

Support

- Die OBVSG übernimmt den First-Level-Support für den Dienst und fungiert als alleinige Ansprechpartnerin für die Anwendenden
- OBVSG-Ticketsystem dient der strukturierten Kommunikation zwischen den KundInnen und der OBVSG

Systemadministration

- Gewährleistung des technischen Betriebs des Systems
- Wartung und Betreuung der Datenbank und der Applikation
- Wartung und Betreuung von Hardware und Betriebssystem
- <https://www.obvsg.at/services/digitale-langzeitarchivierung>

- Vorhandensein eines Repositoriums bzw. Erschließungssystems wird vorausgesetzt
- Kein Zugriff für Endbenutzer*innen – nur für berechnigte Mitarbeiter*innen über dezidierte Schnittstellen (Dark Archive)
- „Eigentümer“ der digitalen Daten bleibt die anliefernde Institution
- Speicher für AIPs kann bei OBVSG und/oder bei Kunden liegen
- Cave: LZA-System ist kein Katalogisierungs- und kein Discoverysystem
- „Kustodiale“ Arbeit bis zur Generierung des SIP bleibt beim Kunden

WORKFLOW DER WUW

= WORKFLOW FÜR ALLE WEITEREN KUNDEN

VORTEILE DES KONSORTIALEN DIENSTES

- Weniger Investitionen in Hard- und Software für jeden einzelnen Mandanten
- Kein eigenes Personal für Programmierung nötig
- Know-how der OBVSG
- Homogenes SIP (Nutzung aller eigens entwickelten Komponenten für unterschiedliche Mandanten)
- Organisation des Zusammenspiels von Rechenknoten, verschiedenen Speicherorten und institutionellem Zugriff
- In der Summe eine Effizienzsteigerung für alle teilnehmenden Einrichtungen

VIELEN DANK FÜR IHRE AUFMERKSAMKEIT!

silvio.wiese@obvsg.at
office@obvsg.at