

KONSORTIALE LANGZEITARCHIVIERUNG IM OBV: EINE DIENSTLEISTUNG VON UND FÜR BIBLIOTHEKEN

LAVAH-WORKSHOP, 25.09.2025

FOLIEN: SILVIO WIESE ; BERNHARD HAMPEL-WAFFENTHAL ; BETTINA KANN

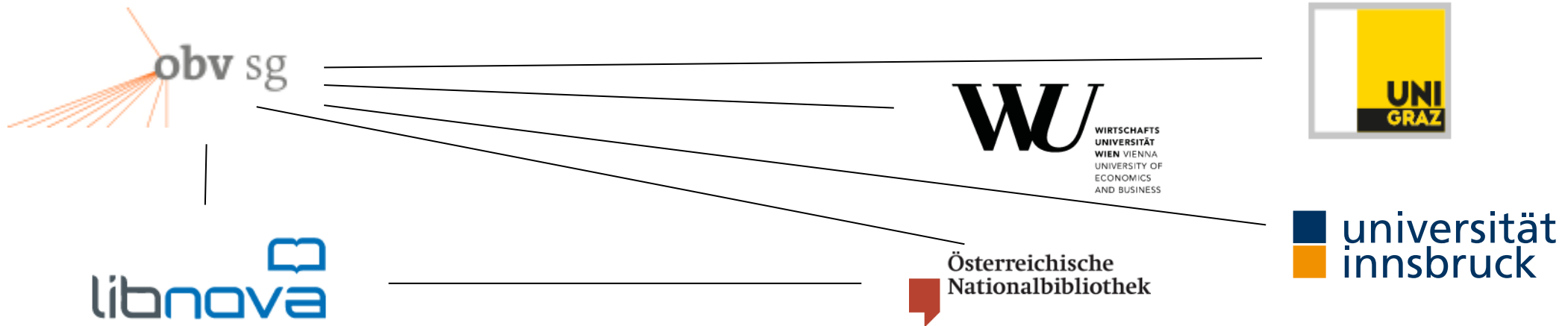
- Langzeitarchivierung mit Libsafe Advanced Pro
 - Ein Überblick
 - Der österreichische Bibliothekenverbund (OBVSG)
 - LZA in der OBVSG: Wie alles begonnen hat
 - Ziele und Ausgangsbedingungen / Grundsatzentscheidungen
 - Libsafe Advanced Pro: Kunden, Workflows und angeschlossene Systeme
 - Einfluss der Libsafe Architektur auf unsere Workflows
 - Einblicke in die Technische Umsetzung
- Ausblick

DIE ÖSTERREICHISCHE BIBLIOTHEKENVERBUND UND SERVICE GMBH (OBVSG)

WER SIND WIR

- 1. Die OBVSG ist die Verbundzentrale für den Österreichischen Bibliothekenverbund (OBV)
 - Verbundkatalogisierung (Über 90 Bibliotheken arbeiten im Verbund zusammen)
 - Nationaler URN-Resolver
 - Und vieles weitere
- 2. Die OBVSG ist auch als Application Service Provider tätig
 - Betreuung von lokalen Bibliothekssystemen (Alma für derzeit ca. 45 Institutionen)
 - Betreuung lokaler Suchmaschinenportale (Primo derzeit ca. 50 Einrichtungen)
 - Betrieb der Visual-Library-Serviceplattform (15 Teilnehmer)
 - Betrieb von LIBSAFE als digitales Langzeitarchivierungssystem im konsortialen Umfeld (4 Teilnehmer)

- **2018 – 2020** Ausschreibung gemeinsam mit ÖNB; Gewinner: LIBNOVA mit LIBSAFE Advanced Pro (On-premise)
- **2020 – 2022** Projekt: Implementierung für die Österreichische Nationalbibliothek (ÖNB)
- **2022 – 2023** Projekt: Anbindung der Wirtschaftsuniversität Wien (WUW)
- **2024 – 2025** Projekt: Anbindung der Universität Graz
- **2025 – 2026** Implementierung Universität Innsbruck



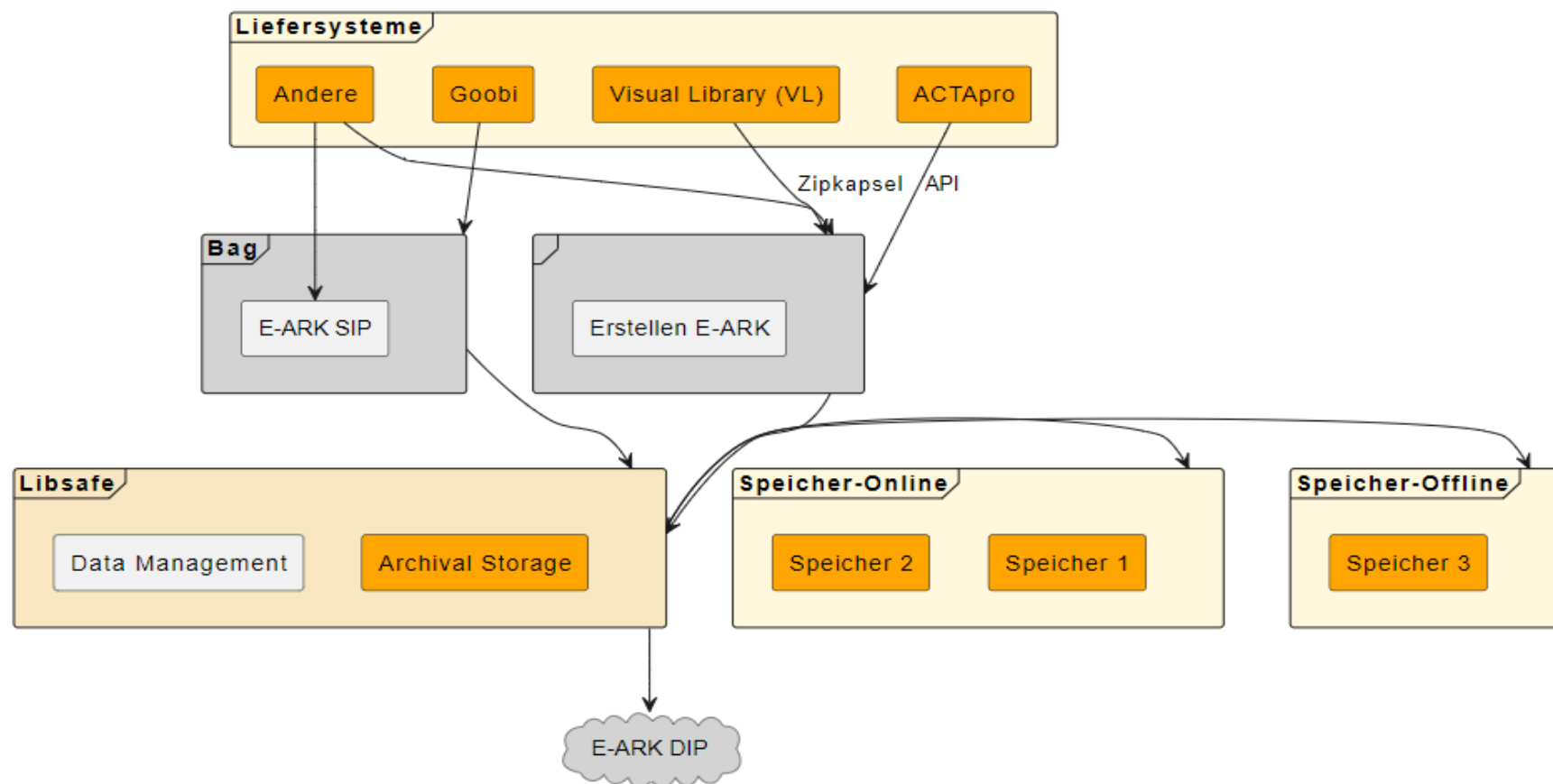
- **Ziele:**

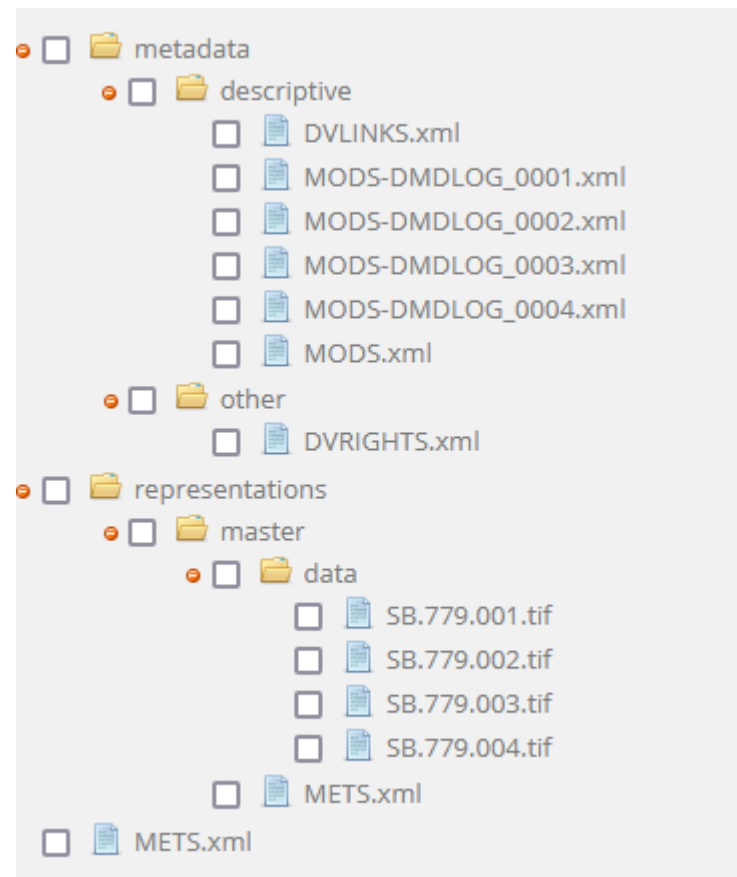
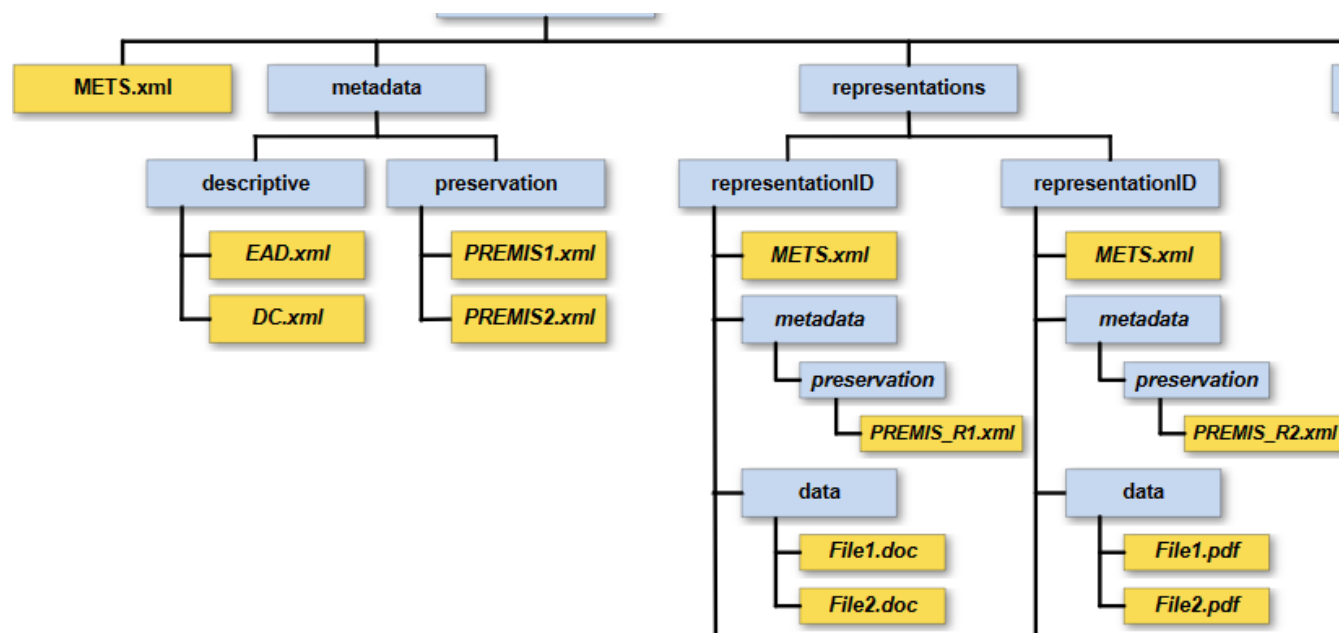
- Aufbau eines konsortialen LZA-Dienstes für Österreich
 - Kompetenzbündelung
 - Synergien durch die Nachnutzbarkeit von anfallenden Entwicklungen
 - Hohes Interesse an Findung passender Paketstandards (SIP, AIP, DIP), auf denen Eigenentwicklungen aufbauen können
 - Was validiert werden kann soll valide sein
 - Nutzung als Dark Archive

- Mandantenfähigkeit
- Datenhoheit und -souveränität
- Interoperabilität
- Skalierbarkeit
- Erweiterbarkeit
- Zugriffsrechte und Rollenmanagement
- Sicherheit und Zugriffskontrollen
- Exit-Strategie

- Windows-basiertes System
- System besteht aus einer oder mehreren Nodes mit Services
- Ingests durchlaufen mehrere Services, manche rufen externe Prozesse auf
- Präprozessoren, Metadatenextraktoren und "Ingestion Checks" sind erweiterbar
- Formaterkennung mit DROID, Formatvalidierung mit JHOVE
- Erweiterung mittels ausführbarer Programme - wir verwenden für deren Programmierung C# (u.a. Virus Checks, SIP-Validierung, Metadatenextraktoren)
- "Transferkonnektoren" als Erweiterungsmodule zur Anbindung anderer Systeme
- Metadaten werden für systeminterne Suche sowohl strukturiert aus (deskriptiven) Metadaten abgeleitet, andere Typen auch als "Extended Metadata" unverändert als im Freitext durchsuchbare Elemente (beides unabhängig vom Objekt versionierbar)
- <https://libnova.com>
- <https://docs.libnova.com/libsafe-advanced>

Langzeitarchivierung in der OBVSG





- keine inkrementelle Versionierung auf File-Ebene
 - Bei Änderungen am Objekt immer ein vollständiges neues SIP
 - Verwendung des Mechanismus von Extended Metadata (versionierbar)
 - Dadurch Entstehung eines "logischen" AIPs
 - Vollständiges DIP: „Einsammeln“ der Versionen der Extended Metadata
- Fehlerhandling (JHOVE-Fehlermeldungen)
 - Das Verhalten von Libsafe bei JHOVE-Fehlermeldungen ist nicht granular konfigurierbar
 - JHOVE-Output wird nicht strukturiert in LIBSAFE abgelegt (Improvement Request)
 - Bei Fehler Ingest erlauben oder nicht erlauben
 - Workaround:
 - 1. Alles Einspielen mit „strengem“ Preservation Plan (keine JHOVE-Fehler erlaubt)
 - 2. Analyse der Fehler
 - 3. Einspielen mit Preservation Plan, der JHOVE-Fehler ignoriert

LIBSAFE ADVANCED PRO: TECHNIK KOMPONENTEN

11

- Angelpunkt einer LIBSAFE-Installation ist die Datenbank (eine pro Mandant, MySQL)
- Für mandantenübergreifendes Reporting eine "konsolidierte" Datenbank
- Ein LIBSAFE-System besteht aus mindestens einer Windows-Node mit Windows-Services
- Webinterface & API
- Suchindex für strukturierte Suche in der Datenbank, zusätzlich erweiterte Suche mit Elasticsearch
- Reporting (auch) mit integrierter Metabase-Instanz

LIBSAFE ADVANCED PRO: TECHNIK STORAGE

12

- LIBSAFE unterstützt für den Archivspeicher die Anbindung von:
 - "regulärem" Speicher (lokal, SMB Netzwerk-Shares)
 - S3
 - "LIBSAFE" Storage (hosted)
- Wir verwenden bisher nur Netzlaufwerke über SMB, plus Offline-Bandsicherung
- Organisiert als Disks in Storage Groups
- Regelmäßige Audits von wechselnden Teilbeständen
- Speicher kann von der Institution oder uns betrieben werden

LIBSAFE ADVANCED PRO: TECHNIK

PRESERVATION AREAS

13

- Ein LIBSAFE-System wird logisch in Preservation Areas unterteilt
- Preservation Areas haben immer einen aktiven Preservation Plan
 - Neue Pläne können abgeleitet werden
 - Der aktive Plan kann anlassbezogen temporär gewechselt werden
- Rechteverwaltung daran geknüpft
- Dies ermöglicht Trennung der Sichtbarkeit unterschiedlicher Sammlungen

LIBSAFE ADVANCED PRO: TECHNIK

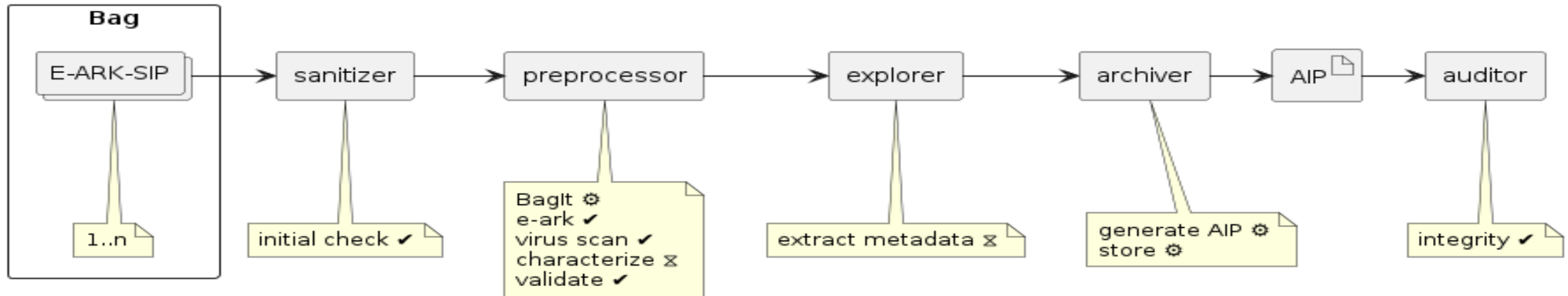
PRESERVATION PLANS

14

- "Preservation Plans" sind die Definition der Verarbeitungsschritte, die ein SIP durchläuft
- Am Ende ein AIP, das auf alle Disks der im Plan konfigurierten Storage Groups transferiert wird
- Überprüfung der Prüfsummen (MD5, SHA-1 oder SHA-256) und Audit von LIBSAFE am Ende des Transfers
- Validierungen vorab abhängig von konkret konfigurierten Präprozessoren oder Ingestion Checks
- Formatidentifizierung und –validierung mit konfigurierbarer Policy
 - Validierungspolicy per PRONOM-ID konfigurierbar
- Metadatenextraktion (Schema + Extended Metadata)
 - Extended Metadata auch für technische Metadaten verwendet

LIBSAFE ADVANCED PRO: TECHNIK INGEST (BEISPIEL WUW)

15



1. Entpacken der Bag-Archive (tar, gzip oder zip)
2. Virenprüfung mit ClamAV (an zweiter Stelle wegen eines ClamAV-Bugs unter Windows)
3. BagIt-Validierung (inkl. BagIt-Profiles) - wenn valide werden enthaltene Objekte herausbewegt
4. Umwandlung in E-ARK SIP, falls vom Kunden nicht schon als solches angeliefert
5. Validierung des E-ARK SIP
6. Formaterkennung (PRONOM IDs via DROID) und -validierung (JHOVE)
7. Metadatenextraktion (technisch, deskriptiv & preservation aus E-ARK Struktur)

Jeder Verarbeitungsschritt ist erfasst und wird bei DIP-Export mittels eines selbst entwickelten Profils als PREMIS angereichert.

LIBSAFE ADVANCED PRO: TECHNIK

PRESERVATION TOOLS

16

- Erweiterbare "Preservation Tools": Transferkonnektoren, Präprozessoren, (Advanced) Metadata Filters, Ingestion Checks, DIP-Profiles
- DROID zur Formatidentifikation und JHOVE zur Formatvalidierung fest verankert und nicht über das Webinterface konfigurierbar
- Bisher entwickelt:
 - Präprozessoren: Virus Checker (mit ClamAV), Archive Extractor (zip, tar oder tar.gz), Bag Processor (Validierung und "Entpacken" von Bags), E-ARK SIP-Validator (mit commons-ip), Konverter (Visual-Library-Kapseln -> E-ARK SIP, ACTApro API -> E-ARK SIP)
 - Metadatenfilter: XSLT (für strukturierte Metadaten), E-ARK Metadatenextraktor (Extended Metadata)
 - DIP

- Die LIBSAFE-Plattform selbst produziert keine Events im PREMIS-Format
- Präprozessoren und Metadatenfilter können optional PREMIS-Events verzeichnen lassen
- Alle anderen Prozesse am Objekt müssen bei Erstellung eines DIP zusammengetragen werden
- Retrieval Jobs dazu optional mit "DIP-Profil"
- Wir versuchen im Rahmen der Möglichkeiten der API alle Informationen als PREMIS-Event-XML über ein selbst entwickeltes Profil in ein E-ARK DIP mitzunehmen
- Dies soll auch frühere Versionen von Extended Metadata und vielleicht auch Objektversionen umfassen

DIP–Dissemination Information Package

- **DIP** kann ausgespielt werden als:
 - Vollständiges AIP
 - AIP mit einzelner Repräsentation
 - Dateien des AIP in abgeleiteten (Präsentations-) Formaten (z.B. .jpg aus .tif)
 - Dateien des AIP + abgeleitete Formate
- Provenienz-/Eventdaten aus dem Archivsystem (PREMIS, inklusive aus Ursprungssystem übernommener Provenienzdaten) immer vorhanden

Fachliche Betreuung

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

Support

- First-Level-Support

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

- Weniger Investitionen in Hard- und Software für jeden einzelnen Mandanten
- Kein eigenes Personal für Programmierung / Technik bei Einrichtungen nötig
- Know-how – Transfer
- Homogene IPs (Nutzung aller entwickelten Komponenten für alle Mandanten)
- Organisation des Zusammenspiels von Rechenknoten, verschiedenen Speicherorten und institutionellem Zugriff
- In der Summe eine Effizienzsteigerung für alle teilnehmenden Einrichtungen

- Installation weiterer Mandanten und damit Einbindung weiterer Datenquellen
- Ausbau der Anzahl der Dateiformate für die Validierungen möglich sein soll
- Ausbau von Know-how
- Weiterentwicklung Geschäftsmodelle für weitere Datenquellen (Forschung)
- Erstellung von Dokumentationen
 - zu den einzelnen Workflows
 - Guidelines
 - Sammeln von Best Practice Beispielen

VIELEN DANK FÜR DIE AUFMERKSAMKEIT!