



Rapport d'atelier archives numériques de la Bibliothèque nationale suisse

Basil Marti et Nadia Lai, Community Day docuteam, 21.11.2024



Archives numériques à long terme de la Bibliothèque nationale Suisse

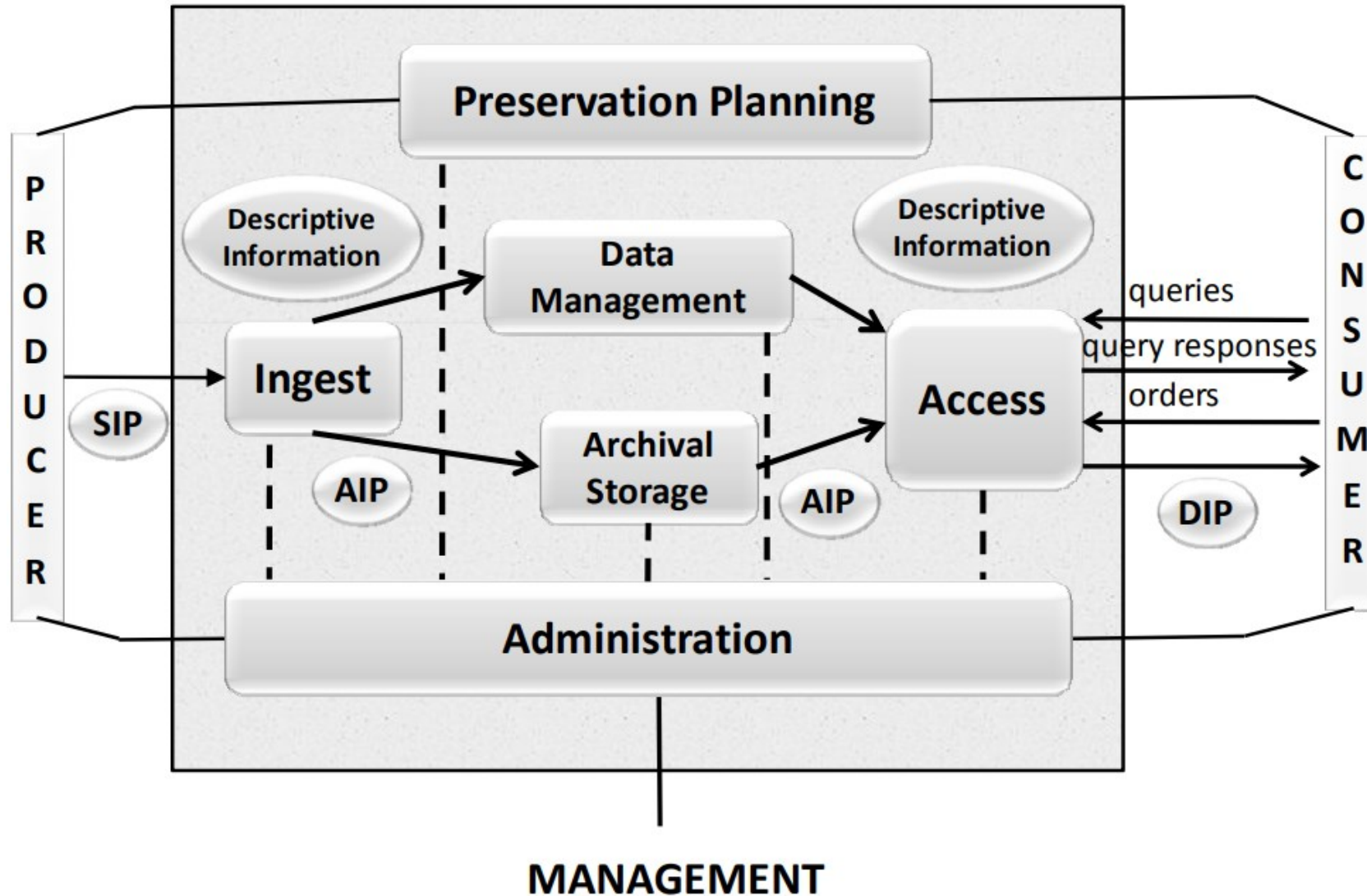
Situation actuelle

- **Mandat de la Bibliothèque nationale suisse**
La Bibliothèque nationale a pour mandat de collectionner, de répertorier, de conserver, de rendre accessible et de faire connaître les informations imprimées ou conservées sur d'autres supports que le papier, ayant un lien avec la Suisse.
- **Sources**
Maisons d'édition, auto-éditeurs, production propre (numérisation), Internet
- **Données**
Monographies, revues, normes, thèses, sites internet, documents sonores, ...
- Principaux systèmes d'archivage à long terme (modèle OAIS) à la BN
Deposit, WAU (WebArchiveUtilities), Ingest, Stockage, Access
(tous sont des développements individuels)

Projet LZA (Langzeitarchivierung)

- **Pourquoi changer?**
 - Les contrats pour e-Helvetica et l'Archivio multimediale sont arrivés à échéance fin 2021.
Ils ont été prolongés jusqu'à la fin du projet (fin 2025).
 - Du point de vue de l'architecture, **la BN a deux systèmes différents**, qui effectuent des tâches similaires.
- **Objectifs principaux du projet**
 - Les systèmes d'archivage des collections numériques pour e-Helvetica et la Phonothèque sont remplacés et **réunis en une seule application**.
 - Le système est en mesure de gérer **toutes les collections** ou des parties de collections **numériques et numérisés**.
 - Le système répond aux exigences du **Preservation Planning**.
 - Le système peut faire face à la croissance du **volume de données** et à **l'évolution des besoins**.
 - Le système peut être **certifié** en tant que système d'archivage à long terme fiable.
 - Si nécessaire, le système peut être mis à disposition **d'autres institutions**.

Nouvelle architecture de système



Projet LZA (Langzeitarchivierung)

- **Planification actuelle / Etapes:**

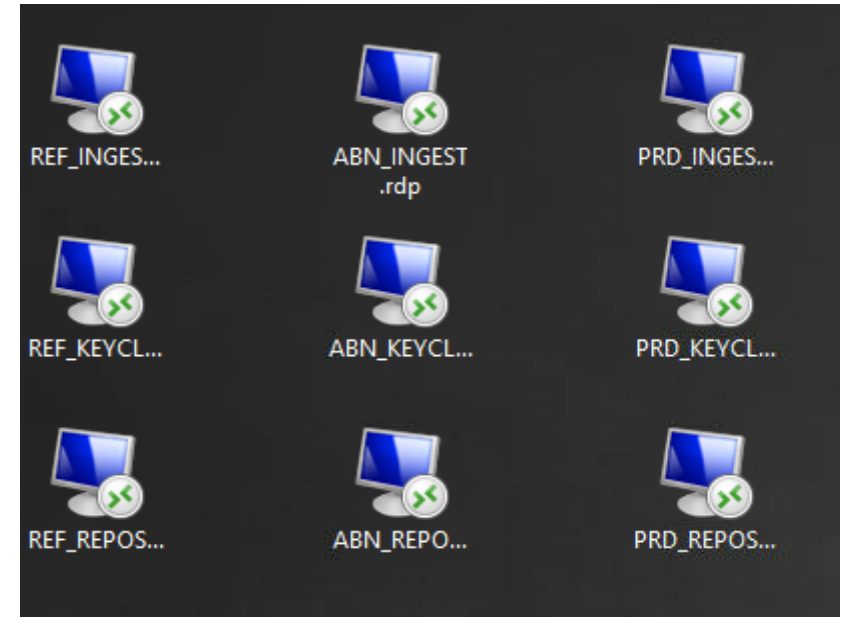
- 4 Appels d'offre (OMC)
Archives digitales, Pre-Ingest, Access, Hosting
- Acceptation (validation) des différents systèmes
Fin 2024 / Début 2025
- **Migration des données:**
Phonothèque: 4 Septembre 2024 jusqu'à ~Janvier 2025
e-Helvetica: ~Janvier 2025 jusqu'à ~Mai 2025
- **Mise en production:**
2^{ème} moitié 2025
- **Remplacement** de l'ancien système:
Fin 2025 les vieilles applications seront arrêtées

Archives numériques à long terme

- Objectifs des archives à long terme selon l'architecture du système :
 - Réception des paquets du Pre-Ingest
 - Pas de paquets Matterhorn METS, mais un format de paquet standardisé
 - Le cas échéant, directement à partir des sources
 - Stockage des données et des métadonnées, indexation pour la recherche
 - Preservation Actions
 - Versement vers Access
 - Accès réservés aux collaborateurs de la BN via feeder et box

Archives numériques à long terme

- Architecture du système
 - Installation de docuteam cosmos standard
 - Installation „on premises“ sur l’infrastructure de l’OFIT, → installation et exploitation par docuteam
 - 3 environnements (REF, ABN, PRD)
 - 3 serveurs Windows par environnement (Ingest, Repository, Keycloak) avec des bases de données postgres
 - Keycloak: Connexion au système d’authentification fédérale (eIAM)
 - Stockage Fedora: Sur un serveur S3
 - Connexion facile à configurer dans Fedora
 - Vitesse pas encore optimale, mais amélioration sensible avec Fedora 6.5.1.
 - Attention lors de l’utilisation d’une système S3 pas fourni par Amazon, en partie incompatible avec le standard (mais fonctionnel avec Dell ECS)













Archives numériques à long terme

- Format du paquet
 - Format de paquet identique pour e-helvetica et Phonothèque (le cas échéant, mandants supplémentaire)
 - Paquet Matterhorn METS:
 - Root Folder (Cote pour FN, URN dans e-helvetica)
 - data: Données primaires, structure quelconque, le cas échéant sous-dossiers
 - metadata: métadonnées supplémentaires au format xml
 - Métadonnées issues du système de gestion des collections (Alma, FN-Base)
 - Métadonnées issues des anciennes archives à long terme



Archives numériques à long terme

CD37244

data

 CD37244.0.JPG	02.09.2009 07:47	JPG File	46 KB
 CD37244_1.mp3	03.09.2009 05:04	MP3 Format Sound	12'587 KB
 CD37244_1.wav	03.09.2009 06:57	Wave Sound	140'422 KB
 CD37244_1.wav.xml	03.09.2009 06:57	XML Document	5 KB
 CD37244_1.xml	03.09.2009 07:00	XML Document	17 KB
 CD37244_2.mp3	03.09.2009 05:06	MP3 Format Sound	7'478 KB
 CD37244_2.wav	03.09.2009 06:59	Wave Sound	83'665 KB
 CD37244_2.wav.xml	03.09.2009 06:59	XML Document	8 KB
 CD37244_2.xml	03.09.2009 07:01	XML Document	17 KB
 CD37244_3.mp3	03.09.2009 05:31	MP3 Format Sound	9'693 KB

metadata

-  CD37244_meta.XML
-  CD37244_rights.XML

 mets.xml

Archives numériques à long terme

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- created with FN-Base32 version 9.xx by Stefano S. Cavaglieri (Fonoteca Nazionale Svizzera, Lugano-CH) -->
<FNBase32Mono xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="https://www.fonoteca.ch/FNBase32Mono.xsd">
  <Objects>
    <Obj uniqueKey="118650.011" lang="it">
      <objSchema>
        <objSchemaFormatK relTable="Files" relKey="16.011">Disco compact</objSchemaFormatK>
        <objSchemaProfileK relTable="Files" relKey="69.011">Pro Helvetia</objSchemaProfileK>
      </objSchema>
      <objDetail headSection="1">
        <objDetailFieldK relTable="Fields" relKey="78.011" dcTags="Titolo;DC - Titolo">Titolo del supporto</objDetailFieldK>
        <objDetailValueR dataType="9">Lucrari de compositori elvetieni interpretate in festivalu international de muzica academica</objDetailValueR>
        <objDetailGrp>000.001</objDetailGrp>
      </objDetail>
      <objDetail headSection="0">
        <objDetailFieldK relTable="Fields" relKey="75.011" dcTags="DC - Formato">Tipo di supporto</objDetailFieldK>
        <objDetailValueR dataType="0" relTable="Headings" relKey="6748.011">Disco compact registrabile</objDetailValueR>
        <objDetailGrp>000.002</objDetailGrp>
      </objDetail>
      <objDetail headSection="0">
        <objDetailFieldK relTable="Fields" relKey="64.011" dcTags="DC - Tipo">Tipo di produzione</objDetailFieldK>
        <objDetailValueR dataType="0" relTable="Headings" relKey="9.011">Demo</objDetailValueR>
        <objDetailGrp>000.003</objDetailGrp>
      </objDetail>
    </Obj>
  </Objects>
</FNBase32Mono>
```

Archives numériques à long terme

- Métadonnées
 - En cas d'ingest ou de la migration, enrichissement des métadonnées METS à partir des fichiers de métadonnées du paquet (XSLT)
 - Mises à jour des métadonnées
 - Manuellement dans docuteam box
 - Mises à jour automatisées à partir des systèmes de gestion des collections
 - Le flux de travail feeder récupère les mises à jour via OAI-PMH depuis Alma ou FN-Base.
 - Nouvelles actions feeder et box API Endpoint pour la mise à jour automatisée des métadonnées dans l'archive

Archives numériques à long terme

- Flux de travail pour l'ingest et la migration
 - Ingest standard, mais enrichi de fonctions nouvellement développées
 - Enregistrement d'URN
 - Un URN est obtenu auprès de la bibliothèque nationale allemande
 - L'URN est enregistré en combinaison avec l'URL
 - Contrôle des doublons via la somme de contrôle des fichiers primaires ou des éléments de métadonnées (comme la cote)
 - Importation de métadonnées vers Alma (API REST pour l'importation de fichiers MARCXML)

Workflow: F4: Migrate FN packages, process package folder, and ingest, transfer

Automatische Seitenaktualisierung alle 10s

Erstellt durch Launch "F.4" after Fonoteca-Delta File is ready on NAS am Dienstag, 22. Okt
Zuletzt bearbeitet von Launch "F.4" after Fonoteca-Delta File is ready on NAS am Dienstag

SIP	//10.187.64.31/Fono_
Startzeit	Dienstag, 05. Novemb
Gesamtdauer	00:00:27

Submission: create SIPs from file or folder (00:00:27) ↻

Quality Assurance: check formats against submission agreement ⌵

Quality Assurance: fixity check ⌵

Ingest: enrich mets.xml from FNBASE metadata files ⌵

Ingest: rename root node based on EAD metadata ⌵

Storage: add PIDs from box ⌵

Storage: convert Matterhorn METS to Matterhorn RDF ⌵

Storage: import Matterhorn RDF into Fedora 6 ⌵

Pre-Ingest

- Pre-Ingest est une plateforme sous forme de divers **développements individuels**
- **Pre-Ingest Deposit**
Plateforme pour le téléchargement de publications par les petits éditeurs ou les auto-éditeurs
- **Pre-Ingest Général**
Reception d'objets et préparation per les archives numériques (par ex. Thèses, Wikipédia, normes, etc.)
- **Pre-Ingest Archives Web**
Webharvesting (Heritrix, Browsertrix, ...), Contrôle de qualité (VQI et manuellement das Replay-Tool (pywb, *OpenWayback*) et préparations pour les archives numériques



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Schweizerische Nationalbibliothek NB
e-Helvetica Deposit

Login Informationen

e-Helvetica Deposit

Auf [e-Helvetica Deposit](#) haben Sie die Möglichkeit, einzelne digitale Bücher und Zeitschriften an die Schweizerische Nationalbibliothek abzuliefern. Bevor Sie digitale Publikationen abliefern können, müssen Sie sich als Verlag einmalig registrieren lassen.

Login

Sie haben noch kein Login?

[Als Verlag registrieren](#)



Pre-Ingest Webarchiv

ABVA

Webarchiv

Websites Crawls Regeln

Neu Crawlten nächste 5 Tage Neu wartet QS light QS U Archiviert **Crawls heute** Crawls morgen

16 Resultate **Filtern**

<input type="checkbox"/>	Website	Status	Neu	Letztes Crawl Datum	Nächstes Crawl Datum	Partner	Collection	Crawls	Archiviert
	https://automnales.ch	Harvesting geplant		26.10.2022	23.10.2024			0	8
	https://les-capucins.ch	Harvesting geplant		19.10.2022	23.10.2024			0	6
	https://www.keramikpanorama.ch	Harvesting geplant		19.10.2022	23.10.2024			0	5
	https://cic-info.ch	Harvesting geplant		19.10.2022	23.10.2024			0	5
	https://www.buchhandlung-hirslanden.ch	Harvesting geplant		19.10.2022	23.10.2024			0	5
	http://www.kirchetrub.ch	Harvesting geplant		17.10.2022	23.10.2024			0	2
	https://www.cfeb.ch	Harvesting geplant		15.10.2022	23.10.2024			0	5
	https://www.emslancy.ch	Harvesting geplant		17.10.2022	23.10.2024			0	2
	http://www.racinotes.ch	Harvesting geplant		17.10.2022	23.10.2024			0	2
	https://www.benediktmeier.ch	Harvesting geplant		17.10.2022	23.10.2024			0	2
	https://montanaire.ch	Harvesting geplant		15.10.2022	23.10.2024			0	10
	https://pro.sankt-martin.ch	Harvesting geplant		17.10.2022	23.10.2024			0	4
	https://www.wuerth-haus-rorschach.ch	Harvesting geplant		21.10.2022	23.10.2024			0	5

Pre-Ingest Webarchiv

ABNA

Webarchiv > Crawls

3 von 4

108 - https://ochlenberg.ch

Finished



Details Crawl Log WARC Dateien

Snapshots

21.10.2024 14:51

Metadaten

Live und Crawl



Start
21.10.2024 14:51

Dauer
00:39:31

Queued
0

Grösse
451 MB

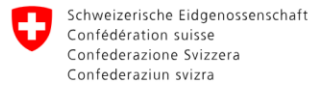
Crawl ID
108

Harvesting URL
https://ochlenberg.ch

Crawled
1826

Status Codes

- e-Helvetica Access sert d'interface de **présentation d'objets numériques** (Sites Web, thèses universitaires, livres, normes et revues, documents sonores)
- Principales fonctionnalités:
 - Interface web en 4 langues avec filtres et facettes.
 - Index de recherche conçu en 4 langues (texte intégrale et métadonnées pour toutes les publications)
 - Résultats de recherche affichés avec des aperçus générés automatiquement
 - Visionneuse entièrement intégrée pour les documents pdf, ePub et document Web.
 - Contrôle des accès flexible au moyen de la définition rôles et droits au niveau des documents.



Schweizerische Nationalbibliothek NB
e-Helvetica

Suche [Vitrinen](#) [Collage](#) [Über e-Helvetica](#)

Digitale Sammlungen NB und Webarchiv Schweiz

Entdecken sie digitales Kulturgut der Schweiz. Historische Websites, digitalisierte Tonträger und Videos, digital publizierte Bücher, Zeitschriften, Hochschulschriften und Normen sowie gedruckte Werke in digitalisierter Form.

[Was ist e-Helvetica](#) →



Stöbern Sie in der digitalen Sammlung der NB

 Zeitschrift 2	 Website 63	 Wikipedia 2	 Norm 3	 Hochschulschrift 62	 Artikel 1.876	 Buch 252
 Audio, Musik 48	 Audio, Gesprochenes 7	 Video, Gesprochenes 4	 Video, Musik 9			

Suche [Vitrinen](#) [Collage](#) [Über e-Helvetica](#)



Erweiterte Suche →

Filtereinstellungen beibehalten

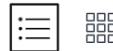
↓ Suche speichern

390 Treffer

25 Treffer pro Seite

Sortieren nach

Relevanz



1 von 16 < >

^ Ergebnisse filtern

- Neu in e-Helvetica
- Dieser Monat 0
- Letzte Woche 0
- Diese Woche 0

- Inhalt ^
- Artikel 326 x
 - Website 42 x
 - Buch 16 x
 - Audio, Musik 3 x
 - Audio, Gesprochenes 1 x
 - Hochschulschrift 1 x
 - Zeitschriftennummer 1 x

Website
Bern
● eingeschränkt verfügbar

Datum 14.08.2018
Verlag Zürich : Bank Julius Bär
Domain www.juliusbaer.com
Signatur bel-1191915
URN urn:nbn:ch:bel-1191915

[Alle Snapshots](#)

<https://www.juliusbaer.com/ch/de/menus/services/standorte/europa/schweiz/bern/>

Bern

22 Treffer →



Connexion d'Access aux archives à long terme

- Livraison de paquets Matterhorn METS à Access
- Mais : certains fichiers ou paquets ne doivent pas être livrés (définition dans l'élément **accessPolicy** de l'EAD).
- Développement d'une nouvelle action feeder qui vérifie le paquet et supprime les fichiers ou dossiers non autorisés.
- Téléversement sur serveur S3 et envoi d'un message
- Pourquoi Access ne va-t-il pas chercher lui-même les données ?
 - Architecture du système, pas d'accès externe au archive numérique possible

Kontakt

Basil Marti

Projektleiter digitale Archivierung

b.marti@docuteam.ch

docuteam AG

Im Langacker 16

CH-5405 Baden

T direkt +41 56 511 11 70

T +41 56 470 03 37

<https://www.docuteam.ch/>

