



Universität
Basel

Universitätsbibliothek

zhb Zentral- und
Hochschulbibliothek
Luzern

ONAS - OCFL Native Archive System

Universitätsbibliothek Basel
Jürgen Enge

Zentral- und Hochschulbibliothek Luzern
Beat Mattmann



Ausgangslage

Basel

- docuteam Infrastruktur (produktiv seit 02'2024)
- Über 200'000 ZIP Kapseln mit Retrodigitalisaten aus e-Rara/-Manuscripta, fast 3 Mio. TIFF-Dateien, ca. 700'000 Volltexte
- Universitätseigene Publikationen
- Spezialsammlungen (auch mit AV-Medien)
- Ca. 150TB Rohdaten

Grundsatzentscheid: Weiter mit bisheriger Infrastruktur oder einen neuen Ansatz wagen?

Luzern

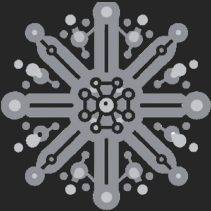
- Verschiedene Abklärungen über Jahre hinweg, doch keine Lösung in Sicht
- Vor allem Budget für externe Entwicklung, jedoch überschaubare Personalressourcen
- Keine Erfahrung & kein Know-How im Thema, jedoch viel Motivation
- Bedarf primär für Digitalisate und Repositories, jedoch kleinere Mengen

Am Scheideweg: Ausschreibung und kommerzieller Anbieter oder Kooperation mit "Branchenpartner"?

Grundprinzipien



«**AIP first**»: Bei diesem Ansatz ist das AIP nicht die Folge der Nutzung eines bestimmten Systems zur Archivierung, sondern der Fokus wird initial auf das AIP gelegt und die Funktionalität des DLZA-Systems daran orientiert.



«**Distributed Resources**»: Verschiedene Stakeholder sollen an dem System partizipieren können, ohne auf zentrale Speichersysteme zurückgreifen zu müssen. Minimalistische Zentralstrukturen verwalten verteilte Prozesse und Ressourcen.



«**Cloud native Technology**»: Einsatz aktueller Cloud-Technologien, um Flexibilität in fast allen Prozessen zu bewahren. Dies bedeutet nicht die zwingende Nutzung von «public» Cloud Providern, sondern kann durchaus auch in einer «private» oder «local» Cloud stattfinden.



«**Zero Trust Architecture**»: Die Systeme, welche im Einflussbereich der DLZA Infrastruktur stehen folgen der Zero Trust Strategie, sodass in einer adäquaten Systemumgebung (Virtualisierung, Netzwerk, Speicher, ...) ein sehr hoher Sicherheitsstandard ermöglicht wird.

Die Zusammenarbeit im Verbund-System

Eigenständige Aufgaben:

- Verantwortlich für eigene Bestände und zu archivierende Datenobjekte
- Betrieb der eigenen Pre-Ingest-Workflows
- Betrieb einer eigenen Ingest-Workbench
- Verantwortlich für die Storage-Bereitstellung und den Storage-Betrieb

Kooperative Aufgaben:

- AIP-Transfer ins gemeinsame Managementsystem und Verteilung auf Storage-Standorte
- Aufbereitung Metadaten und Bereitstellung für das Reporting-GUI
- Preservation Planning – Planung und Durchführung von Migrationen
- Erarbeitung und Pflege von Tools und Know-How

Die Basis für unsere Information Packages: OCFL



OCFL Specifications

This Oxford Common File Layout (OCFL) specification describes an application-independent approach to the storage of digital information in a structured, transparent, and predictable manner. It is designed to promote long-term object management best practices within digital repositories.

Specifically, the benefits of the OCFL include:

- **Completeness**, so that a repository can be rebuilt from the files it stores
- **Parsability**, both by humans and machines, to ensure content can be understood in the absence of original software
- **Robustness** against errors, corruption, and migration between storage technologies
- **Versioning**, so repositories can make changes to objects allowing their history to persist
- **Storage diversity**, to ensure content can be stored on diverse storage infrastructures including conventional filesystems and cloud object stores

Erweiterbarkeit

OCFL in unserem Archiv: gocfl



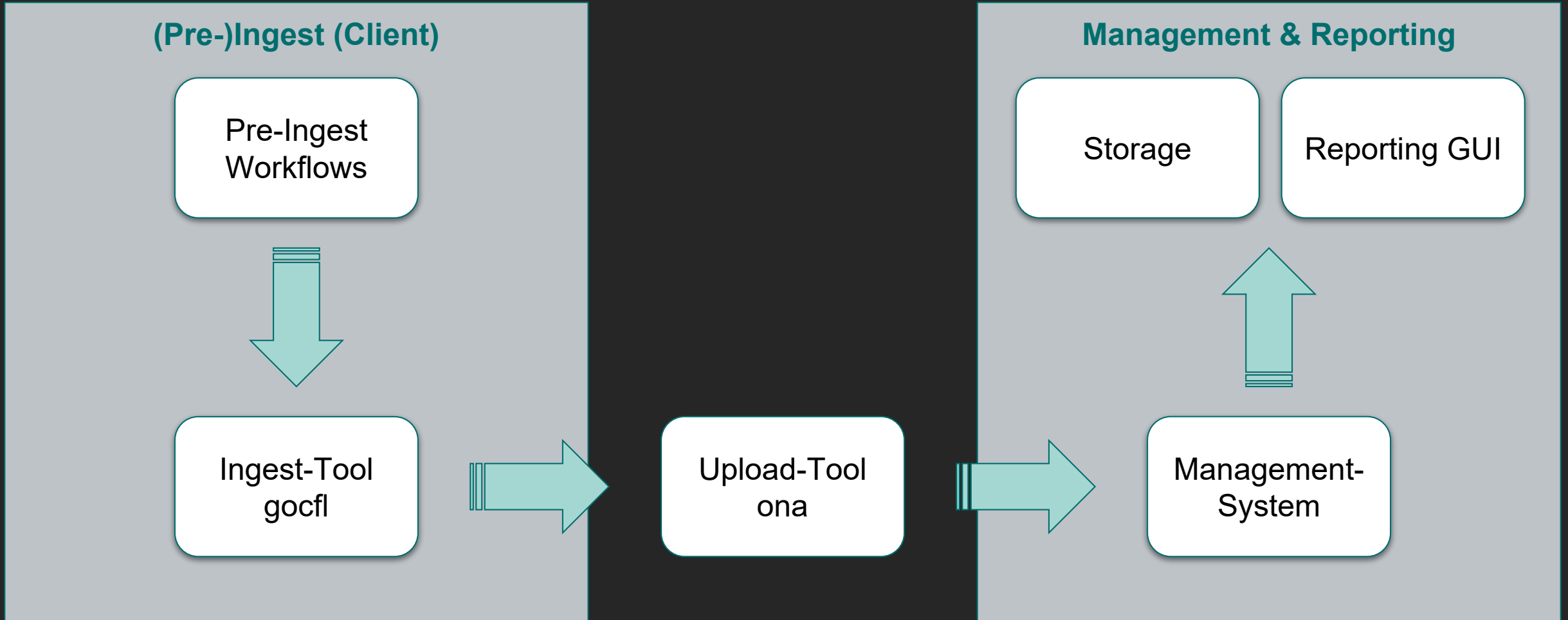
Kommandozeilenwerkzeug zur Erstellung von OCFL kompatiblen Strukturen (Verzeichnisbasiert oder ZIP)

Erweiterungen:

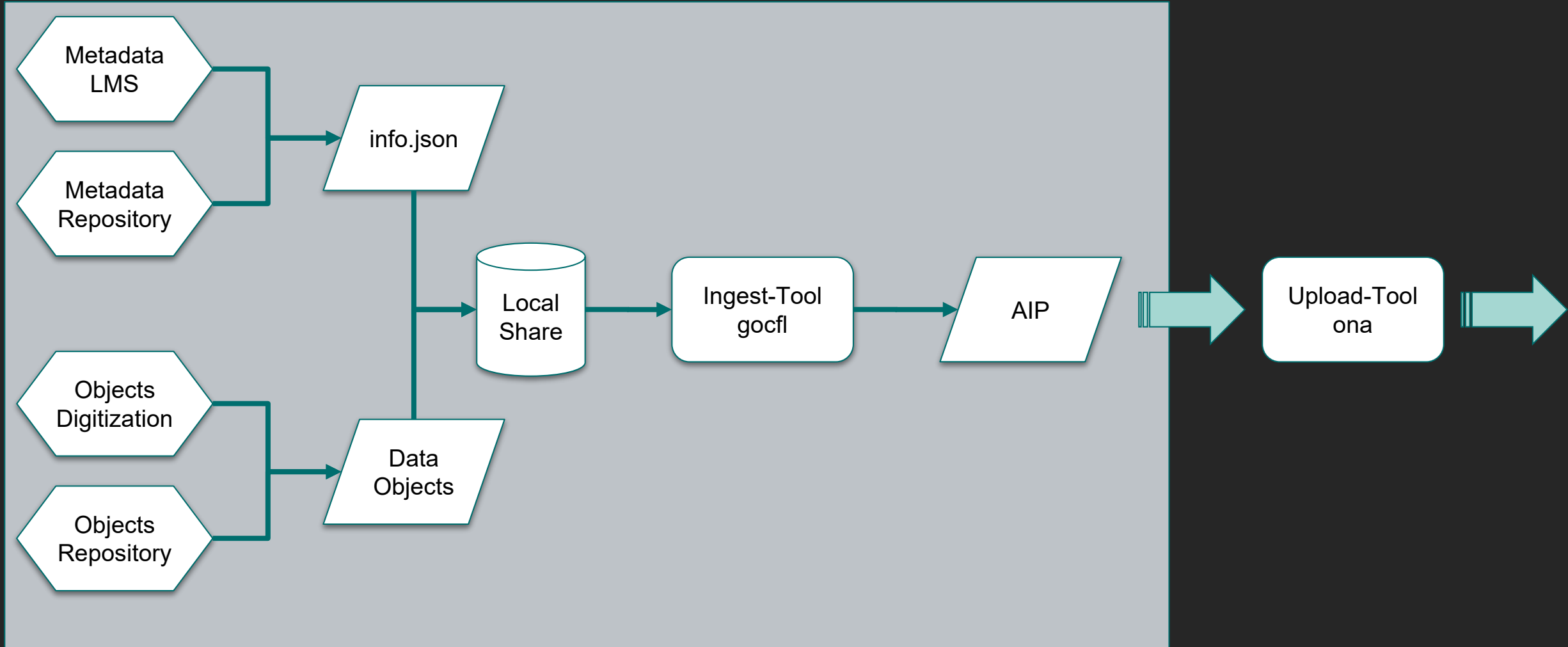
- Formaterkennung/tech. Metadaten via Indexer (Weiterentwicklung des Systems, das vom DLA-Marbach genutzt wird)
→ Siegfried // Tika // ffmpeg // Image Magick
- Generierung von METS/PREMIS angelehnt an DILCIS E-ARK AIP
- Dateisystem Metadaten
- Umbenennung in nachhaltige Pfad- und Dateinamen im AIP
- Thumbnail Generierung
- Automatische Migration (Prototyp)
- Grundlegende semantische Metadaten

Validierung / Reporting / Extraktion / Darstellung

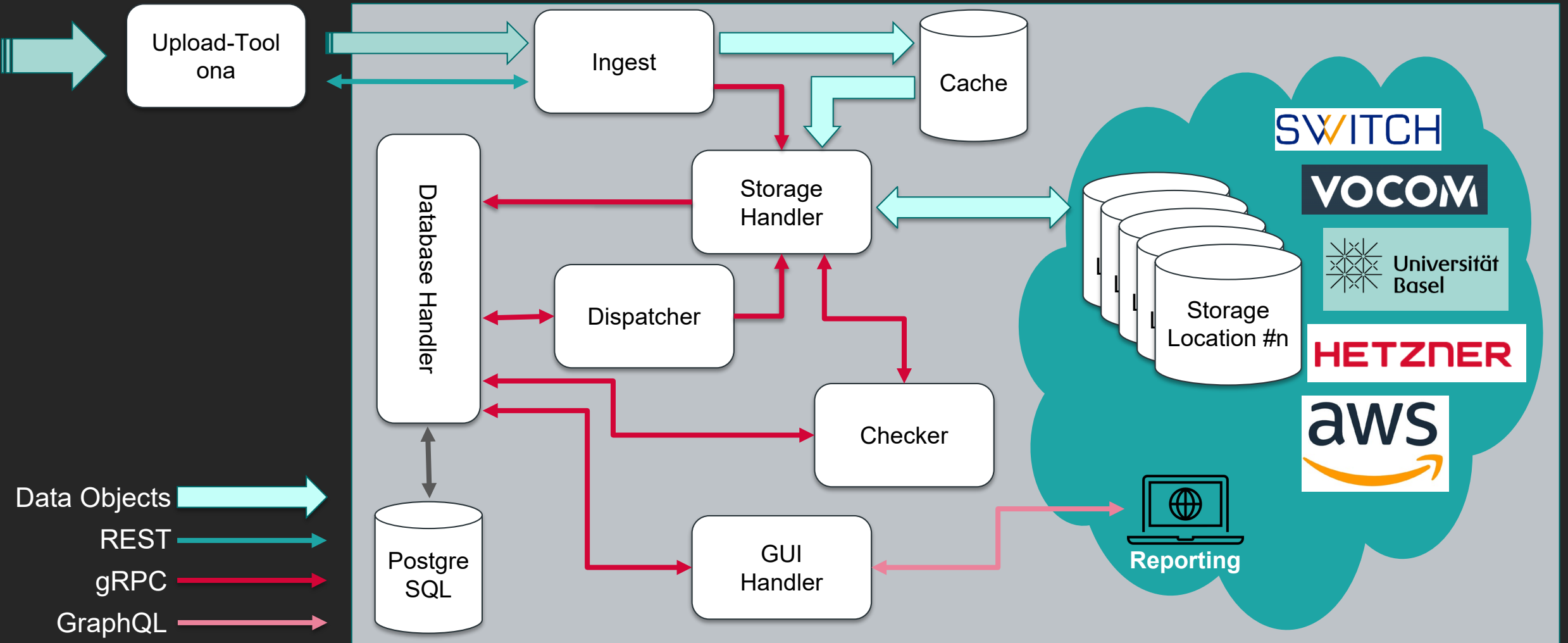
Archiv Management System



Archiv Management System: (Pre-)Ingest



Archiv Management System: Managementsystem & Reporting



Zentral in unserem Ansatz

Transparenz



Jede Organisation muss jederzeit transparente Informationen zu ihren Daten haben:

- Welche Objekte wurden gespeichert?
- Wie sehen die AIP der Objekte aus?
- Wo sind die Kopien abgelegt?
- Wann wurden sie zuletzt überprüft (Prüfsumme)?
- Wie viel Speicherplatz wird aktuell benötigt?
- Welche Medientypen befinden sich in den Objekten?

Zentral in unserem Ansatz

Flexibilität



Flexibilität in allen technischen Bereichen:

- Freie Auswahl von Speicherdienstleistern oder Herstellern (aktuell S3 Protokoll)
- Automatischer Datentransfer bei Speicherausfall oder Abschaltung auf Alternativsysteme
- Pre-Ingest Prozesse spezifisch implementierbar
- Modulare Ingest-Pipelines mit bewährten Tools

gocfl AIP Darstellung

Impression

gocfl

ub:test01

Überblick

Version

ub:test01

State v1

Manifest

Object

Manifest

Version v1

Version v2

Short Report

Full Report

Filesystem

[README.md](#)

fmt/1149 text/plain; charset=ISO-8859-1 171 B

[data/audio/Dragan_Espenschied_-_01_-_Mein_Leben_fr_die_Wissenschaft.ogg](#)

364 kB 2023-11-28T18:03:21+01:00 windows Archive fmt/203 audio/ogg 364 kB

[data/audio/Dragan_Espenschied_-_02_-_P2P.mp3](#)

4.7 MB 2023-11-28T18:03:21+01:00 windows Archive fmt/134 audio/mp3 4.7 MB

[data/audio/Dragan_Espenschied_-_02_-_P2P.ogg](#)

2.3 MB 2023-11-28T18:03:21+01:00 windows Archive fmt/203 audio/ogg 2.3 MB

[data/audio/Dragan_Espenschied_-_08_-_Zombie_Mummy_Theme.mp3](#)

5.0 MB 2023-11-28T18:03:21+01:00 windows Archive fmt/134 audio/mp3 5.0 MB

[data/audio/Dragan_Espenschied_-_09_-_Procrastination_Polka.ogg](#)

459 kB 2023-11-28T18:03:21+01:00 windows Archive fmt/203 audio/ogg 459 kB

[data/audio/Dragan_Espenschied_-_10_-_Оля_Зимой.mp3](#)

2.1 MB 2023-11-28T18:03:21+01:00 windows Archive fmt/134 audio/mp3 2.1 MB

[data/audio/Education_of_the_Noobz_1.doc](#)

60 kB 2023-11-28T18:03:21+01:00 windows Archive fmt/40 application/msword 60 kB

Mimetypes

- 1 application/gzip
- 2 application/illustrator
- 1 application/mp4
- 1 application/msword
- 1 application/octet-stream
- 6 application/pdf
- 1 application/postscript
- 1 application/vnd.oasis.opendocument

gocfl AIP Darstellung Impressionen

Checksums

sha512 79f7273c26222b0b843d19495e9c6c31492bf0d7df4e5d161b8d7

Thumb

Thumbnail

Source

Thumbnail Process Audio#0

Image (metadata/metadata/thumbnails/v1/00017.png)



Indexer

Mimetypes

- audio/mp3
- audio/mpeg

Pronom PUIDs

- [fmt/134](#)

Dimension 128 x 112px

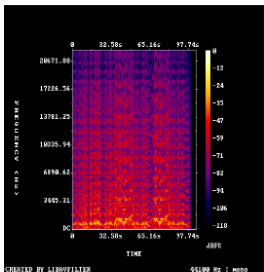
Full Metadata

```
▼ffprobe:  
  ▶format:{ Filename: "fd:", bit_rate: "160000", duration: "", ... }  
  ▶streams:[ { ... }, { ... } ]  
▼siegfried:  
  ▼0:  
    ▶Basis:[ "extension match...", "byte match at [...]" ]  
    Class:"Audio"  
    ID:"fmt/134"  
    MIME:"audio/mpeg"  
    Name:"MPEG 1/2 Audio Layer 3"  
    Namespace:"pronom"  
    Version:""  
    Warning:""  
▼tika:  
  ▼0:  
    Content-Type:"audio/mpeg"  
    ▶X-TIKA:Parsed-By:[ "org.apache.tika...", "org.apache.tika..." ]  
    ▶X-TIKA:Parsed-By-Full-Set:[ "org.apache.tika...", "org.apache.tika..." ]  
    channels:"1"  
    dc:creator:"Dragan Espenschied"  
    dc:title:"Оля Зимой"
```

v2 [data/audio/Dragan_Espenschied - 10 - Оля Зимой.mp3](#)

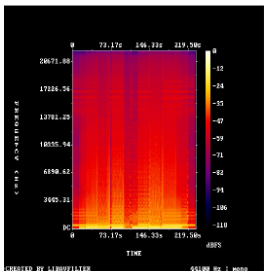


Pronom UID fnt/568
Content-Type image/webp
Dimension 4032x3024px
Filename data/image/IMG_6914.webp
Filesize 2.8 MB
Created 2023-11-28 18:03:21
Modified 2023-11-27 16:53:38
Attributes Archive
OS windows



Pronom UID fnt/134
Content-Type audio/mp3
Dimension 128x112px
Filename data/audio/Dragan_Espenschied_-_10_-_Оля Зимой.mp3
Filesize 2.1 MB
Created 2023-11-28 18:03:21
Modified 2023-11-27 17:49:20
Attributes Archive
OS windows

Filename metadata/indexer_v1.jsonl.gz



Pronom UID fnt/134
Content-Type audio/mp3
Dimension 192x186px
Filename data/audio/Dragan_Espenschied_-_02_-_P2P.mp3
Filesize 4.7 MB
Created 2023-11-28 18:03:21
Modified 2023-11-27 17:48:52
Attributes Archive
OS windows

UB:TEST01

Count	Size
3	23 MB
1	1.9 MB
1	171 B
1	37 MB
2	28 MB
1	2.2 MB
3	12 MB
1	18 kB
1	1.6 MB
1	42 MB
1	547 kB
3	3.2 MB
2	8.0 MB
1	19 MB
1	6.2 MB
1	60 kB
1	36 MB
2	46 kB
1	3.7 MB
4	1.5 MB
1	17 MB





Finanzierungskonzept:

- Sehr niedriger Kostensockel für Betrieb des Managementsystems
- PartnerInnen (lassen) Dienste und Erweiterungen nach eigenem Bedarf entwickeln
- Speicher im Bereich der PartnerInnen (Kosten, Betrieb (Cloud))

Weiterentwicklung:

- Neue Partner - im Verbund und für Dienstleistungen
- Feinabstimmung im Bereich (Pre-)Ingest und Monitoring
- Synergien schaffen bei Erarbeitung und Pflege von Tools und Know-How

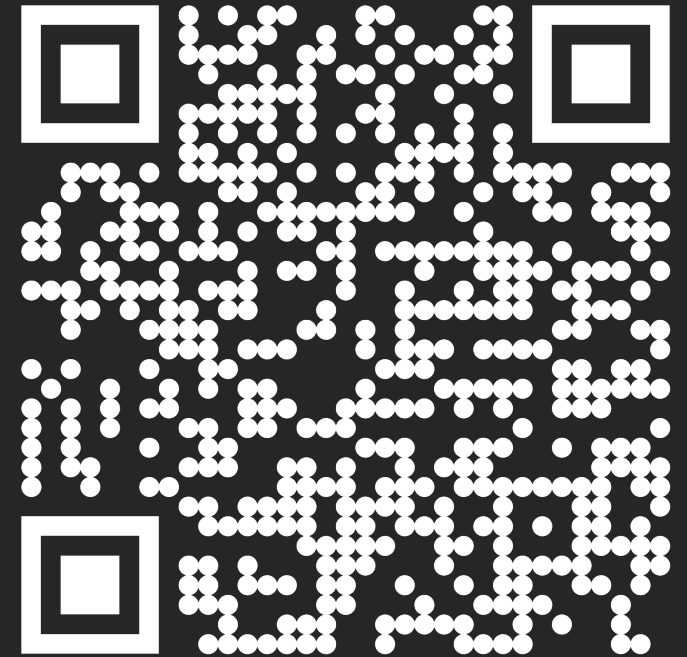
Gerne stehen wir für Ihre Fragen zur Verfügung



Jürgen Enge
Universitätsbibliothek Basel
Leiter Informatik
juergen.enge@unibas.ch



Beat Mattmann
Zentral- und Hochschulbibliothek Luzern
Leiter Library IT / Digitale Dienste
beat.mattmann@zhbluzern.ch



gocfl auf GitHub:
<https://github.com/je4/gocfl>