



Archives
Cantonales
Vaudoises

La préservation numérique sous le capot : État de l'art et Fedora Commons 6.x

Jan Krause

2023-08-31

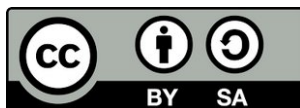


Plan

- **Exemples suisses (15 min)**
- **Un peu de théorie (15 min)**
 - Préservation numérique (OAIS)
 - Linked Data Plaform (W3C LDP)
 - Oxford Common File Layout (OCFL)
- **Démo sous le capot (30 min)**

Supports et matériel en ligne :

<https://tinyurl.com/LibreABC>



Exemples suisses (FCrepo)



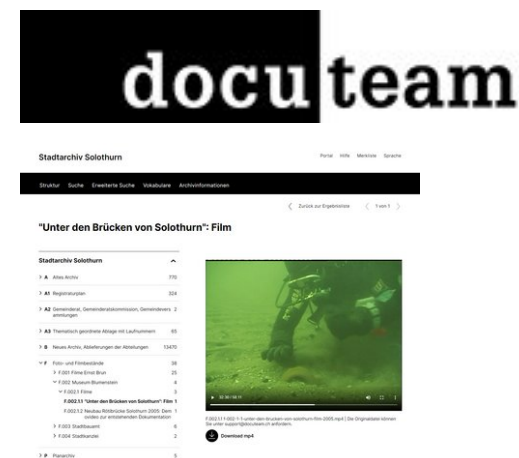
archive-ouverte.unige.ch
lod.unige.ch

Domaine : bibliothèque (IR)
Année : 2022
Interface : maison
FCRepo 5.x
Refonte



memobase.ch

Domaine : audio-visuel
Année : 2021
Interface : maison, libre et publiée
FCRepo 5.x
Documentation , API , GitLab



docuteam.ch/docuteam-cosmos/

Domaine : archives
Année : 2023-10
Interface : libre, SaaS
FCRepo 6.x (octobre)
Documentation ,
Logiciels

Exemples suisses (FCrepo)

Outils les plus répandus (classement indicatif) :



Or, docuteam cosmos et scopelngest utilisent FedoraCommons, ce qui en fait largement le composant de préservation le plus répandu dans le pays.

Préservation numérique



Livre d'heures de Jean de Gingins.

Manuscrit de 1421.

Archives cantonales vaudoises.

<http://dx.doi.org/10.5076/e-codices-acv-P-Sarraz-H-0050>

Que restera-il du
numérique dans
600 ou 1000 ans ?

Préservation: OAIS et PREMIS

Open Archival Information System : ISO 14721

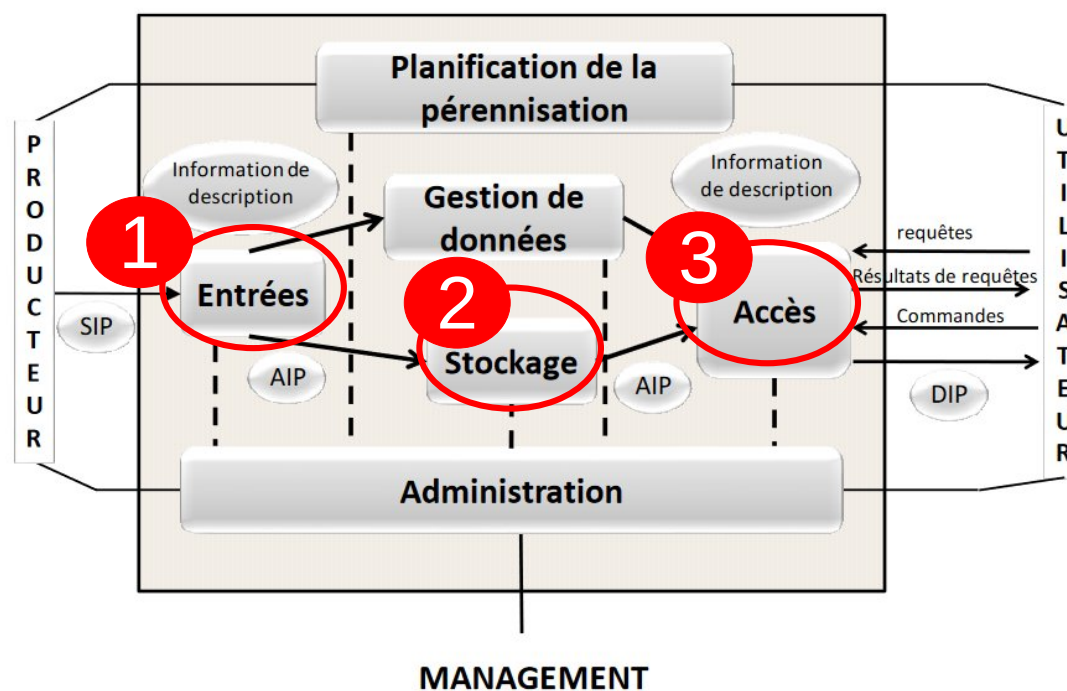


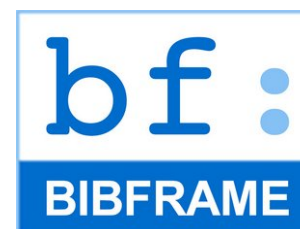
Schéma 4-1 : Entités fonctionnelles OAIS

Version française :

<https://public.ccsds.org/Pubs/650x0m2%28F%29.pdf>

Linked Data

- Métadonnées de préservation
 - PREMIS ontology
- Métadonnées descriptives
 - Bibliothèques : bibframe
 - Archives : Records in Contexts
 - Musées : CIDOC-CRM
- Thesaurii
 - SKOS

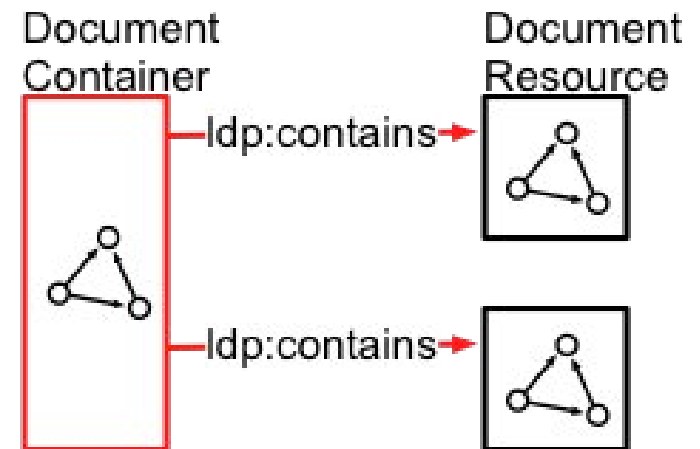


Linked Data Platform (W3C LDP)

- Organiser une plateforme sémantique et binaire
- LDP : <https://www.w3.org/TR/ldp/> (2015)
 - Ressources organisables en containers
 - Deux types de ressources : RDF ou binaires

- API-ReST HTTP

- GET : accéder
- POST : créer
- PUT : mettre à jour
- DELETE : supprimer



Oxford Common File Layout

OCFL (2020) définit :

- la hiérarchie de stockage (organisation des paquets AIP sur le media)
- le format des paquets AIP (objets OCFL)

« 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. »

Source : <https://ocfl.io/>

Oxford Common File Layout

- **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

Les dix dernières années

- Ontologies pour les métadonnées métier
 - descriptives → ontologie RiC-O (2023-10)
 - de préservation → ontologie PREMIS
- ~2015 : Fedora 4.x passe à LDP
- ~2021 : Fedora 6.x passe à OCFL

Et maintenant ?

En parallèle en 2023....

docuteam

intégration en production de ces concepts dans docuteam cosmos

<https://docs.docuteam.ch/introduction/en/matterhornRDF>

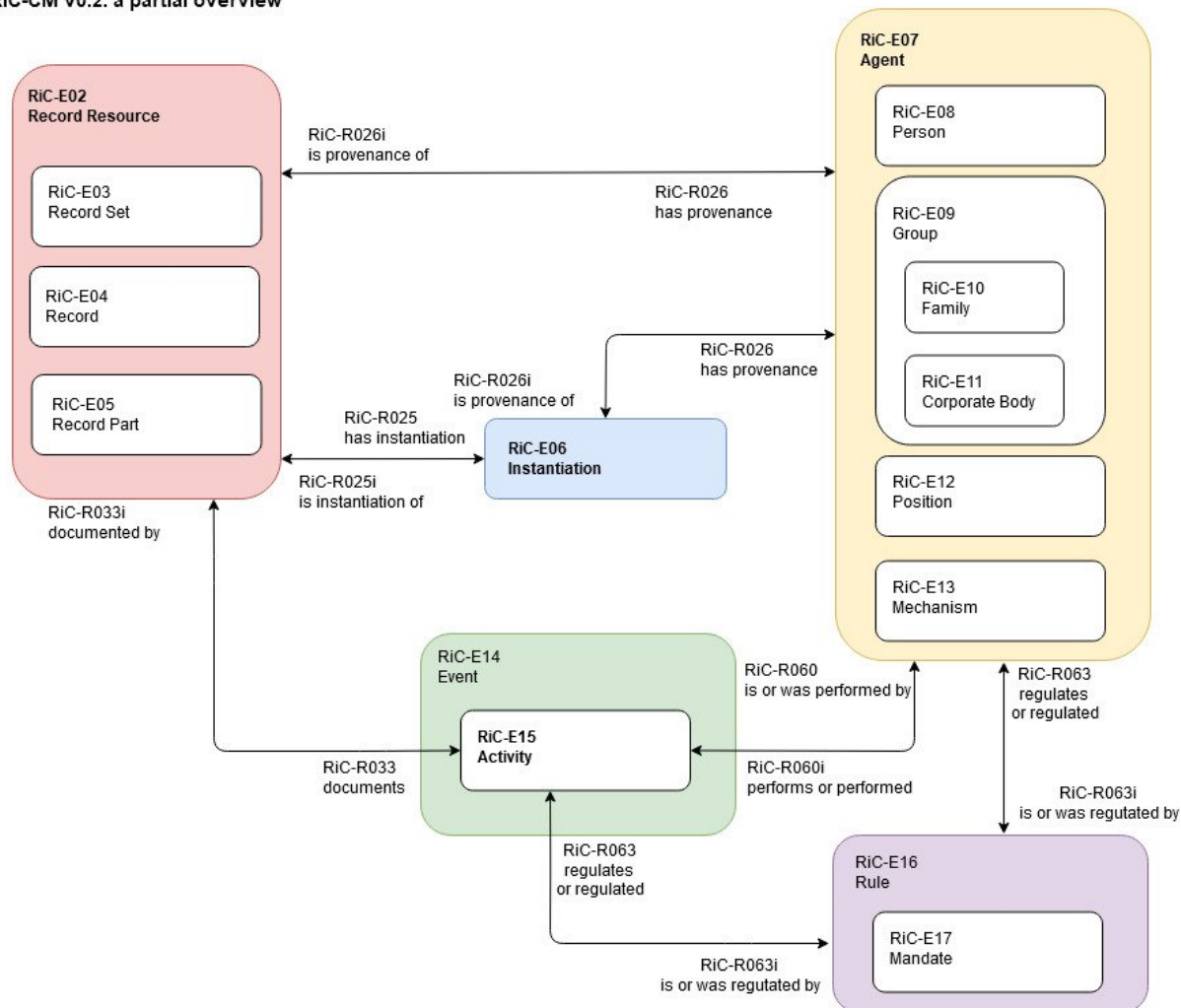


Archives
Cantoniales
Vaudoises

intégration théorique de ces concepts dans le cadre du projet d'archivage numérique vaudois

Demo avec RiC-O et PREMIS

RiC-CM v0.2: a partial overview

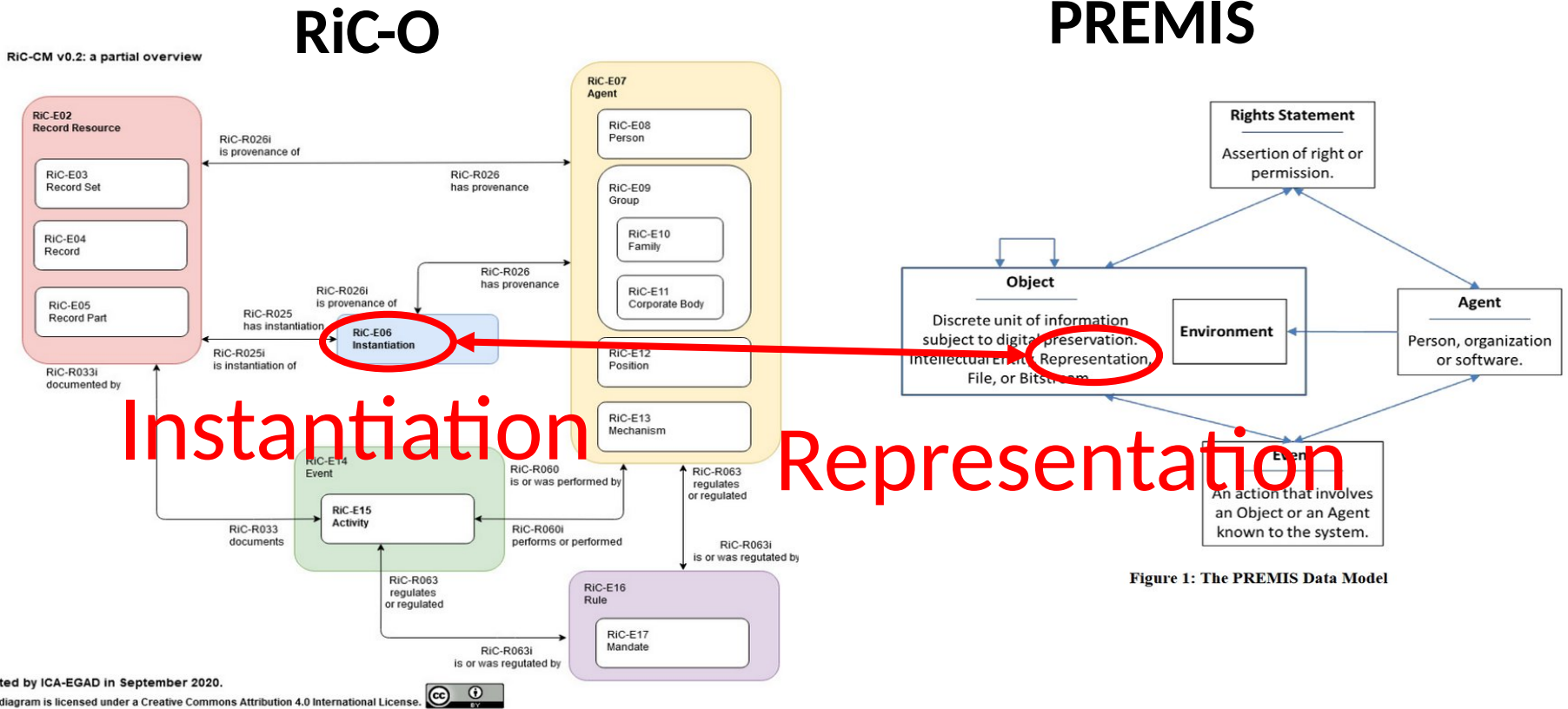


Created by ICA-EGAD in September 2020.

This diagram is licensed under a Creative Commons Attribution 4.0 International License.



Demo avec RiC-O et PREMIS



Démo sous le capot

Matériel en ligne :

<https://tinyurl.com/LibreABC>

Démarrer FCRepo :

Dans le terminal :

```
java -jar fcrepo-webapp-6.3.0-jetty-console.jar
```

Dans le navigateur :

<http://localhost:8080/rest/>

Accéder à une ressource

- GUI, exemple d'un dossier d'affaire
<http://127.0.0.1:8080/rest/record/9a3f45>
- Python :

```
import requests
url = 'http://localhost:8080/rest/record/9a3f45'
r = requests.get(url)
print('Status code:', r.status_code)
print(r.text)
```


Créer une ressource

- GUI, par formulaire.
- Python :

```
import requests

url = 'http://localhost:8080/rest/record/monDossier'
headers = {"Content-Type": "text/turtle"}
auth = ('fedoraAdmin', 'fedoraAdmin')
data = """ <> <rico:title> 'Ceci est le titre'.
          <> <rico:scopeAndContent> 'Voilà la description.'"""

r = requests.put(url, auth=auth, data=data.encode('utf-8'),
headers=headers)

print( 'Status:', r.status_code )

print( r.text )
```

Mettre à jour une ressource

- GUI, par formulaire.
- Python :

```
import requests

url = 'http://localhost:8080/rest/record/monDossier'
headers = {"Content-Type": "text/turtle"}
auth = ('fedoraAdmin', 'fedoraAdmin')
data = """ <> <rico:title> 'Ceci est le titre mis-à-jour.'.
          <> <rico:scopeAndContent> 'Et la description revue'.
          """

r = requests.put(url, auth=auth, data=data.encode('utf-8'),
headers=headers)

print( 'Status:', r.status_code )

print( r.text )
```

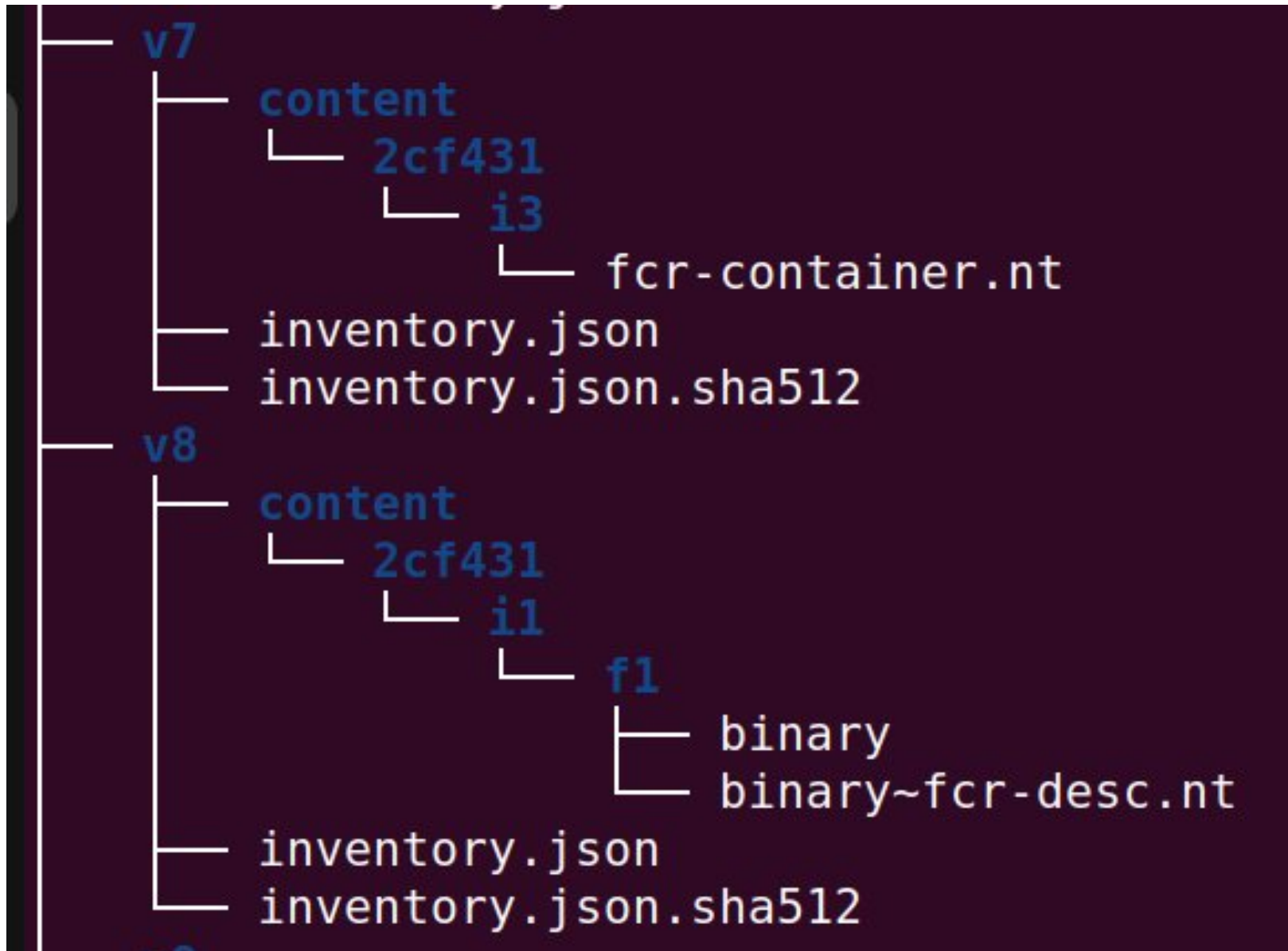
Localiser une ressource dans le stockage

$Z := \text{sha256}(\text{identifiant-de-l-objet})$

Chemin := $Z[1 \text{ à } 3] / Z[4 \text{ à } 6] / Z[9 \text{ à } 12] / Z /$

```
jan@yogan:~/Bureau/ABC/RefGD_fedora$ python3 id2storage.py "record/9a3f45"  
Fedora ID: b'info:fedora/record/9a3f45'  
Hexdigest: b6fb03cbda507dcfbfbb3f4accf2100b87afa0a8f634694a8a2ac783b693dc51  
Storage path: b6f/b03/cbd/b6fb03cbda507dcfbfbb3f4accf2100b87afa0a8f634694a8a2ac783b693dc51
```

Structure OCFL





Indexation

Fedora gère automatiquement les indexes sous formes :

- Base de données relationnelle
- Triple store
- Moteur de recherche

Pour chaque catégorie plusieurs logiciels sont supportés, p.ex. BlazeGraph et Jena .

Résumé : état de l'art

Pour une préservation durable, nous avons enfin tous les standards adaptés :

- Manipulation des données : LDP (W3C)
- Empaquetage des données et abstraction du stockage : OCFL
- Ontologie de préservation : PREMIS
- Ontologie de description : RiC-O, bibframe ...

Tout ceci peut être mis en œuvre sur la base de Fedora Commons !