



UNIVERSITÉ DE GENÈVE

La préservation numérique à UNIGE

Des données de recherche aux données administratives et patrimoniales :
un long fleuve tranquille

Hugues CAZEAUX

hugues.cazeaux@unige.ch

Resp. du pôle e-Research

Service informatique

UNIGE

Agenda

- Introduction
- Le contexte
- La technologie DLCM
- La mise en œuvre
- Aujourd'hui
- Et après...

La préservation numérique à UNIGE

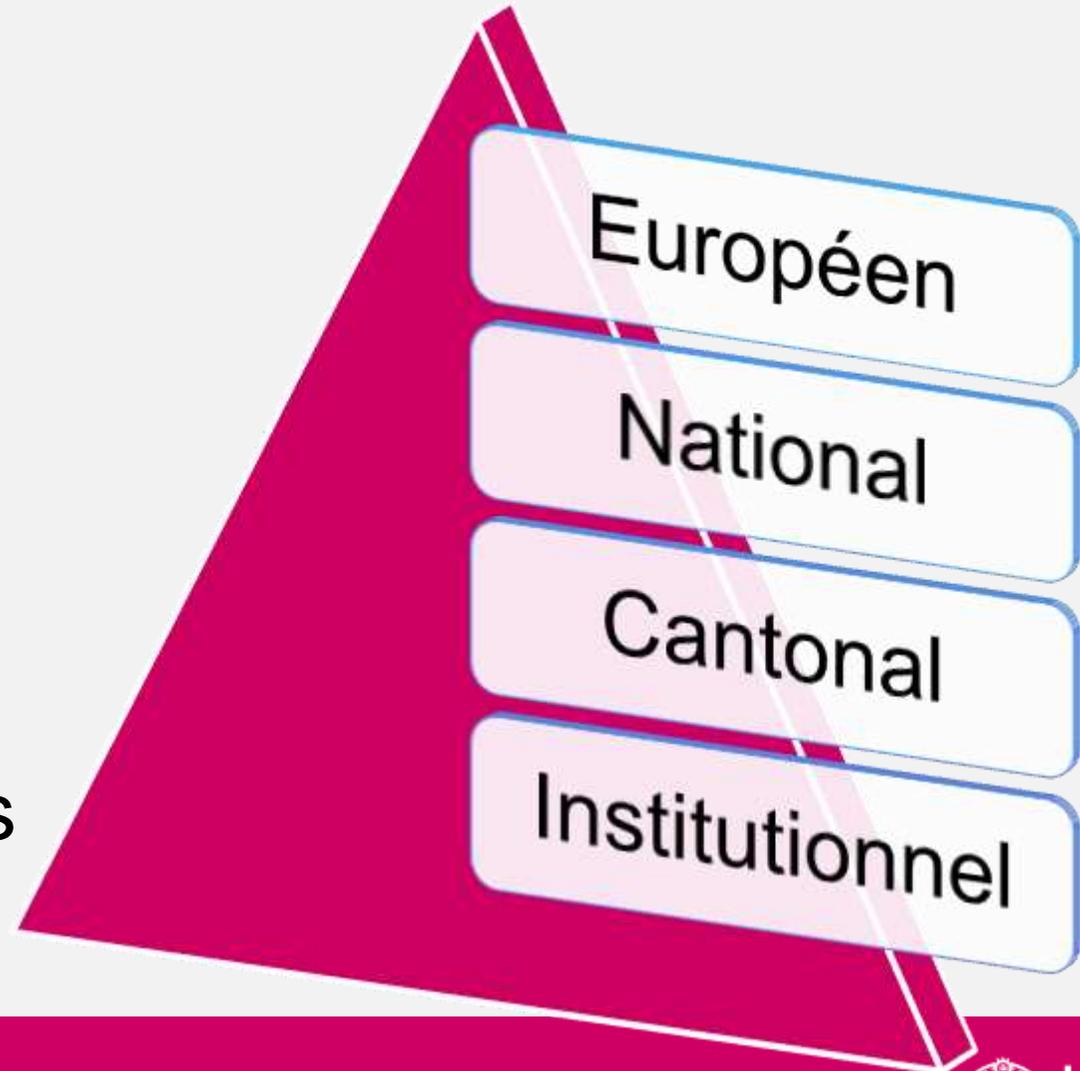
Introduction

Introduction

- Focus sur les données de recherche
- Choix technologique
- Challenge organisationnel
- Extension sur les données administratives et patrimoniales
- Focus sur les aspects opérationnels et techniques

Introduction

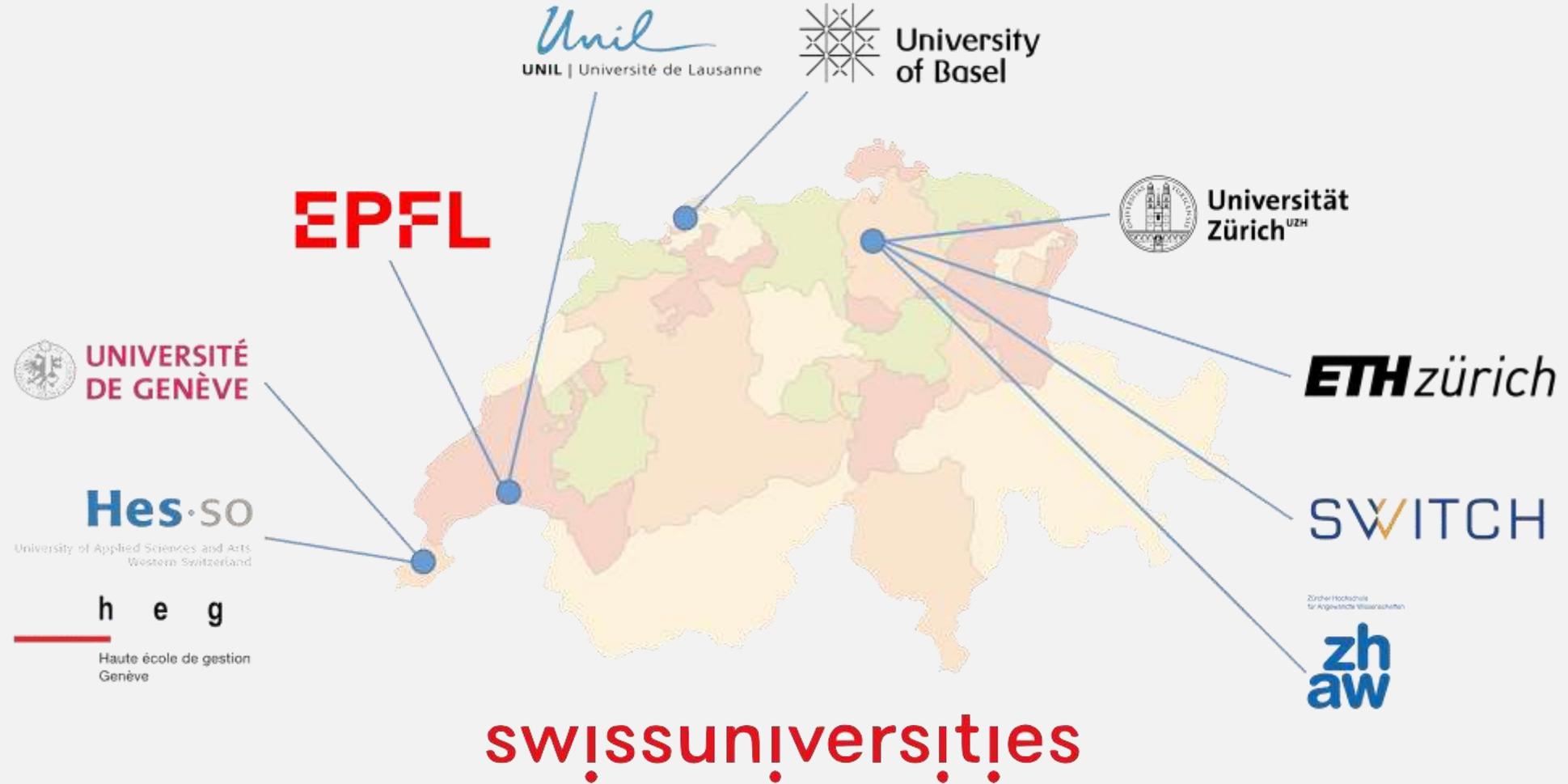
- Différentes initiatives
- Plusieurs projets
- A des échelons différents
- Avec des réglementations propres



La préservation numérique à UNIGE

Le contexte

Projet DLCM - Les partenaires



Projet DLCM – Les phases



Phase #1

2015-2018

- Solution Prototype
- Services
- Institutional policies
- Data Management Plan (DMP)

Phase #2

2018-2020

- DLCM Solution
- National Coordination Desk
- OLOS
- New Partnerships (Switch, CERN, EOSC,...)

Phase #3

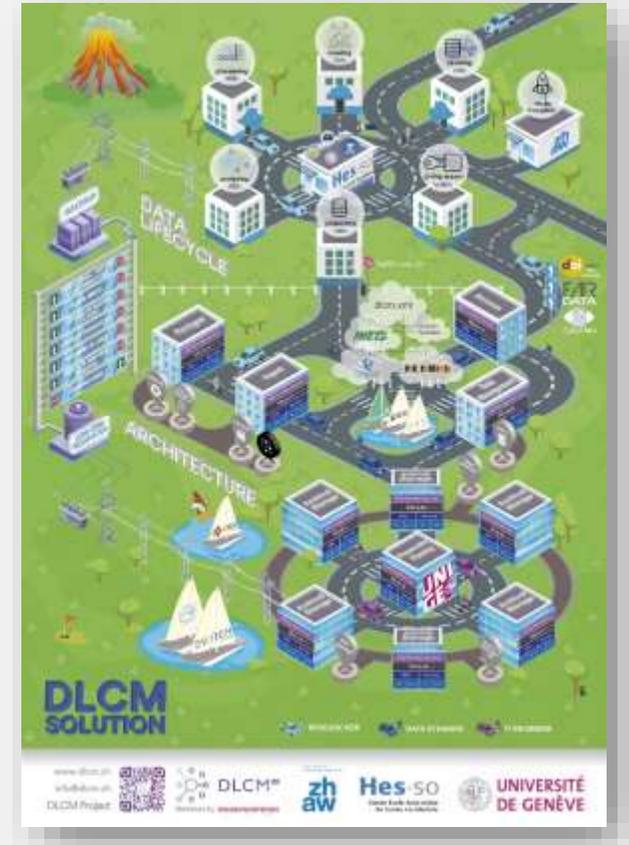
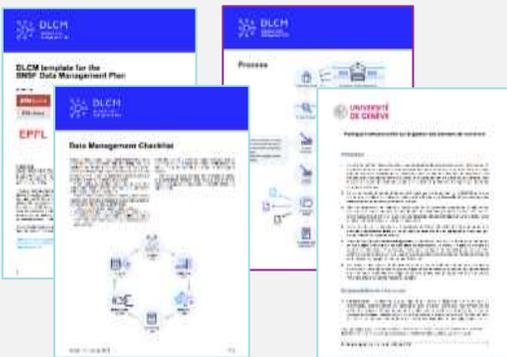
2021

- OLOS Association
- OLOS.swiss launch
- Webinars & Trainings

Projet DLCM – Les livrables



Directives & Modèle de DMP



Présenté à iPRES 2019

Projet de Loi (PL)

■ Titre

« Evolution des infrastructures et services associés de calcul à haute performance et de stockage long terme afin d'optimiser et faciliter l'utilisation des données issues de la recherche dans les hautes écoles universitaires genevoises »

■ Application

- Accepté en 2017
- Durée de 7 ans

■ Définition

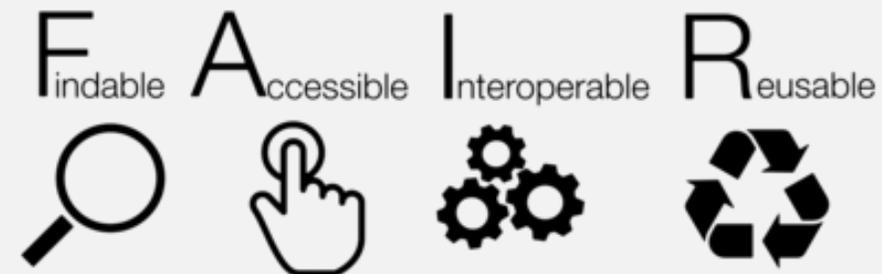
- Un projet de loi est une procédure cantonale qui permet de soumettre une demande de financement à l'Etat de Genève.

Nouvelles règles pour la recherche

- Nouvelles exigences pour obtenir des fonds pour la recherche du FNS (Fonds National Suisse), comme au niveau européen :

- DMP (Data Management Plan) ou PGD (Plan de Gestion de Données obligatoire

- Utilisation d'un dépôt de données FAIR

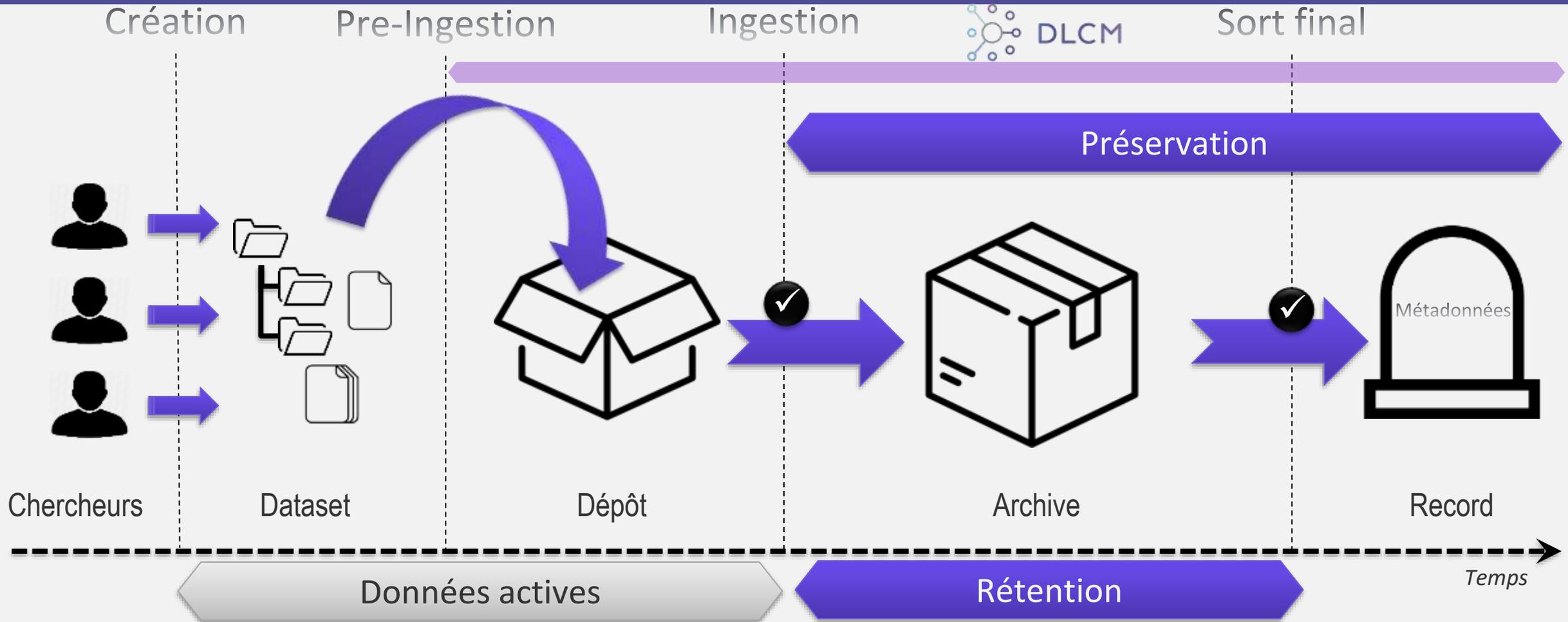


La préservation numérique à UNIGE

La technologie DLCM

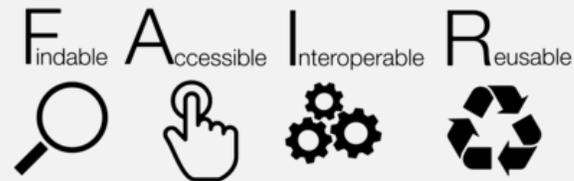
Aspects fonctionnels

L'archivage

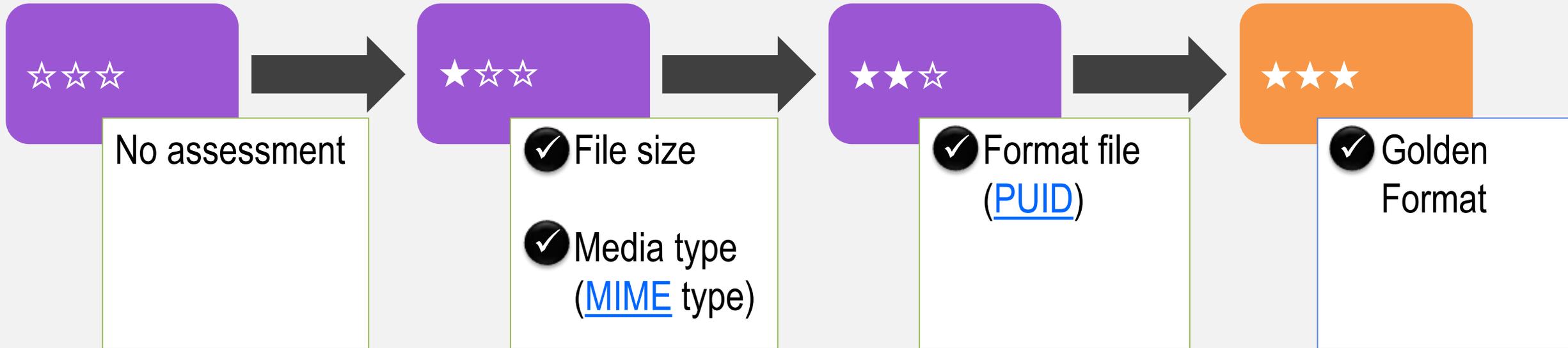


Basé sur le modèle [OAIS](#) (Open Archival Information System, ISO 14721)

- Métadonnées autodescriptives basées sur le schéma DataCite
- Garantir le côté FAIR (“FAIRness”)
 - Support des identifiant DOI avec la fonctionnalité “Réserver DOI”
 - Support des licences ouvertes, incluant CC BY, pour protéger les copyrights



Niveau de compliance



PUID = PRONOM Unique Identifiers
(The National Archives, UK)

<https://www.nationalarchives.gov.uk/PRONOM>

Golden Format = recommended format for
long term preservation
(Library of Congress, USA)

<https://www.loc.gov/preservation/resources/rfs/>

Accès & Usage

Access Level



Qui a **accès** à votre archive ?

License



Data Use Agreement



“Aussi ouvert que possible, autant fermé que nécessaire”

Dataset

Quelles obligations souhaitez-vous **imposer** pour l'usage de votre archive ?

Y a-t-il des **données sensibles** dans votre archive ?

Data Tag



Quels sont les **termes** et les **conditions** pour accéder et utiliser votre archive ?

Data Use Policy





Restriction



Access Level



Public



Restricted



Closed

Data Tag



Not defined
Grey



Public
Blue



Controlled public
Green



Accountable
Yellow



More accountable
Orange



Fully accountable
Red



Maximum restricted
Crimson

Data Use Policy



None



License



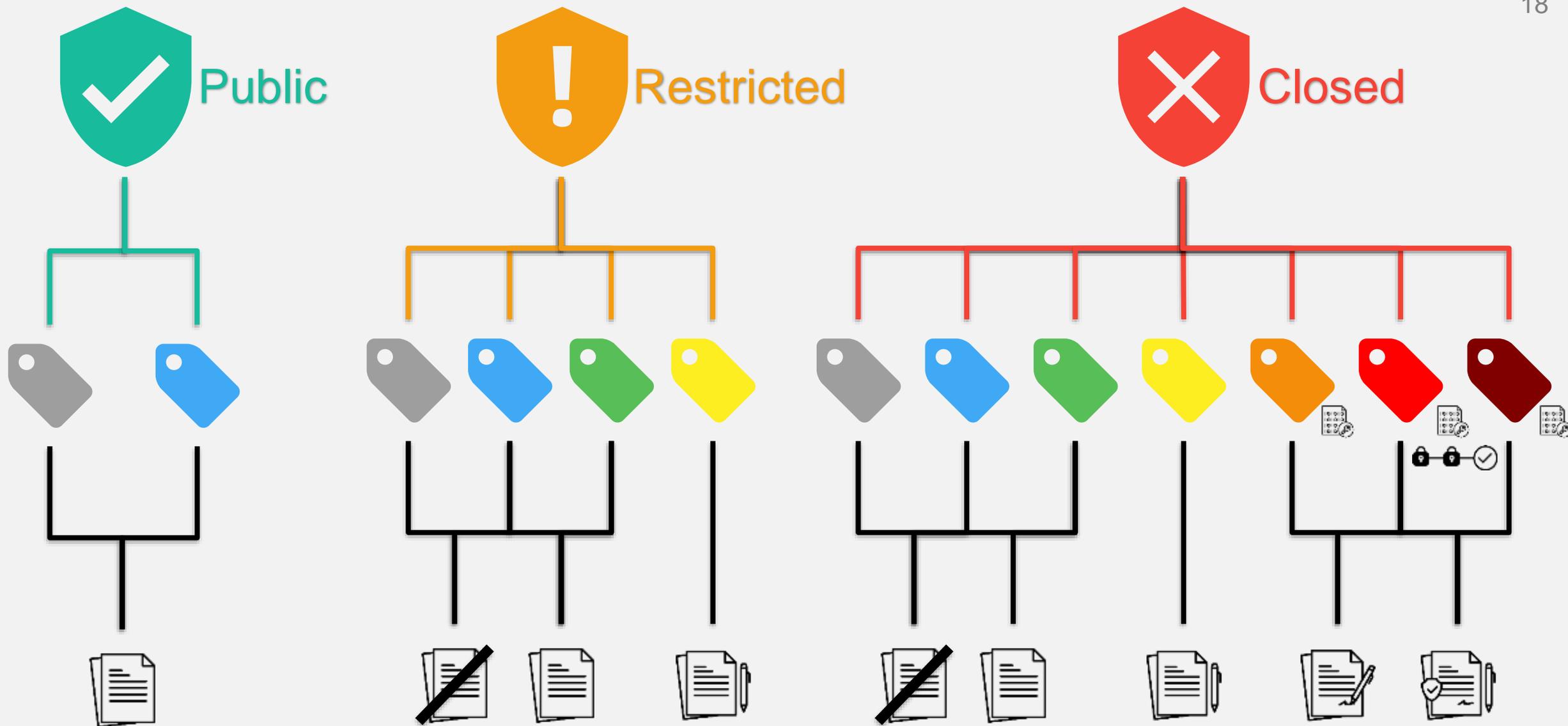
Click-through
DUA



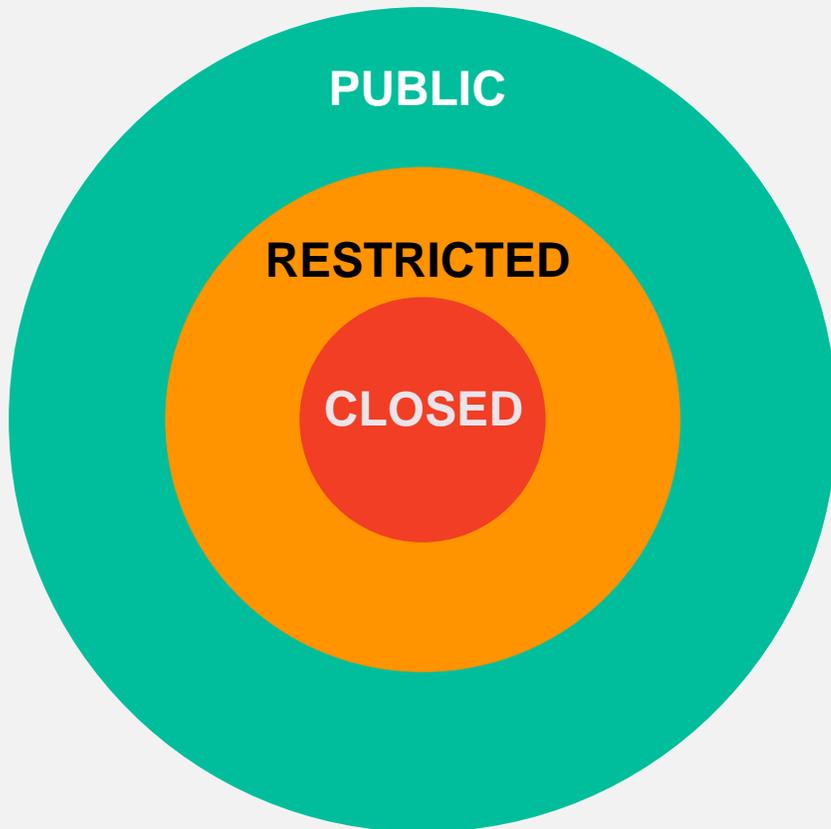
Signed
DUA



External
DUA



Niveaux d'accès & Embargo



- Choisissez le bon niveau d'accès:
 1. **Public**
 - > Open Access
 - >> Tout le monde
 2. **Restrient**
 - > Membres de l'équipe (i.e., Unité Org.)
 - >> Tiers de confiance
 3. **Clos**
 - > Cas par cas
 - >> Individus
- Définissez un embargo pour décaler l'accès public à votre archive

Espace de préservation

Org. Unit. Roles

Institutional Roles 

Manager 

Within an organizational unit
Manage users of this organizational unit

Steward 

Manage deposits
Manage archives

Approver 

Approve or reject a submitted deposit

Creator 

Initiate, edit & submit deposit
Add data files in deposit

Visitor 

Download non "Open Access"
archives

Public

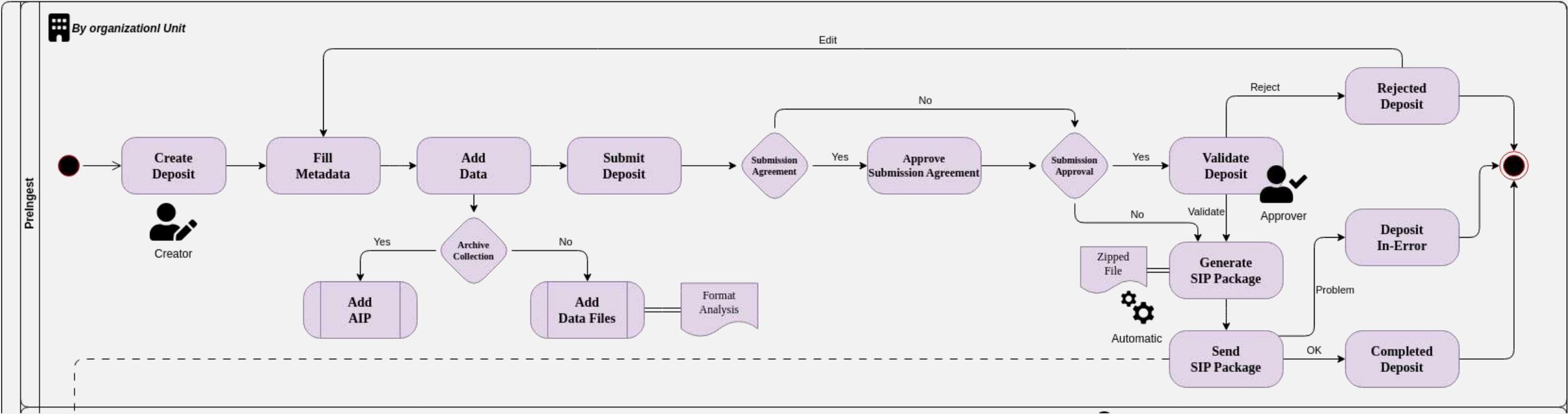
Search & view archive metadata
Download "Open Access"
archives

Unité Organisationnelle

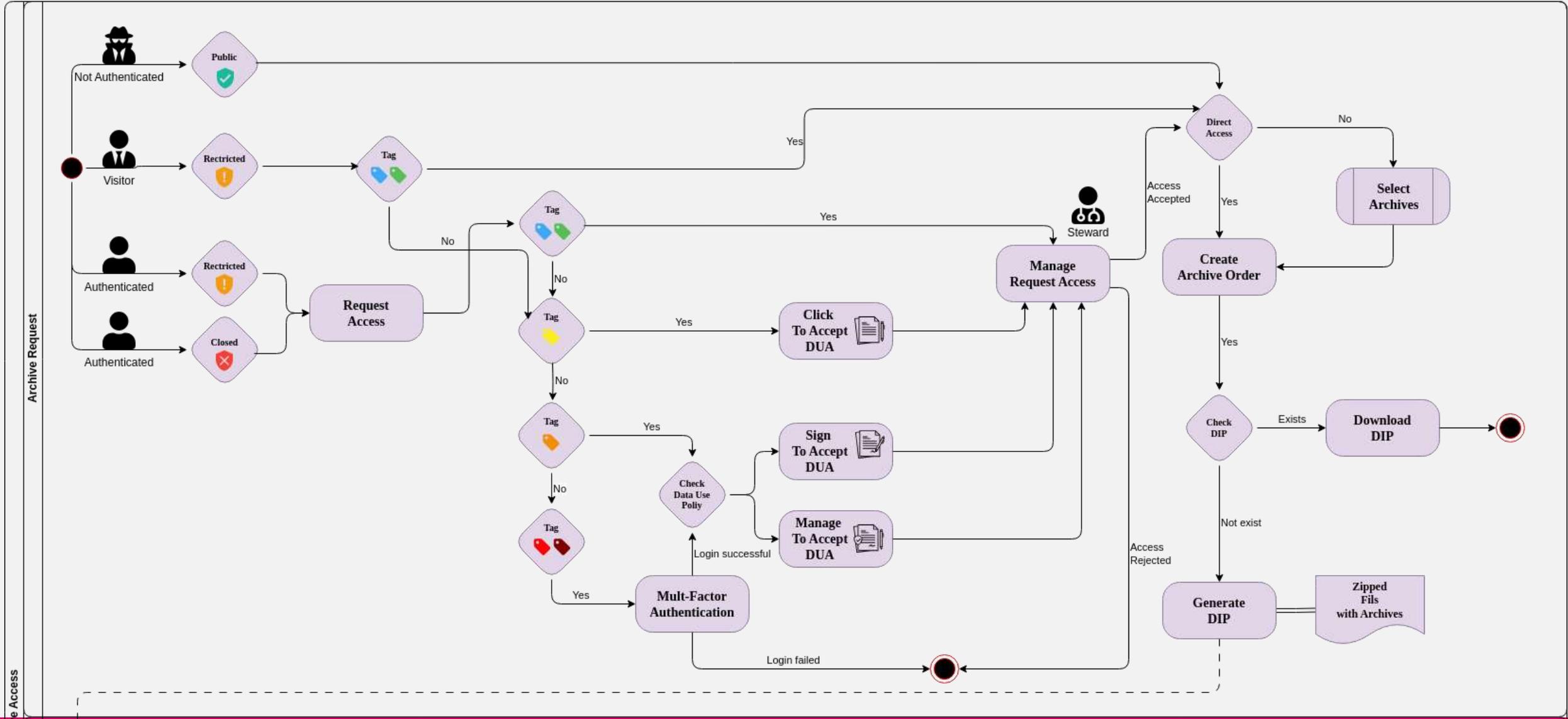
- ~ Espace de préservation
- Contient les archives d'un groupe de recherche, d'un département, d'un laboratoire, etc.

- Rôles hérités d'un niveau institutionnel
 - Un responsable d'une institution peut créer de nouvelles unités org.

Le flux de création d'archive



Le flux d'accès à l'archive



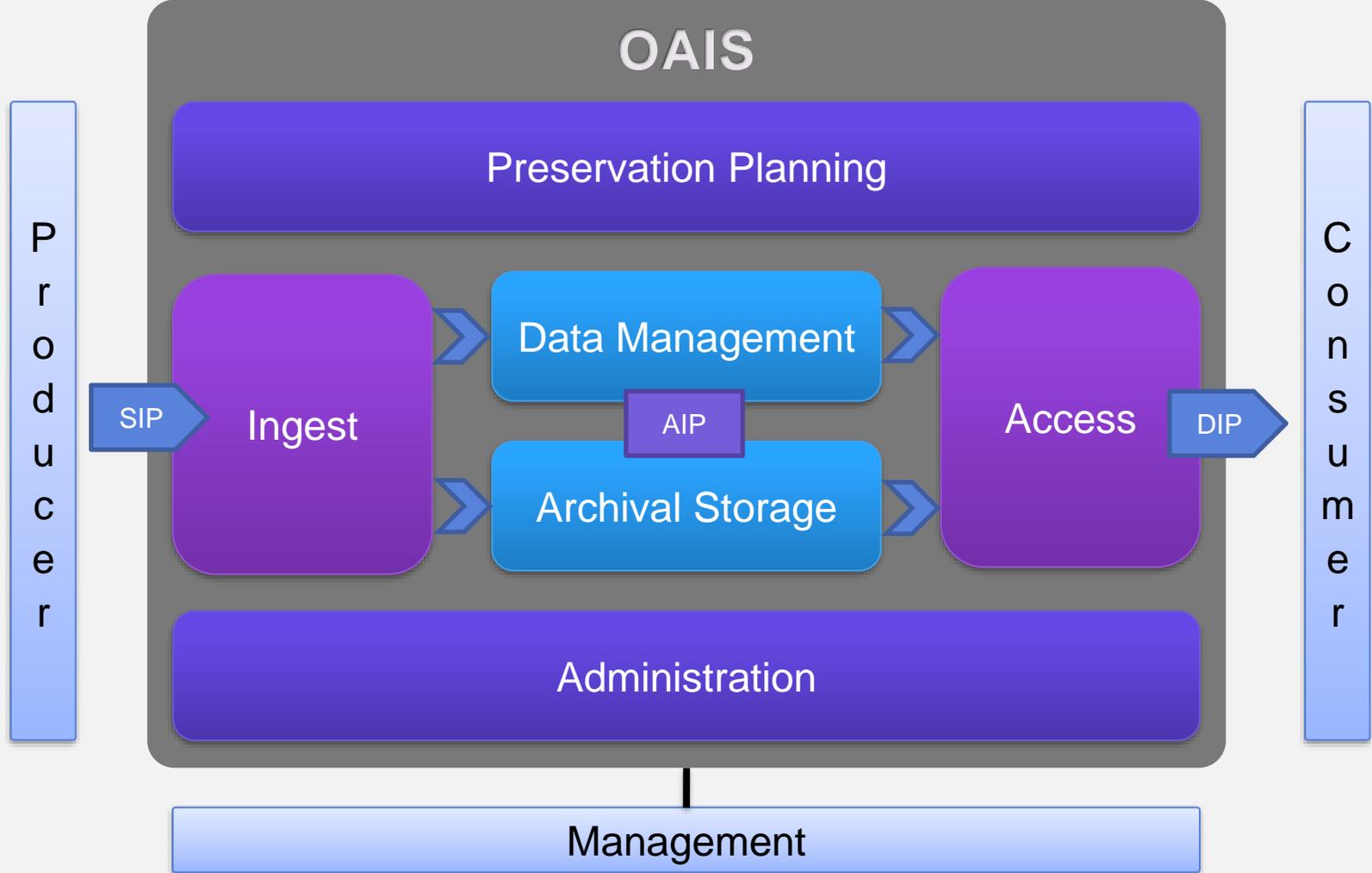
La préservation numérique à UNIGE

La technologie DLCM

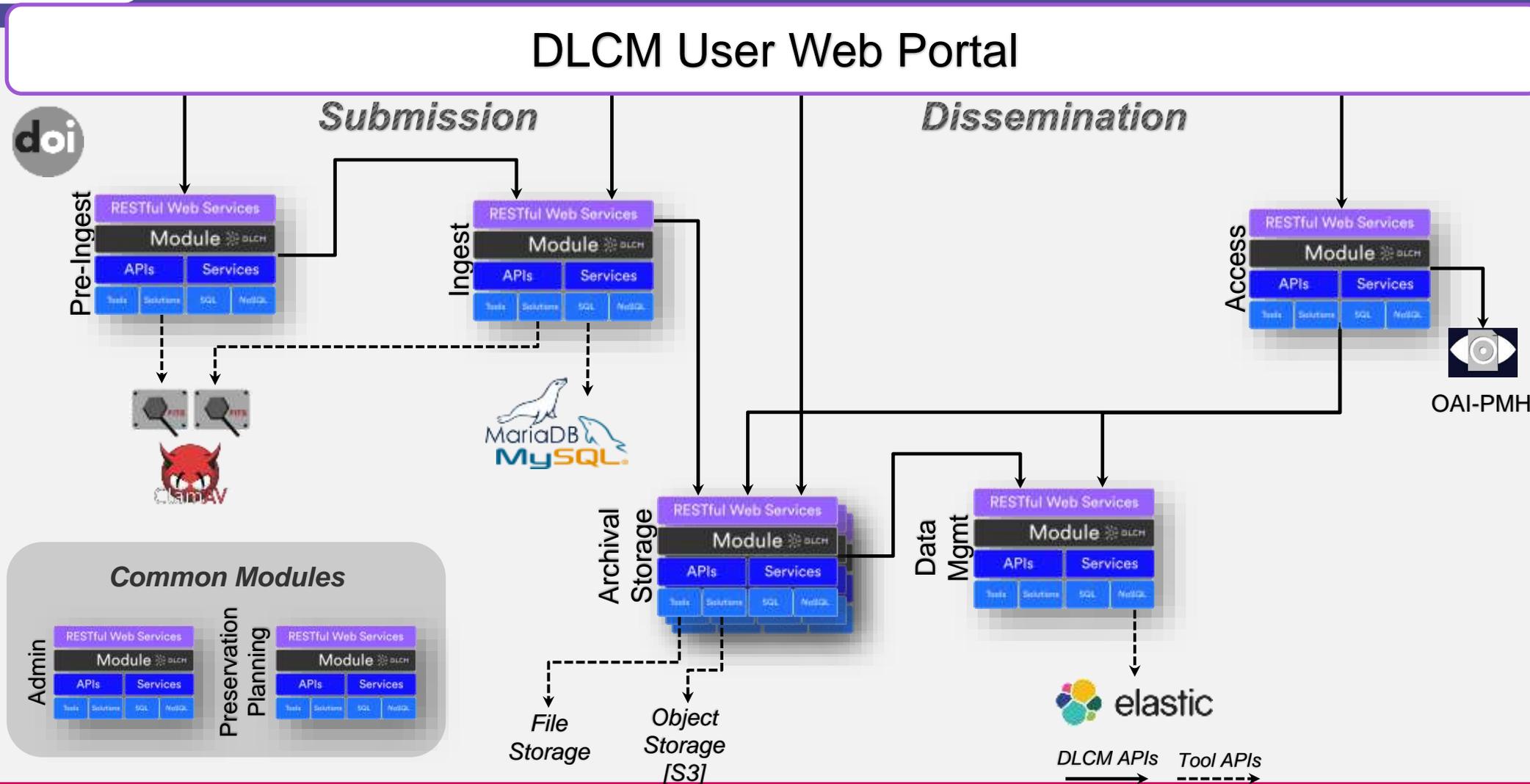
Aspects techniques

Modèle OAIS

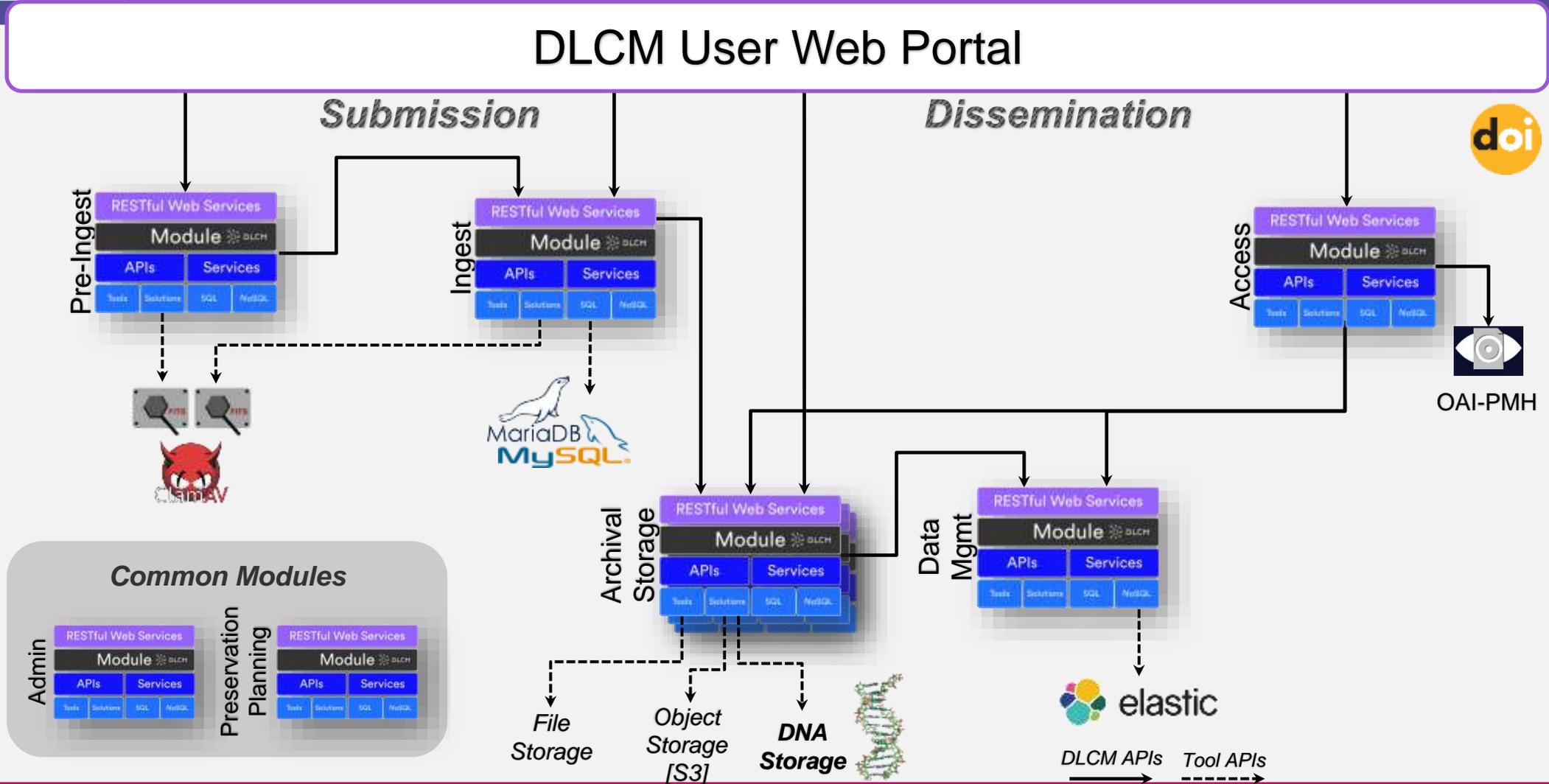
Open Archival
Information System
ISO 14721

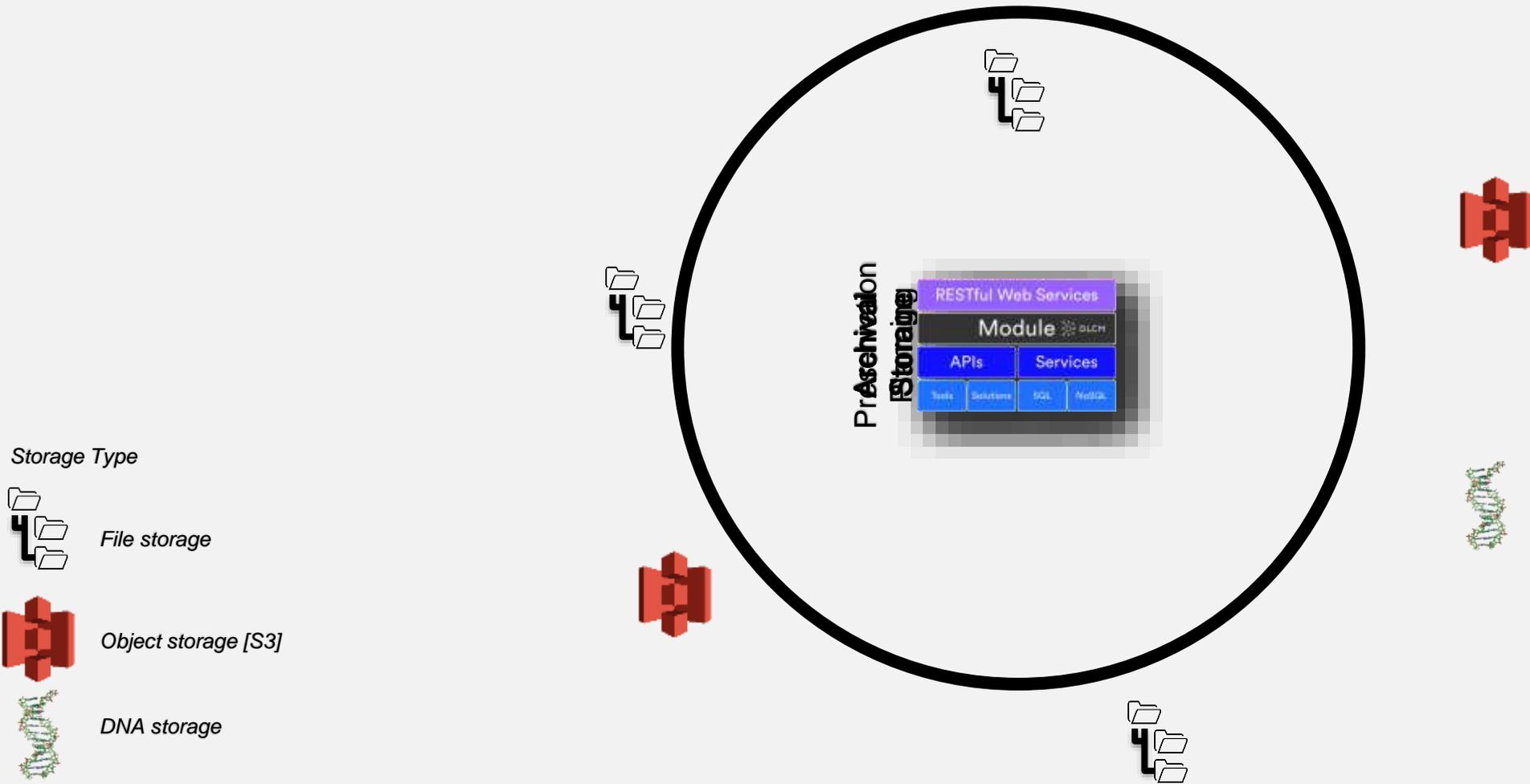


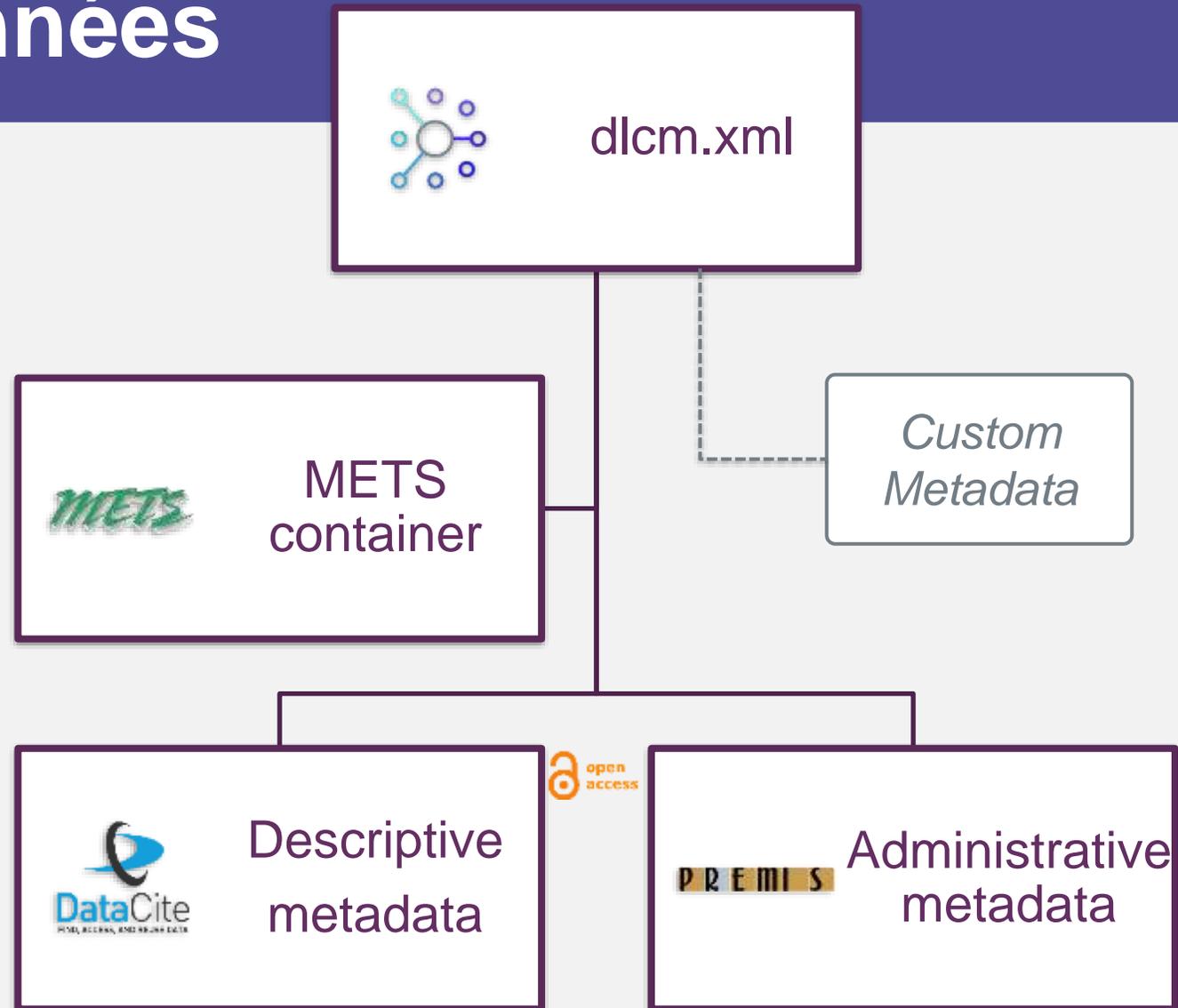
Architecture



Architecture

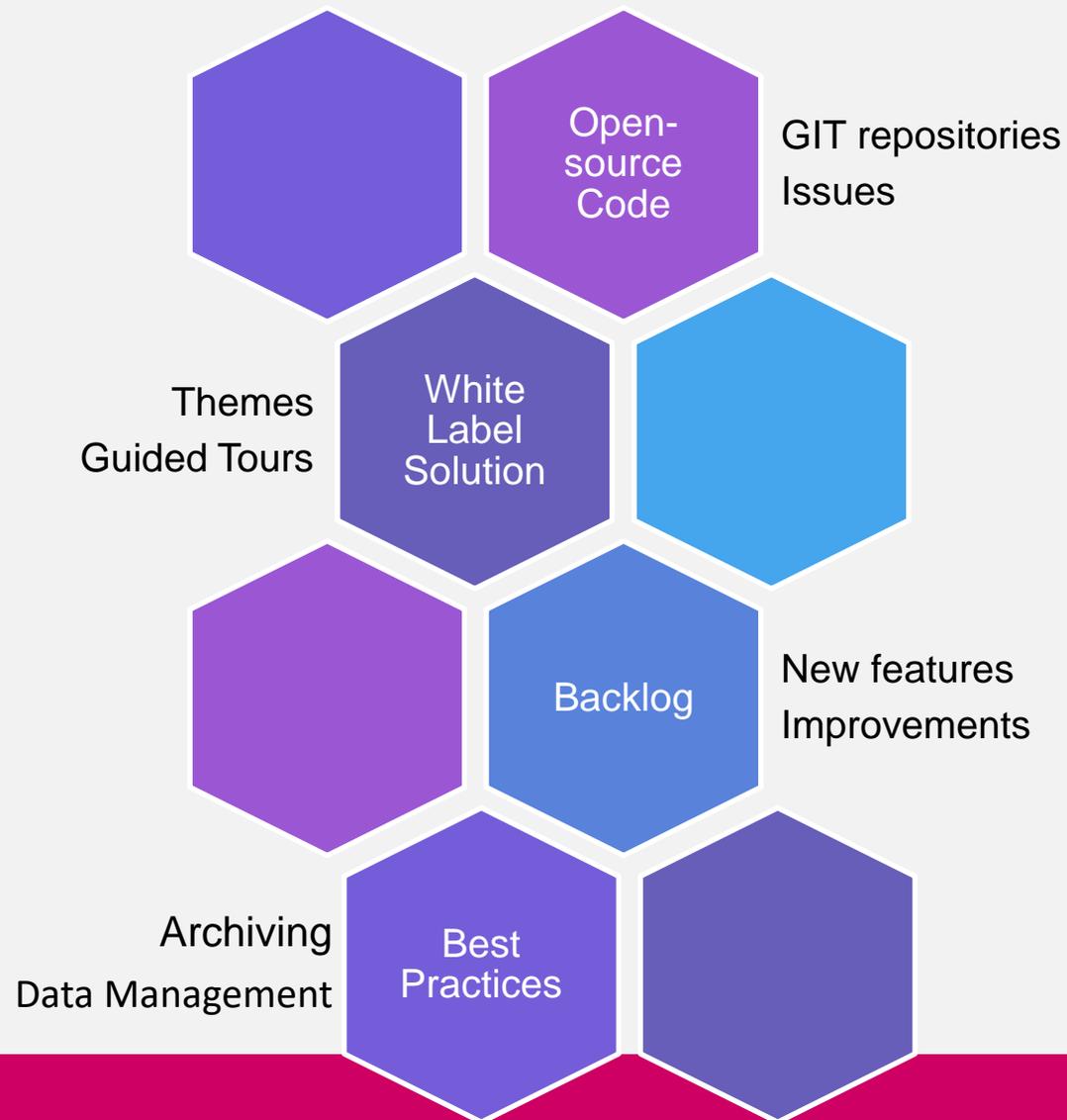








Solution open-source qui use des standards ouverts



La préservation numérique à UNIGE

La mise en œuvre

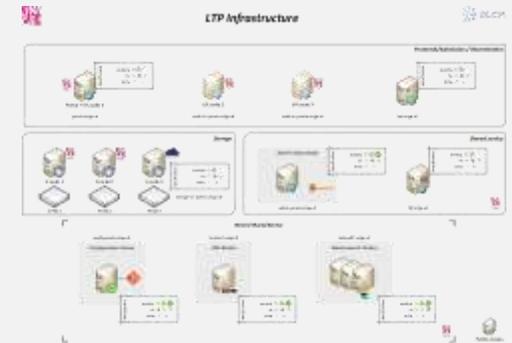
Politique institutionnelle

- Phase de définition de la politique
- Phase de validation par les différentes divisions et le rectorat
- Cela a permis d'officialiser le besoin d'un point vue institutionnel

<https://www.unige.ch/researchdata/fr/propos/politique/>

Service numérique

- Choix de la solution technique → DLCM
- Définition de l'architecture technique
 - Déploiement sur des machines virtuelles
 - Niveau d'exigence technique → disponibilité & backup
- Tarification
- Choix du nom du service



Soutien à la recherche

- Formation organisée régulièrement par les bibliothèques
- Formulaire pour accéder au service
→ création d'une unité organisationnelle
- Organisation du support (1^{er}, 2^e et 3^e niveau)
- Conseil et expertise métier et technique
- Soutien pour les chargements de gros volumes

La préservation numérique à UNIGE

Aujourd'hui

Faits Marquants

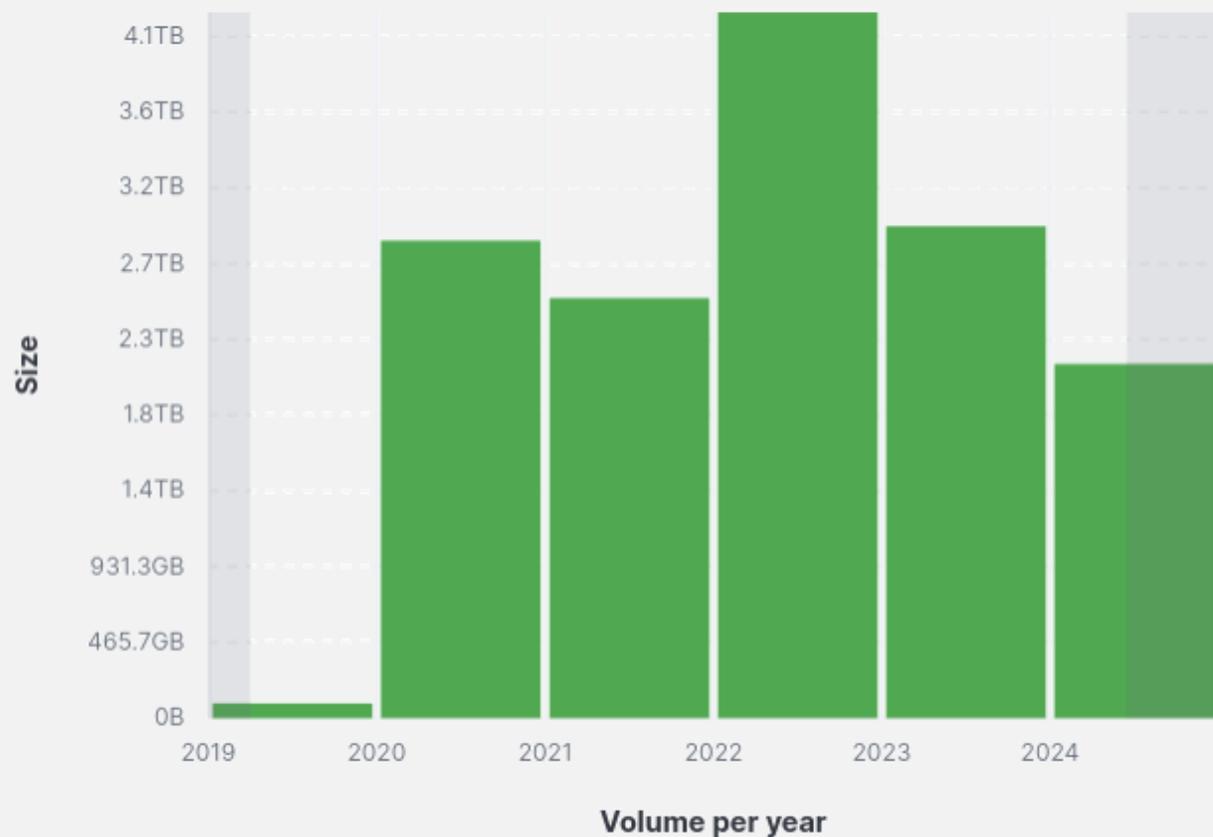
- La **Certification CTS** (<https://www.coretrustseal.org/>) a été obtenue en mars 2024
→ Dépôt fiable de données
- La **1000^e** archive a été créée le 29 février 2024
- **15 To** d'archives
- L'archive la plus volumineuse est de **950 Go**



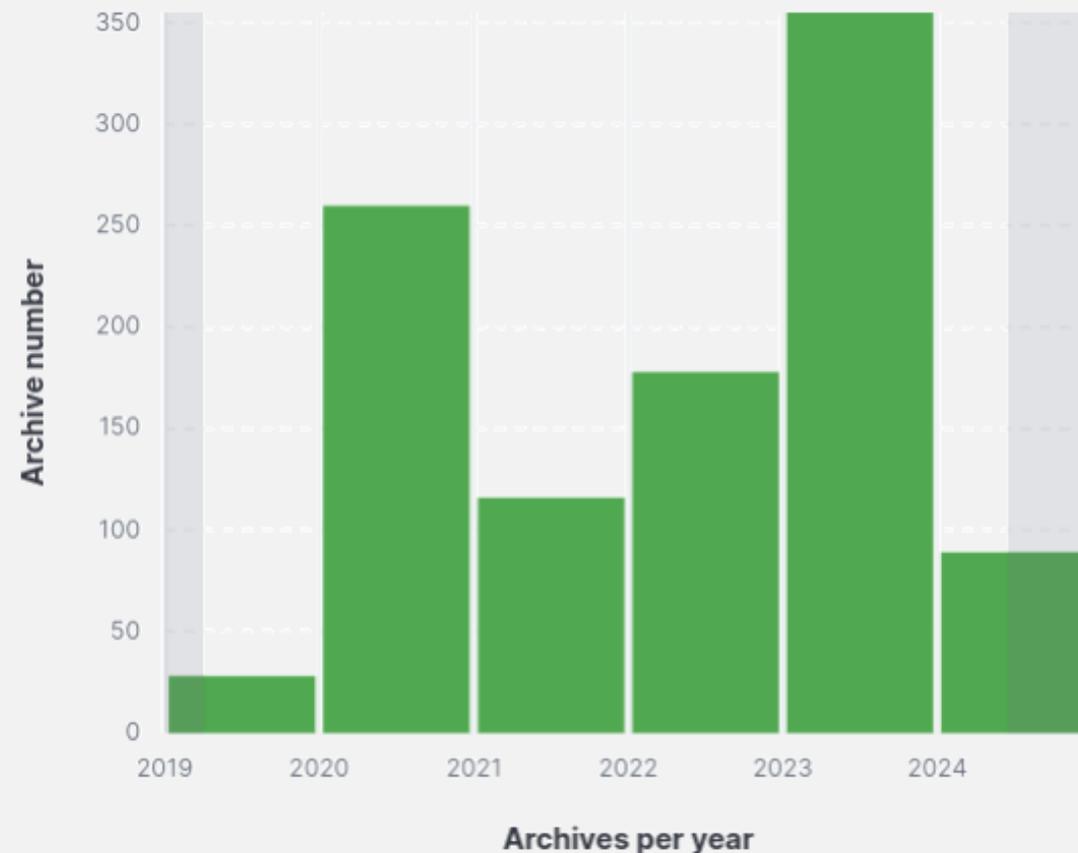


Quelques chiffres

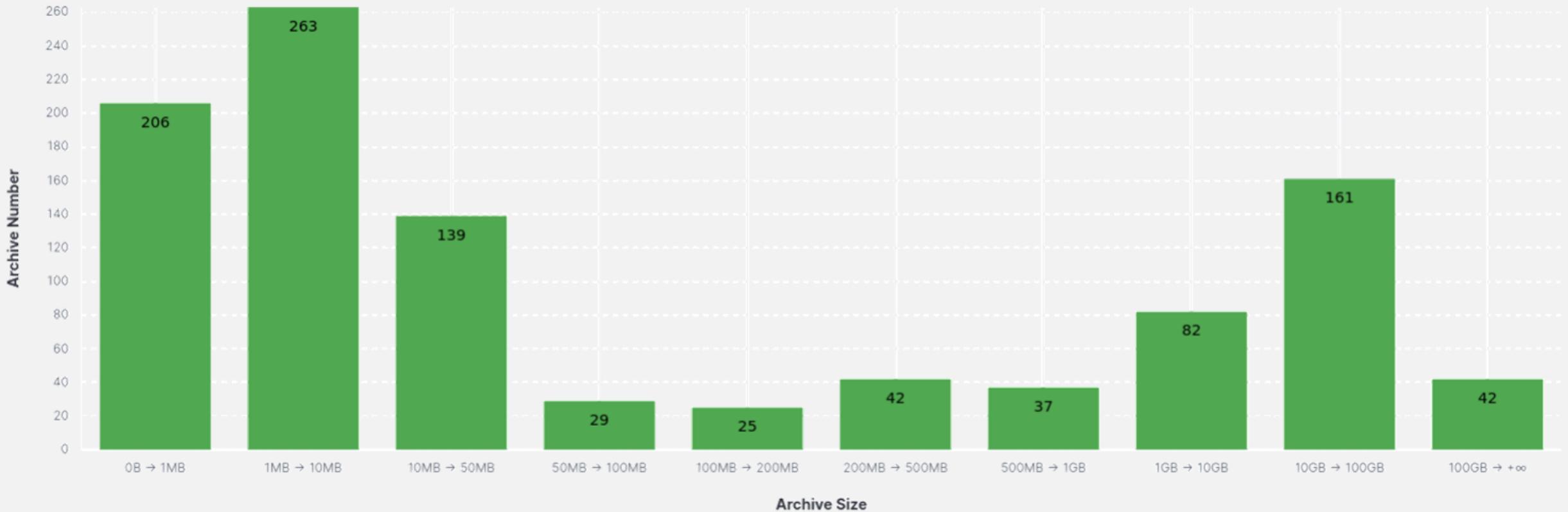
Archive size trend



Archive number trend

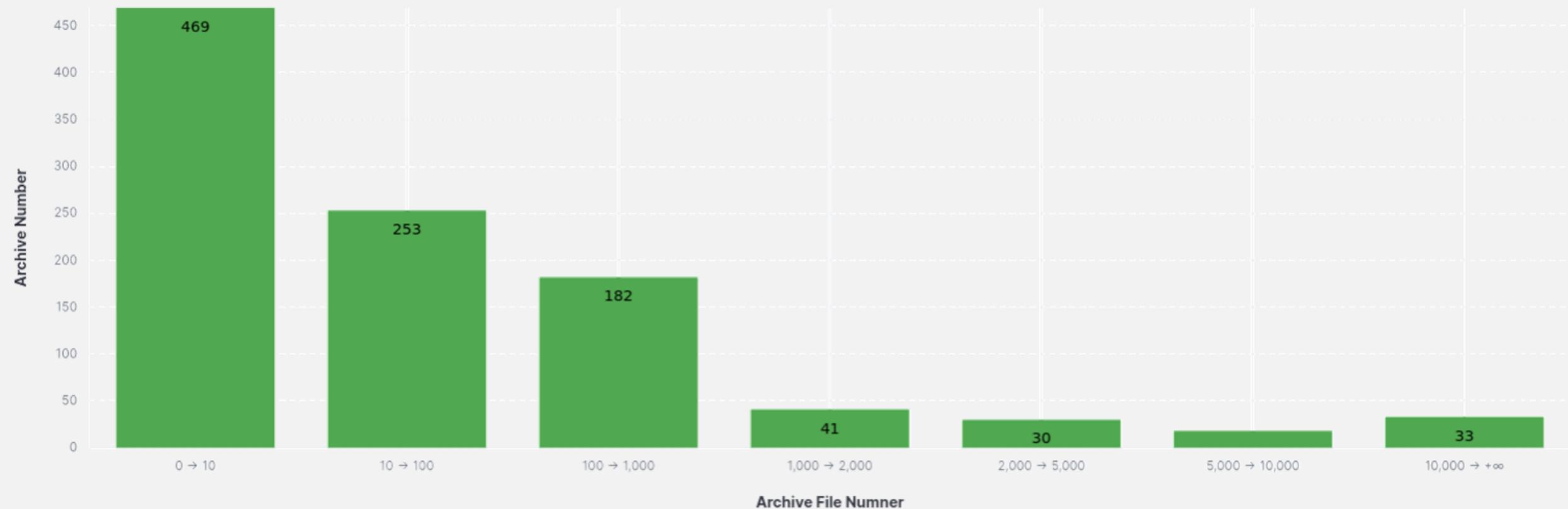


Archive by Size



Quelques chiffres

Archive by File Number



La préservation numérique à UNIGE

Et après...



Communauté DLCM

- On-premise
- SaaS
- Hybrid



2019



2022



2024



Intéressé ?



2021

*Presentations & Discussions
Assessments*

Pilote interne



2023

Testing & Deployment



In progress



Gestion active des données



Import des
données & métadonnées

Accès aux
Données & métadonnées



Web



OAI-PMH

*Cercle vertueux
entre partage et
archivage des
données*

Publié
en
archivant

Partage
en
réimportant

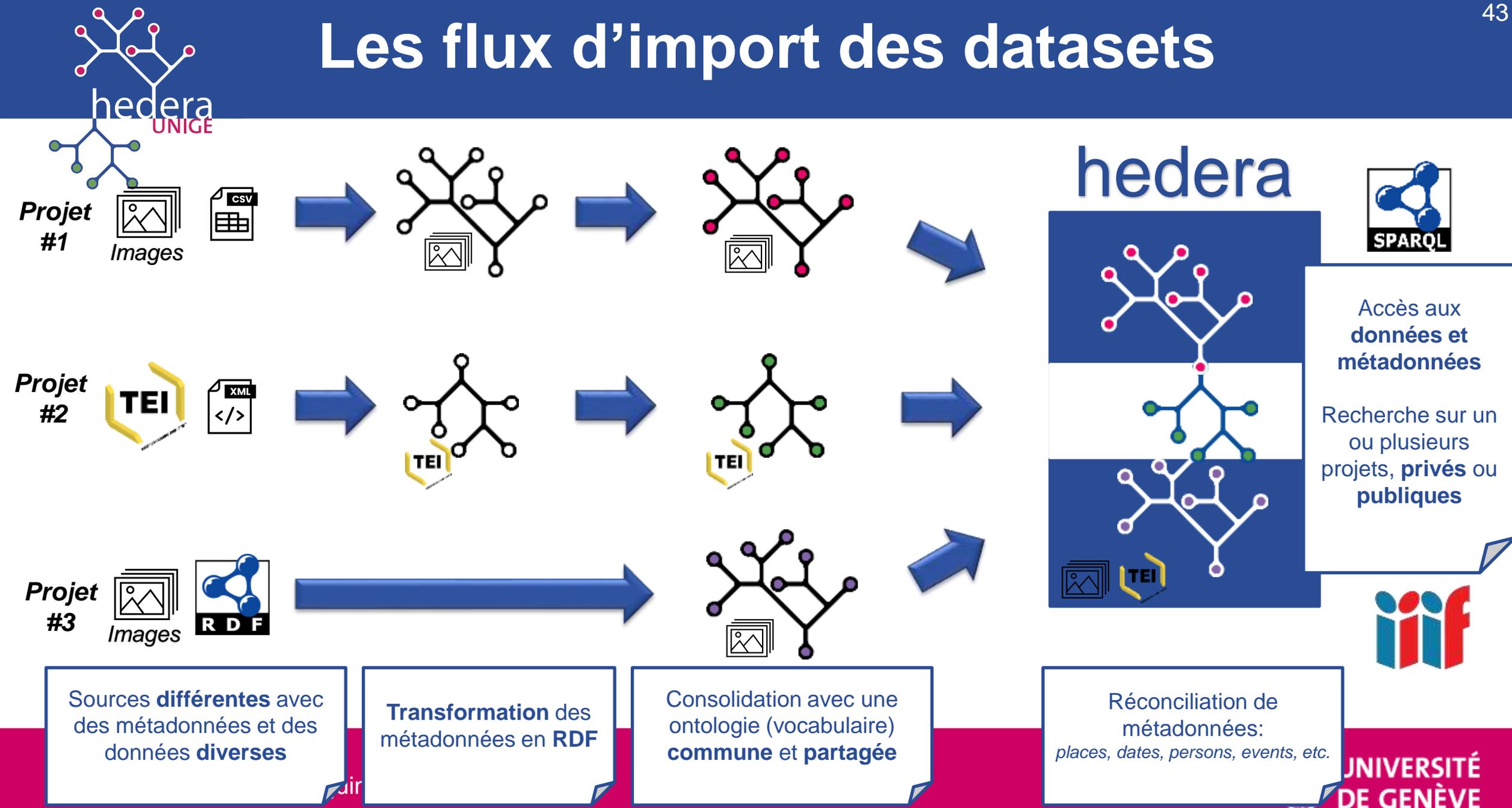
*Lien avec le
nouveau standard
RDF pour
l'archivage RiC-O*



YARETA



Les flux d'import des datasets



Portail hedera



Search data



Home



SPARQL query



Browse

EN


[Home](#) / [Browse](#) / [Project](#) / [Gypso](#) / [Detail](#)

Browse

Projects

Ontologies

[← Back to list](#)


Detail

Research objects

Research data files

Managed objects RDF

IIIF manifests

IIIF collections



Name*

Gypso

Description

Moulages Gypso

Research object Types

author

material

museum

style

technic

Opening

6/6/2024



Closing



URL



Archivage administrative & patrimoniale

- Projet interne avec le secrétariat général de l'Université
- Répondre aux besoins d'archivage du Rectorat
- **Démarche**
 - Socle technologique basé sur DLCM
 - Evaluation des écarts/lacunes avec l'archivage des données de recherche
 - ➔ Développement des aspects manquants (e.g., support des identifiant ARK ou bordereau de versement)
- **Projet (en-cours)**
 - Définition des phases
 - Intégration avec des processus métiers (chaîne de numérisation, SharePoint)

Archivage dans de l'ADN

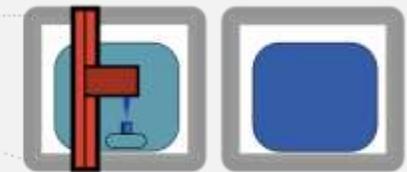
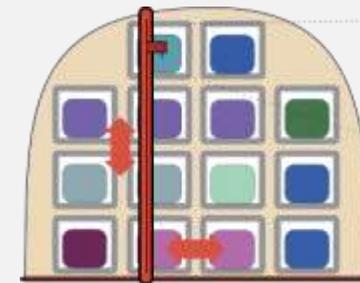
- **Projet européen**
 - <https://dnamic.org/>
 - DNA Microfactory for Autonomous Archiving

- **Objectifs**

Réalisation d'un pilote pour automatiser l'encodage, le stockage et le décodage d'archives dans de l'ADN avec la technologie DLCM et une « micro-factory » pour l'automatisation des opérations.



DLCM



Synthesis, Mixing, Amplification, Sequencing, ... Future technology

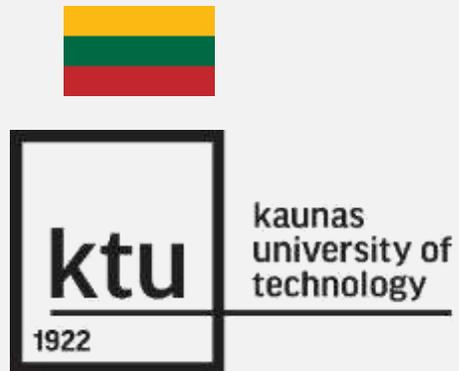
<https://microleanlab.ch>



Les partenaires



GA no. 101115389
dnamic.org



Questions & Réponses

Merci

Références

- **2023**
 - « DLCM Technology » - Congrès ICA 2023 [lien](#)
- **2022**
 - GIT repository: <https://gitlab.unige.ch/dlcm/community/dlcm-backend>
 - Paper for iPRES 2022: « OAIS-compliant digital archiving of research and patrimonial data in DNA » <http://doi.org/10.7207/ipres2022-proceedings>
- **2021**
 - Yareta videos <https://mediaserver.unige.ch/search/yareta>
 - OLOS videos <https://mediaserver.unige.ch/search/olos>
 - « Le projet Data Life-Cycle Management (DLCM) en Suisse : une gestion des données de la recherche pensée pour ses utilisateurs » <http://dx.doi.org/10.3917/i2d.212.0087>
 - « SwissEnvEO: A FAIR National Environmental Data Repository for Earth Observation Open Science » <http://dx.doi.org/10.5334/dsj-2021-022>
- **2019**
 - « DLCM Poster for iPRES 2019 » <http://dx.doi.org/10.26037/yareta:frtet5s3njcubib7vgyg5vvzci>
 - Paper for iPRES 2019: « A Versatile Solution for Long Term Preservation of Research Data - Data Life-Cycle Management (DLCM): The Swiss Way » https://ipres2019.org/static/pdf/iPres2019_paper_32.pdf
 - DLCM movie <https://mediaserver.unige.ch/play/122839>
- **2018**
 - « Le projet DLCM : gestion du cycle de vie des données de recherche en Suisse » <http://dx.doi.org/10.1515/9783110553796-014>
- **2017**
 - « Research data management in Switzerland: National efforts to guarantee the sustainability of research outputs » <http://dx.doi.org/10.1177/0340035216678238>
- **2016**
 - « DLCM : une approche globale pour une meilleure gouvernance du cycle de vie des données de la recherche en Suisse » https://www.researchgate.net/publication/316686422_DLCM_une_approche_globale_pour_une_meilleure_gouvernance_du_cycle_de_vie_des_donnees_de_la_recherche_en_Suisse

Bibliographie

- **Articles scientifiques**

- GIULIANI, Gregory et al. [SwissEnvEO: A FAIR National Environmental Data Repository for Earth Observation Open Science](#). In: *Data Science Journal*, 2021, vol. 20, n° 22. doi: 10.5334/dsj-2021-022

- **Articles professionnels**

- CAZEAUX, Hugues, KRAUSE-BILVIN, Jan, BURGI, Pierre-Yves. [Automatisation de la préservation de données dans le contexte académique](#). In: *Arbido*, 2018, vol. 2.

- **Présentations / Interventions**

- CAZEAUX, Hugues, VONLANTHEN, Mathieu, BURGI, Pierre-Yves. [hedera](#). In: *CIHN'24*. 2024. 11 p.
- CAZEAUX, Hugues, BURGI, Pierre-Yves. [DLCM Technology](#). In: *ICA Congress 2023*. 2023. 31 p.
- GOLD, David Nicolas, CAZEAUX, Hugues. [Coding Dojo: Yareta](#). In: *Coding Dojo recherche - Préserver vos données sur Yareta*. 2023. 15 p.
- GOLD, David Nicolas et al. [Data Science Tools: Yareta](#). In: *[Data Science Competence Center] Data Science Tools. Online*. 2023.
- CAZEAUX, Hugues, VIACCOZ, Cédric, GOLD, David Nicolas. [Data Science Tools: hedera](#). In: *[Data Science Competence Center] Data Science Tools. Remote*. 2023. 12 p.
- CAZEAUX, Hugues, VIACCOZ, Cédric, VONLANTHEN, Mathieu. [DH Day 2023: hedera - Yareta - JupyterHub](#). In: *DH Day 2023. GamMAH*. 2023. 17 p.
- BURGI, Pierre-Yves et al. [OAIS-compliant digital archiving of research and patrimonial data in DNA](#). In: *iPRES - 18th International Conference on Digital Preservation. Glasgow (Scotland)*. 2022.
- BURGI, Pierre-Yves, CAZEAUX, Hugues, ECHERNIER, Lydie. [A versatile solution for long-term preservation of research data: Data Life-Cycle Management: the Swiss Way](#). In: *iPRES - 16th International Conference on Digital Preservation. Amsterdam (The Netherlands)*. 2019. 2 p.

Liens utiles

- Yareta <https://www.unige.ch/eresearch/fr/services/yareta/>
- hedera <https://www.unige.ch/eresearch/fr/services/hedera/>
- Projet DNAMIC <https://dnamic.org/>
- Sandbox DLCM <https://sandbox.dlcm.ch/>
- Projet DLCM <https://www.dlcm.ch/>