

# **Le Matterhorn METS Profile**

## **Un modèle de paquet d'information pour l'archivage numérique en bibliothèque ?**

# Sommaire

1. Un long travail d'appropriation de la matière
2. La plate-forme de pérennisation et d'archivage du Service de la culture
3. Le Matterhorn METS Profile
4. Quelques exemples de paquets d'information

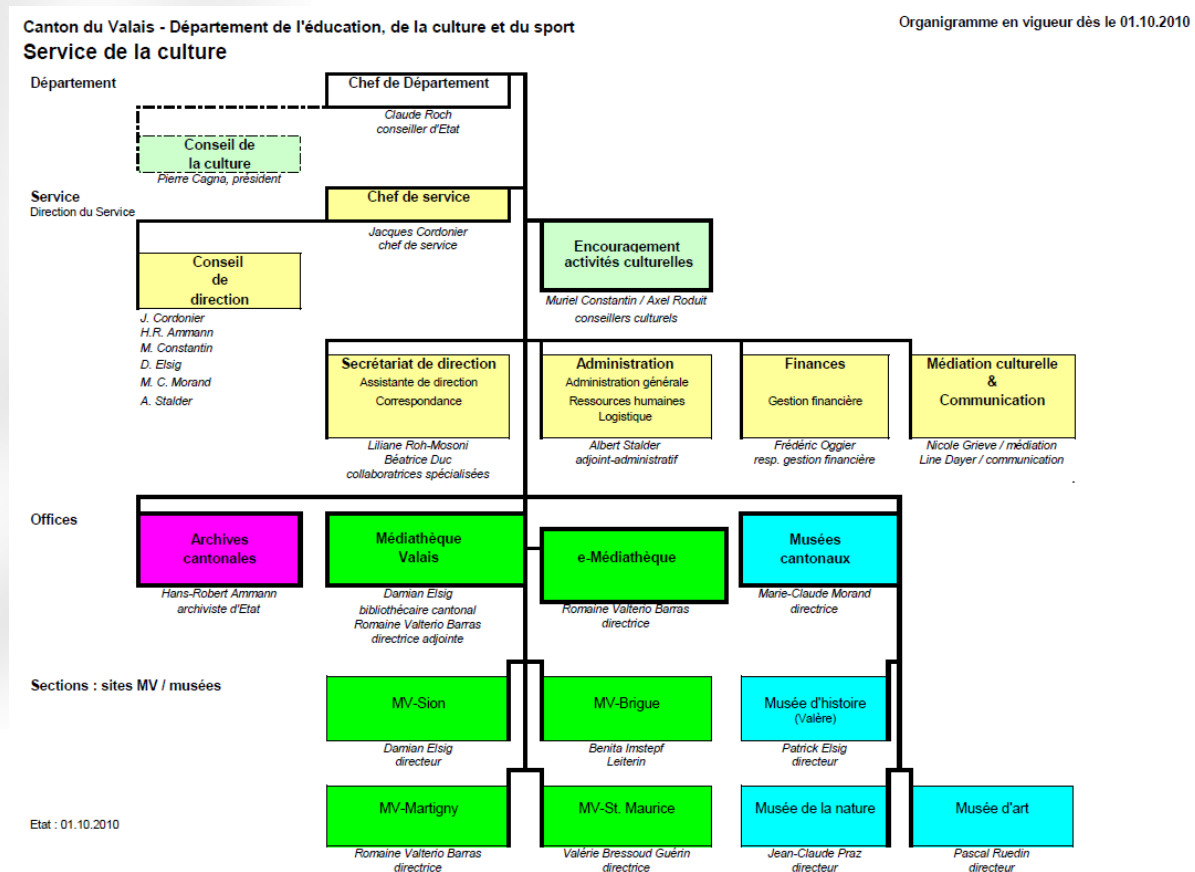
1.

# Un long travail d'appropriation de la matière



# Premiers pas...

## Un élément organisationnel fondateur: la création du Service de la culture



# Premiers pas...

- Point de départ
  - volonté du Chef du Service de la culture de faire le point de la situation en matière de numérisation des collections au sein de son service
- Objectif
  - mettre en œuvre une politique de numérisation commune au sein des Archives de l'Etat du Valais, de la Médiathèque Valais et des Musées cantonaux (projet fédérateur)
- Constitution d'un groupe de travail numérisation composé de:
  - représentants des Archives de l'Etat du Valais, de la Médiathèque Valais et des Musées cantonaux
  - coordinateur informatique du Département de l'éducation, de la culture et du sport

# Premiers pas...

- Mandat du 18.02.2006 – point de situation en matière de numérisation
  - *Le groupe de travail numérisation des collections a pour mandat général de faire le point sur la situation de la numérisation des collections au sein du Service de la culture et de proposer des mesures en vue de renforcer la cohérence de ce secteur.*

*Dans ce but, il*

    - *établit l'inventaire des projets de numérisation réalisés ou en cours dans le cadre du service sur la base d'une fiche descriptive standardisée qu'il établira en accord avec la direction du Service*
    - *établit l'inventaire des projets en cours d'élaboration et des besoins en terme de numérisation à l'intérieur du Service*
    - *établit l'inventaire des équipements et infrastructures de numérisation internes au Service*
    - *formule, à l'attention de la direction du service, les modalités de définition d'une stratégie de numérisation des collections du service.*

# Vers la définition d'une stratégie...

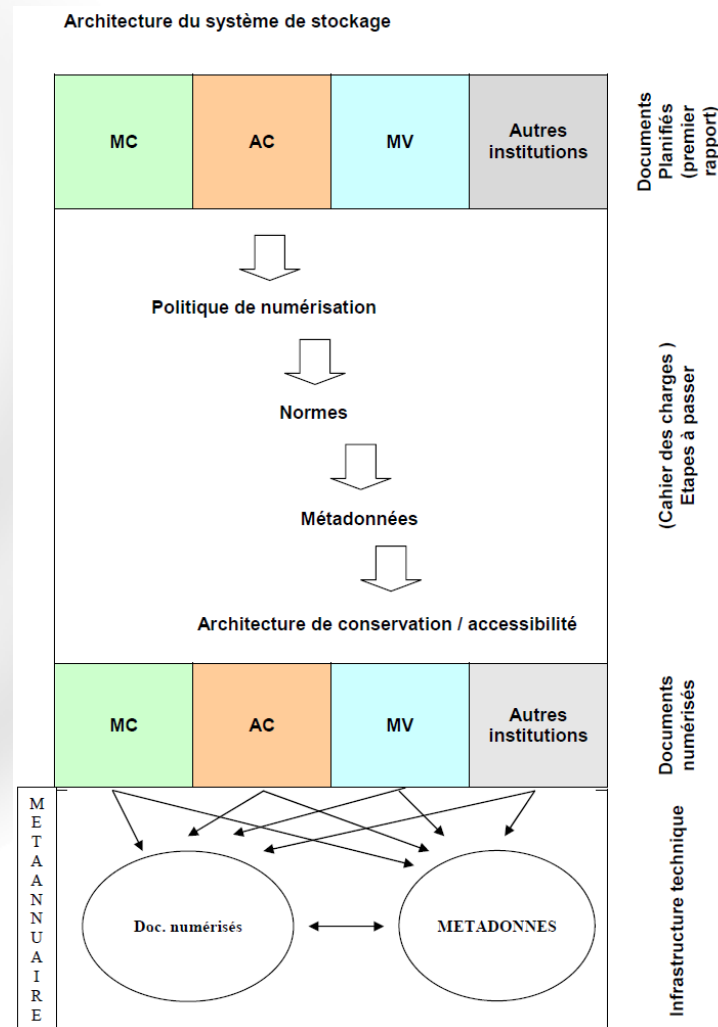
- Rapport final du 25.08.2006
- Nécessité de définir une stratégie en matière de numérisation, de gestion et de conservation à long terme des documents numérisés
  - *Tout ce qui concerne le traitement des données, leur gestion et leur sauvegarde (plus particulièrement à long terme) constitue un grave problème*
  - *L'archivage des données et documents digitaux du Service de la culture doit s'inscrire sur le long terme et faire l'objet d'une éventuelle harmonisation des structures architecturales. Cet objectif de durabilité et d'harmonisation représente un véritable challenge dans un contexte où les techniques digitales ne cessent de progresser, où les usagers se trouvent sans arrêt contraints de convertir et de transférer leurs données pour s'adapter aux nouveaux formats.*

# Vers la définition d'une stratégie...

- Rapport du 11.04.2007
- Définition des différents axes de la stratégie de numérisation
  - *le groupe de travail propose d'élaborer un cahier des charges précis, sur lequel des institutions spécialisées peuvent fournir une offre concrète. Il faudrait préciser dans ce document :*
    1. *la définition d'une politique de numérisation pour chaque unité du SC*
    2. *La définition des normes de numérisation (format, etc.)*
    3. *la définition des niveaux de métadonnées (droits d'accès, publics, internes, délais, etc.)*
    4. *la définition d'une architecture de conservation des données (choix de la méthode de conservation)*



# Vers la définition d'une stratégie...



# Elaboration d'un manuel de numérisation

- Mandat du 4 juin 2007
  - *Le groupe de travail, accompagné par un expert, établit un manuel de numérisation du service, en*
    1. *identifiant et sélectionnant les contraintes minima à remplir pour chaque projet de numérisation, en respectant les aspects évolutifs du NTIC*
    2. *définissant une(des) procédure(s) cohérente(s) pour la gestion (le traitement, l'archivage, la sauvegarde et la publication) des projets de numérisation des collections numérisées (prises de vues, documents, audio, vidéo) au sein du SC en élaborant des outils de réalisation (formulaire, check-list, etc.) accompagnant tout les projets de numérisation*
    3. *proposant une structure de responsabilités / d'accompagnement afin de garantir une gestion efficace des projets de numérisation*
    4. *faisant une analyse du contexte de forces / faiblesses des outils utilisés et propose, en respectant les résultats du points a), des améliorations éventuelles.*

# Elaboration d'un manuel de numérisation

- Le groupe de travail est accompagné par Docuteam GmbH
- Interviews réalisées dans les différents offices du Service de la culture au cours du printemps 2008
- Cinq thèmes de discussion identifiés
  - réflexions stratégiques en matière de numérisation des documents
  - workflows
  - contrôle qualité
  - métadonnées
  - gestion des documents numérisés
- « Bericht zuhanden der Dienststelle für Kultur, Kanton Wallis: Digitalisierung – Strategien, Workflows und Infrastruktur » remis par Docuteam GmbH en septembre 2008

# Elaboration d'un manuel de numérisation

- Rapport remis le 10 mars 2009 par le groupe de travail numérisation
- *Proposition d'une stratégie pour la gestion des données électroniques au sein du Service de la culture*
- Le rapport définit non seulement une politique en matière de gestion des projets de numérisation, mais établit également des normes et standards en matière de conservation à long terme des données électroniques
- Un projet de numérisation doit obéir à l'un des quatre objectifs suivants:
  1. *conservation des collections*
  2. *amélioration de l'accès*
  3. *documentation de travail*
  4. *réponses à des demandes spécifiques de la part des clients*

# Elaboration d'un manuel de numérisation

## ■ Définition des formats d'archivage

Type de données	Format d'archivage	Format d'affichage
Image	TIFF	JPEG, PNG
Texte	PDF/A	PDF/A peut être lu par PDF
Audio	WAVE	Utilisation sur place : WAVE, MP3, Ogg Vorbis Utilisation sur Internet : MP3, Ogg Vorbis
Vidéo	DV (susceptible d'évoluer), DV CAM Digital Video	MPEG2 compression pour DVD et internet

## ■ Définition du standard d'échange pour les métadonnées descriptives

Entité	Systèmes de catalogage et de description	Schémas de métadonnées	Formats d'échange
Archives d'Etat	scopeArchiv	ISAD(G), ISAAR (CPF)	EAD, Dublin Core
Médiathèque Valais	Rero	Marc21	Marc XML, Dublin Core
Musées cantonaux	Klio		Dublin Core

# Elaboration d'un manuel de numérisation

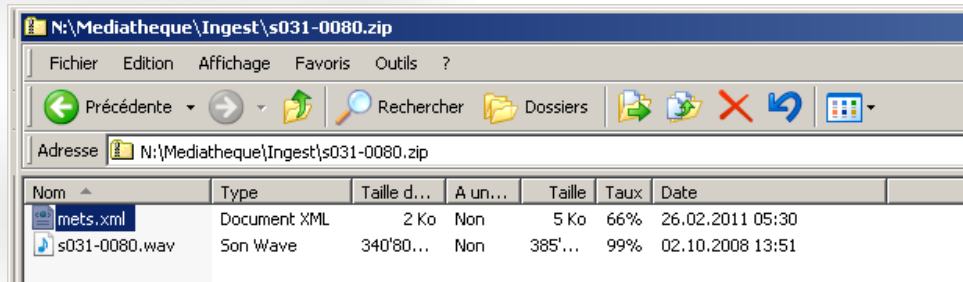
- Définition d'un set minimal de métadonnées descriptives
  - ✓ cote
  - ✓ titre
  - ✓ description
  - ✓ année
  - ✓ créateur (auteur)
  - ✓ volume
  
- Définition d'un set minimal de métadonnées techniques et administratives
  - ✓ cote
  - ✓ format de fichier (y compris la version)
  - ✓ rapport d'audit
  - ✓ mesures conservatoires prises

# Elaboration d'un manuel de numérisation

## ■ Définition d'un paquet d'information standard

fichier .zip composé

- ✓ des données primaires
- ✓ d'un fichier mets.xml contenant notamment
  - les métadonnées descriptives (Dublin Core / EAD)
  - les métadonnées techniques et administratives (Premis)



# Elaboration d'un manuel de numérisation

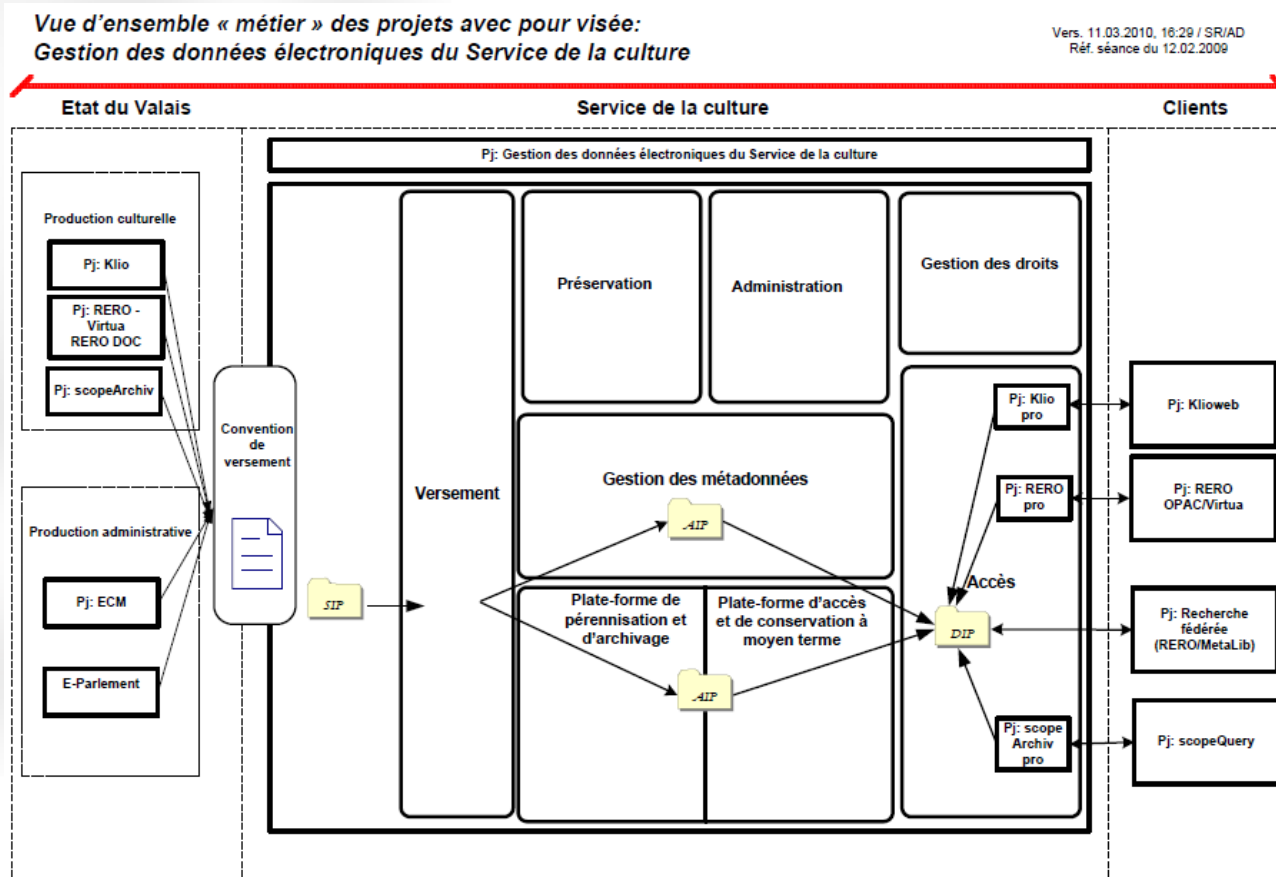
- Absence de stratégie commune en matière de conservation à long terme des données électroniques

Entité	Supports de données
Archives de l'Etat du Valais	serveur, CD
Médiathèque Valais – Sion	disques durs externes, CD, DVD
Médiathèque Valais – Martigny	location de deux SAN auprès de RERO et de Canal 9
Musées cantonaux	CD, disques durs externes



# Elaboration d'un manuel de numérisation

- Définition métier de la plate-forme de pérennisation et d'archivage



# Elaboration d'un manuel de numérisation

- Validation du manuel de numérisation par le comité de pilotage Service de la culture – Service cantonal de l'informatique en juin 2009
- Réflexions stratégiques (2006-2009)
- Mise en œuvre pratique (2009-2011)
  - ✓ projet « Gestion et stockage de masse des documents numériques au sein du Service de la culture » (06.2009 - 04.2011)
    - > mise en production le 1<sup>er</sup> mai 2011
    - > 15 TB issus des collections numériques et numérisées du Service de la culture et du dépôt des publications officielles des autorités cantonales (360 942 objets)

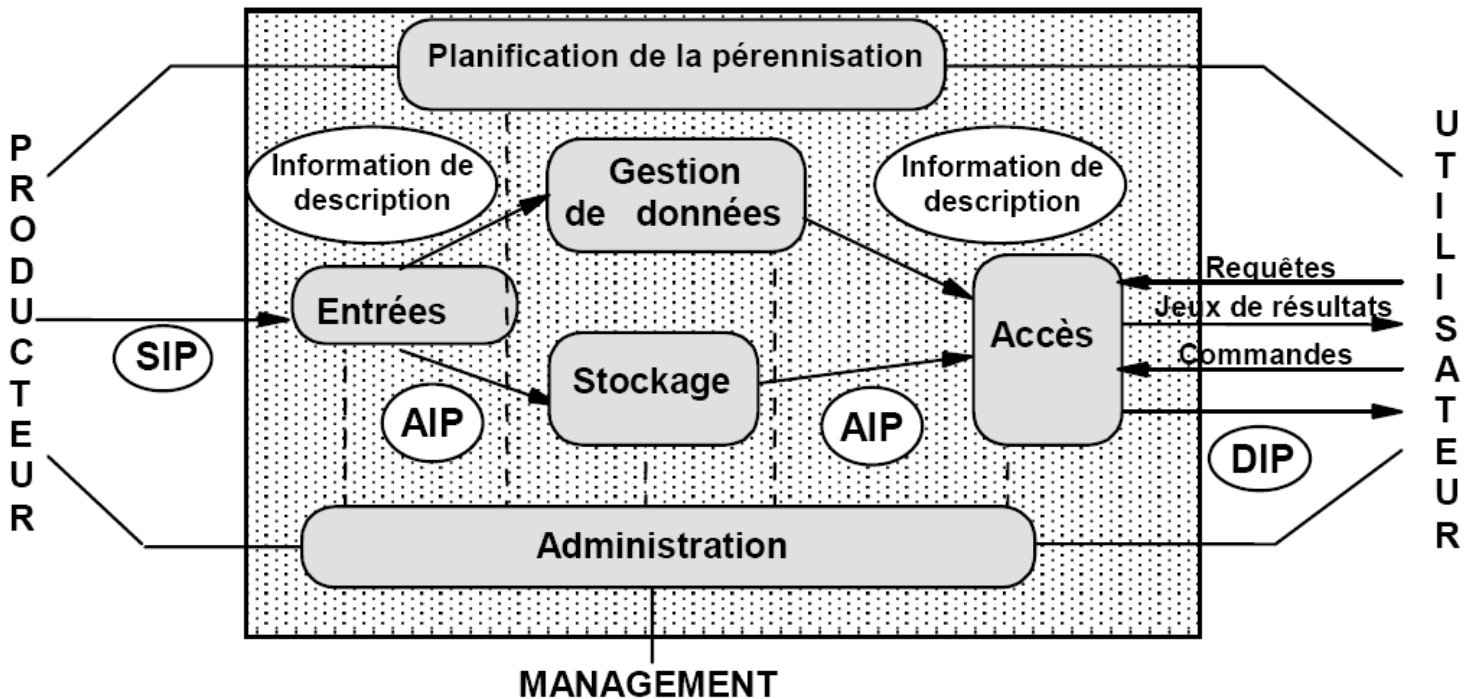
## 2.

# La plate-forme de pérennisation et d'archivage du Service de la culture



# La plate-forme de pérennisation

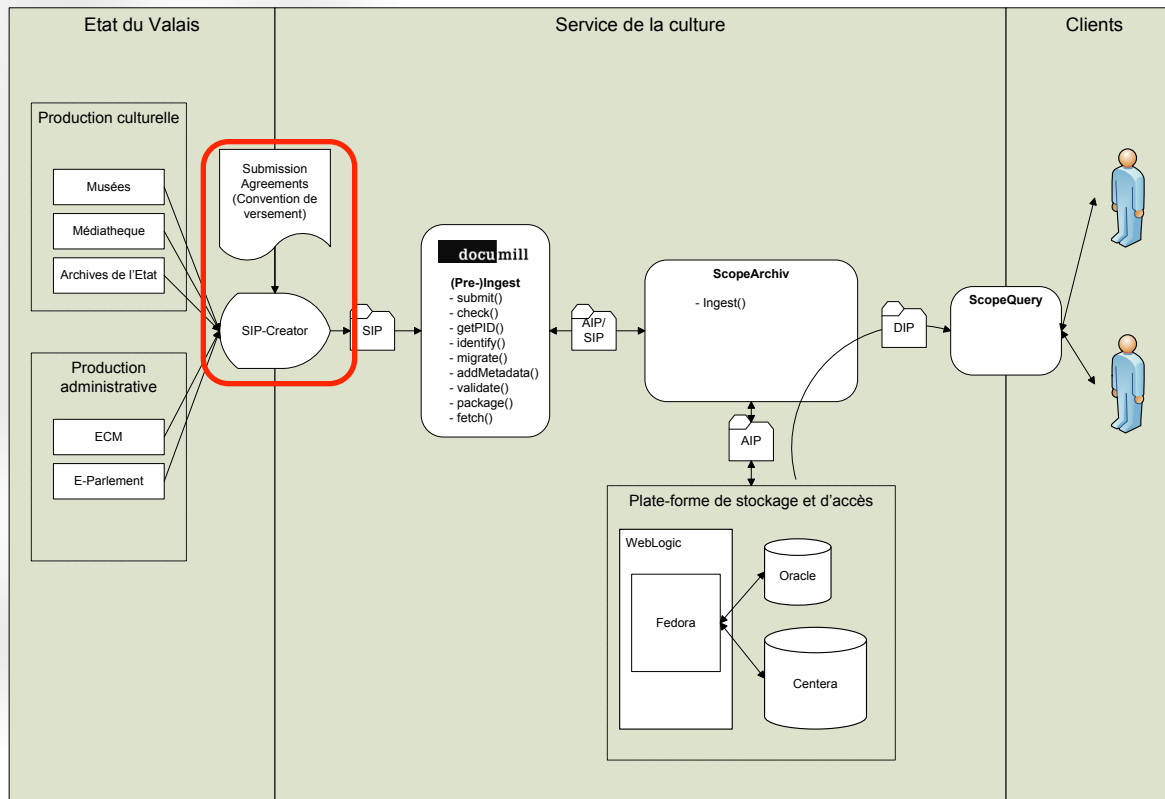
- Modèle de référence OAIS





# Le protocole de versement

## ■ Le protocole de versement



# Le protocole de versement

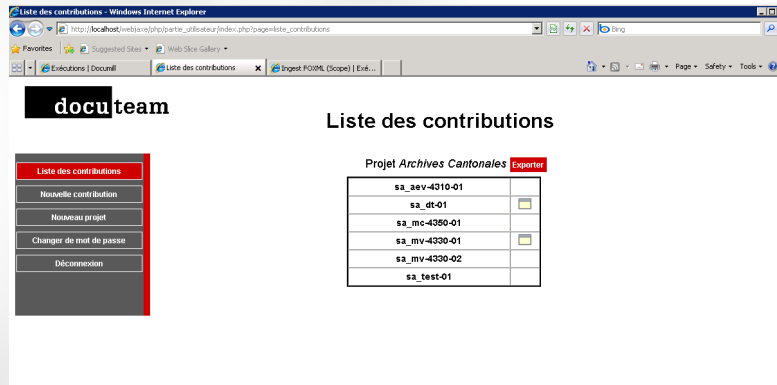
## ■ Objectif

- ✓ tout versement est conditionné par la conclusion préalable d'un protocole de versement entre le producteur et les Archives. Ce protocole de versement mentionne les éléments suivants :

1. informations sur le producteur
2. typologie des documents et des données versés
3. critères d'échantillonnage
4. période de création
5. volume
6. copyright
7. modalités d'accès
8. formats de fichiers acceptés en entrée
9. métadonnées descriptives, techniques et administratives
10. structure du paquet d'information versé (SIP)
11. rôles et responsabilités des intervenants

# Le protocole de versement

- Edition d'un protocole de versement (outil *open source*)
- ✓ Choix du protocole de versement



- ✓ Edition d'une contribution





# Le protocole de versement

## ■ Exemple de protocole de versement (Médiathèque Valais)

Producteur de dossiers

Coordonnées du contact

**Kontakt-ID** prod44330-02

**Organisation** Canton du Valais, Département de l'éducation, de la culture et du sport

**Departement** Médiathèque Valais - Martigny

**Abréviation** MVM

**Person** Besim Berisha

**Telephone** +41 27 722 91 92

**Mail** mv.martigny@mediatheque.ch

Producteur de dossiers

Archive 'CH-000019-X'

Coordonnées du contact

**Kontakt-ID** arch4310-01

**Organisation** Canton du Valais, Département de l'éducation, de la culture et du sport

**Departement** Archives de l'Etat du Valais

**Abréviation** AEV

**Person**

**Telephone** + 41 27 606 47 00

**Mail** archives@admin.vs.ch

Archive 'CH-000019-X'

Vue d'ensemble

Documents

**Description générale**

La Médiathèque Valais numérise le patrimoine audio, vidéo et photographique valaisan à des fins d'études et de conservation et pour des raisons de sécurité en cas d'incendie, d'inondation ou de vol notamment. Ces documents numérisés sont gérés par l'outil Rero. Leur version haute résolution doit par contre être stockée sur la plateforme de pérennisation et d'archivage des Archives de l'Etat du Valais. [Description générale](#)

**Création**

Les collections numériques versées sur la plateforme de pérennisation et d'archivage ont été créées depuis l'année 2000. [Création](#)

**Droit d'auteur/d'utilisation**

Le copyright est propriété de la Médiathèque Valais. Cette mention est intégrée aux métadonnées descriptives du fichier. [Droit d'auteur/d'utilisation](#)

Documents

**Décision de triage** Ces documents doivent être conservés sur le long terme en raison de leur intérêt patrimonial et pour des raisons de sécurité en cas d'incendie, d'inondation ou de vol notamment. [Décision de triage](#)

Vue d'ensemble

Livraison des dossiers

Titre [Si on - Images](#) [Titre](#)





# Le protocole de versement

- Exemple de protocole de versement (Médiathèque Valais)

Livraison des dossiers

Titre Martigny - Films

Calendrier

Date de démarrage 2011-03-28

Fréquence

unique  unique

Nombre 1

Unité Month(s)

Calendrier

Transmission

Cryptographique  Cryptographique

Initiateur prod4330-02

Mode Dropbox

Instructions

Categorie Addition

Type de dossier (DC) MovingImage

Format de fichier

stricte  stricte

PRONOM-PUID fmt/5

PRONOM-PUID fmt/199

PRONOM-PUID x-fmt/152

PRONOM-PUID x-fmt/386

Mimetype

Description

Suffix

Taille de fichiers

Unité Terabyte(s)

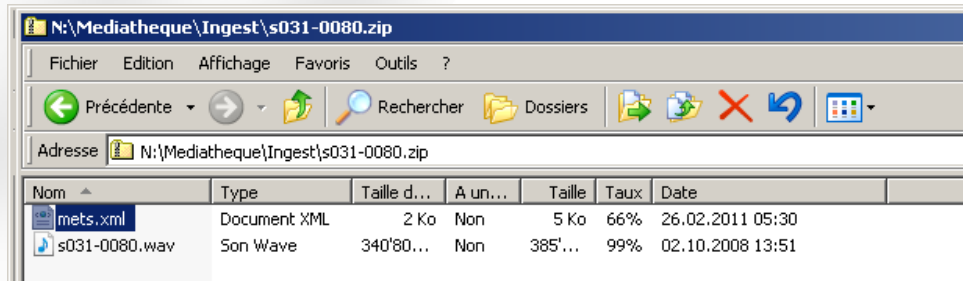
Taille 8

Livraison des dossiers

Convention de versement

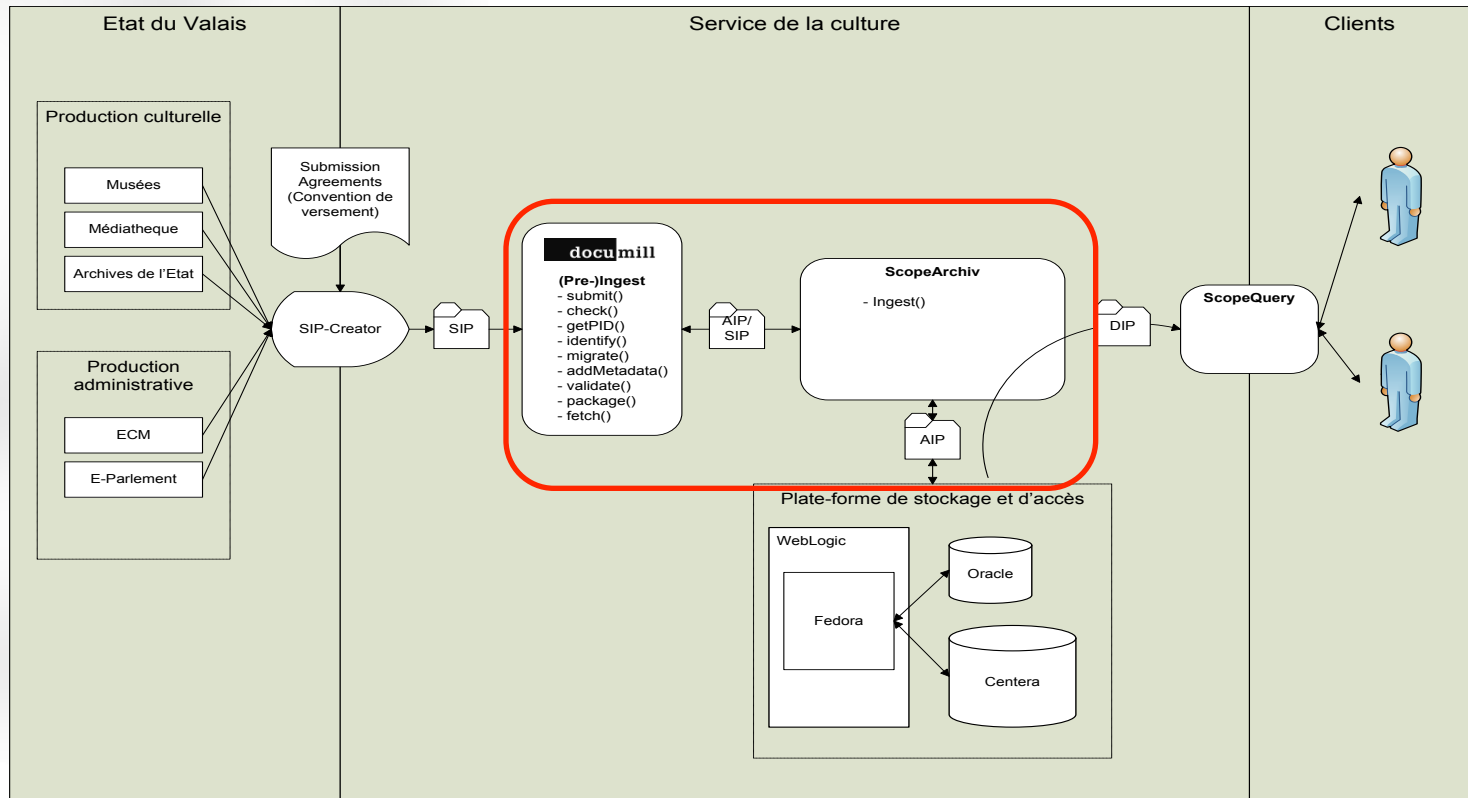
# La création des SIP

- Structure du SIP – fichier .zip composé
  - ✓ des données primaires
  - ✓ d' un fichier mets.xml contenant notamment
    - les métadonnées descriptives (Dublin Core)
    - les métadonnées techniques et administratives (Premis)



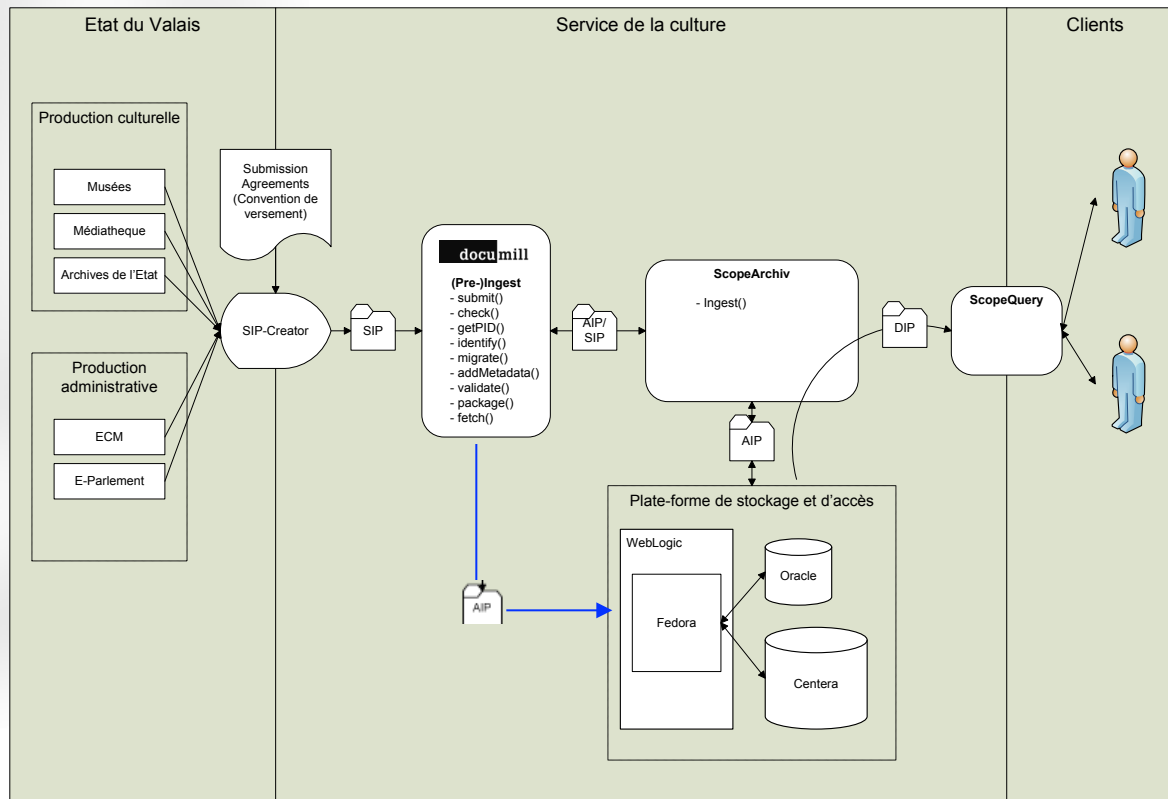
- SIP déposé dans un répertoire dédié (« boîte aux lettres »), accessible tant au producteur qu' aux outils d' *ingest* de la plate-forme de pérennisation et d' archivage (cf. répertoire 'Mediatheque\SIP')

# Les entités « Versement » et « Gestion de données »



# L'Ingest d'un SIP

- Mise en œuvre de deux processus



# L'*Ingest* d'un SIP – processus MV et MC

- Processus d'*Ingest* d'une photographie :
  1. connexion au serveur de fichier Novell, répertoire 'Mediatheque\Ingest'
  2. extraction du SIP à partir du répertoire Novell
  3. vérification des somme de contrôle (*checksum*), qui permet à la plate-forme de pérennisation et d'archivage de vérifier que le SIP reçu est bien celui qui a été envoyé
  4. vérification des formats de fichier (selon le protocole de versement)
  5. migration des fichiers dans migration-config.xml
  6. demande d'un identifiant unique et pérenne (PID) pour le SIP
  7. normalisation des métadonnées
  8. création des fichiers en format Fedora (FOXML)
  9. transfert des fichiers FOXML vers Fedora
  10. rangement des fichiers temporaires
  11. suppression du SIP
  12. envoi du PID au producteur



# L'Ingest d'un SIP – processus MV et MC

- Processus d'Ingest d'une photographie :

The screenshot shows a web browser window with the URL `http://sringest01.exploit.vs.ch:3000/workflow_executions/11227`. The page content is as follows:

**docu mill** Exécutions | Gestions des flux | Pas | Utilisateurs

## MV Martigny - ingest sans garder les originales

créé par admin le 28 décembre, 14:15

**SIP:** 008ph-00963  
**Gestion des flux commencé à:** mercredi, 28 décembre 2011, 14:15  
**Durée totale:** 00:00:35

**Connect Novell**  
**Status:** finit avec Code 0  
**Commande:** IF NOT EXIST \SIION-MEDIA01\USR4\DATA C:\Batch>LoginNovell\FEDCOM.bat  
**Durée:** 00:00:01

**Extraire SIP directement du Novell**  
**Status:** finit avec Code 0  
**Commande:** java ch.docuteam.documill.qualityassurance.SIPExtractor \SIion-media01\usr4\Data\SC\Mediatheque\ngest\008ph-00963.zip  
**Réponse:** INFO 14:15:33.557 (SIPExtractor) Initializing SIP extraction...  
INFO 14:15:33.557 (SIPExtractor) Extracting SIP: 008ph-00963.zip  
INFO 14:15:35.619 (SIPExtractor) Extraction successful  
**Durée:** 00:00:02

**vérification des checksums**  
**Status:** finit avec Code 0  
**Commande:** java ch.docuteam.documill.qualityassurance.SIPFixityCheck 008ph-00963  
**Réponse:** INFO 14:15:36.354 (SIPFixityCheck) Initializing fixity check...  
INFO 14:15:36.354 (SIPFixityCheck) Checking fixity information: c:\workbench2\_work\008ph-00963\mets.xml...  
INFO 14:15:36.635 (NodeFile) Fixity Check: C:\workbench2\_work\008ph-00963\008ph-00963.tif...  
INFO 14:15:39.338 (NodeFile) Checksum OK for file: C:\workbench2\_work\008ph-00963\008ph-00963.tif  
INFO 14:15:39.338 (SIPFixityCheck) Fixity check successful.  
INFO 14:15:39.354 (SIPFixityCheck) mets.xml updated.  
**Durée:** 00:00:04

**vérification des formats selon la convention de versement**  
**Status:** finit avec Code 0  
**Commande:** java ch.docuteam.documill.qualityassurance.SIPSubmissionAgreementCheck 008ph-00963 false  
**Réponse:** INFO 14:15:39.916 (SIPSubmissionAgreementCheck) Initializing submission agreement check...  
INFO 14:15:39.916 (SIPSubmissionAgreementCheck) Checking SIP against submission agreement: c:\workbench2\_work\008ph-00963\mets.xml...  
INFO 14:15:40.135 (SIPSubmissionAgreementCheck) Submission agreement check OK.

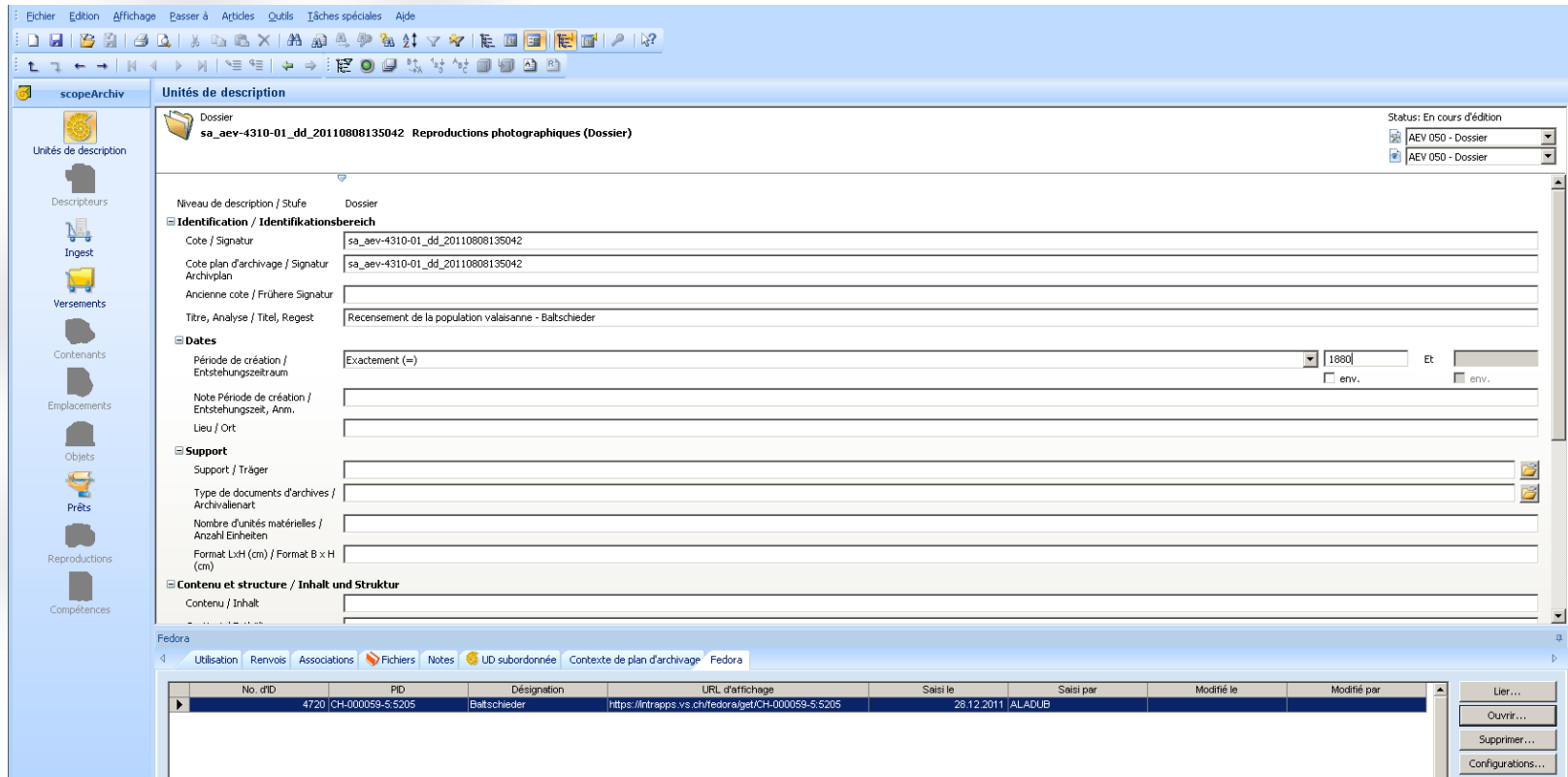
# L'Ingest d'un SIP – processus MV et MC

- Envoi d'un accusé de réception au producteur
  - ✓ il atteste que le SIP est stocké sur la plate-forme de pérennisation et d'archivage sous forme d'AIP
  - ✓ il contient l'identifiant pérenne et unique (PID), qui doit être utilisé ultérieurement par le producteur pour accéder à l'AIP
  - ✓ il est déposé dans un répertoire dédié, accessible au producteur (cf. répertoire 'Mediatheque\PID')

```
008ph-00963.txt
#=====
#
#===== 28.12.2011 14:15:56
#=====
#
# Link to the fedora DB: <http://jbossasp01.exploit.vs.ch:8080/fedora>
#Operation PID File/Folder Errorcode
#-----
-
Store CH-000019-X:147630 008ph-00963.tif 0
```

# L'Ingest d'un SIP – processus AEV

- Lien entre un AIP et son unité de description s'effectue à travers l'onglet Fedora disponible dans le module unité de description



Eichier Edition Affichage Passer à Actides Outils Tâches spéciales Aide

scopeArchiv Unités de description

Dossier sa\_aev-4310-01\_dd\_20110808135042 Reproductions photographiques (Dossier)

Status: En cours d'édition  
 AEV 050 - Dossier  
 AEV 050 - Dossier

Niveau de description / Stufe Dossier

**Identification / Identifikationsbereich**

Cote / Signatur sa\_aev-4310-01\_dd\_20110808135042  
 Cote plan d'archivage / Signatur Archivplan sa\_aev-4310-01\_dd\_20110808135042  
 Ancienne cote / Frühere Signatur  
 Titre, Analyse / Titel, Regest Recensement de la population valaisanne - Baltschieder

**Dates**

Période de création / Entstehungszeitraum Exactly (=) 1880 Et env. env.  
 Note Période de création / Entstehungszeitraum, Anm.  
 Lieu / Ort

**Support**

Support / Träger  
 Type de documents d'archives / Archivallenart  
 Nombre d'unités matérielles / Anzahl Einheiten  
 Format LxH (cm) / Format B x H (cm)

**Contenu et structure / Inhalt und Struktur**

Contenu / Inhalt

Fedora

No. d'ID	PID	Désignation	URL d'affichage	Saisi le	Saisi par	Modifié le	Modifié par
4720	CH-000059-S-5205	Baltschieder	https://intrapps.vs.ch/fedora/get/CH-000059-S-5205	28.12.2011	ALADUB		

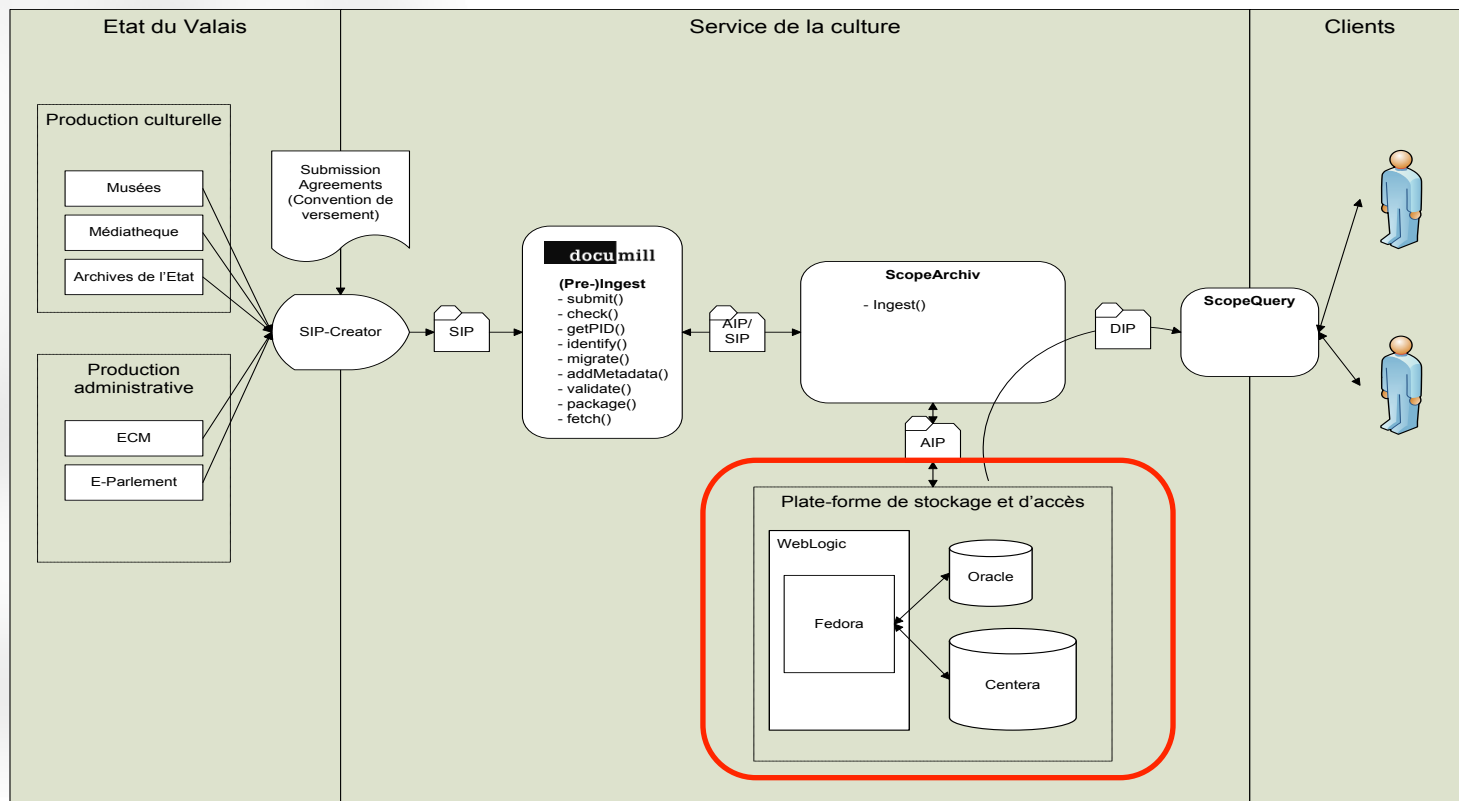
Utilisation Renvois Associations Fichiers Notes UD subordonnée Contexte de plan d'archivage Fedora

Lier...  
 Ouvrir...  
 Supprimer...  
 Configurations...





# L'entité « stockage »



# La plate-forme de pérennisation



## Fedora

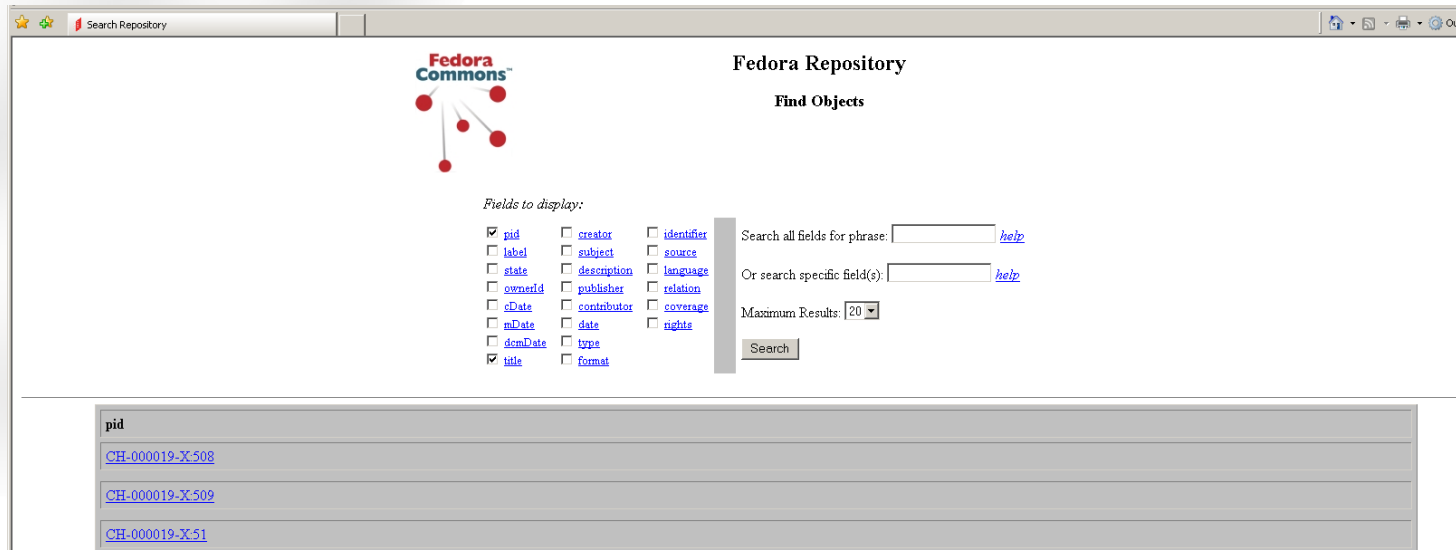
### Repository Information View

**Repository Name:** Archives électronique de l'Etat du Valais

<b>Base URL:</b>	<a href="https://intrapps.vs.ch.443/fedora">https://intrapps.vs.ch.443/fedora</a>
<b>Version:</b>	3.4.1
<b>PID Namespace:</b>	ch-000059-5
<b>PID Delimiter:</b>	:
<b>Sample PID:</b>	ch-000059-5:100
<b>Retain PID Namespace:</b>	*
<b>OAI Namespace:</b>	vs.ch
<b>OAI Delimiter:</b>	:
<b>Sample OAI Identifier:</b>	oai:vs.ch:ch-000059-5:100
<b>Sample Search URL:</b>	<a href="https://intrapps.vs.ch.443/fedora/objects">https://intrapps.vs.ch.443/fedora/objects</a>
<b>Sample Access URL:</b>	<a href="https://intrapps.vs.ch.443/fedora/objects/demo:5">https://intrapps.vs.ch.443/fedora/objects/demo:5</a>
<b>Sample OAI URL:</b>	<a href="https://intrapps.vs.ch.443/fedora/oai?verb=Identify">https://intrapps.vs.ch.443/fedora/oai?verb=Identify</a>
<b>Admin Email:</b>	alain.dubois@admin.vs.ch
<b>Admin Email:</b>	frederic.follonier@admin.vs.ch
<b>Admin Email:</b>	admin@docuteam.ch

# La plate-forme de pérennisation

- Tout AIP
  - ✓ est géré par Fedora Commons
  - ✓ accessible grâce à son identifiant pérenne et unique (PID)



Search Repository

Fedora Commons

Fedora Repository  
Find Objects

Fields to display:

<input checked="" type="checkbox"/> pid	<input type="checkbox"/> creator	<input type="checkbox"/> identifier
<input type="checkbox"/> label	<input type="checkbox"/> subject	<input type="checkbox"/> source
<input type="checkbox"/> state	<input type="checkbox"/> description	<input type="checkbox"/> language
<input type="checkbox"/> ownerid	<input type="checkbox"/> publisher	<input type="checkbox"/> relation
<input type="checkbox"/> cDate	<input type="checkbox"/> contributor	<input type="checkbox"/> coverage
<input type="checkbox"/> mDate	<input type="checkbox"/> date	<input type="checkbox"/> rights
<input type="checkbox"/> dcModifiedDate	<input type="checkbox"/> type	
<input checked="" type="checkbox"/> title	<input type="checkbox"/> format	

Search all fields for phrase:  [help](#)

Or search specific field(s):  [help](#)

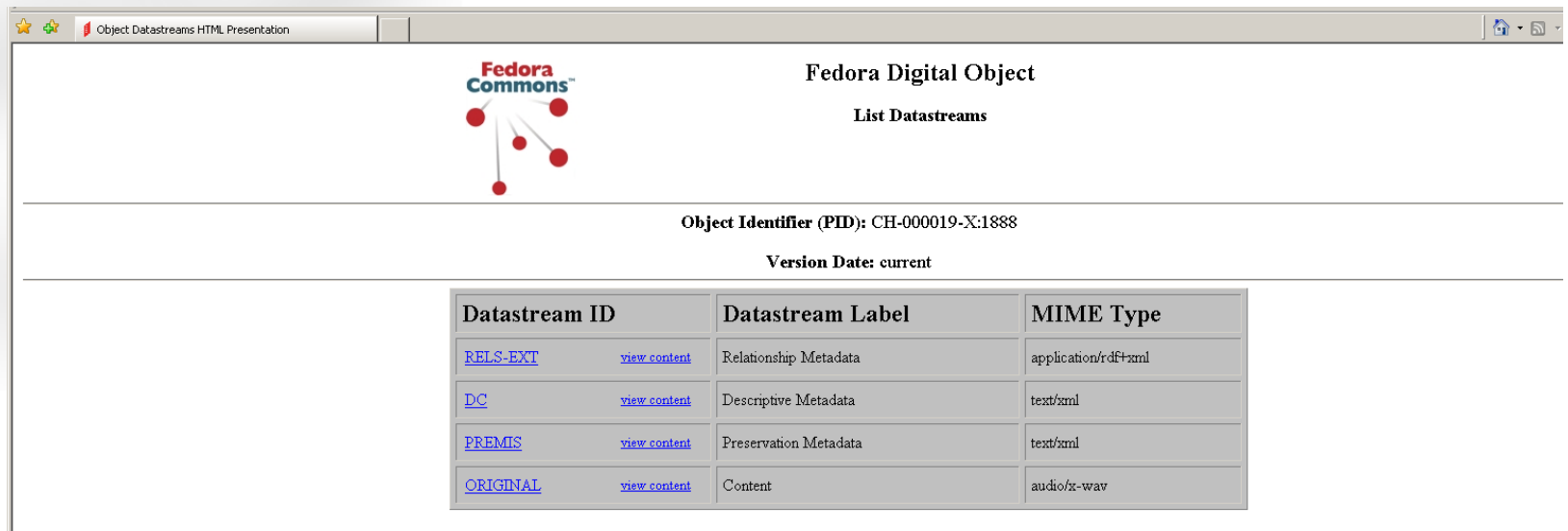
Maximum Results: 20

Search

pid
<a href="#">CH-000019-X-508</a>
<a href="#">CH-000019-X-509</a>
<a href="#">CH-000019-X-51</a>

# La plate-forme de pérennisation

- Pour chaque AIP, il est possible de consulter
  - ✓ ses relations (encodées en RDF)
  - ✓ ses métadonnées descriptives (encodées en Dublin Core ou EAD)
  - ✓ ses métadonnées techniques et administratives (encodées en PREMIS)
  - ✓ son ou ses fichiers originaux



The screenshot shows a web browser window with the title 'Object Datastreams HTML Presentation'. The main content area displays the 'Fedora Commons' logo on the left and the text 'Fedora Digital Object' and 'List Datastreams' on the right. Below this, the 'Object Identifier (PID): CH-000019-X:1888' and 'Version Date: current' are shown. At the bottom, there is a table with four datastreams listed.

Datastream ID	Datastream Label	MIME Type
<a href="#">RELS-EXT</a> <a href="#">view content</a>	Relationship Metadata	application/rdf+xml
<a href="#">DC</a> <a href="#">view content</a>	Descriptive Metadata	text/xml
<a href="#">PREMIS</a> <a href="#">view content</a>	Preservation Metadata	text/xml
<a href="#">ORIGINAL</a> <a href="#">view content</a>	Content	audio/x-wav