

<http://www.creativeaspects.co.uk/>

Bibliothèque nationale de France – SPAR– 19 mai 2008
Laurent Duplouy – Bibliothèque nationale de France / Département des Systèmes d'information

Groupes de travail

- Les groupes de travail fonctionnels: Ils comprennent des utilisateurs représentatifs des départements de la BnF:
 - G1 : groupe transverse de suivi
 - G2 : communauté d'utilisateurs
 - G3 : modèles d'information
 - G4 : gestion des risques
 - G2D : gestion des droits
- Groupe de travail technique (Les membres étant du DSI).

Groupe G2 : les filières

- Numérisation de conservation
 - imprimés, périodiques, images fixes
 - audiovisuel et multimédia
- Numérisation de consultation
- Dépôt légal automatique
- Dépôt légal négocié
- Production administrative/ technique
- Dons et acquisitions
- Tiers archivage

Les filières

- Définissent les politiques (Versement, préservation, accès)
- Sont négociées avec la direction
- Ont une représentation XML → METS profil

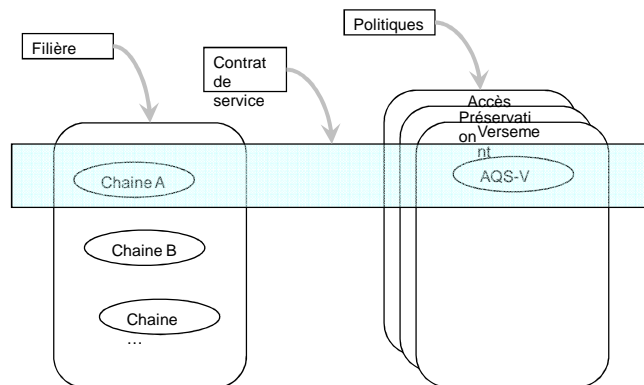
Les filières : définition

- Les filières se définissent par les exigences :
- du système d'archivage vis-à-vis du producteur.
 - Négociation préalable (possible ou non)
 - Cadre législatif et réglementaire (impératif ou non)
 - Conditions d'accès
- du producteur vis-à-vis du système d'archivage
 - Immuabilité des données (modification, version, élimination ou non)
 - Pérennité des données (données, métadonnées, structures et représentations)
 - Accessibilité (directe / applicative, immédiate / différée)

Les chaînes

- 1 filière est décomposée en chaînes
- 1 chaîne = Contrat de service
 - Définit les AQS (Accord sur la Qualité de Service) : AQS-Versement, AQS-Préservation, AQS-Diffusion
 - Négocié avec les producteurs, les utilisateurs, les gestionnaires de collection et les représentants de l'Archive
 - Représentation XML : bnf_aqs

Les filières et les chaînes



AQS de versement

- Les termes d'un AQS-V doivent définir :
 - la structure des SIP : elle comprend le « METS profile » et une description complète cf. « Cahier Technique Détaillé »,
 - la fréquence de livraison,
 - le volume minimum d'une livraison en octets,
 - le volume maximum d'une livraison en octets,
 - le volume moyen d'une livraison en octets (à titre indicatif),
 - le volume total maximum,
 - le temps minimum de prise en compte d'une livraison,
 - le temps maximum de prise en compte d'une livraison,
 - le temps moyen de prise en compte d'une livraison (à titre indicatif),
 - le nombre maximum d'Objets-Données,
le niveau de sécurité (habilitation),
le cryptage éventuel du transfert,
 - la disponibilité de service : plage horaire de disponibilité du service (par exemple, de 2 heures à 8 heures du lundi au vendredi).

AQS de préservation

- Les termes d'un AQS-P doivent définir :
- temps de rétention (éventuellement infini),
- niveau de préservation :
 - préservation du train de bits,
 - surveillance du format (nécessite un format connu),
 - préservation de la représentation des Objets-Données (nécessite un format maîtrisé),
- niveau d'assurance : nombre de copies, localisation des copies (locales ou distantes), fréquence des audits.

AQS de diffusion

- Les termes d'un AQS-D doivent définir :
 - la structure des DIP : elle comprend le « METS profile » et une description complète cf. « Cahier Technique Détaillé »,
 - le niveau de sécurité (habilitation),
 - le cryptage éventuel du transfert,
 - l'authenticité éventuelle des données transférées,
 - le temps minimum de mise à disposition,
 - le temps maximum de mise à disposition,
 - le temps moyen de mise à disposition (à titre indicatif),
 - le nombre minimum de demande d'accès par jour,
 - le nombre maximum de demande d'accès par jour,
 - le nombre moyen de demande d'accès par jour (à titre indicatif),
 - la disponibilité de service : plage horaire de disponibilité du service (par exemple, 24H/24H et 7J/7J).

Groupe G3 : modèle d'information

- Utilisation de standards:
 - Empaquetage via METS :
 - Métadonnées descriptives : DC
 - Métadonnées techniques : dépendent de la filière (MIX, textMD, ...)
 - Métadonnées de préservation : PREMIS
 - Métadonnées de droits : ODRL



Les « niveaux de granularité »

Ensemble

Périodique, Monographie en plusieurs volumes (MPV),
doc. multimédia multi-support (MMS)

Groupe d'objets

Volume, fascicule de périodique, K7, disque, disque vinyle

Objet

Page, fichier HTML, face d'un disque

Fichier

Vue de détail (fichier image),
Fichier texte, fichier vidéo,
Fichier audio,
Image (d'une planche de diapo, d'un fichier html)

Les liens

- **Les métadonnées internes = fortement liées**
 - *Descriptive*
 - *Droit*
- **Les métadonnées externes = faiblement liées**
 - *Informations de représentation*
 - *Les collections*

Le rôle des métadonnées



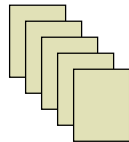
Un **livre numérisé**, c'est une suite de fichiers sans lien entre eux

+



Identifiant unique

+



Métadonnées de structure

- rattacher les fichiers entre eux
- reconstituer la structure du document

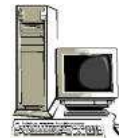
+



Métadonnées descriptives

- rattacher le document à l'original
- donner accès à la copie numérisée

+



Métadonnées administratives

- gérer la collection (historique du fichier numérique)
- gérer les droits d'accès
- préserver les informations techniques pour garantir l'accès



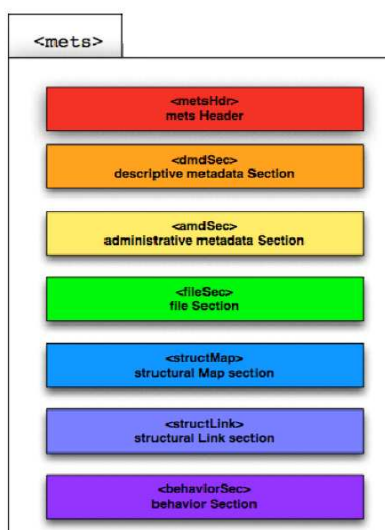
METS, un format de type « capsule »



- Un format XML de type « capsule »
 - Organise les différents blocs de MD servant à gérer un objet numérique
 - Permet de définir la « carte de structure » de l'objet numérique
- Version actuelle : 1.7
- Organisme de maintenance : Library of Congress
 - <http://www.loc.gov/standards/mets/>



Les sections d'un fichier METS



Les sections d'un fichier METS

7 sections :

Header : informations sur le document METS

Descriptive metadata Section : métadonnées descriptives externes ou encapsulées

Administrative metadata Section : métadonnées techniques, droit, source, provenance

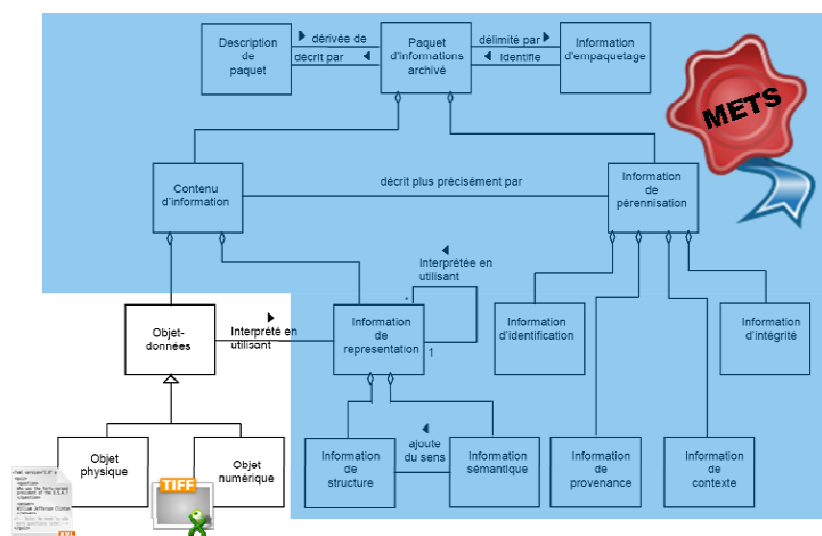
File section : liste des fichiers composant l'objet

Structural map : carte de structure, peut être logique ou physique, répétable

Structural links : liens entre différents éléments de la carte de structure

Behavior : association d'exécutables

METS et le modèle d'information OAIS



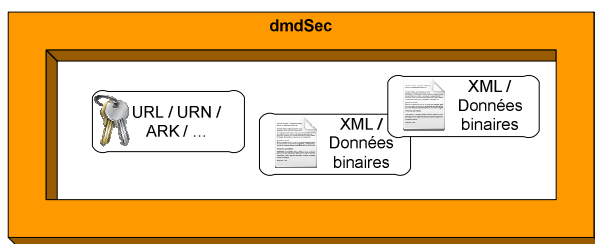


Les sections d'un fichier METS

```
<mets>
  <metsHdr></metsHdr>
  <dmdSec> </dmdSec>
  <dmdSec> </dmdSec>
  <amdSec></amdSec>
    <techMD> </techMD>
    <techMD> </techMD>
    <rightsMD></rightsMD>
  <fileSec></fileSec>
  <structMap></structMap>
  <structLink></structLink>
  <behaviorSec></behaviorSec>
</mets>
```

Les sections <fileSec> et <structMap> sont obligatoires, les autres sont facultatives

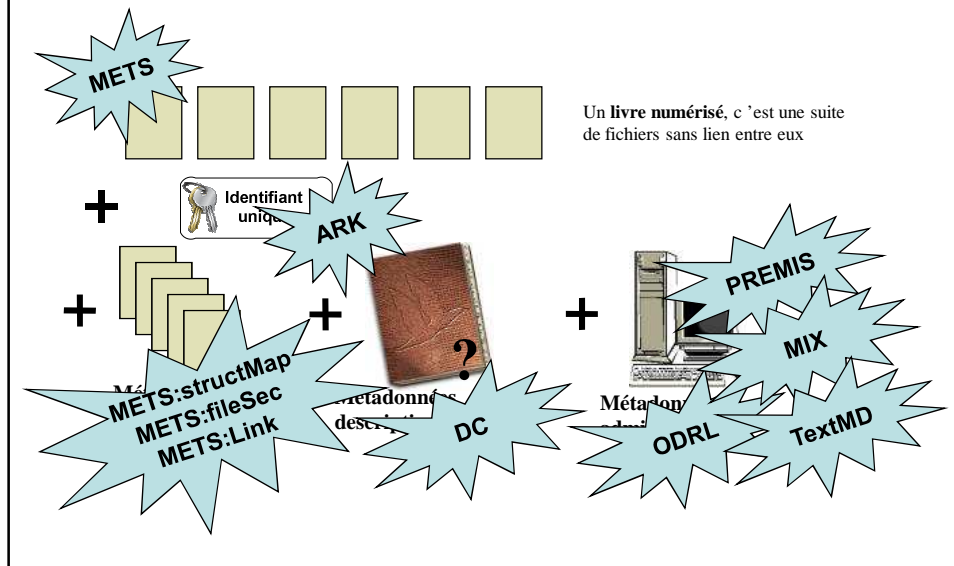
Un format enveloppe



```
<dmdSec ID="DMD.0001">
  <mdRef
    LOCTYPE="ARK"
    xlink:type="simple"
    xlink:href="ark:/12148/bc343631419"
    MDTYPE="DC"
    MIMETYPE="text/xml"/>
</dmdSec>
```

```
<dmdSec ID="DMD.0002">
  <mdWrap MIMETYPE="text/xml"
    MDTYPE="DC">
    <xmlData>
      <dc:dc>
        <dc:title>La Croix</dc:title>
        <dc:date>1883/12/12</dc:date>
        <dc:type>PERIODIQUE</dc:type>
      </dc:dc>
    </xmlData>
  </mdWrap>
</dmdSec>
```

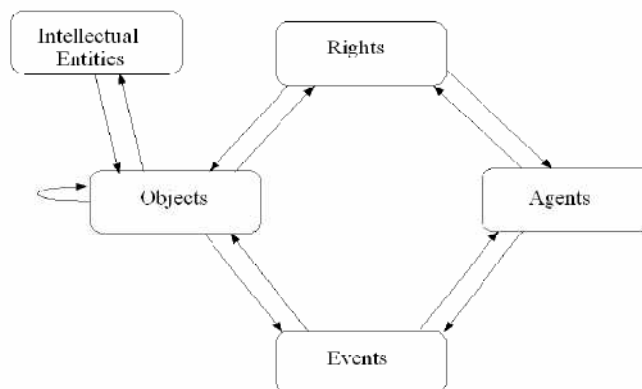
Les dictionnaire de données



Les dictionnaire de données

- **PREMIS**, métadonnées de préservation
- **Dublin Core**, métadonnées descriptives
- **MIX**, métadonnées techniques pour les fichiers image
- **TextMD**, métadonnées techniques pour les fichiers texte...
- **ODRL**, Métadonnées de gestion de droits

- **Preservation Metadata Implementation Strategies** (OCLC/RLG, Bibliothèque du Congrès)
 - Dictionnaire de données comportant plus de 120 éléments de métadonnées, utilisable pour implémenter le modèle d'information de l'OAIS
 - Version 2.0 depuis mars 2008
 - Organisme de maintenance : Library of Congress
 - <http://www.loc.gov/standards/premis/>



Les entités du modèle de données PREMIS

- **Intellectual Entity** : ensemble de contenu considéré comme une seule unité intellectuelle à gérer ou à décrire, par exemple, un livre particulier, une carte, une photo, une base de données.
- **Object (or Digital Object)** : unité discrète d'information sous forme numérique
- **Event** : une action qui implique au moins un Object ou un Agent.
- **Agent** : personne, organisation ou programme logiciel.
- **Rights** : déclaration d'une ou de plusieurs permissions portant sur un Object et ou un Agent.

METS dans SPAR

