



# Normes et archives

Bruno Bachimont, Université de technologie de Compiègne, France

MoReQ, OAIS

# Besoins divers

- Contenus culturels:
  - Assurer leur lisibilité culturelle et technique a au cours du temps sans avoir toujours des cadres réglementaires
    - Cas des archives « historiques et patrimoniales »
    - Plutôt approche « OAIS »
  
- Contenus administratifs:
  - Assurer leur disponibilité comme source de preuve
    - Cas des archives légales, réglementaires, tout ce qui permet de prouver ses droits.
    - Plutôt approche « MoReq »

# Pour nous...

- Gérer les contenus culturels:
  - OAS
  
- Gérer les données administratives les concernant (dossiers de production, contrats, etc;)
  - MoReq

# MOREQ



# Historique

- Norme pour l'archivage électronique:
  - Model Requirements for the Management of Electronic Records
  - Modèle d'exigences pour l'organisation de l'archivage électronique
- 2001 : MoReq
- 2004 : traduction française
- 2008 : MoReq2

# Principes

- Finalité :
  - Conserver les données et informations dont on aura besoin dans l'avenir.
  - S'intéresse aux records: documents, objets d'information dans l'exercice de ses activités, ou de ses obligations légales.
  - Correspond au records management
- Moyen :
  - Système d'archivage
- La finalité est la conservation, le moyen est l'archivage, l'enjeu est la réutilisation.

# Notion de base: record

- Un record est un document :
  - Validé
  - Figé (non modifiable)
  - Sélectionné (critères de risque, disponibilité dans le temps)
  - Qualifié (documenté).
  
- Approche globale
  - Le support est secondaire, ce qui compte est l'information
  - Pas de restriction sur la forme et la granularité.
  
- Archives historiques et patrimoniales sont hors du champ de record management.

# Records management / archivage

## ➤ Définition :

➤ « le records management est une fonction d'organisation, de gestion et de conservation de l'ensemble des documents/ objets d'information produits ou reçus par une personne ou un organisme dans l'exercice de ses activités ou de ses obligations légales ».

## ➤ Objectif :

➤ « permettre à une entreprise ou un organisme de disposer à tout instant et dans des conditions optimales de l'information ou du document dont il a besoin pour conduire ses activités, répondre aux contraintes de l'environnement réglementaire, et construire ou protéger une mémoire d'entreprise ou institution ».



# Périmètre

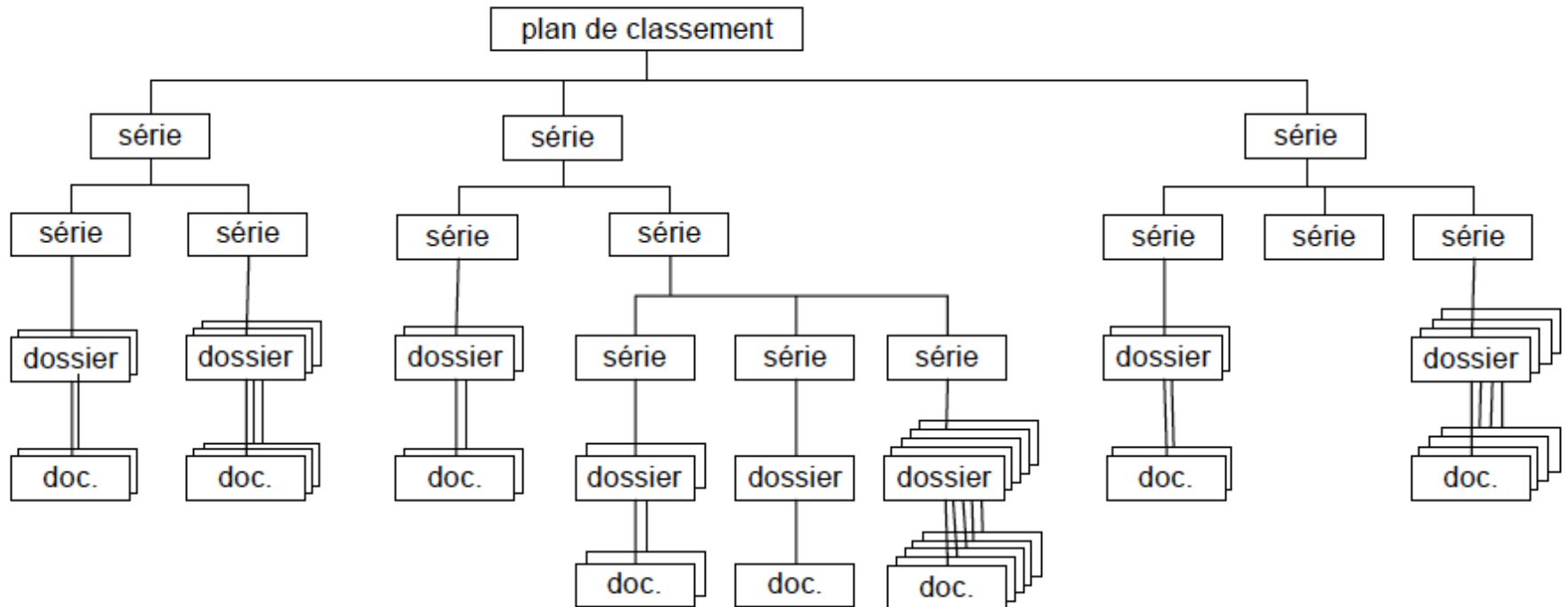
- Documents/objets d'information ayant valeur de preuve, de témoignage ou de traçabilité, ou valeur documentaire dans le temps, que l'organisme aura décidé de préserver au titre de :
  - Obligation légale et réglementaire
  - Moyen de défense en cas de contentieux
  - Preuve de propriété intellectuelle
  - Preuve de savoir et savoir-faire
  - Document pour reprise d'activité en cas de sinistre
  - Élément de construction et de maintenance d'une mémoire historique
  - Document à contenu réexploitable.

- MoReq s'adresse:
  - aux utilisateurs d'un système d'archivage électronique (SAE) pour rédiger leur cahier des charges ou auditer un système existant
  - aux organismes de formation sur l'archivage éditeurs
  - aux prestataires de service en archivage
  - aux utilisateurs des prestations de service
  - aux auditeurs de solutions d'archivage

# Plan de classement

- Classement
  - (dans l'archivage). Identification et ordonnancement systématiques des activités métiers et/ou des documents archivables en catégories selon une structure logique prédéfinie, des méthodes et des règles concrétisées dans un plan de classement (cf ISO 15489)
- Plan de classement
  - (dans MoReq2). Organisation hiérarchique des séries, dossiers, sous-dossiers, volumes et documents
  - un document est rattaché prioritairement à un point du plan de classement mais peut faire l'objet de liens (copies dans d'autres dossiers)
- Code de classement
  - Identifiant attribué à chaque série du plan de classement. Au sein de chaque série, les codes de classement des séries filles sont uniques.

# MoReQ 2



Légende  documents archivés

class = série  
file = dossier

records = documents archivés

# La capture

- Capture :
  - Processus d'enregistrement, de classement, d'ajout de métadonnées et de stockage d'un document dans un système d'archivage.
  
- La capture comporte 4 opérations:
  - enregistrement (attribution d'un identifiant unique dans le système)
  - classement
  - ajout de métadonnées
  - stockage
  
- Modalités
  - Automatique / semi-automatique / Manuelle.
  - Séparer conservation des contenus / conservation des formats de présentation.
  - Gérer la qualité : intégrité / métadonnée / formats / lisibilité

# Règles de conservation / destruction

- Règles associées aux séries ou aux dossiers:
  - durée de conservation (en années) avec le point de départ
  - le devenir de ces documents à l'issue de ce laps de temps (sort final)
- Toute entité archivée (capturée) doit être reliée à au moins une règle de conservation/destruction
- Plusieurs niveaux d'application de la destruction
  - automatique, contrôle, conservation ou non des métadonnées

# Exigences techniques

- Traitement de tous types de documents/données:
  - non structurés mais aussi structurés,
  - tous formats et tous supports,
  - documents hétérogènes ou composites (page web..)
- Automatisation maximale de la capture (des documents et des métadonnées)
- Destruction et export
- Contrôles et alertes
  - en cas de documents non conformes par rapport aux exigences de qualité définies
  - échéance de durées de conservation

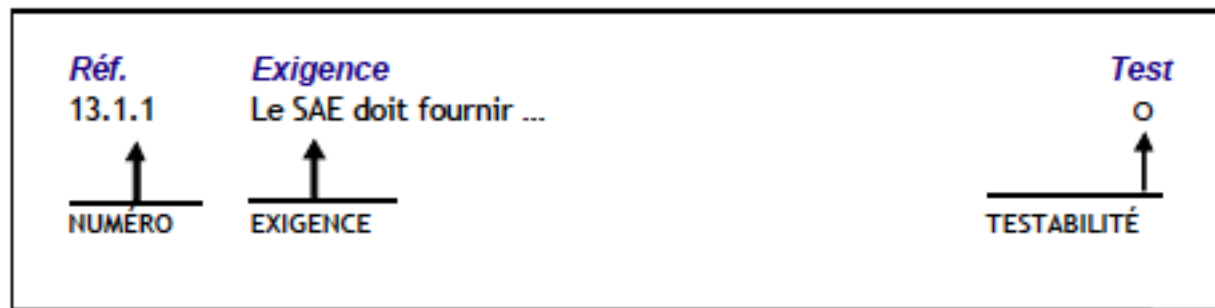
# Sécurité

- Contrôles d'accès
  - un administrateur du SAE
  - des profils utilisateurs
  - des indices de sécurité pour les documents
  
- Traçabilité
  - un historique de tous les événements qui modifient ou affectent les documents archivés (audit trail)
  - tous les changements de localisation sont tracés
  
- Sauvegardes et restauration
  
- Archives vitales
  - cf plan de continuité d'activité



# Tests de conformité

- Batterie de tests logiciels permettant de vérifier la conformité des logiciels avec les exigences fonctionnelles de la norme.
- Destinés
  - aux éditeurs
  - aux utilisateurs



OAIS



# Open Archival Information System (OAIS)

- Open
  - Les standards du modèle de référence sont développés selon une procédure publique et sont librement accessible.
  
- Information
  - Tout type de connaissance qui peut être échangé
  - Indépendant de la forme sous laquelle l'information est représentée
  - Les données sont les formes de représentation des informations.
  
- Archival Information System
  - Hardware, software, et l'organisation qui est en charge de l'acquisition, préservation et distribution de l'information.

# Contenu de la norme

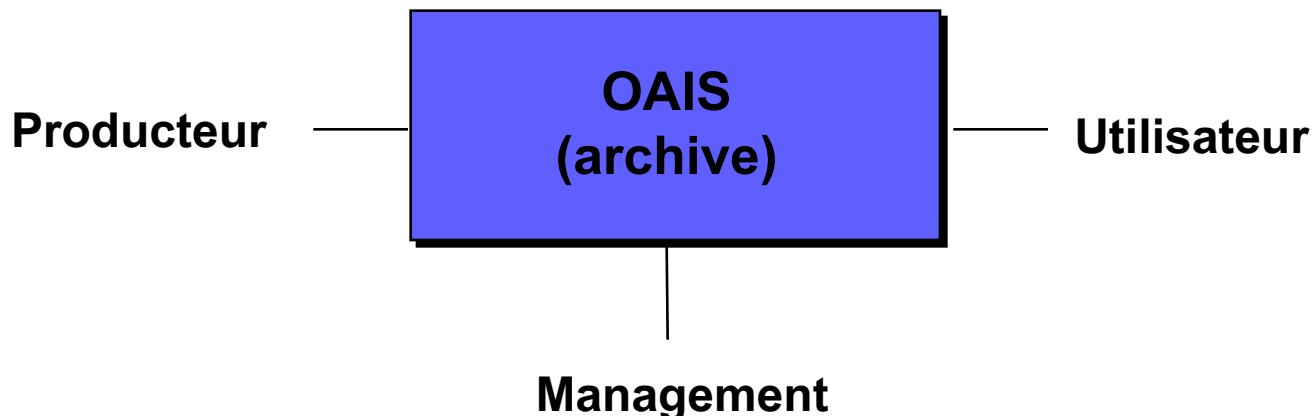
- Introduction
  - But et périmètre, applicabilité, motivations, perspectives, définitions.
- Concepts d'OAIS et Responsabilités
  - Vue de haut niveau sur les modèles fonctionnels et informationnels.
  - L'environnement externe d'OAIS
  - Le minimum pour devenir conforme à OAIS.
- Modèles détaillés
  - Modèles informationnels et fonctionnels mis en perspective.
- Perspectives pour la préservation
  - Migration des média, compression, conversions de format, et préservation des services d'accès.
- Interopérabilité des archives
  - Critères pour distinguer les types de coopération possibles entre les archives.
- Annexes
  - Scénarios sur des archives existantes, compatibilité avec d'autres normes et standards.

# But, périmètre, application

- Cadre pour comprendre et appliquer les concepts nécessaires à la préservation longue durée de contenus numériques.
  - La longue durée est longue dans la mesure où on est concerné par l'évolution technologique
  - Cadre de départ pour aborder l'information non numérique.
- Propose un cadre minimal de responsabilités pour distinguer une archive OAIS des autres.
- Cadre pour comparer les architectures, organisations et gestions des archives actuelles et futures.
- Base pour développer des standards et normes complémentaires.
- Traite la plupart des fonctions archivales.
- Applicable à toutes les archives et les organisations devant traiter des informations susceptibles d'être archivées.
- Ne spécifie aucune implémentation.

# Un environnement OAIS

- Le rôle de Producteur est joué par les personnes, ou les systèmes clients, qui fournissent l'information à conserver.
- Le rôle de Management est joué par les décideurs qui définissent la politique globale de l'archive OAIS, elle même partie d'une politique plus vaste.
- Le rôle d'Utilisateur est joué par les personnes, ou les systèmes clients, qui interagissent avec les services de l'OAIS pour chercher et récupérer les informations conservées qui les intéressent. La Communauté d'utilisateurs cible est une classe particulière d'Utilisateurs : c'est l'ensemble des Utilisateurs en mesure de comprendre les informations conservées.



# Responsabilités OAI

- Négocier et accepter les informations des producteurs
- Obtenir le contrôle nécessaire et suffisant pour assurer la préservation à long terme.
- Déterminer quelles communautés (les communautés désignées) doivent rester capable de comprendre et utiliser l'information conservée.
- Assurer que l'information préservée est intelligible pour les communautés désignées.
- Suivre les politiques et procédures qui assurent que l'information est préservée en fonction des contingences raisonnablement prévisibles.
- Rendre l'information préservée accessible auprès des communautés désignées dans des formes qui leur sont intelligibles.

# Modèles OAIS : principes

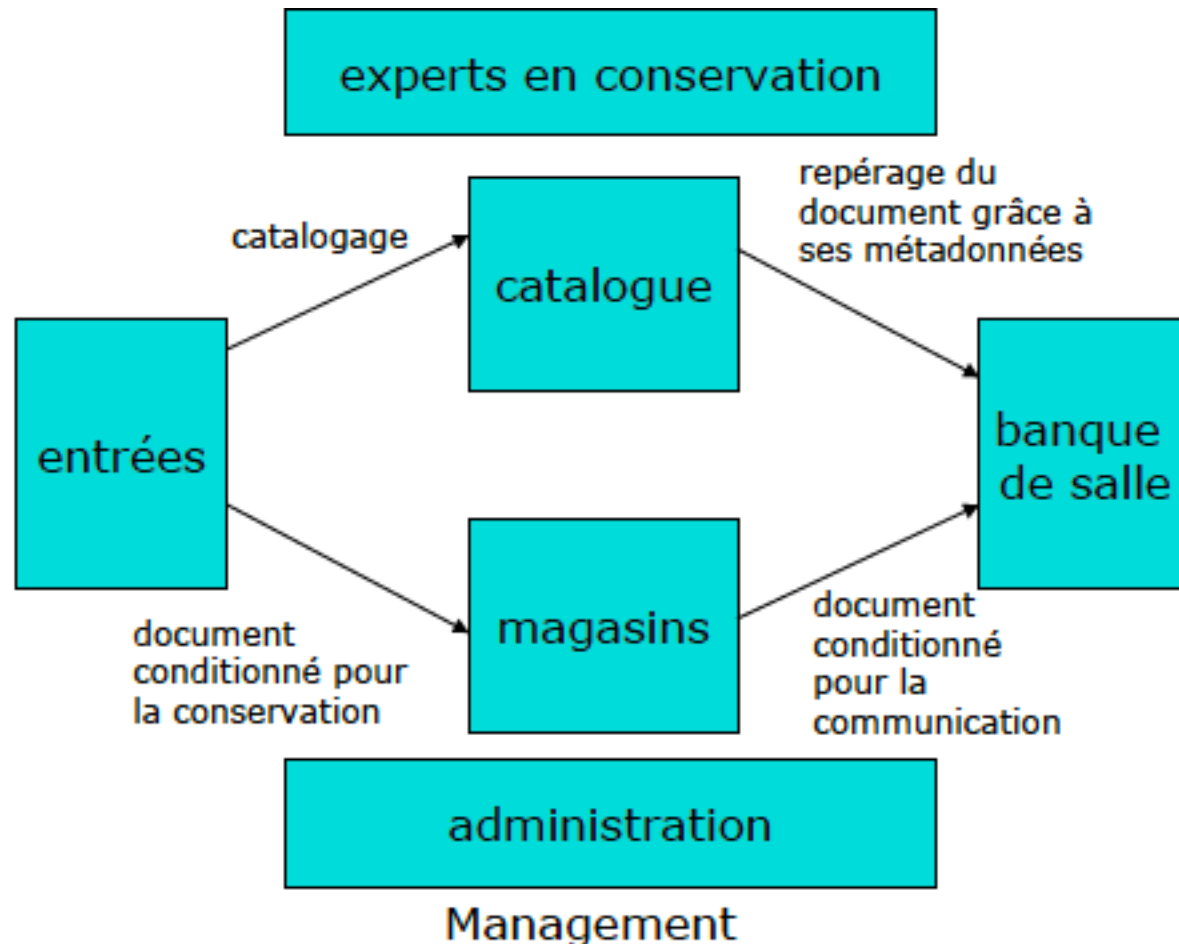
- OAIS dispose d'un modèle fonctionnel et d'un modèle informationnel.
- Les modèles doivent :
  - Utiliser des modèles pour mieux communiquer les concepts OAIS
  - Utiliser une technique bien établie de modélisation formelle
  - Rester autant que faire se peut indépendant de toute implémentation
  - Eviter des conceptions trop détaillées.



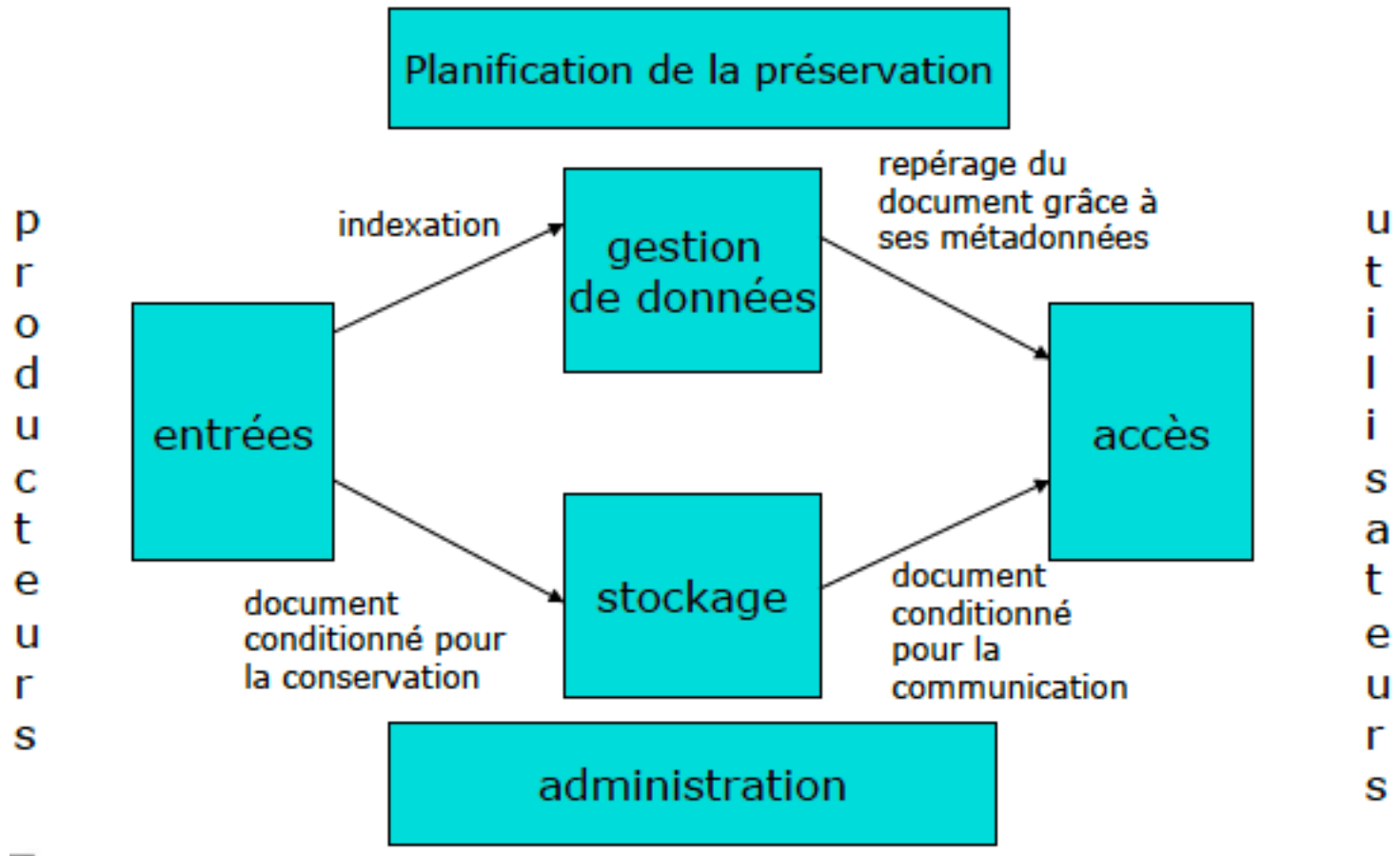
# OAIS : modèle fonctionnel



# Modèle fonctionnel : la réalité



# Et le modèle



# Paquet d'information

## ➤ Un paquet:

- ce qui est manipulé dans OAIS
  - Ce qu'on verse
  - Ce qu'on décrit
  - Ce qu'on archive
  - Ce qu'on communique



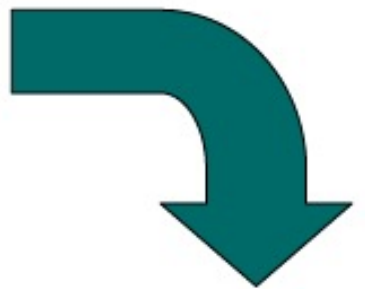
## ➤ Exemple: un périodique numérisé

- Comment définir le paquet :
  - Au niveau du titre ? Au niveau de chaque fascicule ?
    - Choix d'implémentation, non prescrit par la norme
    - Mais question qu'il faut se poser, et la norme le rappelle.

# OAIS : la notion de paquet (2)



Paquets archivés (AIP)

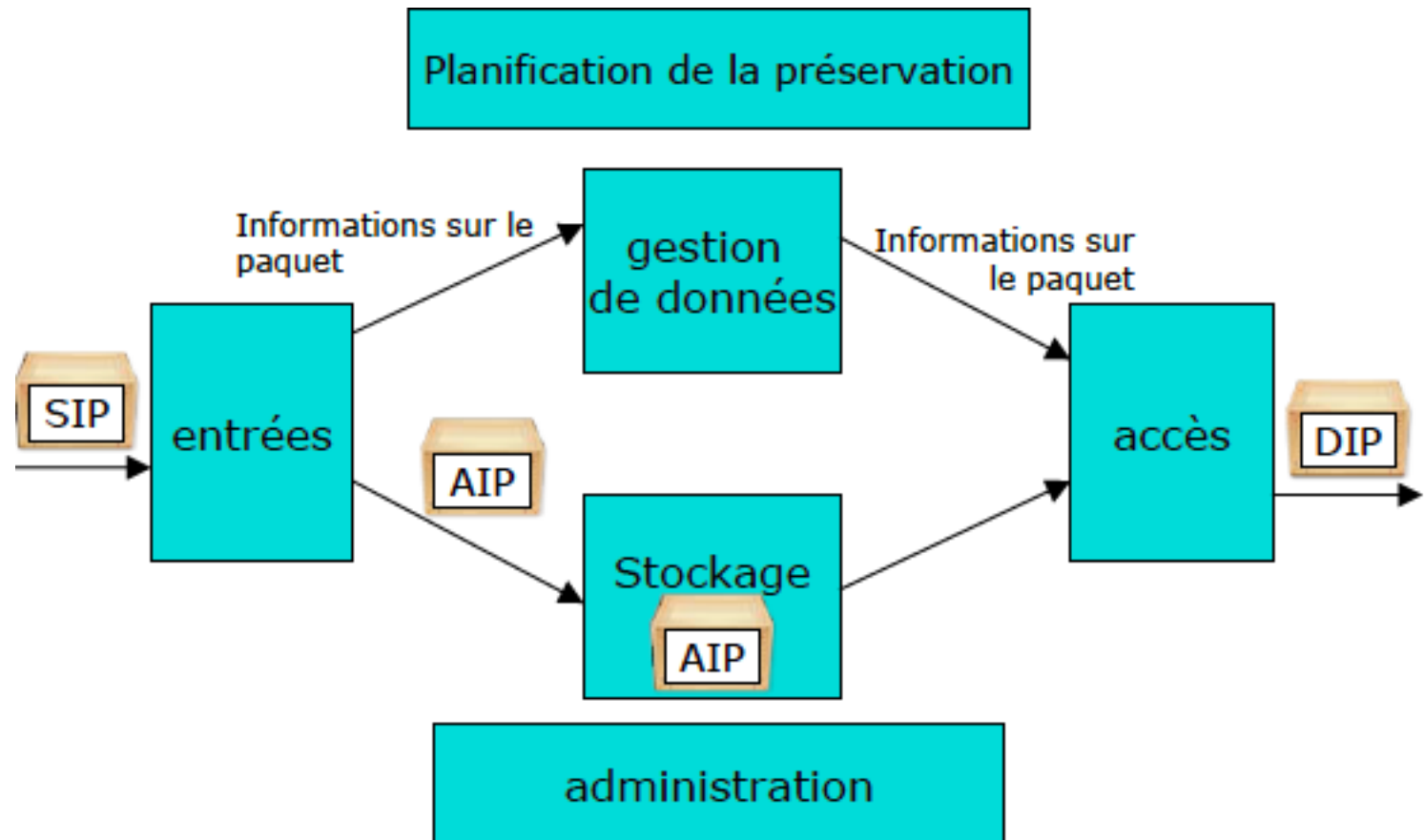


Paquet diffusé (DIP)

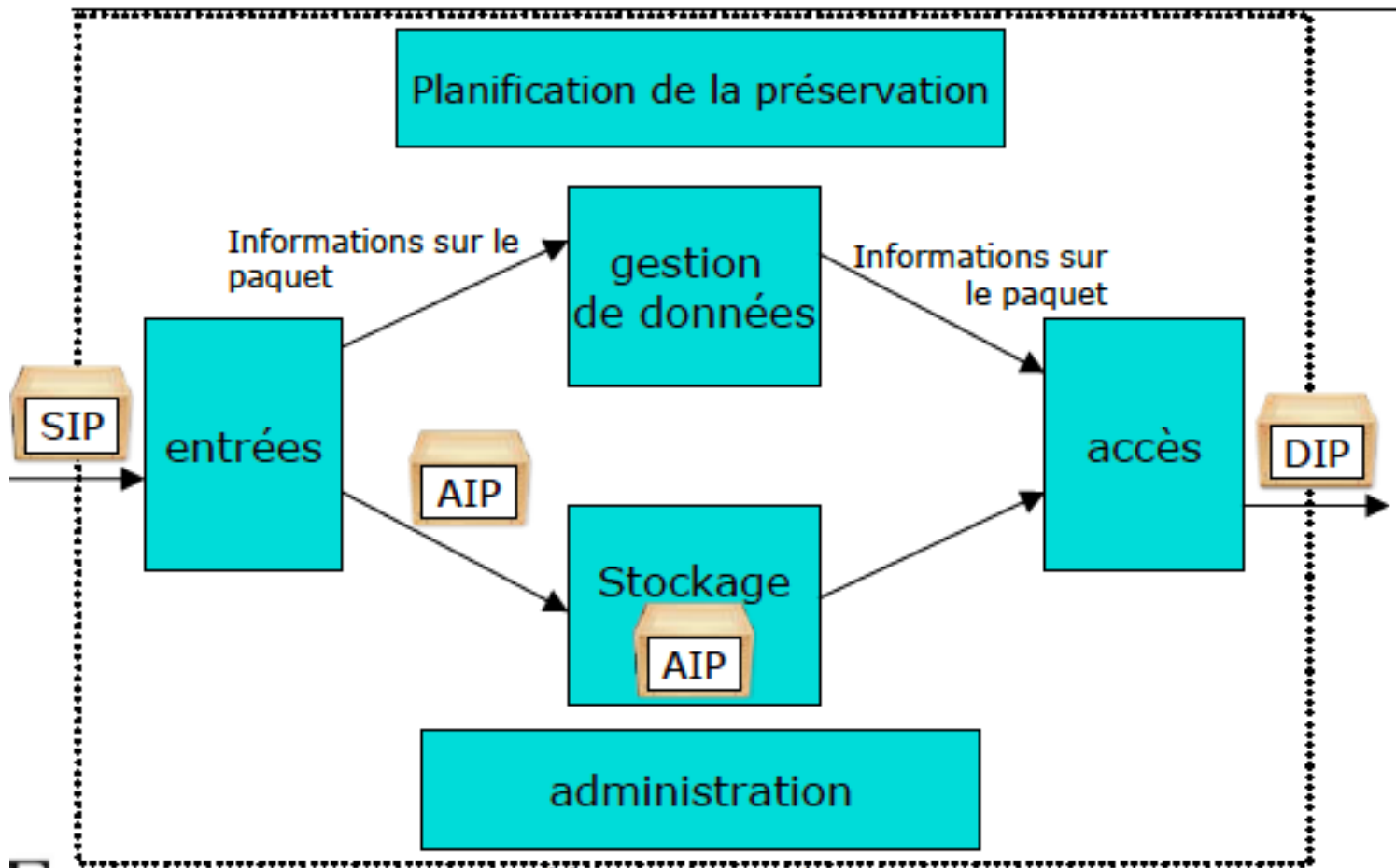


Paquets à verser (SIP)

# Paquets et modèle fonctionnel



# Les contours de l'archive

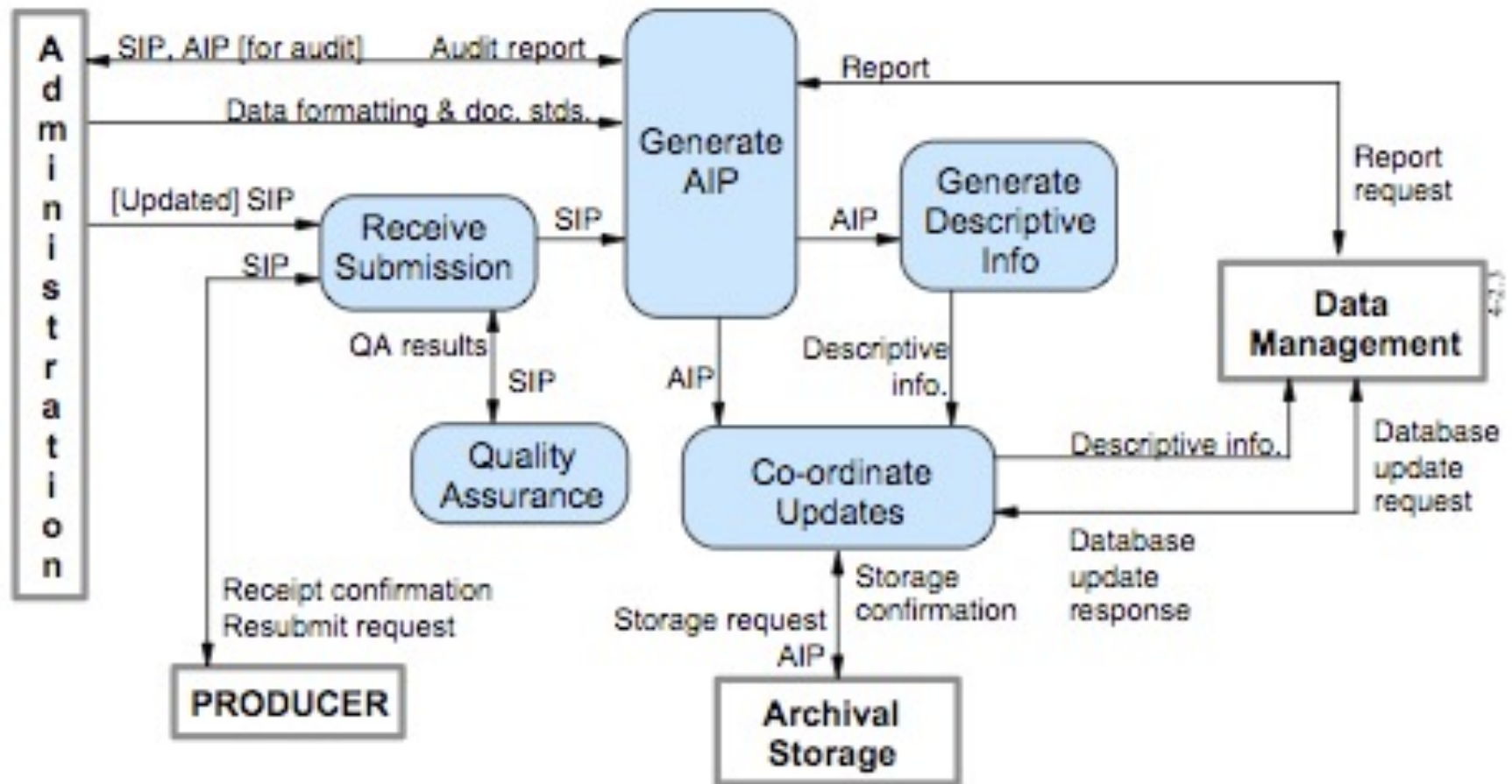


# Les entités fonctionnelles

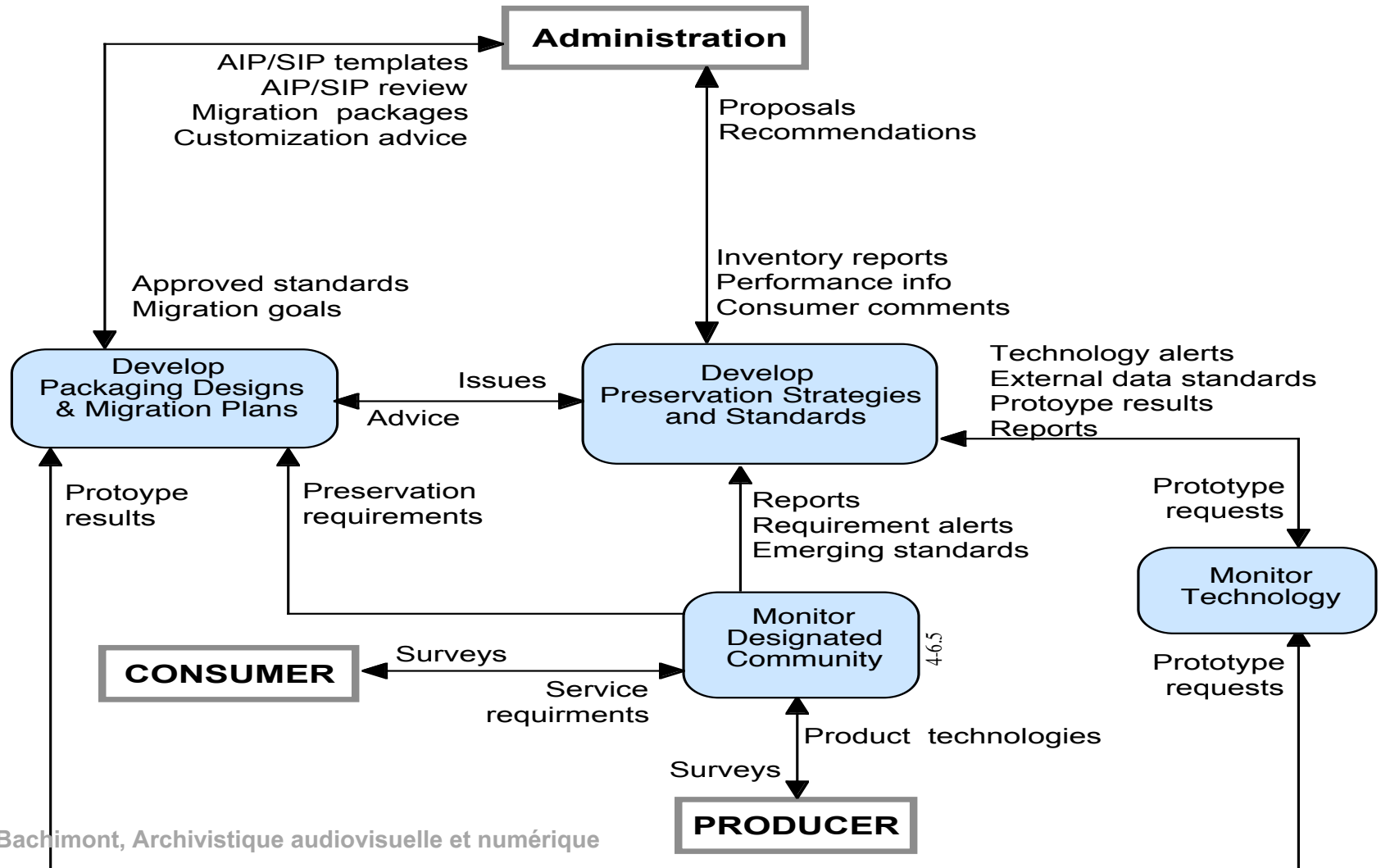
- Ingest:
  - cette entité fournit les services et fonctions nécessaires pour accepter des SIP soumis par les Producteurs, et préparer les contenus pour leur archivage et gestion dans l'archive.
- Archival Storage:
  - cette entité fournit les services et fonctions pour l'archivage, maintenance et récupération des AIP.
- Data Management:
  - cette entité fournit services et fonctions pour créer, maintenir et utiliser les informations de description qui identifient et documentent les contenus archivés et les données liées à la gestion.
- Administration:
  - cette entité gère le fonctionnement global de l'archive.
- Preservation Planning:
  - cette entité pilote l'environnement de l'archive et fournit les recommandations assurant que l'information archivée reste accessible et intelligible pour la communauté désignée sur le long terme.
- Access:
  - cette entité permet aux utilisateurs de découvrir l'information archivée, la commander et l'obtenir.



# Flux de données : ingestion



# Planning de Préservation

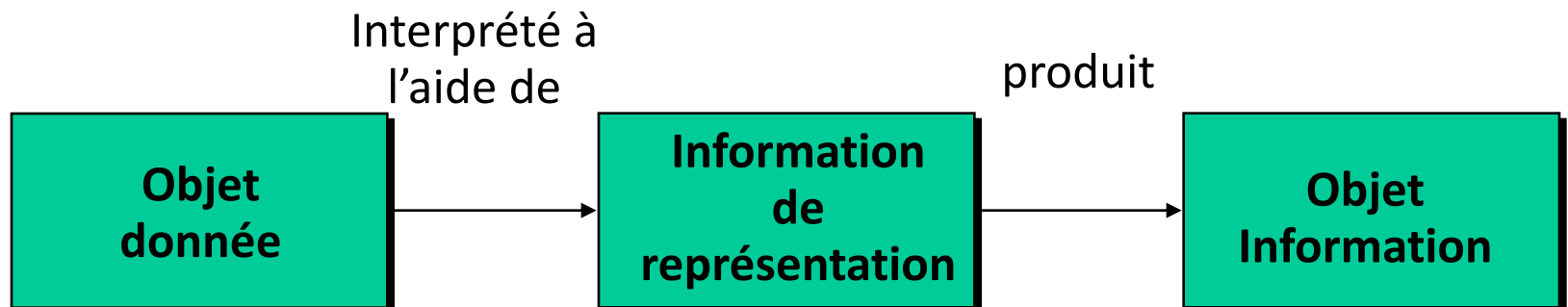


# OAIS: modèle d'information



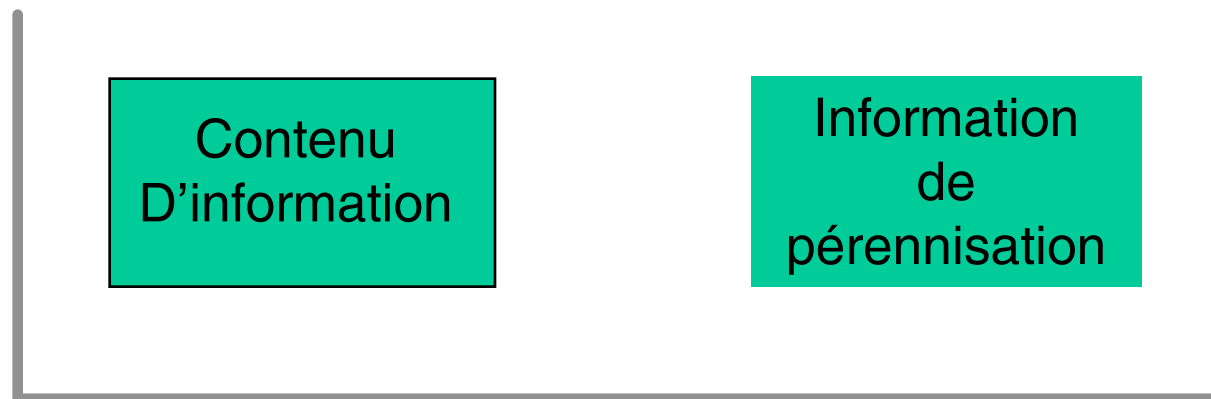
# L'information O AIS

- L'information est toujours exprimée / représentée par des types de données.
- Les données, interprétées en utilisant l'information de représentation, fournissent l'information.
- La conservation de l'objet information nécessite une identification claire et une bonne compréhension de l'objet donnée et de l'information de représentation associée.



# Information Package

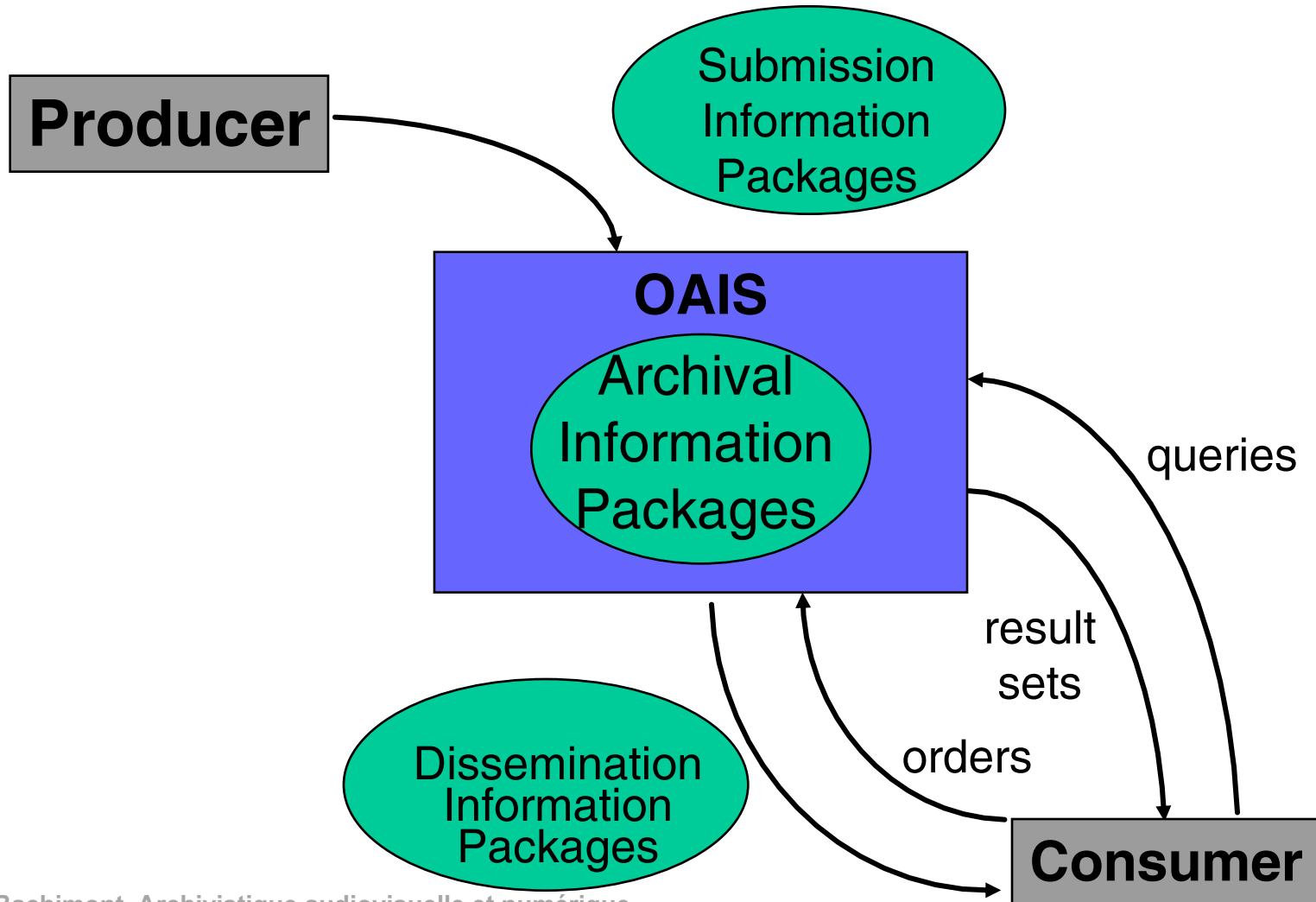
- Un paquet d'information est un conteneur conceptuel rassemblant deux types d'information:
  - Le contenu d'information (content information)
  - L'information de pérennisation (Preservation Description Information - PDI)



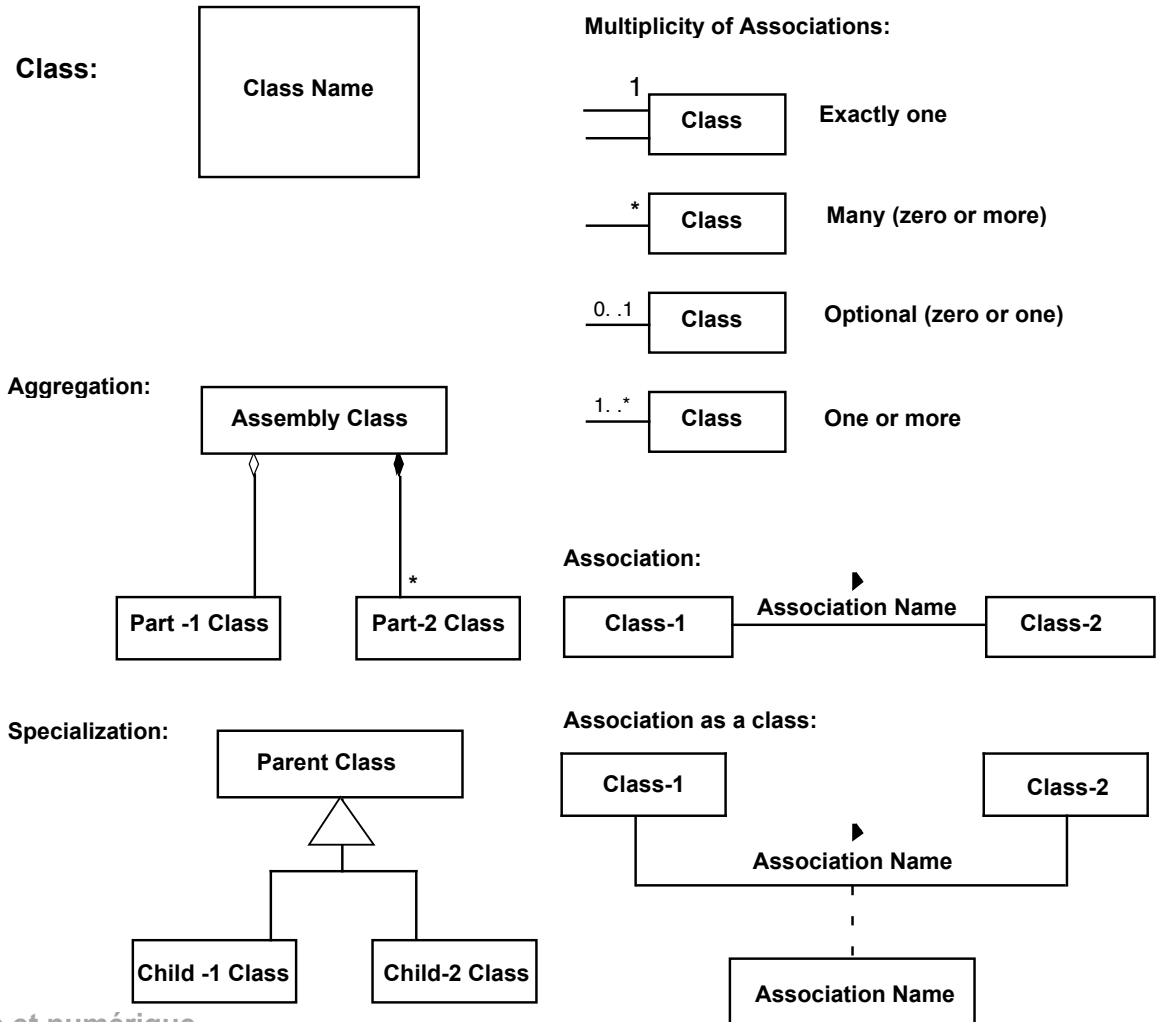
# 3 types d'Information Package

- Submission Information Package : SIP
  - Négocié par le Producteur et l'archive OAIS
  - Envoyée par le producteur à l'archive.
  
- Archival Information Package: AIP
  - C'est le paquet d'information utilisé pour la préservation;
  - Incluse l'ensemble complet des PDI pour les contenus d'information.
  
- Dissemination Information Package : DIP
  - Mobilise tout ou partie des AIP;
  - Envoyé aux utilisateurs par l'archive OAIS.

# Vue des échanges de données

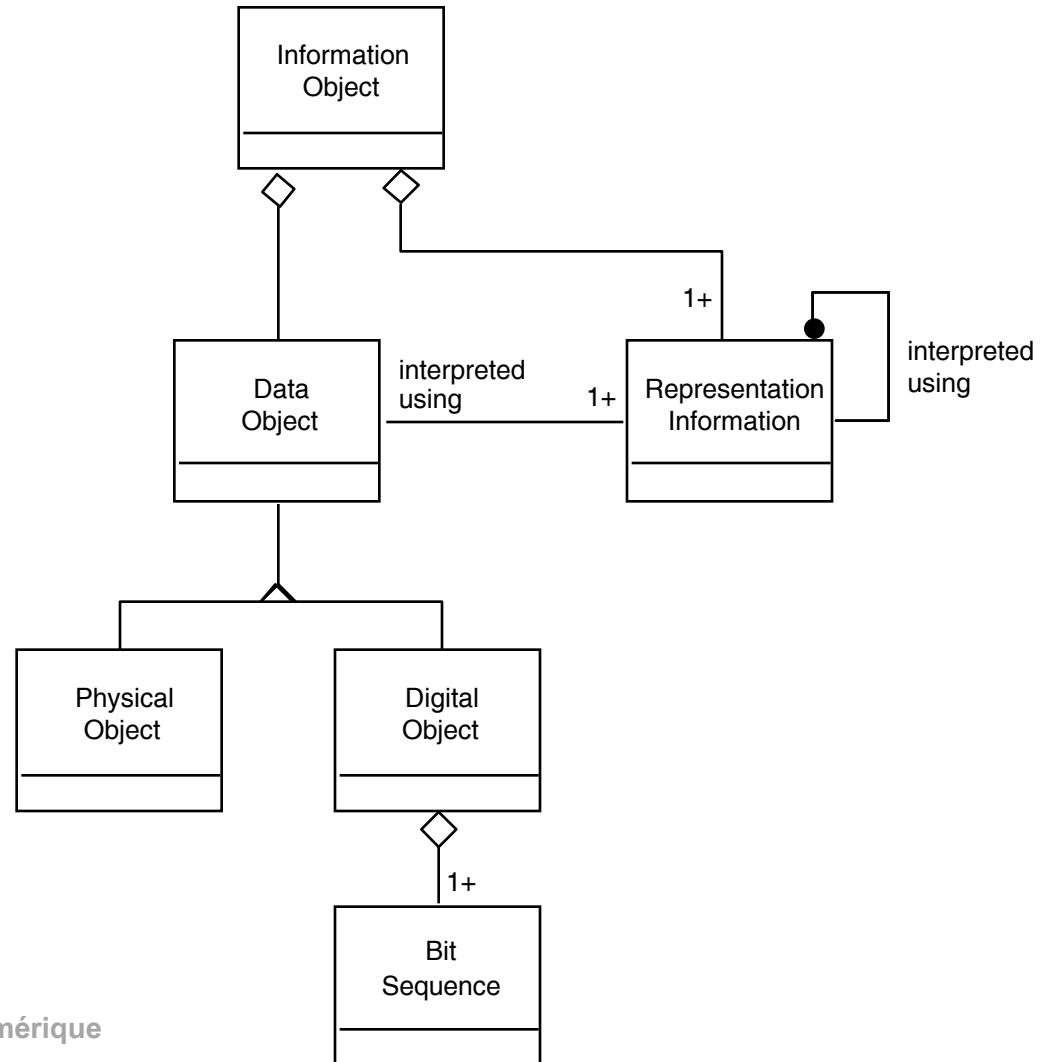


# UML Notation





# Information Objects



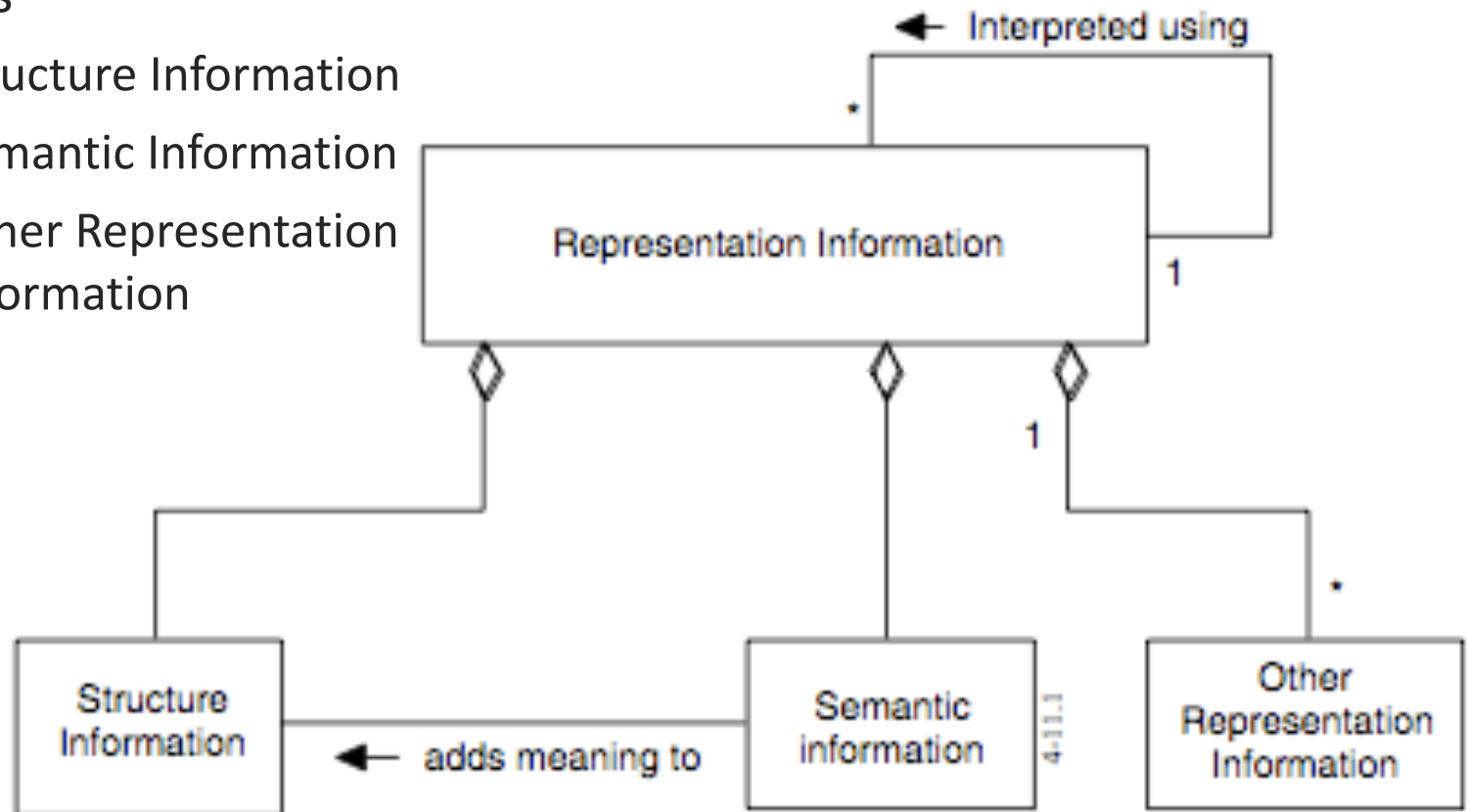
# Representation Information

- L'information de représentation qui accompagne un objet physique (une roche lunaire) peut lui donner une signification supplémentaire :
  - Typiquement, le résultat d'une analyse des propriétés observables de la roche.
  
- L'information de représentation accompagnant un objet numérique est utilisé pour donner une signification supplémentaire:
  - Typiquement, faire correspondre les bits à des données comme des caractères, des entiers ou réel, et à des groupes de données de ce type.
  - L'information de représentation peut associer les objets avec des connaissances plus abstraites qui peuvent avoir des relations complexes entre elles, qu'il faut alors décrire.

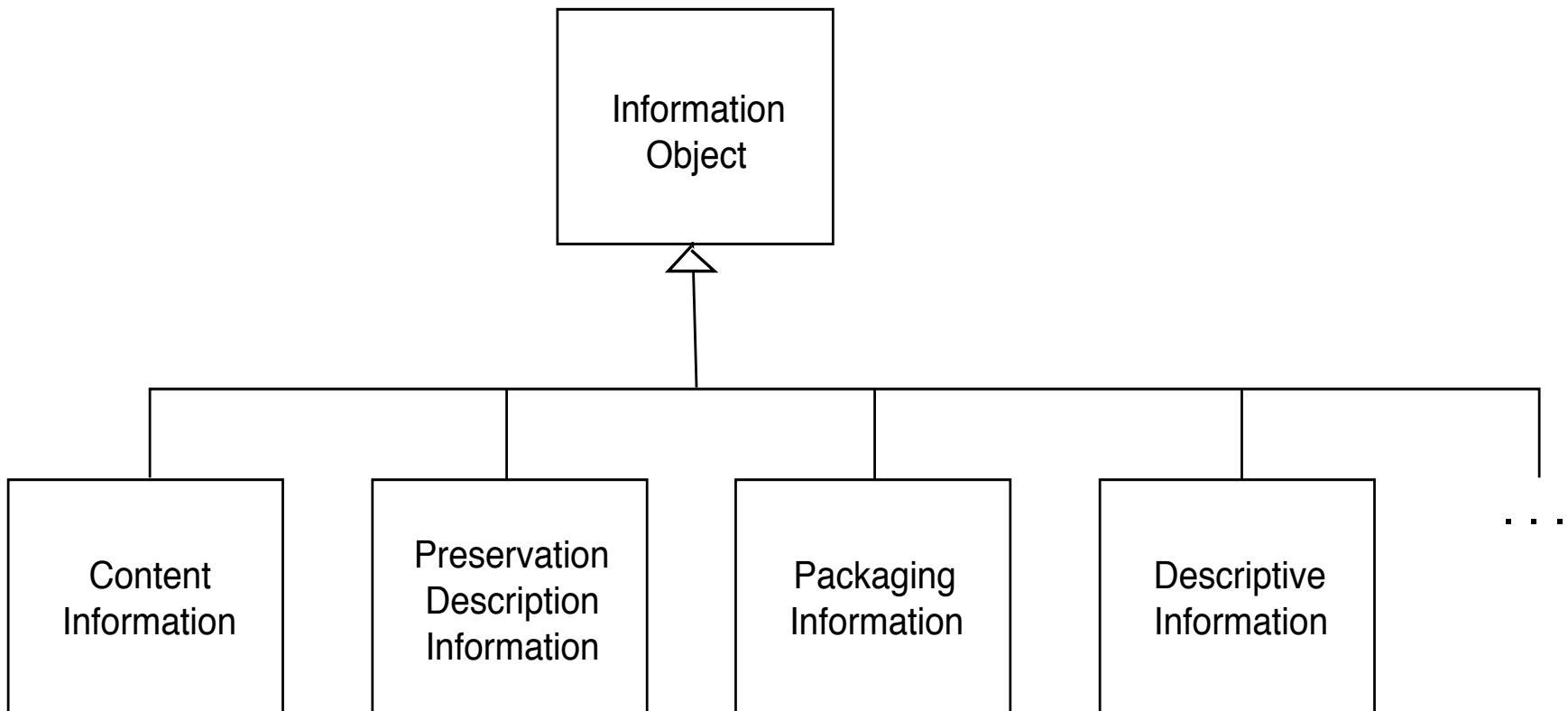
# La nature réursive des Representation Information

## ➤ 3 types

- Structure Information
- Semantic Information
- Other Representation Information



# Les types d'information utilisés dans OAIS



# Content Information

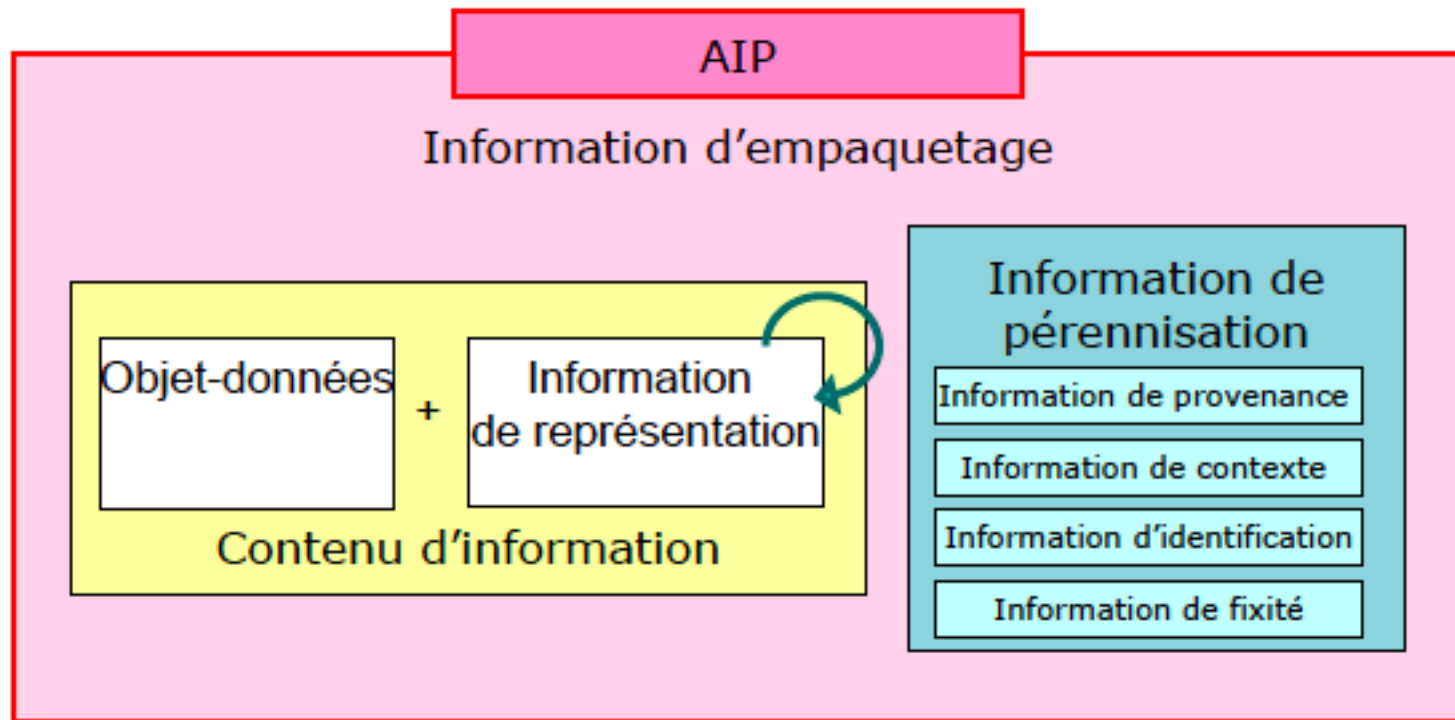
- C'est l'information qui est l'objet premier de la préservation.
- Une instance ou représentation de l'objet d'information est ce que l'archive a pour mission de préserver.
- Décider ce qu'est le contenu d'information n'est pas immédiat et peut nécessiter une discussion avec le Producteur.
- L'objet donné dans le contenu d'information peut être physique ou numérique.

# Preservation Descriptive Information

- Le contenu d'information tout seul ne suffit; je sais pas :
  - Quelle est son histoire :
    - information de provenance
  - Quelles sont ses relations avec d'autres contenus d'information
    - Information de contexte
  - Comment l'identifier
    - Information d'identification
  - S'il n'a pas été corrompu;
    - Information d'intégrité
- Ces informations constituent le PDI : l'information de pérennisation

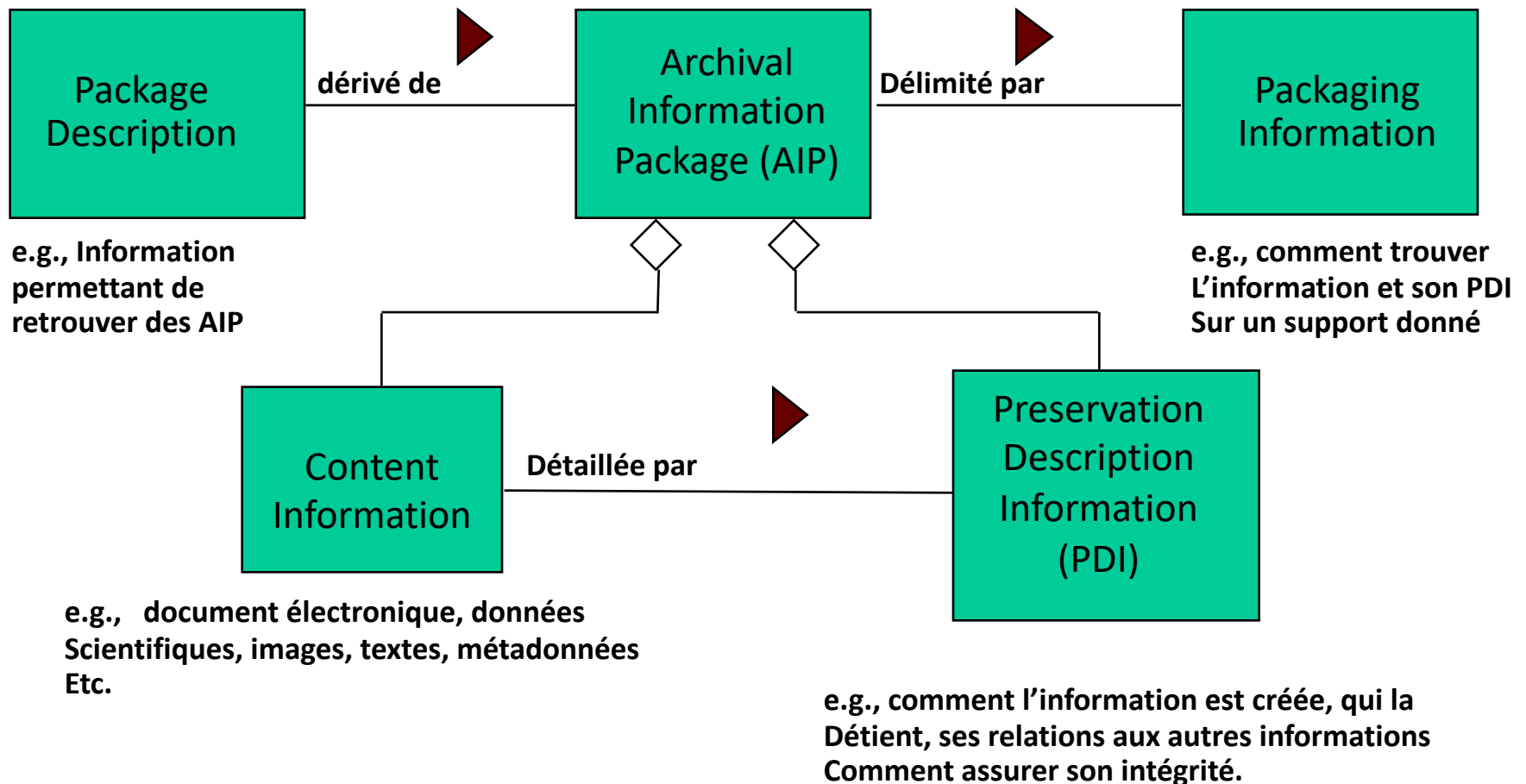
Type de Contenu d'information	Identification	Provenance	Contexte	intégrité
<b>Données scientifiques spatiales</b>	<ul style="list-style-type: none"> <li>• Identificateur de l'Objet</li> <li>• Référence du Journal</li> <li>• Mission, instrument, titre et ensemble d'attributs</li> </ul>	<ul style="list-style-type: none"> <li>• Description d'instrument</li> <li>• Historique du traitement</li> <li>• Description du capteur</li> <li>• Instrument</li> <li>• Mode Instrument</li> <li>• Carte de décommutation</li> <li>• Spécification d'Interface logiciel</li> </ul>	<ul style="list-style-type: none"> <li>• Historique de l'étalonnage</li> <li>• Jeux de données connexes</li> <li>• Mission</li> <li>• Historique du financement</li> </ul>	<ul style="list-style-type: none"> <li>• CRC</li> <li>• <i>Checksum</i></li> <li>• Codage Reed-Solomon</li> </ul>
<b>Information bibliographique</b>	<ul style="list-style-type: none"> <li>• ISBN</li> <li>• Titre</li> <li>• Auteur</li> </ul>	<ul style="list-style-type: none"> <li>• Historique de l'impression</li> <li>• Droits d'auteur</li> <li>• Position dans la série</li> <li>• Références</li> </ul>	<ul style="list-style-type: none"> <li>• Références associées</li> <li>• Système Dewey</li> <li>• Date de publication</li> <li>• Editeur</li> </ul>	<ul style="list-style-type: none"> <li>• Auteur</li> <li>• Signature électronique</li> <li>• Couverture</li> </ul>
<b>Progiciel</b>	<ul style="list-style-type: none"> <li>• Nom</li> <li>• Auteur/origine</li> <li>• Numéro de version</li> <li>• Numéro de série</li> </ul>	<ul style="list-style-type: none"> <li>• Historique de la révision</li> <li>• Titulaire de la licence</li> <li>• Enregistrement</li> <li>• Droits d'auteur</li> </ul>	<ul style="list-style-type: none"> <li>• Fichier d'aide</li> <li>• Guide utilisateur</li> <li>• Logiciel associé</li> <li>• Langage</li> </ul>	<ul style="list-style-type: none"> <li>• Certificat</li> <li>• <i>Checksum</i></li> <li>• Encryptage</li> <li>• CRC</li> </ul>

# L'information d'empaquetage



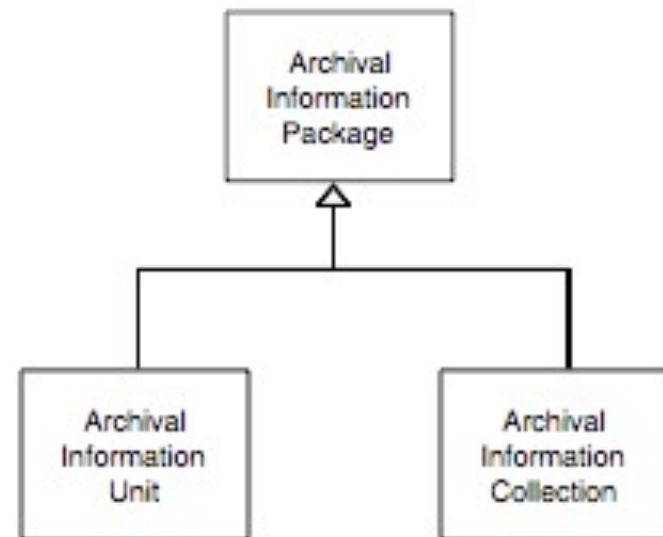


# OAIS Archival Information Package



# Types d'AIP

- Fondés sur les différentes complexités des Content Object.
- Les AIU contiennent un simple Data Object comme un Content Object.
- Les AIC contiennent plusieurs AIP et leur content object:
  - Chaque membre d'un AIC est un AIP contenant un content information et un PDI.
  - L'AIC contient un unique PDI sur la collection comme telle.



# Descriptive Information

- Contient les données nécessaires pour les outils ou documents gérant l'accès à l'archive: Access Aids
- Les Access Aids peuvent servir à localiser, analyser, retrouver ou commander de l'information à partir de l'archive OAIS.

# Package Descriptors and Access Aids

- Les descripteurs de Packets sont utiles pour permettre l'accès et la visibilité des contenus archivés.
- Les descripteurs contiennent 1 ou plus descriptions associées (Associatec descriptions AD) qui décrivent les content information des AIP du point de vue d'une unique aide à l'accès (Access aid)
- Exemples:
  - Finding Aids - aide à la découverte: assister l'utilisateur à la localisation de l'information qu'il recherche
  - Ordering Aids - aide à la commande : permet à l'utilisateur de découvrir le coût et de commander les IAU qu'il désire.
  - Retrieval Aids - aide à l'obtention : permet aux utilisateurs autorisés de récupérer des AIU décrits par le Unit Descriptor de l'archive.

# En résumé

- Un modèle d'objets d'information qui contiennent des données (data object) et des information (representation information).
- 4 types d'information utile pour l'archive: Content Information, PDI, Packaging Information et la Descriptive Information
- Ces classes sont assemblées et articulées dans un AIP pour décrire complètement une instance de Content information.
- Information nécessaire à l'accès, en plus de celle utile pour la préservation.

# OAIS : mise en œuvre



# Mettre en œuvre OAIS

- Définir la communauté désignée :
  - Pour qui préserve-t-on ?
  - Qui valide qu'on a bien fait le travail ?
- Définir les producteurs de l'information :
  - Qui ? Quels contenus ? Comment négocier ? Comment récupérer l'information ?
- Monitorer :
  - État de la communauté
  - État de la technologie
- Restituer :
  - Sous quelle forme restituer ?
  - Quels services ?

# Intérêts et limites d'OAIS

## ➤ Limites

- Indique ce qu'il faut faire, non comment le faire;
- Complexité de la mise en œuvre
- Ce n'est pas une méthodologie ni une norme technique

## ➤ Intérêts

- Cadre normatif permettant de démontrer la qualité de l'archivage effectué
- Lingua franca de la préservation: terminologie commune au niveau international, indépendamment du contexte
- Guide la réflexion sans imposer la solution qui reste liée à chaque institution / situation.



# Questionnaire OAIS

1. Dans votre cas, comment sont les SIP ? Qui sont vos fournisseurs de données ? Quels accords devez vous contractualiser avec eux ? Quel monitoring pour veiller à rester en phase avec leur évolution organisationnelle ou technologique ?
2. Quels sont les AIP que vous avez définis ? comment justifiez vous leur structuration et les choix techniques associés ?
3. Quelle est votre communauté désignée ? Y en-a-t-il plusieurs ?
4. Quel monitoring pour observer le décalage entre votre archive et la communauté désignée ?
5. Comment intégrer les apports de la communauté désignée pour documenter les contenus archivés ?
6. Quels sont les DIP que vous avez identifiés ?
7. Quelle procédure de veille technique avez vous adoptée ?
8. Comment faites vous évoluer vos AIP ?
9. Quels sont les outils d'exploration que vous avez proposés (les Finding Aids) ?
10. Structure des AIP : comment se décompose votre PDI ? Détaillez la structure du PDI et comment vous l'instancier dans votre cas.