

# l'archivage pérenne des documents numériques

Michel Auffret  
CINES  
JRES 2005

# l'archivage pérenne

ce n'est pas :

- une sauvegarde
- une migration HSM
- un stockage ultime avant oubli définitif

# l'archivage pérenne

c'est :

- conserver le document
- le rendre accessible
- en préserver l'intelligibilité
- sur le très long terme
- ... 30 ans et au-delà

# l'obsolescence technologique

- la plupart des fichiers informatiques de plus de 10 ans sont aujourd'hui illisibles
- support physique détérioré
- format de fichier inconnu
- logiciel ou matériel de lecture disparu
- et d'ailleurs... plus personne ne sait ce qu'il y a là-dedans

# contrer l'obsolescence technologique

- support physique détérioré
  - *gestion du vieillissement des médias*
  - *migration physique*
- format de fichier inconnu
  - *privilégier les formats durables*
  - *migration logique*
- logiciel ou matériel de lecture disparu
  - *veille technologique*
- et d'ailleurs... plus personne ne sait ce qu'il y a là-dedans
  - *métadonnées*

# les formats de fichiers durables (\*)

- format publié
- format largement utilisé
- format normalisé si possible
- texte : HTML, OpenDocument, PDF, RTF, TXT, XML
- image : GIF, JPEG, PNG, TIFF
- audio : MP3
- vidéo : MPEG, Quick Time

(\*) ceux d'aujourd'hui !

# les métadonnées

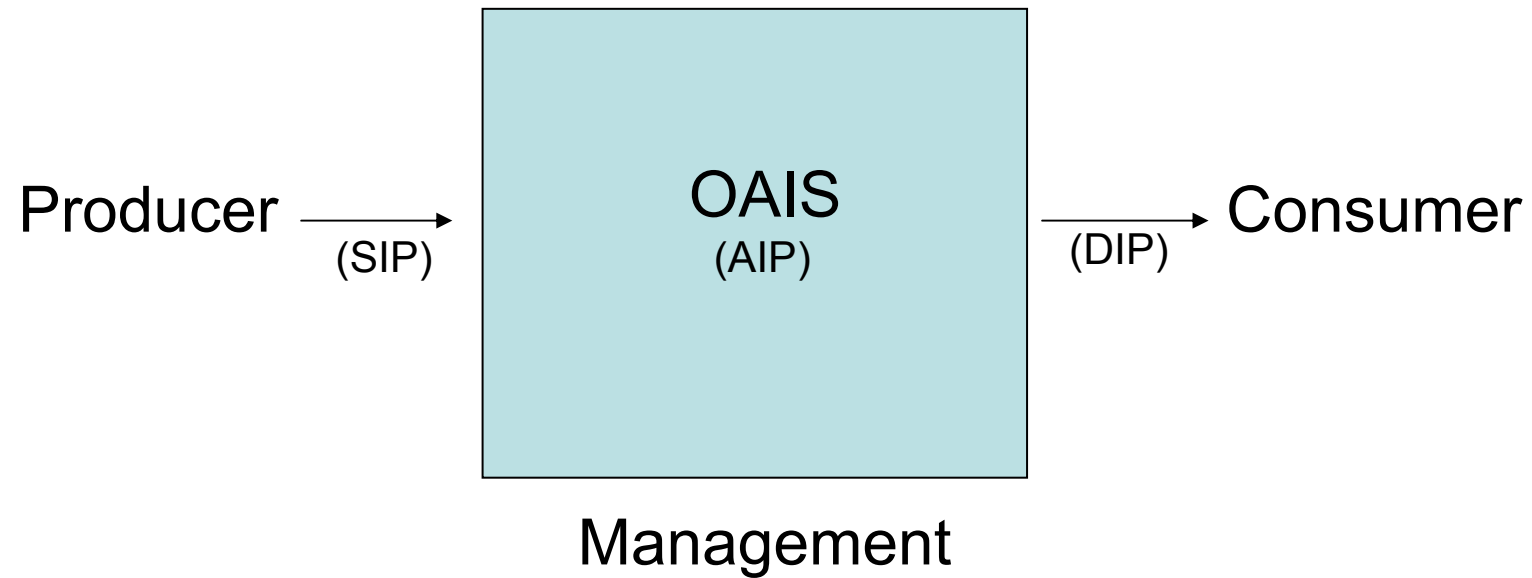
- méta-données = des données sur les données !
- donc = de l'information sur les documents
- tout ce qu'il faut savoir - et tenir à jour ! - sur un document pour pouvoir :
  - le conserver
  - le rendre accessible
  - et intelligible
  - ... sur le très long terme
- identification délicate dans la pratique

# OAIS : un modèle normalisé pour comprendre

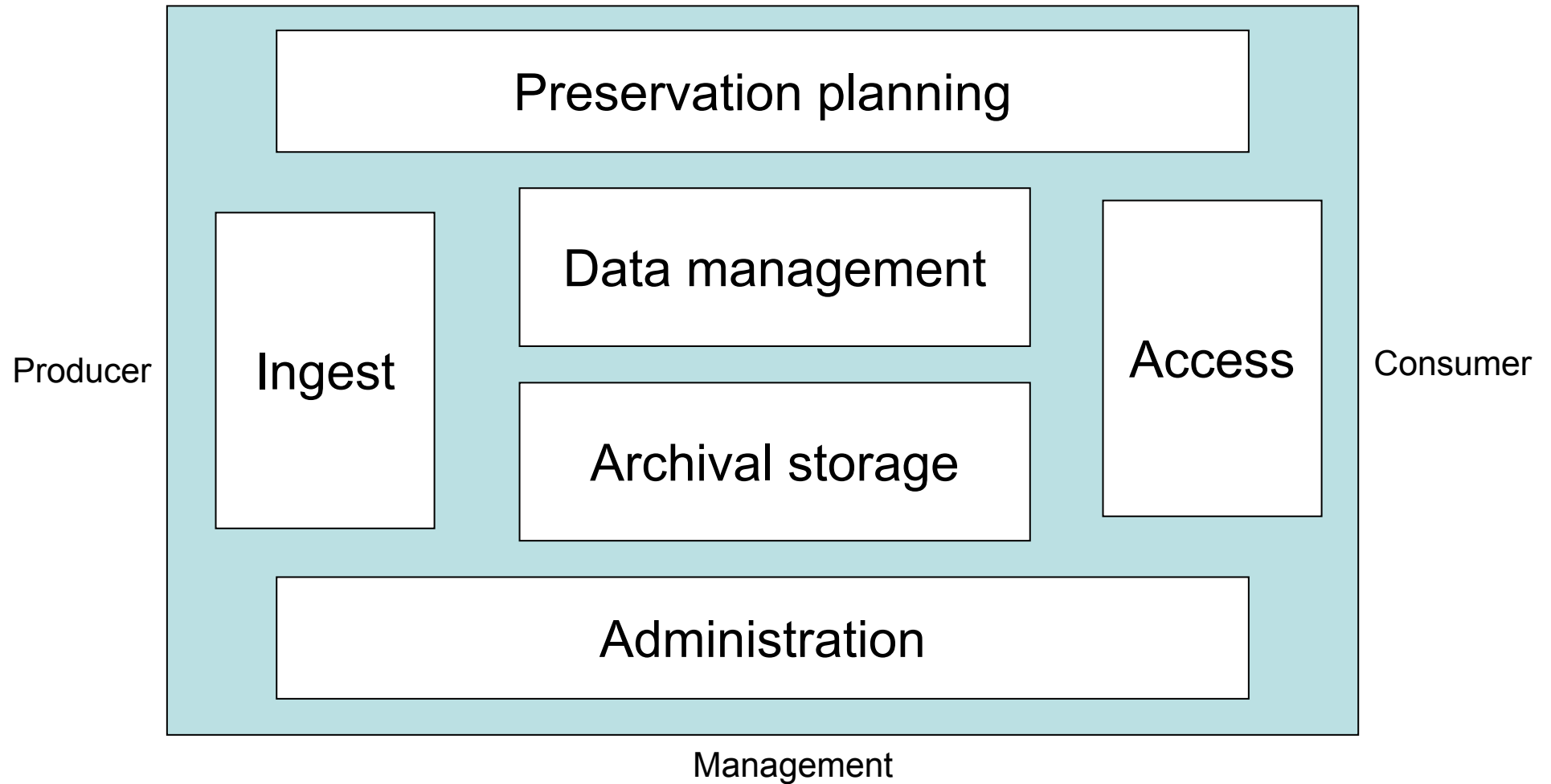
- Reference Model for an Open Archival Information System (janvier 2002)
- initiative CCSDS, rejointe par des archives et des bibliothèques institutionnelles
- modèle purement conceptuel, ne fait aucune recommandation technique
- norme ISO 14721
- traduction française disponible



# OAIS : les acteurs

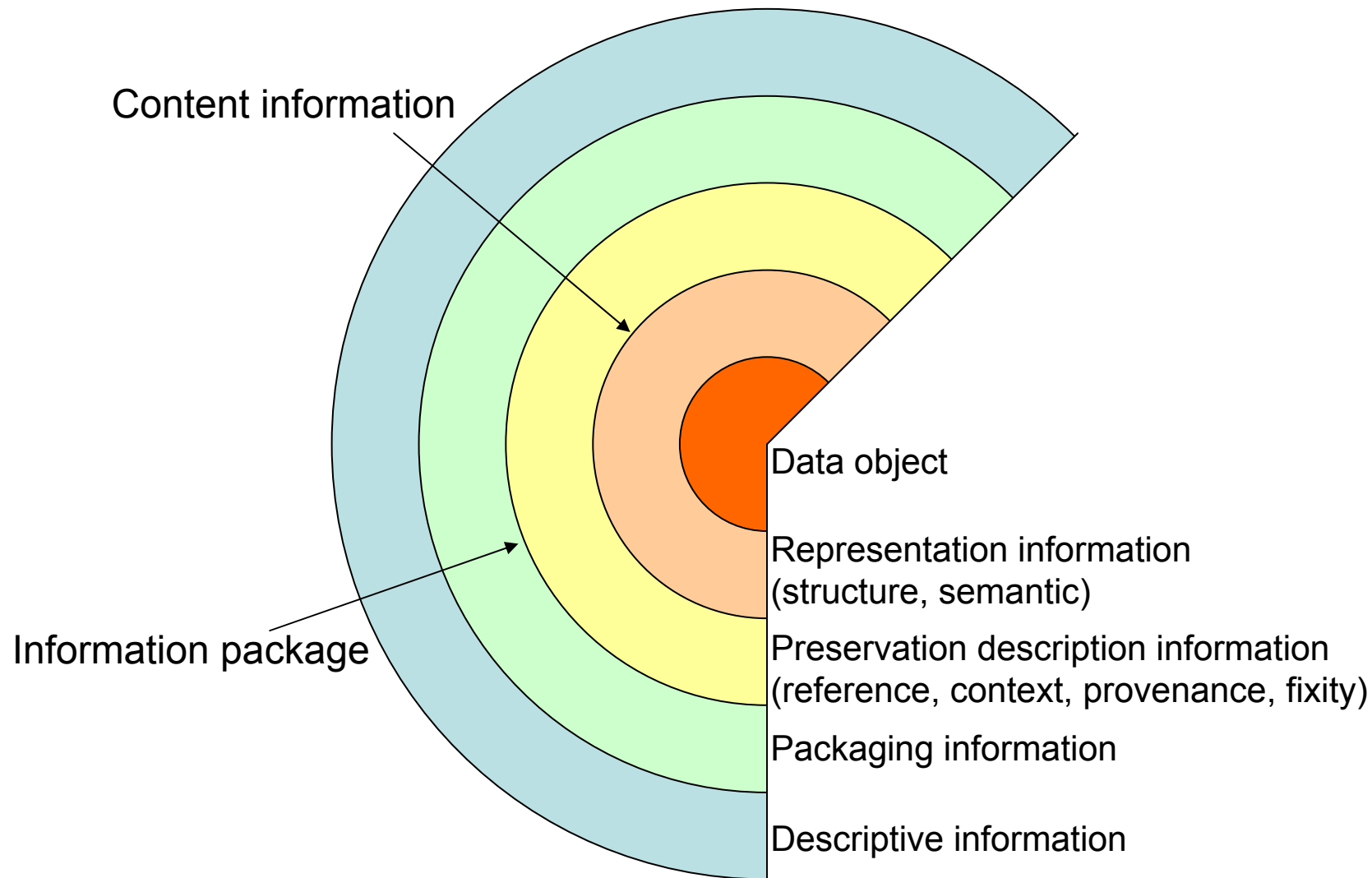


# OAIS : les fonctions



# OAIS : le modèle d'information

(représentation simplifiée)



# la propriété intellectuelle

- le droit d'auteur s'applique aux documents numériques
- une « œuvre de l'esprit » ne tombe dans le domaine public que 70 ans après le décès de son (dernier) auteur
- pour pouvoir archiver un document numérique, l'auteur - ou ses ayants droit - doit avoir cédé le droit de reproduction de l'oeuvre
- pour pouvoir diffuser un document numérique - par exemple sur Internet - l'auteur doit avoir cédé au site de diffusion le droit de représentation de l'oeuvre
- droit d'auteur ≠ copyright anglo-saxon

# les projets d'archives du CINES

- missions nationales du MENESR :
  - archivage des thèses électroniques
  - archivage des revues SHS du projet Persée
- autres projets connexes
  - accès pérenne aux revues scientifiques Elsevier
  - CENS - Centre d'Édition Numérique Scientifique du CNRS

# PAC : plate-forme d'archivage du CINES

- objectif
  - mettre en place un service générique d'archivage de documents numériques pour la communauté Enseignement supérieur et Recherche
- 3 serveurs logiques
  - versement
  - stockage
  - accès

# PAC : le versement des documents

- le producteur soumet un document - fichier(s) + métadonnées - (ou un lot) pour archivage au CINES
- le CINES accuse réception et contrôle la validité de la soumission
- le CINES archive le(s) document(s)

# PAC : le stockage des documents

- pour chaque document archivé, le CINES adresse un certificat d'archivage au producteur
- le producteur peut administrer les métadonnées de ses documents archivés
- Le CINES historise toute intervention sur un document archivé ou ses métadonnées
- le producteur reçoit un état annuel de ses documents archivés



# PAC : l'accès aux documents

- selon le projet d'archives, l'accès au catalogue est libre ou demande une authentification
- l'utilisateur accède aux métadonnées des documents
- selon les droits associés à un document et son niveau d'habilitation, l'utilisateur peut ou non télécharger le document

## pour (ne pas) conclure

- convaincre les décideurs de l'importance de l'enjeu ... *difficulté de la prise en compte du très long terme*
- les coûts ? *ne pas tout archiver*
- un seul site d'archivage ... *danger*
- la qualité de l'archivage d'aujourd'hui se prouvera réellement ... *dans 30 ans*
- en attendant ? *la certification ?*