

La gestion et la conservation des documents numériques

Catherine Lupovici

Département de la bibliothèque numérique

Direction des services et des réseaux

Introduction

- **Système de gestion des collections numériques**
 - Magasin numérique : stocker, conserver, communiquer
 - Gestion informatique des fonctions de magasinage et de communication
- **Historiquement développement de différents types selon les fonctions centrales des systèmes**
 - Gestion de biens numériques (Digital Assets Management)
 - Point fort : gestion des droits d'usage et visualisation
 - Extension plus ou moins intégrée du SIGB (ILMS)
 - Axée ressources électroniques (accès à distance)
 - Axée collections numérisées (numérisation et accès)
 - Entrepôt institutionnel (Institutional repository)
 - Point fort : aide à la publication documentaire et stockage
- **Une question clé pour les institutions patrimoniales**
 - Conservation à long terme du patrimoine numérique

Principes de la conservation du numérique établis depuis 10 ans

- *Report of the Task Force on Archiving of Digital Information/ Commission on Preservation and Access & Research Libraries Group.* - May 1, 1996. – 71 p.
http://www.rlg.org/en/page.php?Page_ID=20442
 - Importance de la conservation du contenu et plus seulement du support
 - Vrai défi : obsolescence technologique de l'ensemble de la plateforme technique de production et de consultation
 - Actions : rafraîchissement des supports, migration des fichiers, émulation des plateformes de consultation
- *Modèle de référence pour un système ouvert d'archivage d'information (OAIS).* Travail initialisé dans les années 96
 - Trad. française mars 2005 de la version anglaise 2003. ISO 14721
[http://public.ccsds.org/publications/archive/650x0b1\(F\).pdf](http://public.ccsds.org/publications/archive/650x0b1(F).pdf)
 - Modèle fonctionnel et modèle d'information

Les stratégies de gestion de la pérennisation

- Une action d'émulation ou de migration modifie la nature des objets
- Choix entre la publication et l'œuvre
 - Conservation de la forme externe (fonction de type musée)
 - Conservation du contenu à un niveau de forme suffisant pour conserver l'intelligibilité (fonction de type documentaire)
- Dans certains cas on doit assurer les deux

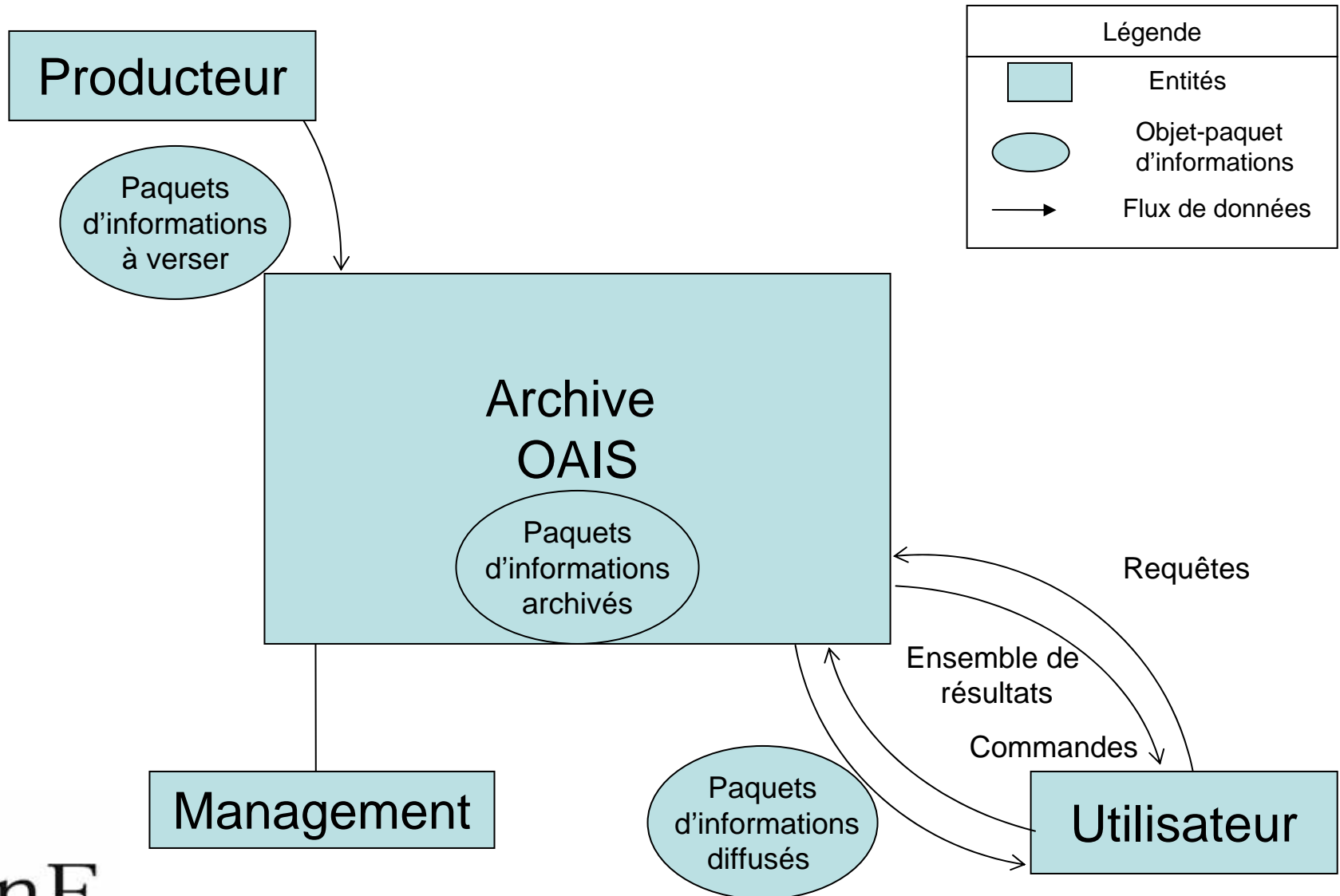
La norme OAIS (Open Archival Information System)

- Comité consultatif pour les systèmes de données spatiales (CCSDS). Red book, mai 1999
 - NASA, ESA, CNES, NARA
 - ISO 14721:2002
- Adoptée dès 1999 par tous les programmes de bibliothèques s'intéressant à la préservation et accès à long terme et par les entrepôts institutionnels
- Groupe de travail et d'échanges français PIN (Pérennisation de l'Information Numérique) créé en 2000 <http://vds.cnes.fr/pin/index.html>
 - Version française de l'OAIS en mars 2005 et normes complémentaires
 - Actions de formation
- Didacticiel : *Gestion de la conservation de collections numériques*
<http://www.library.cornell.edu/iris/tutorial/dpm-french/index.html>

Objectifs et domaine d'application de la norme OAIS

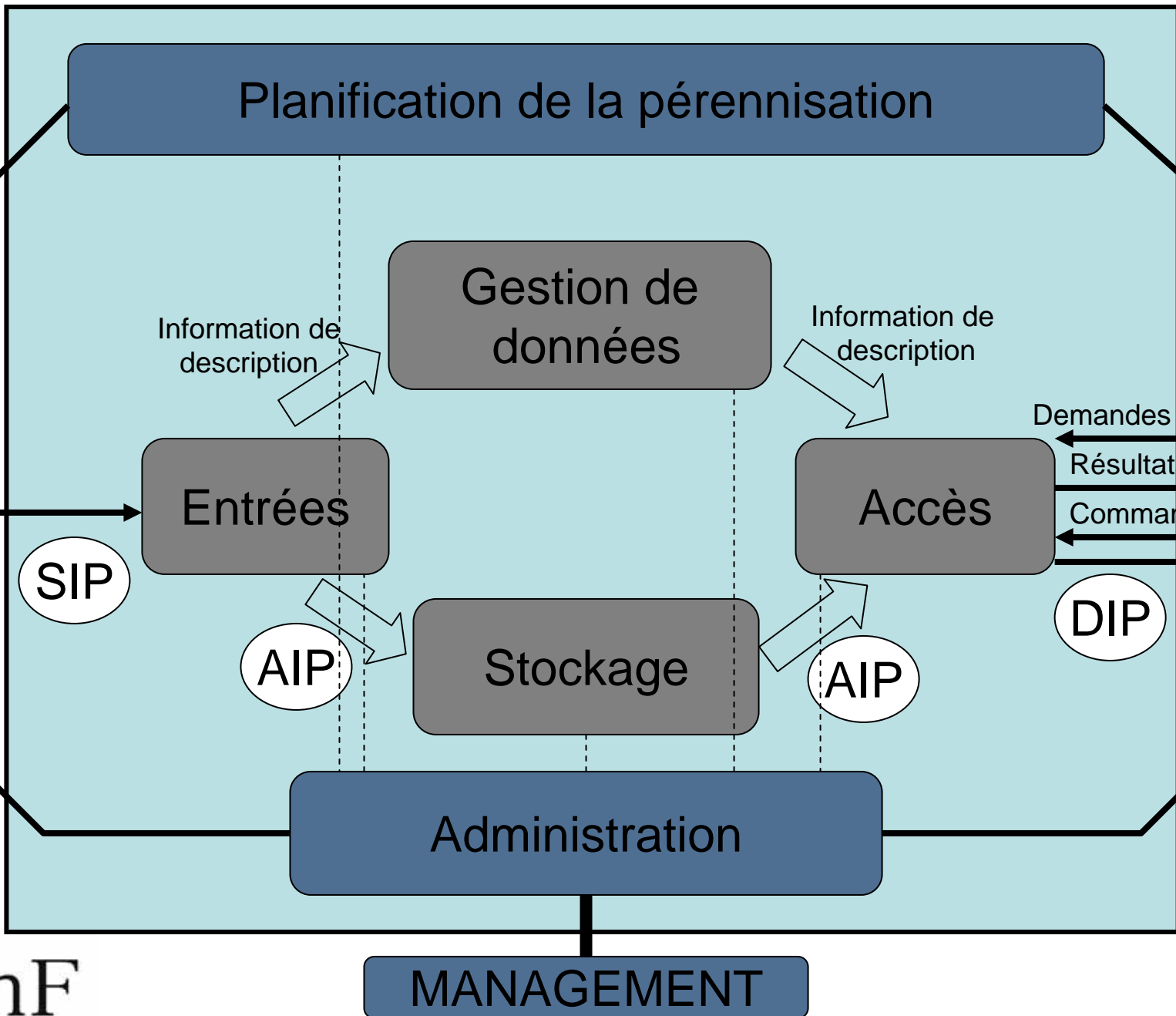
- Archivage d'objets numériques et analogiques mais très élaboré pour le numérique
- Vocabulaire neutre par rapport aux différents secteurs professionnels concernés
- Définit l'ensemble des briques fonctionnelles nécessaires au stockage, à la conservation et à la communication
 - Permet de choisir et d'adapter à ses besoins
 - Ne préjuge pas d'une implémentation particulière
- Cadre de développement pour les normes complémentaires

Flux des données externes à l'OAIS



P
R
O
D
U
C
T
E
U
R

U
T
I
L
I
S
A
T
E
U
R



Métadonnées des objets numériques

- METS (Metadata Encoding & Transmission Standards)
 - Initialisé par la Digital Library Federation. V1 février 2001
 - Schéma XML. Permet de créer un document qui encapsule toutes les métadonnées d'un objet numérique
 - Métadonnées descriptives : en MARC, EAD ou DC ... incorporées dans le fichier METS ou lien vers un autre réservoir
 - Métadonnées administratives de l'objet numérique et de l'objet numérisé : droits, techniques
 - Métadonnées de structure : cartes de structure de l'objet (logique, physique)
 - Section donnant la liste des fichiers et leurs regroupements : vignette, matrice d'archive, version de consultation image, version texte intégral ...
 - Agence de maintenance <http://www.loc.gov/standards/mets/>
- En OAIS peut être utilisé pour les SIP, AIP, DIP
- Norme d'échange d'objets numériques entre des entrepôts

Métadonnées de préservation

- *Preservation metadata and the OAIS Information model*. OCLC/RLG working group on preservation metadata. June 2002. Création du groupe PREMIS en 2003
- PREMIS (Preservation Metadata for Digital Materials) Data dictionary , mai 2005
 - Dictionnaire des métadonnées associées à la chaîne des traitements
 - Pour la description des formats de documents, PREMIS renvoie sur les travaux de description des répertoires de formats
 - Maintenance et suivi des travaux désormais assurés par la Library of Congress

Les répertoires de formats

- Initiatives collectives répondant à différents besoins
- Producteur d'information souhaitant choisir un format en fonction de sa préservabilité
- Identification et caractérisation technique pour la gestion des entrées dans l'archive : création des métadonnées techniques relatives au format
- Evaluation des risques d'obsolescence par l'entité planification de la pérennisation

Certification des systèmes d'archivages à long terme

- Recueil des besoins pour une certification
 - RLG et NARA (Archives nationales USA)
 - Document publié pour commentaires le 31 août 2005 dans le but d'une normalisation internationale ISO
<http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf>
- Éléments du test de conformité
 - Organisation
 - Fonctions, traitements et procédures de l'entrepôt numérique
 - Communauté des utilisateurs cibles et utilisation projetée des contenus
 - Infrastructure technique et technologies mises en œuvre

Audits d'archives numériques

- Projet conduit par le Center for Research Libraries (CRL)
 - Test des éléments d'évaluation RLG/NARA
 - Archives scientifiques des revues d'Elsevier à la Koninklijke Bibliotheek (Pays Bas)
 - Inter University Consortium for Political and Social Research (ICPSR)
 - Système d'archivage Portico pour les revues numériques

Exemples de systèmes conformes à l'OAIS

- Systèmes de gestion d'entrepôts numériques
 - DIAS d'IBM développé très tôt dans la normalisation OAIS
 - DSpace open source développé par le MIT
 - Fedora open source. Cornell University and University Virginia Library. Support de la Fondation Mellon (-sept 2007)
- Systèmes de gestion de bibliothèque
 - Kronos système développé en partenariat NLNZ / Endeavour / Sun Microsystems
 - Endeavour racheté par ExLibris dont le module DigiTool est actuellement orienté uniquement accès
 - VTLS : utilise Fedora

Conservation et accès aux collections numériques à la BnF

Gestion actuelle hétérogène

- Numérisation pour Gallica(1992→)
 - accès en ligne après migration, conservation sur support en verre du master numérique
- Dépôt légal (1994 →)
 - publications sur support : conservation du support (recopies) et émulation (Département de l'audiovisuel)
 - en ligne : migrations dès l'entrée dans certains cas (2002 →), pas d'action de conservation, sauvegarde informatique seule
- Acquisitions
 - accès en ligne ou sur support, pas de conservation (1998 →)
- Archives de la BnF (2006 →)
 - organisation de la production en véritable records management
 - sauvegarde informatique seule

Systeme d'Information du Numérique

- Travail préliminaire (2004-2005)
 - Inventaire des collections estimation des volumes et de l'accroissement
 - Cahier des charges et acquisition des espaces de stockage
 - Révision des recommandations techniques de numérisation pour une numérisation couvrant tous les besoins de communication, de reproduction de sauvegarde et de service commercial
- Réalisation du système d'information (2006-2009)
 - Phase initiale 2006
 - Mise en place du stockage
 - Recueil des besoins et spécifications fonctionnelles du système d'archivage pour la gestion de la conservation et de la communication
 - Évolution et maintenance des applications existantes pour assurer une continuité de service et leur intégration dans le nouveau système global

Estimation des volumes de stockage 2004

		<i>Année 2005</i>	<i>Année 2009</i>	<i>Année 2014</i>
Type de collection	Formats de données			
Département de l'audiovisuel				
Numérisation de sauvegarde	Défini BnF	96	400	460
Migration de CD et DVD			50	280
Dépôt légal nouveaux projets				
Presse quotidienne régionale	PDF	2	9	18
Archivage Web	Tous formats	80	400	800
Numérisation				
Communication (Internet intranet)	TIFF, JPEG	17	80	160
Reproduction commerciale	TIFF, JPEG	3	6	10
Reproduction de sauvegarde	TIFF	2	6	11
Acquisitions En cours d'instruction				
TOTAL en To		200	951	1739

Mise en place du SI Numérique

- Utiliser le modèle fonctionnel et le modèle d'information de l'OAIS (Open Archival Information System)
- Associer tous les utilisateurs du système à la validation des fonctions et aux contrats de service que le système devra garantir
 - Services d'entrées : acquisitions à titre gratuit ou onéreux, dépôts de partenaires publics ou privés, dépôt légal, numérisation, records management
 - Gestion administrative et gestion de la conservation
 - Accès utilisateur final et applications de consultation (BnF ou autres fournisseurs de services)
- Recommandations techniques : formats de métadonnées, registre de formats de données, système global d'identifiant pérenne, sous-système de gestion des droits d'auteur et droits d'accès, gestion des risques

Conclusion

- Un projet global pour un système d'information de conservation et de communication de tous les biens numériques de la BnF
 - En 2006, la BnF y consacre 5 année/homme de bibliothécaires et 12 année/homme d'ingénieurs informaticiens
- Facteur de risque technique fort : objectifs qui pourraient s'avérer difficile à concilier
 - Système de conservation certifié au niveau des procédures qualité garantissant l'intégrité, l'authenticité et la sécurité des collections
 - Système de communication très rapide sur le Web des collections dans le domaine public (Gallica effectue 1,5 millions de communications de documents par mois pour une collection de 150 000 documents)
- Enjeu majeur de gestion du changement
 - Adhésion de l'ensemble de la bibliothèque, évolution des métiers et formations des personnels