



Le modèle OAIS et sa mise en œuvre à la BnF



www.drexel.edu

Les enjeux de la conservation des documents numériques, Bordeaux 23 mai 2008

Frédéric Martin – Bibliothèque nationale de France / Direction des Services et des Réseaux

frederic-d.martin@bnf.fr

- OAIS est le sigle d' **Open archival Information System** (Système ouvert d'archivage d'information)
- OAIS est un modèle de référence ou cadre conceptuel pour la **préservation à long terme de l'information** et sa communication à une communauté d'utilisateurs cible.
- une recommandation du CCSDS *Consultative committee for space data system* (<http://www.ccsds.org>) qui réunit 10 agences spatiales nationales (NASA, CNES...) ; la recommandation est révisable 5 ans après la date d'émission ; l'édition originale date de janvier 2002 ; traduction française par T. Cloarec (BnF) et C. Huc (CNES), de janvier 2005.
- norme ISO depuis 2003 sous la référence ISO 14721:2003



- Principales caractéristiques :



- Utilisable aussi bien pour l'information sur support analogique que numérique (mais surtout utilisé dans le domaine numérique)
- Définit l'ensemble des briques fonctionnelle nécessaires au stockage, à la conservation et à la communication de l'information
 - Quelle que soit la discipline, le type d'institution (choix d'une terminologie neutre)
 - Sans préjuger de la conception ni de la réalisation d'un système particulier
- Sert de cadre à des normes associées (par ex. PAIMAS)

Les responsabilités OAIS

Dans la partie « conformité » sont énumérés les services minima à assurer pour se prévaloir d'une qualification OAIS.

Ces responsabilités obligatoires sont :

- **entrée** des informations dans le système : négocier avec les producteurs les règles d'entrée et valider les données qui entrent ;
- **préservation** : maîtriser l'information au niveau nécessaire pour la pérenniser ;
- **accès** : avoir défini la communauté d'utilisateurs cibles en mesure de comprendre l'information demandée en communication sans assistance des producteurs ;
- **disponibilité** : rendre l'information conservée disponible pour la communauté d'utilisateurs cibles ;
- **authenticité** : garantir l'authenticité et tracer le rapport avec l'original en cas de migration.

L'environnement OAIS

- Un système d'archivage ouvert OAIS s'insère dans un système administratif global de gestion au sein d'une institution.
- Ce n'est pas un système autosuffisant isolé. Il obéit aux objectifs généraux et aux règles générales de gestion de l'institution définis par sa direction. Il a des interactions avec les autres systèmes de gestion de cette institution.
- OAIS définit quatre rôles principaux dans un système d'archivage :
 - un acteur interne, l' « Archive », c'est-à-dire l'opérateur du système d'archivage,
 - et trois acteurs externes : le « Management », les « Producteurs » et les « Utilisateurs ».



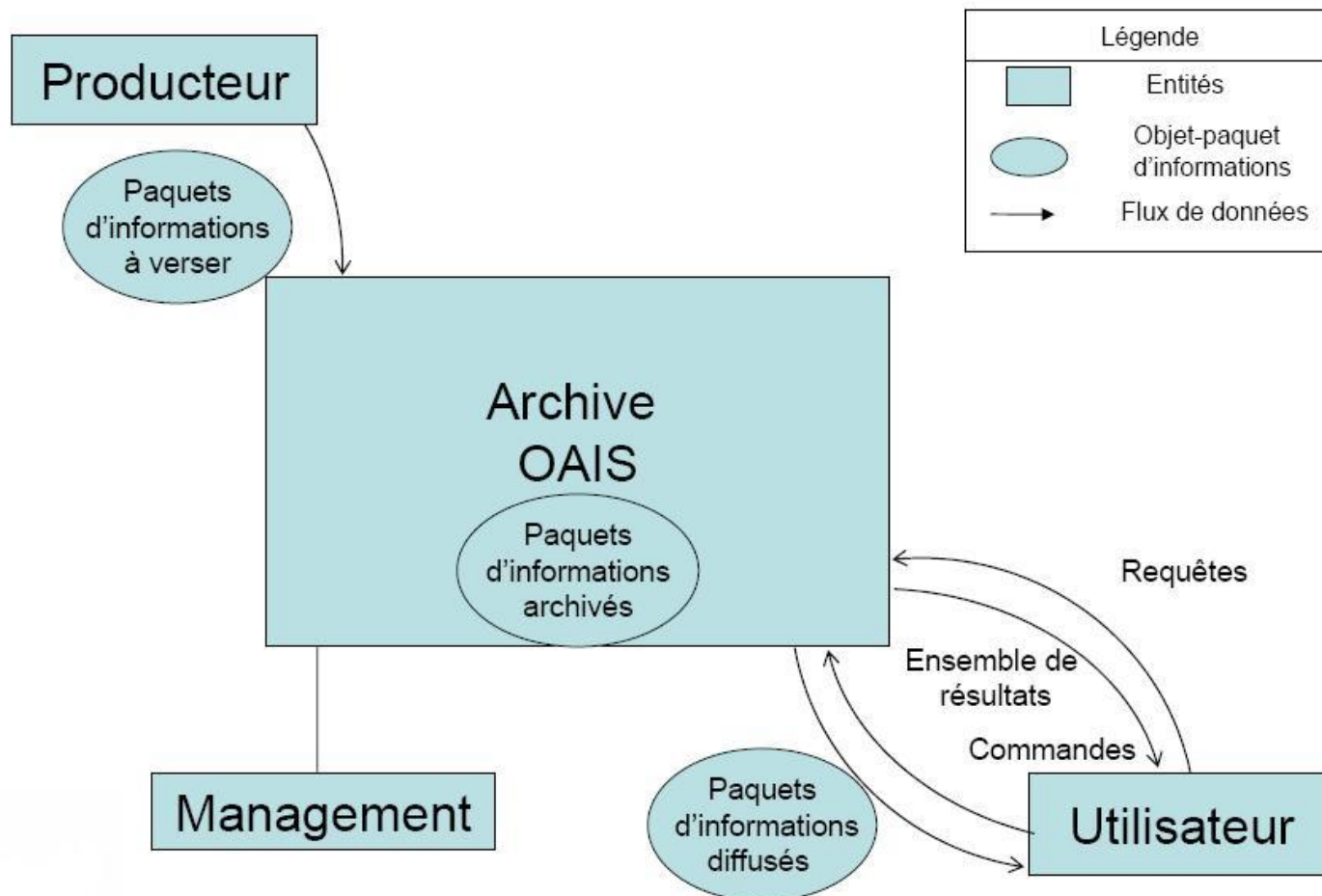
Les acteurs de l'archivage

- **L'Archive OAIS** est l'opérateur du système d'archivage. Elle reçoit des données soumises à l'archivage par les Producteurs de données.
- **Producteur** ne désigne pas nécessairement ici une entité ayant créé l'information mais l'entité, personne ou système qui fournit l'information à conserver.
- **Les Utilisateurs** de l'Archive sont soit des personnes, soit d'autres systèmes qui demandent la communication d'informations contenues dans l'Archive en question.

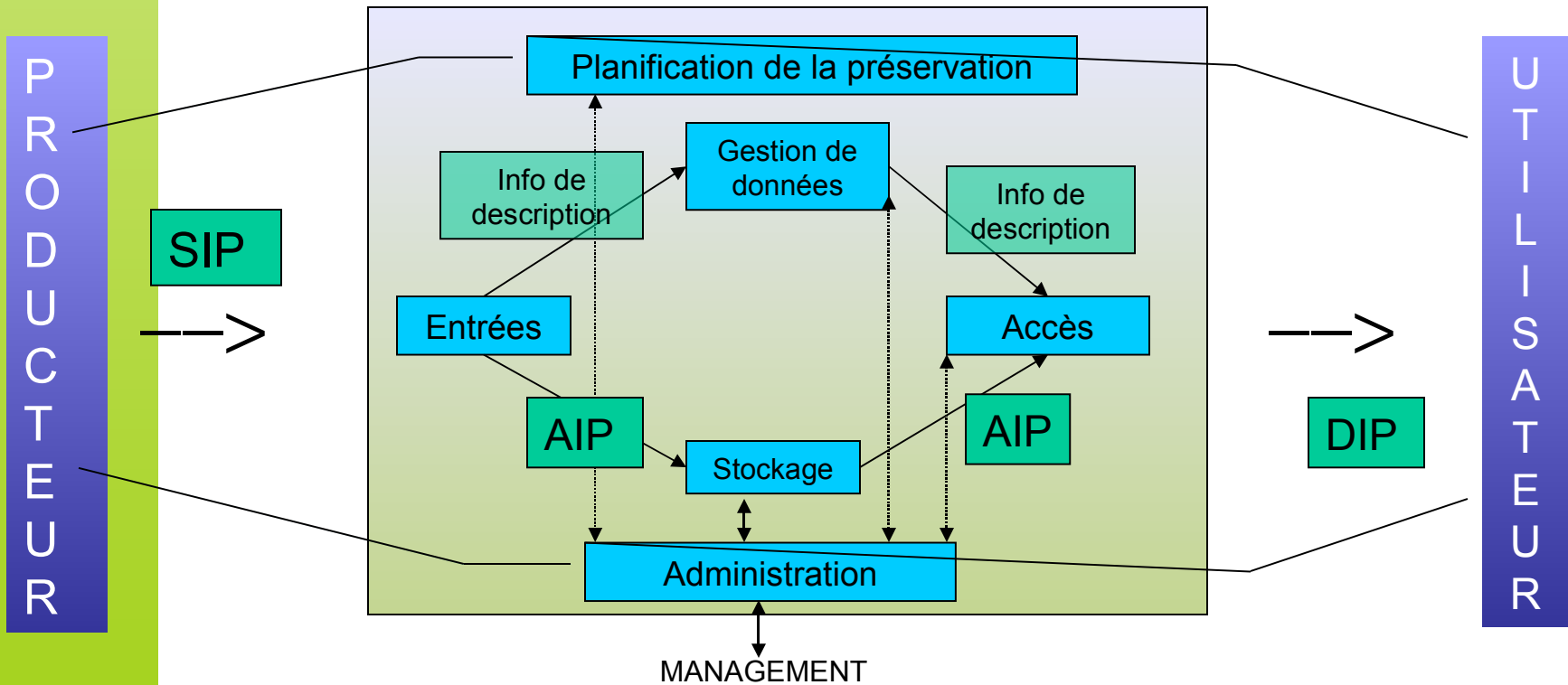
Le système de l'Archive n'est donc *a priori* pas un système techniquement indépendant, mais un système qui pourra **interopérer** avec d'autres systèmes externes avec lesquels certains de ses modules fonctionnels doivent communiquer.



Flux de données externes



Principales fonctions OAIS



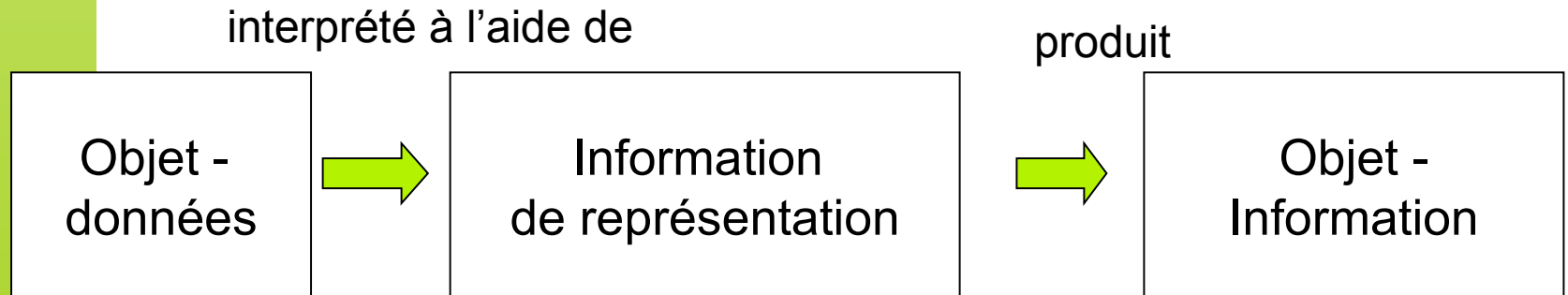
Les 6 principales fonctions OAIS

- **Entrées**: elle reçoit, contrôle et valide les objets à archiver. Les objets eux-mêmes sont transmis à l'entité « stockage », tandis que les informations nécessaires à leur description et à leur gestion dans le temps sont transmises à l'entité « gestion des données ».
- **Stockage**: elle stocke, maintient et restitue les objets archivés.
- **Gestion de données** : elle crée les données descriptives qui documentent les collections archivées et les données administratives internes à l'Archive, les maintient et en permet l'accès.
- **Accès** : elle rend l'information disponible aux utilisateurs.
- **Administration**: elle assure les services généraux et les autres fonctions de gestion de l'Archive OAIS; elle garantit la coordination générale du système. Elle en établit les règles internes. Elle veille à la qualité globale du service rendu et à son amélioration. Elle rend compte au Management.
- **Planification de la pérennisation** : cellule de veille et de planification du système, elle pilote l'environnement de l'Archive et fournit les recommandations d'actions pour que l'information archivée demeure accessible sur le long terme pour la communauté d'utilisateurs cibles, même si l'environnement informatique original devient obsolète.

Les Migrations numériques

Le modèle OAIS distingue 4 types de « migrations numériques » pour pérenniser l'information

- **Le Rafrâichissement de support** : Migration numérique visant à remplacer un support par une copie suffisamment exacte pour que l'Entité « Stockage » (matériel et logiciel) continue à fonctionner comme auparavant.
- **La Duplication** : Migration numérique dans laquelle il n'y a de changement ni dans l'Information d'empaquetage, ni dans le Contenu d'information, ni dans le PDI. Les bits représentant ces Objets-information sont conservés dans le transfert vers un support de même type ou vers un autre support.
- **Le Ré-empaquetage** : Migration numérique au cours de laquelle l'Information d'empaquetage du Paquet d'information archivé (AIP) est modifiée.
- **La Transformation** : Migration numérique au cours de laquelle le Contenu d'information ou l'Information de pérennisation (PDI) d'un Paquet d'informations archivé (AIP) est modifié. Par exemple, la conversion des codes ASCII en UNICODE dans un document textuel est une transformation.



- **L'Objet-données** n'est pas forcément un **Objet-numérique**.
- Récursivité : l'information de représentation est également un objet-données pouvant nécessiter une information de représentation pour produire un **objet-information**.

OAIS Modèle de donnée

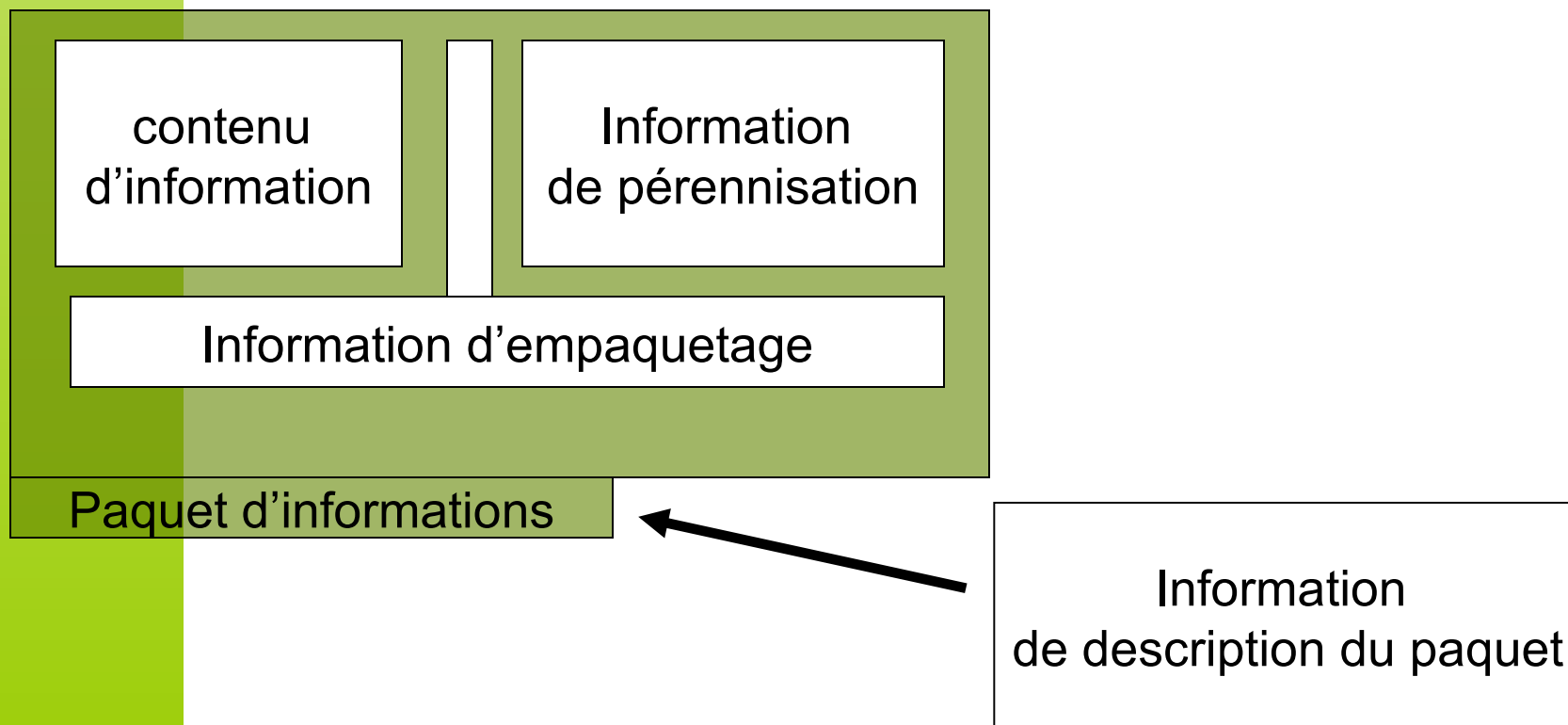
- **Objet-données** : objet physique ou numérique
- **Information de représentation** : information qui traduit un objet-données en des concepts plus explicites. Par ex., la définition du code ASCII décrit comment une séquence de bits (un objet-données) est convertie en caractère
- **Objet-information** : objet-données avec son information de représentation

Information de représentation

=> pour traduire les séquences de bits en information intelligible

- **Information de structure** : le ou les formats de l'objet-numérique
- **Information sémantique** : par ex. la langue d'un texte
- **Réseau de représentation** : par ex. si l'information de représentation possède elle-même une information de représentation

Définition d'un paquet d'informations



Définition d'un paquet d'informations

- **Contenu d'information**
 - ensembles d'informations constituant l'objet principal de la pérennisation ; c'est un objet d'information (objet-données + information de représentation)
- **Information de pérennisation**
 - information nécessaire à la bonne conservation du contenu d'information
- **Information d'empaquetage**
 - information permettant de relier et identifier les composants du paquet d'information
- **Information de description du paquet**
 - pour la recherche et la récupération de paquets d'information

L'information de pérennisation

- **Information d'identification**
 - identifie et éventuellement décrit la ressource
- **Information de contexte**
 - documente les relations du Contenu d'information avec son environnement
- **Information de provenance**
 - documente l'historique des Contenus d'information (origine, source, changements...)
- **Information d'intégrité**
 - fournit des moyens de contrôler l'intégrité des données

Les différents Paquets d'informations

- **paquet d'informations à verser (SIP)**

paquet d'information livré par le Producteur pour l'élaboration d'un ou de plusieurs AIP

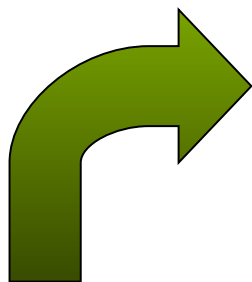
- **paquet d'informations archivé (AIP)**

paquet d'informations conservé dans un OAIS

- **paquet d'informations diffusé (DIP)**

paquet d'informations reçu par l'Utilisateur en réponse à sa requête à l'OAIS ; provient d'un ou de plusieurs AIP

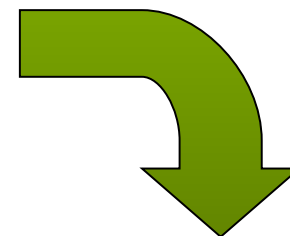
Les Paquets d'informations



Paquets à verser (SIP)



Paquets archivés (AIP)



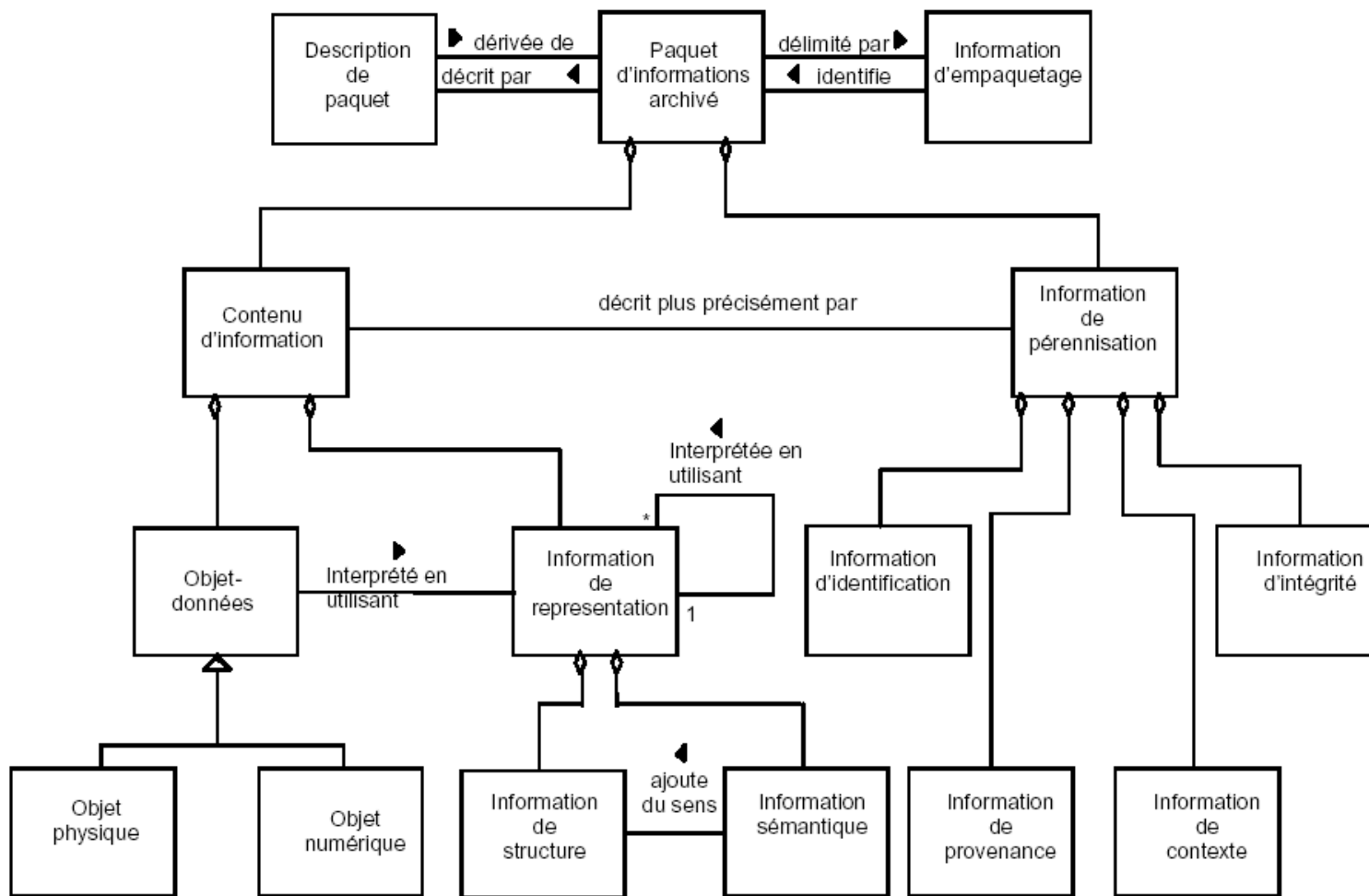
Paquet diffusé (DIP)

Nature des Paquets

- Des Paquets « logiques »
 - Un Paquet par document, par ex. pour les documents provenant de la numérisation.
- Des Paquets « physiques »
 - Dépôt légal du web : les Paquets sont constitués selon le parcours du crawler sur Internet. Les fichiers d'un même site Internet peuvent se retrouver dans différents Paquets.



Modèle d'information



Paquet d'informations archivé (vue détaillée)

Standards autour de l'OAIS

- PAIMAS
 - Guide méthodologique de négociation Producteur / Archive (ISO 20652 : 2006)
- Projets de certification OAIS
 - TRAC
 - DRAMBORA interactive
- Standards de métadonnées et d'identifiants
 - PREMIS (métadonnées de préservation)
 - METS (structuration des objets numériques, emballage des métadonnées)
 - Identifiants pérennes ARK...



Mise en œuvre à la BnF

Le projet SPAR : objectifs

- Prendre la responsabilité de la conservation à long terme de données
- Atteindre une masse critique minimale pour réduire les coûts (moyens matériels, logiciels et humains)
- Permettre la mutualisation de stockage d'archives entre plusieurs établissements ou institutions
- Réaliser un système centralisé pour réduire la diversité (matériel de stockage, formats utilisés, départements ...)
- S'inspirer des bonnes pratiques et standards du domaine
- => Volonté de réaliser un système conforme à l'OAIS
- **SPAR = système de préservation et d'archivage réparti**

Calendrier et démarche

- 2006 – Définition du besoin : les groupes de travail réalisent les choix structurants
- 2007 - Écriture du marché SPAR réalisation
- 2008 - Réalisation de la première tranche
 - Fonctions de base de l'OAIS
 - Filière numérisation de conservation (chaîne A : images, imprimés)
- 2008 - 2010 Intégrer de nouveaux modules et de nouvelles « filières »

Les groupes de travail

- Les groupes de travail fonctionnels: Ils comprennent des utilisateurs représentatifs des départements de la BnF:
 - G1 : groupe transverse de suivi, « Fonctions OAIS »
 - G2 : communauté d'utilisateurs
 - G3 : modèles d'information (formats de fichiers, métadonnées de préservation)
 - G4 : gestion des risques
 - G2D : gestion des droits
- Groupe de travail technique (département des systèmes d'information).

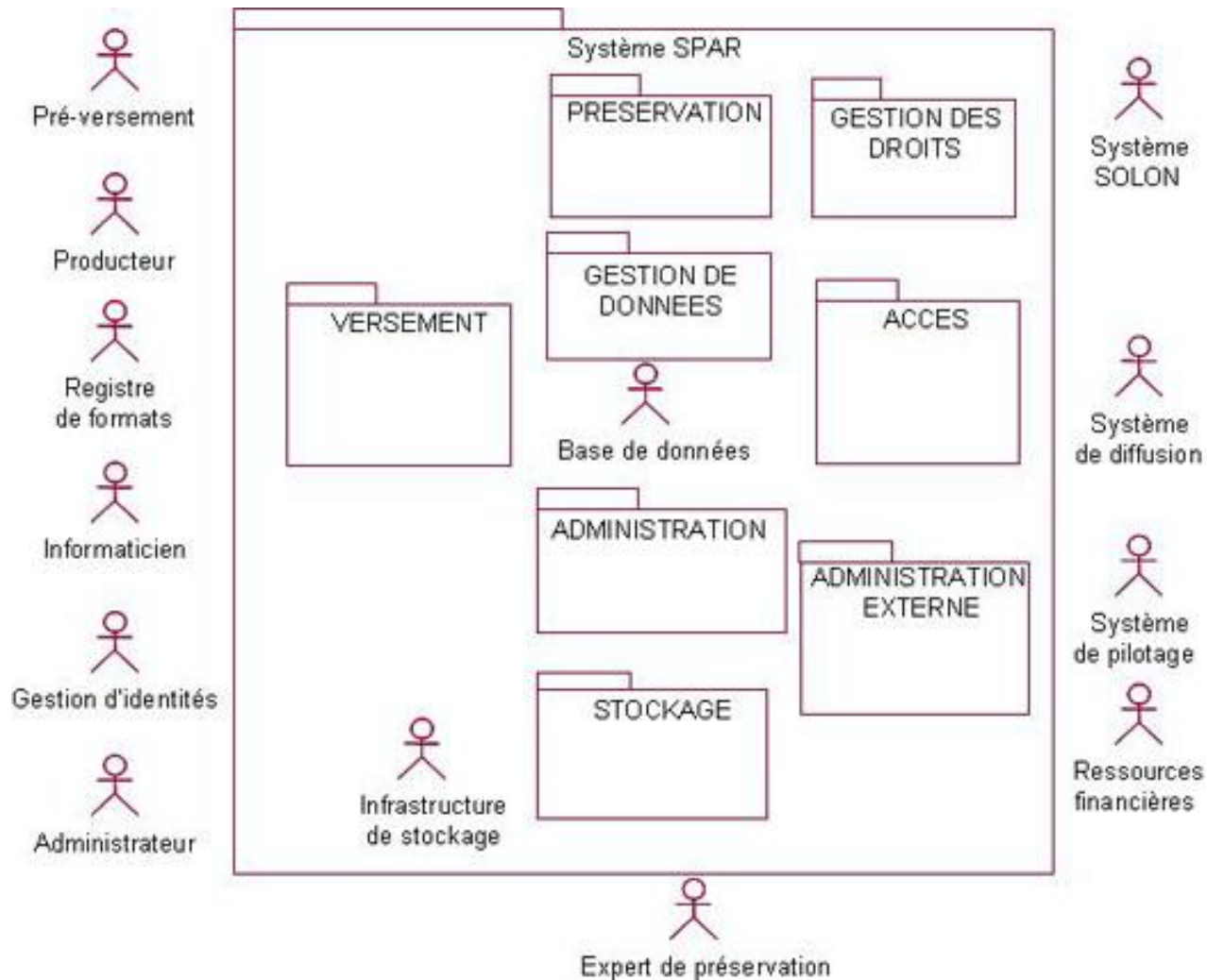
Les exigences techniques

- Ouverture
- Respect des normes établies
- Modularité, distributivité et scalabilité
- Performances, robustesse, haute-disponibilité
- Pérennité
- Journalisation, pour tracer ce qui s'est passé
- Compatibilité avec l'infrastructure
- Intégration de briques logicielles existantes

La solution proposée

- Société ayant remportée le marché: Atos Origin 
- Réalisation en ce moment par une équipe de 12 personnes, hébergée à la BnF
- Projet d'intégration basé sur des briques Open Source
- Rapprochement privilégié vers : 
 - La fondation Fedora Commons
 - Le SDSC 
 - Sun 

SPAR et OAIS : la modularité



Les filières d'entrées

- filière de numérisation de conservation,
- filière de numérisation de reproduction,
- dépôt légal automatique,
- dépôt légal négocié,
- archivage légal des documents administratifs et techniques de la BnF,
- dépôt et tiers archivage dans le cadre de sa politique de coopération nationale,
- acquisition et don de données numériques.

Le modèle de données SPAR

- Les paquets d'archive sont constitués de différents types d'informations, classés dans le système suivant leur contenu et leur rôle dans le fonctionnement de l'Archive.
- Dans SPAR, on distingue essentiellement les concepts suivants:
 - **concept d'objet-données**, qui correspond aux fichiers numériques à conserver
 - **concept de métadonnées**, qui correspond aux informations nécessaires pour rendre intelligible les objets-données (en particulier, l'information de représentation et l'information de pérennisation)
 - **concept d'empaquetage**, qui décrit les liens réels ou logiques des différents composants enregistrés d'un paquet sur un support. Dans SPAR, il est exprimé à l'aide d'un Manifeste METS qui correspond à un format d'empaquetage.



MIX

textMD



P R E M I S

- **Dublin Core** pour les informations de description, c'est-à-dire la description de l'objet que l'on veut archiver,
- **MIX** pour encoder les métadonnées techniques des images fixes,
- **textMD** pour encoder les métadonnées techniques des fichiers textuels,
- **ODRL** pour encoder les licences d'utilisation des objets numériques,
- **PREMIS** pour les informations de provenance, c'est-à-dire le suivi de la vie des objets-données
- **ARK** pour les identifiants pérennes de consultation des objets numériques

Ex. Tramway électrique de Bordeaux-Bouscat au Vigean : projet de prolongement dans Bordeaux : avis sur enquête / Mairie de la ville de Bordeaux , 1894- <http://gallica2.bnf.fr/ark:/12148/bpt6k35364v>

Les niveaux de granularité

Set

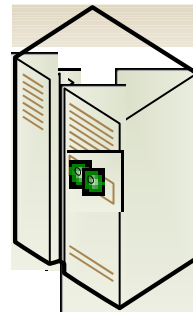
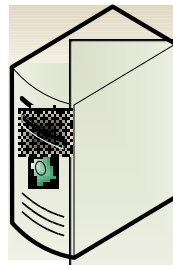
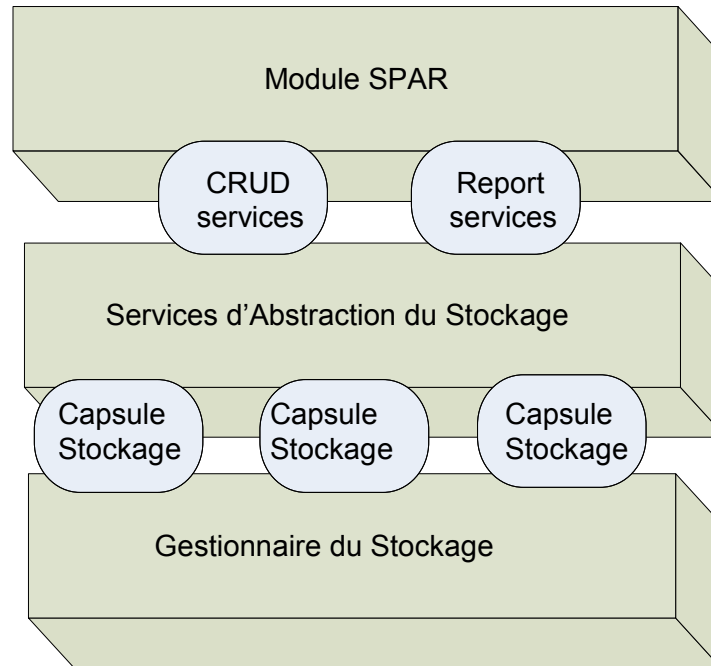
Group

Object

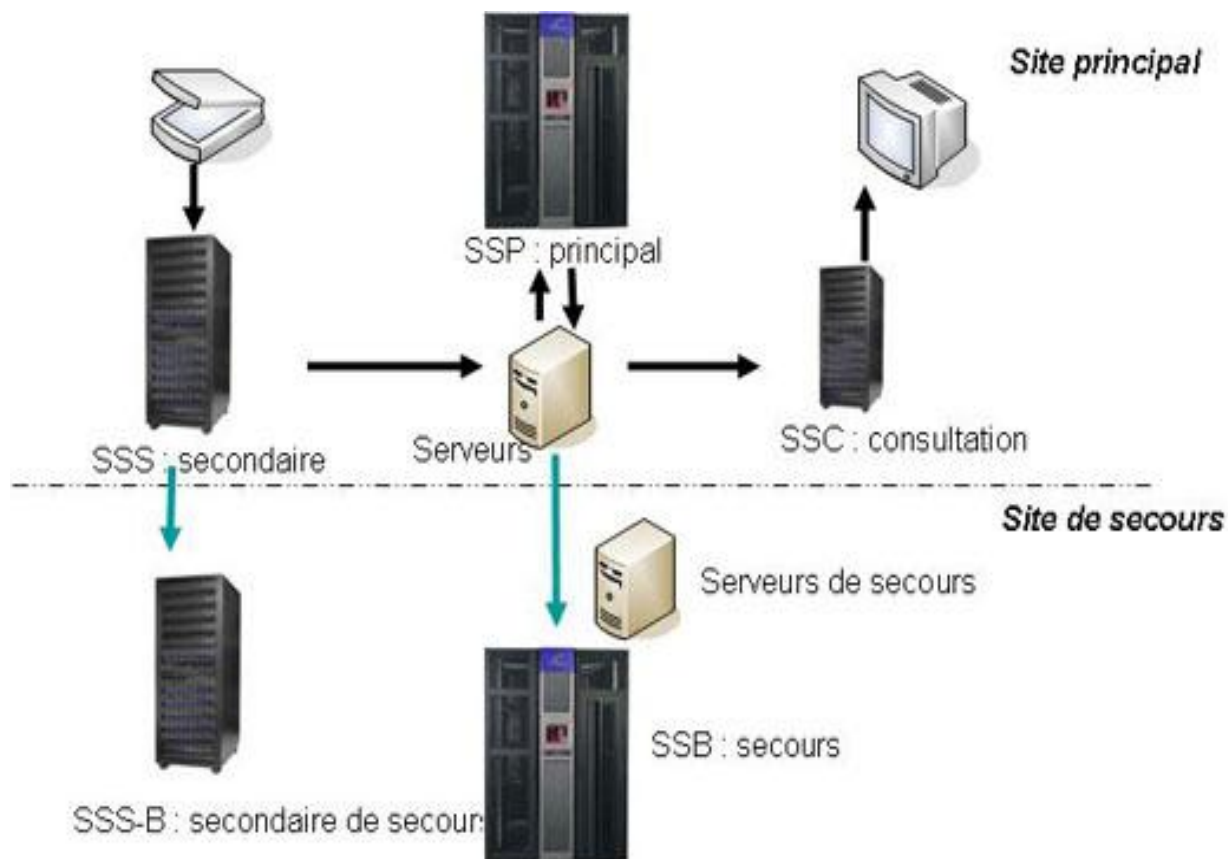
File

- 4 niveaux de granularité :
- **set**: collection d'objets numériques
- **group**: objet numérique (monographie, lot d'images, ...)
- **object** : élément (page...)
- **file**: objet-donnée (fichier numérique ou train de bits)

Abstraction du stockage



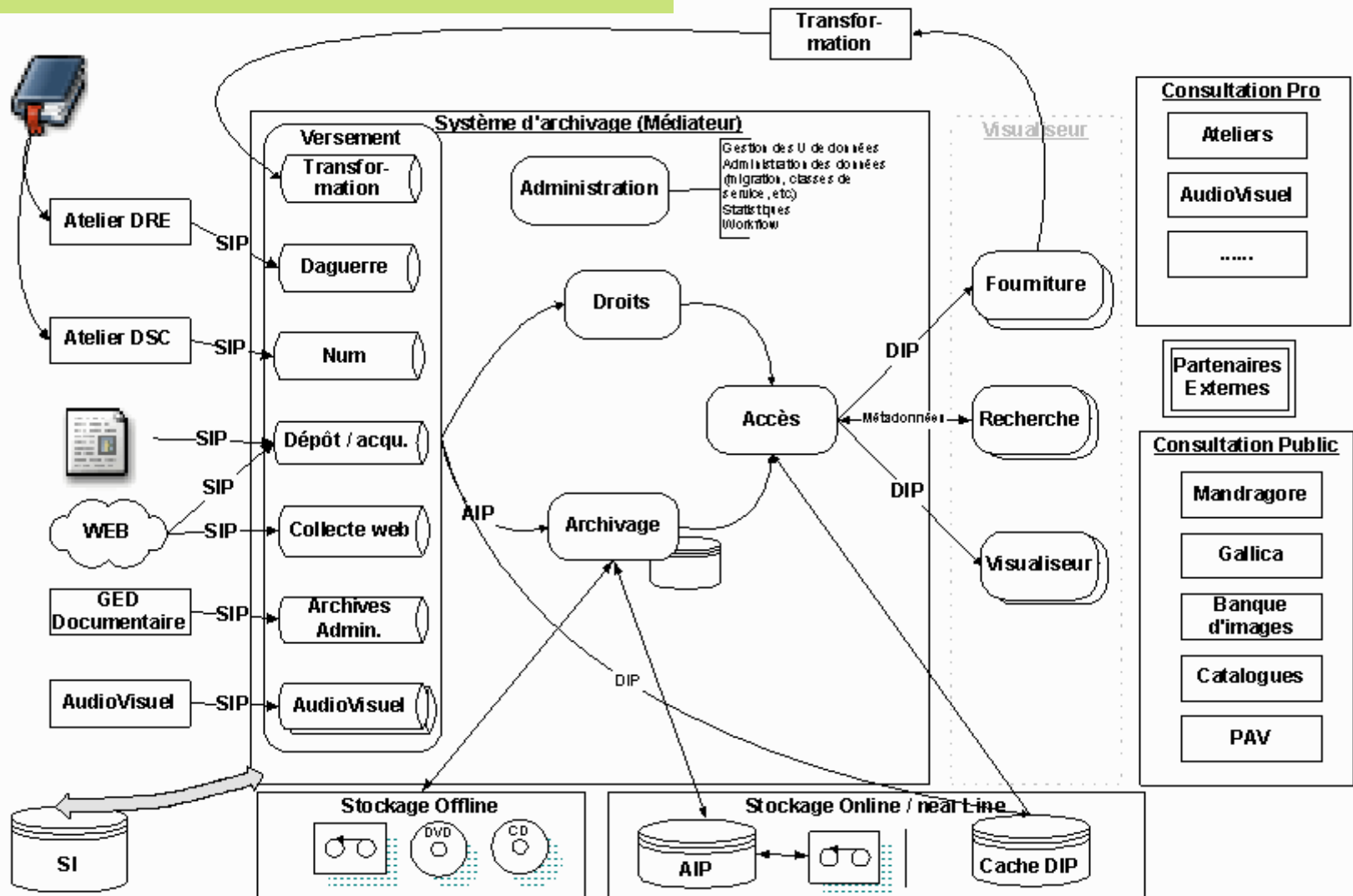
Infrastructure de stockage



Matériel de stockage



Vue d'ensemble



- **La traduction française de la norme :**
http://vds.cnes.fr/pin/documents/projet_norme_oais_version_francaise.pdf
- **Le didacticiel de l'Université de Michigan (en français)**
<http://www.icpsr.umich.edu/dpm/dpm-french/terminology/oais.html>
- **Qu'est-ce que le modèle OAIS ? / Figoblog**
<http://www.figoblog.org/document1089.php>
- **Les travaux du groupe PIN (pérenniser l'information numérique)**
<http://vds.cnes.fr/pin/>
- **Présentation de SPAR**
http://www.bnf.fr/pages/infopro/numerisation/num_spar.htm