

La conservation numérique



Maîtriser les conditions de stockage et de conservation des documents numériques

Stage Médiat Rhône-Alpes 24 mars 2010



Plan

1. Enjeux de la conservation numérique
2. État des lieux
3. Les bonnes pratiques (supports, formats)
4. Politique de conservation (gestion des risques, plan de nommage...)
5. Exemples de plates-formes mis en place



1. Enjeux de la conservation numérique



Quels sont les documents numériques ?

■ Documents issus de la numérisation

- conversion d 'analogique à numérique (images, sons, vidéos)
- copie de numérique à numérique

■ Documents nés numériques

- publications scientifiques
- bases de données
- documents bureautiques
- messages électroniques



Enjeux

- Préserver les originaux
- Préserver les documents n'existant que sous forme numérique
- Pérenniser les investissements réalisés
- A prendre en compte dès début d'un projet de numérisation
- Pas de solutions prêtes à l'emploi



Du changement dans la continuité

Bibliothèque physique

Sélection
Acquisition
Catalogage
Équipement
Magasinage
Conservation
Communication
Valorisation

Bibliothèque numérique

Sélection
Collecte / production
Métadonnées/ indexation
Empaquetage
Gestion
Préservation
Accès
Valorisation



Conserver, préserver

- Conservation des documents numériques, c'est surtout les préserver (en général, avec le numérique, après = trop tard)
- Les méthodologies de conservation sont les mêmes pour le papier et le numérique
 - stratégies préventives,
 - plans de conservation,
 - choix de priorités en fonction de la valeur/fragilité des collections,
 - gestion de risques,
 - alertes et résolutions de problèmes,
 - plans d'urgence...

Risques majeurs



- **Obsolences technologiques** (applications informatiques et formats de fichiers)
- **Fragilité et durée de vie des supports**
- **Ne pas retrouver des fichiers si pas bien décrits**



Politique de conservation numérique

- Importance de l'organisation
(responsabilité de chacun)
- Veille permanente
- Nommage des fichiers
- Formats de pérennisation



2. État des lieux



L'enquête DLL 2006 sur la conservation des documents numérisés en bibliothèques

- Enquête menée auprès de 20 BM concernées de près par la numérisation
- Objectifs :
 - connaître les supports utilisés (CD, DVD, disque dur, serveur...), les formats des fichiers (diffusion et conservation), les capacités de stockage
 - connaître les procédures de sauvegarde et de migration
 - identifier les services informatiques sollicités



Résultat de l'enquête DLL 2006 (2)

Les CD sont le support privilégié à plus de 80%

- Respect des recommandations de livraison et de gravure (support de consultation et de conservation)
- CD-Gold, support de référence pour la conservation
- Parfois les conditions de conservation ne peuvent être respectées, faute de moyens : pas de double lieu de stockage, de copies régulières et transferts de supports (norme NF Z 42-013)



Résultat de l'enquête DLL 2006 (3)

- Acquisition de serveurs de stockage et mise en place de système de stockage dans les grandes BM
 - implication des informaticiens
 - bonne coopération avec le DSI de la collectivité
 - mutualisation des moyens au niveau de la collectivité

Exemples : médiathèque de l'agglomération troyenne, BM de Lyon et de Lille

- Les petites et moyennes BM se trouvent plus démunies



Résultat de l'enquête DLL 2006 (4)

Paysage en mutation rapide, à 2 vitesses

- Signes encourageants :
 - Prise de conscience de l'importance de la conservation numérique
 - Partage des rôles et compétences entre la bibliothèque et les services informatiques
 - Apparition de plate-formes d'archivage numérique
- Nécessité d'information, de soutien et de coopération pour les moyennes et surtout les petites structures



Évolutions constatées lors du Schéma numérique des bibliothèques en 2009

- Évolution du stockage des fichiers de conservation : la solution support magnétique / serveur se développe
- Choix d'un prestataire de service pour la conservation et maintenance des fichiers de consultation en légère progression
- Question de l'externalisation envisagée pour la maintenance des fichiers de conservation, et plus seulement la gestion en interne
- Dans les nouveaux projets, la question de la conservation à long terme est prise en compte :
 - Stockage sur 2 sites distants,
 - transfert de support
- Mutualisation des moyens au niveau de la collectivité



Le Schéma numérique des bibliothèques

Enquête sur les bonnes pratiques

■ Volumétrie :

- 50 % < 1 To
- 31,7 % > 1 à 5 To (< 5 To pour 81,7 %)
- 16,7% > 5 à 20 To
- 1,7 % > 20 et 100 To
- 0% > 100 To

■ Fonds concernés à 75 % pour numérisation rétrospective patrimoniale

■ Formats les plus utilisés : PDF, JPEG, TIFF (77 %), mais on trouve des formats propriétaires (Microsoft)

■ Méconnaissance des normes et standards de préservation à plus de 50 %



Le Schéma numérique des bibliothèques

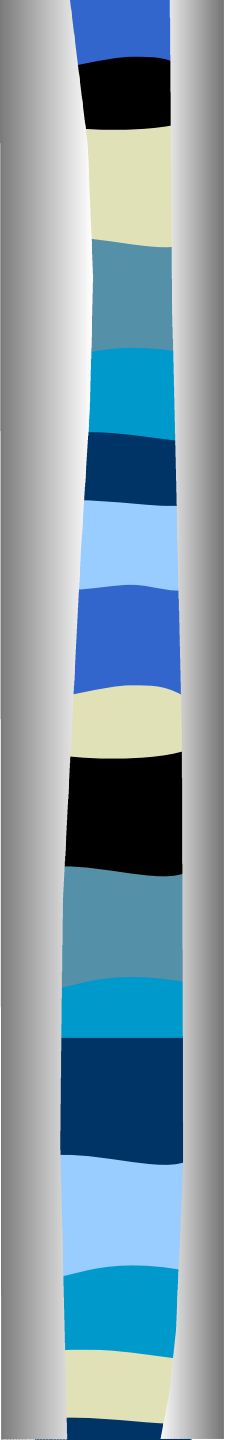
Enquête sur les bonnes pratiques

- Forte attente en guide de bonnes pratiques
 - Besoins pragmatiques en matière de guide pratiques : formats et métadonnées (usage et conservation)
- Peu de connaissances des normes et standards : effort communication et formation
 - Mettre en place des bonnes pratiques : inclure la conservation numérique dès le début du projet, impliquer les services informatiques, faire une copie sur site distant...



Les recommandations du Schéma numérique des bibliothèques en matière de conservation numérique

- Créer un consortium national (GIP, association, fondation...) pour coordonner la conservation numérique en bibliothèque
- Mutualisation des équipements
- Diffuser les bonnes pratiques (guides, formations)
- Cartographie des compétences
- Soutenir développement logiciel Open source
- Établir un modèle d'évaluation des coûts



3. Les bonnes pratiques :
les supports
les formats



3.1 Les supports

Les grands types de supports

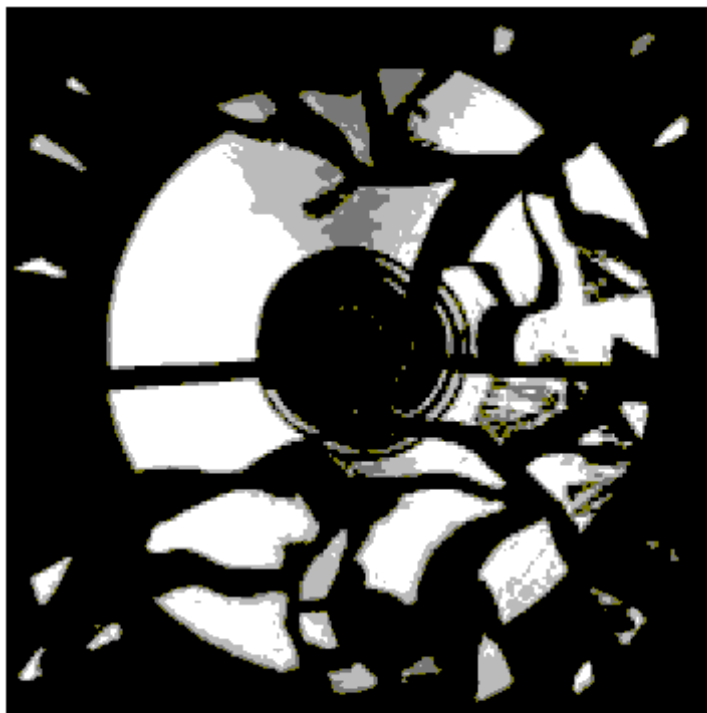
Supports magnétiques	Bandes magnétiques	LTO (400 Go) Support professionnel Qualité contrôlée Support assez utilisé Haute fiabilité Obsolescence rapide
	Disques magnétiques internes / externes	Disque dur Support professionnel Qualité contrôlée Support assez utilisé Risque technologique élevé (fiabilité et pérennité peu connues)
Supports optiques	CD-R (ou CD-WORM) = compact disc recordable : disque compact inscriptible. CD-RW = compact disc rewritable : disque compact réinscriptible.	Capacité : 700 Mo Support grand public Qualité inégale Risque technologique faible Fabrication doit répondre à la norme 42-011-2 gravure doit répondre à la norme ISO 9660
	DVD-R, DVD-RW = digital versatil disc, recordable, rewritable : disque numérique polyvalent, inscriptible, réinscriptible	Capacité : 4,7 Go Support grand public Qualité très inégale Risque technologique faible Double couche ou double face (capacité 17 Go)



Choix du support de stockage

- Les critères de choix du support de stockage se font en fonction :
 - du volume à stocker,
 - du coût,
 - de la durée de conservation,
 - des caractères critiques des données à conserver, de la confidentialité de l'information et de la sécurité,
 - de la fréquence d'utilisation de l'information,
 - de l'évolutivité (possibilités de migrations)

Les supports sont fragiles



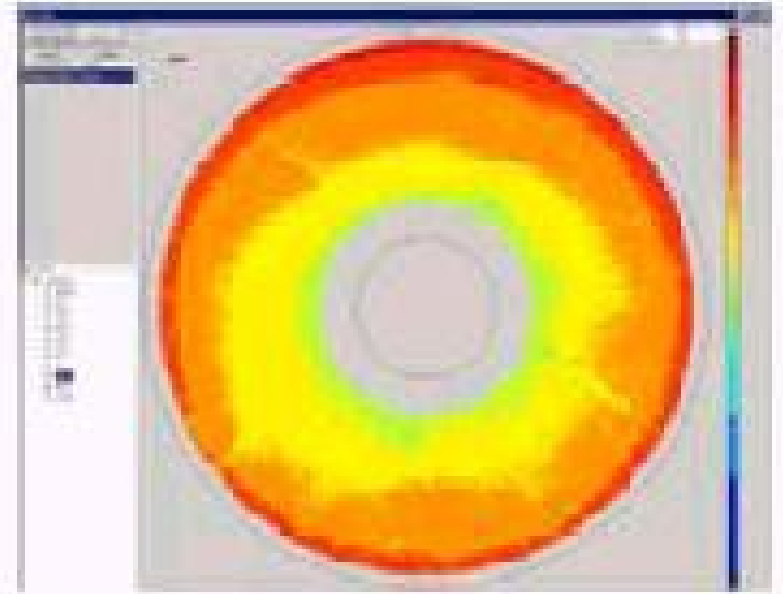
*Une très grande vitesse de rotation
peut provoquer l'éclatement du disque
dans le lecteur-enregistreur !*

© LAM - JMF

Des mauvaises conditions de stockage ou d'utilisation peuvent aboutir à une destruction rapide des supports :

- d'un mauvais environnement (température, humidité, lumière, champ magnétique),
- d'erreurs de manipulation (rayures, produits toxiques, effacement accidentel).

Les supports se dégradent



Les supports ont aussi tendance à se dégrader naturellement par exemple par oxydation.

En général, la durée de vie des supports ne dépasse guère 5 ou 10 ans.



Résultats des tests de CD-R conservés dans les services publics d'archives *enquête effectuée en 2005 par DAF et le LNE*

- Echantillonnage limité : un peu plus de 200 CD testés
- Plus de 20 % présente un risque important ou très élevé quant à leur lisibilité future
- Résultats « relativement préoccupants »
- Les CD sont de qualité variable et nécessitent des précautions indispensables :
 - vérification de la qualité des disques gravés en interne et éventuel renouvellement de l'équipement des ateliers de numérisation
 - mise en place d'une stratégie de recopie et de transferts de supports
- A l'exception de la médiathèque de l'agglomération troyenne, les BM sollicitées n'ont pas identifié de fichiers défectueux ou illisibles dans leurs collections de CD



Actions du ministère de la Culture sur les CD-R et DVD-R

- La DAF propose sur son [site](#) un ensemble de *Recommandations* relatives :
 - au choix des CD-R et DVD-R et graveurs du marché (analyse des couples CD/graveurs)
 - à la gravure, la conservation et l'évaluation des CD-R,
 - un guide pour la réalisation de la migration de stocks de CD-R

<http://www.archivesdefrance.culture.gouv.fr/gerer/archives-electroniques/stockage/>

- La constitution en 2005 d'un Réseau national de recherche sur la conservation des informations enregistrées sur disques optiques numériques (GIS-DON)

http://www.lne.fr/fr/r_et_d/gis-don/conservation-donnees-numeriques-gis-don.asp



Que faire ?

■ Il faut donc :

- choisir des supports robustes, normalisés, produits par plusieurs fabricants,
- procéder à des renouvellements réguliers de supports (renouvellement recommandé tous les 5 ans pour tous ces supports),
- surveiller les conditions de stockage, l'état des supports ainsi que leur disponibilité sur le marché,
- créer au moins une copie de sauvegarde de préférence sur un support différent (inconvenient : on double au minimum les volumes et les coûts de stockage)



Quelques recommandations relatives à la gravure, à la conservation et à l'évaluation des CD-R

- choisir de préférence une métallisation or
- modèles dédiés à la conservation longue durée
- remplir le disque à 90 % de sa capacité au maximum
- éviter de graver à trop grande ou à trop basse vitesse
- vérifier de temps en temps l'état des disques
- respecter les conditions de stockage (température entre 16 et 23°, humidité entre 30 et 50%, luminosité)
- graver au moins 2 exemplaires (un master d'archivage, une copie de consultation)
- stockage distant des exemplaires



Les périphériques de stockage

Les types de périphériques de stockage sont différents suivant le support utilisé et suivant la capacité de stockage de la plate forme :

- Pour les CD-R et DVD-R : jukebox
- Pour les bandes magnétiques : lecteur de bandes ou «auto-loader» pour les petits volumes, des librairies intégrant plusieurs drives pour les gros volumes
- Pour les disques durs : baies de stockage



Préconisations au niveau support et périphérique de stockage selon la volumétrie

- Utilisation de DVD-R et CD-R pour les très petits volumes, en raison de leur coût
- Utilisation de plate-forme automatisée uniquement à partir d'un certain volume, à partir de 2 To de données, mais cela dépend aussi des moyens humains.



Exemples de solutions possibles selon la volumétrie

■ Pour les petits volumes (50 Go) :

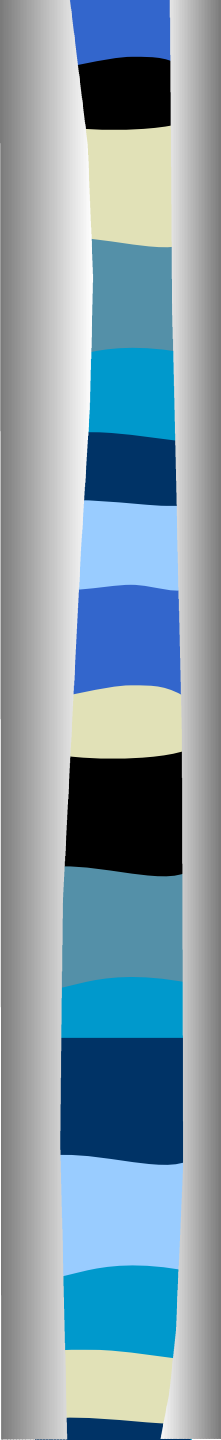
- utilisation de DVD-R ou CD-R pour le stockage (en 2 exemplaires)
- contrôle régulier des supports et une recopie à prévoir sous 5 à 7 ans
- Faire tester un échantillon de CD-R régulièrement (par société privée ou par logiciel dédié)
- Faire appel à un tiers-archivageur ?

■ Pour les volumes moyens (5 To)

- Copies multiples en diversifiant les technologies de stockage
- => conserver sur site 1 copie disque et 1 copie bande et sur site distant 1 copie bande
- archivage de la totalité des fonds sur serveur (disque dur) avec copie sur bande LTO (acquisition d'un auto-loader) en architecture réseau.



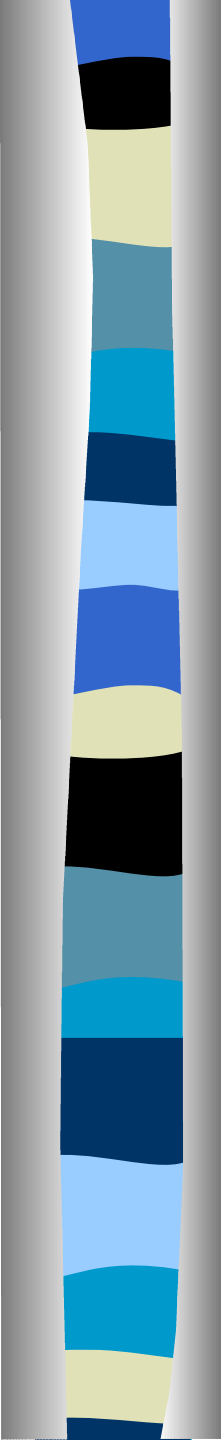
3.2 Les formats

- 
- Il est essentiel de choisir les bons formats de numérisation, tant pour la qualité de la numérisation que la préservation à long terme des documents numériques.
 - Il est important de distinguer le format de conservation du format de diffusion pour des raisons notamment d'accès aux documents en terme de temps d'affichage, de téléchargement et de distinction entre la consultation et la préservation.

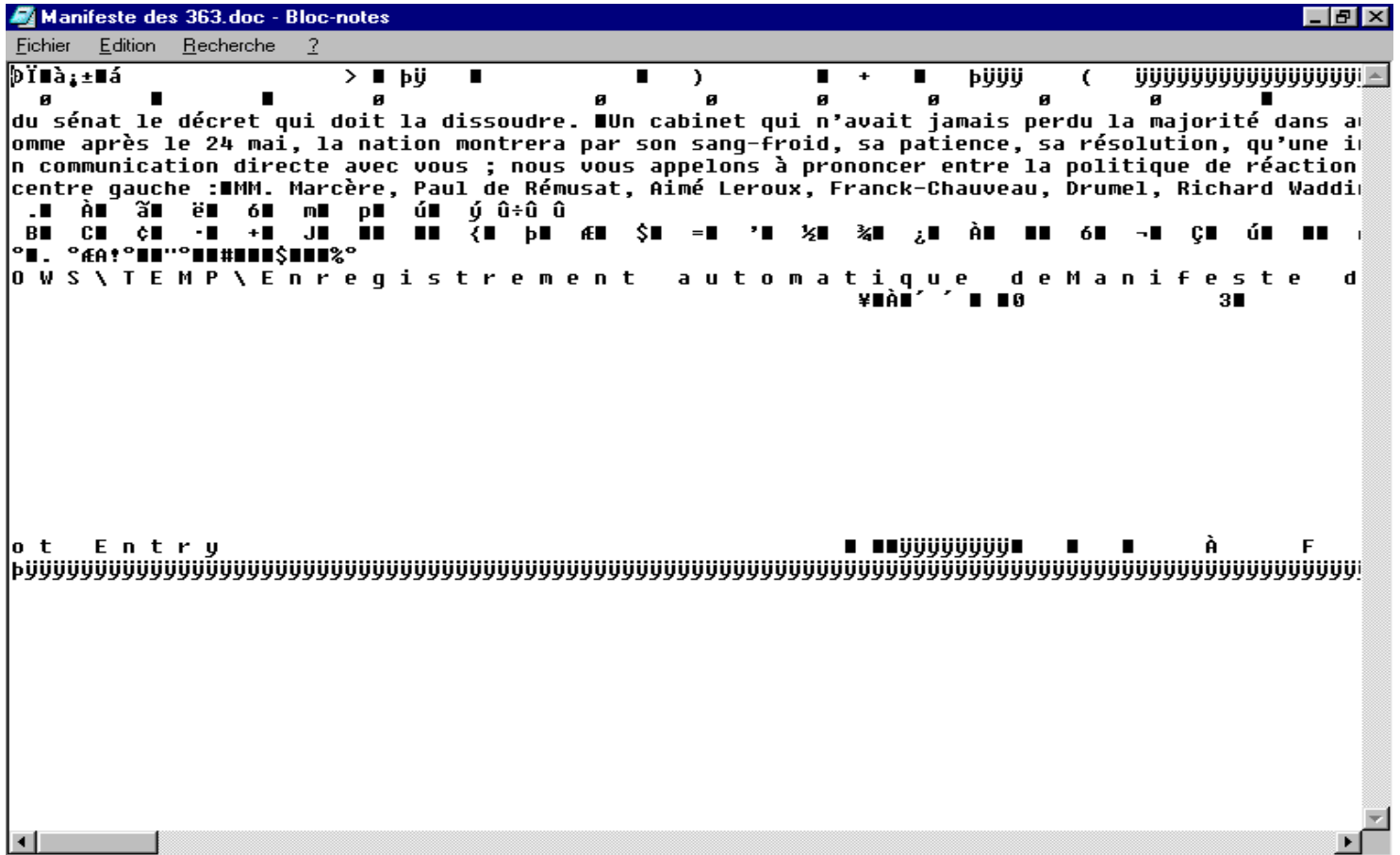


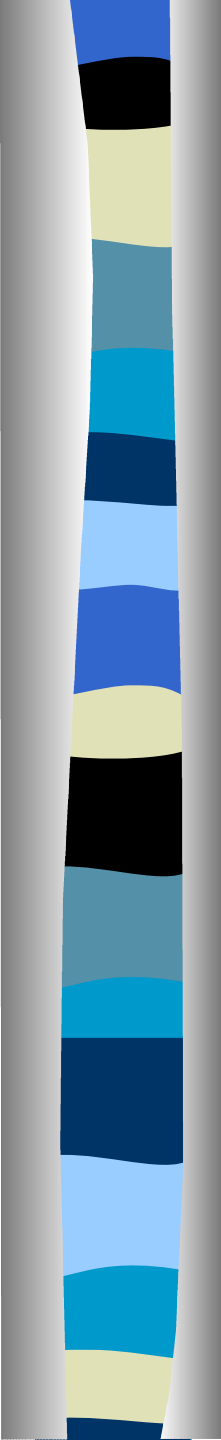
L 'information numérique

```
001111000100001000111110001111000100011001
001111010011100101010000100000010000110100
111101001100010011110101001000111101001000
100010001101000110010001100011011000110110
001100000011000000100010001111100100001001
101111011011100110101001101111011101010111
001000100000001000010011110000101111010001
100100111101001110010101000011111000111100
0010111101000010001111110
```

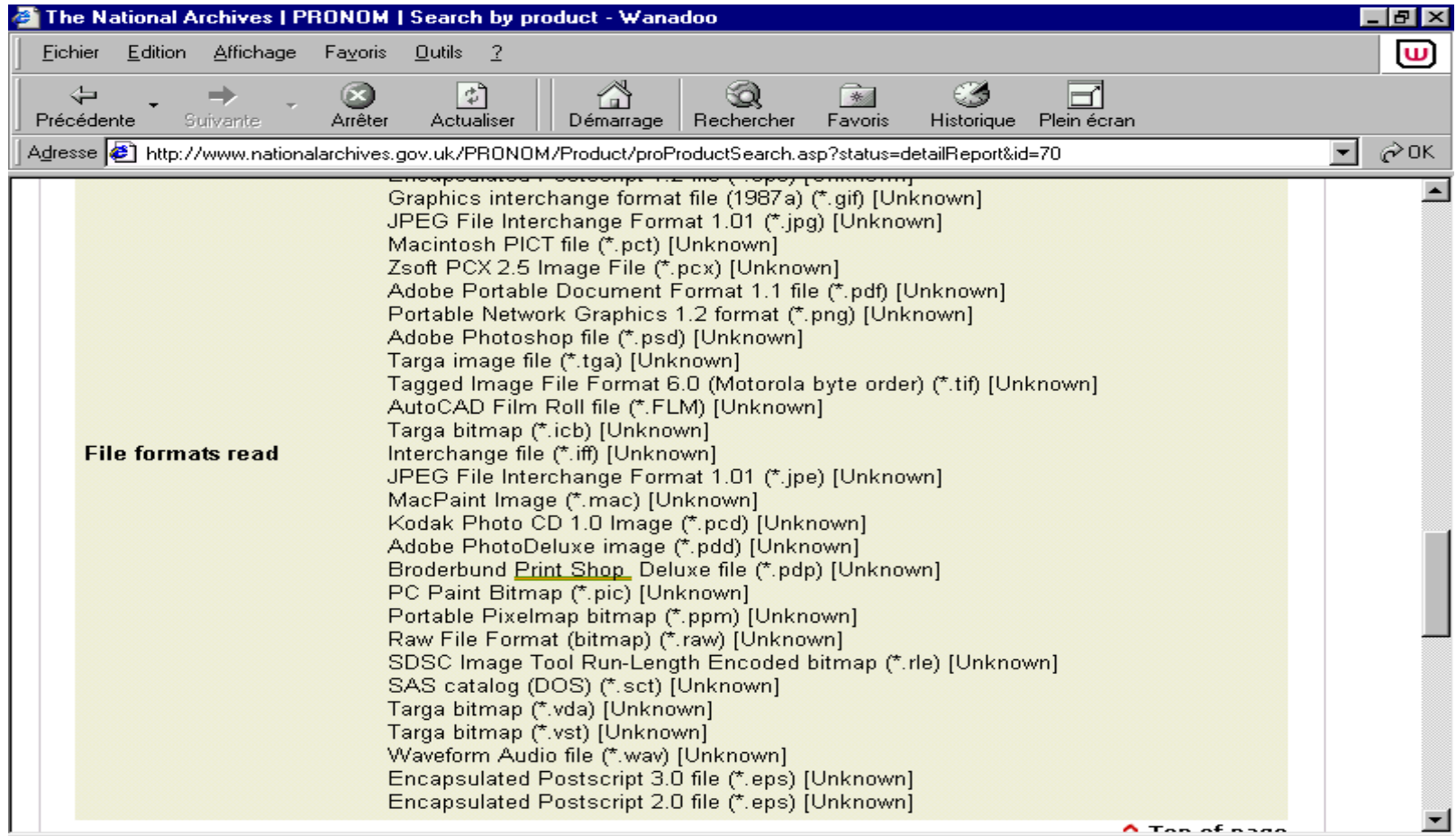
- 
- Les formats sont des systèmes de codage de l'information, qui permettent de passer d'informations intelligibles à des données binaires, et vice versa.
 - Ces opérations de codage et de décodage sont généralement assurées par des logiciels (par ex. Photoshop, Dreamweaver, Mozilla...)

Certains formats sont secrets



- 
- Si les logiciels évoluent ou disparaissent, le sens des données binaires stockées risque de devenir inaccessible.
 - C'est le cas lorsque le processus de codage est secret et que seul le logiciel est capable de le réaliser

Certains formats sont rares





Que faire ?

- Il est donc indispensable d'utiliser des formats dont les spécifications sont publiques, et de préférence des formats reconnus comme normes internationales.
- Il est également conseillé d'utiliser des formats courants.
- Des migrations seront nécessaires si les documents sont à conserver longtemps.



Le format doit être :

- un format **ouvert**, c'est à dire un format dont la documentation est librement accessible. Un format peut être secret mais il devra être suffisamment documenté pour en permettre une large utilisation
- un format largement **utilisé**,
- un format **normalisé** si possible
- un format **indépendant** (notamment des autres formats et des plateformes)



Le Référentiel général d'interopérabilité (RGI) préparé par la
Direction générale de la modernisation de l'état (DGME) [http://
www.references.modernisation.gouv.fr/rgi-interoperabilite](http://www.references.modernisation.gouv.fr/rgi-interoperabilite)

Type de document	Format recommandé pour la conservation
Documents	XML PDF/A PDF 1.7
Images	PNG JPEG TIFF
Son	MP3 WAV
Vidéo	MPEG-2 MPEG-4
Base de données	XML CSV



■ *Autre format :*

Le format JPEG2000 peut être considéré comme un bon compromis pour le master de conservation dans la mesure où

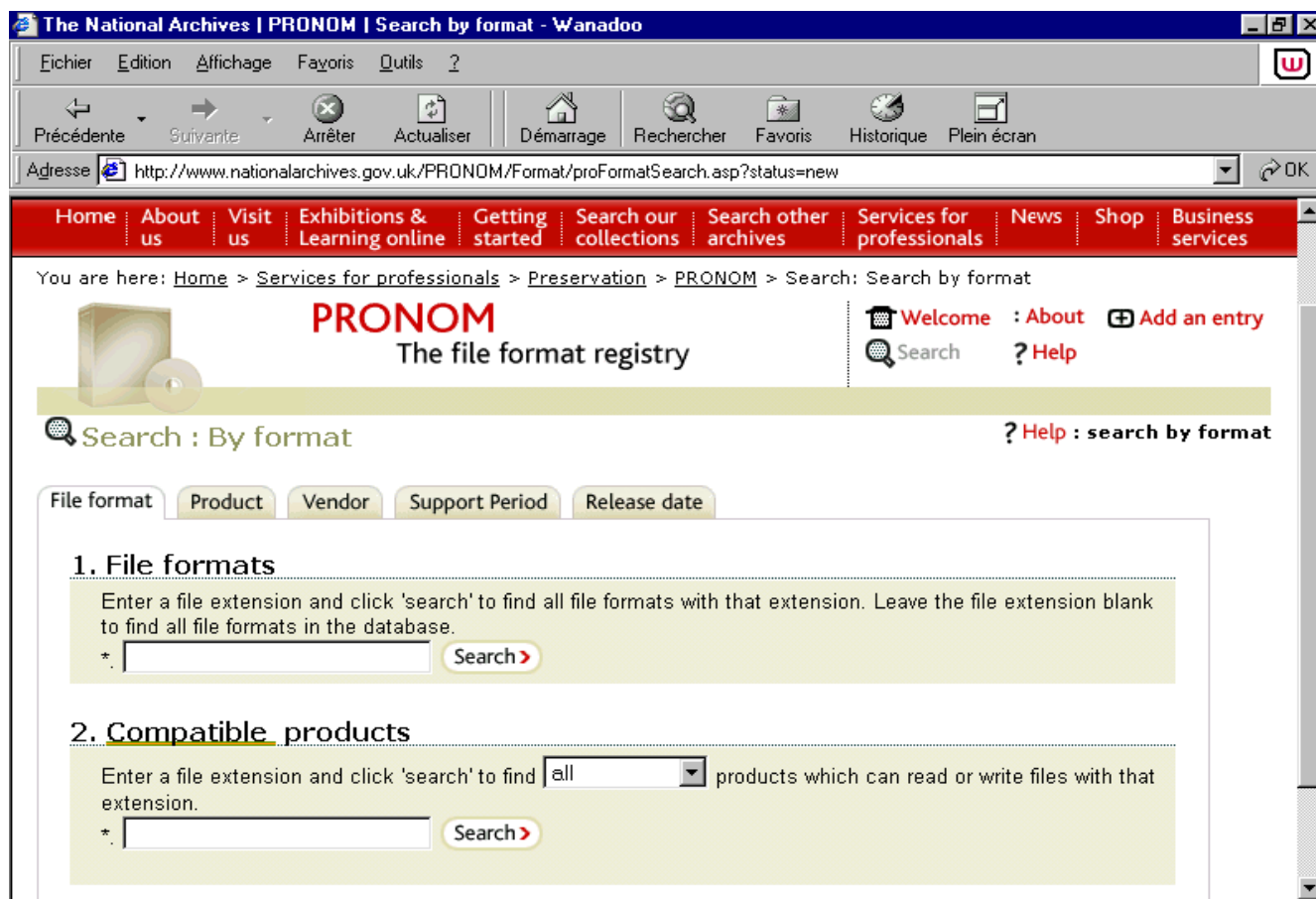
- permet une compression de meilleure qualité que le JPEG
- Il est utilisé par la Bibliothèque du Congrès pour certains documents (cartes...) depuis 2005, mais aussi pour la diffusion (mais là on se heurte au problème des visualiseurs), et par la Bibliothèque royale des Pays-Bas (Koninklijke Bibliotheek) et la BN Norvège pour la conservation. La BnF étudie la possibilité d'utiliser JPEG2000 en format d'archivage.

■ *Autres listes de formats :*

- Le Cines publie sur son site une [liste](#) des formats archivables sur sa plate forme PAC
- La BnF a réalisé un [mémento](#) sur les formats d'images, présentant une sélection de formats et leurs principales caractéristiques, permettant à l'utilisateur d'en avoir une connaissance générale.

Où trouver de l'information sur les formats ?

La base PRONOM



The screenshot shows a web browser window titled "The National Archives | PRONOM | Search by format - Wanadoo". The browser's address bar displays the URL: <http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.asp?status=new>. The browser's menu bar includes "Fichier", "Edition", "Affichage", "Favoris", and "Outils". The toolbar contains navigation buttons: "Précédente", "Suivante", "Arrêter", "Actualiser", "Démarrage", "Rechercher", "Favoris", "Historique", and "Plein écran".

The website's navigation menu is located at the top and includes: [Home](#), [About us](#), [Visit us](#), [Exhibitions & Learning online](#), [Getting started](#), [Search our collections](#), [Search other archives](#), [Services for professionals](#), [News](#), [Shop](#), and [Business services](#).

The main content area shows the breadcrumb path: "You are here: [Home](#) > [Services for professionals](#) > [Preservation](#) > [PRONOM](#) > Search: Search by format". Below this is the PRONOM logo, which consists of a CD-ROM icon and the text "PRONOM The file format registry". To the right of the logo are links for "Welcome", "About", "Add an entry", "Search", and "Help".

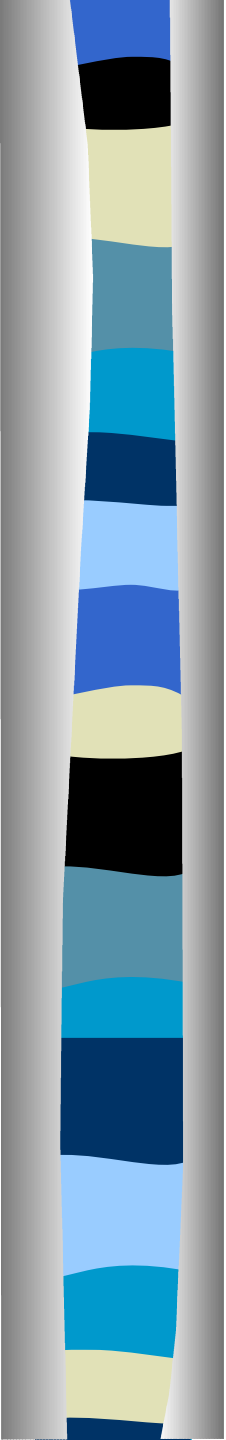
The search section is titled "Search : By format" and includes a "Help : search by format" link. Below the title are five tabs: "File format", "Product", "Vendor", "Support Period", and "Release date". The "File format" tab is currently selected.

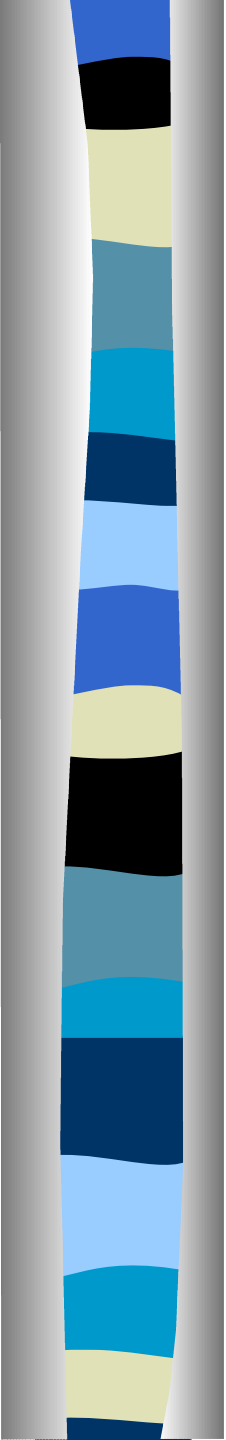
Under the "File format" tab, there are two search sections:

- 1. File formats**
Enter a file extension and click 'search' to find all file formats with that extension. Leave the file extension blank to find all file formats in the database.
*
- 2. Compatible products**
Enter a file extension and click 'search' to find products which can read or write files with that extension.
*



4. La politique de conservation numérique

- 
- Les problèmes rencontrés par les établissements dans leurs stratégies de préservation sont essentiellement les capacités insuffisantes de stockage, la perte de données, les données devenues illisibles et enfin l'expertise insuffisante.

- 
- Il convient de mettre en place d'**une politique de conservation numérique à son niveau**, comprenant notamment une politique claire de nommage des fichiers, de formats, de stockage et de migration
 - Elle précisera clairement les responsabilités des acteurs (bibliothèque ou archives, service informatique)



Rôle central des métadonnées

Les métadonnées sont des données sur les données.

Elles permettent de :

- retrouver les documents (métadonnées descriptives. Exemple : Dublin Core) ;
- gérer le cycle de vie et les droits d'accès aux documents (métadonnées administratives) ;
- connaître les caractéristiques techniques des documents en vue de migrations (métadonnées techniques ou de structure. Exemple : METS) ;
- enregistrer les traitements effectués sur les documents.

Il s'agit du poste le plus coûteux a priori (création manuelle / génération automatique ; contrôle et validation / mise à jour). Mais il est indispensable à la pérennité.



Une veille permanente

- Surveillance des supports
- Surveillance des formats
- Surveillance des procédures de gestion

Un document électronique qui dort court le risque d'être rapidement illisible

La conservation est une démarche active



L'importance de l'organisation

Une bibliothèque/ un service d'archive doit définir :

- les fonctions à remplir par les différents acteurs :
entrées, gestion de données descriptives, stockage ,
administration, accès, planification de la
pérennisation
- les règles organisationnelles et techniques à suivre



Politique de conservation numérique

Dans le cas de la numérisation, une politique de numérisation abordera par exemple :

- Le nommage des fichiers
- Les formats
- Le stockage (distinction support de consultation, support de conservation)
- La migration
- L'outil de suivi



Rôles et obligations des bibliothèques

- Préparer les données et les métadonnées
 - Fournir les éléments nécessaires à la construction de l'arborescence des fichiers
 - Fournir les informations techniques permettant de surveiller et de faire migrer les données
 - S'assurer que les informations sur les droits d'accès sont présentes
- Mesurer le facteur de risques lié au degré d'incomplétude des métadonnées
- Respecter les formats de données garantissant la pérennité et permettant de couvrir la majorité des besoins



Rôles et obligations des informaticiens

- la gestion de grands ensembles de fichiers stockés, dupliqués sur différents types de supports
- les technologies réseau à haut débit permettant de communiquer avec les usagers du service
- les technologies de stockage à grande capacité, robots de stockage..., les supports de stockage, leurs caractéristiques, leur fiabilité
- les moyens de surveillance de l'état des supports
- la capacité à maintenir en fonctionnement opérationnel un système ouvert très largement et à le faire évoluer en fonction des progrès de la technologie et des montées en charge.



Des concepts utiles : le modèle OAIS

Le modèle OAIS (Open Archival Information System) – norme ISO 14721

Un ensemble de concepts sur :

- L'information numérique et les métadonnées nécessaires à sa préservation
- Les différentes fonctions à remplir pour assurer la conservation (le versement des données, le stockage, la gestion des données, l'accès et la diffusion, mais aussi la migration d'informations numériques vers des supports et des formats nouveaux...)



Autres normes :

- Z-42-013 Archivage électronique - Spécifications relatives à la conception et à l'exploitation de systèmes informatiques en vue d'assurer la conservation et l'intégrité des documents stockés dans ces systèmes :

Cette norme fournit un ensemble de spécifications concernant les mesures à mettre en oeuvre pour l'enregistrement, le stockage et la restitution de documents électroniques afin d'assurer la conservation et l'intégrité de ceux-ci.

- Records management : Description du processus de gestion documentaire depuis la création du document.
Permettre la gestion des archives courantes et intermédiaires

(Voir aussi les [circulaires sur le Record Management](#) éditées par la DAF)



Solutions de tiers archivage

- Des offres de tiers-archivage numérique (ou archivage électronique) sont proposées à la fois par des établissements publics et par des entreprises privées.
- Du côté du privé, l'offre actuelle d'archivage numérique se structure principalement autour de la notion d'archivage sécurisé à valeur probante

Sur le site de la FNTC, liste des opérateurs et prestataires de services de confiance adhérents :

http://www.fntc.org/component/option,com_fntcmembre/task,college/Itemid,



Notions à prendre en compte pour choisir une solution pérenne de tiers-archivage :

- Une gestion sur le long terme : format d'archivage, migration, respect des normes en terme de archivage pérenne
- La taille des fichiers à archiver : possibilité d'archiver de grande masse tel que les masters.
- La place des métadonnées et l'indexation.

Les coûts 7 000 euros le To /an. Ces coûts varient en fonction des services rendus, de la fréquence de l'accès aux données, en fonction de la mutualisation possible



Tiers stockage (court et moyen terme) / Tiers archivage (moyen et long terme)

Un système d'archivage n'est pas

- un système de sauvegarde ou de secours,
- un rangement définitif des données

Il s'agit d'un ensemble de processus et de moyens concourant à la pérennisation des informations

Système de stockage = plateforme ou élément de base sur lequel s'appuie le système d'archivage

Avantages de l'externalisation du stockage

- recours à des acteurs disposant des compétences
- économies d'échelle, compte tenu de l'importance des coûts fixes (personnel, matériel)



A retenir...

- La numérisation ne peut se concevoir aujourd'hui sans une prise en compte de la conservation à long terme des documents numérisés
- Il n'y a pas de solution unique, mais des solutions en fonction d'objectifs
- Un document numérique est un document vivant, en constante évolution
- La préservation à long terme est avant tout affaire de bonnes pratiques



5. Exemples de plates-formes mises en place



CINES

Centre informatique national de l'Enseignement supérieur

- Etablissement public basé à Montpellier créé en 1999 sous tutelle du ministère de l'Enseignement supérieur et de la recherche
- 2 Principales missions :
 - Calcul intensif
 - Archivage pérenne des documents électroniques (scientifique, patrimoniale, administratif)
- Plus d'information : <http://www.cines.fr/>



CINES : mission d'archivage

- Mise en place du projet PAC (Plate-forme d'Archivage au Cines) :
 - dotation d'une plate-forme et d'un service numérique pérenne
- Dates clés
 - démarrage du projet en 2004
 - mise en production de la 1ère version en 2007
 - 2ème version en 2008 (capacité à gérer des grands volumes)
 - 1er To archivé en 2009



CINES

- Objectifs : solution performante, économique et sécurisée
- Contraintes :
 - solution générique, basée sur normes du domaine
 - facilité de veille technologique et de migration
- Données concernées :
 - scientifiques (résultats d'observation et de calcul)
 - patrimoniales
 - administratives (archives intermédiaires et courantes)

CINES : la plate-forme

■ **Volumétrie** : 1 To archivé, capacité 20 To extensible à 40 To

■ **Principes organisationnels** :

- Basée sur les standards du domaine : Modèle OAIS, protocole standard d'échange de données pour l'archivage, métadonnées Dublin Core, Identifiant pérenne (ARK), calcul d'empreinte numérique
- Liste des formats de fichier acceptés volontairement limitée : publiés, largement utilisés, normalisés et possibles

Type	Format
Texte	HTML, PDF, TXT, XML, ODT
Image	GIF, JPEG, TIFF, PNG, SVG
Audio	WAV, AIFF, AAC, VORBIS
Vidéo	MJPEG2000, MPEG4, THEORA



CINES

■ Qui peut verser ?

- Ensemble de la communauté de l'enseignement supérieur et de la recherche
- Avoir un corpus de documents à archiver reconnu d'intérêt national
- Avoir un système informatique interfaçable avec la plate-forme PAC

■ Comment ?

- Phase préliminaire (identification des informations, liste des données et métadonnées...)
- Phase de définition (contrat-convention)



CINES : les projets

■ En exploitation

- Centre national officiel d'archivage des thèses électroniques (STAR)
- Archivage des revues SHS diffusées par le portail PERSEE
- Les documents numérisés de la BIUM et BUPMC

■ En cours

- Archivage de documents sonores issues de la recherche dans le domaine de l'oral
- Archivage de cours universitaire de Canal-U
- Archivage des documents déposés dans les Archives Ouvertes (HAL)



BnF - SPAR

Le Système de Préservation et d'Archivage Réparti

- Début 2006 : définition des besoins (groupes de travail interne) et étude technique des solutions possibles
- Marché passé en 2007
- Réalisation par itération à partir de 2008

Plus d'information :

<http://www.bnf.fr/pages/zNavigat/frame/infopro.htm?ancre>



SPAR Enjeux : problématique technique

- **Croissance des volumes** : plus en plus de documents numérisés, archivage du Web, passage à la numérisation de masse (de 200 To en 2006 à 1 Po en 2009)
- **Difficultés de maîtrise des estimations**
 - Les besoins de stockage de documents numérisés sont dispersés
 - Insuffisance de la maîtrise des nouveaux programmes
- **Difficulté liée des achats successifs**
 - Un parc hétérogène, difficile à maintenir.



SPAR enjeux

- Production documentaire uniquement sous forme numérique : Dépôt légal : Web, Presse, etc.
- Dématérialisation des supports analogiques
Opérations de numérisation de masse des collections papiers
- Numériques suite à l'obsolescence voire la disparition des équipements de restitution
Numérisation de sauvegarde des collections audiovisuelles
- Disparition progressive des moyens de productions de microforme
Remplacement des moyens de capture analogique



Objectifs de la BnF

- Prendre la responsabilité de l'archivage 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 global pour réduire la diversité (matériel de stockage, formats utilisés, départements ...)
 - S'inspirer des bonnes pratiques et standards du domaine
- => Volonté de réaliser un système sur le modèle de l'OAIS.



Les filières concernées

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



Politique d'archivage

- Politique de versement : caractéristiques de la négociation et du protocole de versement
 - Qui négocie ?
 - Que négocie-t-on ? : droit, format, volumétrie, flux
- Politique d'archivage: caractéristiques de la conservation
 - Que conserve-t-on ? : données, métadonnées, systèmes de représentation... original, master, produits dérivés
 - Comment conserve-t-on ? : à l'identique (train original des bits), émulation, migration
 - Combien de temps conserve-t-on ?



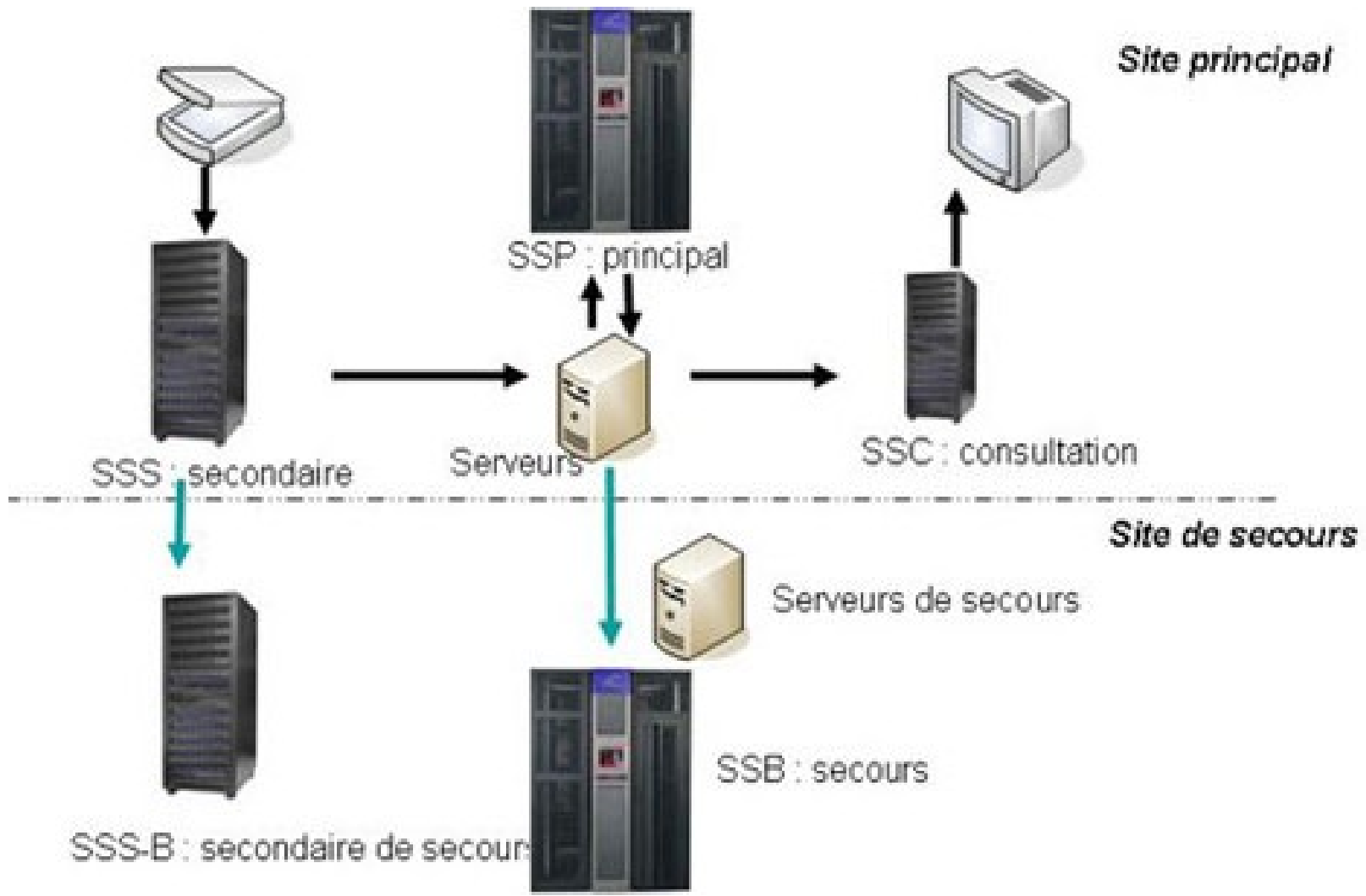
Politique d'archivage

- Politique d'accès: caractéristiques de l'accès
 - Avec ou sans restriction
 - Avec ou sans services supplémentaires (ex. veille des formats)
 - Immédiat ou différé
 - Direct (applicatif DSI) ou indirect
 - Sur place, à distance
 - Volumétrie des transactions



SPAR Principes organisationnels :

- Basée sur les standards du domaine :
Modèle OAIS, dictionnaire de métadonnées (METS, DC, PREMIS, ODRL), Identifiant pérenne (ARK), calcul d'empreinte numérique
- Infrastructure basée sur du matériel SUN, applicatif développé par Atos Origin basé sur Fedora-Commons



La salle informatique





Le Projet Pilae - DAF

- Etude sur les coûts d'une plate-forme d'archivage électronique à la demande de la DAF, en 2005.
 - Ouverture du nouveau centre des Archives nationales de Pierrefitte sur Seine et de sa plate-forme d'archivage électronique
 - Développement de l'administration électronique
- En 2006, lancement d'un marché pour la réalisation d'une plate-forme pilote
 - Un noyau réutilisable par les collectivités territoriales
 - Implémentation des standards
 - Utilisation du pilote par les Archives nationales
- Plate-forme attendue pour 2012



Réalisation d'un pilote national de plateforme d'archivage électronique sur la base d'outils existants

- un coffre-fort électronique (disposant en interne d'outils logiciels assurant les mécanismes de scellement, signature, horodatage, journalisation ad-hoc propres à assurer le maintien de l'intégrité des archives prises en charge) : édité par la société Security.com
- une infrastructure de stockage IBM assurant les fonctions ad-hoc de sécurisation des supports, de réplication, de sauvegarde



Projet Pilae

- Repose sur la norme OAIS
 - Description fonctionnelle du pilote
 - Structuration des données (distinction entre les paquets d'archives à soumettre / à conserver / à diffuser)
- Repose sur les normes de description des archives (norme ISAD-G)
- Repose sur les référentiels (politique d'archivage, standard d'échange de données pour l'archivage)



Projet Pilae

Types de données concernées

- données extraites de bases de données
- documents bureautiques issus des outils de traitement de texte/tableurs/présentations
- documents issus de GED et décrits par une base de donnée
- messageries électroniques
- flux de données sécurisées...



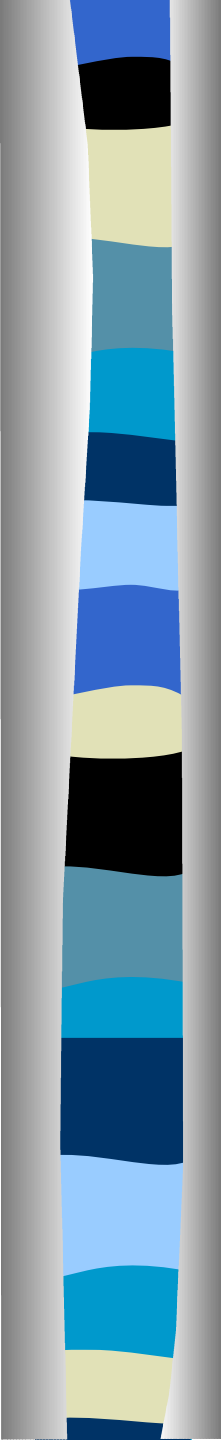
Vue d'ensemble

■ Objectif

- recevoir, valider, conserver, communiquer les données et documents nativement numériques produits par les services centraux de l'Etat
- pour l'instant les utilisateurs de l'application sont :
 - les archivistes des services d'archives dans les ministères
 - les Archives nationales qui reçoivent, contrôlent, valident et conservent

■ Non ouvert :

- aux producteurs
- au grand public



Au-delà des archives nationales, le modèle de PIL@E pourra être utilisé et repris dans d'autres environnements :

- chez les producteurs
- chez les archivistes
- dans les ministères, dans les collectivités territoriales (plateformes d'archivage intermédiaire)



Médiathèque de l'agglomération troyenne

- Mise en place d'un système de stockage électronique en interne (2005-2008)
 - 3 baies des stockage, 1 robot de sauvegarde
- Coûts :
 - Total investissement pour espace prévisionnel de stockage de 27 To (27000 Go) : 173 000 € TTC (installation et paramétrage sur site inclus)
 - Coût de maintenance pour 3 ans : environ 30 000 € TTC
 - Coût de maintenance en 2009 : 13 000 €



La BM de Lille

- Solution externe
- Coût de l'hébergement de la bibliothèque numérique :
 - Frais d'accès : 6 500 € HT
 - 1ère année : 9 500 € / autres années : 3 000 €
- Prestations choisies :
 - 2 serveurs, Espace disque utile : 146 Go
 - Sauvegarde totale
 - Administration complète / Supervision / Accès SSH / FTP ...



Plates-formes mutualisées

■ Besançon : Projet commun BM/BU

- Archivage sur serveur par le service informatique de la Ville
- L'enregistrement se fait via le réseau de la ville, sur 2 sites distincts.
- Volumétrie de 4 To, à terme nécessité de 15 To

■ AD Yvelines (CG78) / BM Versailles

- Numérisation par la BM qui conserve les fichiers de diffusion
- Mise en ligne et archivage par l'AD qui conserve les fichiers de conservation

Pour en savoir plus :

■ Ouvrage de référence :

Françoise Banat-Berger, Laurent Duploux et Claude Huc. *L'archivage numérique à long terme : les débuts de la maturité ?*, Paris, La documentation française, 2009. (Manuels et guides pratiques / Direction des archives de France)

■ Sites de référence :

— Site « Numérisation du patrimoine culturel »

<http://www.culture.gouv.fr/culture/mrt/numerisation/>



Conservation à long terme des documents numérisés

un guide technique publié en 2008 et réalisé en coordination avec le comité de pilotage de numérisation (normes et standards, politique de conservation, production des documents numérisés, supports et infrastructure de stockage, questions juridiques, projets et expériences)

<http://www.culture.gouv.fr/culture/mrt/numerisation/fr/technique/documents/conservation.pdf>

— Site de la Bibliothèque nationale de France

<http://www.bnf.fr/fr/professionnels/numerisation.html>



Pour en savoir plus :

- Site d'information des Archives de France, Archives électroniques

<http://www.archivesdefrance.culture.gouv.fr/gerer/archives-electroniques/>

- Site du groupe PIN (Pérennisation de l'information numérique) de l'association Aristote

<http://www.aristote.asso.fr/PIN/index.html>

- site de la Bibliothèque de l'université Cornell : didacticiel

<http://www.library.cornell.edu/preservation/tutorial-french/contents.html>