

Thierry Claerr et Isabelle Westeel (dir.)

## Numériser et mettre en ligne

Presses de l'enssib

---

# Conserver les documents numérisés

Patricia Le Galèze et Isabelle Westeel

---

DOI : 10.4000/books.pressesenssib.439  
Éditeur : Presses de l'enssib  
Lieu d'édition : Presses de l'enssib  
Année d'édition : 2010  
Date de mise en ligne : 4 avril 2017  
Collection : La Boîte à outils  
ISBN électronique : 9782375460320



<http://books.openedition.org>

### Référence électronique

LE GALÈZE, Patricia ; WESTEEL, Isabelle. *Conserver les documents numérisés* In : *Numériser et mettre en ligne* [en ligne]. Villeurbanne : Presses de l'enssib, 2010 (généré le 01 février 2021). Disponible sur Internet : <<http://books.openedition.org/pressesenssib/439>>. ISBN : 9782375460320. DOI : <https://doi.org/10.4000/books.pressesenssib.439>.

---

## CONSERVER LES DOCUMENTS NUMÉRISÉS

*par Patricia  
Le Galèze  
et Isabelle  
Westeel*

### CONTEXTE ET ENJEUX

\*\*\*\*\*

La pérennité de l'information numérique est une question complexe. Il n'existe pas de solutions prêtes à l'emploi dans ce domaine. En effet les technologies de préservation sont à renouveler régulièrement, le matériel informatique étant en évolution perpétuelle. L'accroissement exponentiel de l'information numérique est une difficulté supplémentaire.

Dans un souci de préservation des originaux comme de maîtrise des coûts, la préservation des documents numérisés cherche à éviter de devoir numériser à nouveau les originaux. Il s'agit de pérenniser l'investissement réalisé. La préservation numérique doit donc être prise en compte dès le début d'un projet de numérisation. Ceci afin de protéger les données numériques, mais aussi de les garder accessibles le plus longtemps possible.

La préservation numérique se heurte cependant à trois risques majeurs :

- l'obsolescence des applications informatiques et des formats de fichiers, qui incitent à garder les outils ou à utiliser la migration des données ou l'émulation des systèmes ;
- la fragilité et la durée de vie relativement faible des supports ;
- la difficulté de retrouver des fichiers informatiques s'ils n'ont pas été bien décrits à l'origine et bien suivis. Pour cela, les services informatiques ont besoin d'un maximum d'informations, et donc de métadonnées. Même s'il s'agit d'informations coûteuses, elles se révèlent indispensables à la pérennité.

L'enjeu est double :

- la conservation du support ;
- l'accès aux données en fonction des formats, de la structuration du document et de l'encodage de l'information 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 donc de mettre en place une politique de conservation numérique à son niveau, comprenant notamment une politique claire de nommage des fichiers, de formats, de stockage et de migration :

- préparer des données et des métadonnées ;
- rassembler 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 risque lié au degré d'incomplétude des métadonnées ;
- respecter les formats de données garantissant la pérennité et permettant de couvrir la majorité des besoins.

Il sera également nécessaire d'associer autant que possible les services informatiques, y compris des informaticiens spécialisés et des prestataires qui se verront attribuer les rôles et les obligations suivants :

- 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, supports de stockage avec leurs caractéristiques et 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.

## **RECOMMANDATIONS TECHNIQUES ET CONSEILS POUR UNE NUMÉRISATION DURABLE**

\*\*\*\*\*

En matière de conservation numérique, le Référentiel général d'interopérabilité (RGI)<sup>68</sup> définit des recommandations réparties entre différents volets :

- volet organisationnel : politique d'archivage sécurisé dans le secteur public ;
- volet sémantique : schéma d'échange de données pour l'archivage (visant à faciliter les échanges entre services producteurs d'archives et services d'archives) ;
- volet technique : préconisations en matière de formats pour la conservation ainsi qu'en matière de supports.

Ces recommandations constituent les objectifs à atteindre pour favoriser l'interopérabilité. Le RGI n'émet que des préconisations en matière de normes, de standards et de bonnes pratiques. Aucune norme ni standard n'est interdite ou déconseillée.

---

68. RGI Version 1.0, 12 juin 2009, version en vigueur. [En ligne]  
 < <http://www.references.modernisation.gouv.fr/rgi-interoperabilite> > (consulté le 22 janvier 2010) et  
 < [http://references.modernisation.gouv.fr/sites/default/files/RGI\\_Version1\\_0.pdf](http://references.modernisation.gouv.fr/sites/default/files/RGI_Version1_0.pdf) > (consulté le 22 janvier 2010).

### sites de référence

- Ministère de la Culture et de la Communication. *Conservation à long terme des documents numérisés* : guide technique publié en 2008 et réalisé en coordination avec le comité de pilotage numérisation : normes et standards, politique de conservation, production des documents numérisés, supports et infrastructure de stockage, questions juridiques, projets et expériences. [En ligne] < <http://www.culture.gouv.fr/culture/mrt/numerisation/fr/technique/documents/conservation.pdf> > (consulté le 22 janvier 2010).
- Site de l'association Aristote. *Pérennisation des Informations Numériques (PIN)*. [En ligne] < <http://www.aristote.asso.fr/PIN/> > (consulté le 22 janvier 2010). Créé en 2000, le groupe de travail et d'échanges français PIN (Pérennisation de l'Information Numérique) est un lieu d'échange de savoirs, de savoir-faire et d'actions dans le domaine de la pérennisation des informations représentées sous forme numérique.
- Site de la Direction générale de modernisation de l'État (DGME) : la DGME regroupe, depuis le 1<sup>er</sup> janvier 2006, l'ensemble des services œuvrant pour la réforme et la modernisation de l'État, dont l'ancienne agence pour le développement de l'administration électronique (ADAE), devenue au sein de la DGME, le service pour le développement de l'administration électronique (SDAE). C'est avec ce service que la direction des Archives de France met en œuvre l'action 103 du programme ADELE, sur l'archivage électronique. [En ligne] < <http://www.modernisation.gouv.fr/> > (consulté le 22 janvier 2010).

### FORMATS

Il est non seulement essentiel de choisir les bons formats de numérisation, tant pour la qualité de la numérisation que pour la préservation à long terme des documents numériques, mais il est aussi 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.

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. Adobe® Photoshop® ...). Certains formats sont secrets (cf. Microsoft® Word), d'autres sont rares. 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 d'origine est capable de le réaliser. 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.

Le format doit donc être :

- un format ouvert, c'est-à-dire un format dont la documentation est librement accessible (un format peut être propriétaire 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).

type de document	format	commentaires
Documents	XML	Format dont les spécifications sont normalisées par l'ISO
	PDF/A	ISO 19005-1 Archivage des documents bureautiques statiques non révisables
	PDF 1.7	Conservation des documents bureautiques dynamiques
Images	GIF	
	PNG	ISO 15948
	JPEG	ISO 10918
	TIFF	
	DNG	
Son	MP3	ISO 11172-3 Conservation de séquences sonores de qualité ordinaire
	WAV	Codage PCM Sans compression - Conservation de séquences sonores de haute qualité sonore
Vidéo	MPEG-2	ISO 13818 Conservation de séquences vidéo en basse définition
	MPEG-4	ISO 14496 Conservation de séquences vidéo en haute définition
Objet graphique en 2D	SVG 1.1	Recommandation W3C
Objet et univers 3D	X3D	Normalisé par l'ISO
Bases de données	XML CSV	

Utilisé par la Bibliothèque du Congrès et par la Bibliothèque royale des Pays-Bas, le format JPEG 2000 peut être considéré comme un bon compromis pour le master de conservation. La BnF étudie la possibilité d'utiliser ce format pour l'archivage même si ce format qui n'est pas supporté par tous les visualiseurs, connaît des difficultés dans le domaine de la diffusion.

### sites de référence

- De même que le Centre informatique de l'enseignement supérieur (CINES) publie sur son site une liste de formats archivables sur sa plateforme PAC<sup>69</sup>, la BnF a réalisé un mémento sur les formats d'images. [En ligne]  
< [http://www.bnf.fr/fr/professionnels/num\\_metadonnees/s.num\\_metadonnees\\_formats.html?first\\_Art=non](http://www.bnf.fr/fr/professionnels/num_metadonnees/s.num_metadonnees_formats.html?first_Art=non) > (consulté le 22 février 2010).
- La base PRONOM des Archives nationales du Royaume-Uni permet de mettre à disposition des informations sur les formats de fichiers, les produits logiciels et autres composants techniques nécessaires pour assurer la préservation à long terme des documents électroniques. [En ligne]  
< <http://www.nationalarchives.gov.uk/PRONOM/> > (consulté le 22 janvier 2010).
- La Bibliothèque du Congrès propose des pages consacrées à l'étude et à l'évaluation des formats. [En ligne]  
< <http://www.digitalpreservation.gov/formats/> > (consulté le 22 janvier 2010).

## NORMES ET STANDARDS

Des normes, des standards et des référentiels ont vu le jour et commencent à trouver des applications concrètes au niveau national.

### normes généralistes

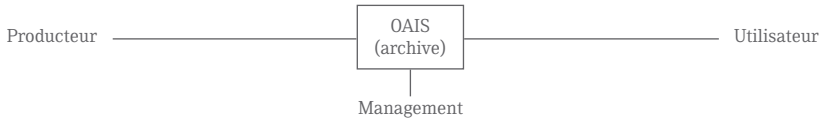
**OAIS (Open Archival Information System/système ouvert d'archivage d'information)** : norme ISO 14721 : 2003<sup>70</sup> (Système de transfert des informations et données spatiales – Système ouvert d'archivage de l'information – Modèle de référence)

69. Cines. [En ligne] < <http://www.cines.fr/spip.php?article685> > (consulté le 22 janvier 2010).

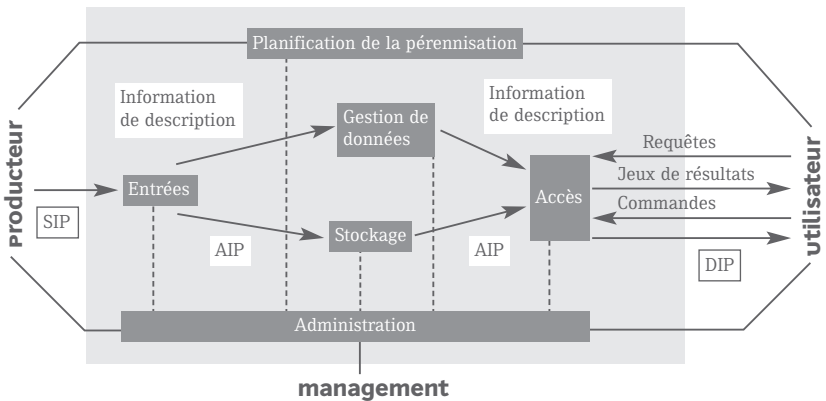
70. [En ligne] < [http://www.iso.org/iso/fr/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=24683](http://www.iso.org/iso/fr/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24683) > (consulté le 22 janvier 2010). Norme en ligne < [http://www.aristote.asso.fr/PIN/documents/OAIS\\_650\\_Ob1.pdf](http://www.aristote.asso.fr/PIN/documents/OAIS_650_Ob1.pdf) > (consulté le 22 janvier 2010).



**SCHÉMA** **Modèle d'environnement d'un OAIS (traduction française de l'OAIS p. 2-2)**



**SCHÉMA** **Entités fonctionnelles OAIS (traduction française de l'OAIS p. 4-1)**



Norme généraliste mise au point par les principaux centres d'études spatiales du monde dont le Centre national d'études spatiales (CNES) en France, OAIS est un modèle de référence<sup>71</sup> pour la définition des concepts indispensables à la compréhension de l'archivage numérique : l'organisation, l'architecture et le mode de fonctionnement des archives existantes et futures. Ce système a été adopté dès 1999 par tous les programmes de bibliothèques s'intéressant à la préservation et à l'accès à long terme et par les entrepôts institutionnels.

71. Traduction française, en cours de normalisation, consultable sur le site PIN. [En ligne] < [http://www.aris-tote.asso.fr/PIN/documents/norme\\_oais\\_version\\_francaise.pdf](http://www.aris-tote.asso.fr/PIN/documents/norme_oais_version_francaise.pdf) > (consulté le 22 janvier 2010).

Ce modèle aborde non seulement toute une série de fonctions liées à la conservation d'informations archivées comprenant 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, les modèles de données utilisés pour représenter l'information, le rôle du logiciel dans la conservation de l'information, ainsi que l'échange d'informations numériques entre archives. Il identifie les interfaces internes et externes des fonctions d'archivage ainsi qu'un certain nombre de services de haut niveau relatifs à ces interfaces. Il détermine les responsabilités minimales auxquelles une archive doit souscrire pour être qualifiée d'OAIS.

Il définit :

- les objets d'information, les métadonnées nécessaires à leur préservation et l'organisation à mettre en place pour leur archivage, leur conservation et leur communication ;
- un vocabulaire et un ensemble de concepts permettant d'appréhender, de façon globale et complète, la question de l'archivage à long terme des données sous forme numérique quels que soient les secteurs professionnels concernés ;
- deux modèles complémentaires : un modèle d'information et un modèle fonctionnel détaillé.

Il propose également une classification des types de migration et des différents modes de coopération possibles entre archives. La norme ne préjuge pas d'une implémentation particulière et constitue un cadre de développement pour des normes complémentaires.

**La norme NF Z 42-013 (mars 2009) : 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.** La première version de la norme datait de 2001. Une nouvelle version, élaborée dans le cadre de la commission de normalisation 171

(« Archivage - Cycle de vie du document et sécurité des données »), a été adoptée le 4 février 2009 par le conseil d'administration de l'AFNOR. Cette norme fournit un ensemble de spécifications concernant les mesures à mettre en œuvre pour l'enregistrement, le stockage et la restitution de documents électroniques afin d'assurer la conservation et l'intégrité de ceux-ci. Cette nouvelle version reconnaît les supports réinscriptibles grâce aux garanties et protections offertes par les moyens technologiques reconnus tels la signature électronique et l'horodatage. Elle étend le périmètre d'application aux documents sonores et audiovisuels. Elle définit l'organisation des audits des systèmes d'archivage, ainsi que l'intervention des tiers dans le processus d'archivage (tiers archiveurs et autres prestataires de services).

Le **Records Management : norme ISO 15489** (*Information and documentation – Records management* octobre 2001, en français en 2002) qui a pour objectif la gestion des archives courantes et intermédiaires. Cette norme a pour « finalité de permettre à l'organisme de disposer à tout instant du document dont il a besoin pour conduire ses activités, répondre aux exigences légales et réglementaires, et se protéger en cas de contentieux<sup>72</sup> ». La norme définit très précisément les processus à mettre en place depuis la création du document. Elle permet de prendre des décisions à propos des étapes de création, de capture, de contrôle, de conservation et d'archivage des documents<sup>73</sup>.

Les **spécifications Moreq 2** : modèle européen pour la mise en œuvre du Records Management pour les documents électroniques et précisant les exigences de l'archivage électronique à des fins de preuve, disponible sur < <http://www.archivesdefrance.culture.gouv.fr/gerer/records-management-et-collecte/> >.

---

72. Groupe métier AAF – ADBS « Records management », Analyse de la norme ISO 15489 au regard des pratiques archivistiques françaises. Version 2 – 8 avril 2005, p. 4. [En ligne] < <http://www.adbs.fr/groupe-de-travail-aaf-adbs-records-management-29392.htm?RH=ACCUEIL> > (consulté le 4 mars 2010).

73. Voir le site des Archives de France. [En ligne] < <http://www.archivesdefrance.culture.gouv.fr/gerer/records-management-et-collecte/> > (consulté le 22 janvier 2010).

**standards et normes relatifs à l'interface Producteur-Archive**

PAIMAS (*Producer-Archive Interface Methodology Abstract Standard*) : norme ISO 20652. Cette norme concerne la méthodologie de conduite de l'interface Producteur-Archive et permet de disposer d'un guide standard.

PAIS (*Producer-Archive Interface Specification*) – XFDU (*XML Formatted Data Units*) spécifie des mécanismes formels (automatisation) permettant la définition, le transfert et la validation des documents à archiver et permet de mettre en œuvre PAIMAS.

Le protocole standard d'échange DGME/DAF<sup>74</sup> spécifie les processus et la mise en œuvre des transferts depuis le producteur vers l'utilisateur.

**certification/audit**

TRAC (*Trustworthy Repositories Audit and Certification*) définit des exigences de certification des archives en vue de la normalisation de ces exigences par l'ISO. Il fournit une liste de critères pour l'étude de la fiabilité des capacités d'une Archive à conserver des documents numériques de façon pérenne.

DRAMBORA (*Digital Repository Audit method Based on Risk Assessment*) est une méthodologie d'audit d'archives numériques basée sur l'évaluation des risques liés à l'archivage de documents numériques. Elle fournit des outils pour faciliter l'audit interne d'une archive en vue d'identifier ses forces et faiblesses.

**Pour en savoir plus**

Site de l'association Aristote. Pérennisation des Informations

Numériques (PIN). [En ligne]

< <http://www.aristote.asso.fr/PIN/> > (consulté le 22 janvier 2010).

---

74. Voir le site des Archives de France. [En ligne] < <http://www.archivesdefrance.culture.gouv.fr/gerer/archives-electroniques/standard/> > (consulté le 22 janvier 2010).

## **métadonnées**

Les métadonnées techniques permettent d'identifier et de caractériser les formats de représentation de l'information ou formats de données. (Pour une présentation complète des métadonnées, se reporter à la Partie V). Le dictionnaire PREMIS basé sur le format XML est, dans ce domaine, un outil très précis. Il vise à établir, à partir du modèle OAIS, une liste des métadonnées gérées dans le processus de conservation numérique<sup>75</sup>. Il permet la spécification d'une stratégie pour le codage, le stockage et la gestion de métadonnées de préservation pour les systèmes d'archivage.

## **RECOMMANDATIONS PRATIQUES POUR LE STOCKAGE**

\*\*\*\*\*

### **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 l'accessibilité ;
- de la fréquence d'utilisation de l'information ;
- de la fiabilité par rapport aux mécanismes d'enregistrement, de la stabilité du support, de sa robustesse et de la large diffusion de la technologie ;
- de l'évolutivité (possibilités de migrations).

---

75. PREMIS : *Preservation metadata and the OAIS information model*. Groupe de travail OCLC/RLG sur les métadonnées de conservation, juin 2002. PREMIS Data Dictionary, mai 2005. Maintenance et suivi des travaux assurés par la Bibliothèque du Congrès. [En ligne] < <http://www.loc.gov/standards/premis> > (consulté le 22 janvier 2010).

Les supports sont variés (supports optiques, supports magnétiques) mais il n'y a pas aujourd'hui de support pérenne. Même si on constate une baisse continue du prix de ces supports, on se heurte à l'augmentation régulière des volumes.

Les supports sont fragiles et se détériorent. Des mauvaises conditions de stockage ou d'utilisation peuvent aboutir à une destruction rapide des supports, que ce soit en raison :

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

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.

Il faut donc :

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

**TABLEAU SUPPORTS DE STOCKAGE**

SUPPORTS DE STOCKAGE		
Supports magnétiques	Bandes magnétiques	LTO Support professionnel Qualité contrôlée Support assez utilisé Risque technologique élevé
	Disques magnétiques internes/externes	Disque dur Support professionnel Qualité contrôlée Support assez utilisé Risque technologique élevé
Supports optiques	CD = <i>compact disc</i> : disque compact non inscriptible i.e. pressé ou gravé en usine. CD-R (ou CD-WORM) = <i>compact disc recordable</i> : disque compact inscriptible. CD-RW = <i>compact disc rewritable</i> : disque compact réinscriptible.	Capacité : 700 Mo Support grand public Qualité inégale Risque technologique faible La fabrication doit répondre à la norme 42-011-2 La gravure doit répondre à la norme ISO 9660
	DVD, DVD-R, DVD-RW = <i>digital versatil disc, recordable, rewritable</i> : 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

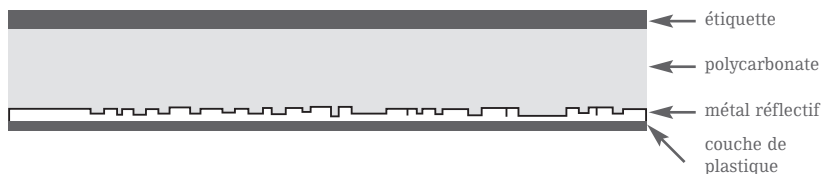
**LES SUPPORTS OPTIQUES**

Il est important que ce type de supports respecte :

- la norme NF Z 42-013 concernant les conditions et procédures de conservation (tests, copies régulières et transferts de supports) qui doivent être décrites en détail ;

- la norme d'enregistrement ISO 9660 (niveau 2). Cette norme garantit la lisibilité des noms sur tout type de plateforme. Les noms de répertoires et de fichiers (y compris le séparateur et l'extension) sont limités à 31 caractères qui doivent obligatoirement être des lettres capitales, des chiffres et des tirets bas ( \_ ). L'arborescence des fichiers est limitée à huit niveaux de répertoires.

#### SCHÉMA *Composition d'un CD ou DVD*



Pour rappel, la fabrication d'un CD-R doit répondre à la norme 42-011-2 <sup>76</sup>. Un CD ou DVD, quel que soit son type, se compose en général d'une couche métallique réfléchissante (or, argent ou aluminium, suivant la qualité du support) et d'une couche de plastique à base de polycarbonate. Ces deux éléments sont immuables et servent à la fois à rendre le CD lisible, en réfléchissant la lumière du laser de lecture, et à solidifier le support en le rendant plus épais. La composition de la zone de données dépend du type de média utilisé : laque où les données sont pressées pour les CD-Rom, DVD-Rom et CD-Audio, couche de teinture réagissant à un type précis de laser pour les CD-R et DVD-R (inscriptibles). Les supports réinscriptibles (CD et DVD RW) utilisent un alliage à changement de phase, chauffé par un laser, pour stocker les données. Pour y effacer des informations, il suffit alors de redonner à l'alliage en question son état initial sur certaines sections du disque. Ces différences de structures suivant les différentes technologies de support impliquent bien entendu des méthodes de conservation, et des précautions d'usage adaptées.

76. Cette norme a été retirée par l'AFNOR mais constitue encore une référence.



Ces dernières années, plusieurs études ont été menées pour aboutir à des recommandations relatives à la conservation des supports optiques.

- Diffusion en 2005 par la Direction des Archives de France de Recommandations relatives à la gravure, à la conservation et l'évaluation des CD-R<sup>77</sup>. Les principales recommandations sont les suivantes :

- > choisir des 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 deux exemplaires : un master d'archivage et une copie de consultation ;
- > stocker les exemplaires dans deux endroits séparés et suffisamment distants.

- Constitution en 2005 d'un « Réseau national de recherche sur la conservation des informations enregistrées sur disques optiques numériques » sous la forme d'un Groupement d'Intérêt Scientifique (GIS-DON<sup>78</sup>) donnant de nombreuses informations utiles sur les perspectives existantes dans ce domaine.

- Mise en place d'un observatoire permanent de la qualité des CD-R du marché avec le lancement d'une étude en 2006 sur les CD-R (analyse des couples CD/graveurs) dont les résultats sont disponibles sur le site des Archives de France<sup>79</sup> qui propose aussi un guide pour la réalisation de la migration de stocks de CD-R.

---

77. [En ligne] < <http://www.archivesdefrance.culture.gouv.fr/gerer/archives-electroniques/stockage> > (consulté le 22 janvier 2010).

78. [En ligne] < [http://www.lne.fr/fr/r\\_et\\_d/gis-don/conservation-donnees-numeriques-gis-don.asp](http://www.lne.fr/fr/r_et_d/gis-don/conservation-donnees-numeriques-gis-don.asp) > (consulté le 22 janvier 2010).

79. [En ligne] < <http://www.archivesdefrance.culture.gouv.fr/gerer/archives-electroniques/stockage> > (consulté le 22 janvier 2010).

Dans le cahier des charges, il conviendra de demander au prestataire de suivre ces différentes recommandations. On pourra distinguer le support de livraison et le support de conservation.

Le prestataire indiquera la marque et la capacité des disques utilisés ainsi que la nature du film métallique et du colorant organique. Il indiquera le nom du fabricant des supports et les numéros de série des disques utilisés. Il donnera également des indications précises sur le graveur et le logiciel de gravure utilisés.

Pour des questions de contrôle et de traçabilité, les pochettes des CD-R pourront comporter les indications suivantes :

- Nom de l'établissement ;
- Nom du CD-R selon les recommandations de la Mission de la recherche et de la technologie du ministère de la Culture et de la Communication<sup>80</sup> ;
- Titre et cote du document ;
- Date de production ;
- Logiciels utilisés pour la production ;
- Résolution et compression ;
- Nombre et poids total des fichiers ;
- Nom des fichiers extrêmes.

On pourra demander au prestataire la livraison d'un autre jeu d'images sur un support de disque dur externe connectables en USB 2.0. Le prestataire indiquera le type de disque utilisé et précisera la structuration du disque dur (*filesystem*) au moment du formatage : NTFS, FAT32, ext3, ext2...

Cette prestation est surtout demandée pour des questions de commodité de chargement de la base de données.

---

80. [En ligne] < [http://www.culture.gouv.fr/mrt/numerisation/fr/gestion\\_fonds\\_images/gstim8.htm](http://www.culture.gouv.fr/mrt/numerisation/fr/gestion_fonds_images/gstim8.htm) > (consulté le 22 janvier 2010).

## LES PÉRIPHÉRIQUES DE STOCKAGE

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

- pour les CD-R et DVD-R : stockage par *jukebox* ;
- pour les bandes magnétiques : lecteur de bandes ou « autoloader » pour les petits volumes, bibliothèques intégrant plusieurs drives pour les gros volumes ;
- pour les disques durs : baies de stockage.

### Les types d'architecture de stockage

- DAS (*Direct Attached Storage*) : architecture la plus simple et la plus répandue, en attachement direct (un serveur avec un disque attaché) mais peu évolutive et peu adaptée à la volumétrie importante ;

- NAS (*Network Attached Storage*) : architecture en attachement réseau. Les données sont stockées sur des serveurs de fichiers reliés entre eux par le réseau local :

> avantages : simplicité, évolutivité, disponibilité des données, augmentation de la sécurité des données ;

> inconvénients : le réseau local partagé empêche la garantie des performances, administration complexe.

- SAN (*Storage Area Network*) : réseau informatique dédié au stockage, permettant la mutualisation des ressources de stockage :

> avantages : performances garanties et administration simplifiée ;

> inconvénients : coût très élevé, solution réservée aux grandes organisations, nécessite des compétences pointues.

Pour le réseau de stockage, les systèmes de type NAS et SAN ne sont prescrits que pour les plus grandes plateformes (gérant plusieurs dizaines de To).

## LE STOCKAGE HIÉRARCHIQUE

Le stockage hiérarchique (HSM = Hierarchical Storage Management) permet d'optimiser l'utilisation d'un espace de stockage pour des volumes importants de données. Le principe consiste à utiliser des supports performants (disques durs) pour les données à haute criticité ou demandant un accès fréquents et des supports moins rapides (disques durs, bandes magnétiques...) pour les données demandées moins souvent ou pouvant être sollicitées par un accès plus lent. Ce fonctionnement est tout à fait transparent pour l'utilisateur. Des procédures automatiques permettent de dynamiser le système en migrant au besoin les fichiers les moins sollicités vers les supports adéquats, tout en conservant la trace. Répondant à des questions d'économie, de gestion de performances et de bonne administration des capacités de stockage, cette possibilité est utilisée depuis quelques années dans les administrations et les collectivités. Elle peut répondre aux problématiques de stockage mais également d'archivage dans certains cas.

## LA SAUVEGARDE DES DONNÉES ET L'ARCHIVAGE

Un système de stockage devra être accompagné d'outils de sauvegarde permettant de sécuriser les données. On pourra équiper les machines de disques en RAID (*Redundant Array of Independant Disks*)<sup>81</sup>. Les technologies RAID offrent différents niveaux de sécurisation de données. Il est conseillé d'utiliser des machines équipées de disques durs en RAID 1 ou 5. Le RAID 1 consiste à dupliquer en permanence l'information sur plusieurs disques selon un principe de miroir. La fiabilité du système est excellente mais le coût du stockage est doublé. Le RAID 5 est un système dans lequel les données sont réparties sur plusieurs disques selon le principe de l'agrégation avec parité répartie. Ce système moins coûteux offre une sécurité très élevée. Des rythmes de sauvegarde (quotidiens, hebdomadaires...) sont à mettre en place. De plus, il est parfois conseillé de stocker régulièrement à l'extérieur des sauvegardes sur bandes.

---

81. Le système RAID permet de stocker les données sur plusieurs disques. Ils sont redondants, en totalité ou en partie, pour assurer une sauvegarde en cas de défaillance du système.

## 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 d'une plateforme automatisée à partir d'un certain volume (à partir de 2 To de données), en fonction des moyens humains.

### exemples de solutions 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 dans un délai de 5 à 7 ans ;
- faire tester un échantillon de CD-R régulièrement (par société privée ou par logiciel dédié).

Pour les volumes moyens (5 To) :

- copies multiples en diversifiant les technologies de stockage ;
- conserver sur site une copie disque et une copie bande et sur site distant une copie bande ;
- archivage de la totalité des fonds sur serveur en RAID 5 avec copie LTO (acquisition d'un auto-loader) en architecture de type DAS, NAS.

- le microforme<sup>82</sup> comme archive informatique : microfilm argentique impressionné par un laser numérique, le microfilm numérique s'impose désormais comme un moyen sûr d'archiver à long terme des informations numériques. Ces microfilms sont réalisés par des traceurs micrographiques ou traceurs de fichiers numériques sur microfilm 35 mm et 16 mm.

---

82. La commission CGZ43 C (Archivage des données électroniques) est chargée de l'élaboration d'une norme française sur l'archivage des données électroniques sur microformes COM (Computer Output Microforms, sortie d'ordinateur sur microfiche) y associant éventuellement la technologie COLD (Computer Output to Laser Disk, sortie d'ordinateur sur disque laser).

## 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 comprenant la dématérialisation, la sécurisation des échanges et la certification avec les tiers de confiance chargés de garantir l'authenticité des documents (signature électronique) et la sécurité des échanges.

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 d'archivage pérenne ;
- la taille des fichiers à archiver : possibilité d'archiver de grandes masses de données (ex. les masters) ;
- la place des métadonnées et l'indexation.

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

### **tiers stockage (court et moyen terme) / tiers-archivage (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.

Un système de stockage est une plateforme ou un élément de base sur lequel s'appuie le système d'archivage. Le recours à l'externalisation du stockage permet de faire appel à des acteurs disposant de compétences et de faire des économies d'échelle, compte tenu de l'importance des coûts fixes de ce type de projet (personnel, matériel).

**ENCADRÉ****PLATEFORMES D'ARCHIVAGE D'ÉTABLISSEMENTS PUBLICS**

**Nom :** Plateforme d'Archivage du Cines (PAC)

**Établissement :** Centre informatique national de l'enseignement supérieur (Cines)

**Types de données archivées :** Thèses électroniques déposées dans STAR (Abes), revues en SHS du portail Persée, documents sonores issus de la recherche dans le domaine de l'oral (Projet pilote CRDO dans le cadre du programme SHS du TGE-Adonis)

**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 fichiers acceptés volontairement limitée : formats publiés, largement utilisés, normalisés si possible ;
- architecture basée sur du matériel SUN, le logiciel Arcsys et des logiciels libres (Java, MySQL, Jhove, ImageMagick).

Volonté de se positionner comme tiers archiveur à destination du monde de l'enseignement supérieur et de la recherche.

**Lien :** < <http://www.cines.fr/-D-I-S-T-.html> >

**Nom :** Système de Préservation et d'Archivage Réparti (SPAR)

**Établissement :** Bibliothèque nationale de France

**Types de données archivées :**

- production issue de la numérisation de conservation et de consultation (imprimés, images, audiovisuel, manuscrits...) ;
- production issue du dépôt légal automatique et négocié (support, Web, presse quotidienne régionale) ;
- production administrative et technique (documents internes).

**Volumétrie :** 400 To archivés, 1 Po attendu en 2010

**Principes organisationnels :**

- basée sur les standards du domaine : modèle OAIS, dictionnaire de métadonnées (METS, Dublin Core, PREMIS, ODRL), identifiant pérenne (ARK), calcul d'empreinte numérique ;
- infrastructure basée sur du matériel SUN, applicatif développé par Atos Origin basé sur Fedora-Commons

Volonté de se positionner comme tiers archiveur

**Lien :**

< <http://bibnum.bnf.fr/spar/index.html> >

En 2009, l'Association des développeurs et des utilisateurs de logiciels libres pour l'administration et les collectivités territoriales (ADULLACT) crée une coopérative API pour proposer de l'infogérance d'archivage électronique aux collectivités.

**ENCADRÉ****PLATEFORMES D'ARCHIVAGE PRIVÉES****Nom : Rosetta**

**Établissement :** ExLibris, développé en collaboration avec la bibliothèque nationale de Nouvelle-Zélande (depuis 2008).

**Principes organisationnels :**

- progiciel à destination des bibliothèques ;
- se veut un système évolutif, prenant en compte les collections des bibliothèques de toute taille (documents numérisés, électroniques, audiovisuel, sites Web), tenant compte du cycle de vie du document, de l'évolutivité des formats ;
- système basé sur les standards du domaine tel que le modèle OAIS.

**Lien :** < <http://www.exlibrisgroup.com/category/ExLibrisRosettaOverview> >

**Nom : E-coffre fort**

**Établissement :** Office Français pour la Sécurité et l'Archivage des Documents (OFSAD)

**Types de données archivées :** documents papier et documents électroniques.

**Volumétrie :** capacité du coffre illimité et durée de conservation illimitée

**Principes organisationnels :**

offre s'adressant aux particuliers, aux associations et aux entreprises

- conversion automatique en PDF avec engagement à suivre les éventuelles modifications du format PDF ;
- engagement de restituer des copies numériques des documents d'origine papier toujours lisibles et exploitables plus de 30 ans après leur dépôt ;
- diversité des modes de dépôt (par imprimante virtuelle, par téléchargement par Drag & Drop (et copier-coller), par dépôt initial sur CD-Rom ;
- cryptage et empreinte numérique ;
- archivage de trois originaux sur deux sites distants ;
- certification ;
- référence à la norme NF Z 42-013

Archivage électronique.

**Tarifs :** offres distinctes pour particuliers (de 3 € HT à 12 € HT par mois) et professionnels (de 12 € HT à 28 € HT par mois). Labellisé par la FNCT en 2003

**Lien :** < <http://www.e-coffrefort.fr/index.htm> >



...

**Nom : Le coffre fort électronique<sup>®</sup>****Établissement :** CDC Arkhinéo**Types de données archivées :**

tout document

**Principes organisationnels :**

- enregistrement de quatre originaux sur plusieurs disques magnétiques répartis sur deux sites distants, *back up* automatique quotidien sur disques magnétiques sur un 3<sup>e</sup> site distant ;
- constitution d'une enveloppe logique contenant le calcul d'empreinte, l'horodatage et le document reçu ;
- attribution automatique d'un identifiant unique d'archive et signature électronique de l'enveloppe d'une Public Key Infrastructure (PKI) interne.

Tarifs non communiqués, aucune migration de format semble prévue.

**Lien :** < <http://www.cdcarckhineo.com/index.php?page=accueil> >

**Nom : Système d'Archivage Électronique****Établissement :** Locarchives**Types de données archivées :**

tout document

**Principes organisationnels :**

- contrôle automatique des fichiers ;
- indexation des documents reçus ;
- conversion automatique en PDF-A,
- signature électronique (certification, empreinte, horodatage) ;
- certificat de réception ;
- stockage on line et off line (CD-R) sur deux sites distants ;
- utilisation de supports et formats normalisés : CD-R et PDF-A ;
- réalisation éventuelle de migration de supports en cas de non-lisibilité et possibilité d'assurer la conversion vers des formats pérennes.

Tarifs non communiqués, référence à la norme NF Z 42-013 Archivage électronique  
**Lien :** < <http://www.locarchives.fr/solutions/archivage-documents-electroniques.html> >

Sur le site de la Fédération nationale des tiers de confiance (FNTC), liste des opérateurs et prestataires de services de confiance adhérents : < [http://www.fntc.org/component?option=com\\_fntcmembre/task,college/Itemid,76/idCollege,3/](http://www.fntc.org/component?option=com_fntcmembre/task,college/Itemid,76/idCollege,3/) >

**ENCADRÉ**

**PLATEFORMES D'ARCHIVAGE OPEN SOURCE**

---

**Nom : Flexible Extensible Digital Object Repository Architecture (Fedora)**

Développement par l'université de Cornell et l'université de Virginie

**Lien :** < <http://fedora-commons.org/> >

**Nom : D-Space**

Développement en 2002 par les bibliothèques du MIT et Hewlett-Packard.

Système tourné vers le monde académique.

**Lien :** < <http://dspace.org/> >

> **Plateformes** libres d'accès pouvant être adaptées aux spécificités de chaque service. Depuis 2009, ces deux systèmes sont unis dans Dura-Space.

**Lien :** < <http://duraspace.org/index.php> >

---

**Nom : Lots of Copies Keep Stuff Safe (LOCKSS)**

Mis en place par l'université de Stanford, ce système a pour objectif de fournir aux bibliothèques des outils et supports pour la préservation numérique.

Système basé sur le modèle OAIS, migration de formats, réparation des données.

**Lien :** < <http://www.lockss.org/lockss/Home> >

---