



CONSERVATION A LONG TERME DES DOCUMENTS NUMERISES

MINISTERE DE LA CULTURE ET DE LA COMMUNICATION

2008

1	CONTEXTE GLOBAL	P. 3
2	NORMES ET STANDARDS	P. 5
3	POLITIQUE DE CONSERVATION ORGANISATION, ACTEURS, INFRASTRUCTURE TECHNOLOGIQUE	P. 7
4	PRODUCTION DE DOCUMENTS NUMÉRISÉS FORMAT DE DONNEES – METADONNEES DE CONSERVATION	P. 9
5	SUPPORTS ET INFRASTRUCTURE DE STOCKAGE SUPPORTS OPTIQUES – SUPPORTS MAGNETIQUES	P. 13
6	QUESTIONS JURIDIQUES	P. 15
7	PROJETS ET EXPÉRIENCES EN COURS	P. 16

1 CONTEXTE GLOBAL

Le ministère de la Culture et de la Communication et ses établissements publics sont entrés depuis quelques années dans une période de numérisation massive de leurs collections voire de productions directes numériques (prises de vues par les photographes). Des collections numériques sont ainsi constituées, d'un volume souvent très considérable et qu'il importe par conséquent, de pérenniser étant donné les investissements réalisés. Le fait que les originaux existent, n'exonère pas de la conservation des images numériques, en raison de leur propre dégradation, patente notamment pour les supports analogiques audiovisuels.

Dans un premier temps le ministère de la Culture et de la Communication a su développer des savoir-faire et des référentiels en matière de description, de numérisation des collections et de leur diffusion ; dans un deuxième temps il ouvre le chantier, complexe, de la problématique de la conservation à long terme des images produites

Les années récentes ont vu se multiplier tant au niveau national qu'europpéen et international les études, projets, standards, normes, recommandations et initiatives de mises en place de services d'archivages à long terme d'informations numériques.

De plus, les méthodologies, référentiels, outils développés actuellement en matière de pérennisation des informations numériques, portent surtout sur des fichiers numériques nativement produits. Toutefois, ceux-ci peuvent également s'appliquer aux images numérisées, au moins pour leur stockage pérenne. Les domaines scientifiques ont été les premiers concernés ainsi qu'au sein du ministère de la Culture, la direction des Archives de France et des grands établissements publics comme la Bibliothèque nationale de France et l'Institut national de l'audiovisuel.

Il s'agit par conséquent ici de sensibiliser les acteurs du ministère de la Culture ainsi que les autres acteurs culturels notamment dans les collectivités territoriales.

Il ne s'agit pas ici d'écrire un manuel sur le sujet mais de présenter thématiquement des ressources déjà publiées sur les réseaux.

Ressources documentaires sur le serveur du groupe PIN

Le groupe de travail et d'échanges français PIN (Pérennisation de l'Information Numérique) créé en 2000 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. Il se préoccupe également de l'accès aux informations qui ont été pérennisées. Il est constitué d'un ensemble de représentants d'organismes et d'entreprises qui ont des vocations fort différentes : activités d'étude et de recherche scientifiques et techniques, services publics, organismes à vocation patrimoniales... Néanmoins, tous se retrouvent sur un même terrain qui est celui de la préservation d'informations numériques.

<http://vds.cnes.fr/pin/index.html>

L'originalité et la force de PIN résident dans le caractère transverse de sa composition articulée sur plusieurs pôles : des institutions patrimoniales, des organismes à caractère scientifique et technique, de grands groupes nationaux publics ou privés, des experts et chercheurs indépendants reconnus du domaine¹.

¹ AVENTIS, Archives de France, Archive-17, Bibliothèque nationale de France, Commissariat à l'Énergie Atomique (CEA), Centre des archives contemporaines, Centre Informatique National de l'Enseignement Supérieur (CINES), Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Centre National d'Études Spatiales (CNES), Francetelecom, Groupe Médéric, Institut Français de Recherche pour l'Exploitation de la MER (IFREMER), Institut National de l'Audiovisuel (INA), INRIA, INSEAD, ministère de l'Agriculture, de l'Alimentation, de la Pêche et des Affaires rurales, ministère de l'Équipement, ministère de la Justice, Ville de Paris, Institut Pasteur

Les pages ressources du site, régulièrement mises à jour, constituent une mine d'informations indispensables pour suivre les évolutions technologiques, normatives, organisationnelles dans un contexte en mutation rapide

http://vds.cnes.fr/pin/pin_ressources.html

Le didacticiel : Gestion de la conservation de collections numériques

<http://www.library.cornell.edu/iris/tutorial/dpm-french/index.html>

Le didacticiel « Gestion de la conservation de collections numériques » produit par la société des archivistes américains et l'Université de Cornell constitue une introduction aux principes fondamentaux de la conservation d'objets numériques. Il est particulièrement adapté aux besoins des bibliothécaires, archivistes, conservateurs, gestionnaires et spécialistes techniques. Vous y trouverez des définitions, des concepts importants, des conseils pratiques, des exercices et des références à jour. La version française du site a été lancée en avril 2006.

Autre module de formation en ligne :

Le portail international archivistique francophone (PIAF), modules de formation en ligne, chapitre : « **Gestion et archivage des documents numériques** » :

http://www.piaf-archives.org/sections/formation/module_07/

2. NORMES ET STANDARDS

Présentation d'une réflexion de synthèse sur les normes et standards, C. Huc, 2005

http://vds.cnes.fr/pin/presentations/2005/reflexions_normes_et_standards.pdf

Présentation de la situation des organisations de normalisation et des comités concernés par l'archivage électronique. Les points suivants sont abordés : typologie de l'archivage électronique, stockage, formats, intégrité et authenticité, empaquetage, identification.

Zoom sur la norme généraliste du domaine :

La norme OAIS (Open Archival Information System)

La norme ISO 14721:2003 (Systèmes de transfert des informations et données spatiales - Système ouvert d'archivage de l'information - Modèle de référence), plus connue sous le nom de modèle OAIS (Open Archival Information System) est consultable à l'adresse suivante :

http://vds.cnes.fr/pin/documents/OAIS_650_0b1.pdf

Une traduction française, en cours de normalisation, est accessible à l'adresse suivante :

http://vds.cnes.fr/pin/documents/norme_oais_version_francaise.pdf

Cette norme conceptuelle, mise au point par les principaux centres d'études spatiales du monde dont le CNES (Centre National d'Études Spatiales), 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. Cette norme définit un vocabulaire et un ensemble de concepts permettant d'appréhender de façon globale et complète, la question de l'archivage long terme de données sous forme numérique. Elle définit deux modèles complémentaires : un modèle d'information et un modèle fonctionnel détaillé. Elle propose également une classification des types de migration et des différents modes de coopération possibles entre archives.

Adoptée dès 1999 par tous les programmes de bibliothèques s'intéressant à la préservation et accès à long terme et par les entrepôts institutionnels, elle définit :

- un archivage d'objets numériques et analogiques
- un vocabulaire neutre par rapport aux différents secteurs professionnels concernés
- un ensemble de briques fonctionnelles nécessaires au stockage, à la conservation et à la communication
- des facilités d'adaptation aux besoins des institutions

Elle ne préjuge pas d'une implémentation particulière et constitue un cadre de développement pour des normes complémentaires.

La gestion et conservation documents numériques

L'intervention présente le cadre général et l'implémentation d'OAIS à la Bibliothèque nationale de France. Catherine Lupovici, Entretiens de la BnF, décembre 2006

<http://www.bnf.fr/pages/infopro/journeespro/pdf/entetiens06/Lupovici.pdf>

Référentiel général interopérabilité

http://www.synergies-publiques.fr/rubrique.php?id_rubrique=71

Le référentiel général d'interopérabilité (RGI) s'inscrit dans l'action de l'État visant à promouvoir le développement de l'administration électronique, portée par le nouvel environnement juridique initié par la loi du 13 mars 2000 relative au droit de la preuve et à la signature électronique.

Il est institué par une ordonnance en date du 8 décembre 2005 relative aux échanges électroniques entre les usagers et les autorités administratives et entre les autorités administratives. Le décret n° 2007-284 du 2 mars 2007 fixe les modalités d'élaboration, d'approbation, de modification et de publication du référentiel général d'interopérabilité.

Le RGI vise à définir les conditions techniques dans lesquelles s'organisent les échanges d'informations au sein des administrations. Il déterminera un ensemble de normes et de règles destinées à faciliter l'échange de données et d'informations des services publics entre eux et vers les usagers. Il permettra également la mutualisation des systèmes d'information de l'État et des investissements nécessaires à leur développement. Enfin, il facilitera le développement de nouveaux services aux usagers.

Rappel

http://synergies.modernisation.gouv.fr/IMG/pdf/Referentiel_General_Interoperabilite_Volet_Technique_V0.98c.pdf

Le référentiel général d'interopérabilité présente ainsi un ensemble de règles visant à améliorer l'interopérabilité des systèmes d'information du service public. Par règle, il faut comprendre la recommandation ou l'injonction d'utiliser, ou non, tel répertoire, de s'appuyer ou non, sur telle référence, norme ou standard...

Ces règles sont inscrites dans le temps et accessibles par thématiques génériques, techniques ou fonctionnelles (identification, archivage...). Elles peuvent aussi concerner un service spécifique (plate-forme de dématérialisation des appels d'offre, plate-forme de transmission sécurisée...).

Les règles portent sur des objets de natures très différentes, et donc expriment des exigences très diverses, dans des domaines tels que les normes et standards techniques, les modèles de données, les nomenclatures, les composants mutualisés (contrôle des adresses postales, souche logiciel libre de gestion de l'identité), ainsi que les politiques de mise en oeuvre de systèmes (clauses de politique d'archivage).

Des règles organisées en volets

Le référentiel couvrira tous les domaines utiles à l'obtention d'une interopérabilité effective des systèmes d'information du secteur public :

- l'interopérabilité technique (protocoles IP, supports et formats d'encodage des différentes catégories d'informations numériques : jeux de caractères, texte, image fixe, image animée, son et vidéo, formats structurés, documents techniques...),
- l'interopérabilité sémantique (les formats des messages, la structuration, le sens et la référence de validation des éléments les constituant),
- l'interopérabilité organisationnelle (les rôles des entités et des acteurs en interaction avec les systèmes d'information).

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

Décret n° 2007-284 du 2 mars 2007 fixant les modalités d'élaboration, d'approbation, de modification et de publication du référentiel général d'interopérabilité

<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT00000271296&dateTexte=>

Une nouvelle version de la norme **NF Z 42-013** (décembre 2001) sur les systèmes informatiques destinés à la conservation des documents numériques est en cours de révision à l'AFNOR (CN 171) et devrait être publiée officiellement mi-2008.

3. POLITIQUE DE CONSERVATION

ORGANISATION, ACTEURS, INFRASTRUCTURE TECHNOLOGIQUE

Instruction DITN/RES/2006/005 du 13 septembre 2006. Publication de l'étude commanditée par la direction centrale de la sécurité des systèmes d'information (DCSSI) sur l'archivage électronique sécurisé dans le secteur public

http://www.archivesdefrance.culture.gouv.fr/fr/circAD/DITN_RES_2006_005.pdf

<http://www.ssi.gouv.fr/fr/confiance/archivage.html>

Les services du Premier Ministre ont réalisé, avec la direction des Archives de France, une étude sur l'archivage sécurisé. Plusieurs livrables ont été publiés :

- une étude sur les questions légales liées à l'archivage électronique,
- un modèle de cahier des charges fonctionnel et technique pour un système d'archivage,
- un modèle de politique d'archivage. Ce document définit les exigences minimales, en termes juridiques, fonctionnels, opérationnels, techniques et de sécurité, qu'une autorité d'archivage doit respecter afin que l'archivage électronique mis en place puisse être regardé comme fiable. Ainsi cette PA Type repose sur des contraintes "standard" à mettre en place dans les domaines suivants que sont l'identification et l'authentification de l'origine de l'Archive, l'intégrité des objets archivés, leur intelligibilité et lisibilité, leur pérennité, leur disponibilité et accessibilité, ainsi que la traçabilité des opérations les concernant. La PA Type constitue donc un référentiel de la sécurité de l'archivage électronique pour qu'il puisse être qualifié de fiable.

LES COÛTS DE L'ARCHIVAGE ELECTRONIQUE

Question loin d'être résolue et claire, recul insuffisant, cependant : il existe des modèles et des études ont été menées sur le sujet. Des retours d'expérience en vraie grandeur devraient être disponibles dès 2008 ; ces coûts varient en fonction des services rendus et des contraintes propres aux contenus : besoin de préserver des éléments de preuve d'authenticité par exemple, besoin d'accéder plus ou moins rapidement aux informations. Ils varient également suivant les matériels, équipements et infrastructures dont on dispose déjà soit directement soit mutualisés avec d'autres services

Dans le secteur privé, depuis deux ou trois ans, se développe une offre autour de « coffres-forts » électroniques qui permettent d'ingérer les données, de les contrôler et vérifier, de garantir leur intégrité (empreintes) et de les envoyer vers un espace de stockage (coûts de licences variant suivant la volumétrie à gérer) : offres de Cecurity.com, d'Omnikles. Parallèlement à cette offre autour du concept du coffre-fort, des briques logicielles permettant le versement et la restitution des archives numériques, commencent à être proposées : produit Arcsys chez Storagetek, briques génériques de EADS. Par ailleurs, il existe une offre en matière de services d'infrastructures de stockage dans le monde des « stockeurs » (IBM, Storagetek, EADS, Article Circle, Centera de EMC) qui commencent à intégrer des notions d'ILM (notions de durées de conservation, d'indexation...). Sinon, pour la restitution (et plus généralement le système documentaire), les actuels éditeurs de progiciels spécialisés pour la description archivistique et la recherche documentaire des archives papier, pourraient évoluer pour offrir des produits fonctionnant pour les archives numériques. Il existe de ce point de vue là également une offre chez les éditeurs de produits de GED et de gestion des contenus.

Une offre de tiers-archivage émerge également : partie de sociétés venant du monde de l'édition (Asphéria, Orsid, Astérian avec son offre e-Star) et une offre basée sur des supports optiques non réinscriptibles., on la retrouve aujourd'hui par exemple chez CDC-Arkhinéo (filiale de la Caisse des dépôts et consignations) dont le produit est basé sur la technologie Zantaz au départ spécialisé dans la gestion des courriels et qui utilise des disques magnétiques (baies EMC avec un dispositif complexe de contrôle et verrouillage logiciel), ou bien encore chez des tiers-archivageurs jusque là spécialisés dans le stockage des archives papier (comme Locarchives qui met en œuvre le coffre-fort électronique de Cecurity.com). Certains offrent ces deux types de prestations.

ÉTUDE RELATIVE A LA REALISATION DE PLATE-FORMES D'ARCHIVAGE ELECTRONIQUE

- Note d'information DITN./RES/2005/007 du 13 octobre 2005

<http://www.archivesdefrance.culture.gouv.fr/fr/circAD/DITN.2005.007.pdf>

Cette étude, commandée par la direction des Archives de France en 2005, met en lumière les fonctions qui doivent être remplies par un système d'archivage électronique (basées sur le modèle OAIS), définit plusieurs scénarii d'organisation (infrastructure centralisée ou décentralisée) et évalue les coûts de chaque scénario (personnel, logiciels, matériel).

- Étude de l'existant et des besoins, Parker Williborg, 2005 -Direction des Archives de France – Plate-forme d'archivage électronique

<http://www.archivesdefrance.culture.gouv.fr/fr/circAD/DITN.2005.006.etudeexistant.pdf>

- Étude des scénarios, Parker Williborg, 2005 – Direction des Archives de France – Plate-forme d'archivage électronique

<http://www.archivesdefrance.culture.gouv.fr/fr/circAD/DITN.2005.007.Etude%20des%20sc%E9narios%20-%20V3.0.pdf>

LA CERTIFICATION ET LES AUDITS

La réflexion démarre tant au niveau national qu'international. L'établissement d'une grille d'audit est le préalable à la mise en œuvre de systèmes de certification/homologation de services ou d'outils. L'étude menée par la DCSSI comprend déjà une grille d'audit basée sur les principaux éléments de la politique d'archivage.

Au niveau international, un projet porté en 2006 par Research Library Group (RLG) et les archives nationales américaines (NARA) a permis l'établissement d'un outil pour évaluer une plate-forme d'archivage électronique, conforme à la norme OAIS

RLG Checklist : Certification d'un entrepôt de préservation conforme à la norme OAIS

<http://www.oclc.org/programs/ourwork/past/repositorycert.htm>

Une version révisée vient d'être publiée par le Center for Research Libraries *Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)*. Version 1.0 février 2007

<http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91>

DRAMBORA (Digital Repository Audit Method Based on Risk Assessment)

DRAMBORA est une initiative de Digital Curation Centre (Grande-Bretagne) et du projet européen DigitalPreservationEurope. Il s'agit de développer un outil pour l'audit et l'auto-évaluation d'entrepôt numérique, basé sur la maîtrise des risques. La première version de cet outil est disponible depuis mars 2007. Une version interactive de l'outil a été mise en ligne en avril 2008

<http://www.repositoryaudit.eu/>

Travaux du CCSDS

ISO Standard for audit and certification of repositories of digital information

Dans la lignée OAIS. Version préliminaire de livre blanc prévu mars 2008

<http://nssdc.gsfc.nasa.gov/nost/conf/archive21st/titles.html>

Intervention dans *Science Archives in the 21st Century* de David Giaretta : ***Toward an international standard for Audit and Certification of Digital***

<http://wiki.digitalrepositoryauditandcertification.org/bin/view>

Wiki de travail sur le projet de norme : « Welcome to the Digital Repository Audit and Certification Wiki »

4. PRODUCTION DE DOCUMENTS NUMÉRISÉS

Manuscrits, documents reliés, plans, dessins photographies, microformes.

Concernant la production des documents numérisés, la Direction des Archives de France vient d'éditer un guide technique *Écrire un cahier des charges pour la numérisation du patrimoine*, mars 2008.

http://www.archivesdefrance.culture.gouv.fr/fr/archivistique/guide_numerisation.pdf

Il s'applique aux manuscrits, documents reliés, plans, dessins photographies, microformes.

Les chapitres du guide indiqués ci-dessous concernent plus spécifiquement la conservation :

- Nommage des fichiers
- Métadonnées des fichiers images
- Livraison des images numériques
- Contrôle des prestations par le commanditaire
- Recommandations sur les caractéristiques techniques des fichiers images : les parties relatives aux « fichiers de conservation », à consulter dans l'annexe technique.

Ce guide technique sera progressivement enrichi pour la numérisation de documents sonores, de documents audiovisuels et la numérisation 3D.

Les étapes de la chaîne de numérisation à la Bibliothèque nationale de France

Les techniques retenues et la charte technique, mars 2005

http://www.bnf.fr/pages/infopro/numerisation/num_technique.htm

Cahier des charges de conversion OCR de la presse périodique (BnF)

http://bibnum.bnf.fr/numerisation/charte_technique_ocr_presse.pdf

Le plan de sauvegarde et de numérisation de l'INA : audiovisuel

<http://www.ina.fr/entreprise/activites/archives-sauvegarde-numerisation/index.html>

Documents oraux

Corpus oraux : Guide des bonnes pratiques (action DGLFLF)

http://www.dglf.culture.gouv.fr/recherche/corpus_parole/Corpus_Oraux_GBP%202006_version_imprimee.pdf?CV=5584&type1=Ouvrage

Sous la direction d'Olivier Baude, Paris, CNRS, 2006, 192 p.

L'ouvrage a pour vocation d'éclairer la démarche des chercheurs, de repérer les problèmes et les solutions juridiques et de favoriser l'émergence de pratiques communes pour la conservation, l'exploitation et la diffusion des corpus oraux.

IDENTIFIANTS ET NOMMAGE

Normalisation des noms d'images - Normalisation des noms de supports CD-R

Prescriptions destinées aux services de la culture par Jean-Louis Pascon et Frédéric Rolland
Images

http://www.culture.gouv.fr/mrt/numerisation/fr/gestion_fonds_images/gstim9.htm

Supports

http://www.culture.gouv.fr/mrt/numerisation/fr/gestion_fonds_images/gstim8.htm

Identifiants uniques des documents numériques, site INIST

Forum de recherche d'information et de discussion, 6 juillet 2006, ARTIST, Jacques Ducloy, Jean-Paul Ducasse, Marin Dacos

http://artist.inist.fr/article.php3?id_article=348

Les identifiants pérennes à la BnF

Les identifiants pérennes sont des adresses Internet particulières qui permettent de citer un document numérique à l'aide d'un lien hypertexte, tout en ayant la garantie que ce lien ne risque pas de changer. Il existe différents systèmes permettant de créer des identifiants pérennes. Le système choisi par la BnF est l'Archival Resource Key (ARK).

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

Emmanuelle Bermés, 5 mai 2006

<http://bibnum.bnf.fr/identifiants/identifiants-200605.pdf>

Implementing persistent identifiers, Hise et Kothe, CERL, novembre 2006

Le rapport explique les principes relatifs aux identifiants pérennes et accompagne les institutions dans le choix d'un schéma correspondant à ses besoins. Il présente Handle, le Digital Object Identifier (DOI), Archival Resource Keys (ARK), le Persistent Uniform Resource Locator (PURL), Uniform Resource Names (URN), National Bibliography Number (NBN), et les URL ouverts, donnant de nombreux exemples et des références pour chaque identifiant.

Ce rapport, préparé par le département R&D de l'Université de Göttingen à la demande de l'Advisory Task Group (ATG) du Consortium des bibliothèques de recherche d'Europe (CERL), a été publié par le CERL et la Commission européenne pour la préservation et l'accès (ECPA).

http://www.cerl.org/web/en/publications/report_on_persistent_identifiers?s=kothe

4.1 FORMATS DE DONNEES POUR LA CONSERVATION

Le principe général est de choisir pour la conservation, des formats ouverts reposant sur des normes et standards, dont les spécifications sont publiques.

Dans certains cas, la distinction qui est effectuée traditionnellement entre format de conservation / de diffusion ne s'impose plus : suivant les types de documents ou la technologie de diffusion proposée, l'exploitation immédiate peut être assurée à partir des fichiers haute définition. Dans d'autres cas, par exemple pour les documents microfilmés, le format de conservation est le microfilm lui-même et il sera inutile de prévoir en sus un format de conservation numérique. À l'inverse, dans d'autres cas, la distinction entre les deux types de formats se justifie. Ainsi, il arrive couramment que seules des vignettes ou des versions dégradées des images numériques soient diffusées, afin de protéger la propriété intellectuelle ou des droits d'exploitation.

L'indication du format de fichier doit préciser le nom du format et sa version (exemple : TIFF version 6).

Il importe, une fois le choix du format arrêté, de maintenir ce choix tout au long de l'opération de numérisation des documents concernés. En effet, un changement de format (par exemple de TIFF à JPEG) au milieu d'une opération, posera des problèmes de traçabilité, nommage, voire des difficultés en cas de retraitements ultérieurs (par exemple pour des opérations de reconnaissance de forme).

On se reportera également utilement, pour plus de détails techniques, sur les paragraphes du Référentiel général d'accessibilité (RGI) traitant des formats (interopérabilité technique) :

- Formats des images fixes
- Formats des images animées
- Formats pour le son et la vidéo
- Formats pour l'audiovisuel et la vidéo HD
- Formats des données graphiques

http://synergies.modernisation.gouv.fr/IMG/pdf/Referentiel_General_Interoperabilite_Volet_Technique_V0.98c.pdf

Répertoires de formats

Il existe plusieurs initiatives en ce sens. On mentionnera particulièrement :

PRONOM, service opérationnel produit par les Archives nationales du Royaume-Uni sur les formats de fichiers de données et les logiciels qui les supportent.
<http://www.nationalarchives.gov.uk/pronom/>

GDFR (Global Digital Format Registry), projet d'une durée de deux ans conduit par la bibliothèque de l'Université de Harvard qui vise à rassembler des informations sur les formats de fichiers. 2006
<http://hul.harvard.edu/gdfr/>

4.2 METADONNEES

Les métadonnées sont un ensemble de données qui caractérisent d'autres données en vue d'en permettre la recherche, la gestion, la conservation.

Les métadonnées qui s'appliquent aux fichiers d'images numériques peuvent être de plusieurs sortes :

- descriptives (sujet, date, lieu, auteur de l'information...),
- techniques (format de fichier, mode de numérisation...),
- administratives (droits d'utilisation des images...).

Ces métadonnées peuvent être soit enregistrées à l'intérieur même du fichier informatique auxquelles elles s'appliquent, soit stockées à l'extérieur du fichier, en lien avec lui.

Les deux techniques présentent des avantages et des inconvénients. L'enregistrement des métadonnées à l'intérieur du fichier permet de maintenir un certain nombre d'informations (origine, propriétaire, sujet...) associées au fichier, même lorsque celui-ci est extrait de son contexte ordinaire d'utilisation. En revanche, ces métadonnées sont souvent limitées en nombre et assez difficiles à mettre à jour. De plus, il arrive qu'elles soient altérées lorsque le fichier est converti dans un autre format. Le stockage des métadonnées à l'extérieur du fichier permet l'utilisation de jeux de métadonnées beaucoup plus riches, dont la mise à jour est très aisée. Par contre, le lien entre les métadonnées et l'image auxquelles elles se rapportent est plus fragile et l'image, lorsqu'elle est manipulée isolément, n'est pas documentée.

Il est donc recommandé d'associer les deux méthodes :

- certaines métadonnées fondamentales, comme l'identifiant et le titre du document, le nom du service d'archives qui conserve le document, le nom du prestataire et la date de numérisation, la taille du document original ou la résolution de numérisation, doivent être enregistrées par le prestataire dans le fichier lui-même ;
- ces métadonnées doivent aussi, avec toutes les autres métadonnées jugées utiles, figurer dans une ou plusieurs bases de données extérieures aux fichiers images.

Il existe de nombreux formats de métadonnées. Certains sont généraux (Dublin Core), d'autres sont davantage spécialisés (par exemple l'EAD pour la description des documents d'archives).
http://www.archivesdefrance.culture.gouv.fr/fr/archivistique/guide_numerisation.pdf

Certains de ces formats sont orientés pour la conservation. Parmi ceux-ci, on mentionnera :

METS Métadonnées des objets numériques (Metadata Encoding & Transmission Standards) Initialisé par la Digital Library Federation. V1 février 2001

Schéma XML. Il permet de créer un document qui encapsule toutes les métadonnées d'un objet numérique :

- Métadonnées descriptives : en MARC, EAD ou DC...incorporées dans le fichier METS ou lien vers un autre réservoir
- Métadonnées administratives de l'objet numérique et de l'objet numérisé : droits , techniques
- Métadonnées de structures : cartes de structures de l'objet numérisé (logique, physique)
- Section donnant la liste des fichiers et leurs regroupements : vignette, matrice d'archive, version de consultation image, version texte intégral

Agence de maintenance

<http://www.loc.gov/standards/mets/>

En OAIS peut être utilisé pour les SIP, AIP, DIP

Norme d'échange d'objets numériques entre des entrepôts

<http://www.bnf.fr/pages/infopro/journeespro/pdf/entetiens06/Lupovici.pdf>

PREMIS : Preservation metadata and the OAIS information model. Groupe de travail OCLC/RLG sur les métadonnées de conservation, juin 2002

<http://www.oclc.org/research/projects/pmwg/>

Agence de maintenance

PREMIS (Preservation Metadata for Digital Materials) Data Dictionary, mai 2005

- Dictionnaire des métadonnées associées à la chaîne des traitements
- Pour la description des formats de documents, PREMIS renvoie sur les travaux des répertoires de formats
- Maintenance et suivi des travaux désormais assurés par la Library of Congress

<http://www.loc.gov/standards/premis/>

Metadata for Digital Libraries : State of the Art and Future Directions

Le JISC a publié, en avril 2008, un état de l'art sur les métadonnées :

Metadata for Digital Libraries: State of the Art and Future Directions

http://www.jisc.ac.uk/media/documents/techwatch/tsw_0801pdf.pdf

5. SUPPORTS ET INFRASTRUCTURE DE STOCKAGE

Aujourd'hui, un grand nombre de types de supports est disponible pour l'archivage. On distingue généralement deux grandes familles : les supports magnétiques et les supports optiques. S'agissant des supports magnétiques nous distinguerons les bandes magnétiques, des disques magnétiques et des nouvelles technologies qui en sont issues. En ce qui concerne les supports optiques nous citerons successivement les CD, puis les DVD et enfin les disques magnéto-optiques.

Par rapport à cette grande diversité de supports tant magnétiques qu'optiques, il est clair que les critères de choix devront se porter essentiellement sur la capacité des supports, leur fiabilité par rapport aux différents mécanismes d'enregistrement et bien sûr leur coût. Le critère du taux de transfert (assimilable à la vitesse d'écriture sur le média) est quelque peu marginal dans la mesure où en règle générale les procédures d'archivage sont effectuées à des périodes non critiques.

Comme autres critères, il ne faudra pas oublier l'évolutivité (problème des compatibilités ascendantes), les possibilités de migrations, la sécurité offerte.

Même si dans l'absolu le support idéal existait, ce qui est loin d'être le cas, encore ne faudrait-il pas oublier de prendre en considération les aspects économiques de façon globale. En effet sur ce dernier point, il est nécessaire de raisonner non pas sur l'achat ponctuel de tel ou tel support ou technologie, mais sur une exploitation simulée de plusieurs années afin de prendre en compte l'ensemble des paramètres : administration, maintenance, remplacement, ...

Quoiqu'il en soit, le type de support sera avant tout choisi en fonction de critères précis comme la durée de conservation, le caractère critique des données à conserver, l'accessibilité, la volumétrie et le coût.

Nous pouvons également ajouter comme exigences vis-à-vis des supports qu'ils aient les qualités suivantes :

- Stabilité intrinsèque du support et robustesse,
- large diffusion de la technologie et offre multi-constructeurs ou reposant sur des normes publiques,
- existence d'outils de contrôle des supports,
- chemin d'accès aux données protégé,
- simplicité des opérations de recopie,
- protection contre l'effacement accidentel.

On se reportera au Référentiel général d'accessibilité (RGI) – volet technique – Supports d'archivage.

http://synergies.modernisation.gouv.fr/IMG/pdf/Referentiel_General_Interoperabilite_Volet_Technique_V0.98c.pdf

5.1. SUPPORTS OPTIQUES

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

Afin d'aider les services d'archives qui stockent des données sur des CD, la direction des Archives de France a publié des *Recommandations sur la gravure, la conservation et la surveillance des CD-R*(2005).

Instruction DITN/RES/2005/004 du 29 mars 2005

<http://www.archivesdefrance.culture.gouv.fr/fr/textenorme/index.html>

Recommandations

<http://www.archivesdefrance.culture.gouv.fr/fr/textenorme/index.html>

Mémento pratique

<http://www.archivesdefrance.culture.gouv.fr/fr/textenorme/index.html>

Étude sur des CD-R et des graveurs du marché

En 2006, plusieurs CD-R et graveurs de CD ont été comparés et les plus adaptés pour la conservation ont été identifiés.

<http://www.archivesdefrance.culture.gouv.fr/fr/circAD/DITN.2006.008.pdf>

Note d'information DITN/RES/2006/008 du 20 décembre 2006

Rapport de synthèse de l'étude :

<http://www.archivesdefrance.culture.gouv.fr/fr/textenorme/index.html>

Étude sur les CD-R conservés par les services publics d'archives

Instruction DITN/RES/2006/003 du 14 juin 2006

<http://www.archivesdefrance.culture.gouv.fr/fr/textenorme/index.html>

Rapport de synthèse de l'étude :

<http://www.archivesdefrance.culture.gouv.fr/fr/textenorme/index.html>

GIS DON Réseau national de recherche sur la conservation des informations enregistrées sur disques optiques numériques- Création en décembre 2004 pour une durée de quatre ans

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

Il vise à apporter aux acteurs du domaine (fabricants, utilisateurs et normalisateurs) une expertise scientifique destinée à les guider dans le choix des supports et à préciser leurs conditions d'utilisation.

Laboratoire d'acoustique musicale (LAM), unité mixte CNRS

Laboratoire de photochimie moléculaire et macromoléculaire (LPMM), unité mixte CNRS

Laboratoire national d'essais (LNE)

Conseil scientifique composé de onze personnalités compétentes

Programme établi en concertation avec la mission de la recherche et de la technologie pour les aspects patrimoniaux

L'archivage sur CD-R : Acquérir - Graver - Contrôler - Conserver

Paris, Eyrolles, 2006, 94 p.

ISBN : 2-212-11799-X

<http://www.editions-eyrolles.com/>

Ouvrage écrit par des experts (des supports optiques et des normes, des professionnels du domaine technique du CD et des utilisateurs), qui dresse un état des lieux très complet des principes et des recommandations indispensables à la qualité et à la sécurité de tout projet d'archivage numérique sur CD-R, du choix des supports à la gravure et au contrôle, sans oublier les questions cruciales de conservation et de suivi dans le temps. Le chapitre 5 : Les procédures et les moyens de contrôle des CD-R est accessible en ligne.

5.2. SUPPORTS MAGNETIQUES

Étude d'architecture de stockage de données et de sauvegarde : état de l'art des techniques et technologies de stockage et de protection des données pertinentes dans le contexte du Ministère réalisée par la société Hi-Stor en 2006 pour définir une infrastructure de stockage pérenne pour les images numérisées du Ministère.

Disponible sur demande auprès de la direction des Archives de France.

6. QUESTIONS JURIDIQUES

Fiches juridiques <http://www.culture.gouv.fr/mrt/numerisation/fr/dll/index.htm>
Questions juridiques liées à l'exploitation des documents numériques

Note d'information DITN/RES/2006/002 du 30 mars 2006. Recommandation du forum des droits sur l'Internet sur la conservation des documents électroniques
<http://www.archivesdefrance.culture.gouv.fr/fr/circAD/DITN.2006.002.pdf>

Loi n° 2006-961 du 1er août 2006 relative au droit d'auteur et aux droits voisins dans la société de l'information (texte publié au Journal officiel).
<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000266350&dateTexte=>

7. PROJETS ET EXPERIENCES EN COURS

Le serveur PIN a une rubrique dans laquelle on trouve des dizaines de projets et d'expériences en cours.

http://vds.cnes.fr/pin/pin_projets.html

Projet SPAR de la Bibliothèque nationale de France

Pour faire face à l'accroissement de ses collections numériques, la BnF a décidé de se doter d'un système capable d'en assurer la préservation sur une période indéfinie. Ce système, appelé SPAR (Système de Préservation et d'Archivage Réparti), doit permettre aux générations futures d'accéder aux archives d'Internet, aux documents imprimés, audiovisuels, et aux images fixes numérisés à des fins de sauvegarde et de consultation. Il sera capable non seulement de stocker des masses importantes de données (dépassant le million de Gigaoctets en 2010), mais aussi d'appliquer les opérations nécessaires à leur pérennisation : le transfert de support, la conversion des formats de données, l'émulation. Des exigences fortes ont été définies, en matière de respect des normes internationales (OAIS), d'interopérabilité (Dublin Core, METS, PREMIS), de modularité, d'extensibilité et de performances.

Une orientation très nette a été prise en faveur du logiciel libre, afin de garantir un niveau maximal d'indépendance. Ce choix se traduit en particulier par le choix du logiciel libre Fedora.

Après le déploiement de l'infrastructure matérielle, puis le travail de spécifications fonctionnelles et informationnelles, le projet est entré dans une nouvelle phase avec le lancement début 2008 de la réalisation logicielle, confiée sur appel d'offres à la société Atos origin.

<http://www.aristote.asso.fr/sem/sem0804.d/009-ATOS-groupe-PIN.pdf>

ATOS Origin, avril 2008

http://vds.cnes.fr/pin/presentations/2007/Presentation_SPAR.pdf

BnF, septembre 2007

Projet PIL@E

Le projet Pil@e, plate-forme PILote d'Archivage Electronique, est né de ce contexte. Le pilote permet de tester un dispositif complet d'archivage depuis la réception des données et leur contrôle, le stockage de façon sécurisée, jusqu'à la consultation par les services internes ou le grand public.

La plate-forme a pour vocation d'accueillir les archives nativement numériques produites par les services centraux de l'État et par conséquent de favoriser l'automatisation de certaines tâches et une meilleure sécurité de la conservation. Les enjeux sont notamment de mettre en œuvre le cadre normatif déjà défini antérieurement mais aussi de tester les points structurants d'un système d'archivage électronique repris point par point ci-dessous.

Note d'information DITN/RES/2006/006 du 11 septembre 2006. Lancement par la direction des Archives de France du marché concernant la réalisation d'une plate-forme d'archivage électronique pilote.

<http://www.archivesdefrance.culture.gouv.fr/fr/circAD/DITN.2006.006.pdf>

http://www.aristote.asso.fr/sem/sem0804.d/002-DAF_pilae_10042008.pdf

Plate-forme d'archivage électronique aux Archives départementales des Yvelines

Conséquence de tous les projets d'administration électronique, la question de l'archivage électronique se pose de manière aiguë: les documents électroniques doivent être conservés sur le long terme et restitués en cas de besoin. C'est pourquoi le conseil général des Yvelines a mis en place une plate-forme d'archivage électronique aux archives départementales des Yvelines : cette plate-forme permet de recevoir, de conserver et de restituer l'ensemble des

documents électroniques qui n'ont plus d'originaux papier. Cette plate-forme est également accessible à l'ensemble des communes du département pour qu'elles puissent verser leurs archives électroniques. Une expérimentation débutera en 2008 avec le versement sur cette plate-forme des délibérations du conseil général et de quelques communes des Yvelines.

PAC : plate-forme d'archivage du CINES (Centre Informatique National de l'Enseignement Supérieur)

Archivage pérenne de documents électroniques du patrimoine scientifique. Février 2008

http://www.cines.fr/pac/download/2008-02-04_05.PAC.pdf

Projet CASPAR

<http://www.casparpreserves.eu/>

Accès, préservation et restitution de ressources culturelles, œuvres artistiques et scientifiques

Début : avril 2006. Durée 3 ans et demi

Financement UE : 8.8 M€

L'IRCAM et l'INA sont partenaires de ce projet qui vise à établir des modèles de préservation des contenus et connaissances numériques basés sur des standards pérennes.

Un cadre pour la préservation des ressources patrimoniales numériques

L'objectif de ce projet pilote est la constitution d'un cadre à la pointe des technologies et des normes actuelles ou émergentes pour la préservation, l'accès et la diffusion de ressources patrimoniales numériques. Sur la base de la norme OAIS1, il prévoit le développement d'un environnement technique générique, adapté à la préservation de données numériques complexes dans différents contextes, et prenant en compte différents aspects tels que la gestion du cycle de vie de préservation/accès, ou l'ingénierie et la formalisation de connaissances, généralement implicites, liées à l'utilisation des données préservées. Cet environnement fera l'objet d'une spécification et d'une validation à travers le développement et l'implantation de plates-formes, dans trois principaux domaines d'application :

- les données scientifiques (notamment spatiales, à travers l'Agence spatiale européenne)
- les données culturelles (UNESCO) ;
- les œuvres artistiques interactives et/ou liées aux arts du spectacle (musique, danse, installations interactives, multimédia, réalité virtuelle) ; cette dernière plate-forme est coordonnée par l'IRCAM en collaboration avec le Groupe de recherches musicales (GRM) à l'INA, l'université de Technologie de Compiègne (UTC), l'université de Leeds et l'International Centre for Art and New Technology (CIANT).

Le projet est actuellement en phase initiale : établissement d'un état de l'art, définition des besoins utilisateurs, spécifications générales d'architecture et spécifications fonctionnelles.

Coordination de la plate-forme artistique

Le rôle de l'IRCAM dans le projet est de coordonner et de mettre en œuvre la plate-forme artistique, à travers notamment la préservation du corpus d'œuvres interactives produites à l'IRCAM. La plate-forme réalisée sera étendue à d'autres domaines artistiques connexes tels que la danse, les arts vidéo, voire les jeux vidéo.

Le rôle de l'INA, à travers le Groupe de recherches musicales, est de participer à la mise en œuvre de la plate-forme artistique et d'assurer la préservation des œuvres électroacoustiques. L'INA a participé au projet Mustica qui préfigurait CASPAR sur les contenus artistiques. Pour CASPAR, il assure la coordination vis-à-vis de la communauté artistique électroacoustique. L'INA contribue à la définition de modèles de préservation des contenus et des informations associées pour les œuvres de support.

Enfin, l'université de Technologie de Compiègne (UTC) intervient à travers l'unité mixte de recherche Heudiasyc (UMR CNRS 6599) au côté de l'INA et de l'IRCAM pour apporter un soutien méthodologique et scientifique. L'UTC participe à la modélisation des œuvres musicales, à leur intégration dans le cadre d'OAIS et enfin au développement des outils d'ingénierie des connaissances (ontologies et outils d'inférence RDF2) nécessaires pour la

préservation non seulement des contenus eux-mêmes mais aussi des informations descriptives liées.

Projet PRESTOSPACE

<http://prestospace-sam.ssl.co.uk/>

Sauvegarde, stockage, accès : pratiques normalisées pour l'archivage des contenus audiovisuels

L'Institut national de l'audiovisuel (INA) coordonne ce projet qui vise à développer des solutions intégrées et une approche industrielle pour rationaliser la sauvegarde et la numérisation des archives audiovisuelles et pour en réduire les coûts, afin de les rendre accessibles aux grands comme aux petits détenteurs de collections. Le Centre de recherches sur la conservation des collections (CRCC) est aussi un partenaire de Prestospace.

8 centres d'archives audiovisuelles + 17 partenaires industriels

Durée 4 ans

Financement UE : 9M€

L'objectif du projet est de fournir des systèmes intégrés et des solutions techniques pour la sauvegarde en numérique de tous types de collections audiovisuelles.

Le projet prévoit de fournir des résultats tangibles dans le domaine de la sauvegarde, de la restauration, de la gestion du stockage et des archives, de la description des contenus, de la livraison et de l'accès aux contenus. On veillera particulièrement à intégrer les facteurs économiques dans la mise en œuvre des services de sauvegarde.

Le but principal est d'aider à construire des usines de sauvegarde, permettant de fournir des services accessibles à tous conservateurs d'archives afin de gérer et de donner accès à leurs collections.

PrestoSpace oriente sa démarche et évalue ses résultats selon trois critères :

- la qualité ;
- la rapidité ;
- la réduction des coûts.

« PrestoSpace » vise à développer un service complet :

- identifier les besoins de chaque utilisateur ;
- concevoir une solution pour toutes les collections audiovisuelles, quel que soit leur taille ;
- améliorer la technologie de lecture et de numérisation pour la vidéo, le film et l'audio ;
- intégrer la technologie dans une chaîne de traitement efficace ;
- sauvegarder mieux, plus vite et à moindre coût ;
- faciliter l'exploitation et proposer des services intégrés adaptés à chaque besoin.

Le projet se décompose en quatre domaines de recherche :

- la sauvegarde : coordonnée par l'Ina, elle permet d'assurer le transfert des archives analogiques vers un ou des formats numériques.
- La restauration : coordonnée par le Joanneum Research Institut de Graz, en Autriche, elle vise à améliorer la qualité des documents.
- Le stockage et le management des archives : coordonnés par la BBC, afin d'optimiser les choix techniques de conservation et de gestion.

DPE Digital Preservation Europe

<http://www.digitalpreservationeurope.eu/>

DPE est une action concertée au niveau européen qui fait suite à ERPANET et regroupe les principaux acteurs de la recherche.

Voir aussi :

http://www.digitalpreservationeurope.eu/publications/reports/Repository_Planning_Checklist_and_Guidance.pdf