

# L'archivage pérenne des documents numériques

Michel Auffret  
CINES  
auffret@cines.fr

## Résumé

*Le constat de départ : la plupart des fichiers informatiques créés il y a dix ans sont aujourd'hui illisibles. Alors que la production et la publication des documents scientifiques tendent vers le tout numérique, une telle situation n'est plus acceptable. Si l'obsolescence technologique est évidemment un facteur inéluctable, il est tout à fait possible d'en freiner et d'en maîtriser les effets néfastes.*

*Si les solutions techniques aux problèmes posés restent aujourd'hui encore très partielles et insatisfaisantes, le « matériel » disponible au plan méthodologique et conceptuel permet de cerner pleinement la problématique et les enjeux, et pose les bases des solutions à venir. Le modèle OAIS - norme ISO 14721 - fournit un cadre abstrait indispensable pour bien comprendre, et pour se poser toutes les bonnes questions.*

*Aujourd'hui, la prise de conscience par les politiques et les décideurs, des enjeux et des risques liés à la perte de l'information numérique se fait rapidement. Les textes récents visant à promouvoir l'économie numérique, et les obligations d'archivage légal qui en découlent, y contribuent fortement.*

*Le CINES, qui a reçu mission du MENESR d'assurer l'archivage pérenne des thèses électroniques et des revues numérisées du portail Persée, met actuellement en place un service générique d'archivage numérique qu'il propose à l'ensemble de la communauté Enseignement supérieur et Recherche.*

*Un nouveau métier se dessine au carrefour de l'informatique, de l'archivistique et de la bibliothéconomie. Si nombre de questions de tous ordres sont encore aujourd'hui sans réponse, la multiplication des initiatives et des projets, dans un contexte d'échange et de coopération, incite résolument à l'optimisme.*

## Mots clefs

Archivage, document numérique, OAIS

## 1 Introduction

Il est tout d'abord essentiel de bien s'accorder sur le sens des mots.

Dans le contexte informatique, le terme « archivage » recouvre aujourd'hui plusieurs sens. Aucun d'eux ne

convient à notre propos d'aujourd'hui. Disons donc rapidement ce que n'est pas l'archivage pérenne. L'archivage pérenne n'est pas une sauvegarde. L'archivage pérenne n'est pas un service HSM - hierarchical storage manager – où l'on migre les fichiers sur bande pour faire de la place sur disque. L'archivage pérenne n'est pas non plus l'ultime étape du stockage des données avant oubli ou perte définitive.

L'archivage pérenne du document numérique a pour objet d'assurer une triple fonction :

1. conserver le document,
2. le rendre accessible,
3. en préserver l'intelligibilité.

Ces trois services étant conçus sur le très long terme. Le très long terme ? Pour fixer les idées, disons 30 ans et au-delà.

Conserver le document dans le temps ? C'est la fonction la plus évidente que l'on demande à un service d'archivage.

Donner accès au document ? En effet, une conservation sans communication serait parfaitement stérile.

Préserver l'intelligibilité du document ? Cet objectif peut sembler à la fois ambitieux et hors de propos. Ambitieux, il l'est sans aucun doute. Mais il se situe bien au cœur du métier d'archiviste. Que le document - ou l'archiviste - soit aujourd'hui « numérique » n'y change rien.

L'échelle de temps est ici un paramètre majeur eu égard au problème posé. Si on se situe à un horizon de l'ordre de 10 ans, le problème est - relativement - simple à traiter. En effet, un stockage informatique de bonne qualité et sécurisé garantit sans doute contre la perte accidentelle du document. L'évolution des technologies n'aura sans doute pas conduit à un bouleversement tel que le document sera devenu irrémédiablement illisible. Et enfin, la communauté des utilisateurs potentiels du document sera vraisemblablement assez proche, scientifiquement et culturellement, de celle qui aura créé le document 10 ans plus tôt. Si on se place à un horizon de l'ordre de 30 ans ou plus, rien de tout cela n'est plus assuré si personne ne s'est préoccupé d'accompagner le document dans le temps. C'est bien cet horizon du très long terme qui constitue le cœur du défi de l'archivage numérique.

L'ère numérique actuelle voit chaque jour croître de façon démesurée, et incontrôlable, la quantité d'information disponible. Dans le même temps, nous sommes témoins du caractère de plus en plus volatil de cette information, comme en témoigne le nombre de liens morts rencontrés sur le web. Toute la production scientifique mondiale est aujourd'hui nativement numérique. Peut-on accepter de laisser perdre cette connaissance quelques mois, ou au mieux quelques années, après sa création ? Evidemment non. Dans le domaine de la recherche, comme partout ailleurs, il est essentiel et il devient urgent

de prendre toutes les dispositions nécessaires pour assurer l'archivage pérenne des documents produits.

## 2 L'obsolescence technologique

Le premier facteur qui entre en jeu dans la problématique de l'archivage pérenne est celui de l'obsolescence technologique.

### 2.1 Un phénomène inéluctable

Mieux que quiconque, nous, informaticiens, sommes tous les jours confrontés au progrès des technologies de l'information. Depuis le début des années 1960 jusqu'à aujourd'hui, c'est à peu près tous les 5 ans qu'une nouvelle génération technologique succède à l'ancienne. Ce dynamisme est certes le signe très positif d'un secteur d'activité toujours en plein essor, mais il se paie chèrement en retour dès qu'on se place dans une perspective de préservation de l'information sur le long terme. Progrès et obsolescence technologiques sont évidemment les deux faces d'une même médaille.

Dressons un petit inventaire des difficultés auxquelles nous aurions à faire face pour relire, par exemple après 10 ans, un fichier informatique pour lequel n'aurait été prise aucune précaution de conservation.

### 2.2 Votre fichier a 10 ans

Tous les supports de l'information numérique, quels qu'ils soient, se dégradent avec le temps. Du plus éphémère - la disquette magnétique - au plus durable - le disque de verre ? - aucun support n'est éternel. Si votre fichier de 10 ans a été conservé sur un support courant, sans recopie et sans précaution de stockage, il a toutes les «chances» d'être aujourd'hui illisible.

L'évolution des technologies de stockage s'accompagne bien sûr du renouvellement des équipements de lecture et d'écriture et dans les logiciels. Si votre fichier de 10 ans est sur une disquette 3,5 pouces - et si cette dernière est encore lisible ! - il est plus qu'urgent de changer de support : les lecteurs de disquettes commencent à se faire rares sur les micro-ordinateurs neufs. Mais, s'il s'agit par exemple d'un document bureautique, il n'est en outre pas du tout certain que la version actuelle de votre traitement de texte soit capable de relire correctement ce document ancien : le principe de compatibilité ascendante atteint assez rapidement ses limites dans la réalité.

Autre difficulté possible : le format de fichier peut avoir disparu et être aujourd'hui complètement inconnu. Si votre fichier de 10 ans a été créé via un logiciel propriétaire et d'usage restreint, il y a tout lieu d'être inquiet sur ce point.

Dernier écueil, et non des moindres : votre fichier de 10 ans, hérité peut-être d'un collègue récemment parti à la retraite, est parfaitement lisible techniquement, mais il est absolument inutilisable car il n'est pas documenté !

## 3 Comment lutter ?

Ce petit tour d'horizon des différents pièges nous permet de mettre en évidence autant de précautions à prendre.

Contre la dégradation et le vieillissement des médias, il faudra prévoir de conserver des copies multiples des documents archivés, tout en diversifiant les technologies de stockage. Dans l'idéal, on peut par exemple souhaiter conserver sur site une copie disque et une copie bande, et une seconde copie bande sur un site distant. Ces seules précautions ne suffiront pas. Il faudra en outre rafraîchir (i.e. renouveler) régulièrement les supports en remplaçant les supports anciens par des supports neufs. Ici également, on veillera à appliquer le principe de précaution. Si tel fabricant réputé sérieux de tel support en garantit la lisibilité sur 10 ans, c'est sans doute qu'il le sait fiable sur une durée de 20 ans. En vertu de quoi, on procédera par précaution au rafraîchissement régulier de tous les supports vieux de... 5 ans, soit la demi durée de vie garantie par le fournisseur.

Pour se prémunir contre la disparition soudaine de matériels ou de logiciels, on mettra en place des dispositifs d'alertes de type veille technologique et économique. En l'occurrence, et contrairement à la veille classique qui s'intéresse aux phénomènes émergents, on s'intéressera ici surtout aux technologies éprouvées en vue d'anticiper leur obsolescence. D'une manière générale, dans ce domaine comme ailleurs, on cherchera à privilégier systématiquement les normes pour éviter les dépendances vis-à-vis de solutions propriétaires.

Pour éviter l'impasse de la disparition des formats de fichier, on optera pour les formats durables. Qu'est-ce qu'un format durable ? C'est avant tout un format publié. C'est-à-dire un format dont les spécifications internes sont librement accessibles. Seul un format publié peut en effet permettre, dans le plus extrême des cas, à un développeur d'écrire un programme spécifique de relecture du fichier. Un format durable est aussi un format très utilisé, ou appelé à le devenir. Un format durable, c'est enfin et si possible un standard voire une norme.

Au regard de ces critères, on peut dresser une liste des principaux formats durables d'aujourd'hui :

- pour les documents texte, on retiendra prioritairement les formats HTML, OpenDocument<sup>1</sup>, PDF, RTF, TXT et XML valide<sup>2</sup>

---

<sup>1</sup> Norme OASIS mise en œuvre dans les suites bureautiques Open Office et Star Office.

<sup>2</sup> Conforme à un schéma ou à une DTD de référence.

- pour les documents image : GIF, JPEG, PNG et TIFF
- pour les documents audio : MP3
- pour les documents vidéo : MPEG et Quick Time

Pour surmonter le dernier obstacle - le manque de documentation - la solution tient en un mot : les métadonnées. Les métadonnées, étymologiquement «les données sur les données », représentent toute l'information sur le document, information qu'il va falloir conserver avec le document pour en assurer l'intelligibilité dans le futur.

## 4 Les métadonnées de pérennisation

L'identification a priori de toutes les métadonnées qui devront accompagner un document numérique est un travail crucial mais difficile. Crucial, car il va conditionner très directement la qualité du service d'archivage et d'accès futur au document. Difficile, car les retours d'expérience en archivage numérique sur le très long terme sont évidemment quasi inexistantes aujourd'hui, et nous évoluons donc dans un domaine encore plein d'inconnu où l'on ne peut avancer qu'en faisant des hypothèses et en prenant quelques risques assumés.

Le point d'ancrage de tout travail autour des métadonnées est très certainement l'ensemble de métadonnées de référence du Dublin Core [1]. Les métadonnées du Dublin Core sont constituées de 15 éléments fondamentaux destinés à décrire toute ressource, au sens large, disponible sur Internet. Ces 15 éléments sont : contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title, type. Ce noyau constitue un premier ensemble de métadonnées descriptives de la ressource, c'est-à-dire ici du document.

Mais à ce premier niveau généraliste de métadonnées descriptives, il est nécessaire d'ajouter des métadonnées plus techniques, spécialisées dans l'activité de pérennisation proprement dite. Plusieurs recommandations existent dans ce domaine. Extraites de normes ou d'études expertes diverses, elles préconisent généralement des ensembles de métadonnées très - trop ? - complets, impossibles à mettre en œuvre dans la réalité.

En pratique, l'identification d'un ensemble de métadonnées d'archivage relève plutôt d'une approche de type « gestion de risques ». Partant d'une ou plusieurs listes très larges, on élimine chaque métadonnée jugée inutile au vu des services futurs prévus ou prévisibles.

A titre d'exemple, on trouvera en annexe la liste des métadonnées retenues pour le système d'archivage du CINES.

## 5 Le modèle OAIS [2]

Le modèle de référence OAIS – Reference Model for an Open Archival Information System - est le résultat d'un travail du CCSDS - Consultative Committee for Space Data Systems (organisme international de normalisation des agences spatiales) - réalisé à la demande l'ISO. Ce travail, auquel ont été associés des représentants de bibliothèques et d'archives institutionnelles, a abouti en 2002 à la rédaction d'un document qui spécifie de manière très générale l'architecture logique et les fonctionnalités d'un système d'archivage. OAIS ne préjuge de la nature des objets à archiver, mais le modèle est particulièrement bien adapté aux objets numériques. Le modèle de référence OAIS est aujourd'hui une norme internationale ISO (ISO 14721). Sa traduction en français, projet de norme ISO, est disponible depuis le printemps 2005 [3].

OAIS est un modèle abstrait. Il définit une terminologie et des concepts. Il identifie les acteurs, décrit les fonctions et les flux d'information, et propose un modèle d'information adapté à la problématique de l'archivage numérique. OAIS n'est pas un recueil de spécifications techniques destinées à être directement implémentées. C'est un guide qui cerne la problématique dans son ensemble et qui contraint à se poser toutes les bonnes questions.

En tant que modèle conceptuel de référence, OAIS s'est aujourd'hui largement imposé, au niveau international, auprès de tous les acteurs institutionnels de l'archivage numérique.

### 5.1 Les acteurs de l'archivage

Le modèle OAIS identifie quatre rôles principaux dans un système d'archivage :

un acteur interne, l'« archive », c'est-à-dire l'opérateur du système d'archivage,

et trois acteurs externes : le « management », les « producteurs » et les « utilisateurs ».

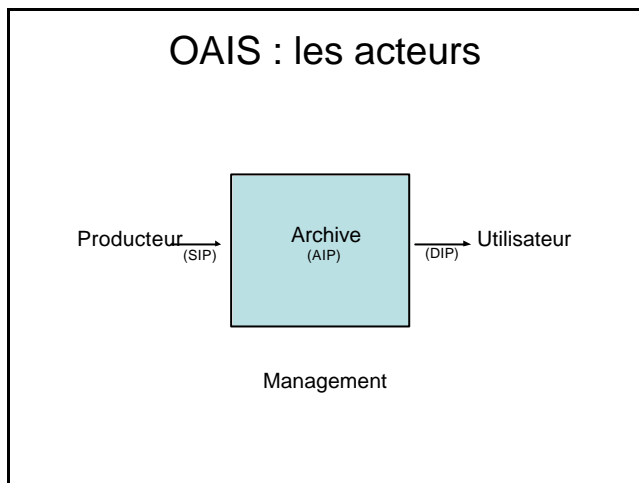


Figure 1 - Les acteurs selon le modèle OAIS

Le « management » assure la fonction de décideur politique. Pour le management, le système d'archivage s'insère dans un plan stratégique global dont il n'est qu'un élément parmi d'autres. C'est évidemment au management qu'il revient de soutenir le dispositif, politiquement, financièrement, et sur le très long terme.

Les « producteurs » sont les personnes, ou plus vraisemblablement les organismes, qui fournissent les objets à archiver. Les objets numériques sur lesquels travaillent les producteurs avant archivage sont les SIP (submission information package). Une fois archivés, ils deviennent les AIP (archival information package) objets internes à l'Archive.

Les « utilisateurs », quant à eux, sont les organismes et les personnes qui ont accès aux objets archivés. Les objets numériques mis à la disposition de utilisateurs sont les DIP (dissemination information package). OAIS identifie une classe particulière d'utilisateurs - la « communauté d'utilisateurs cible » - comme étant la population prioritairement bénéficiaire du service d'archivage. Les services à fournir seront distincts selon que la communauté d'utilisateurs cible est nombreuse ou non, experte ou « grand public », etc.

## 5.2 Les fonctions de l'archivage

OAIS cartographie un système d'archivage selon 6 grands domaines fonctionnels.

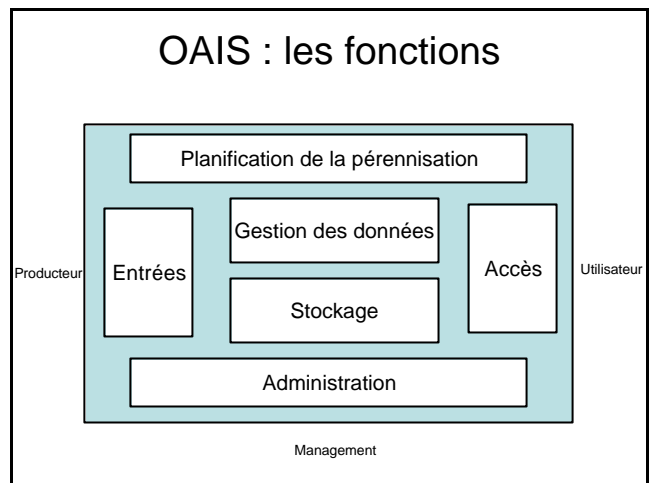


Figure 2 - Les 6 fonctions du modèle OAIS

L'entité « **entrées** » reçoit, contrôle et valide les objets à archiver. Les objets eux-mêmes sont transmis à l'entité « **stockage** », tandis que les informations nécessaires à leur description et à leur gestion dans le temps sont transmises à l'entité « **gestion des données** ».

L'entité « **stockage** » assure la conservation physique des objets archivés. Elle tient les objets archivés à la disposition de l'entité « **accès** ». Conformément aux règles établies par l'entité « **administration** », elle prend en charge la réalisation des copies multiples et le renouvellement des supports anciens.

L'entité « **gestion des données** » prend en charge la tenue à jour de toutes les informations internes - base de données - nécessaires au système d'archivage. Elle fournit aux autres entités du système les informations descriptives des objets archivés (notamment à l'entité « **accès** ») et toutes les informations de gestion techniques et archivistiques nécessaires.

L'entité « **administration** » assure la coordination générale du système. Elle en établit les règles internes. Elle veille à la qualité globale du service rendu et à son amélioration. Elle rend compte au management externe.

L'entité « **planification de la pérennisation** » est la cellule de veille et de planification du système. Elle « **écoute** » l'environnement extérieur et émet des recommandations en vue de procéder aux évolutions nécessaires, notamment aux évolutions technologiques. Elle prépare et planifie ces évolutions. Elle est également responsable du suivi des changements qui peuvent s'opérer dans la communauté d'utilisateurs cible en vue de garantir que le service d'accès reste conforme aux attentes nouvelles des utilisateurs.

L'entité « **accès** » regroupe tous les services qui sont en interface directe avec les utilisateurs. Outre les fonctions de contrôle d'accès, il s'agit principalement de permettre aux utilisateurs de rechercher dans le « catalogue » des objets archivés, et de leur fournir les objets dont ils passent commande.

### 5.3 Les migrations numériques

La préservation sur le très long terme d'un objet numérique ne peut se garantir que grâce à une succession d'opérations de migration. Ces migrations peuvent s'opérer à plusieurs niveaux. Elles peuvent être de simples recopies, ou bien des opérations plus complexes de reformatage.

Le modèle OAIS distingue quatre types de migrations numériques :

Le **rafraîchissement** : copie bit à bit de l'information, dans laquelle un support généralement ancien est remplacé par un support identique généralement neuf.

La **duplication** : recopie des objets archivés vers un nouveau type de support, sans changement de l'organisation logique du stockage. Un exemple typique de duplication est le transfert des objets archivés vers des supports similaires plus capacitifs.

Le **ré-empaquetage** : recopie des objets archivés vers un nouveau type de support, nécessitant une nouvelle organisation logique du stockage. Typiquement, il peut s'agir d'une recopie d'objets archivés sur des supports disques on-line, vers des supports bandes off-line.

La **transformation** : il ne s'agit plus d'une copie, mais d'une réelle modification du contenu de l'information, portant notamment sur sa forme. La transformation peut ou non être réversible. Un exemple de transformation pourrait être la conversion d'un document au format Word de MS-Office vers le format PDF.

*nb : au terme de cette présentation rapide du modèle OAIS, il peut être utile de préciser un point de terminologie qui est parfois source de confusion. Bien que très voisins, les sigles OAIS et OAI désignent des réalités tout à fait indépendantes l'une de l'autre. OAI désigne tous les projets et les réalisations en faveur de l'Open Access Initiative, c'est à dire de l'accès libre à l'information « académique ». Le sigle OAI est parfois développé en Open Archive Initiative (« archives ouvertes » en français), ce qui ne manque pas d'entretenir la confusion avec OAIS. Le terme « Open » de OAIS signifie simplement que le modèle, c'est-à-dire le contenu du document de référence, n'est pas sous « copyright » et qu'il est donc librement diffusable et utilisable. Mais, le modèle OAIS ne préjuge pas du contrôle ou non de l'accès aux informations archivées, et il est donc complètement indépendant des initiatives OAI.*

## 6 D'autres normes

En complément de la norme OAIS, un certain nombre de documents normatifs plus spécialisés viennent éclairer le contexte de l'archivage numérique.

### 6.1 PAIMAS [4]

Tout comme le modèle OAIS, le projet de norme internationale PAIMAS - Producer-Archive Interface Methodology Abstract Standard - est le résultat d'un travail impulsé par le CCSDS.

S'inscrivant directement dans le cadre OAIS, la « norme » PAIMAS est un guide méthodologique pour la conduite formalisée des échanges entre des producteurs et une archive (terme par lequel OAIS désigne le système d'archivage dans son ensemble).

Les expériences réelles d'archives numériques ont mis en évidence que l'une des grandes difficultés de la mise en place d'un service d'archivage se situe dans la non conformité des objets numériques fournis en entrée au système d'archivage. Ces difficultés trouvent elle-mêmes leur origine dans l'insuffisance du travail préparatoire entre producteur et archive. L'objectif de PAIMAS est de pallier cette insuffisance en proposant une démarche structurée de dialogue préalable.

### 6.2 AFNOR NF 42-013 [5]

La norme AFNOR NF 42-013 est le fruit d'un travail qui a réuni les professionnels français de la GED (Gestion électronique de documents).

Elle établit des règles techniques précises, et contraignantes, pour l'archivage « légal » des documents d'entreprise. L'archivage dit légal recouvre toutes les obligations d'archivage qui incombent aux entreprises, et à tous les acteurs économiques, depuis la loi de mars 2000 qui donne au document électronique la même valeur juridique probante que le document papier.

En l'état actuel, la norme NF 42-013 ne reconnaît comme support possible de l'archivage « légal » que les supports optiques non ré-inscriptibles de type WORM (Write Once Read Many). Cette exigence exclusive n'est plus conforme à la réalité des technologies de stockage où le « WORM logique » sur supports magnétiques offre aujourd'hui des garanties d'intégrité comparables au « WORM physique ». Par ailleurs, ce parti pris technologique de la norme lui interdit sans doute de viser le statut de norme internationale

Difficile à mettre en œuvre et « sectaire » quant à ses choix technologiques, la norme NF 42-013 contient cependant de nombreuses recommandations utiles.

### 6.3 ISO 15489 [6] et MoReq [7]

Pour compléter ce bref panorama des normes sur l'archivage, il convient de mentionner la norme ISO 15489 (Records Management) et le document européen MoReq. L'un comme l'autre sont le résultat de travaux récents menés par des professionnels de l'archivistique. Tous les deux sont empreints de la conception anglo-saxonne du processus d'archivage, conçu et mis en œuvre dès le début

du cycle de vie du document. Ces deux documents sont parfaitement adaptés au contexte de l'archivage des objets numériques pour lequel ils fournissent de très précieuses recommandations.

## 7 Le contexte

La problématique de l'archivage pérenne des documents numériques est sortie d'une relative confidentialité dans le début des années 2000. Est-ce une conséquence indirecte du passage réussi à l'an 2000 - et maintenant face à l'« éternité » que peut-on faire ? Toujours est-il qu'on observe depuis quelques années une réelle « ébullition » autour de cette question.

En Europe et en France, plusieurs textes de loi ont contribué à la sensibilisation des acteurs économiques. La loi de mars 2000 qui donne à l'écrit numérique la même valeur de preuve que l'écrit papier, les décrets de mars 2001 et d'avril 2002 relatifs à la signature électronique, et enfin la loi de juin 2004 pour la confiance dans l'économie numérique (LEN).

Parallèlement, les initiatives des gouvernements successifs en faveur du développement de l'e-administration vont dans le même sens : déclaration de revenu en ligne, dématérialisation des marchés publics, programme ADELE de l'ADAE, etc.

Aux Etats-Unis, les attentats du 11 septembre 2001, le scandale Enron, la loi Sarbanes-Oxley, ... sont venus renforcer les exigences sur les plans de reprise d'activité et sur les besoins de traçabilité.

Le droit de la propriété intellectuelle vient également interférer dans ce panorama pour tout ce qui touche aux documents scientifiques et aux oeuvres patrimoniales. Il n'est sans doute pas inutile de rappeler que le document électronique est soumis au même droit de la propriété intellectuelle que l'écrit papier. Il s'ensuit que, par exemple, une « œuvre numérique » ne tombe dans le domaine public que 70 ans après le décès du dernier auteur. Et donc qu'avant cette échéance, toute copie du document - y compris pour archivage - nécessite la cession explicite d'un droit de reproduction, et toute diffusion du document - par exemple sur Internet - nécessite la cession explicite d'un droit de représentation.

Dans ce contexte foisonnant, face à des questions aussi nombreuses qui mobilisent des compétences très diverses, la mise en commun des expériences est indispensable. En France, le groupe PIN - pérennisation des informations numériques - de l'association Aristote, constitue un lieu privilégié d'échange. Ce groupe expert rassemble de nombreux organismes : le CNES, la BnF, les Archives de France, le CEA, ... Le groupe PIN tient des réunions trimestrielles et propose un site web [8].

## 8 Le service d'archivage du CINES

Le MENESR a donné mission au CINES d'assurer l'archivage pérenne des thèses électroniques et des revues numérisées du portail Persée. D'autres projets en cours de réalisation - le CENS et la diffusion des abonnements Elsevier - ont également des besoins d'archivage approchants. Cette multiplication de projets d'archivage a incité le CINES à concevoir et à mettre une œuvre une solution d'archivage générique qui est aujourd'hui proposée à l'ensemble de la communauté Enseignement supérieur et Recherche, dès lors que le corpus de documents numériques à archiver est reconnu d'intérêt national. Les premiers tests d'archivage pérenne des thèses électroniques sont prévu d'ici la fin de l'année 2005.

Fonctionnellement, le système d'archivage est vu de l'extérieur comme un ensemble de trois serveurs logiques :

- un serveur de versement
- un serveur de stockage
- un serveur d'accès

### 8.1 Le serveur de versement

- Le **producteur** (personne ou système informatique) intervient dans le cadre d'un **projet d'archives** donné, et soumet un document (i.e. fichier(s) et métadonnées) **ou un lot de documents** au CINES,
- Le CINES **accuse réception** de la soumission,
- Le CINES procède à un ensemble de **contrôles de validité technique**<sup>3</sup> de la soumission (validation et complétude des métadonnées, validation du format, cohérence entre fichier(s) et métadonnées, ...liste fermée à établir), phase pendant laquelle des échanges complémentaires peuvent avoir lieu avec le producteur,
- A tout moment durant cette phase, le producteur peut souhaiter **annuler la soumission** (envoi d'un avis d'annulation au CINES), ou le CINES peut être contraint de **rejeter la soumission** (envoi d'un avis de rejet au producteur),
- Le producteur peut **interroger le serveur de versement** pour connaître la liste de ses documents en attente d'archivage,
- Dès que la soumission est valide, et au plus tard dans un délai d'une semaine après l'envoi de l'accusé de réception, le document (ou le lot) est transmis au serveur de stockage.

---

<sup>3</sup> Les contrôles de validité « intellectuelle » sont faits en amont et sont de la responsabilité du producteur.

## 8.2 Le serveur de stockage

- Dès réception d'un document, le serveur de stockage envoie le **certificat d'archivage du document** au producteur, l'informant notamment de l'**identifiant unique et pérenne du document** dans le système d'archivage,
- Durant la vie du document dans le système d'archivage, 3 types d'opérations peuvent intervenir :
  1. des **opérations d'exploitation courante** : le producteur n'en est pas informé (ex. sauvegarde des métadonnées, migration physique de support),
  2. des **opérations avec avis** : le producteur est informé de la réalisation de l'opération (liste à convenir au cas par cas pour chaque projet d'archives),
  3. des **opérations sur accord** : le CINES ne procède à l'opération que s'il a reçu l'accord préalable du producteur (liste à convenir au cas par cas pour chaque projet d'archives),
- Tout événement, action, ... sur un document archivé est **horodaté, historisé et traçable** (au minimum : quand ? qui ? quoi ? et si possible : pourquoi ? comment ?),
- Le producteur peut **interroger le serveur de stockage** pour connaître la liste de ses documents archivés,
- Le producteur **peut tenir à jour les métadonnées** de ses documents archivés (celles dont il a la responsabilité), **toute modification de métadonnée est historisée**,
- Le CINES doit tenir à jour les métadonnées qui sont de sa responsabilité, toute modification étant historisée de même,
- Le CINES adresse annuellement au producteur la **liste de ses documents archivés**.

## 8.3 Le serveur d'accès

- **L'accès au serveur d'accès est contrôlé**,
- Un **utilisateur** (personne ou système informatique) peut **consulter le catalogue complet** du système d'archivage, ou le **sous-catalogue** d'un projet d'archives particulier,
- Un utilisateur autorisé à consulter le catalogue complet (ou un sous-catalogue) du système d'archivage, peut visualiser ou **obtenir une copie des métadonnées** d'un document archivé,
- Selon les droits associés à un document archivé, un utilisateur peut être autorisé à **passer commande d'un document** pour en **recevoir une copie** sur son poste de travail (ou système informatique distant).

## 9 Un nouveau métier pour un défi nouveau

L'archivage pérenne des documents numériques représente un défi nouveau. Les difficultés à surmonter sont réelles, et pour beaucoup d'entre elles encore méconnues. L'adversaire - le temps - est coriace et sort souvent gagnant des combats qu'il engage. Les solutions techniques - informatiques - aux problèmes posés sont aujourd'hui partielles et insatisfaisantes.

Mais à la réflexion et avec un peu de recul, on s'aperçoit que les efforts à faire en priorité ne sont pas d'ordre technique. Aujourd'hui, les premières réponses à apporter sont :

1. d'ordre culturel : un renforcement des collaborations entre informaticiens, archivistes et bibliothécaires est indispensable,
2. d'ordre organisationnel : de nouveaux processus métier et de nouvelles méthodes sont à définir, à mettre en oeuvre et à maîtriser.

Rien n'empêche d'entreprendre dès aujourd'hui ces deux démarches tout en mettant au banc d'essai des solutions techniques encore incomplètes dont sait de toute façon qu'elles auront une durée de vie limitée.

Rendez-vous donc dans ... 30 ans !

## Annexe

Ensemble des métadonnées du système d'archivage du CINES :

Métadocument	Métadonnées fichier	Définition
contributor		contributeur(s), autre(s) que l'auteur principal, au document
coverage		périmètre ou domaine d'application du document
creator		créateur (généralement auteur principal) du document
date		date du document
description		description, résumé ou sommaire du document

Métadatas document	Métadatas fichier	Définition
<b>format</b>		manifestation physique ou numérique du document
<b>identifiant</b>		référence unique du document dans le système d'archivage du CINES
<b>langage</b>		langue (principale) du contenu du document
<b>publisher</b>		entité responsable de la mise à disposition du document
<b>relation</b>		référence(s) à un(des) document(s) connexe(s) archivé(s) (au CINES ou ailleurs)
<b>rights</b>		droits associés au document
<b>source</b>		référence(s) à une(des) ressource(s) dont est dérivé le document
<b>subject</b>		sujet du document ou liste de mots clés
<b>title</b>		titre du document
<b>type</b>		nature ou genre du contenu du document
<b>authenticité</b>		attestation de l'authenticité du document
<b>dateArchivage</b>		date et heure de l'archivage du document
<b>duréeConservation</b>		durée prévue de conservation du document
<b>noteDocument</b>		note complémentaire relative au document
<b>producteur</b>		identification de l'organisme qui a versé le document

Métadatas document	Métadatas fichier	Définition
<b>projet</b>		identification du projet d'archives CINES pour lequel le document a été versé
<b>sortFinal</b>		ce qu'est devenu le document après son retrait du système d'archivage du CINES
<b>structureDocument</b>		description de la structuration du document en fichiers (peut être l'identifiant d'un document archivé qui décrit la structure)
<b>URI</b>		URI du document
<b>version</b>		version du document
<b>versionPrécédente</b>		référence de la version précédente archivée du document (au CINES ou ailleurs)
	<b>compression</b>	algorithme ou logiciel de compression / décompression du fichier (si compression externe au format)
	<b>empreinte</b>	valeur de l'empreinte numérique du fichier dans son format courant
	<b>encodage</b>	type et version de l'encodage des caractères dans le fichier
	<b>formatFichier</b>	nom et version du format du fichier
	<b>idDocument</b>	identifiant du document auquel est rattaché le fichier
	<b>migration</b>	description de la dernière migration effectuée sur le fichier
	<b>nomFichier</b>	référence complète (chemin d'accès, nom) du fichier au CINES



Métadocument	Métadocument fichier	Définition
	<b>noteFichier</b>	note complémentaire relative au fichier
	<b>structureFichier</b>	description de la structure du fichier (peut être l'identifiant d'un document archivé qui décrit la structure)
	<b>tailleEnOctets</b>	taille du fichier en octets dans son format courant

## Bibliographie

- [1] <http://dublincore.org/>
- [2] <http://www.ccsds.org/documents/650x0b1.pdf>
- [3] [http://vds.cnes.fr/pin/documents/projet\\_norme\\_oais\\_version\\_francaise.pdf](http://vds.cnes.fr/pin/documents/projet_norme_oais_version_francaise.pdf)
- [4] <http://www.ccsds.org/CCSDS/documents/651x0b1.pdf>
- [5] AFNOR. Recommandations 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. 2001.
- [6] AFNOR. Records management - Gestion et conservation des documents en toute sécurité. 2004.
- [7] [http://www.adbs.fr/site/publications/texte\\_ref/Moreq.pdf](http://www.adbs.fr/site/publications/texte_ref/Moreq.pdf)
- [8] <http://vds.cnes.fr/pin/>