

Archivage électronique

Problématique et normes

par **Marie-Anne CHABIN**

Expert en stratégie et méthodologie d'archivage

Présidente du cabinet Archive 97

Professeur associé au Conservatoire national des arts et métiers CNAM

1. Qu'est-ce que l'archivage ?	H 7 276v2 - 2
2. Enjeux de l'archivage dans le contexte électronique	— 3
3. Environnement réglementaire et besoins métiers	— 4
4. Réponses techniques	— 5
5. Aspects organisationnels et acteurs de l'archivage électronique	— 6
6. Coûts de l'archivage électronique	— 7
7. Présentation des quatre principales normes pour l'archivage électronique	— 8
7.1 Norme ISO 15489 (2001).....	— 8
7.2 Modèle européen MoReq2 (2008)	— 9
7.3 Norme NF Z42-013 (2008).....	— 10
7.4 Modèle de référence OAIS (2001).....	— 11
8. Conclusion	— 12
Pour en savoir plus	Doc. H 7 276v2

Une partie des documents et des données qui sont produits au quotidien dans les entreprises doit être archivée pour être conservée pendant quelques années ou quelques décennies, afin de se conformer aux contraintes réglementaires ou pour constituer une documentation interne pour les équipes. C'était vrai dans l'environnement papier et c'est encore plus vrai dans l'environnement électronique. En effet, si les papiers savent plus ou moins se conserver tous seuls, les fichiers électroniques livrés à eux-mêmes sont rapidement inexploitable et illisible. L'archivage électronique doit répondre aux problèmes posés par l'**obsolescence des supports et des formats de codage**, ainsi qu'à la nécessité d'identifier et de décrire les documents et les données numériques.

Pour l'entreprise, il s'agit d'anticiper le risque de non-disponibilité des données en cas de problème juridique ou technique, et de le faire au meilleur coût.

L'archivage électronique est aujourd'hui encadré par plusieurs normes qui traitent chacune un ou plusieurs aspects de la question : la **norme internationale ISO 15489** propose des principes directeurs, des procédures et une démarche pour maîtrise de l'archivage ; le **modèle européen MoReq2** détaille les spécifications fonctionnelles d'un système d'archivage électronique (Electronic Records Management System) ; la **norme française NF 42-013** définit le cadre et les procédures de conservation des documents numériques natifs ou numérisés ; la **norme OAIS** définit un modèle d'information et un modèle fonctionnel pour la pérennisation de l'information numérique.

Notation et acronymes	
AIP	<i>Archival Information Package</i>
DIP	<i>Dissemination Information Package</i>
DTD	Définitions de Types de Documents
GED	Gestion électronique de documents
ILM	<i>Information Lifecycle Management</i>
InterPARES	<i>International Research on Permanent Authentic Records in electronic Systems</i>
LOCKSS	<i>Lots of Copies Keep Stuff Safe</i>
MoReq	<i>Model Requirements for the Management of Electronic Records</i>
OAIS	<i>Reference Model for an Open Archival Information System</i>
PAIMAS	<i>Producer-Archive Interface Methodology Abstract Standard</i>
PDF	<i>Portable Document Format</i>
PDI	<i>Preservation Description Information</i>
PIN	pérennisation de l'information numérique
PIST	Pérennité, Intégrité, Sécurité, Traçabilité
SAE	Système d'archivage électronique
SIP	<i>Submission Information Package</i>
SPAR	système de préservation et d'archivage réparti
WARC	<i>Web ARChive file format</i>
W3C	<i>World Wide Web Consortium</i>
XML	<i>eXtensible Markup Language</i>

1. Qu'est-ce que l'archivage ?

Au premier abord, archiver, c'est simple : on stocke dans un lieu retiré les documents qui ne sont plus d'usage courant. Au premier abord seulement, car archiver n'est pas stocker et c'est bien dans la différence entre ces deux mots que l'on peut le mieux appréhender la problématique de l'archivage.

On stocke généralement des matériaux, des produits, des fournitures, des imprimés, c'est-à-dire des objets existants en X exemplaires en attendant de les utiliser, de les transformer, de les distribuer les uns après les autres, dans le cadre d'une activité industrielle, commerciale, administrative, etc. Chaque modèle est identifié et les exemplaires sont numérotés ou code-barrés pour une meilleure gestion des stocks, justement. Si dans un lot, un exemplaire est défectueux, on le remplace par le suivant car les différents éléments d'une série sont interchangeables.

Archiver, c'est autre chose. On archive des documents, c'est-à-dire des écrits, des traces qui résultent d'une transaction, d'une opération, d'une réflexion, qui s'est produite à un moment donné, dans un contexte donné, dans un lieu donné, avec des acteurs donnés. Chaque archive est donc un objet unique, qui a des coordonnées dans le temps et dans l'espace et qui, par conséquent, n'est pas interchangeable avec un autre. Archiver a pour but de préserver pour demain ou après-demain la valeur de preuve et de mémoire d'une information consignée aujourd'hui sur un support et dotée des signes de validation qui la rende authentique. Si un document n'a pas d'intérêt pour l'avenir, il n'y a aucune raison de le conserver, et donc aucun intérêt à l'archiver.

La loi du 3 janvier 1979 (révisée en 2008) donne la définition d'archives : « Les archives sont l'ensemble des documents, quels que soient leur date, leur forme et leur support matériel, produits ou reçus par toute personne physique ou morale, et par tout service ou organisme public ou privé, dans l'exercice de leur activité » et précise que « la conservation de ces documents est organisée dans l'intérêt public tant pour les besoins de la gestion et de la justification des droits des personnes physiques ou morales, publiques ou privées, que pour la documentation historique de la recherche ».

Ce besoin de conserver la trace d'aujourd'hui pour agir demain n'est pas nouveau. Les éleveurs mésopotamiens du deuxième millénaire avant J.-C. notaient au fil du temps sur des tablettes d'argile le décompte de leurs troupeaux ; ces tablettes, archivées, c'est-à-dire rangées dans un endroit sûr, permettaient de mieux gérer les exploitations. De surcroît, 4 000 ans plus tard, elles nous racontent comment vivaient ces gens.

Pourquoi archive-t-on aujourd'hui ? En premier lieu, **pour être en règle avec la loi** car l'activité des entreprises est encadrée par des lois qui exigent la tenue et la conservation de certains documents, comme les procès-verbaux du Conseil d'administration ou les déclarations fiscales. Ensuite, on archive **pour se prémunir contre une procédure judiciaire** : toute fourniture de biens ou de services, toute action de rémunération est sujette à contestation tant que la prescription légale n'est pas échu. Pour tous les documents transactionnels, le délai de prescription équivaut à la durée d'archivage : 5 ans en matière de salaires, 5 ou 10 ans en matière comptable et commerciale, 30 ans après la fin de l'affaire pour toute opération mettant en cause des biens immobiliers. Pour des actes simples, on archive des pièces ; pour des actes plus complexes, on archive des dossiers, c'est-à-dire des ensembles de pièces (contrats, courriers, études, avec leurs justificatifs) qui peuvent apporter leur témoignage, individuellement et collectivement, en cas de besoin de reconstitution de l'affaire.

Enfin, on archive **pour la mémoire** : garder la trace de ce qui a été dit et fait pour ne pas avoir à le refaire dès lors que l'on est susceptible d'avoir de nouveau besoin des mêmes informations. C'est le cas des dossiers de brevets ou de projets non aboutis à une époque mais qui pourraient être relancés ; c'est le cas des études prospectives, des statistiques, des analyses scientifiques, etc.

À noter que la valeur d'un document archivé peut être à la fois légale et informative. Les plans d'un immeuble locatif sont soumis à la prescription civile de 30 ans mais sont aussi des documents de travail indispensables lorsque l'on veut réhabiliter l'immeuble vingt ans après. En n'archivant pas ces plans, on prend un double risque : risque juridique et risque de devoir payer (cher) la réfection des plans *a posteriori*.

Quand les documents archivés ont perdu leur valeur probante ou opérationnelle pour l'entreprise, ils conservent, pour une petite part d'entre eux, une valeur de mémoire collective, de témoignage sur l'entreprise, sur les personnes qui l'ont fait vivre, et participent à ce titre aux sources de l'histoire de la société. Ils sont alors versés aux archives historiques et, selon les délais de communicabilité définis par la loi, accessibles à tout public.

L'archivage est souvent associé à la **GED** (gestion électronique de documents) dans une relation pas toujours très claire : « archiver dans la GED », « GED documentaire ou GED d'archivage », « mettre en œuvre une solution de GED/archivage complète », etc. Si l'on conçoit l'archivage, basiquement, comme le stockage, dans un outil partagé d'une information avec sa fiche d'identité, la fonctionnalité d'archivage apparaît comme un prolongement de la période d'utilisation, afin de ne pas supprimer le document « au cas où ». Mais l'archivage au sens fort (avec cette notion ancestrale d'archives constituant des traces originales) est plus que cela : il doit permettre de maîtriser l'information jusqu'à sa destruction éventuelle, et donc de contrôler les critères de cette destruction (ou de la non-destruction). De ce point de vue, l'archivage s'oppose à la GED.

L'archivage est décrit dans ce dossier comme le processus qui va prendre en charge un document, au moment où il est terminé

et validé et l'accompagner, le maintenir en bon état tout au long de sa vie, jusqu'à l'échéance de sa durée de conservation. Les étapes cruciales de l'archivage sont :

- l'**identification** du document ou du dossier pertinent à archiver ;
- la détermination de sa **durée de conservation** et du point de départ de ce délai ;
- la **description minimale** de cet objet d'information pour pouvoir accéder rapidement à son contenu ;
- le **stockage** proprement dit (dans un lieu adéquat) ;
- la **maintenance de l'accessibilité matérielle et intellectuelle** pendant la période requise, qu'il s'agisse de quelques années ou de plusieurs décennies, voire de plusieurs siècles.

Il faut voir **deux étapes clés** dans le processus d'archivage : la première, plutôt courte, dynamique, qui consiste à faire passer le document à archiver de son environnement natif vers le système d'archivage et l'autre, plus consistante et qui peut être assez longue, qui vise à maintenir l'information archivée dans le temps. L'article « archivage électronique » de Wikipédia (qui convient du reste à l'archivage non électronique) distingue ces deux étapes par deux séries de verbes éclairantes :

- l'archivage de contenus électroniques est l'ensemble des actions, outils et méthodes mis en œuvre pour **réunir, identifier, sélectionner, classer et conserver** des contenus électroniques, sur un support sécurisé, dans le but de les exploiter et de les rendre accessibles dans le temps, que ce soit à titre de preuve (en cas d'obligations légales notamment ou de litiges) ou à titre informatif ;
- la conservation est l'ensemble des moyens mis en œuvre pour **stocker, sécuriser, pérenniser, restituer, tracer, transférer voire détruire**, les contenus électroniques archivés.

2. Enjeux de l'archivage dans le contexte électronique

Dans l'**environnement papier**, le processus d'archivage est rigoureux mais simple et il ne devient complexe que lorsque les archives elles-mêmes sont nombreuses, hétérogènes et de conservation longue. Les documents à archiver sont identifiés à l'aide d'un référentiel d'archivage qui recense les types de documents produits et leur durée de conservation, puis décrits et mis en boîtes ; les boîtes sont numérotées, placées éventuellement dans des conteneurs et stockées sur les rayonnages *ad hoc* dans des locaux présentant les conditions requises en termes de température, hygrométrie, éclairage, protection contre l'incendie et l'intrusion. Les mouvements d'entrée (versements) et sortie (élimination, communication) sont suivis à l'aide d'un outil informatique.

Avec l'**électronique**, la question de l'archivage doit être repensée et repensée. En effet, avec 90 % de l'information globale produite *via* des outils électroniques (bureautique, bases de données, sites Web, messagerie), l'électronique va devenir peu à peu le support naturel d'archivage. Il est logique, comme on l'a toujours fait, d'archiver les documents sous leur forme initiale de production et de validation.

Or, la forme des documents électroniques se révèle plus délicate à gérer que la feuille de papier ou le registre traditionnels. Certes, les systèmes électroniques produisent des fichiers simples, texte ou image, dont la logique de gestion sérielle est comparable à la gestion d'archives dans l'environnement analogique, mais ils produisent aussi, et de plus en plus, des données complexes parce que composites et hétérogènes, accompagnées de métadonnées :

- d'un côté, le **contenu** mêle du texte, de l'image, du son, codés chacun avec un format spécifique ;
- de l'autre, la **structure et la présentation** de ces données qui peuvent se référer à une feuille de style, à un modèle extérieur par lequel il faudra passer pour recomposer le document et le donner à voir à l'utilisateur.

On note aussi que la notion de **dossier**, matérialisée dans l'environnement papier par une chemise cartonnée, est complètement virtualisée dans l'environnement électronique : l'appartenance au dossier se gère par l'attribution d'un code ou identifiant de dossier commun aux différentes pièces du dossier sans que celles-ci aient besoin d'être physiquement regroupées.

Autre élément à prendre en compte : plus encore que dans l'environnement papier, l'**information** que l'on veut ou doit pérenniser n'est pas seulement le contenu du document ; ce peut être la date d'envoi ou la date de validation (qu'est-ce qui a été envoyé à qui et quand ?). Il faut donc gérer non seulement les contenus documentaires mais aussi le contexte de production et de diffusion, suivre le document à la trace au long de son cycle de vie.

D'une manière générale, et plus encore pour l'archivage, la grande révolution de l'électronique est que cette technologie dissocie le **contenu** et le **support** en faisant deux éléments distincts dont l'agrégation n'est plus naturelle : l'écrit est décomposé en code binaire et les trains de bit correspondants enregistrés sur une bande ou un disque, avec processus de recomposition du code en écrit intelligible lors de la lecture. De plus, l'utilisateur perd la vision globale de l'ensemble puisque, lorsqu'il voit le contenu sur un écran, il ne voit pas le support et inversement, lorsqu'il voit un disque, il ne peut en appréhender, même sommairement, le contenu.

Tout irait bien si ces processus de décomposition-recomposition étaient stables, disons au moins aussi stables que l'encre sur les parchemins médiévaux. Mais le grand défi du document électronique pour l'archivage est son **instabilité dans le temps**.

L'instabilité de l'information électronique se manifeste dans les **formats et supports**. La durabilité des uns et des autres est souvent plus courte que les durées de conservation requises. Les supports du type disque dur ou disquette ne sont pas du tout adaptés au stockage moyen ou long terme. Les cartouches et bandes magnétiques se maintiennent entre 15 à 30 ans. Certains supports optiques (cédérom, CD-WORM, DVD, DON WORM) annoncent une durabilité intrinsèque de 20 à 50, voire 100 ans pour le « *Century disc* » en verre trempé mais ce n'est pas le cas des formats.

À noter que ce problème n'est pas propre à l'électronique puisque la vidéo présente également une durée de vie limitée (quelques décennies) et des supports non lisibles à l'œil humain (ce qui reste malgré tout possible pour un film ou un microfilm). De ce fait, les premières bandes magnétiques 2 pouces de l'Institut national de l'audiovisuel (INA), en passe de devenir inexploitable, ont été numérisées il y a quelques années, bien que cette opération ne puisse être considérée comme définitive.

La **conservation des supports** n'est pas qu'un problème technologique. Des choix économiques peuvent intervenir chez les éditeurs soit en fabricant à moindre coût des produits de moindre qualité, soit en introduisant des composantes aux effets attractifs dans l'immédiat mais néfastes à la pérennisation, à l'instar des fabricants de papier qui utilisent du chlore pour le blanchir au détriment de sa conservation. Seules des normes peuvent interdire ou limiter les dérives du marché.

La **durabilité des systèmes d'encodage de l'information**, plus courte que celles des supports, est donc encore plus grave. Les formats d'enregistrement n'ont pas toujours inclus dans le passé, ni même encore aujourd'hui, la compatibilité avec les versions précédentes des logiciels et on trouve de nombreux exemples de disquettes ou de cédéroms en bon état physique mais illisible après 3 ou 5 années seulement de « non-consultation ».

S'ajoute à cela l'évolution des systèmes d'exploitation et des outils de lecture des données électroniques archivées. On trouve le cas de données stockées en bon état avec un système de codage connu et interprétable mais pour lesquelles le logiciel qui permet d'accéder aux supports n'est pas maintenu par l'éditeur... L'archivage, rappelons-le, est comptable du bon accès à l'information jusqu'à l'échéance de son intérêt probant ou informatif.

Un autre défi est la **sécurité du document** pour la préservation de son intégrité. Il est indispensable de sécuriser le document

contre les manipulations indues lors de sa confection, de sa validation, de son stockage et de sa consultation, c'est-à-dire tout au long de son cycle de vie. La production et la conservation des archives papier requièrent aussi des mesures de sécurité et les siècles passés ont connu de nombreuses affaires de faux et de vols d'archives. Mais l'électronique, avec la malléabilité de l'écrit électronique et le développement des réseaux, exige de nouvelles mesures.

Enfin, avec l'électronique, le volume des données a explosé. Certes, la plus grande partie de l'information produite peut être éliminée à brève échéance (information périmée, éphémère, redondante...) mais il en reste, proportionnellement, des volumes toujours plus importants à archiver.

En résumé, l'archivage électronique, comparé à l'archivage traditionnel, s'est enrichi de trois aspects essentiels à un archivage de qualité :

1. une prise en compte des documents archivables dès leur production, selon les principes du **records management** (cf. § 7.1) afin de contrôler valablement le cycle de vie ;
2. la **pérennité** des données numériques qui doivent rester lisibles pendant plusieurs années ou décennies, grâce, essentiellement, à une maîtrise permanente des formats ;
3. un dispositif de **sécurité** qui permette de préserver l'authenticité des documents émis et transmis, et d'assurer l'intégrité des fichiers numériques dans le temps, grâce à un contrôle rigoureux du processus de capture du système de production vers le système d'archivage et à une traçabilité au sein du système d'archivage.

3. Environnement réglementaire et besoins métiers

Pour l'archivage électronique de documents natifs (l'archivage au sens fort du terme), le point de départ législatif est la **loi n° 2000-230 du 13 mars 2000** portant adaptation du droit de la preuve aux technologies de l'information et relative à la **signature électronique**, loi qui transpose dans l'hexagone la directive du 13 décembre 1999 du Parlement européen sur un cadre communautaire pour les signatures électroniques.

Cette loi modifie en profondeur le code civil français dont l'article 1316-1 est devenu « *L'écrit sous forme électronique est admis en preuve au même titre que l'écrit sur support papier, sous réserve que puisse être dûment identifiée la personne dont il émane et qu'il soit établi et conservé dans des conditions de nature à en garantir l'intégrité* ».

Un an plus tard est mis en place le dispositif de **certificats électroniques** lié à l'exercice et au contrôle de la **signature électronique** (décret n° 2001-272 du 30 mars 2001). Le certificat électronique est joint à la signature et en garantit l'origine ; il est fourni par un tiers « prestataire de services de certification électronique » accrédité. Il établit le lien entre l'identité du signataire et la clé publique qu'il utilise.

Depuis, la réglementation française s'enrichit régulièrement de nouveaux textes qui précisent pour tel ou tel domaine ou action les conditions d'admissibilité en preuve d'écrits électroniques archivés et conservés électroniquement.

Les principaux textes sont les suivants.

– Le nouveau code des **marchés publics** (voir décret n° 2002-692 du 30 avril 2002 relatif à la dématérialisation des procédures de passation des marchés publics) prévoit que, depuis 2005, les collectivités et services publics doivent être équipés pour diffuser leurs appels d'offre sous forme électronique et pour recevoir les

offres des fournisseurs sous forme numérique. Sont concernés tant les documents transmis aux prestataires (règlement de consultation, cahier des charges, documents complémentaires) que les réponses (dossiers de candidature et offres).

– Depuis juillet 2003, les **factures** peuvent être émises et conservées sous forme électronique, grâce à l'usage d'un certificat électronique non qualifié (moins lourd techniquement et administrativement parlant que le certificat initial). L'article 289 bis du code général des impôts CGI précise : « Les factures transmises par voie électronique dans les conditions prévues aux articles 289-V et 289 bis du CGI tiennent lieu de facture d'origine pour l'application des articles 286 et 289 du même code. Elles doivent donc être conservées dans leur format original dans les délais et conditions prévus par l'article L. 102 B du livre des procédures fiscales.

– La loi du 21 juin 2004 pour la confiance dans l'économie numérique (LCEN) apporte quant à elle un complément à l'article 1108 du code civil : « *Lorsqu'un écrit est exigé pour la validité d'un acte juridique, il peut être établi et conservé sous forme électronique dans les conditions prévues aux articles 1316-1 et 1316-4 et, lorsqu'un acte authentique est requis, au second alinéa de l'article 1317.* »

– Un décret du 16 février 2005 stipule que « les **contrats** conclus par voie électronique dont le montant est supérieur à 120 euros doivent obligatoirement faire l'objet d'un archivage ». Il est précisé que « la durée de conservation est de 10 ans à compter de la conclusion du contrat lorsque la livraison du bien ou de la prestation est immédiate ».

– Diverses ordonnances de l'année 2005 s'efforcent de simplifier l'accomplissement de certaines formalités contractuelles par voie électronique ou de faciliter les échanges électroniques entre les usagers et les autorités administratives (ce que l'on appelle la **e-administration**).

– En 2005 toujours, un décret du 10 août prévoit la possibilité de passer des **actes notariés** sans le recours au papier : « Le notaire qui établit un acte sur support électronique utilise un système de traitement et de transmission de l'information agréé par le Conseil supérieur du notariat et garantissant l'intégrité et la confidentialité du contenu de l'acte. L'acte doit être signé par le notaire au moyen d'un procédé de signature électronique sécurisée... L'acte établi sur support électronique doit être conservé dans des conditions de nature à en préserver l'intégrité et la lisibilité ».

– Un décret du 4 janvier 2006 fixe à 20 ans la durée de conservation des **données de santé** stockées chez des hébergeurs.

– Enfin, l'instruction fiscale du 11 janvier 2007 précise les conditions dans lesquelles les entreprises qui créent et conservent, sous forme électronique, des factures qu'elles transmettent à leurs clients sur support papier peuvent être dispensées de l'obligation de conserver sous forme papier le double des factures ainsi transmises. Ce texte présente un intérêt tout particulier car il distingue de manière très pertinente l'**intégrité technique** (le fichier numérique ne doit pas être modifié d'un *iota*) et l'**intégrité de l'information** en permettant que l'exemplaire conservé par l'émetteur de la facture se limite aux seules informations de contenu se rapportant aux mentions obligatoires devant figurer sur les factures et abandonne les logo, bandeau publicitaire ou fond de page qui figurent sur l'original expédié au client mais qui alourdissent le procédé et le coût de la conservation.

Il est convenu de distinguer les **documents techniques et scientifiques** des documents de nature juridique, bien que la frontière entre les deux ensembles soit loin d'être étanche. En effet, un dossier de conception d'une automobile, ou un rapport d'exploration du sous-sol, ou encore un dossier médical est un document qui sert d'abord aux équipes métiers mais qui peut être aussi produit comme pièce justificative lors d'un contentieux. Les qualificatifs de techniques ou scientifiques et juridiques ne sont pas exclusifs et il ne serait pas logique de gérer les uns et les autres de manière totalement déconnectée.

Les exigences d'archivage pour les **documents métiers** (juridiques et techniques) recoupent celles des documents des fonctions supports (comptabilité, dossiers du personnel, gestion des locaux...), notamment en terme de sécurité et de droit d'accès

(niveau de sécurité et de confidentialité variable selon les contenus). Les durées de conservation sont parfois liées à des durées réglementaires, parfois définies par les équipes, avec des durées de quinze ou vingt ans qui ne sont pas des durées légales. Les deux différences essentielles entre les documents métiers et ceux des fonctions support sont, d'une part, les formats des données, qui sont souvent spécifiques à l'activité et beaucoup plus variés et, d'autre part, le besoin d'accès permanent et rapide, ce qui rapproche la démarche d'archivage des documents métiers de celle de la documentation technique des projets de gestion des connaissances (*knowledge management*).

4. Réponses techniques

La conservation des supports d'archives a pour but la conservation des informations qui y sont enregistrées dans une optique de restitution à l'identique à l'utilisateur à une date ultérieure plus ou moins éloignée. Dans l'environnement électronique, le **support physique** existe toujours (disque, bande...) mais la conservation du support d'archives est passée d'un problème de stockage dans des conditions physiques appropriées à une problématique de maintenance d'une chaîne matérielle et logicielle potentiellement complexe.

Sont donc concernés : les systèmes de codage et d'organisation de l'information (contenu propre, métadonnées), les supports de stockage et les outils de gestion et de restitution des données archivées (sécurité, pilotage des migrations, accès). Différents scénarios sont possibles qui doivent être arbitrés en fonction du rapport qualité/coût, le terme qualité englobant ici la satisfaction des exigences légales et les besoins de disponibilité des données dans l'entreprise.

En ce qui concerne les **formats de codage**, la question se pose autant en termes de versions successives des produits qu'en termes d'éditeur et de propriété des formats. Le caractère propriétaire des logiciels de codage et les risques de disparition des éditeurs font planer une ombre certaine sur la pérennité des archives dépendantes d'un logiciel non maîtrisé.

Le recours à des formats de codage, d'échange ou de présentation des données qui sont soit des normes validées par les instances internationales (ASCII, JPEG, HTML, XML, etc.), soit des standards de fait en raison de leur ouverture (PDF, DXF) est donc un minimum.

Les deux formats les plus souvent cités pour l'archivage sont XML et PDF.

■ Le format **PDF** (*Portable Document Format*) est initialement un format propriétaire mais qui a su s'imposer, la publication des sources ayant gagné la confiance des utilisateurs. Une déclinaison du format pour l'archivage a été lancée il y a quelques années et le format **PDF/A** (A pour Archive) a été validé en 2005 par l'ISO avec la norme ISO 19005-1. Le PDF/A est basé sur la version PDF 1.4 qui, entre autres, exclut les objets dynamiques, interdit le recours au chiffrement des objets archivés et impose l'utilisation de métadonnées. Il existe deux niveaux de formats PDF/A : PDF/A – 1a, complet, et PDF/A – 1b restreint et plus léger. ISO 19005-1 précise également la façon de développer un outil de visualisation de fichiers conformes au format PDF/A. Les spécifications détaillées représentent près de 1 000 pages... Le format PDF/A est aujourd'hui reconnu et adopté à l'échelle internationale. En France, il figure en bonne place pour l'échange et la conservation de documents dans le Référentiel général d'interopérabilité RGI de la DGME (Direction générale de la modernisation de l'état).

■ **XML** (*eXtensible Markup Language* ou langage de balisage extensible), diffusé par le *World Wide Web Consortium* (W3C) est le format privilégié, là encore, aussi bien pour l'échange que pour la conservation des documents à contenus balisés. XML fête en

2008 ses 10 ans, décennie qui a vu le développement de nombreuses DTD (*Définitions de Types de Documents*) et de nombreux schémas XML.

L'archivage et la conservation à long terme des fichiers en 3D, notamment dans l'industrie et dans les sciences géographiques et géophysiques, est encore au stade expérimental.

Le choix du ou des **supports successifs d'archivage** repose sur plusieurs arguments :

- en premier lieu, la durée de conservation légale ou souhaitée de l'information archivée car on ne procède pas de la même façon pour les documents éliminables au bout de 5 ou 10 ans et pour les documents à conservation illimitée ;

- en second lieu, les droits d'accès, la fréquence de consultation et l'urgence de l'accès (certaines données archivées peuvent supporter un temps d'accès de quelques heures, selon le contexte de la consultation).

Les volumes de données entrent également en ligne de compte.

Enfin, le choix est aussi contraint ou favorisé par l'offre du marché qui continue à se développer et à se préciser.

En matière de supports d'archivage, il existe aujourd'hui deux types de solutions plus ou moins concurrentes :

1. les supports optiques WORM (DON, CD, DVD...) ou **WORM physique** ; c'est la solution initialement préconisée par la norme NF Z 42-013 mais plus la seule aujourd'hui (cf. § 7.3) ; le disque optique non réinscriptible permet la conservation sans modification possible du document ; la durabilité du support est bonne (plusieurs décennies) mais doit tenir compte de la durabilité plus faible des formats et des lecteurs (environ 5 ans) ;

2. les baies de stockage à base de disques magnétiques non réinscriptibles ou **WORM logique** ; cette solution, plus récente, présente notamment des avantages de rapidité d'accès aux données et la possibilité de supprimer des fichiers (les données archivées doivent être détruites à l'issue de leur durée réglementaire de conservation ou lorsqu'elles sont périmées) ; la sécurité est garantie par une duplication systématique des données ; le contrôle d'intégrité des fichiers est garanti par un dispositif de cryptage.

Le concept de « **coffre-fort électronique** » constitue également une réponse à la problématique du stockage électronique sécurisé, d'où l'offre croissante des tiers-archivistes dans ce sens.

La stratégie de conservation dépend aussi de la fréquence de consultation, selon les bons principes de l'**ILM** (*Information Lifecycle Management*). Pour les fichiers clos, peu volumineux mais fréquemment consultés, l'**archivage en ligne**, c'est-à-dire la pérennisation des fichiers archivés sur les serveurs actifs, peut se révéler intéressant, les archives bénéficiant des mêmes opérations de sauvegarde et des mêmes opérations de mise à jour des systèmes.

Pour le **très long terme**, on est aujourd'hui encore dans l'incertitude et une des solutions est d'archiver les données « à plat », en format ASCII, accompagnées de leur documentation, afin de pouvoir « remonter » les bases de données ultérieurement lors des besoins de consultation. Mais ce système peut paraître bien lourd. C'est toutefois la solution retenue prudemment par les Archives nationales (Centre des archives contemporaines de Fontainebleau) pour l'archivage des grands fichiers statistiques nationaux soumis à un délai de communicabilité lui-même très long : recensements INSEE, bases de données ministérielles visées par le secret statistique pendant 100 ans.

Les **outils logiciels de gestion** des données archivées font partie du système d'archivage dans la mesure où ils sont indispensables à l'identification des fichiers et à la **restitution**, écran ou papier, de l'image des documents, restitution qui peut passer par une recombinaison d'éléments stockés séparément. Dans ce cas, la qualité du support physique de stockage et la maîtrise du décodage des données sont insuffisantes à garantir une représentation fidèle du document, si le dispositif de restitution n'est pas adapté.

La question des **logiciels libres**, *versus* les logiciels propriétaires, reste ouverte car on n'a pas aujourd'hui le recul nécessaire pour

juger de la performance et du coût des diverses solutions sur le long terme. Par ailleurs, on peut se demander comment sera assumée la responsabilité des évolutions apportées progressivement à ces logiciels. Quels que soient les systèmes, l'archivage électronique exige une vision globale des différentes composantes de la chaîne matérielle et logicielle, une démarche de maintenance et aussi une réactivité à l'évolution de l'environnement technologique.

La pratique de l'archivage traditionnel nous a habitués à rechercher un support « pour la vie » avec l'idée que, idéalement, il ne doit pas y avoir de changement matériel entre le moment de la fixation de l'information, moment de la création de l'archive, et celui de son utilisation ultime, cette permanence étant le gage même du caractère authentique de l'information archivée, de sa valeur d'« original ». Toutefois, l'environnement électronique remet en cause cette vision des choses car, avec la dissociation du support et du contenu, la notion de **document original** est transposée dans la possibilité de garantir l'intégrité et la fidélité de l'information. Le groupe de travail international InterPARES (*International Research on Permanent Authentic Records in electronic Systems* – Université de Colombie-Britannique à Vancouver) a établi le principe selon lequel « Conserver un document électronique est à proprement parler impossible ; seule la capacité de le reproduire peut être préservée ». L'évolution est d'importance car, dès lors, ce n'est plus tant le support parfait que l'on recherchera mais l'intégrité du contenu d'information, garanti par la traçabilité des objets et des systèmes. Comme le dit remarquablement Me Isabelle Renard : « L'original est mort, vive la trace numérique ! ».

La pérennité de l'information numérique est indissociable aujourd'hui de la **migration sécurisée** des données, d'un support à l'autre, d'un format à l'autre, autant de fois que la durée totale de conservation l'exigera. La **migration d'archives** consiste à « transférer des documents d'un système à un autre en préservant leurs caractéristiques, c'est-à-dire « leur authenticité, leur intégrité, leur fiabilité et leur exploitabilité » (définition norme ISO 15489, cf. § 7.1). Sur les différents types de migrations, voir plus loin le modèle OAIS (§ 7.4).

À partir du moment où le transfert physique des données d'un système ou d'un support d'origine à un autre est admis, la qualité de la conservation dépend du contrôle des opérations de transferts : pour que l'objet restitué soit recevable, il faut pouvoir prouver qu'aucune altération de structure, de contenu et donc de sens n'a été effectuée. Il s'agit donc de tracer toutes les étapes, non seulement d'enregistrement initial mais aussi de migration, et également de consultation. C'est le rôle de l'historique des événements ou journal des événements que prévoit aussi bien le modèle MoReq2 (cf. § 7.2) que la norme NF Z 42-013 (cf. § 7.3).

5. Aspects organisationnels et acteurs de l'archivage électronique

Face à cet environnement technique mouvant, les aspects organisationnels de l'archivage ne doivent pas être négligés. La question de la capacité de stockage et de la durabilité des supports n'est pas la toute première à poser.

Il convient de définir avant toute chose ce que l'on veut archiver et pourquoi, c'est-à-dire quels groupes de données et pour quelle finalité. Cette démarche suppose de connaître le périmètre de l'information produite au sein de l'entreprise, par qui et pour quelle raison elle est émise et, par conséquent, sa valeur de preuve et de mémoire. Une partie seulement de la production (interne et externe) requiert d'être archivée ; parmi cet ensemble, une part requiert un archivage court terme (2-5 ans), une autre part un archivage moyen terme (une dizaine d'années) et une troisième partie un archivage long terme ou très long terme (30 ans ou indéfini).

On peut ainsi cartographier, sous forme de cercles concentriques, les différents niveaux d'archivage à organiser et adapter les solutions techniques et les classes de service aux besoins de l'entreprise.

La mise en œuvre de ce processus global d'archivage présente quelques aspects critiques :

- le **recensement** des documents et objets de données produits ou reçus, tâche d'autant plus délicate que la taille de l'entreprise est importante et sujette à des réorganisations et que les applications sont nombreuses et pas toujours contrôlées par la DSI (Direction des systèmes d'information) ;
- l'**évaluation de la valeur juridique** des documents car tous les documents d'entreprise n'ont pas une durée légale de conservation ; dans un certain nombre de cas, la durée de conservation est établie après analyse de la valeur d'information du document (qui peut être très faible dans les années qui suivent la production du document et « se réveiller » plus tard) ; cette décision est souvent prise par analogie avec d'autres documents au sein d'un même « dossier » (pièces conservées en tant que pièces justificatives d'un processus de décision) ;
- la **redondance** de l'information entre plusieurs documents et la multiplication des exemplaires : exemplaires de référence et exemplaires de travail, versions successives, mises à jour par écrasement ou par accumulation des données ;
- le **repérage de séries** documentaires logiques dans la mesure où la forme même des documents évolue au gré des habitudes de travail des équipes et des outils de production ou de diffusion de l'information ; fiche cartonnée, formulaire bureautique, courrier électronique peuvent se succéder dans le temps pour tracer la même opération technique ou administrative, masquant ainsi l'homogénéité de la série documentaire qui continue, quel que soit le support, à relever des mêmes règles d'archivage ;
- le **sort final** des données archivées ; à la fin de la durée de conservation, la majeure partie des documents ou données est détruite tandis que les plus pertinents vont alimenter les archives historiques. Il est important pour l'économie de l'archivage d'éviter les coûts de stockage inutiles et la présence d'informations non fiables dans les résultats de recherche.

La mise en œuvre de l'archivage électronique nécessite l'intervention de **plusieurs acteurs**.

Tout d'abord, l'archivage a besoin d'un **coordinateur**, d'un chef d'orchestre, pour piloter le projet, établir les documents de référence, rapporter à la hiérarchie les questions de risques et de coûts qui méritent arbitrage, sensibiliser les utilisateurs, accompagner le déploiement de la solution, auditer le fonctionnement du système et les résultats obtenus.

L'élaboration des règles de gestion des documents archivés requiert l'intervention, d'une part des **métiers** et des **qualiticiens**, d'autre part des **juristes**. Les règles de sélection des informations à archiver, de conservation, d'accès, de destruction, doivent tenir compte à la fois de l'environnement réglementaire, de l'analyse de risque dans le contexte spécifique de chaque entreprise et des besoins d'information des métiers, étant entendu que tous ces paramètres peuvent évoluer avec le temps.

Troisième groupe d'acteurs : les **techniciens** qui, tout au long du cycle de vie des données, contrôlent l'adéquation et la performance des outils de stockage et de gestion, et réalisent les opérations de capture, de sécurisation, de migration, de restitution et de destruction.

On pourrait ajouter à la liste les utilisateurs du système qui ont un rôle à jouer et une part de responsabilité dans l'identification des informations à archiver et dans le respect des règles de consultation.

Compte tenu du caractère transverse de l'archivage dans l'entreprise, l'engagement de la **direction générale**, seule légitime à empêcher le cloisonnement des processus, est indispensable à la réussite du projet.

À noter que les acteurs peuvent être internes ou externes, selon les tâches, et peuvent intervenir de manière permanente ou ponctuelle.

6. Coûts de l'archivage électronique

Certes, l'archivage a un coût pour l'entreprise et on entend souvent dire que l'archivage électronique est onéreux. Mais le « non-archivage », avec ses conséquences juridiques et organisationnelles a également un coût.

Le coût de l'archivage recouvre plusieurs composantes et ne peut être apprécié comme un tout. Il convient de distinguer les coûts techniques, les coûts humains et le risque financier encouru par un défaut d'archivage.

- Les **coûts techniques** de l'archivage électronique comprennent :
 - l'acquisition et l'enregistrement des objets de données dans le système d'archivage ;
 - les dispositifs de sécurité (signature, horodatage...) ;
 - les supports de stockage, supports initiaux et supports de remplacement lors des migrations ; ce poste de coût, sans être négligeable, bénéficie théoriquement de la baisse générale des coûts des supports de stockage, couplée à la croissance exponentielle des capacités de stockage ; toutefois, cette baisse peut être « compensée » par l'accroissement encore plus phénoménal du volume des données archivées... ;
 - les opérations de migrations technologiques, qui sont d'autant plus coûteuses si, faute d'anticipation dans l'identification des fichiers, elles ne peuvent être automatisées ; ces opérations incluent les tests de contrôle de la qualité des migrations ;
 - les outils de gestion des opérations d'archivage et de maintenance (et la maintenance même de ces outils...).

Il faut y ajouter les aspects sécurité qui sont globalement les mêmes que ceux du système d'information dans son ensemble, le système d'archivage se présentant comme un sous-ensemble du système d'information.

L'audit du système d'archivage électronique en vue d'une certification a également un coût.

- Les **coûts humains** liés à l'archivage électronique sont :
 - l'identification et/ou l'élaboration de référentiels et de procédures pour l'identification des documents et objets de données à archiver et la définition de leur durée de conservation, ainsi que la maintenance de ces documents de référence face à l'évolution juridique et organisationnelle de l'entreprise ;
 - la définition des métadonnées pertinentes pour l'archivage, en fonction du contexte de création et de l'usage prévisible des données ;
 - l'indexation manuelle, lorsque l'indexation ne peut être automatisée : résumés, choix de mots-clés, signalement des documents liés ;
 - les recherches complexes de documents ou de données (documents anciens, utilisation secondaire des documents, évolution du vocabulaire, etc.) ;
 - la mise en œuvre du sort final des documents et des données : sélection, validation, contrôle de la destruction (destruction des originaux et des copies) ; transfert à une institution archivistique ;
 - le contrôle des procédures d'acquisition d'enregistrement, d'identification, de recherche, de destruction ;
 - le pilotage du projet et l'audit du système.

Le programme paraît ambitieux et il l'est, car les archives d'une entreprise forment un ensemble de données considérables. Surtout, cet ensemble n'est pas monolithique, toutes les informations archivables ne sont pas d'égale importance, au plan légal, stratégique ou opérationnel. La politique d'archivage doit prendre en compte la valeur (ou plutôt les valeurs) des documents et données à archiver, les hiérarchiser et établir des priorités globales de traitement. Il s'agit d'arbitrer entre le coût positif de l'archivage

et le coût négatif de l'indisponibilité des données, par suite d'un non-archivage ou d'un archivage insuffisamment sécurisé.

■ Cette approche fait ressortir la notion d'**archives vitales**, définies comme les archives absolument nécessaires à une entreprise pour redémarrer son activité au lendemain d'un sinistre. Les archives vitales ne se confondent pas avec les **archives légales** : un fichier de clientèle n'est pas un document légal mais sa valeur stratégique est évidente. Les critères importants ici sont l'originalité de l'information, sa valeur ajoutée et la confidentialité.

À côté des archives vitales, on peut distinguer les **archives utiles**, qu'il serait possible, le cas échéant de reconstituer, moyennant un certain coût (un acte notarié dont le notaire pourrait, le cas échéant, fournir une nouvelle expédition, un document déposé auprès de l'Administration dont on peut obtenir une copie, etc.), et les **archives de confort**, dont la disparition serait gênante pour les collaborateurs mais finalement peu dommageable (documentation de travail, archives à faible valeur ajoutée).

Vis-à-vis de ce critère de valeur stratégique, l'archivage fonctionne un peu comme une assurance : l'investissement matériel et intellectuel de l'archivage est la garantie de la restitution de l'information au cas où l'on en a besoin. L'importance de cet investissement est donc indexée sur l'importance des documents pour la vie de l'entreprise.

Les **coûts négatifs** incluent les sanctions consécutives à un audit non satisfaisant et les conséquences financières de la perte d'un procès due à l'impossibilité de produire une pièce devant une juridiction, à quoi s'ajoute le déficit d'image de marque consécutif à l'échec.

Enfin, le coût global d'archivage doit intégrer, en positif, les gains de temps réalisés dans la recherche d'information lorsque l'archivage est pertinent, par opposition au temps « perdu » à rechercher des documents mal archivés parce que mal identifiés, mal référencés ou enregistrés sur un support déficient.

■ Un **plan global d'archivage** tient compte de l'hétérogénéité des archives dans l'entreprise : certains documents requièrent un archivage électronique hautement sécurisé ; dans d'autres cas, une duplication des données est aussi efficace et moins onéreuse qu'une conservation unique ultrasécurisée ; pour les archives de très longue conservation avec une consultation très faible (comme les bulletins de salaires conservés 50 ou 60 ans pour la justification des droits à la retraite), le microfilm peut se révéler moins onéreux que l'électronique, surtout si on peut coupler les deux technologies pour la consultation ; enfin, même pour des documents électroniques natifs, il ne faut pas exclure l'archivage papier qui, pour des documents peu volumineux à consultation linéaire (du premier au dernier), peut s'avérer le plus intéressant ou le moins risqué.

Il n'y a pas de solutions d'archivage électronique *a priori* : il y a des solutions technologiques et des démarches organisationnelles à mettre en œuvre en fonction de besoins spécifiques.

Les problématiques d'archivage comportent souvent la gestion d'un **stock non maîtrisé**, cas de figure classique avec les archives papier entassées dans les caves, mais de plus en plus fréquemment également avec les piles de disques amovibles ou de bandes de sauvegarde que l'on ne sait plus relire avec les outils courants. La récupération et le traitement de ce stock, de ce passif, s'ils sont jugés nécessaires, doivent faire l'objet d'un sous-projet à part, avec son coût spécifique.

Le traitement du stock s'avère souvent lourd et long (on s'approche de l'archéologique numérique...) ; sauf événement critique, il est rarement vital de lui donner la priorité. Du reste, commencer par là conduit en général à ralentir le vrai projet d'archivage vivant et à laisser se constituer un nouveau passif...

7. Quatre principales normes pour l'archivage électronique

La dernière décennie a procuré aux professionnels et aux praticiens de l'archivage électronique une série de textes de référence, comblant ainsi un vide qui commençait à se faire sentir.

Nous présentons ici quatre textes :

- la norme internationale ISO 15489 sur le *Records management* ;
- le modèle européen MoReq2 pour la maîtrise de l'archivage électronique au sens du *records management* ;
- la norme française 42-013 sur la conservation et l'intégrité des documents numériques ;
- le modèle OAIS sur l'archivage des données spatiales.

7.1 Norme ISO 15489 (2001)

La norme ISO 15489 s'intitule (en français également...) « norme sur le *records management* ». Mais c'est bien d'archivage dont il est question, d'archivage en général, donc aussi d'archivage électronique.

Il s'agit de prendre en charge les documents et données résultant de l'activité de l'entreprise, depuis leur validation jusqu'à l'extinction de leur valeur probante et/ou de leur utilité interne. Il s'agit d'organiser l'archivage dès la validation de l'information, dans un but de responsabilité et d'efficacité au service de l'entreprise. Dans le langage courant, l'expression « archivage légal » est utilisée dans ce sens. Cependant, *stricto sensu*, l'archivage n'est pas « légal » ; certes la loi impose la conservation de certains documents et de certaines données mais ne se prononce pas sur les moyens. Il est donc plus exact de parler d'« archivage à des fins de preuve ».

Cette norme internationale sur le *records management* trouve son origine dans la norme nationale australienne sur le même thème, publiée en 1995 et présentée à l'ISO en raison de son succès.

La définition officielle du *records management* (dans la norme) est : « Champ de l'organisation et de la gestion en charge d'un contrôle efficace et systématique de la création, de la réception, de la conservation, de l'utilisation et du sort final des documents, y compris des méthodes de fixation et de préservation de la preuve et de l'information liées à la forme des documents ».

En amont, le *records management* ne gère pas la documentation de travail (non validée) ; en aval, il ne gère pas les archives historiques.

Pratiqué dans les pays anglo-saxons depuis plusieurs décennies, le *records management* tend aujourd'hui à s'imposer universellement. En effet, avec l'électronique, on ne peut plus attendre que les documents ou les données ne soient plus d'usage courant pour réfléchir à leur archivage. C'est bien au moment de la création des données que l'archivage doit être pensé. Ainsi, pour des bases de données, les contraintes dues à l'archivage (identification, métadonnées, migrations) ont intérêt à être prises en compte dès la définition du modèle de données, voire de l'architecture technique du système, faute de quoi, l'archivage exigera de modifier ou de reconstituer des données, avec les risques et les coûts associés.

ISO 15489 n'est pas une norme technique. Sa cible est très large : les décideurs, les professionnels de l'information, les responsables des systèmes et, plus généralement, toutes les personnes physiques ayant la responsabilité de produire ou de conserver des informations de preuve et de mémoire. Les responsabilités et les compétences en matière de *records management* sont définies et attribuées à tous les niveaux de collaborateurs.

Les objectifs de la norme sont assez ambitieux :

- organiser la traçabilité des documents produits ou reçus par tout organisme ou entreprise dans l'exercice de ses activités ;
- encadrer la responsabilité des entreprises ou organismes vis-à-vis de la loi, de ses collaborateurs, de ses partenaires, de la collectivité ;
- s'assurer que les systèmes d'information produisent des documents fiables en accord avec les normes qualité **ISO 9001 et ISO 14000** ;
- gérer l'information au meilleur coût.

Au terme de la norme, le *records management* a pour tâche :

- de définir quels documents doivent être créés pour chaque type d'activité et quelles informations doivent y figurer ;
- de dire sous quelle forme et quelle structure les documents doivent être produits et archivés, et quelles technologies doivent être utilisées ;
- de définir les métadonnées nécessaires ;
- de déterminer les besoins de consultation et les durées de conservation en se conformant aux exigences légales et réglementaires ;
- d'évaluer les risques attachés à la non-disponibilité de documents probants ;
- de rechercher l'amélioration de la rentabilité, l'efficacité et la qualité des méthodes, des décisions et des opérations de création, d'organisation ou de gestion des documents.

Le système d'archivage (*records management system*) doit notamment présenter des caractéristiques de fiabilité (tous les documents de l'entreprise sont intégrés systématiquement), d'intégrité (contrôle d'accès, de la validité, de la destruction et de la sécurité) et de conformité avec l'environnement réglementaire et les exigences de l'entreprise.

La norme préconise une démarche de mise en œuvre d'un projet de *records management* en huit étapes, selon la méthode élaborée par les *records managers* australiens et appelée DIRKS.

1. Enquête préliminaire : le *records management* tient compte de l'environnement réglementaire défini par la législation dans les domaines administratif et judiciaire et de la réglementation spécifique du secteur d'activité concerné, ainsi que des lois et règlements relatifs aux archives, à l'accès à l'information, au respect de la vie privée, à la notion de preuve, au commerce électronique et à la protection des données ; s'y ajoutent les normes en vigueur et les procédures internes.

2. Analyse des activités : identifier et documenter toutes les fonctions, activités et opérations et les hiérarchiser dans un plan de classement des activités ; lister les documents qui procèdent de chaque grande fonction, de chaque activité au sein d'une fonction, de chaque opération au sein d'une activité.

3. Identification des exigences archivistiques : analyse de l'environnement réglementaire de l'organisme ou de l'entreprise en matière de conservation pour évaluer le risque qu'il y aurait à ne pas produire ou à ne pas conserver les documents.

4. Évaluation des systèmes existants (systèmes d'archivage et autres systèmes de gestion de l'information liés).

5. Identification de la stratégie pour la satisfaction des exigences archivistiques : adopter principes, normes, procédures et pratiques requises vis-à-vis de ces exigences.

6. Conception d'un système d'archivage qui intègre la stratégie et les procédures de la norme ; le cas échéant, redéfinir des méthodes de travail et les systèmes opérationnels et de communication.

7. Mise en œuvre d'un système d'archivage selon une démarche de gestion de projet.

8. Contrôle a posteriori : audit du projet.

Les trois premières étapes sont spécifiques au *records management*, les autres sont plus ou moins communes à tout projet de gestion de l'information. Il faut souligner que cette méthode vise à

la mise en place d'un système d'archivage électronique pour le flux et n'intègre pas la gestion du passif qui doit être géré à part (voir ci-avant).

Concernant les métadonnées du *records management*, il convient de mentionner la récente norme ISO 23081 encore peu diffusée.

7.2 Modèle européen MoReq2 (2008)

MoReq2, publié par la Commission européenne en février 2008 constitue une mise à jour d'un premier document de référence publié en 2001 sous le titre MoReq pour *Model Requirements for the Management of Electronic Records* «Modèle d'exigences pour l'organisation de l'archivage électronique».

MoReq2 peut être consulté en anglais à l'adresse : <http://www.moreq2.eu>. La traduction française prise en charge par la Direction des archives de France sera disponible à partir de décembre 2008.

MoReq décrit les **fonctionnalités d'un système d'archivage électronique (SAE)** selon les principes du *records management*, c'est-à-dire ce que doit savoir-faire un outil d'archivage pour assurer la conservation, la traçabilité et l'intégrité des documents numériques qui présentent une valeur de preuve ou de mémoire pour une entreprise.

L'exigence d'archivage électronique peut se résumer à une simple phrase : identifier les informations dont on aura besoin demain, les capturer dans un système et les y conserver de manière fiable et sécurisée pendant toute la durée requise en permettant et en contrôlant l'accès aux informations archivées.

La mise en œuvre de cette exigence globale peut en revanche conduire à la prise en compte de très nombreux éléments, d'autant plus nombreux si l'entreprise est importante, si l'environnement réglementaire est contraignant ou si les documents archivés sont hétérogènes.

MoReq s'appuie sur quelques concepts fondamentaux parfaitement en phase avec les principes du *records management*, notamment les notions de :

- « **record** », document validé, sélectionné, archivé et sécurisé pour sa valeur de preuve et de mémoire, notion opposée à celle du document non figé ou non validé, utile à l'utilisateur mais sans valeur particulière pour l'entreprise ;
- **plan de classement** qui ordonne les dossiers de manière structurée ; les bonnes pratiques veulent que cette structure reflète les activités de l'entreprise ou de l'organisme ;
- **capture** qui est le processus d'entrée des documents dans le système, avec quatre étapes : l'enregistrement (attribution d'un identifiant unique), le rattachement au plan de classement, l'ajout de métadonnées et l'affectation d'une zone de stockage dans le système d'archivage ;
- **métadonnées** d'archivage qui sont des informations structurées ou semi-structurées qui permettent la création, la gestion et l'utilisation des documents archivés au cours du temps ;
- **historique des événements** qui sont des informations sur les opérations ou autres événements qui concernent ou modifient les entités gérées (les données et les métadonnées), suffisamment détaillées pour permettre une reconstitution des opérations passées.

Le système d'archivage électronique vise potentiellement la totalité des documents de l'entreprise. Aujourd'hui, les documents électroniques peuvent être de formats variés, produits par des auteurs différents et peuvent être reçus sous forme de pièces isolées ou sous forme de dossiers. Ils peuvent arriver *via* divers canaux de communication, c'est-à-dire le réseau local (LAN), un réseau étendu, la messagerie, le fax, le courrier postal (avec scan), à des fréquences et avec des volumes variables. La capture des documents requiert un système souple doté de contrôles pertinents, de sorte que les diverses exigences soient respectées.

Le **processus de capture** du SAE doit posséder les fonctionnalités suivantes :

- enregistrer et gérer tous types de documents quelles que soient la méthode de codage et les caractéristiques techniques de production ;
- s'assurer que les documents archivés sont reliés au plan de classement et rattachés à un ou plusieurs dossiers ;
- prendre en compte l'application qui a produit le document ;
- contrôler et valider la saisie des métadonnées dans le SAE.

Lorsqu'un document possède plusieurs versions, le SAE doit permettre aux utilisateurs de choisir au moins une des actions suivantes :

- archiver toutes les versions du document comme une seule entité ;
- archiver une seule version du document ;
- archiver chaque version séparément.

Le SAE doit comporter la possibilité d'extraire automatiquement des métadonnées des documents au moment de leur capture. Le système d'archivage doit permettre de définir différents jeux de métadonnées pour différents types de documents électroniques.

Les évolutions organisationnelles doivent être prises en compte dans les fonctions de maintenance du système : changement du nombre d'utilisateurs, accroissement de la demande en capacité de stockage, restauration en cas de défaillance du système, surveillance des erreurs système. Par ailleurs, le SAE doit fournir à l'administrateur des **outils de reporting** souples pour évaluer par exemple :

- le nombre de dossiers, sous-dossiers et documents archivés pendant telle période ou relativement à telle activité ;
- les pratiques et fréquences de consultation pour les dossiers, sous-dossiers et documents.

Le succès d'un système ne repose que sur ces fonctionnalités. Certaines exigences non fonctionnelles sont indispensables à la réussite du système : facilité d'utilisation, performance et évolutivité, disponibilité du système, normes techniques, conformité à l'environnement législatif et réglementaire, externalisation et recours à des tiers, conservation à long terme et obsolescence technologique.

MoReq2 met à jour et complète MoReq. Le nouveau texte compte 330 pages et près de 800 exigences pour un système d'archivage électronique complet... Cependant, MoReq2, à travers l'exposé des concepts fondamentaux de l'archivage, peut se révéler très utile dès le démarrage d'un projet.

En matière de fonctionnalités du système, les points forts de MoReq2 sont les suivants.

1. Le système d'archivage électronique doit gérer les documents archivés dans un cadre structuré, au travers d'un **plan de classement hiérarchique** de toutes les données de l'entreprise. Ce plan de classement identifie les dossiers à archiver et les organise en séries, ou les subdivise en sous-dossiers ou en volumes qui contiennent eux-mêmes les documents ; chaque entité documentaire reçoit, lors de sa capture dans le système d'archivage, un identifiant unique et est rattachée à une **règle de conservation/destruction** qui précise son devenir à échéance de la durée de conservation ; le plan de classement et les règles de conservation/destruction des documents archivés doivent être élaborés par l'entreprise préalablement à la mise en place du système d'archivage ; un exemple de plan de classement est donné figure 1.

2. Le SAE doit pouvoir prendre en charge toutes les informations archivables de l'entreprise, quelle que soit leur forme et quel que soit leur support ou format ; MoReq2 insiste notamment sur les conditions de capture dans le système de **documents composites** : une page web constituée d'un document HTML et des images JPEG, un document bureautique composé d'un texte avec un hyperlien vers un tableur dans un fichier, un courriel avec ses pièces jointes, etc.

3. La **sécurité** est particulièrement importante dans le SAE et présente plusieurs aspects :

- les indices de sécurité attribués aux documents et les habilitations attribuées aux utilisateurs ; les contrôles d'accès par un ou plusieurs administrateurs ;

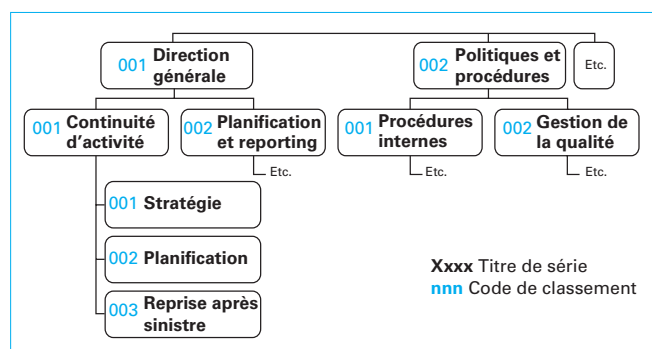


Figure 1 – Plan de classement MoReq2

- l'historique des événements (enregistrement complet de toutes les opérations sur chaque document) ;
- les sauvegardes régulières avec une distinction entre les sauvegardes intégrales et les sauvegardes vitales (cf. le plan de continuité d'activité de l'entreprise) ;
- l'automatisation d'un maximum de tâches ;
- la traçabilité des processus d'export, de destruction ou de gel des destructions, avec la notion de métadonnées-témoins.

4. Le SAE doit **s'intégrer au mieux au système d'information de l'entreprise** ; MoReq2 décrit les modalités de l'interfaçage entre le SAE et les autres outils de gestion de l'information :

- gestion des archives physiques ;
- destruction des archives physiques ;
- GED et travail collaboratif ;
- workflow ;
- gestion de dossiers ;
- intégration avec les systèmes de gestion de contenu ;
- signature électronique ;
- chiffrement ;
- gestion des droits numériques ;
- systèmes distribués ;
- travail hors ligne et à distance ;
- intégration du fax ;
- gestion de la sécurité.

Concernant la GED, MoReq et MoReq2 insistent sur la différence entre la gestion des documents non figés, non validés, non diffusés et qui ne portent donc pas de valeur probante pour l'entreprise (en anglais le *Document management*) et le *Records management* où les documents sont figés, validés, non modifiables, diffusés et obéissent à des règles communes.

Par ailleurs, la grande nouveauté de MoReq2 en matière de normalisation de l'archivage est de compléter les spécifications par une batterie de tests logiciels qui permettent aux éditeurs et aux entreprises utilisatrices, d'évaluer la conformité à MoReq2 des logiciels du marché et des solutions internes. Ce n'est pas encore la certification des solutions d'archivage mais on s'en approche.

7.3 Norme NF Z 42-013 (2008)

La norme NF Z 42-013, promue par l'Association des professionnels du numérique (APROGED) et adoptée par l'AFNOR en 1999, révisée en 2001 et de nouveau en 2008, contient les **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**, pendant toute la durée de leur conservation et de leur utilisation.

Le postulat de base est que, pour qu'un organisme ou une entreprise, qui utilise ou qui souhaite stocker et restituer des documents électroniques, soit en mesure de montrer que ces documents

peuvent être considérés comme des représentations ou des copies fidèles des documents d'origine, il doit réaliser ou faire réaliser périodiquement des audits de ses systèmes et de ses procédures.

La norme NF Z 42-013 énonce quatre niveaux d'exigences pour l'archivage, symbolisé par quatre mots dont les initiales forment le sigle PIST (Pérennité, Intégrité, Sécurité, Traçabilité) facile à retenir :

1. la **pérennité**, avec la recommandation d'utiliser des formats ouverts, normalisés ou standardisés, facilitant les nécessaires migrations ;
2. l'**intégrité**, avec une sécurisation optimale des processus de stockage et de destruction ;
3. la **sécurité**, avec une duplication des données, un processus d'authentification, le chiffrement et déchiffrement des documents ;
4. la **traçabilité**, touchant l'horodatage, la tenue d'un dossier technique, le journal du cycle de vie des archives et le journal des événements.

Le **dossier de description technique du système** comprend notamment la liste des matériels y compris les équipements de connexion et de sécurité, un modèle conceptuel des données du système, la liste des logiciels et leurs documentations associées et les conditions physiques pour la bonne conservation des supports.

La norme évoque la possibilité de définir au moment de chaque dépôt un **profil d'archivage**, c'est-à-dire un ensemble de règles applicables à des documents partageant les mêmes critères de confidentialité, de durée de conservation, de destruction et de droits d'accès pour déposer, consulter ou détruire. Plus généralement, l'organisme ou l'entreprise doit mettre en place des procédures pour l'enregistrement, le stockage, la communication et la restitution des documents. Ces procédures doivent être décrites dans le dossier technique.

La **sécurité** est un point fort de la norme. Sont abordées les questions d'administration et d'organisation de la sécurité, de sécurité physique, de sécurité des matériels et logiciels, ainsi que les aspects de maintenance et d'évolution des systèmes et de migration des supports. Par ailleurs, un enregistrement doit être réalisé pour chaque événement lié à l'exploitation du système ou au cycle de vie des archives. Les journaux doivent permettre de démontrer la continuité de la conservation quels que soient les supports utilisés.

Le système doit assurer les opérations de sauvegarde et, le cas échéant, de restauration.

Concernant les **supports**, la norme décrit les procédures relatives à l'utilisation des supports de conservation, en distinguant les supports amovibles (WORM physique ou logique) et les supports fixes. On a donc trois grands cas de figure :

1. les systèmes basés sur des supports WORM amovibles ;
2. les systèmes basés sur des supports WORM logiques ;
3. les systèmes basés sur des supports réinscriptibles, avec différents niveaux de sécurisation.

Le chapitre suivant décrit la procédure de **capture des documents** dans le système avec l'enregistrement des fichiers à archiver et la gestion des métadonnées, avec les contrôles afférents. Lorsque la capture vise des documents papier, le dispositif de numérisation doit respecter certaines règles, notamment pour le traitement des images, avec un contrôle qualité comprenant une journalisation précise des opérations. Il en va de même pour la numérisation du son et de l'audiovisuel, avec les recommandations en matière de compression (sans pertes, avec pertes) et de migration.

Un chapitre est consacré à l'**exploitation des archives**, c'est-à-dire la consultation, la restitution et l'élimination des documents. Il est recommandé que les applications de consultation soient indépendantes des outils de création des documents conservés. À la fin de la durée de conservation indiquée dans les métadonnées, les archives doivent être supprimées et le système doit garantir l'inaccessibilité définitive des données. D'une manière générale, les méthodes utilisées doivent être décrites dans le dossier de description technique du système.

Pour l'**horodatage**, la norme envisage deux possibilités :

1. un dispositif interne produisant des contremarques avec une source de temps de référence et un système sécurisé et contrôlé régulièrement ;
2. l'intervention d'un tiers horodateur délivrant des contremarques de temps conformes aux normes ou à une politique d'horodatage contrôlable.

Plus largement, le système et toutes les procédures associées sont soumis régulièrement à un audit (interne ou externe).

Les derniers chapitres concernent les tiers archiveurs et les autres types de prestations. Le **tiers archiveur** peut assurer toute la chaîne d'archivage ou simplement une prestation de stockage. Il est tenu de garantir la confidentialité des données et métadonnées qu'il reçoit en dépôt, de fournir des attestations de dépôt pour chaque opération, de fournir un journal de cycle de vie des archives. La norme propose un contrat de service-type.

7.4 Modèle de référence OAIS (2001)

Le *Reference Model for an Open Archival Information System* (OAIS) vise l'archivage à long terme des informations existant sous une forme numérique. Il est le résultat d'un travail normatif approfondi entrepris en 1995 par le CCSDS (Consultative Committee for Space Data Systems) à la demande de l'ISO (International Standard Organization) en vue d'établir des recommandations techniques générales pour encourager et faciliter l'**échange des données spatiales**. Le CNES (Centre national d'études spatiales) a participé à l'élaboration de cette norme.

Le modèle OAIS a été porté au statut de norme internationale en 2003 sous le numéro ISO 14721. Sa traduction française est disponible en ligne sur le site du groupe PIN (cf. [Doc. H 7 276v2]).

Dans le titre du modèle, *Open* signifie que ce modèle a été développé dans le cadre d'un processus public et qu'il est disponible et utilisable par ce public, et non que l'accès aux informations archivées se fasse sans restriction. Le terme *information* désigne toute connaissance qui peut être échangée, indépendamment de la forme (physique ou numérique) que prend cette information alors que la « donnée » est définie comme une forme de représentation de l'information. Quant au système d'archivage, il est constitué par un ensemble de moyens matériels, logiciels et humains en charge de l'acquisition, de la conservation et de la distribution de l'information archivée. La norme utilise le terme « archive » (au singulier) pour désigner l'organisation (personnes et systèmes) chargée de conserver des informations et de les rendre accessibles à une communauté d'utilisateurs cible.

7.4.1 Modèles d'information et fonctionnel

L'OAIS comporte deux modèles : un modèle d'information et un modèle fonctionnel.

■ Le **modèle d'information** (figure 2) insiste sur la nécessité de bien identifier ce que l'on veut conserver. Le point de départ est l'« Objet-Données » qui contient l'information principale à pérenniser (le fichier) ; il est complété par « l'information de représentation » qui lui est associée (fichiers de structure, dictionnaire de données, etc.). L'ensemble constitue un « Objet-Information » qui permet de traduire les données en information significative.

Les informations qui précisent le contenu d'information sont diverses : information d'identification, information de contexte, information de provenance, information de description, information de structure, information d'empaquetage, information d'intégrité, information de pérennisation ou PDI (*Preservation Description Information*). Elles sont produites soit par le producteur de l'information, soit par le système d'archivage et constituent de fait des métadonnées, terme que le modèle OAIS n'utilise pas. Ces

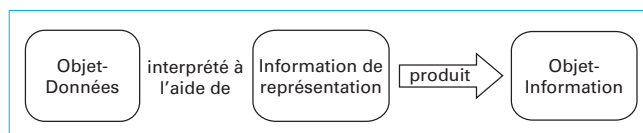


Figure 2 – Modèle d'information de l'OAIS

informations permettent de constituer des **paquets d'informations** appropriés à chaque opération ; on distingue ainsi :

- le paquet d'informations archivés AIP (*Archival Information Package*), constitué d'un contenu d'information et de l'information de pérennisation associée ;
- le paquet d'informations à verser SIP (*Submission Information Package*) livré par le producteur à l'OAIS pour l'élaboration d'un ou plusieurs paquets d'informations archivés (AIP) ;
- le paquet d'informations diffusés DIP (*Dissemination Information Package*), reçu par l'utilisateur en réponse à sa requête à l'OAIS et formé à partir d'un ou de plusieurs paquets d'informations archivés (AIP).

L'OAIS introduit également la notion de collection, définie comme un AIP constitué d'un agrégat logique d'autres AIP.

■ Le **modèle fonctionnel** identifie les trois acteurs avec lesquels le système d'archivage est en relation :

- le producteur de données, qui fournit l'information à pérenniser ;
- le management, qui fixe les règles générales applicables à l'OAIS ;
- l'utilisateur qui est en relation avec les services de l'OAIS pour chercher et se procurer des informations.

Le modèle fonctionnel distingue six entités fonctionnelles selon le schéma de la figure 3 :

- l'entité « Entrées » assure le contrôle et la préparation des paquets d'informations à verser (SIP) provenant des producteurs ;
- l'entité « Stockage » assure la maintenance et la récupération des AIP : réception des AIP, affectation d'un lieu de stockage, gestion de la hiérarchie du stockage, renouvellement des supports, contrôles, sauvegarde et plan de reprise d'activité, transmission des AIP à l'entité « Accès » en fonction des commandes ;
- l'entité « Gestion de données » s'occupe de l'enrichissement, de la conservation et de l'accès aux informations de description et aux données administratives associées ;
- l'entité « Administration » pilote l'exploitation d'ensemble du système d'archivage. Elle est en relation avec les producteurs et les communautés d'utilisateurs. Elle contrôle la conformité des opérations et fournit les moyens de fonctionnement aux entités de stockage et de gestion des données ;
- l'entité « Planification de la pérennisation » a pour rôle de surveiller l'environnement de l'OAIS dans un but d'accessibilité sur le long terme et propose les migrations nécessaires, en relation avec l'entité « Administration » ;
- l'entité « Accès » est en relation avec les utilisateurs qu'elle aide à repérer les informations recherchées : réception des demandes, contrôles d'accès, traitement des demandes, transmission des informations (paquets d'informations diffusés).

La norme présente des schémas détaillés de fonctionnement des différentes entités. Le chapitre 4 définit en outre des taxinomies pour les Objets-Informations utilisés par l'OAIS et des exemples de types d'information de pérennisation (PDI).

Les relations entre le producteur et l'OAIS sont définies dans un **protocole de versement** qui précise notamment le mode de transmission et le contenu de l'opération de versement.

Un OAIS n'a pas toujours l'obligation de conserver l'information versée exactement dans le format de versement ; une transformation peut être souhaitable, notamment par suite de l'obsolescence des formats.

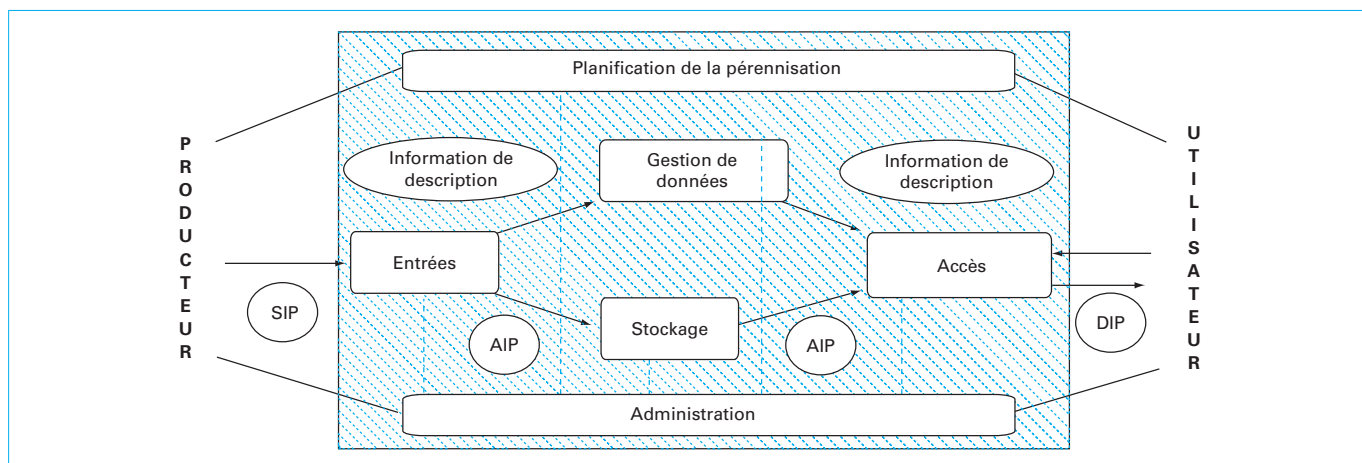


Figure 3 – Modèle fonctionnel de l’OAIS

L'utilisateur, de son côté, établit avec l’OAIS un **protocole de commande** pour recevoir des informations (transmission de supports matériels ou télétransmission). Le repérage des informations intéressantes pour l'utilisateur peut nécessiter l'aide d'un outil de recherche.

7.4.2 Pérennisation et interopérabilité

Le chapitre 5 est intitulé « Perspectives de la pérennisation » et traite des migrations. La norme identifie trois facteurs majeurs pouvant conduire à une migration numérique des paquets d'informations archivés au sein d'un OAIS : la dégradation des supports, de nouvelles exigences de service ou une meilleure rentabilité.

Le modèle OAIS distingue quatre principaux types de **migration numérique**, présentés par ordre croissant de risque de perte d'information :

1. le rafraîchissement de support consistant à remplacer le support initial par un support du même type par copie bit à bit ;
2. la duplication qui ne modifie ni l'information d'empaquetage, ni le contenu d'information, ni le PDI ;
3. le ré-empaquetage qui modifie l'information d'empaquetage ;
4. la transformation qui conserve l'intégralité des contenus d'information mais produit quelques changements dans les trains de bits.

Le chapitre 6 traite de l'interopérabilité des archives OAIS, en distinguant quatre catégories d'associations d'archives en fonction du niveau d'interaction mis en œuvre :

1. archives indépendantes, sans interaction technique ou de gestion ;
2. archives coopérantes, ayant en commun des producteurs, des standards de versement et de diffusion, mais pas d'outil de recherche ;
3. les groupements d'archives qui partagent une communauté d'utilisateurs intéressée par les fonds de plusieurs archives OAIS qui ont mis en commun un ou plusieurs outils de recherche ;
4. archives aux ressources partagées, pour améliorer leur offre ou réduire les coûts.

La norme OAIS est complétée par la norme **ISO 20652** (2006) dite norme **PAIMAS** (*Producer-Archive Interface Methodology Abstract Standard*). PAIMAS prend en compte la terminologie et les concepts du modèle de référence OAIS. Elle définit une méthodologie structurée en plusieurs étapes pour préparer et conduire dans les meilleures conditions les transferts d'objets numériques entre le producteur et l'archive.

La participation du CNES à l'élaboration du modèle OAIS a conduit à la création en 2000, au sein de l'association Aristote, d'un groupe de travail pour faire connaître le modèle et poursuivre la réflexion. Il s'agit du **groupe PIN** (pérennisation de l'information numérique). Les réunions du groupe sont l'occasion de présentations techniques ou de retours d'expérience très intéressants, comme le suggèrent ces quelques titres :

- Plate-forme d'archivage numérique au CNES : le SIPAD-NG, système générique pour l'archivage et l'accès aux données scientifiques ;
- Présentation détaillée du format WARC (*Web ARChive file format*) ;
- LOCKSS (*Lots of Copies Keep Stuff Safe*) pour la préservation de l'accès aux revues numériques ;
- Projet de l'IRCAM pour la préservation des données numériques dans le cas de la création artistique ;
- Présentation du Projet Pil@e de la Direction des archives de France (sur la base d'un « Standard d'échange de données pour l'archivage » inspiré notamment de l'OAIS) ;
- Projet SPAR – système de préservation et d'archivage réparti de la BnF, qui s'appuie également sur l'OAIS.

8. Conclusion

L'archivage électronique a connu une avancée décisive depuis quelques années avec la loi de mars 2000, sur la **signature électronique** et la production de **normes**, tant pour l'organisation et la structuration de l'information à archiver que pour la conservation numérique pérenne, et avec les développements technologiques au service de la **sécurité**, de la **traçabilité** et de **migration** de l'information.

Toutefois, on compte plus de projets que de réalisations et de retours d'expérience, notamment en matière de conservation à long terme. Le contexte réglementaire, le nombre d'acteurs et de compétences impliquées dans un projet global et la grande variété des types d'information à gérer s'ajoutent aux défis proprement technologiques d'**obsolescence des supports et des formats** ou d'**interopérabilité des systèmes**.

Les années à venir verront nécessairement d'autres progrès. Les plus attendus concernent l'harmonisation des normes, la maîtrise des formats (tant leur nombre que les formats spécifiques ou complexes comme les fichiers 3D), une meilleure visibilité des risques et des coûts, avec une évolution de l'offre logicielle pour y répondre.