

Le champ normatif de l'archivage électronique

Stéphanie Roussel

Citer ce document / Cite this document :

Roussel Stéphanie. Le champ normatif de l'archivage électronique. In: La Gazette des archives, n°228, 2012. Normalisation et gestion des documents d'activité (records management) : enjeux et nouvelles pratiques pour notre profession. pp. 59-76;

http://www.persee.fr/doc/gazar_0016-5522_2012_num_228_4_4984

Document généré le 15/03/2017

Le champ normatif de l'archivage électronique

Stéphanie ROUSSEL

Introduction

Corollaire d'une accélération permanente de la dématérialisation des processus, l'archivage électronique ne cesse de gagner du terrain et tend à devenir une problématique de mieux en mieux comprise et de plus en plus partagée¹. Souvent initiés pour pérenniser des contenus électroniques dont il s'avère indispensable de préserver la valeur de preuve et de garantir l'intégrité, les projets d'archivage électronique se multiplient et contribuent à l'amélioration permanente des connaissances et des savoir-faire sur ce domaine encore en pleine évolution mais commençant néanmoins à atteindre une certaine maturité.

Dans ce contexte, textes normatifs et standards tiennent un rôle de premier plan car ils établissent un fondement stable et relativement pérenne dans un environnement technologique qui ne cesse de se transformer. Ils permettent, d'autre part, aux projets d'archivage électronique de s'adosser à des socles partagés et reconnus.

L'objet du présent article est donc de présenter un panorama du champ normatif de l'archivage électronique en se plaçant du point de vue de la pérennisation de l'information et des caractéristiques attendues du système d'archivage électronique (SAE). Il laissera donc de côté les normes relatives au *records management* (ISO 15489, 30300 et 30301) et les spécifications fonctionnelles de type ISO 16175 (ICA-Req) ou MoReq car elles abordent la question de l'archivage électronique sous l'angle du *management* de

¹ Voir à ce sujet par exemple l'étude réalisée par Markess International au mois d'avril 2012, *Archivage de contenus électroniques à vocation probatoire – France, 2012-2014*.

l'information, ceci ne signifiant pas que dans le cadre de la mise en œuvre d'un projet, il ne faudrait pas les prendre en compte.

L'une des principales difficultés pour comprendre l'environnement normatif de l'archivage électronique est la multitude de textes qui, de près ou de loin, peuvent entrer dans ce champ. La norme centrale en la matière est sans conteste le modèle OAIS, auquel on peut associer d'autres normes généralistes telles que PAIMAS¹ ou la norme NF Z 42-013 (devenue récemment une norme internationale sous le numéro ISO 14641). Mais il faut savoir que gravitent autour d'elles de nombreuses normes traitant de sujets spécifiques tels que le stockage, les formats, les métadonnées, l'intégrité et l'authenticité, l'audit et la certification, etc.

Une norme centrale : l'ISO 14721 « systèmes de transfert des informations et données spatiales – système d'archivage d'information, modèle de référence OAIS »

Le modèle de référence pour un système ouvert d'archivage d'information – ou OAIS, c'est-à-dire *Open Archival Information System* – est un modèle conceptuel destiné à la gestion, à l'archivage et à la préservation à long terme de l'information numérique. Il définit dans les grandes lignes les fonctions, les responsabilités et l'organisation d'un système (Archive OAIS) qui voudrait conserver et préserver de l'information sur le long terme, pour en garantir l'accès à des communautés d'utilisateurs identifiées. Le long terme est défini comme suffisamment long pour être soumis à l'impact des évolutions technologiques. Il convient de souligner que le modèle fournit également une base générale permettant de couvrir la pérennisation de l'information qui n'est pas sous forme numérique telles que des archives papier, collections d'objets, maquettes d'architectes, etc.). En effet, pendant une phase transitoire, les services d'archives pourront être amenés à gérer des informations aussi bien sous forme numérique que sous forme physique.

Ce texte est une norme fonctionnelle généraliste sur laquelle s'appuient de nombreuses autres normes et standards. Son importance est considérable, dans la mesure où elle décrit le cadre conceptuel sur lequel s'appuie la majeure partie des systèmes d'archivage électronique mis en œuvre aujourd'hui.

¹ *Producer archive Interface methodology abstract standard*, ISO 20652.

Historique de l'élaboration de la norme

La mise au point de l'OAIS a été pilotée par le Comité consultatif pour les systèmes de données spatiales (CCSDS), organisme international de normalisation des agences spatiales, qui joue également le rôle de sous-comité 13 (systèmes de transfert des données et des informations spatiales) du comité technique 20 (aéronautique et espace) de l'ISO. Une première version de l'OAIS est parue en 2002 et a été reçue comme norme par l'ISO sous le numéro 14721:2003 avec le nom officiel *Space data and information transfer systems, Open archival information system, Reference model*; elle a été publiée en français en 2009 (*Systèmes de transfert des informations et données spatiales - Système ouvert d'archivage d'information - Modèle de référence*)¹. Une version révisée est en cours d'élaboration (ISO/DIS 14721).

Au départ développé dans le domaine aérospatial, le modèle OAIS a ensuite servi de base à de nombreux travaux menés dans le domaine de l'archivage des données et documents nativement numériques notamment au sein des grandes institutions scientifiques et patrimoniales. C'est notamment sur l'OAIS que s'appuient les politiques d'archivage², le Standard d'échange de données pour l'archivage (SEDA) et la plupart des plates-formes d'archivage électronique conçues ces dernières années. Le modèle OAIS fait également partie des recommandations du Référentiel général d'interopérabilité (RGI) pour la mise en place de systèmes d'archivage électronique³.

L'environnement OAIS

Bien qu'étant conçu dans un cadre spécifique, le modèle OAIS a un champ d'application bien plus large. En effet, l'OAIS est un modèle abstrait, qui ne donne aucune spécification technique, mais offre plutôt un vocabulaire et un cadre théorique pour penser différents cas de figure de l'archivage. Sa démarche peut donc être partagée par des institutions de nature et de préoccupations très diverses : bibliothèques nationales, institutions

¹ Disponible à : http://pin.association-aristote.fr/lib/exe/fetch.php/public/documents/norme_oais_version_francaise.pdf

² Voir les recommandations pour une politique d'archivage sécurisée sur le site de l'Agence nationale de la sécurité des systèmes d'information : http://www.securite-informatique.gouv.fr/gp_article188.html

³ La version 1.1 du RGI (12 juin 2009) est disponible à : <http://references.modernisation.gouv.fr/rgi-interoperabilite>

scientifiques et archivistiques ou encore industries ayant à conserver sur le long terme des volumes importants d'informations numériques.

La norme OAIS a pour objectif de définir les responsabilités et les différentes fonctions de l'ensemble des acteurs impliqués dans le processus d'archivage électronique, et en premier lieu l'Archive, entendue ici au sens de service d'archives, dont la responsabilité est de pérenniser l'information qu'elle reçoit, c'est-à-dire de la conserver et de la rendre accessible et compréhensible sur le long terme. La norme établit un certain nombre de concepts désignés par un vocabulaire choisi et adapté, dont la définition et l'explication occupent une part importante du texte. Certains termes sont à comprendre dans une acception différente de leur sens usuel, aussi est-il d'usage de signaler par une majuscule les concepts OAIS pour éviter toute confusion.

Principaux concepts

L'introduction de la norme est largement dédiée à l'établissement de la terminologie utilisée. L'Archive ou Archive OAIS (au singulier, à ne pas confondre avec les archives au sens des documents d'archives) est notamment définie comme une « organisation chargée de conserver l'information pour permettre à une communauté d'utilisateurs cible d'y accéder et de l'utiliser » : c'est l'opérateur du système d'archivage.

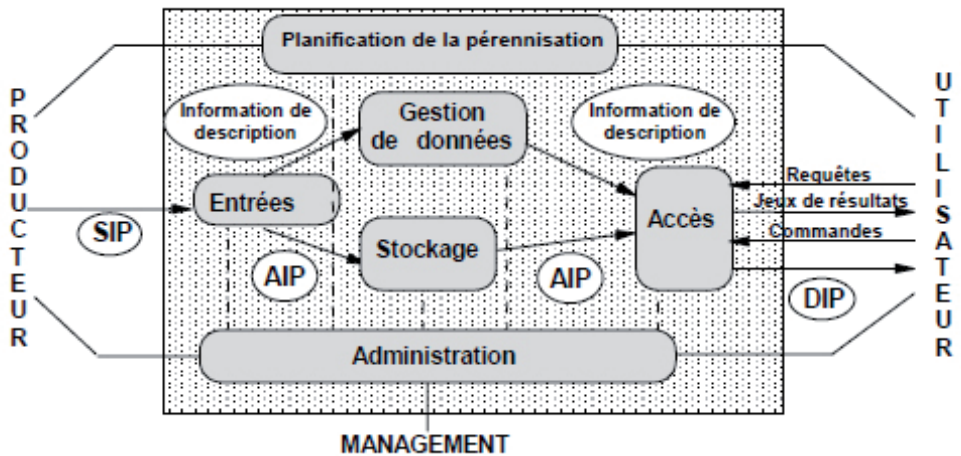
Les six responsabilités minimales d'une Archive OAIS sont les suivantes :

- négocier avec les Producteurs d'information pour s'assurer que les Contenus d'information et Informations de pérennisation (PDI) associés qu'elle va recevoir correspondent bien à sa mission et aux besoins de la Communauté d'utilisateurs cible ;
- acquérir une maîtrise suffisante de l'information fournie, au niveau requis pour pouvoir en garantir la Pérennisation ;
- extraire – ou obtenir par d'autres moyens – une Information de description suffisante pour que la Communauté d'utilisateurs cible puisse trouver le Contenu d'information qui l'intéresse ;
- déterminer quelles communautés doivent constituer la Communauté d'utilisateurs cible en mesure de comprendre l'information fournie et assurer que l'information à conserver est compréhensible par cette communauté (c'est-à-dire, sans l'assistance des experts ayant produit ces informations) ;

- appliquer une stratégie et des procédures documentées garantissant la conservation de l'information contre tout imprévu dans les limites du raisonnable, et permettant la diffusion d'une information, copie authentifiée de l'original ou permettant de remonter à l'original ;
- rendre l'information conservée disponible pour la Communauté d'utilisateurs cible.

Ces responsabilités recouvrent et complètent celles des archives papier traditionnelles en raison des risques accrus de perte de l'intelligibilité de l'information sous forme numérique.

En amont de l'Archive se situe le Producteur (*Producer*) qui fournit les informations à conserver, et en aval l'Utilisateur (*Consumer*) qui peut en demander l'accès. L'Archive est enfin placée sous la tutelle du Management (*Management*), qui définit son champ d'action et oriente son travail en fonction de ses attentes.



Entités fonctionnelles OAIS¹

¹ Schéma extrait de *Recommendation for Space Data System Practices. Reference model for an open archival information system (OAIS). Recommended practice, CCSDS 650.0-M-2. Magenta book*, publié par le *Consultative Committee for Space Data Systems (CCSDS)*, juin 2012 (<http://public.ccsds.org/publications/archive/650x0m2.pdf>)
Schéma reproduit avec l'aimable autorisation du CCSDS.

Le modèle d'information

Le modèle OAIS repose sur l'idée que l'information constitue des paquets et que ces paquets ne sont pas les mêmes suivant que l'on est en train de produire l'information, d'essayer de la conserver ou de la communiquer à un utilisateur. On a donc trois sortes de paquets :

- les objets numériques sur lesquels travaillent les Producteurs avant archivage sont les SIP (*Submission Information Package* ou Paquets d'informations à verser) ;
- une fois archivés, les SIP deviennent des AIP (*Archival Information Package* ou Paquets d'informations à archiver), objets internes à l'Archive ;
- les objets numériques mis à disposition des Utilisateurs sont les DIP (*Dissemination Information Package* ou Paquets d'informations à diffuser), transformés par l'Archive à partir des AIP dans une forme plus facile à communiquer, notamment sur le réseau.

Si l'on veut pérenniser l'information contenue dans un objet numérique, il n'est pas suffisant de conserver cet objet. Il est indispensable de conserver, avec cet objet, un ensemble d'informations qui permettront de passer des bits constituant l'objet numérique au contenu informationnel de cet objet. C'est ainsi que dans chaque paquet, à chaque stade, on va trouver des fichiers informatiques correspondant à l'objet ou au document que l'on veut conserver, et des informations sur ce document, c'est-à-dire des métadonnées. Un Paquet d'informations (SIP, AIP ou DIP) est donc un conteneur conceptuel de deux types d'informations : le Contenu d'Information et l'Information de pérennisation.

Le Contenu d'information (*Content information*) regroupe l'Objet-données (*Data object*), qui peut être physique (tel qu'un livre ou document) ou numérique (il s'agit dans ce dernier cas d'une suite de bits, 1 ou 0, écrite sur un support numérique quel qu'il soit) et son Information de représentation (*Representation information*), c'est-à-dire son format, sa structure, sa signification, etc. nécessaire à la compréhension de cet objet par la Communauté d'utilisateurs cible.

La notion d'Information de représentation ne se borne pas à celle de format de fichier : on peut y inclure des Informations de structure (*Structure information*) plus générales comme la norme utilisée pour le codage des caractères ou des Informations sémantiques (*Semantic information*) qui éclairent le sens de l'information, comme la langue employée ou la signification d'abréviations et de codages conçus par l'homme.

Les Informations de représentation ont également besoin de leurs propres Informations de représentation, par un mécanisme récursif. Le nombre d'Informations de représentation versées, conservées ou ajoutées est déterminé par l'Archive en fonction des besoins de sa Communauté d'utilisateurs cible (*Designated community*), c'est-à-dire de personnes susceptibles de recourir aux Informations conservées par l'Archive et qui partagent une Base de connaissances (*Knowledge base*) commune. La Communauté cible comme sa Base de connaissances sont susceptibles d'évoluer dans le temps, et nécessitent dans ce cas que l'Archive anticipe la multiplicité et l'hétérogénéité des Utilisateurs dans le temps.

Le Contenu d'information et le PDI sont identifiés et encapsulés par une Information d'emballage (*Packaging Information*). Le paquet qui en résulte peut être retrouvé grâce à une Description de paquet (*Package description*). Cette Description synthétise à la fois la teneur du Contenu d'information et celle des Informations de pérennisation sous une forme qui permet aux Utilisateurs la recherche des ressources qui les intéressent d'après un certain nombre de critères.

Il est indispensable, pour assurer la Pérennisation d'une Information, de conserver à la fois le Contenu d'information, l'Information de pérennisation, l'Information d'emballage et l'Information de description du paquet, et pour chacune, l'Objet-données et ses Informations de représentation. Dans la pratique, les différents types d'information qui aident à la compréhension et à la pérennisation du Contenu d'information sont répartis entre différents formats de métadonnées spécifiquement adaptés (pour les métadonnées descriptives, techniques, administratives ou d'emballage) ou compilés dans des formats d'encapsulation qui permettent d'intégrer des métadonnées de plusieurs formats comme METS (*Metadata encoding and transmission standard*).

Les entités fonctionnelles

Le modèle de référence décrit de manière très précise toutes les fonctions qui, au sein de l'Archive, sont nécessaires pour assurer la Pérennisation de l'Information depuis sa prise en charge jusqu'à sa communication. Ces fonctions sont regroupées en Entités fonctionnelles, au nombre de six : l'Administration, les Entrées, le Stockage, la Gestion des données, l'Accès et la Planification de la Pérennisation.

L'entité Administration (*Administration*) a pour rôle d'assurer le suivi constant des activités de l'OAIS en demandant à chaque Entité fonctionnelle de

produire des rapports documentant toutes les étapes du processus d'archivage, assurant ainsi sa traçabilité par la tenue de journaux d'évènements conservés d'une manière pérenne, au même titre que les ressources prises en charge pour archivage. Elle a aussi un rôle de communication avec l'extérieur. Elle reçoit ses missions et son budget du Management, auquel elle fournit des rapports d'activité. C'est également elle qui négocie avec les Producteurs les protocoles de versement des Paquets d'informations dans leurs aspects techniques (date du versement, périodicité éventuelle, format des fichiers, nature des informations, supports utilisés, etc.). Elle s'occupe enfin, le cas échéant, de la facturation de ses services auprès des Utilisateurs ainsi que de l'évaluation de leurs besoins.

L'Entité fonctionnelle « Entrées » (*Ingest*) est responsable de l'accueil des SIP par le Producteur dans l'Archive. Elle doit d'abord s'assurer que le transfert respecte les termes du protocole négocié précédemment. À partir du SIP, elle génère un AIP (*Archival Information Package* ou Paquets d'informations à archiver). Elle a en charge la rédaction de l'Information de description de l'AIP qu'elle communique à l'entité « Gestion des données ». L'AIP est ensuite transmis à l'entité « Stockage ». Il est important de noter qu'à un SIP ne correspond pas nécessairement un AIP équivalent : on peut, en effet, procéder par exemple à des regroupements de versements, on peut également ajouter dans l'AIP les fichiers convertis, à partir des fichiers sources, vers un format pérenne, ou bien encore, dans le cadre d'une opération de numérisation, les fichiers images dans un format destiné à la conservation à long terme et des fichiers dans des formats compressés adaptés à la consultation. De même, on peut trouver dans l'AIP des métadonnées supplémentaires permettant de rendre la ressource plus explicite aux communautés d'utilisateurs.

Les fonctions de l'entité « Stockage » (*Archival Storage*) couvrent la gestion du parc de supports selon la politique décidée par l'Administration, ce qui comprend le suivi de leur vieillissement, leur remplacement, la gestion de l'espace disponible, la correction d'éventuelles erreurs d'écriture, l'écriture des fichiers à archiver sur des sites distants (duplication ou réplique synchrone ou asynchrone) et si possible sur des types de supports différents, les procédures de sauvegarde. Les technologies permettant d'assurer l'intégrité des fichiers conservés (scelllements numériques) sont également incluses dans ces fonctions. Enfin, dans le cas où l'Archive cesse ses activités, un plan de reprise d'activité qui puisse assurer la continuité de la prise en charge des Informations archivées doit impérativement être mis en place.

L'entité « Gestion des données » (*Data Management*) a en charge la gestion de la base de données qui réunit l'ensemble des descriptions des AIP détenus par l'Archive et en assure les mises à jour en collaboration avec les entités « Entrées » (qui transmet les descriptions) et « Administration » (qui transmet les mises à jour du système). De fait, c'est également cette entité « Gestion de données » qui exécute les requêtes dans la base de données relayées par l'entité « Accès ».

L'entité « Accès » (*Access*) joue le rôle d'interface entre les Utilisateurs et les entités de l'OAIS qui entrent en jeu pour satisfaire leurs demandes. Après avoir reçu et transmis les requêtes de l'Utilisateur à l'entité « Gestion de données », elle reçoit une commande de tout ou partie du résultat de ces requêtes. Elle se tourne alors vers l'entité « Stockage » pour demander les AIP concernés. À partir du transfert d'AIP et de l'Information de description, elle génère un DIP qui est finalement transmis à l'utilisateur. Là encore, le DIP n'est pas forcément l'équivalent de l'AIP, de la même façon que, dans le monde papier, des lecteurs peuvent demander la consultation de dossiers ou d'extraits de dossiers issus de plusieurs fonds, séries, etc. Le DIP peut être ainsi constitué à partir de plusieurs parties d'AIP, le contexte de production devant bien évidemment être conservé lors de l'affichage des résultats.

L'entité « Planification de la pérennisation » (*Preservation planning*) a un rôle déterminant : c'est elle qui élabore la stratégie de l'Archive en matière de conservation à long terme des données numériques. Pour cela, elle effectue un triple suivi : une veille technologique visant à anticiper l'obsolescence technologique des logiciels, matériels, systèmes d'exploitation, des supports, des formats, ainsi qu'un suivi de l'activité de la communauté des Producteurs et enfin un suivi des Utilisateurs, qui peuvent formuler des exigences nouvelles. En fonction de ces observations, l'entité propose à l'Administration des stratégies générales de pérennisation. Par exemple, elle définit des standards de métadonnées nécessaires à la pérennisation des AIP, ou bien encore des plans de migration des Informations numériques qui peuvent concerner aussi bien les supports que les formats, auxquels la norme consacre un chapitre.

Impacts du modèle OAIS

L'OAIS est désormais un modèle mondialement utilisé, tant par des bibliothèques numériques que par des services d'archives institutionnels, des centres de données scientifiques ou des entreprises industrielles comme dans l'aéronautique.

L'OAIS est par conséquent devenu une base conceptuelle pour définir des normes de mise en pratique comme le PAIMAS (ISO 20652) ou le standard XFDU¹ qui traite de l'empaquetage de données et métadonnées pour les échanges et pour l'archivage.

Autres normes généralistes fondées sur le modèle OAIS

Si l'OAIS est le modèle conceptuel fondamental de l'archivage électronique et mérite donc d'être considéré comme une base en la matière, deux autres normes généralistes peuvent lui être associées : la norme PAIMAS (*Producer archive Interface methodology abstract standard*, ISO 20652) et la norme NF Z 42013 (ISO 14641).

- La norme PAIMAS (ISO 20652)

Dans un processus d'archivage numérique, la phase de versement est l'une des plus critiques. Lorsque l'Archive a reçu des documents numériques conformes à ce qu'elle attend et convenablement décrits, une part importante des obstacles a été franchie. Un certain nombre de normes interviennent dans la phase de versement, dont les normes d'empaquetage. La norme PAIMAS vise pour sa part à définir, pour l'Archive et le Producteur, une méthodologie de travail commune et structurée en quatre phases, elles-mêmes caractérisées par une série d'actions à conduire. La norme se présente comme un guide générique qui peut être spécialisé en fonction du cadre particulier de l'application et des spécificités des partenaires. La norme PAIMAS est dérivée directement du modèle OAIS dont elle utilise le vocabulaire et les concepts.

La première phase est une phase dite préliminaire qui consiste à établir un premier contact entre le Producteur et l'Archive, à évaluer la faisabilité et à établir un accord préalable.

Vient ensuite une phase de définition formelle qui a pour objectif de négocier le « protocole de versement ». Ce protocole doit inclure une description complète, précise et formelle des objets numériques à livrer à l'Archive par le Producteur et un planning. Cette description doit faire l'objet d'un modèle formel, c'est-à-dire d'un modèle construit à l'aide d'un formalisme permettant à un individu ou une personne d'analyser et de comprendre sa cohérence et permettant également à la machine, à l'ordinateur, d'en faire un usage automatisé au cours du processus de transfert. Ce formalisme peut lui-même s'appuyer sur un dictionnaire décrivant les différentes classes d'objets numériques qui seront transférées ainsi que leurs principales caractéristiques techniques et sémantiques.

¹ XML *Formatted Data Unit*.

La phase suivante, dite phase de transfert, correspond au transfert effectif des paquets d'informations à verser. Elle inclut les tests du transfert puis sa mise en œuvre opérationnelle.

Enfin, la phase de validation comprend l'homologation des paquets reçus par l'Archive et toutes les actions que celle-ci peut impliquer. Différents types de validation des paquets sont envisagés : validation au fil de l'eau de chaque paquet, validation globale d'un ensemble de paquets constituant un tout logique, validation systématique, validation approfondie par échantillonnage.

À l'issue de cette dernière, les Objets sont normalement prêts à être insérés dans l'Archive.

- La norme NF Z 42013 (ISO 14641)

La norme NF Z 42013 « 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 », désormais portée à l'international sous le numéro ISO 14641, est antérieure à la norme OAIS puisque sa première version date de 1999. Cette norme a été révisée en 2009, ce qui a permis notamment d'élargir son périmètre à tous les types de supports numériques et non plus aux seuls disques optiques numériques non réinscriptibles.

Il s'agit plus d'une norme technique que d'une norme fonctionnelle, se concentrant sur les caractéristiques du système informatique sur lequel l'Archive numérique va appuyer ses activités. Elle met tout particulièrement l'accent sur la traçabilité de tous les processus en œuvre dans l'archivage électronique comme la numérisation de documents, l'horodatage, les communications, etc. ainsi que sur les exigences du système en matière de sécurité et d'accès. Son importance tient aussi au fait qu'elle définit les clauses nécessaires dans un contrat de service passé avec un tiers archiveur.

De plus, cette norme insiste largement sur la question de l'intégrité des documents archivés et sur la capacité de l'Archive à apporter la preuve de cette intégrité. Un point essentiel abordé dans ce texte est celui de la nécessité de disposer d'une description détaillée de tous les constituants du système d'archivage et d'une traçabilité exhaustive de tous les changements au sein de l'Archive. Il doit être possible de démontrer la continuité de l'enregistrement de tous les événements tracés.

Chaque document archivé sera rattaché à un « profil d'archivage » défini comme un ensemble de règles applicables aux documents partageant les mêmes critères de confidentialité, de durée de conservation, de destruction et de droits d'accès.

Une gradation des niveaux d'exigence permet d'adapter les procédures au niveau du risque que l'Archive accepte de prendre. Un recours différencié à l'horodatage, au calcul d'empreintes, au cryptage et à la signature électronique sera utilisé en fonction des besoins.

Cette norme apporte donc, au plan de la mise en œuvre, un grand nombre de principes, de règles et de solutions techniques de nature à donner au système d'archivage toutes les garanties de sécurité et d'intégrité qui peuvent être exigées pour les documents numériques.

Textes spécialisés entrant dans le champ de l'archivage électronique¹

Autour des normes généralistes que nous venons de décrire, gravitent de nombreux textes qui peuvent entrer dans le champ de l'archivage électronique et concernent des domaines nombreux et variés (stockage, formats, métadonnées, intégrité et authenticité, emballage, identification, audit et certification, etc.). S'il n'est pas imaginable qu'une même personne soit en mesure de connaître en profondeur tous ces domaines, il sera cependant indispensable que tout responsable d'une Archive numérique dispose d'une vue d'ensemble et puisse s'appuyer sur des compétences spécialisées pour chacun d'eux en tant que de besoin. Ceci est d'autant plus vrai qu'aux normes proprement dites, doivent être ajoutés d'autres textes tels que les standards, les bonnes pratiques ou recommandations.

Les lignes qui suivent ont pour objet d'en dresser un rapide panorama, non exhaustif.

¹ Le projet InterPares 3 a abouti à la publication d'un document contenant une liste des normes et standards utiles à la préservation de l'information numérique *General Study : international standards relevant to the InterPARES 3 project*. Disponible à l'adresse : http://www.interpares.org/display_file.cfm?doc=ip3_gs04_international_standards_v1-4p.pdf

Stockage

Du stockage relève tout ce qui concerne la préservation des bits. Il existe, d'une part, des guides et recommandations sur la stratégie de stockage, l'organisation d'un service de stockage et, d'autre part, des normes sur les supports d'enregistrement, comme la norme ISO 9660 pour les cédéroms ou les normes ISO 13421 et 13962 pour les DLT (*Digital Linear Tape*). Ces normes sur les supports d'enregistrement sont utilement complétées par les évaluations de certains de ces supports, telles que celles réalisées par le Laboratoire national de métrologie et d'essais (LNE)¹.

Formats de représentation de l'information

Qu'il s'agisse de textes, d'images, d'images 3D, de graphiques vectoriels, de documents multimédia, d'audio, de vidéo, de données techniques ou de données scientifiques, toutes les informations numériques sont représentées au moyen d'un format. Il existe des formats généralistes et des formats métier, certains sont normalisés, d'autres ne le sont pas.

Parmi les formats normalisés, citons notamment HTML (ISO 15445:2000), PDF/A (ISO 19005-1) et PDF1.7 (ISO 32000-1) – deux formats propriétaires d'Adobe –, ODF (*Open Document Format*, normalisé ISO 26300), PNG (ISO 29500), etc.

En matière de formats, il convient également de distinguer les formats libres, généralement recommandés car les spécifications du format sont publiques, des formats normalisés qui ne sont pas nécessairement libres. Ainsi, un format normalisé n'est pas nécessairement garant de pérennité. C'est notamment le cas du format Office Open XML de Microsoft dont les spécifications, bien que normalisées (ISO/IEC 29500), sont d'une complexité telle que la normalisation n'est pas garante de pérennité.

Formats d'échanges

Depuis 2006, les Archives de France ont élaboré en collaboration avec la direction générale de la modernisation de l'État (DGME) le Standard d'échange des données pour l'archivage (SEDA), actuellement stabilisé dans sa version 1.0².

¹ Disponible à : <http://www.lne.fr/>

² Disponible à : <http://www.archivesdefrance.culture.gouv.fr/seda/>

Le SEDA modélise les différentes transactions qui peuvent avoir lieu entre des acteurs dans le cadre de l'archivage de données. Ces transactions sont au nombre de six : le transfert, la demande de transfert, la modification, l'élimination, la communication et la restitution. Les acteurs sont eux au nombre de cinq : le service producteur, le service versant, le service d'archives, le service de contrôle et le demandeur d'archives.

Le SEDA définit de manière formelle d'une part les transactions par des scénarios (diagrammes de séquences en UML) dans lesquels les acteurs s'échangent des messages et d'autre part la forme des messages échangés au cours de ces transactions par des schémas XML.

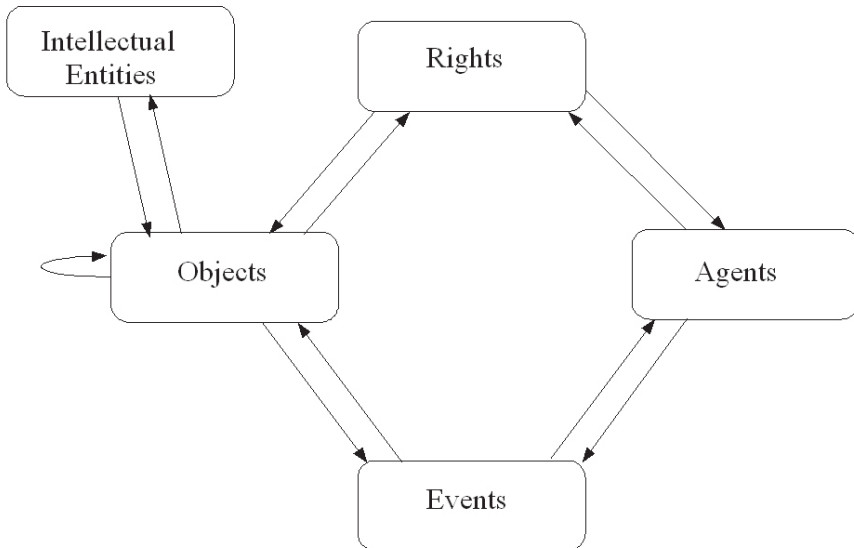
Métadonnées

Les métadonnées sont des données qui caractérisent d'autres données. Avec le numérique, des besoins nouveaux sont apparus en termes de métadonnées, comme par exemple la nécessité de disposer de métadonnées techniques garantissant la capacité future à restituer une information compréhensible à partir d'un ensemble de séquences de bits ou encore de retrouver immédiatement les documents correspondant à un besoin donné, au sein de millions d'autres documents, éventuellement stockés de façon répartie sur des sites géographiquement distants.

La généralisation du numérique a entraîné l'émergence de très nombreux formats de métadonnées, qu'ils soient à caractère généraliste comme le Dublin Core (ISO 15836) ou orientés métier comme la norme ISO 19115 de métadonnées géographiques qui propose plus de 400 rubriques descriptives. La norme ISAD(G), qui permet de décrire des documents d'archives, peut être rattachée à ce domaine normatif, tout comme la norme PREMIS relative aux informations de préservation.

PREMIS¹ est en effet un standard pour l'expression des métadonnées de préservation. Ce modèle concerne autant ceux qui conservent que ceux qui transmettent de l'information à conserver. Il définit principalement un dictionnaire de données utile pour décrire les métadonnées de préservation. Il est possible d'être conforme à PREMIS si on utilise le vocabulaire pour exprimer les métadonnées même si on n'utilise pas le schéma XML qui en est proposé. Il est possible d'utiliser PREMIS dans le cadre de METS (*Metadata encoding and transmission standard*) ce que font déjà quelques éditeurs de systèmes d'archivage électronique. Tout comme METS, ce standard est hébergé et maintenu par la Bibliothèque du Congrès.

¹ Disponible à : <http://www.loc.gov/standards/premis/>



PREMIS¹

Intégrité et authenticité

Il s'agit des normes qui vont permettre d'apporter des preuves de l'intégrité et de l'authenticité des documents. Ces normes vont couvrir différents domaines d'application : la cryptographie, les algorithmes de calcul d'empreinte, les protocoles de communication sécurisés, les processus de signature électronique.

Un certain nombre de standards a été proposé par le projet NESSIE (*New European Schemes for Signatures, Integrity, and Encryption*) de la Commission européenne. Le NIST (*National Institute for Standards and Technology*), américain, propose également un ensemble de standards et, en définitive, l'ISO a entériné les choix du NIST et de NESSIE, et normalisé les principaux algorithmes de calcul d'empreinte numérique.

¹ <http://www.loc.gov/standards/premis/> : reproduit avec l'autorisation de PREMIS - Bibliothèque du Congrès (États-Unis).

Empaquetage

Les standards d'empaquetage permettent d'associer, au sein d'une structure globale cohérente et portable, des objets numériques pouvant être composés d'un fichier ou d'un ensemble de fichiers, de métadonnées caractérisant ces objets et ces fichiers, des liens existant entre les fichiers eux-mêmes ou entre les composants de l'objet et les métadonnées.

Ces standards permettent aussi de définir des mécanismes applicables aux objets numériques : ils peuvent exiger, par exemple, que tel objet soit ouvert en utilisant telle application, ou encore que les différents fichiers de l'objet soient ordonnancés et traités d'une façon telle à pour pouvoir restituer leur contenu. Les standards d'empaquetage vont ainsi permettre d'identifier, de décrire, de manipuler les composants numériques élémentaires d'un objet complexe.

Le standard METS de la Bibliothèque du Congrès relève de ce domaine. METS est un standard définissant un format d'empaquetage, c'est-à-dire un format permettant d'organiser un ensemble d'informations liées de manière explicite. Ce standard définit un schéma XML qui est utilisé dans des systèmes d'archivage électronique pour donner une forme aux paquets d'informations, en particulier pour l'AIP (le paquet qui est conservé au sein du système). Le schéma permet d'encapsuler et de placer dans des catégories distinctes des métadonnées d'usage divers telles que des métadonnées EAD, Dublin-core ou encore PREMIS.

Identification

La problématique consistant à identifier un objet numérique de manière unique dans un domaine au sein de l'Archive ou plus souvent aujourd'hui, dans un domaine beaucoup plus vaste, n'est pas nouvelle et a déjà été résolue par exemple avec l'usage de l'ISBN (*International Standard Book Number*) pour les ouvrages publiés ou de l'ISSN (*International Standard Serial Number*) pour les revues.

La question de l'identification des documents se pose de manière particulièrement aiguë dans le domaine numérique. Des standards d'identification (URI, *Uniform Resource Identifier*) et de nommage (URN, *Uniform Resource Name*) ont été définis par le consortium W3C¹, mais aujourd'hui les navigateurs ne savent pas les interpréter. Ils ne savent pas non plus interpréter

¹ *World Wide Web Consortium* : organisme de normalisation à but non-lucratif chargé de promouvoir la compatibilité des technologies du *World Wide Web* (Wikipédia).

que les URL (*Uniform Resource Locator*) sont des standards de localisation. Or, chacun sait aujourd'hui à quel point les URL ne sont pas pérennes, alors que le besoin de posséder des identifiants stables va croissant. Dans ce cadre, différents services d'affectation d'identifiants pérennes comme ARK (*Archival Resource Key*) ou les DOI (*Digital Object Identifier*) ont vu le jour.

La certification des archives numériques

La question qui se pose ici est de savoir sur quels éléments s'appuyer pour accorder une réelle confiance à une archive numérique.

D'une manière générale, lorsqu'une entreprise ou une institution veut démontrer qu'elle fait bien son travail, qu'elle est apte à fournir des produits ou des services satisfaisant aux exigences des clients et à la réglementation applicable, qu'elle vise en permanence à accroître ses performances et la satisfaction de ses clients, elle met en place un système de contrôle de la qualité et entreprend un processus de certification tel que ISO 9001. Cette certification est délivrée pour une durée définie par des auditeurs indépendants, eux-mêmes accrédités par un organisme agréé pour délivrer ces accréditations.

Dans le domaine de l'archivage numérique, plusieurs approches émergent en matière de certification :

- une approche liée à un audit de conformité du service d'archivage numérique par rapport à une spécification technique existante comme la norme NF Z 42-013 ;
- une approche plus ambitieuse et couvrant réellement la totalité de la problématique de l'archivage numérique, basée sur la norme ISO 16363 (*Audit and Certification of Trustworthy Digital Repositories*) et la norme ISO 16919 sur les exigences requises pour auditer et certifier un service d'archivage numérique, norme en cours d'élaboration.

Autres domaines normatifs

Enfin, parmi les autres domaines normatifs à prendre en compte dans l'archivage numérique, il convient de citer également tout ce qui concerne les systèmes de *management* de la sécurité de l'information – série des normes ISO 27 000 (incluant la confidentialité, la protection des informations sensibles, les plans de continuité d'activité) –, ainsi que les questions relatives à l'externalisation de prestations d'archivage (NF Z 40-350).

Conclusion

L'environnement normatif s'est considérablement enrichi en quelques années et constitue un apport irremplaçable pour l'archivage numérique.

Le cadre normatif reste cependant difficile à appréhender en raison de la diversité des points de vue et des périmètres couverts, mais également en raison des recouvrements partiels existant entre certaines normes. Cette situation complexe a pour origine essentielle la multiplicité des secteurs d'activités ayant à faire face à des nécessités d'archivage numérique. Chacun de ces secteurs se retourne naturellement vers les organismes de normalisation et, au sein de ces organismes, vers les comités techniques dont ils relèvent habituellement alors qu'une approche transverse serait probablement bénéfique pour tous.

Stéphanie ROUSSEL
Conservateur du patrimoine
Service interministériel des Archives de France