

Module 7C - Section 4 : Mettre en place, administrer et auditer un système d'archivage pour les documents d'archives sur support numérique

Édouard Vasseur @AIAF - PIAF

VF 02/12/2024



Table des matières

Objectifs	4
Introduction	6
1. Des approches différenciées	8
1.1. Un guide normatif, la norme NF Z42-013	8
1.1. Les grandes fonctionnalités du système d'archivage que préconise la norme	9
1.2. La norme souligne les problématiques transversales de la gestion d'un système d'archivage.....	10
1.2. Quel système pour les organisations de petite taille et aux ressources (humaines, financières, techniques) limitées ?.....	10
1.2.1. Constat	10
1.2.2. Procédure à adopter dans une petite structure	11
1.3. Les approches plus complexes : 3 stratégies	12
Première stratégie : l'acquisition d'une solution logicielle « sur étagère »	12
Deuxième stratégie : la conception et le développement d'une plateforme ad hoc	12
Troisième stratégie : l'acquisition d'un service d'hébergement ou d'archivage	12
Conclusion	13
2. Standards d'échange	14
2.1. Metadata Encoding and Transmission Standard (METS)	14
1.1. Description du standard METS.....	14
1.2. La souplesse du standard MELTS.....	15
2.2. Protocole d'échange de données pour l'interopérabilité et la préservation (Depip) et Standard d'échange de données pour l'archivage (SEDA)	16
2.1. Le standard SEDA offre 5 types d'échanges (transactions).....	17
2.2. Structure d'un schéma SEDA.....	17
2.3. À noter.....	17
3. Solutions logicielles	19
3.1. Fonctions des solutions logicielles	19
1.1. Automatiser le fonctionnement des systèmes d'archivage.....	19
1.2. Mettre en oeuvre des tâches spécifiques au sein des systèmes d'archivage.....	19
3.2. Exemples de solutions logicielles utilisées pour la réalisation de systèmes d'archivage	20
2.1. 3.2.1. Archivematica.....	20
2.2. 3.2.2. Asal@e.....	20
2.3. 3.2.3. Preservica.....	20
2.4. 3.2.4. Vitam	21
4. Audit et certification des systèmes d'archivage	22
4.1. Considérations générales	22
4.2. Audit ISO 16363 et la certification d'entrepôts numériques de confiance	22

4.3. CoreTrustSeal	23
4.4. Méthodologie pour la mise en place d'une opération d'audit et de certification.....	24
4.1. Les questions à se poser	24
4.2. Se préparer en amont... ..	24
5. Le système d'archivage au quotidien	25

Objectifs



Description du module :

La préservation des documents d'archives sur support numérique – ce que les Québécois nomment documents technologiques – constitue désormais un enjeu quotidien des archivistes. L'archiviste dispose désormais d'un important panorama de normes, de standards, d'outils et de retours d'expérience pour lui permettre d'appréhender les documents d'archives sur support numérique et envisager leur préservation dans le temps.

Le but du module est de :

- aider à évaluer la situation en matière de préservation des documents d'archives sur support numérique ;
- permettre de concevoir et mettre en œuvre une politique de préservation des documents d'archives sur support numérique.

L'apprenant doit être en mesure de :

- appréhender les spécificités en matière de préservation des documents d'archives sur support numérique ;
- dresser un état des lieux d'ensembles de documents d'archives sur support numérique ;
- définir et mettre en œuvre une politique de préservation des documents d'archives sur support numérique ;
- concevoir, mettre en œuvre et administrer un environnement permettant de gérer dans le temps les documents d'archives sur support numérique, quels que soient les moyens disponibles ;
- appréhender les différentes catégories de formats de fichiers numériques ;
- savoir comment aller plus loin dans la réflexion.

Positionnement :

Ce module s'inscrit naturellement dans la chaîne archivistique. S'il se concentre sur les questions de planification de la préservation, de mise en œuvre de la préservation et de stockage des documents d'archives sur support numérique, il fournit également des éléments à prendre en compte lors de la mise en place de politiques et procédures de gouvernance de l'information et de gestion de l'archivage/gestion des documents d'activité/gestion des documents institutionnels/records management, de collecte de documents d'archives définitifs et d'accès à ceux-ci.

Il ne s'intéresse en revanche pas à la numérisation de documents d'archives sur support physique ou d'enregistrements sonores et audiovisuels sur support analogique, sauf dans le cas où l'opération de numérisation vise à substituer la version du document sur support numérique à celle sur support physique ou analogique.

Point sur le vocabulaire employé :

- Le terme “préservation” est entendu comme recouvrant « les fonctions de conservation préventive et matérielle » [Direction des Archives de France, Dictionnaire de terminologie archivistique, 2002] ;
- Sont distingués :
 - **les documents d’archives sur support physique**, où l’information est directement accessible à l’œil humain ou ne nécessite, pour le devenir, que l’emploi d’un appareil optique (projecteur) permettant de faciliter son agrandissement
 - **les documents d’archives sur support analogique**, où l’information, pour être intelligible, a absolument besoin de la médiation d’un appareil pour permettre à l’utilisateur de prendre connaissance de l’information (projecteur, lecteur, etc.) ;
 - **les documents d’archives sur support numérique**, qu’ils aient été directement produits avec des outils numériques ou soient le produit de la numérisation de documents d’archives sur support physique ou analogique. L’information, pour être intelligible, a absolument besoin de la médiation d’un environnement matériel et logiciel pour permettre à l’utilisateur de prendre connaissance de l’information ;
- “Document d’archives” est l’expression utilisée pour identifier toute information sur un support qui a besoin d’être prise en charge et conservée, soit pour sa valeur de preuve, soit pour sa valeur informationnelle, soit pour sa valeur patrimoniale ou de recherche. En fonction du contexte, l’expression pourra concerner des documents, des records ou des archives au sens anglo-saxon des termes ;
- “Service d’archives” est l’expression utilisée pour désigner toute structure ou organisme souhaitant mettre en place une politique de préservation de documents d’archives sur support numérique. Ce service d’archives peut être
 - interne à une organisation productrice et en charge de la gouvernance de l’information et de la gestion de l’archivage/gestion des documents d’activité/records management ou de la gestion d’archives intermédiaires ;
 - externe à une organisation productrice, soit qu’il s’agisse d’un prestataire de tiers archivage, soit d’un service d’archives définitif.

Les notions abordées dans ce module peuvent être complétées par :

- le module 9 - Section 2 : Numériser les documents qui présente les techniques de base de transfert de support vers le numérique
- le module 5 Gestion et traitement des archives courantes et intermédiaires

Il est vivement conseillé de prendre connaissance du module 7B Gestion des documents numériques au stade courant avant d’entamer la lecture du module présentement proposé. Certaines notions de base, activées ici, sont exposées plus longuement dans ce premier module.

Le glossaire du PIAF doit être consulté pour les définitions des termes spécifiques.

Introduction



La réception, la prise en charge, le stockage, la conservation et l'accès aux documents d'archives sur support numérique nécessitent, pour un service d'archives, la mise en œuvre et la maîtrise d'un système d'archivage adapté.

Système d'archivage



Par système d'archivage, on entend ici tout l'environnement organisationnel, humain, matériel et logiciel – donc une combinaison de personnes, de processus et de technologies qui fournit les moyens de recevoir, prendre en charge, stocker, conserver et donner accès aux documents d'archives sur support numérique – permettant de mettre en œuvre des procédures d'archivage adaptées aux documents d'archives sur support numérique.



La question sémantique prend ici une certaine importance. Longtemps, dans le monde francophone, on a utilisé l'expression *système d'archivage électronique (SAE)*. Cette expression traduit deux concepts anglo-saxons :

- les *electronic records management systems*, à savoir les systèmes d'archivage électronique adaptés à la prise en charge et à la conservation de ce que les Anglo-saxons désignent sous le terme de *records* jusqu'à la mise en œuvre du *sort final*. Le maintien du caractère digne de confiance et de la valeur probante des documents d'archives (c'est-à-dire leur capacité à servir de preuve en cas de contentieux), ainsi que la gestion du cycle de vie de ceux-ci jusqu'à leur sort final, constituent les finalités principales de ce type de système ;
- les *digital repositories*, à savoir des systèmes d'archivage adaptés à la prise en charge, à la conservation et à la préservation sur le long terme des documents d'archives sur support numérique. Ces systèmes ont davantage pour finalité l'accès à long terme aux documents d'archives sans nécessairement s'assurer du maintien de la valeur probante et de la mise en œuvre d'un sort final, et prennent en charge les opérations de préservation (ex. caractérisation, migration de format).

Nous utiliserons de préférence l'expression système d'archivage adapté aux documents d'archives sur support numérique dans la mesure où :

- les services d'archives doivent gérer un système d'archivage global, capable de prendre en charge et de conserver des documents d'archives sur supports physiques, analogiques et numériques ;
- ce système d'archivage peut avoir pour finalité la conservation de records ou d'archives au sens anglo-saxon des termes ;
- certaines fonctionnalités sont communes à tous les documents, quel que soit leur support, tandis que d'autres sont spécifiques aux documents d'archives sur support numérique.

Le présent chapitre a pour objectifs de donner un aperçu des fonctionnalités qui doivent être celles d'un système d'archivage adapté aux documents d'archives sur support numérique, de présenter les standards d'échange qui ont été élaborés pour automatiser les échanges avec ces systèmes d'archivage ainsi que les outils logiciels qui ont été conçus et développés pour aider les services d'archives à conserver les documents d'archives sur support numérique et, enfin, d'exposer l'audit et la certification de ces systèmes.

1. Des approches différenciées



Introduction

S'il y a des constantes dans les attendus des systèmes d'archivage adaptés aux documents numériques, les chemins pour les atteindre diffèrent en fonction de nombreuses variables touchant à la fois aux compétences, aux moyens, aux volumes ou aux choix stratégiques.

1.1. Un guide normatif, la norme NF Z42-013

Un système d'archivage adapté aux documents d'archives sur support numérique pour les recevoir, les prendre en charge, les stocker, les conserver, les préserver et leur donner accès est un dispositif réalisable de différentes manières.

Plusieurs modèles existent, incluant ou pas un ou plusieurs systèmes d'information dédiés :

- des modèles pour **organisations de petite taille** et des modèles pour **organisations de grande taille** ;
- des modèles **internalisés** dans les services d'archives ou des modèles **externalisés** auprès de prestataires ;
- des modèles **propres aux services d'archives** ou des modèles **mutualisés** avec d'autres services (services d'archives ou autres services chargés d'une fonction de préservation numérique comme des bibliothèques ou des musées) ;
- des modèles s'appuyant sur un environnement matériel et logiciel **simple** ou des modèles s'appuyant sur des solutions logicielles **complexes** ;
- des modèles opérables majoritairement sous forme manuelle ou des modèles permettant une automatisation de certaines tâches.

La France dispose avec la norme NF Z42-013, approuvée en 2009 et revue en 2020, d'un outil pouvant constituer un guide pour formaliser les recommandations et exigences fonctionnelles, organisationnelles et de sécurité utiles à la conception, à la réalisation et à l'exploitation d'un système d'archivage adapté à la prise en charge et à la conservation des documents d'archives sur support numérique.



La norme internationale ISO 14641-1 est une transposition de la norme NF Z42-013 dans sa version 2009. Elle ne tient pas compte de la réactualisation de 2020.

La norme NF Z42-013, dans sa version de 2020, est organisée selon le schéma suivant.

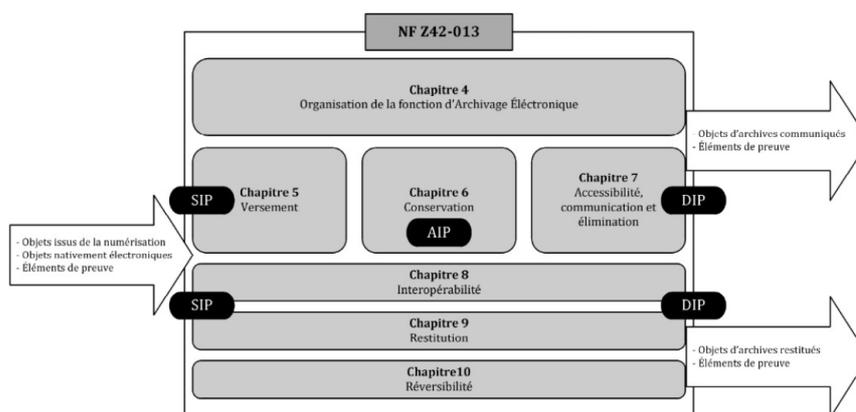


Fig. 1 : Représentation schématique des exigences et recommandations de la norme NF Z42-013 (version 2020), p. 9

2.1. Les grandes fonctionnalités du système d'archivage que préconise la norme

La norme présente les grandes fonctionnalités d'un système d'archivage adapté aux documents d'archives sur support numérique :

- **les fonctionnalités de versement**, qui rassemblent tout ce qui a trait à la création des documents – y compris des recommandations et exigences relatives à la signature des documents, à leur datation, à la numérisation de documents d'archives sur support physique ou analogique en vue de leur transformation en documents d'archives sur support numérique –, au transfert des documents dans le système d'archivage – création des SIP, migration de format de fichiers, maintien de l'intégrité pendant toutes ces opérations –, réception et prise en charge par le système d'archivage – contrôles à effectuer lors de la réception, production des AIP, écriture sur les supports de conservation, enregistrement des métadonnées dans le système ;
- **les fonctionnalités de conservation**, qui rassemblent tout ce qui a trait au stockage des documents – choix du type de support, identification des sites physiques de conservation, contraintes d'hébergement, cloisonnement physique des fonds, stockage physique, gestion de la disponibilité et de la sécurité, chiffrement, contrôle et surveillance des dysfonctionnements, contrôle d'exhaustivité, journalisation et contrôle d'intégrité des journaux, traitement des incidents –, aux opérations de migration de supports et de conversion de formats pendant la période de prise en charge par le système d'archivage – y compris les cas spécifiques des documents chiffrés ;
- **les fonctionnalités d'accès et de mise en œuvre du sort final**, qui rassemblent tout ce qui a trait à la description et aux outils intellectuels d'organisation et de gestion du cycle de vie des documents – plans de classement, règles de conservation, droits des individus et des systèmes, gestion des utilisateurs –, à la recherche et à la communication des documents – modalités de recherche, identification des documents répondant à la demande des utilisateurs, constitution des DIP, gestion des droits de consultation – et à la mise en œuvre du sort final – éliminations, restitution au service producteur ou transfert à un autre service d'archives.

2.2. La norme souligne les problématiques transversales de la gestion d'un système d'archivage

La norme insiste également sur les problématiques transversales de la gestion d'un système d'archivage :

- **la documentation du système** : les procédures, la documentation fonctionnelle et technique, les documents d'organisation du service d'archives, la description de l'architecture du système, la description des dispositifs de sécurité ;
- **les niveaux de service proposés par le système** : horaires d'ouverture, disponibilité pendant les horaires d'ouverture, temps de réponse en accès, délais de transfert et de prise en charge, temps nécessaire pour basculer vers des systèmes de secours, perte de données admissible ;
- les dispositifs à mettre en place pour assurer **l'amélioration continue du système** ;
- **les modalités de sécurité informatique** : politique de sécurité et son amélioration continue, analyse de risques, réalisation régulière de tests de sécurité ;
- **les modalités de sécurité physique des locaux** ;
- **les conditions d'audit du système** : définition et mise en œuvre d'un programme d'audit interne ou externe ;
- **la production et la conservation des éléments de preuve** : traces informatiques (logs) et journaux relatifs aux opérations réalisées sur les documents, conservation de l'intégrité des journaux, modalités de production des attestations.

Cette norme définit des recommandations et des exigences qui peuvent être utiles, quel que soit le modèle de réalisation retenu.

Elle prend à la fois en compte les fonctionnalités de préservation de la valeur probante et de préservation de l'accès à long terme.



Elle est cependant davantage pertinente pour les systèmes d'archivage destinés à prendre en charge des documents d'archives à valeur probante du type systèmes d'archivage électronique gérant des documents d'archives numériques avant application du sort final ou *electronic records management systems*.

3. 1.2. Quel système pour les organisations de petite taille et aux ressources (humaines, financières, techniques) limitées ?

3.1. 1.2.1. Constat

Il n'est pas obligatoirement nécessaire de mettre en place un environnement matériel et logiciel complexe pour disposer d'un système d'archivage adapté à la prise en charge des documents d'archives sur support numérique et satisfaisant aux principales recommandations et exigences normatives. Pour les institutions de petite taille, aux ressources humaines, financières, techniques limitées, et n'ayant d'un petit nombre de documents d'archives sur support numérique à conserver, des solutions simples existent.

3.2. 1.2.2. Procédure à adopter dans une petite structure

Une bonne combinaison des procédures, matériels et outils suivants constitue déjà un socle solide, adapté aux structures qui ne disposent pas des ressources humaines et financières similaires à celles des grands services d'archives et permettant de concevoir et développer des systèmes d'archivage sophistiqués :

a) Définir clairement les rôles et responsabilités

Cette étape est essentielle.

b) S'assurer pour opérer le transfert et la prise en charge de :

- disposer d'un **espace de stockage séparé** pour la production des documents d'archives du service, pour le traitement des documents sur support numérique reçus et pour la conservation des documents pris en charge ;
- disposer d'un **antivirus à jour** pour examiner les fichiers correspondant aux documents reçus ;
- **faire en sorte de recevoir sous forme de tableur une liste des fichiers destinés à être transférés**, avec leur empreinte, ne serait-ce qu'établie avec des outils d'identification de format comme DROID qui permettent de générer cette information. À réception des fichiers, utiliser le même outil d'identification de format pour générer une liste similaire, et comparer les deux listes pour identifier le nombre de fichiers présents (et repérer ainsi les éventuelles pertes liées aux transferts), ainsi que les empreintes des fichiers, et vérifier ainsi qu'ils sont restés intègres pendant le transfert ;
- procéder à un **examen aléatoire des fichiers reçus** en les ouvrant manuellement ou au moyen d'un petit script pour identifier d'éventuels problèmes d'accès. Les messages d'erreur proposés par les outils d'identification peuvent constituer un bon moyen de repérer des fichiers problématiques ;
- **tenir un registre des fichiers reçus**, alimenté à partir des informations récupérées et exportées par les outils d'identification de format (DROID par exemple), complétées avec les informations liées aux supports de stockage retenus. Ce registre peut prendre la forme d'un tableur ou d'une petite base de données ;
- **copier les documents pris en charge sur différents supports de stockage**, avec une copie facilement accessible sur un disque non amovible, une copie supplémentaire sur un support de stockage moins accessible et moins coûteux et entreposé dans un lieu géographique différent. Si les documents pris en charge ne présentent pas de contrainte particulière en matière de confidentialité et de sécurité, le recours à une offre d'hébergement externalisée de type cloud peut être envisagé.

c) Pendant la période de prise en charge :

- réaliser périodiquement des vérifications sur les supports de stockage, soit manuellement soit au moyen d'un petit script : sont-ils toujours lisibles ? Tous les fichiers qui sont censés être stockés sur ceux-ci sont-ils toujours présents ? Les empreintes des fichiers présents sont-elles bien les mêmes que celles enregistrées dans le registre des fichiers ?
- En cas de problème d'intégrité, remplacer les fichiers défectueux par les fichiers qui ne le sont pas, en tirant partie des différentes copies existantes.

d) En matière d'accès

- pour les documents non librement communicables, permettre l'accès à une de leurs copies sur un poste dédié, déconnecté de tout réseau informatique externe, et sans offrir la possibilité de reproduire avec l'appareil photographique d'un téléphone portable leur contenu ou d'en prendre copie via l'écriture sur un support amovible ;
- pour les documents librement communicables, utiliser les services de transfert à distance gratuits.

Conclusion

Dans tous les cas, les opérations ci-dessus peuvent être mises en œuvre avec des moyens financiers et humains limités. Pour ceux qui peuvent disposer de ressources informatiques minimales, ces procédures peuvent être partiellement automatisées en recourant à des petits scripts informatiques.

4. 1.3. Les approches plus complexes : 3 stratégies

Cependant, dans la mesure où le nombre d'archives à transférer et à prendre en charge a tendance à croître, il peut être intéressant de recourir à des approches plus complexes, mettant en œuvre des solutions matérielles et logicielles plus intégrées, qu'elles soient déployées au sein même de l'organisation ou acquises auprès de fournisseurs extérieurs.

Les services d'archives, depuis une vingtaine d'années, ont mis en œuvre des stratégies multiples, les plus gros recourant à la mise en œuvre de plateformes relativement complexes.

Trois stratégies émergent principalement :

4.1. Première stratégie : l'acquisition d'une solution logicielle « sur étagère »

L'acquisition d'une solution logicielle « sur étagère » auprès d'un éditeur du marché, plus ou moins paramétrable en fonction des demandes du service. Cette solution peut reposer sur les logiciels ouverts ou sur des logiciels propriétaires.



Cette stratégie peut nécessiter des paramétrages coûteux, notamment si le fonctionnement attendu s'écarte trop de celui proposé par défaut par la solution existante.

4.2. Deuxième stratégie : la conception et le développement d'une plateforme ad hoc

La conception et le développement d'une plateforme ad hoc, utilisant des logiciels développés spécifiquement.



Cette stratégie peut se révéler gourmande en développements, même en cas de recours à des logiciels ouverts, et risquée sur le long terme, faute de disposer des moyens pour administrer et faire évoluer la plateforme.

4.3. Troisième stratégie : l'acquisition d'un service d'hébergement ou d'archivage

L'acquisition d'un service d'hébergement ou d'archivage auprès d'un fournisseur de services externe, qui peut s'appuyer ou non sur une solution du marché, mutualisée ou non avec d'autres clients via un cloisonnement des services et espaces de stockage offerts aux différents clients.



Cette stratégie nécessite une relation contractuelle de confiance et n'est pas nécessairement applicable dans des environnements où les documents pris en charge sont confidentiels ou soumis à des exigences de sécurité fortes.



Si la mutualisation présente des avantages (partage des coûts, recours à des expertises pointues), elle nécessite un accord entre tous les partenaires sur de nombreux points :

- modalités de cloisonnement des services et espaces de stockage ;
- administration des données de référence partagées (ex. registre des formats de fichiers) ;
- accord sur les services fournis en commun (hébergement, opérations de préservation, etc.).

Il est par ailleurs indispensable d'anticiper avant même l'ouverture du service le terme contractuel de celui-ci, afin de savoir dans quelles conditions les documents d'archives sur support numérique confiés au prestataire de service pourront être récupérés. Enfin, toute apparition d'une divergence de vues entre partenaires ou retrait d'un d'entre eux peut s'avérer préjudiciable à l'ensemble.

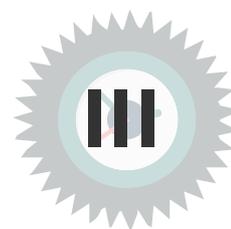
4.4. Conclusion

Dans tous les cas, il s'agit d'une opération complexe qui implique l'identification et la collaboration de multiples acteurs aux compétences variées :

- experts archivistes, chargés de définir les fonctionnalités attendues ;
- experts informaticiens, chargés de proposer des solutions pour mettre en œuvre les fonctionnalités attendues ;
- experts en sécurité ;
- experts en achats, en gestion des ressources humaines, en finances.

Dans bien des cas, l'approche retenue consistera en un mélange des différentes stratégies.

2. Standards d'échange



Introduction

Pour faciliter les échanges d'archives entre systèmes et normaliser le format attendu des SIP à l'entrée des systèmes d'archivage, des standards d'échange ont été élaborés.

Ces standards consistent en la modélisation des :

- interactions d'un système d'archivage avec des systèmes extérieurs ;
- objets échangés par le système d'archivage avec les systèmes extérieurs.

Deux standards principaux ont émergé et sont actuellement utilisés par les services d'archives :

- *Metadata Encoding and Transmission Standard (METS)* ;
- Protocole d'échange de données pour l'interopérabilité et la préservation (Depip) et Standard d'échange de données pour l'archivage (SEDA).

Néanmoins, l'utilisation de ces deux standards n'a rien d'obligatoire. Tout dépend de la plateforme utilisée, des solutions logicielles mises en œuvre et des exigences fonctionnelles et techniques imposées par celles-ci pour assurer le transfert et la prise en charge des documents d'archives.

2.1. Metadata Encoding and Transmission Standard (METS)

Le *standard d'échange METS* (Metadata Encoding and Transmission Standard) a été élaboré à partir de 2001 sous l'égide de la Digital Library Federation, avant d'être repris et maintenu par la Bibliothèque du Congrès américaine, sous l'égide d'un comité éditorial.



METS a au départ pour finalité de définir un modèle de SIP permettant d'intégrer l'ensemble des métadonnées descriptives, de gestion et de structure correspondant à des objets numériques à transférer à un **entrepôt numérique**. Plus largement, il est utilisé pour définir les modalités d'échanges entre entrepôts numériques. **Il est basé sur le langage XML.**

2.1.1. Description du standard METS

Le cœur du standard METS est constitué par un schéma XML à la structure normalisée et comprenant les **sept sections** suivantes :

- un en-tête (METS header), qui donne des informations sur le schéma METS lui-même (date de création, date de dernière modification, créateur du fichier, identifiant du fichier, etc.) ;
- une section relative aux *métadonnées descriptives* (descriptive metadata) ;
- une section relative aux *métadonnées de gestion* (administrative metadata), divisée en quatre sous-sections, qui permet d'indiquer comment les fichiers ont été créés et stockés, de préciser les droits de propriété intellectuelle applicables aux fichiers, de fournir des informations sur l'objet source dont l'objet numérique transféré dérive, de préciser les opérations de transformation subies par l'objet ;

- une section permettant de lister tous les fichiers correspondant à l'objet numérique transféré (file section), avec les informations techniques essentielles (taille, empreinte, emplacement du fichier) ;
- une section relative à la structure de l'objet (structural map), qui constitue le cœur du schéma. Elle précise la structure hiérarchique de l'objet numérique et permet de faire le lien avec les fichiers listés dans la section précédente. Il s'agit de la seule section obligatoire du schéma ;
- une section relative aux liens structurels (structural links) qui permet d'indiquer l'existence d'hyperliens entre les différentes parties de l'objet numérique décrites dans la section relative à la structure. Cette section est plus particulièrement utile en cas de transferts de sites internet ;
- une section relative au comportement (behaviour section) qui permet de préciser les prérequis techniques de restitution de l'objet numérique, en associant les objets aux programmes permettant de les lire.

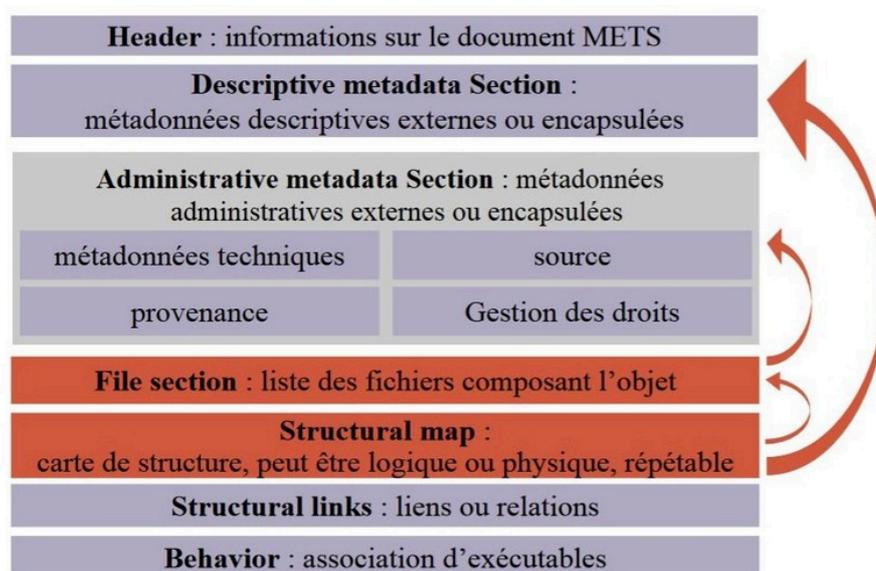


Fig. 2 : structure d'un fichier METS (source : Emmanuelle Bermès)

? Exemple

Exemple de document : <https://repository.library.brown.edu/studio/item/bdr:95917/>

Exemple de fichier METS correspondant : <https://library.brown.edu/metsrecords/109898358915625.xml>

Autres exemples : <https://www.loc.gov/standards/mets/mets-examples.html>

2.1.2. La souplesse du standard METS

Comme tout standard d'échange, le standard METS laisse une grande liberté de mise en œuvre à ses utilisateurs :

- Dans les sections relatives aux métadonnées descriptives et aux métadonnées de gestion, la manière d'exprimer les informations n'est pas imposée. C'est-à-dire qu'il est possible soit d'encapsuler dans le schéma une description exprimée dans le langage XML, en utilisant des standards externes (ex. EAD,), soit de faire référence à une description externe (ex. identifiant d'une notice dans un catalogue) ;
- les sections relatives à la structure de l'objet laissent également une large marge de manœuvre à l'utilisation ;

- la manière dont les identifiants sont exprimés dans le schéma n'est pas spécifiée dans le standard.

NB :

Sur l'EAD ou Description archivistique encodée, module 6-Traitement des archives définitives, section 8 (Publication électronique des instruments de recherche).

Le standard METS est largement utilisé par les solutions logicielles développées depuis le début des années 2000.



Le schéma étant souple, ses mises en œuvre peuvent fortement différer entre les différents entrepôts numériques, ce qui peut limiter leur capacité à s'échanger des objets numériques.

Pour résoudre cette difficulté, le standard METS permet de développer des profils, c'est-à-dire des déclinaisons particulières du schéma pour une communauté particulière ou un type d'objets particulier, spécifiant des restrictions par rapport aux possibilités qu'il offre.



Comme tout standard d'échange, METS spécifie l'organisation des fichiers et métadonnées transmis à un entrepôt numérique mais ne spécifie pas la manière dont ces fichiers et métadonnées sont gérés dans l'entrepôt lui-même.

2.2. Protocole d'échange de données pour l'interopérabilité et la préservation (Depip) et Standard d'échange de données pour l'archivage (SEDA)

Le *Standard d'échange de données pour l'archivage (SEDA)* a été élaboré à partir de 2006 par la direction des Archives de France et est toujours maintenu aujourd'hui par le ministère de la Culture/Service interministériel des Archives de France, sous l'égide d'un comité éditorial.



Le SEDA a initialement pour finalité de modéliser les différents échanges de documents d'archives sur support numérique entre les services d'archives publics et leurs partenaires (services producteurs, services versants, autorités de contrôle, tiers archiveurs, utilisateurs). Depuis la version 2.0. publiée en 2015, il permet de modéliser les échanges pour des documents d'archives sur tous supports.

Normalisé à l'AFNOR en 2014 dans la **norme NF Z44-022 Modélisation des échanges de données pour l'archivage (Medona)** – après une généralisation permettant de le rendre applicable à des services d'archives privés –, il a été porté à l'ISO et est à la source de la norme NF ISO 20614 Protocole d'échange de données pour l'interopérabilité et la préservation (Depip). **Comme le standard METS, il est basé sur le langage XML.**

2.2.1. Le standard SEDA offre 5 types d'échanges (transactions)

Le standard SEDA et ses normalisations à l'AFNOR (Medona) et à l'ISO (Depip) offrent tout d'abord une modélisation sous forme de diagrammes et de messages des échanges possibles entre un système d'archivage et ses partenaires.

Cinq échanges (ou transactions) ont ainsi été définis :

1. le transfert d'archives (*Archive Transfer*) ;
2. la communication d'archives (*Archive Delivery*) ;
3. la modification d'archives (*Archive Modification Notification*) ;
4. l'élimination d'archives (*Archive Destruction Notification*) ;
5. la restitution d'archives (*Archive Restitution*).



Complément

Les normalisations à l'AFNOR et à l'ISO se contentent de définir la structure globale des objets échangés (SIP ou DIP) sans entrer dans le détail précis de leur organisation.

Le SEDA en tant que tel, déclinaison des normalisations à l'AFNOR et à l'ISO pour les services d'archives publics, va plus loin en proposant un schéma et une organisation des objets échangés (SIP ou DIP).

2.2.2. Structure d'un schéma SEDA

Le schéma est structuré de la manière suivante :

- **l'entête du message**, qui donne des informations sur l'échange en cours (type, identifiant, etc.) ;
- une section relative aux **supports échangés**, qu'ils correspondent à des documents d'archives sur support physique (cartons, volumes reliés, etc.) ou analogique (bandes, cassettes, etc.) ou à des documents d'archives sur support numérique (fichiers). Pour les fichiers correspondant à des documents d'archives sur support numérique, cette section permet d'exprimer les métadonnées techniques relatives à ces fichiers, qu'elles soient génériques (taille, empreinte, nom du fichier, date de dernière modification, etc.) ou spécifiques (en fonction de schémas propres à chaque catégorie technique) ;
- une section relative aux **métadonnées de gestion** applicables à l'ensemble des documents faisant l'objet de l'échange ;
- une section relative à **l'organisation intellectuelle des documents**, permettant également de donner les métadonnées descriptives et de gestion applicables à chaque document ou à chaque ensemble de document. Des identifiants permettent de faire le lien entre les documents d'archives et les supports échangés.

2.2.3. À noter...

Comme tout standard d'échange, le SEDA spécifie l'organisation des fichiers et métadonnées échangés par un système d'archivage avec ses partenaires, mais ne spécifie pas la manière dont ces fichiers et métadonnées sont gérés dans le système d'archivage lui-même.

Le schéma étant souple, ses mises en œuvre peuvent fortement différer entre les différents systèmes d'archivage, ce qui peut limiter leur capacité à s'échanger des documents d'archives. Pour résoudre cette difficulté, le standard SEDA permet de développer des profils, c'est-à-dire des déclinaisons particulières du schéma pour une communauté particulière ou un type d'objets particulier, spécifiant des restrictions par rapport aux possibilités offertes par le schéma. Des générateurs de profils comme Pastis (<https://git.cines.fr/dad/pastis-dad-externe>) ou Gaspar (<https://gaspar.mintika.fr/>) ont été développés ces dernières années et sont disponibles en ligne.

2. Standards d'échange

Le standard SEDA est largement utilisé par les solutions logicielles développées en France depuis le début des années 2010.

3. Solutions logicielles



3.1. Fonctions des solutions logicielles

Le marché de la préservation numérique est en pleine expansion depuis le début des années 2000 et a donné lieu à la conception, la réalisation, la distribution et la maintenance de nombreux outils logiciels permettant d'aider les organisations à assurer la préservation des objets numériques en général et des documents d'archives en particulier dont elles assurent la conservation.

3.1.1. Automatiser le fonctionnement des systèmes d'archivage

Des solutions logicielles ont notamment été développées pour automatiser le fonctionnement de systèmes d'archivage, soit par des éditeurs commerciaux soit par des institutions patrimoniales elles-mêmes, selon des logiques propriétaires ou open source.

Voici quelques-unes des solutions logicielles qui ont été développées :

- Fedora (1997)
- Eprints (2000)
- Dspace (2002)
- Safety Deposit Box (2003)
- Rosetta (2008)
- Archivematica (2009)
- Asal@e (2009)
- DuraCloud (2011)
- Preservica (2012)
- Vitam (2016).

3.1.2. Mettre en oeuvre des tâches spécifiques au sein des systèmes d'archivage

D'autres solutions logicielles ont été développées pour mettre en oeuvre certaines tâches soit au sein des systèmes d'archivage, soit en dehors des systèmes d'archivage :

- logiciels permettant d'auditer un système de fichiers (ex. Archifiltre) ;
- logiciels permettant de procéder à l'identification ou à la validation de formats de fichiers (cf. la section 2 de ce module., ex. DROID, VeraPDF, JHOVE, etc.) ;
- logiciels permettant de constituer des SIP conformes aux spécifications d'un système d'archivage (ex. BagIt, DocuTeam Packer, OCTAVE) ;
- logiciels permettant de procéder à la conversion de formats ;
- logiciels permettant d'émuler le fonctionnement d'anciens environnements matériels et logiciels ;
- logiciels permettant de restituer des documents d'archives sous format numérique à des utilisateurs (visualiseurs, DBPTK Database Preservation Toolkit).

3.2. Exemples de solutions logicielles utilisées pour la réalisation de systèmes d'archivage

3.2.1. Archivemática

Archivemática est une suite intégrée de logiciels ouverts (open source) éditée par la société Artefactual Systems et destinée à permettre la préservation sur le long terme de documents d'archives sur support numérique, depuis leur prise en charge jusqu'à leur communication, au moyen de microservices.

Fondée sur le modèle de référence OAIS, cette suite logicielle utilise le standard d'échange METS et prend en compte les standards de métadonnées PREMIS et Dublin Core.

Le code de ces logiciels est mis à disposition sous licence GNU Affero General Public License (A-GPL 3.0), et toute la documentation est publiée sous la licence internationale Creative Commons Attribution-ShareAlike 4.0.

Des services d'assistance technique sont également fournis par la société Artefactual Systems.



Pour en savoir plus : <https://www.archivemática.org/fr/>

3.2.2. Asal@e

Asal@e est un logiciel édité par la coopérative française Libriciel, destiné à aider administrations et collectivités territoriales à préserver sur le long terme des documents d'archives sur support numérique, depuis leur prise en charge jusqu'à leur communication.

Fondée sur le modèle de référence OAIS et compatible avec la norme NF Z42-013, la solution Asal@e utilise le standard d'échanges SEDA.

La solution Asal@e s'articule avec les autres logiciels développés et mis à disposition par la société Libriciel pour accompagner ces organisations dans le développement de l'administration électronique (parapheur électronique, télétransmission de documents, etc.), dans la préparation des documents d'archives à verser dans Asal@e (Vers@e), de la gestion du cycle de vie des documents d'archives (Ref@e) et dans la supervision des processus d'archives (Pastell).



Pour en savoir plus : <https://www.libriciel.fr/>

3.2.3. Preservica

Preservica est une société commerciale qui propose des solutions propriétaires et des services permettant de préserver sur le long terme des documents d'archives sur support numérique.

Fondée sur le modèle de référence OAIS, la solution Preservica Essentials prend en charge les standards de métadonnées Dublin Core ou MODS, tout en permettant aux utilisateurs de créer leurs propres schémas.



Pour en savoir plus : <https://preservica.com/>

3.2.4. Vitam

Vitam est une solution logicielle ouverte (open source) éditée par l'État français sous licence CECILL-2.1¹ permettant la prise en charge, la conservation et la consultation sécurisée de documents d'archives sur support numérique. La documentation associée est disponible sous licence V2.0.

Fondée sur le modèle de référence OAIS et compatible avec la norme NF Z42-013, elle utilise le standard d'échanges SEDA.

La solution logicielle Vitam est utilisable de manière mutualisée.

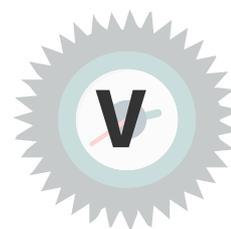
Elle s'adresse en premier lieu aux structures ayant à gérer de grosses volumétries de documents d'archives sur support numérique, avec des besoins d'accès fréquents.



Pour en savoir plus : <https://www.programmevitam.fr/>.

¹. https://github.com/ProgrammeVitam/vitam/blob/master_6.rc.x/LICENSE.md

4. Audit et certification des systèmes d'archivage



4.1. Considérations générales

Des méthodes d'audit et de certification des systèmes d'archivage adaptées à la prise en charge et à la conservation des documents d'archives sur support numérique ont été définies depuis une vingtaine d'années.



Il s'agit de procédures formelles mises en œuvre par des organisations accréditées, permettant d'attester qu'un système d'archivage respecte ou non les recommandations et exigences édictées par les normes nationales ou internationales en vigueur.

La finalité de ces procédures d'audit et de certification est de renforcer la confiance des services producteurs et des utilisateurs dans les services offerts par les systèmes d'archivage.



Des procédures d'auto-évaluation existent également, permettant aux services d'archives d'évaluer leur degré de conformité avec certaines de ces normes.

Les deux outils présentés dans la section 2, chapitre 1- Connaître l'environnement de préservation...- de ce module (NDSA Levels of Preservation et Rapid Assessment Model de Digital Preservation Coalition) peuvent être considérés comme des outils d'auto-évaluation, même s'ils ne sont pas adossés à une norme particulière.

Aujourd'hui, après une dizaine d'années de réflexion et de foisonnement, **trois procédures d'audit et de certification sont principalement utilisées dans le monde de la préservation numérique :**

- la norme allemande DIN 31644 Information et documentation – Critères pour les archives numériques fiables (2008) ;
- la norme ISO 16363:2012 – Audit et certification des entrepôts numériques fiables —, suite à la normalisation de la norme OAIS (2012) ;
- le CoreTrustSeal (2018).

4.2. Audit ISO 16363 et la certification d'entrepôts numériques de confiance

La norme d'audit ISO 16363:2012 – Audit et certification des entrepôts numériques fiables — a été élaborée suite à la publication du modèle de référence OAIS et à sa normalisation à l'ISO, en 2011, sous la référence ISO 14721:2012 – Systèmes de transfert des informations et données spatiales — Système ouvert d'archivage d'information (SOAI) — Modèle de référence.

Elle prend pour source les préconisations sur l'audit édictées par le Comité consultatif pour les systèmes de données spatiales, auteur du modèle de référence OAIS.



La norme ISO 16363 s'applique à l'ensemble de l'organisation désignée sous le terme d'entrepôt et pas seulement à l'environnement matériel et logiciel permettant la prise en charge et la conservation des objets numériques.

Les 100 paramètres constituant le socle de la procédure d'audit et de certification sont regroupés en trois domaines :

1. l'infrastructure organisationnelle : gouvernance, structure organisationnelle, effectifs, répartition des rôles et responsabilités, cadre politique, viabilité financière, gestion des contrats et licences ;
2. la gestion des objets numériques : transfert et prise en charge des objets numériques, planification de la préservation numérique, création et préservation des AIP, gestion de l'information, accès aux objets numériques ;
3. la gestion des risques liés à l'infrastructure et à la sécurité : infrastructure technique, gestion des risques, gestion des risques spécifiquement liés aux questions de sécurité.

La terminologie employée par cette norme est objectivement alignée sur celle du modèle de référence OAIS.



La procédure d'audit et de certification est relativement lourde, ce qui explique qu'une seule organisation ait obtenu une certification officielle par des auditeurs externes et agréés, l'U.S. Government Publishing Office. Plus nombreuses sont les organisations à utiliser la norme ISO 16363 comme un outil d'auto-évaluation interne.

4.3. CoreTrustSeal

Le CoreTrustSeal résulte de l'harmonisation des pratiques de deux précédentes entreprises de certification, le Data Seal of Approval (DSA) et le World Data System (WDS) de l'International Science Council, principalement orientées vers la certification d'entrepôts voués à la préservation numérique des données de la recherche. Initié en 2013, le projet a abouti en 2018 à la définition d'un noyau d'exigences communes permettant de certifier des entrepôts numériques prenant en charge et conservant sur le long terme des données numériques comme étant dignes de confiance. Le CoreTrustSeal est géré par une fondation à but non lucratif de droit néerlandais.

La dernière version des exigences émises par le CoreTrustSeal pour la période 2023-2025, couvre les aspects suivants :

- Aspects organisationnels : périmètre et objectifs du système, gestion des droits, continuité de service, aspects juridiques et éthiques, gouvernance et ressources, expertise et conseils ;
- Gestion des objets numériques : provenance et gestion de l'authenticité, transfert et prise en charge (y compris l'évaluation), planification de la préservation, assurance qualité, processus de travail, recherche et identification, réutilisation ;
- Aspects techniques et de sécurité : stockage et gestion de l'intégrité, infrastructure technique, sécurité.



À ce jour, 130 entrepôts ont fait l'objet d'une certification par le CoreTrustSeal, mais ils sont peu nombreux dans le monde francophone.

4.4. Méthodologie pour la mise en place d'une opération d'audit et de certification

4.4.1. Les questions à se poser

Avant d'entamer une opération d'audit et de certification d'un système d'archivage adapté à la prise en charge et à la conservation de documents d'archives sur support numérique, il convient de se poser un certain nombre de questions :

- **Quel est l'objectif poursuivi en lançant une opération d'audit et de certification ?** S'agit-il d'obtenir une certification d'une autorité externe ? De disposer simplement d'un regard extérieur dans le cadre d'un processus d'amélioration continue ? Une opération d'audit et de certification peut se révéler longue et coûteuse en ressources tant humaines que financières ;
- **Quelles sont les ressources disponibles pour mettre en œuvre ce processus d'audit et de certification ?**
- **Quel type d'audit et de certification est le plus adapté aux objets numériques pris en charge par le système d'archivage ?** Alors que le CoreTrustSeal est plus approprié aux entrepôts de données, l'ISO 16363 a un périmètre plus large. Une certification associée à la norme française NF Z42-013 convient de son côté mieux aux systèmes d'archivage d'archives à valeur probante, mais moins aux systèmes d'archivage de documents d'archives à valeur informationnelle ou à valeur patrimoniale ou de recherche ;
- **Quel cadre d'audit et de certification est le plus adapté au système d'archivage ?** Un cadre national dans la mesure où le service d'archives qui l'administre a une portée nationale ? Un cadre international dans la mesure où le service prend en charge des documents d'archives produits dans plusieurs pays et soumis à différentes contraintes en matière juridique – ce qui est le cas d'entreprises multinationales ?

4.4.2. Se préparer en amont...

Il faut également se rappeler que tout processus d'audit et de certification constitue un projet dont on doit :

- approuver le périmètre et les objectifs ;
- définir le calendrier et les échéances ;
- identifier les acteurs et parties prenantes, internes comme externes ;
- définir les modalités d'organisation et de gouvernance ;
- définir les modalités de communication en interne comme en externe ;
- anticiper les phases de mise en œuvre des recommandations, y compris les aspects de conduite de changement induits par celles-ci.

5. Le système d'archivage au quotidien



Au terme de ce chapitre, il est important d'insister sur le fait que la mise en place d'un système d'archivage adapté à la prise en charge et à la conservation d'archives sur support numérique est une chose, mais qu'administrer ce système au quotidien en est une autre.



Fondamental

Un système mis en place sans gestion au quotidien et sans financement pérenne ne sert à rien et constitue une dépense inutile.

Trop souvent, tous les moyens et ressources des services d'archives sont focalisés sur l'acquisition et la mise en œuvre du système, sans se donner les moyens de l'administrer et de le faire vivre au quotidien.

Tout service d'archives qui veut faire vivre sur le long terme un système d'archivage adapté aux documents d'archives sur support numérique doit impérativement prendre en compte les points suivants :

- mettre en place et installer un système d'archivage mobilisent **des ressources et compétences variées qu'il s'agit de préserver**. Trop souvent, les personnes qui mettent en œuvre le système acquièrent, pendant la phase de mise en œuvre, des connaissances et des compétences qu'ils emportent avec eux à la fin du projet. Capitaliser les connaissances et compétences acquises et les transmettre aux personnes chargées d'administrer le système au quotidien est essentiel, que ce soit lors de la mise en œuvre du système ou au fil de l'arrivée des nouveaux utilisateurs ;
- **disposer d'une bonne documentation du système à sa mise en production est essentiel**, mais cette documentation doit aussi être mise à jour au fur et à mesure des évolutions opérées. Trop souvent, seule la documentation initiale est conservée, et il est difficile, plusieurs années après, de comprendre le fonctionnement du système avec une documentation obsolète ;
- **définir les rôles et responsabilités dans l'administration et l'utilisation au quotidien du système** est indispensable. L'administration du système lui-même (octroi et retrait des droits de lecture, d'écriture, de modification ou de suppression sur tout ou partie des documents, mise à jour des données de référence, etc.) est trop souvent négligée ;
- **définir les modalités d'identification et de résolution des incidents** susceptibles de survenir lors de la mise en œuvre du système, qu'ils résultent de dysfonctionnements humains, matériels ou logiciels ;
- tout système d'archivage repose sur l'utilisation de référentiels (ex. registre des formats PRONOM) qui évoluent dans le temps. Il convient de **veiller à la mise à jour de ces référentiels**, voire de rejouer certaines opérations s'appuyant sur ceux-ci, dans la mesure où les résultats qu'ils fournissent peuvent avoir été affinés suite à leur mise à jour (ex. identification d'un format auparavant non permise d'après le registre) ;
- un système d'archivage est un système complexe, même s'il est manuel, et, comme tout système, il est soumis à **une obsolescence partielle ou totale**. Matériels et logiciels doivent faire l'objet d'une veille permanente sur leur évolution, d'une maintenance et d'une amélioration continue régulières, voire d'un remplacement total ou partiel régulier ;

- dans la mesure où les modalités de création des documents d'archives sur support numérique et les attentes des utilisateurs évoluent régulièrement, il est évident que tous les besoins n'auront pas été couverts au moment de la conception et de la mise en production du système. **Il faut conserver une marge de manœuvre suffisante pour prendre en compte de nouveaux besoins**, ce qui implique de disposer des ressources suffisantes pour réaliser, tester et mettre en production ces évolutions ;
- les menaces en matière de sécurité évoluent elles aussi très rapidement. Il importe donc de **procéder régulièrement à des exercices et à des audits de sécurité**, pour garantir qu'un niveau d'alerte suffisant est maintenu chez tous les acteurs intervenant dans l'utilisation ou l'administration du système ;
- dans la mesure où le système s'appuie sur des **fournisseurs externes** (par exemple en matière d'hébergement), il convient de vérifier régulièrement la capacité du service à se substituer à eux ou à transférer le service à d'autres fournisseurs, par exemple au terme de la durée d'un contrat ;
- dans le cas où le système d'archivage est mis à la disposition d'un nombre croissant de clients, **il convient de vérifier que la venue de tout nouveau client ne vienne pas perturber le fonctionnement du service** offert aux précédents clients (dégradation des opérations, trop grande affectation de ressources au bénéfice de ce nouveau client au détriment des ressources prévues pour les anciens clients) ;
- enfin, assurer la gestion au quotidien d'un système d'archivage adapté aux documents d'archives sur support numérique nécessite **un financement adapté et pérenne**. Les coûts d'exploitation d'un système sur un temps long et/ou la mise en place de nouveaux projets d'archivage portant sur des types de documents nouveaux, sont trop souvent minimisés.