

# Introduction à l'archivage électronique

## Sommaire

- 1 Introduction à l'archivage électronique à valeur probante
  - ◆ 1.1 Les archives et les archives numériques
    - ◇ 1.1.1 Qu'est-ce qu'une archive ?
    - ◇ 1.1.2 Le cycle de vie des archives
    - ◇ 1.1.3 Le document numérique
    - ◇ 1.1.4 Ce que n'est pas un SAE
      - 1.1.4.1 Le SAE n'est pas une GED
      - 1.1.4.2 Le SAE n'est pas la sauvegarde
      - 1.1.4.3 Le SAE n'est pas un coffre-fort électronique
    - ◇ 1.1.5 Les systèmes de gestion
  - ◆ 1.2 Références normatives
    - ◇ 1.2.1 OAIS - Open Archival Information System
      - 1.2.1.1 Modèle fonctionnel
      - 1.2.1.2 Modèle d'information
    - ◇ 1.2.2 NF Z42-013
      - 1.2.2.1 Politique d'archivage
      - 1.2.2.2 Exigences
    - ◇ 1.2.3 MoReq
    - ◇ 1.2.4 SEDA et MEDONA
      - 1.2.4.1 Cas d'usage
      - 1.2.4.2 Séquences des échanges
      - 1.2.4.3 Modèle de données
  - ◆ 1.3 Spécifications du système d'archivage
    - ◇ 1.3.1 Le modèle de données
      - 1.3.1.1 Données
      - 1.3.1.2 Métadonnées
      - 1.3.1.3 Acteur producteur et classement
    - ◇ 1.3.2 Les fonctions
      - 1.3.2.1 L'entrée
      - 1.3.2.2 Le classement
      - 1.3.2.3 Le stockage
      - 1.3.2.4 Les métadonnées
      - 1.3.2.5 L'accès
      - 1.3.2.6 Le management
      - 1.3.2.7 La pérennisation
      - 1.3.2.8 Les fonctions complémentaires
  - ◆ 1.4 Annexes
    - ◇ 1.4.1 Références
    - ◇ 1.4.2 Bibliographie
    - ◇ 1.4.3 Liens externes

## Introduction à l'archivage électronique à valeur probante

Il y a quelques années encore, le papier était le seul support d'enregistrement, de diffusion et de conservation de l'information reconnu comme preuve. L'archivage était une étape finale dans le cycle de vie des documents, mise en ?uvre lorsque les traitements métiers avaient été réalisés, en grande partie à cause de la nature physique du support original à conserver, indissociable du contenu. Le développement des technologies de l'information et de la communication a totalement révolutionné les modes de production, de diffusion et de conservation de l'information numérique. Avec l'évolution de la législation, le document numérique constitue aujourd'hui une preuve, et représente une part de plus en plus importante des données à forte valeur juridique, stratégique et patrimoniale des administrations publiques et des organisations privées.

L'archivage électronique commence dès la création des documents et permet à l'organisation productrice de l'information de les exploiter pour toute la durée de leur cycle de vie. Elle met en ?uvre un ensemble d'actions, d'outils et de méthodes pour collecter, identifier, classer, conserver, communiquer ? et finalement détruire le cas échéant ? les contenus numériques. C'est un processus qui fait appel à de multiples domaines de la gouvernance informatique :

- la gestion électronique de documents pour la classification, l'indexation, la recherche multicritères, sémantique, plein-texte, etc
- la sécurité des systèmes, l'interopérabilité, la preuve, la journalisation, la veille réglementaire
- le stockage sur les infrastructures (matérielles et logicielles)
- les outils de conservation à long terme du numérique, la veille technologique

Cet article a pour objectif de présenter ce domaine complexe qui comporte à la fois des aspects métier, technologiques et réglementaires.

## Les archives et les archives numériques

---

### Qu'est-ce qu'une archive ?

La loi n° 79-18 du 3 janvier 1979 définit les archives comme

**« l'ensemble des documents, quels que soient leur date, leur forme et leur support matériel, produits ou reçus par toute personne physique ou morale, et par tout service ou organisme public ou privé, dans l'exercice de leur activité. »**

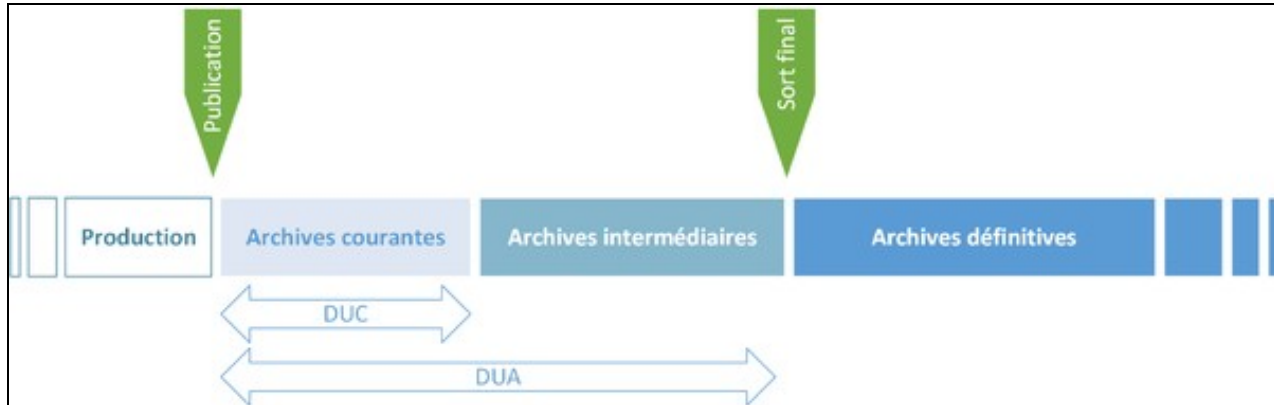
On voit donc à quel point la notion d'archives concerne tout un chacun quelle que soit la sphère considérée. La question n'est pas tant de définir ce qu'est une archive que de savoir pourquoi, comment et combien de temps la conserver. Les raisons d'archiver sont nombreuses :

- **Respecter des obligations** : les textes législatifs définissent le périmètre documentaire qui doit légalement être conservé à des fins de preuve, fournit les durées de conservation associées ainsi que les exigences spécifiques au domaine considéré. Les archives publiques sont principalement régies par le code du patrimoine; pour les archives privées, chaque code fournit les informations sur la conservation des documents produits par les activités qu'il couvre (code du commerce, code du travail, code général des impôts).
- **Protéger ses droits** : les documents archivés peuvent être ressortis dans le cadre de litiges et le règlement de contentieux. Les tribunaux considèrent désormais les documents électroniques comme des preuves recevables, à condition que leur conservation respecte les normes et réglementations !

- **Maîtriser les risques et préserver la pérennité de l'activité** : l'archive est sécurisée et pérennisée, elle garantit l'accès aux contenus nécessaires à la continuité de l'activité en cas de sinistre.
- **Conserver la mémoire** : lorsque les documents revêtent un aspect historique ou patrimonial, l'archivage est le garant de la mémoire de l'organisation ou de la personne.
- **Faciliter l'exploitation des documents** : un archivage efficace et opérationnel permet de rendre les documents disponibles et exploitables dans les meilleures conditions.

## Le cycle de vie des archives

Le schéma ci-dessous représente les étapes du cycle de vie de l'archive :



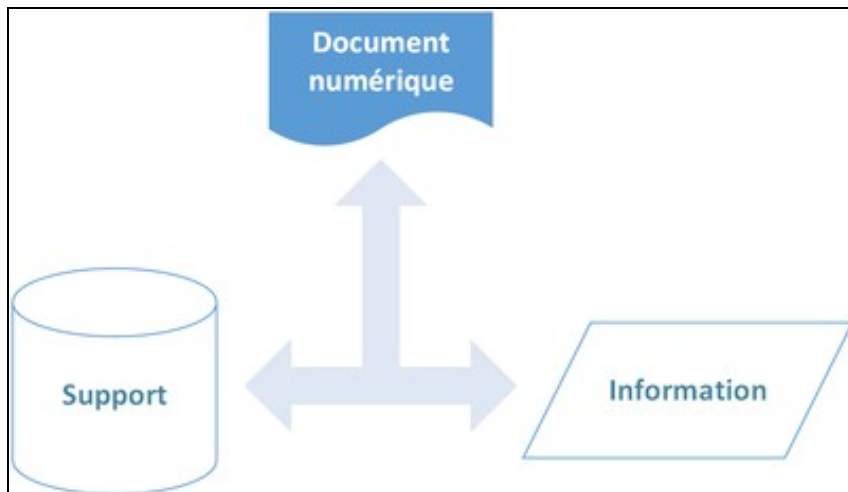
Le document constitue une archive dès sa production ou sa réception par la personne ou l'organisme dans le cadre de son activité. Dès lors, son cycle de vie peut être découpé en trois étapes :

1/L'archive est dite "**courante**" lorsqu'elle est utilisée par les différents services pour les activités liées au domaine. Elle peut circuler entre les acteurs pour qu'ils en prennent connaissance et utilisent l'information dans des processus métier. Par exemple, la facture d'achat reçue d'un fournisseur va circuler dans les différents services de l'organisation entre le moment de sa réception et la fin des traitements avec le règlement du fournisseur, avec des étapes comme la validation par le service acheteur, le rapprochement avec les achats, le traitement en comptabilité. La période couverte est appelée **durée d'utilisation courante** ou DUC, qui prend fin dès lors que le document n'est plus nécessaire aux processus qu'il a déclenché ou qu'il concerne.

2/L'archive est dite "**intermédiaire**" à la fin de son utilisation courante, lorsqu'elle est effectivement versée dans l'archive (papier ou électronique), classée et sécurisée, pour être conservée dans les conditions et pour la durée requise. Sauf dérogation, son accès est restreint au service qui a produit le document et à ceux qui assurent sa gestion. Les conditions et la durée de conservation sont définies par la législation en vigueur ou les exigences ou recommandations liées à la typologie documentaire. Par exemple, la facture du fournisseur doit être conservée durant 10 ans après son édition. La période couverte commence (comme la DUC) à la production du document et est appelée **durée d'utilisation administrative** ou DUA.

3/A l'issue de la DUA, l'archive est soumise à une règle de **sort final** qui détermine si elle peut être détruite ou si elle doit être conservée parce que l'information revêt un intérêt historique ou patrimonial. Si l'archive est conservée, elle entre dans la troisième et dernière étape de son cycle de vie et devient une "**archive définitive**". La conservation sans limitation de durée concerne plus souvent les archives publiques, les documents et dossiers sont alors transférés vers des services d'archive tels que les Archives Départementales, les Archives Nationales ou la Bibliothèque Nationale.

## Le document numérique



Le document numérique dissocie le support et l'information contenue  
Un document est défini comme **un support et une information enregistrés de manière persistante**.

Lorsqu'il s'agit d'archives physiques, le support papier ou microfiche assure la pérennité de l'information qui y est inscrite. Dans toute l'histoire de l'humanité, l'activité humaine a toujours produit des documents physiques, pour lesquels le support et l'information étaient conservés de pair, depuis les premiers dessins dans les grottes jusqu'aux documents imprimés en passant par les tablettes de cire, la toile et le parchemin.

**Mais il n'en va pas de même avec le document numérique, qui dissocie le support et l'information contenue.**

De nos jours le support peut être un disque dur magnétique, un stockage mémoire (SSD), un disque optique, une bande magnétique, et demain sera peut-être un bloc de quartz ou une molécule d'ADN! L'information, elle, est binaire. C'est une série de bits (de zéro et de un) organisés selon une

structure définie par les spécifications d'un format pour permettre de représenter le contenu.

L'activité humaine produit de plus en plus d'information sous forme numérique, principalement selon 3 grands axes :

- La numérisation : Ce sont des originaux papiers, le plus souvent reçus, qui sont numérisés (par un scanner, demain par cliché photographique d'un terminal mobile ?)
- La dématérialisation : De plus en plus de processus métiers s'appuient sur des logiciels, qui réduisent l'utilisation de supports physiques au profit de fichiers informatiques pour la production de leurs informations.
- Les activités nativement numériques : De plus en plus d'activités sont basées sur la réception ou la production de documents multimédia, de courriels, pages web, etc. Elles ne savent plus être ni réceptrices ni génératrices de documents physiques.

On produit de plus en plus d'information numérique, mais, paradoxalement, la nature numérique des documents est inféodée à des environnements de conservation de moins en moins pérennes: La durée de vie d'un disque dur ou d'un DVD n'excède par une dizaine d'année même dans des conditions optimales d'utilisation, un format de fichier est rendu obsolète par les évolutions de ses spécifications et l'impossibilité d'exécuter les logiciels associés dans les nouveaux environnements système. L'exploitation de l'information numérique fait intervenir à la fois des systèmes matériels (centraux/périphériques) et des systèmes logiciels (OS/drivers/afficheurs) très divers. Tous ces systèmes connaissent des évolutions de plus en plus rapides et à des rythmes différents, ce qui au final entraîne un risque d'obsolescence technologique, le support et le contenu étant soumis à des cycles de vie différents.

## Ce que n'est pas un SAE

L'archivage électronique est souvent confondu avec d'autres systèmes informatiques, dont il intègre effectivement certains concepts mais qui ont un périmètre fonctionnel différent, parfois partiellement commun, parfois inclus à celui du SAE.

### Le SAE n'est pas une GED

Le but de la GED est de faciliter l'acquisition, le classement et l'exploitation des documents pour optimiser les flux de travail centrés sur les documents reçus ou produits. Elle ne s'intéresse pas à leur conservation dans le temps ou selon des exigences de sécurité et d'intégrité, mais uniquement à leur exploitation dans le cadre de l'activité qu'ils concernent.

### Le SAE n'est pas la sauvegarde

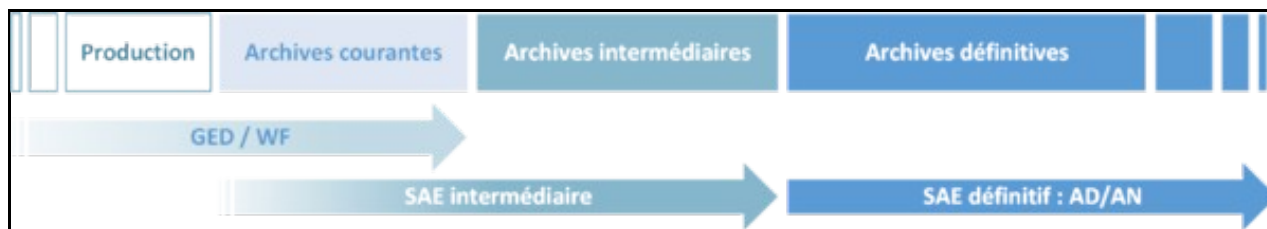
La sauvegarde duplique les données à un instant T pour éviter la perte et assurer la continuité de l'activité en cas d'incident technique sur les infrastructures matérielles ou logicielles. Elle ne fait pas la différence entre les documents d'archive et les autres fichiers informatiques, notamment en terme de cycle de vie, car les documents peuvent être en cours de création (non finalisés) et donc ne pas avoir été produits.

### Le SAE n'est pas un coffre-fort électronique

Un coffre-fort électronique est un dispositif technique qui vise à sécuriser les objets numériques déposés pour garantir leur intégrité, par des mécanismes d'identification des accédants et de traçabilité. Il ne s'intéresse pas au cycle de vie de ces documents ni à leur classement. C'est une sorte de sauvegarde en ligne de documents d'archive qui est une composante possible du SAE.

## Les systèmes de gestion

Le schéma ci-dessous montre les systèmes logiciels utilisables au cours du cycle de vie de l'archive :



- **Les logiciels de GED** peuvent inclure les fonctionnalités en amont de la production du document (la composition, la collaboration, la validation, la chaîne d'édition documentaire) ainsi celles qui permettent l'exploitation de l'information contenue (workflow métier, production de rapports, constitution de dossiers de traitement) durant la phase d'utilisation courante.
- **Les systèmes d'archivage intermédiaire** prennent en charge l'archivage sécurisé des documents pour la durée de leur utilisation administrative. Cette phase recouvre tout ou partie de leur durée d'utilisation courante, il est donc possible que le document soit à la fois en circulation dans une GED pour son utilisation courante et dans le SAE pour sa pérennisation. La GED peut alors par exemple s'appuyer sur des services d'accès aux contenus du système d'archivage, ne pas conserver elle-même les documents numériques mais seulement les métadonnées nécessaires aux traitements métier.
- **Les systèmes d'archivage définitifs** sont dédiés à la conservation sans limitation de durée des documents d'archives dont le sort final est la conservation. Ils reçoivent les archives transférées par les systèmes d'archivage intermédiaire.

## Références normatives

Il existe quatre grandes familles de normes et standards qui fournissent un ensemble de recommandations et d'exigences relatives à la conservation des objets numériques. L'idée n'est pas ici de détailler le contenu des textes mais de dégager les points-clés pour en comprendre la portée.

### OAIS - Open Archival Information System

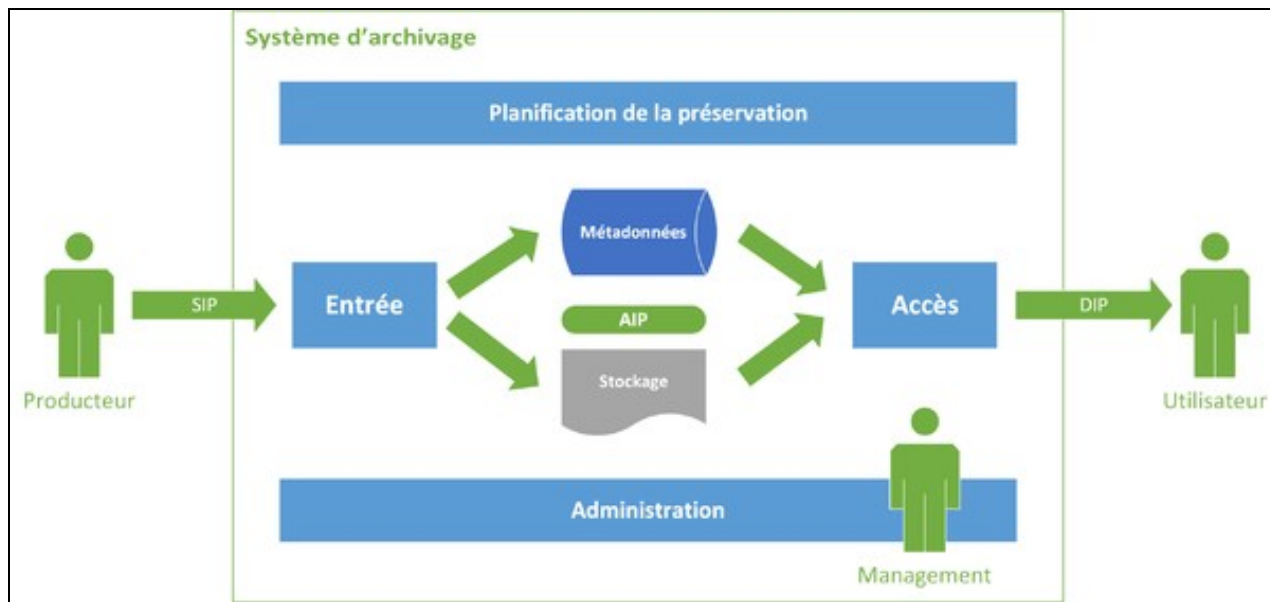
Il s'agit d'un modèle pour l'archivage. Défini au départ par le Consultative Committee for Space Data System (CCSDS), l'organisme international de normalisation de l'aérospatiale, il a fait l'objet d'un travail de normalisation qui a produit en 2003 la norme ISO 14721:2003.

La norme OAIS adresse les besoins de l'archivage des données numériques selon trois axes :

- **Le pourquoi** : elle décrit les exigences intrinsèques de l'archivage comme étant la pérennisation des contenus et l'interopérabilité avec les autres systèmes d'information
- **Le quoi et le qui** : elle définit les grands concepts de l'archivage et répartit les responsabilités vers trois acteurs que sont le producteur, l'accédant et le management.
- **Le comment** : elle fournit un modèle fonctionnel et un modèle d'information qui posent les bases jugées nécessaires à la réalisation d'un système d'archivage.

## Modèle fonctionnel

Le schéma ci-dessous représente d'une manière très générique le modèle fonctionnel de l'OAIS :



Il représente 6 fonctions principales, 3 acteurs et 3 entités.

Les 6 fonctions principales du système d'archivage sont:

- **L'entrée** : Le système d'archivage permet au producteur de déposer les archives numériques.
- **La gestion du stockage** : Elle assure les opérations d'enregistrement, de lecture et de destruction des objets numériques sur les supports de stockage.
- **La gestion des métadonnées** : Elle assure les opérations de mise à jour des référentiels du système avec les métadonnées d'archives et associées.
- **L'accès** : Le système d'archivage restitue l'information numérique à l'utilisateur au travers de copies numériques, d'affichage ou d'impression.
- **L'administration** : C'est un ensemble de fonctions qui permettent de piloter le système pour assurer la conservation, la sécurité, la traçabilité, l'exploitation...
- **La planification de la préservation** : Le système d'archivage permet de constituer une base de connaissance et de mettre en ?uvre les mécanismes pour assurer la pérennité des archives

Les 3 acteurs représentés et leurs responsabilités sont :

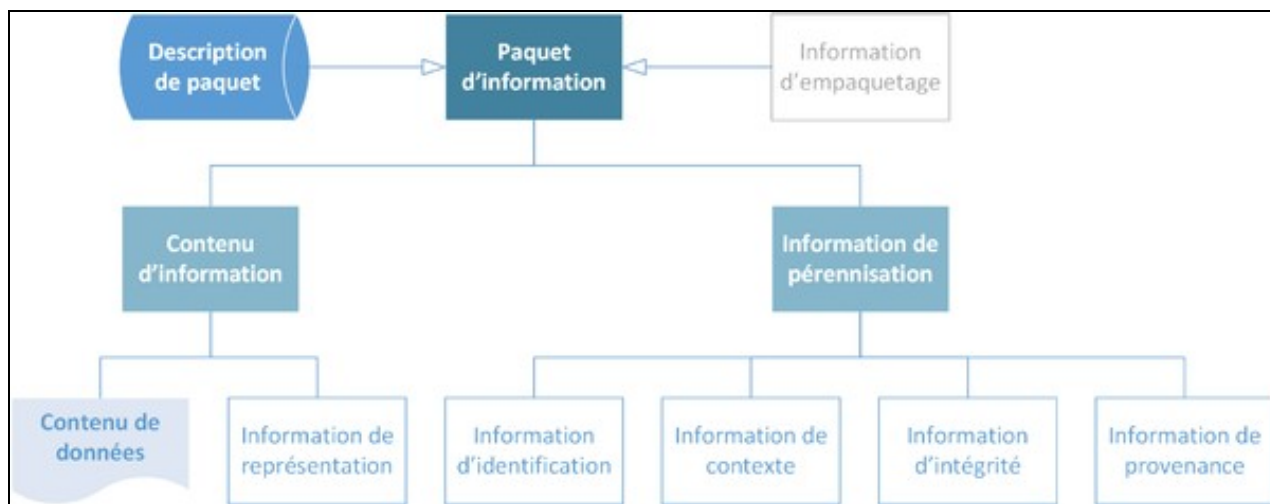
- **Le producteur** de l'archive accède aux fonctions d'entrée (ou de dépôt, on parle aussi de déposant), par lesquelles il transfère les données d'archives au système.
- **L'utilisateur** accède aux fonctions de sortie (ou d'accès, on parle aussi d'accédant ou de demandeur), par lesquelles le système lui restitue les données d'archives demandées.
- **Le management** assure l'administration du système et la planification de la pérennisation.

Les 3 entités sont :

- **SIP** (Submission Information Package) : c'est le paquet d'information versé par le producteur au système d'archivage via la fonction d'entrée. Il sert de source de données pour que le système puisse extraire l'information de description qui sera inscrite dans le référentiel et décrira le paquet archivé.
- **AIP** (Archival Information Package) : c'est le paquet d'information archivé par le système d'archivage via la fonction de stockage, décrit par les métadonnées et objet de la pérennisation.
- **DIP** (Dissemination Information Package) : c'est le paquet d'information diffusé à l'utilisateur via la fonction d'accès. Il est constitué du contenu d'information et d'une partie du paquet archivé et des métadonnées, selon le nombre et la nature de l'information à transmettre.

## Modèle d'information

Le schéma ci-dessous représente le modèle de données d'un paquet OAIS:



Pour la norme OAIS, le paquet archivé doit être auto-portant, c'est-à-dire qu'il doit être exploitable en-dehors du système d'archivage sans le logiciel ou les métadonnées descriptives contenues dans le référentiel. Il ne suffit donc pas de stocker le contenu numérique sur les supports, il faut l'accompagner de toute l'information nécessaire à sa restitution et à sa compréhension.

Le paquet d'information est donc constitué de deux parties principales, elles-mêmes composées de sous-ensembles :

- **Le contenu d'information** est la partie qui permet de restituer l'information à l'utilisateur. Elle se compose de deux sous-ensembles :
  - ◆ Le contenu de données (ou CDO pour Content Data Object) est le flux binaire, le document numérique lui-même.
  - ◆ L'information de représentation décrit le mode de structuration de l'information contenue, notamment le format d'encodage, l'organisation des éventuelles parties (mime), les caractéristiques du flux media, etc. *Notez que certains formats permettent nativement d'intégrer des métadonnées, et notamment des métadonnées de représentation. Cela dit, elle ne pourront être accessibles que si le contenu binaire est restitué... il faut donc dans tous les cas extraire cette information et la rendre accessible directement.*
- **L'information de pérennisation** est la partie qui permet de comprendre et d'exploiter le contenu d'information. Elle se compose de quatre sous-ensembles :
  - ◆ L'information d'identification fournit, comme son nom l'indique, les moyen d'identifier l'archive dans l'écosystème.
  - ◆ L'information de contexte place l'information par rapport aux autres archives du système ou d'autres systèmes.
  - ◆ L'information d'intégrité fournit les moyen de contrôler que le contenu numérique est conforme et qu'il n'a pas été altéré.
  - ◆ L'information de provenance décrit l'origine de l'archive et l'historique de sa conservation.

Le paquet d'information est accompagné de

- **l'information d'empaquetage** qui décrit l'organisation des données à l'intérieur du paquet, et de
- **l'information de description** qui alimente le référentiel du système d'archivage et permet au management d'assurer la conservation sans avoir à accéder au paquet.

## NF Z42-013

C'est LA norme de référence de l'AFNOR. Son titre actuel est "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". Il s'agit donc d'un recueil d'exigences qui régissent le fonctionnement d'un système d'archivage de données numériques, notamment à destination des opérateurs de tiers-archivage.

La norme a été publiée en 1999 et a fait l'objet d'une révision en 2009, traduite en anglais, internationalisée et publiée en 2012 sous la référence ISO 14641-1. Depuis 2010 elle s'accompagne d'un guide d'application sous la référence AFNOR GA Z42-019 et permet une certification à la marque AFNOR NF461 intitulée "Système d'Archivage Electronique".

### Politique d'archivage

L'exigence principale de la norme est la définition par l'opérateur d'archivage d'une politique d'archivage qui "définit les exigences en termes juridiques, fonctionnels, opérationnels, techniques et de sécurité". La politique d'archivage est décrite au niveau fonctionnel, indépendamment des implémentations opérationnelle et technique. Elle classe les exigences en quatre familles :

- les prestations fournis par l'opérateur : les niveaux de service, les modalités de fonctionnement, les quotas en termes de volumétrie et de fréquence des dépôts, etc.
- les obligations, principalement des service d'archives mais aussi des autres acteurs (déposants et accédants) afin d'assurer le bon fonctionnement de l'archive
- les fonctionnalités décrivent l'organisation du travail et les flux d'information entre les acteurs
- les principes de sécurité déclarent l'organisation de la sécurité, la mise en œuvre de procédures, les aspects techniques.

Surtout, cette vision démontre que le système d'archivage ne se limite pas à la mise en œuvre d'une plateforme technique mais englobe des considérations beaucoup plus larges sur les rôles et responsabilités et l'organisation de l'activité d'archivage, incluant les aspects juridiques, organisationnels, techniques, les locaux, les assurances, les procédures, etc.

### Exigences

La norme répartir les exigences du système d'archivage en quatre grandes familles :

- **Intégrité** : Le système doit assurer la protection des contenus archivés contre toute altération fortuite ou malveillante, ou à défaut permettre de détecter les substitutions ou modifications, par des mécanismes d'empreinte numérique, de journalisation, de contrôle, de sécurisation du stockage, etc.
- **Pérennité** : Le système doit assurer la pérennité de l'accès aux contenu d'information par l'utilisation de formats normalisés ou standardisés utilisables librement, leur description par des métadonnées, des procédures de migrations de support, de conversions de format, etc.
- **Sécurité** : Le système doit assurer la sécurité de l'archive par l'identification des personnes et des processus qui accèdent aux données, la journalisation des événements ayant trait à la sécurité, des copies de sécurité des données et métadonnées.
- **Traçabilité** : Le système d'archivage doit tracer l'ensemble des événements applicatifs, tenir un journal des événements du cycle de vie de l'archive, et produire des attestations électroniques de l'activité de l'opérateur, conservés dans les mêmes conditions que les archives qu'ils

concernent.

## MoReq

Il s'agit d'une norme européenne intitulée aujourd'hui "Modular Requirements for Records Systems" (qui n'est pas le titre d'origine). Elle fournit une approche opérationnelle de la norme ISO 15489 Records Management, qui constitue un guide pour l'organisation et la gestion des documents d'archives.

La norme originelle **MoReq** a été publiée en 2001 par la Commission européenne et le DLM Forum, puis révisée en 2008 sous l'appellation **MoReq2**; jugée trop complexe, elle a de nouveau été révisée et a donné **MoReq2010** publiée en 2011.

Dans sa version actuelle, elle fournit

- un recueil d'exigences fonctionnelles et non fonctionnelles pour les systèmes d'archivage électronique,
- un schéma XML de métadonnées pour les archives
- une architecture de services à implémenter dans les systèmes d'archivage électronique.

MoReq2010 définit les responsabilités de 9 services principaux :

- utilisateur et groupe,
- modélisation des rôles,
- classement,
- enregistrement,
- modélisation des métadonnées,
- échéancier de destruction,
- recherche,
- exportation.

Elle y adjoint 4 modules complémentaires :

- Interface utilisateur
- API
- Classement hiérarchique
- Composants électroniques

## SEDA et MEDONA

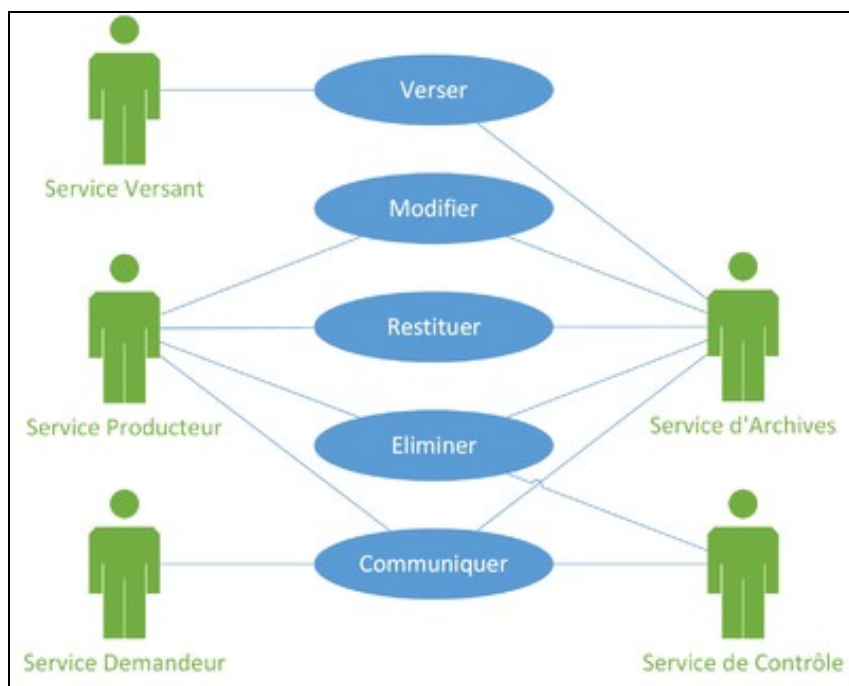
Ce sont des modèles pour les transactions d'échange entre les acteurs du SAE.

Le SEDA est le "Standard d'Echange de Données pour l'Archivage" relatif aux données d'archives publiques. Il a été créé en 2006 par la Direction des Archives de France et la Direction Générale pour la Modernisation de L'Etat, et révisé en 2010 (v0.2), en 2012 (v1.0) et plus récemment en 2015 (v2) pour assurer sa compatibilité avec la norme MEDONA. Elle comporte à la fois le modèle fonctionnel et un modèle de données pour les échanges de données d'archive de la sphère publique : description, acteurs, règles de gestion, référentiels métier et techniques.

La norme AFNOR NF Z44-022, intitulée "Modèle d'Echange de DONnées pour l'Archivage" ou MEDONA a été publiée en 2014. Elle est basée sur le SEDA, dont elle reprend le modèle fonctionnel, elle s'ouvre à tous les standards de description des archives et à tous les référentiels.

### Cas d'usage

SEDA et MEDONA partagent donc un même modèle fonctionnel qui décrit 5 cas d'usage principaux et 5 acteurs:



Les acteurs sont les suivants :

- **Le Service d'Archive** assure la gestion des archives dans le respect des conditions légales, réglementaires ou contractuelles. Il peut s'agir d'un prestataire de service de tiers-archivage.
- **Le Service Versant** transfère les archives à un Service d'Archives lors d'une opération de versement ou de transfert.

- **Le Service Producteur** a produit les archives, c'est-à-dire qui les a créées ou reçues dans le cadre de son activité.
- **Service de Contrôle** est celui qui, le cas échéant, autorise ou non la Communication, l'Élimination ou la Modification des archives en exerçant un contrôle scientifique et technique.
- **Le Service Demandeur** est toute personne physique ou morale qui souhaite consulter les informations conservées par le Service d'Archives dans le respect des conditions légales, réglementaires ou contractuelles en vigueur.

Les cas d'usage sont les suivants :

- **Verser** : C'est l'action par un Service Versant de transmettre des données d'archives à un Service d'Archives en vue de lui en confier la conservation.
- **Communiquer** : C'est l'action par un Service d'Archives de transmettre des données à un Service Demandeur, avec l'autorisation, le cas échéant, du Service Producteur et du Service de Contrôle compétent.
- **Modifier** : C'est l'action par un Service d'Archives de notifier un Service Producteur des modifications apportées sur les informations transférées. Ces modifications peuvent être nécessaires afin d'assurer une bonne conservation des informations (par exemple conversion de format ou ajout, correction, mise à jour des métadonnées).
- **Restituer** : C'est l'action par un Service d'Archives de transmettre des archives à un Service Producteur en vue de lui rendre la responsabilité de leur conservation.
- **Éliminer** : C'est l'action par un Service d'Archives de supprimer des informations d'archive dont il avait la responsabilité de la conservation. Le Service d'Archive notifie le Service Producteur de la destruction. L'élimination peut être précédée, le cas échéant, d'une Demande d'autorisation d'élimination au Service Producteur et d'une demande d'autorisation au Service de Contrôle.

## Séquences des échanges

Ces cas d'usage sont réalisés au travers de transactions d'échange qui se composent de séquences d'envoi et de réception de messages de requête, de réponse, de notification et d'accusé de réception.

Séquence	Message à utiliser
<b>Transférer</b>	
Demande de transfert	ArchiveTransferRequest
Accusé de réception de demande de transfert	Acknowledgement
Réponse à la demande de transfert	ArchiveTransferRequestReply
Accusé de réception de la réponse à la demande de transfert	Acknowledgement
Transfert d'archives	ArchiveTransfer
Accusé de réception de transfert	Acknowledgement
Réponse au transfert	ArchiveTransferReply
Accusé de réception de la réponse au transfert	Acknowledgement
<b>Communiquer</b>	
Demande de communication	ArchiveDeliveryRequest
Accusé de réception de demande de communication	Acknowledgement
Réponse à la demande de communication	ArchiveDeliveryRequestReply
Accusé de réception de la réponse à la demande de communication	Acknowledgement
<b>Modifier</b>	
Notification de modification	ArchiveModificationNotification
Accusé de réception de notification de modification	Acknowledgement
<b>Éliminer</b>	
Notification d'élimination	ArchiveDestructionNotification
Accusé réception de la notification d'élimination	Acknowledgement
<b>Restituer</b>	
Demande de restitution	ArchiveRestitutionRequest
Accusé de réception de la demande	Acknowledgement
Réponse à la demande de restitution	ArchiveRestitutionRequestReply
Accusé de réception de la réponse à la demande de restitution	Acknowledgement
<b>Demander l'autorisation du Service producteur</b>	
Demande d'autorisation au service producteur	AuthorizationOriginatingAgencyRequest
Accusé de réception de demande d'autorisation	Acknowledgement
Réponse à la demande d'autorisation au service producteur	AuthorizationOriginatingAgencyRequestReply
Accusé de réception de réponse à la demande d'autorisation	Acknowledgement
<b>Demander l'autorisation du Service de contrôle</b>	
Demande d'autorisation au service de contrôle	AuthorizationControlAuthorityRequest
Accusé de réception de demande d'autorisation	Acknowledgement
Réponse à la demande d'autorisation au service de contrôle	AuthorizationControlAuthorityRequestReply
Accusé de réception de réponse à la demande d'autorisation	Acknowledgement

## Modèle de données

La norme et le standard fournissent un modèle de données pour les messages échangés.

Le modèle comporte la structure des différents messages avec pour chacun globalement le schéma suivant :

```
Type de message
|_ identifiant du message
|_ date du message
|_ information de service émetteur
|_ information de service destinataire
|_ code de réponse (pour les réponses)
|_ 0..* commentaires
|_ 0..* identifiant d'objet (pour les demandes)
|_ 0..* objet (pour les échanges de données lors des versements, communication, restitution)
```

Le SEDA fournit un modèle complet et détaillé adapté aux données d'archives publiques pour

- les objets échangés
- la description des objets échangés
- la description des services (émetteur, destinataire, producteur dans la description des objets)
- la description des règles de conservation et de communicabilité
- les référentiels associés pour les codes de conservation, de communicabilité, le format, les unités de mesure, les langues...

La norme MEDONA fournit un modèle détaillé pour les objets échangés mais laisse libre l'implémentation du modèle pour l'adapter au besoin de la communauté ou du domaine pour :

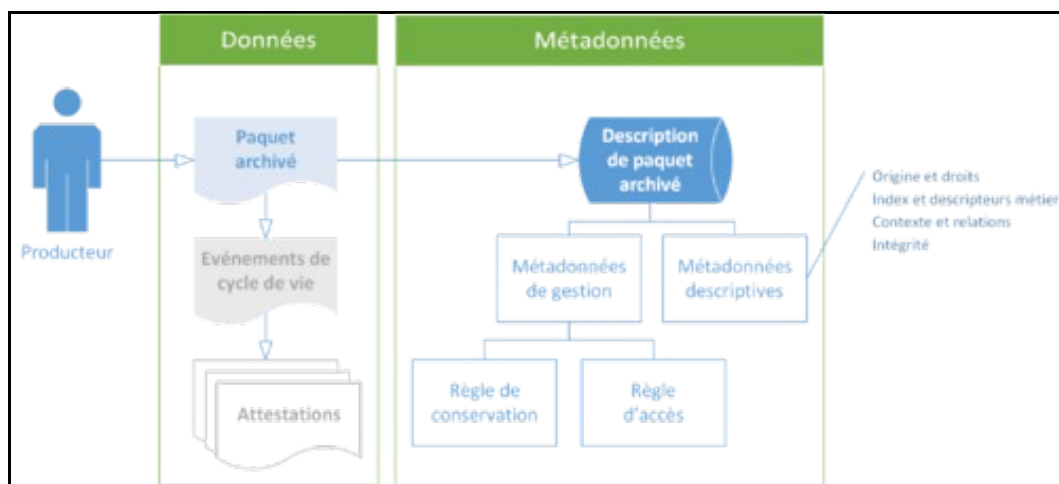
- la description des objets échangés
- la description des services (émetteur, destinataire, producteur dans la description des objets)
- la description des règles de communicabilité
- les référentiels associés pour les règles de communicabilité, le format, les unités de mesure, les langues, les codes réponse, les algorithmes de hachage...

## Spécifications du système d'archivage

Ce chapitre propose, en se basant sur les concepts et les exigences décrites précédemment, une description fonctionnelle du système d'archivage en tant que solution logicielle intégrée à un dispositif plus large qui vise la conservation de l'information sous la forme de documents numériques.

### Le modèle de données

Voici un modèle conceptuel très simplifié des données gérées par le système d'archivage électronique :



Au niveau macroscopique, on voit tout de suite que le SAE répartit l'information en deux grands domaines :

- **les données**, qui sont l'objet de la conservation, et
- **les métadonnées**, qui sont utilisées pour la conservation.

Faisons le parallèle avec un service d'archives papier :

- les données sont les dossiers stockés dans les rayonnages, qui doivent être classés et conservés dans les conditions requises pour assurer leur pérennité, et
- les métadonnées sont les lignes inscrites dans les registres du Service d'Archives, consultables dans les bureaux ou au comptoir.

### Données

Si l'on respecte le modèle de données de la norme OAIS, nous avons déjà vu que les données sont structurées en paquet d'information archivé ou AIP, avec son contenu d'information et son information de pérennisation. Le choix d'un modèle pour la structuration des métadonnées dépend largement du domaine et des besoins de la communauté adressée. Il existe de nombreuses initiatives qui visent à créer des standards de description des métadonnées :

- Bibliothèque du Congrès U.S.
  - ◆ METS (Metadata Encoding & Transmission Standard)
  - ◆ EAD (Encoded Archival Description)
  - ◆ EAC (Encoded Archival Context)
- Dublin-core de la Dublin Core Metadata Initiative
- Conseil International des Archives (ICA)
  - ◆ ISAD (G) (Archival Description - General)
  - ◆ ISAAR (CPF) (Archival Authority Record - Corporate bodies, Persons and Families)
  - ◆ ISDF (Describing Functions) pour les fonctions des producteurs
  - ◆ ISDIAH (Describing Institutions with Archival Holdings) pour les institutions

Pour les archives papier, il s'agirait d'une fiche descriptive jointe au dossier.

Selon la norme NF Z42-013, les données sont constituées non seulement des paquets archivés provenant des producteurs, mais aussi des journaux du cycle de vie et attestations produits par le Service d'Archives au cours de son activité, ainsi que du journal de l'application retraçant tous les événements applicatifs.

Les paquets archivés sont stockés sur des supports de stockage. La norme NF Z42-013 exige qu'au moins une copie de sécurité soit réalisée sur un site distant avant de délivrer une attestation de dépôt au déposant. Il y a donc deux copies de chaque paquet archivé.



## Métadonnées

Les métadonnées constituent la description du paquet archivé, utilisée afin d'exploiter l'information.

La norme MEDONA et le SEDA fournissent des modèles de description pour les échanges, donc plutôt des représentations des paquets versés et des paquets diffusés. Comme on le verra, il existe bien d'autres modèles plus ou moins génériques ou spécifiques à des domaines, mais tous ont en commun une structuration logique en deux parties des métadonnées :

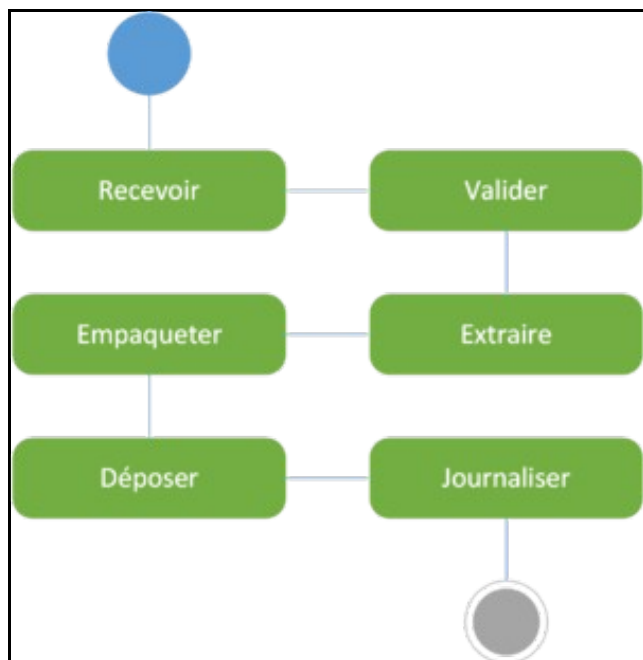
- les métadonnées de gestion permettent au Service d'Archive de gérer la conservation des documents. Elles comportent à minima une règle de conservation qui définit la durée de conservation, la date de départ du calcul et le sort final, ainsi qu'une règle d'accès qui définit les droits des personnes à réaliser les opérations (dépôt, communication, modification, élimination, restitution...) sur les archives.
- les métadonnées descriptives fournissent la description de l'information archivée et peuvent varier en quantité et en qualité : son origine et les droits associés, son identification grâce à des index et descripteurs métier, son historique, ses relations avec d'autres archives présentes dans le système ou une autre archive, etc.

## Acteur producteur et classement

La norme MoReq définit la nécessité d'un classement, qui constitue le moyen de catégoriser l'information et de l'attribuer à une activité du producteur. Bien que le producteur soit un acteur et non une entité de données, la position de classement représente un composant du modèle conceptuel, qui peut être reflété soit par l'organisation physique du stockage (un espace dédié et structuré selon le plan de classement qui reflète l'activité du producteur), soit par l'adjonction de métadonnées descriptives au travers d'un plan de classement virtuel.

## Les fonctions

### L'entrée



### Étapes d'une entrée d'archive

Sur un plan fonctionnel, il s'agit pour le déposant de transférer la responsabilité de la conservation des documents numériques à un Service d'Archives, qui lui fournit en réponse une attestation que le dépôt est réalisé selon les termes de l'accord passé entre les parties et conformément à la politique d'archivage.

Sur un plan technique, il s'agit d'enregistrer l'objet numérique dans le système d'archivage. Pour ce faire, il faut prendre en compte tous les systèmes qui produisent de l'information archivable : les logiciels bureautique, chaînes éditique, applications métier, progiciels et ERP, les bases de données, les sites internet, les chaînes de numérisation, les messageries et autres systèmes de communication... La liste est non exhaustive et évolue fréquemment. C'est une fonction plus complexe qu'il n'y paraît, qui peut être modélisée en 6 étapes:

- **Recevoir** : Le message du déposant est reçu et transfère des données d'archive de manière unitaire ou par lot, via des protocoles d'échange, des connecteurs, etc. Le dépôt est stocké dans un sas d'entrée sécurisé, pour permettre à la procédure d'être interrompue et reprise ultérieurement sans perte d'information.
- **Valider** : Le système valide la complétude et la conformité des données reçues, par rapport à la politique d'archivage, aux accords entre les parties, à un profil de typologie documentaire. Il peut aussi inclure une phase d'acceptation par le Service d'Archives, voire de modification du paquet reçu pour le mettre en conformité.
- **Extraire** : Les métadonnées sont extraites à partir du paquet versé, c'est-à-dire toute l'information nécessaire à la constitution de son référentiel et du paquet archivé. Cette étape pourrait être intervertie avec la précédente si l'extraction s'avère nécessaire pour valider l'information.
- **Empaqueter** : Cette étape génère les entités conservées sous la forme du paquet d'information archivé et des métadonnées de description.
- **Déposer** : Le paquet archivé est enregistré sur les supports de stockage, le référentiel de l'archive est mis à jour avec les métadonnées et l'adresse de stockage.
- **Journaliser** : Le système trace le dépôt dans le journal de l'application et le journal du cycle de vie de l'archive et produit une attestation de dépôt à destination du déposant.

## Le classement

Il s'agit de rattacher le document à l'activité du producteur, afin de lui permettre de le retrouver, de gérer les accès, de gérer le cycle de vie de l'archive. Le plan de classement doit donc refléter l'activité et positionner l'information dans une structure hiérarchique faisant apparaître l'organisation, les grands domaines d'activité, les processus qui produisent l'information numérique, les éventuels dossiers de regroupement et enfin les typologies de documents.

Le diagramme ci-dessous représente la structuration des niveaux de classement, avec pour chacun une proposition de rattachement à un niveau de description de l'activité :

Fonds d'archive : organisation  
Sous-fond : division de l'organisation (facultatif)  
Série : domaine d'activité  
Sous-série : processus (facultatif)  
Dossier  
Document

Le fonds d'archives représente la racine du plan, à partir de laquelle les branches de plus en plus précises se terminent par la typologie documentaire qui est la feuille, fournissant ainsi un ensemble d'informations relatives à

- la nature de l'information contenue par le document
- la règle de conservation : DUC, DUA, sort final, date de référence du calcul
- les règles d'accès : droits d'accès, communicabilité, confidentialité
- les index et métadonnées descriptives associées

La définition du plan de classement est l'une des premières étapes dans la mise en œuvre d'un système d'archivage, car elle fournit le modèle de données spécifique aux besoins de l'activité et l'ensemble des règles de gestion à implémenter.

### Le stockage

Il s'agit de conserver les paquets archivés sur les supports, qui peuvent être réinscriptibles (HDD, SAN/NAS), WORM (Write Once Read Many) physiques ou logiques (CD/DVD, systèmes propriétaires), ou encore amovibles.

Le stockage assure les opérations suivantes :

- **Écriture** : lors des entrées, des migrations support et de format
- **Lecture** : lors des accès pour consultation, communication, restitution et les contrôles d'intégrité
- **Suppression** : lors des opérations d'élimination et de restitution

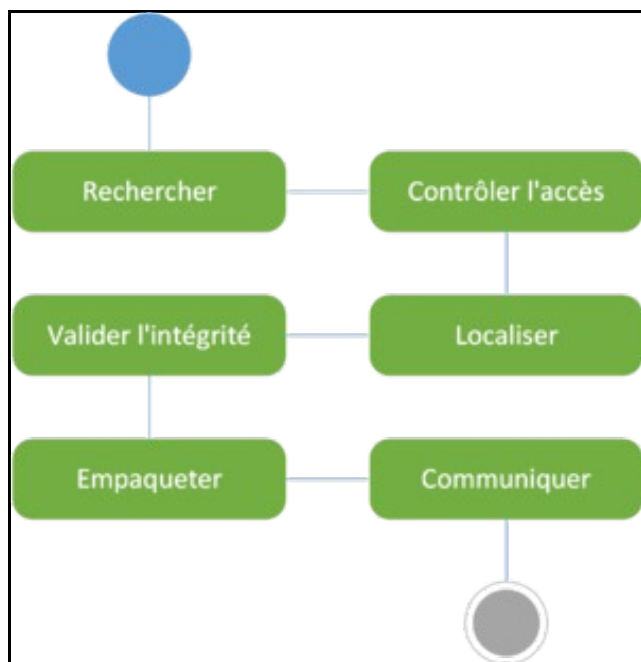
### Les métadonnées

Il s'agit de gérer le référentiel des données qui permettent d'exploiter les documents.

Les métadonnées sont gérées dans une base de données qui assure les opérations suivantes :

- **Enregistrement** : lors des entrées (dépôt, conversions de format)
- **Lecture** : pour la recherche et la consultation, la production de rapports et de statistiques
- **Mise à jour** : lors de la modification des règles de gestion, le gel ou dégel de l'application du cycle de vie, les migrations
- **Suppression** : lors des élimination et restitution

### L'accès



#### Les étapes de l'accès

Dans la vision fonctionnelle, il s'agit de délivrer au demandeur un contenu d'information ayant une valeur probatoire, avec toute l'information pour le comprendre et l'exploiter.

D'un point de vue technique, l'accès restitue l'objet numérique à l'utilisateur, accompagné de tout ou partie du paquet archivé et des métadonnées. Il permet de prendre connaissance du contenu d'information via les moyens nécessaires pour visualiser, télécharger, imprimer.

Il s'agit comme pour l'entrée d'un processus plus complexe qu'il n'y paraît, qui peut être découpé en 6 actions :

- **Recherche** : Elle permet à l'utilisateur de retrouver l'information dans l'archive, au travers de formulaires de recherche multi-critères, de recherche à facette, par la navigation dans un plan de classement archivistique ou virtuel, etc. Elle doit être simple, intuitive, ergonomique, pertinente.
- **Contrôler l'accès** : Le système doit s'assurer de l'identité du demandeur, vérifier ses droits d'accès aux métadonnées et aux données, prendre en compte les éventuelles informations sur la communicabilité ou encore la confidentialité de l'archive.
- **Localiser** : Le système doit retrouver le paquet archivé via son adresse de stockage enregistrée dans le référentiel, mais aussi vérifier sa disponibilité (pour les supports amovibles), déterminer, s'il y a plusieurs copies, laquelle ou lesquelles doivent être restituées.
- **Contrôler l'intégrité** : Il faut vérifier que le document numérique restitué est bien conforme, qu'il n'a subi aucune altération ou substitution, afin de garantir la valeur probatoire du contenu, via un contrôle d'empreinte numérique ou encore la validation d'une signature électronique

- **Empaqueter** : Cette étape consiste à organiser les données et les métadonnées communicables pour générer le paquet à diffuser dans la forme attendue par l'utilisateur.
- **Communiquer** : C'est un processus technique qui assure le transport du paquet diffusé vers le demandeur via un protocole de communication ou le dépôt dans un sas de sortie par exemple. Pour une consultation simple via le SAE, celui-ci devra prévoir les dispositifs techniques de restitution de l'information (visualiseurs), de téléchargement ou encore de réimpression de contenus.

## Le management

C'est plus un ensemble de fonctions, dont l'objectif est de définir les règles de fonctionnement du système d'archivage et d'en assurer le bon fonctionnement selon 4 axes :

- **Gestion** : Quotidiennement, le management doit administrer les référentiels liés à l'archive et assurer le traitement des opérations d'échange de données avec les tiers (déposants et accédants) lors des entrées et des accès.
- **Conformité** : Il faut mettre en œuvre et maintenir les processus dans le respect des normes et standards en vigueur, mais aussi assurer un contrôle continu (à la demande, périodique, systématique ou par échantillonnage) de la conformité des archives.
- **Sécurité** : Le management doit définir les règles de contrôle de l'accès aux données et métadonnées, procéder à des copies de sécurité, des sauvegardes, des restaurations, assurer la traçabilité par la journalisation des événements et la sécurisation des journaux produits.
- **Exploitation** : Le management a en charge de gérer la configuration technique et fonctionnelle, de contrôler l'exploitation et les performances du système, de détecter et de traiter les anomalies qui surviennent.

## La pérennisation

Au cœur du système d'archivage, la stratégie de pérennisation vise à assurer la continuité de l'accès aux contenus d'information pour toute la durée de leur conservation. On l'a vu, le document numérique dissocie le support et le contenu binaire, chacun étant soumis à des risques liés à l'obsolescence des technologies qui permettent leur conservation. On peut donc séparer les tâches en deux grandes catégories correspondant aux composantes du document que sont le stockage et le contenu.

Pour le stockage, il s'agit d'assurer la persistance de l'information enregistrée en assurant un contrôle permanent de l'état des supports, en procédant à des opérations de maintenance, et le cas échéant à des migrations vers d'autres supports :

- **le rafraîchissement** est la réalisation d'une copie exacte du contenu vers un support équivalent neuf lorsque l'état du support actuel n'est plus satisfaisant. Par exemple, lorsque un support réinscriptible comme un disque dur magnétique comporte des secteurs défectueux ou que sa capacité maximale est pratiquement atteinte, son contenu copié vers un nouveau disque plus récent ou de capacité supérieure.
- **la duplication** est aussi une copie exacte du contenu, à la différence que le support cible n'est pas équivalent au support d'origine, car il utilise une autre technologie pour les accès. Par exemple, le contenu d'un disque dur magnétique est copié vers un disque optique WORM de type DVD-R sans changer la structure du stockage mais en changeant les processus d'accès, qui incluront désormais une opération de montage du DVD pour remettre le contenu en ligne.
- **le ré-empaquetage** est une réorganisation du contenu. Par exemple, une restructuration des répertoires de stockage pour fusionner ou au contraire éclater les regroupements logiques.

Pour les contenus d'information, l'objectif est de pérenniser l'accès à l'information à partir du contenu binaire stocké. De par la nature numérique du contenu, sa restitution à l'utilisateur pour qu'il prenne connaissance du contenu d'information nécessite, outre l'accès au flux binaire, l'interprétation de ce dernier par un logiciel. Ce logiciel lui-même est soumis à l'obsolescence du code (non maintenu par l'éditeur), mais aussi de l'environnement nécessaire à son exécution, donc des bibliothèques tierces, du système d'exploitation, et par extension des matériels compatibles avec leur exécution.

Le contenu de données stocké est interprété par un logiciel dans son environnement d'exécution, qui restitue l'information à l'utilisateur.:



Les stratégies de pérennisation se définissent sur deux axes :

- la conservation de la technologie liée à l'environnement logiciel capable de restituer le contenu <--> la conservation du contenu d'information lui-même.
- la spécificité à un format ou une famille de format <--> la prise en compte générique de tous les formats archivés

Le choix d'une stratégie est guidé par l'analyse des risques et la faisabilité technique, mais aussi par la nature des contenus et leur utilisation (s'ils sont accédés fréquemment ou non, s'il sont conservés 10ans ou 500...). Parmi elles, en voici 3 qui nous semblent significatives :

- **La conversion de format** réorganise le contenu binaire d'un document d'un format obsolète (ou reçu non pérenne) vers un nouveau format d'encodage pérenne. C'est une stratégie risquée qui crée une nouvelle copie à conserver et ne doit pas avoir d'impact sur le contenu informationnel.
- **Le portage** permet de pérenniser le logiciel en lui permettant de fonctionner dans de nouveaux environnements d'exécution plus récents.
- **La virtualisation** conserve l'intégralité de l'environnement logiciel dans une machine virtuelle embarquant le système d'exploitation, les bibliothèques tierces et le ou les logiciels qui peuvent restituer l'information. Cette stratégie nécessite que le document soit "transféré" dans l'environnement virtuel afin d'être restitué, ce qui est à réserver aux documents dont l'exploitation n'est pas nécessairement immédiate.

## Les fonctions complémentaires

Aux 7 fonctions principales détaillées ci-dessus, on peut ajouter rapidement les suivantes qui peuvent compléter le périmètre fonctionnel du système d'archivage:

- La GED, workflow et collaborative
- La gestion de dossiers sériels (clients, RH, banque/assurance/immobilier)
- La signature électronique des journaux et contenus
- Le chiffrement des données et des documents
- La gestion des droits numériques (DRM)
- Le travail hors ligne et à distance (synchronisation de base documentaire)
- ...

## Annexes

## Références

- Norme ISO 14721:2003 OAIS : Open Archival Information System
- Norme AFNOR NF Z42-013 : 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, guide d'application GA Z42-019, et règles de certification à la marque NF461
- Norme AFNOR NF Z44-022 MEDONA : Modèle d'Echange pour les DONnées d'Archives
- Standard SEDA : Standard d'Echange pour les Données d'Archives
- Norme ISO 15489 : Records Management
- Norme MoReq2010 : Modular Requirements for Records Systems
- Norme ICA-Req : Principes et exigences fonctionnels pour l'archivage dans un environnement électronique

## Bibliographie

- "Dématérialisation et archivage électronique" de Jean-Marc Rietsch, Marie-Anne Chabin, Eric Caprioli ? Editions DUNOD
- "L'archivage numérique à long terme. Les débuts de la maturité ?" de la DAF ? La Documentation Française

## Liens externes

- Modernisation de l'Etat : <http://references.modernisation.gouv.fr/archivage-numerique>
- Archives de France : <http://www.archivesdefrance.culture.gouv.fr/gerer/archives-electroniques>
- Groupe PIN : <http://pin.association-aristote.fr/doku.php>
- CINES : <https://www.cines.fr/archivage/>
- Série d'articles "Dématique et SAE" par Antony Belin sur le site Convergence de l'Association des archivistes du Québec <https://archivistesqc.wordpress.com/author/belinantony>
- Bibliothèque du Congrès : <http://www.digitalpreservation.gov/>
- Archives Nationales du Royaume-Uni : <http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records>
- Archives Nationales de Nouvelle-Zélande : <https://digitalpreservation.natlib.govt.nz>