Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
Premier ministre
Direction interministérielle des systèmes d'information et de communication

Archivage électronique

Guide de bonnes pratiques

Fiches annexes

Table des matières

Mise en place d'une stratégie d'archivage dans le système d'information d'une organisation.....5

Enjeux de l'archivage et cycle de vie des documents.....5

Modalités d'intervention sur l'ensemble du système d'information de l'organisation.....6

Comment définir une stratégie d'archivage face aux différentes applications métier existantes ?.....6

1. Définition des priorités d'archivage à partir de la cartographie : répondre à la question "quoi archiver?".....6
2. Répondre à la question « comment archiver ? » 8
3. Les différentes stratégies d'archivage.....8
4. Mise en œuvre du scénario d'archivage.....10

Pour tout nouveau projet d'application métier, intégration de l'expertise archives.....11

1. Où s'insère l'expertise archives ?.....11
2. Focus sur chaque étape du projet.....11
3. Vie de l'application.....13
4. Intégration de l'expertise archives au cycle de vie "projet": vision synoptique.....14

Trois cas d'usage.....15

L'accompagnement des services dans la gestion de leurs fichiers bureautiques et de leurs serveurs partagés.....15

1. Mise en œuvre du plan de classement dans un univers bureautique.....15
2. Le nommage des dossiers et des fichiers.....16
3. Les modèles de document.....17

L'accompagnement des services dans l'archivage de leurs courriels.....18

1. Contexte juridique.....18
2. Archivages des courriels : les approches existantes.....19
3. Sélection des courriers électroniques.....20
4. Mise en œuvre concrète de solutions d'archivage.....21

Intégrer la gestion du cycle de vie dans un projet de numérisation.....23

1. Définir le projet de numérisation.....23
   - 1.1. Définir le périmètre de la numérisation.....23
   - 1.2. Choix des dossiers à numériser.....23
   - 1.3. Réflexion sur l'organisation de la numérisation.....23
   - 1.4. La création des dossiers dans le système d'information 24
2. Les étapes du projet.....24
   - 2.1. Déterminer la valeur juridique des documents.....24
   - 2.2. Déterminer les durées d'utilité administrative des dossiers et des pièces.....24
   - 2.3. Déterminer le sort final des dossiers et des pièces.....24
   - 2.4. Établir le plan de classement des pièces constitutives du dossier et la liste des métadonnées :.....24
   - 2.5. Conduire et organiser les opérations de numérisation.....25
3. Choix techniques.....26
   - 3.1. Adopter des formats ouverts.....26
   - 3.2. Définir des règles de nommage des fichiers numérisés ainsi que des dossiers numériques 26
   - 3.3. Privilégier la fidélité au document (couleur ou niveaux de gris plutôt que noir et blanc)..26
   - 3.4. Limiter les traitements sur les images.....26
   - 3.5. Sécuriser l'application de production 27

3.6. Protéger les données à caractère personnel....27
3.7. Sécuriser le stockage....27
4. Contrôles....28
4.1. Contrôles dans le cadre de la numérisation du stock....28
4.2. Contrôles dans le cadre de la numérisation du flux....28
5. Intégrer les fichiers nativement numériques ?....29

Présentation théorique d'un système d'archivage électronique....30

Présentation de la norme OAIS....30
1. L'environnement OAIS....30
2. Le modèle d'information....32
3. Les entités fonctionnelles....35
4. Perspectives de collaboration entre Archives....37
5. Évolutions en cours....38

Fonctionnalités d'un système d'archivage électronique....40
1. Fonctionnalités du SAE....40
F1. Fonction versement....41
F2. Fonction stockage....43
F3. Fonction gestion des données descriptives....45
F4. Fonction consultation / communication....46
F5. Administration du Système d'archivage électronique....47
2. Architecture du SAE....48

Les conditions de l'interopérabilité et de la préservation sur le long terme : le respect du cadre normatif....50

Les normes généralistes de l'archivage électronique....50
1. Organisation et processus....51
2. Pérennisation....53

Les normes spécialisées de l'archivage électronique....55
1. Les formats de représentation de l'information....55
Critères à retenir dans le choix des formats....55
Formats bureautiques et non structurés....56
Formats image....57
Formats d'archivage pour les bases de données....57
Formats d'archivage pour les documents audiovisuels....57
Outils d'identification des formats....57
Outils de validation des formats....58
2. Les formats d'échanges....58
Présentation du standard d'échange de données pour l'archivage (SEDA)....59
L'utilisation du standard d'échange de données pour l'archivage....61
Les ressources sur le SEDA et sa boîte à outils....62
3. Les supports....63
4. Les formats de métadonnées....64

Le cadre juridique de l'archivage électronique....66

Le cadre juridique de l'administration électronique....66
1. Intégrité des données....66
2. Les fondements juridiques de l'authenticité des données numériques....68
3. Une contrainte juridique supplémentaire de l'environnement numérique : la protection des données personnelles....71

Le cadre juridique de l'archivage....74
1. Définition et justification des archives....74

2. Le réseau des archives de l'État....75
3. Exercice du contrôle scientifique et technique de l'État sur les archives publiques 75
4. Modalités de collaboration entre les services producteurs et les services d'archives. Versements dans les services publics d'archives....76
5. Accès aux archives et régime de communication des archives (Art. L212-1 à L 212-5)....78
6. Sanctions pénales....78
La politique d'archivage dans le secteur public....80
1. Le référentiel sur l'archivage numérique sécurisé....80
2. Structure du référentiel : politique d'archivage et répartition des responsabilités....80
3. Répartition des rôles et des responsabilités....81
Glossaire....82

Mise en place d'une stratégie d'archivage dans le système d'information d'une organisation

Cette fiche a pour but d'aider à la prise en compte des problématiques d'archivage dans un SI. Elle s'intéresse uniquement aux applications (bases de données métier, gestions électroniques de documents...). La question de l'archivage des documents bureautiques et des courriels est par ailleurs traitée dans deux fiches pratiques spécifiques.

Plan de la fiche :

- Enjeux de l'archivage et cycle de vie des documents
- Modalités d'intervention sur l'ensemble du système d'information de l'organisation
- Comment définir une stratégie d'archivage face aux différentes applications métier existantes ?
  - Répondre à la question "comment archiver?"
  - Les différentes stratégies d'archivage
  - Mise en œuvre du scénario
- Pour tout nouveau projet d'application métier, intégration de l'expertise archives
  - Où s'insère l'expertise archives ?
  - Focus sur chaque étape du projet
  - Vie de l'application

Enjeux de l'archivage et cycle de vie des documents

Les documents à archiver produits ou reçus par une personne physique ou morale sont à conserver à titre de preuve, conformément aux obligations légales ou d'information nécessaire à l'exercice de l'activité courante. Ces documents peuvent se présenter sous différents formats. Une fois cette durée d'utilité administrative expirée, certains d'entre eux revêtent un intérêt patrimonial qui justifie leur prise en charge à titre définitif par le service public d'archives.

Généralement, le service ayant recours à une application pour produire des documents et gérer des données a besoin de :

- accéder immédiatement et efficacement à l'information pertinente ;
- garantir le statut des documents ou données (validé ou non) ;
- garantir la complétude et l'intégrité des documents ;
- maîtriser les risques de perte ou de destruction intentionnelles ou non intentionnelles de données ou documents ;
- détruire de manière contrôlée les données ou documents qui n'ont plus d'utilité ou dont la durée de conservation est échue ;
- verser les archives historiques à l'autorité archivistique compétente.

Une bonne gestion de l'archivage permet de répondre à l'ensemble de ces besoins.

Prendre en compte le cycle de vie des données et des documents et organiser leur archivage nécessite une collaboration forte entre les services d'archives, les directions des systèmes d'information et les services métiers.

Modalités d'intervention sur l'ensemble du système d'information de l'organisation

Il conviendra :

- d'obtenir de la direction **un soutien clair (sponsor)** pour la mise en place d'une politique globale d'archivage englobant tant la production traditionnelle que la production numérique (applications métier, documents bureautiques, courriels). Ce soutien fort doit se manifester notamment :
  - par une lettre de mission annonçant la mise en place de cette politique, diffusée à toute l'organisation ;
  - par la désignation d'un responsable de la politique en matière de qualité de l'information au sein de l'organisation ;
  - par des moyens humains et financiers clairement définis ;
  - par la mise en œuvre de plans de formation adaptés pour les agents de l'organisation.
- d'élaborer cette **politique d'archivage globale** pour l'organisation. On peut envisager de décliner des politiques d'archivage spécifiques à des typologies documentaires et à des familles d'applications.
- de faire intégrer dès que cela est possible dans le **schéma directeur informatique de son organisation ou de son entité,** le bloc fonctionnel "gestion du cycle de vie de l'information et archivage".
- de conduire des **opérations de communication** régulières vis à vis des agents de l'organisme qui seront des utilisateurs, des dirigeants.

Comment définir une stratégie d'archivage face aux différentes applications métier existantes ?

Une fois cette première étape concernant l'ensemble du système d'information de l'organisation franchie, il convient de définir une stratégie d'archivage pour les différentes applications métiers existantes.

Deux modalités d'actions existent : d'une part définir une stratégie d'archivage face à applications métier déjà existantes, d'autre part intégrer l'expertise archives dans la conduite de projet informatique de son organisation, dès sa conception.

Une fois ces étapes réalisées, il conviendra de veiller à leur mise en œuvre.

1. Définition des priorités d'archivage à partir de la cartographie : répondre à la question "quoi archiver?"

a) Amorcer le travail d'expertise des systèmes d'information à partir de la cartographie des SI du département ministériel.

La norme ISO 15-489 recommande le recensement à caractère systématique des objets numériques et l'évaluation de chaque objet numérique dans son contexte (Partie 1. point 8.4.).

Si cette cartographie est en projet, il est nécessaire d'intégrer le groupe projet du service des systèmes d'information en charge de sa rédaction.

L'objectif est en effet le recensement des applications existantes, leur évaluation

archivistique et technique et l'établissement d'un diagnostic et de recommandations pour la mise en conformité de ces applications (plan d'action et calendrier). Ainsi, au ministère de la culture, sur la base de la cartographie revue, 75 applications ont été retenues pour l'étude dont 25 prioritaires. L'utilité de cette démarche était double : pour l'archiviste, initier la dynamique d'archivage numérique, pour la DSI, disposer d'une cartographie enrichie et mise à jour.

Au ministère en charge de l'écologie, le travail d'évaluation archivistique se baserait ainsi sur le cadre stratégique d'urbanisation des SI qui englobe à la fois la liste des applications par domaines et sous-domaines, la liste des applications par entités administratives et enfin les fiches « application », soit des éléments directement utilisables pour une macro-évaluation : identification du SI (enjeux, utilisation, référentiel), type d'application, objets métiers, fonctionnalités, contexte juridique, organisation mise en place, ressources existantes, description détaillée (technique), état d'avancement.

Cette évaluation archivistique à partir d'une cartographie permet ainsi d'intégrer dans les outils de pilotage de l'archivage (tableaux de gestion ou calendriers de conservation)[1] aux côtés de la production traditionnelle, la production bureautique et ainsi de pouvoir avoir un début d'analyse comparative possible entre les différents flux produits, une compréhension de la généalogie des processus (papier, mécanisation, dématérialisation...), une identification de la MOA et enfin un outil de dialogue entre les différents « métiers ».

La limite de ce type de dispositif est d'une part le nombre parfois très élevé des systèmes d'information de l'organisation (près de 500 au ministère en charge de l'écologie, près de 2000 au ministère de la défense) ainsi que la complexité de certains flux qui convergent vers plusieurs processus.

b) Si cette cartographie n'existe pas, il est nécessaire d'en réaliser une pour les systèmes d'information au plus près du cœur de métier de chaque direction.

Pour chaque application, pour une première approche, il convient également d'avoir une connaissance précise des utilisateurs et de leurs besoins en termes de finalités de l'archivage (probatoire, patrimonial) ainsi que des critères de recherche et d'accès.

Il est conseillé, pour ce faire, de s'appuyer sur l'ensemble des outils de vulgarisation et de présentation de l'application existantes (manuels utilisateurs, actions de communication….).

Nota : Il convient de porter une attention particulière aux données à caractère personnel qui requièrent généralement un traitement spécifique

Pour cette étape, il convient de s'appuyer sur les urbanistes SI de son organisation ainsi que sur les archivistes, en complétant bien sûr avec les informations fournies par les services métier.

A l'issue de cette phase, un premier état des lieux aura été fait, à partir duquel un **audit plus circonstancié des principales applications repérées** devra être mené afin de pouvoir prioriser les SI à forte valeur juridique et/ou patrimoniale.

Les spécifications IcaReq (module 3) - voir fiche normes généralistes- peuvent utilement être utilisées pour auditer un SI. De même, une grille outils a été définie dans le cadre de la méthode Astaré (voir ci-dessous).

- Voir : « Une *méthodologie pour l'audit des applications métier avec le module* d'ICA-Req » à destination des archivistes"[2] boîte à outils de la norme IcaReq
- Voir la grille Outils (60 questions) élaborée dans le cadre de la méthode ASTARE

1 Voir en annexe le calendrier de conservation de la DARES.
2 Rédigée par Lourdes Fuentes-Hashimoto (MAEE).

2. Répondre à la question « comment archiver ? »

La réponse à la question *Comment archiver les données et documents numériques produits dans mon organisation* ?, suppose l'analyse :

- des besoins prioritaires en termes d'archivage ;
- des acteurs du processus d'archivage, du point de vue leur degré de sensibilité à la problématique et aux enjeux de l'archivage numérique ;
- des outils, sous l'angle de la sécurité du système d'information et des capacités d'archivage des applications métier dans lesquelles sont créés les données et documents.

Il sera alors possible de définir **la stratégie d'archivage (et la détermination des responsabilités d'archivage)** et une planification des interventions suivant les urgences. On pourra définir des stratégies d'archivage différentes suivant les applications.

La boite à outils Astaré (Analyse stratégique de l'archivage électronique) permet de répondre à cette question en identifiant la stratégie la mieux appropriée dans le contexte du système d'information (SI) analysé[3].

3. Les différentes stratégies d'archivage

a. Typologie des stratégies

Il s'agit :

- soit de permettre l'archivage des données durant leur durée de conservation au sein de l'application métier (la fin du processus étant l'élimination réglementaire des données avec le visa de l'administration des archives) ;
- soit d'organiser, dès leur validation ou durant le cours de leur durée d'utilité administrative, le versement vers un SAE (adoption dans ce cas du format d'échange SEDA).

Dans ce second cas, l'archivage électronique sera géré :

- soit par l'administration productrice dans son SAE ;
- soit sera pris en charge au sein d'un SAE mutualisé avec d'autres services, d'autres administrations, voire dans le SAE géré par le service public d'archives pour ses archives définitives ;
- soit en mode externalisé par un tiers archiveur agréé.

http://www.archivesdefrance.culture.gouv.fr/gerer/records-management-et-collecte/agrements/

Du point de vue d'un DSI, il sera effectivement utile d'évaluer l'intérêt d'utiliser l'offre existante interne ou externe à son organisation et d'analyser l'opportunité de développer un SAE ou de le mutualiser en étudiant les interfaces à développer avec des applications métier

Si l'archivage des données est réalisé au sein d'un SAE, il faudra faire attention aux conditions d'accès aux documents et données. Pour les données et documents auxquels on souhaite accéder rapidement et fréquemment, il pourra être utile de prévoir des fonctionnalités de recherche dans le SAE directement depuis l'application métier, de manière transparente pour l'utilisateur.

Concernant les documents d'intérêt scientifique, statistique ou historique, ceux ci peuvent être versés dans le SAE du service d'archives public compétent à plusieurs moments :

3 Cette boite à outils a été réalisée par le groupe de travail de l'association des archivistes français (AAF) « conduire un projet d'e-archivage dans la sphère publique », en partenariat avec les Archives de France.

-dès la validation des documents et données ;
-selon une périodicité régulière, définie au moment de la phase de cadrage ;
-à la fin de la durée de responsabilité de l'organisation sur les données (terme de la durée d'utilité administrative).

Dans les deux premiers cas, les filières d'archivage des données pour les besoins de l'organisation et pour les besoins patrimoniaux seront parallèles et concomitantes. Au plus tard à la fin de la DUA, sera effectué le versement dans un service public d'archives pour archivage définitif.

Pour cette étape, le choix de la stratégie sur la base de l'expertise de l'archiviste sera effectué au sein des instances décisionnelles du système d'information de l'organisation.

b. Quelques exemples de choix pouvant être effectués

Archivage dans l'application métier

Privilégier l'archivage au sein de l'application métier pour des données de gestion dont la durée de conservation est de faible durée.

Pour des données et documents auxquels on souhaite accéder rapidement et fréquemment, privilégier si possible l'archivage au sein de l'application ou du moins l'export vers un SAE à haute disponibilité.

Archivage hors de l'application métier ou au sein d'un SAE mutualisé

Ne pas déployer en interne un SAE si la structure est petite ou s'il existe des solutions de mutualisation interministérielle.

Ne pas déployer en interne un SAE si la maturité des acteurs en termes d'archivage numérique n'est pas suffisante.

Pour des données et documents signés électroniquement, privilégier le versement dans un SAE (maintien de leur force probante) dès leur validation.

Pour des documents à forte valeur patrimoniale, il est toujours possible de verser dans le SAE des archives nationales ou départementales un exemplaire de ces archives dès leur validation.

Externalisation

Ne pas choisir a priori d'externalisation de ses données et documents chez un tiers archiveur pour des données et documents à très longue conservation pour des raisons de coûts (lors des changements de tiers archiveurs lors des changements de marchés).

Ne pas choisir a priori d'externalisation de ses données et documents chez un tiers archiveur pour des données sensibles ou hautement confidentielles.

Ne pas externaliser chez un tiers archiveur des archives définitives (interdiction du Code du patrimoine).

Stockage

Ne pas choisir des supports de stockage type bandes pour des données et documents fréquemment consultés et à l'inverse ne pas hésiter à choisir des supports de stockage type bandes moins onéreux pour des documents peu consultés.

Une fois le scénario (ou les scénarios) d'archivage défini(s), il faut faire **évoluer les différentes applications priorisées pour intégrer les fonctionnalités relatives au cycle de vie de l'information et pour préparer les éliminations et/ou les exports pour archivage définitif.**

Sinon, il faut **mettre en œuvre un SAE** (voir ci-dessous).

4. Mise en œuvre du scénario d'archivage

Dans tous les cas, il convient de :

- Établir un contrat de service avec les producteurs et les services informatiques précisant les rôles, missions et responsabilités de chacun des partenaires.
- Définir les durées de conservation et sort final des données et documents et développer les fonctionnalités correspondantes.
- Définir éventuellement sa stratégie en matière d'archivage de fichiers signés : outils de vérification des signatures, de production de rapports de vérification et d'archivage de ces rapports.
- Définir les règles de nommage des fichiers.
- Définir sa stratégie en matière de formats des fichiers acceptés par les applications métier, récupérer les métadonnées techniques relatives à ces formats afin de permettre les futures migrations. Les formats ouverts et si possible normalisés seront privilégiés..
- Définir les modalités (réglementaires et matérielles) de destruction des données et documents arrivés à terme de leur durée de conservation et à la traçabilité de ces destructions et faire développer les fonctionnalités nécessaires à les mettre en œuvre.
- Définir les formats d'échanges (format SEDA) en cas de versements et les profils de données correspondants et faire développer les fonctionnalités relatives aux exports..
- Dans le cas des données à archiver qui sont à extraire de bases de données, on privilégiera le format SIARD
- En cas de données personnelles, introduire les modalités d'archivage dans la déclaration CNIL.
- Mettre en œuvre des modalités d'audits de son système et de son service.

Voir *infra* la partie sur le respect du cadre normatif.

En cas de mise en œuvre d'un SAE, il convient également de :

- *Versement* : Établir pour chaque profil de données, un protocole de versement : prise en charge manuelle (archives non structurées : bureautiques notamment) ou prise en charge automatisée (export à partir de l'application métier au format du SEDA).
- *Politique de sécurité* : bien veiller à s'insérer dans la politique de sécurité informatique de l'établissement. Décliner les fonctionnalités de gestion des droits d'accès, de gestion du cycle de vie des documents archivés, d'intégrité (empreintes), de signature électronique (des bordereaux de prises en charge, des bordereaux d'éliminations), de traçabilité (journaux des cycles de vie, journaux des événements et déterminer leur format et leurs modalités d'exploitation).
- *Stockage* : choisir une stratégie de stockage adaptée à ses besoins, modeste si les volumes sont faibles mais évolutive et avec une couche d'abstraction afin de pouvoir faire évoluer son infrastructure sans changer l'ensemble du système. Bien préciser les modalités de duplication des archives, ainsi que celles de sauvegarde et de redondance. Au ministère de l'intérieur, la DSIC propose aux services une offre de service : hébergement et exploitation d'application
- *Pérennisation* : Élaborer en cas de longue durée de conservation (plus de 10 ans), des stratégies de pérennisation (migrations de supports, migrations de formats).

Voir *infra* la partie sur le respect du cadre normatif

Pour cette étape de définition des spécifications et de mise en œuvre, les acteurs principaux seront l'archiviste et le service métier pour la définition des spécifications

archivistiques (durées de conservation, sort final des données et documents, modèles de description (bordereaux de versement et d'élimination) tandis que la mise en œuvre (développement des fonctionnalités, interfaces, work-flow.) relèvera plutôt des équipes de la DSI : services projets, services réseau et exploitation...On s'appuiera également sur les expertises existantes au sein de l'organisation en matière de sécurité des systèmes d'information ainsi que sur celles afférentes par exemple à la protection des données personnelles.

Voir également le livrable concernant les offres d'archivage électronique intermédiaires présentées lors des travaux du groupe « Mandat DISIC/archivage électronique », dans lequel des précisions sur le coût et les ressources humaines nécessaires à l'exploitation de ces SAE sont données.

Pour tout nouveau projet d'application métier, intégration de l'expertise archives

1. Où s'insère l'expertise archives ?

L'expertise archives **s'insère à toutes les étapes du projet, dès sa conception** :

- Au niveau de la gouvernance des projets : la notion d'archivage au sein des systèmes d'information doit être intégrée dans l'organisation et le suivi de chaque projet.
- Au niveau de l'étude préalable des projets : opportunité et faisabilité.
- Dans la conception du projet : au niveau de l'analyse fonctionnelle et de la conception détaillée.
- Dans les opérations de validation des spécifications fonctionnelles
- Dans la participation à l'analyse des candidatures
- Dans la qualification des recettes
- Concernant la question de la reprise des données : aide à la qualification des données, gestion des données non reprises (purge ou archivage définitif)
- Au niveau de la mise en production, afin d'accompagner la mise en œuvre du scénario d'archivage défini.
- Lors du retrait de service du système d'information, afin d'appliquer les exigences fonctionnelles liées au projet et la réglementation archivistique en vigueur quant à la destruction des données.

Au ministère de la défense, il est prévu que dès la conception d'un projet informatique, le futur outil de gestion des systèmes d'information et de communication, qui rentrera en production au printemps 2012, intègre un onglet « archivage ». Lui sera associé systématiquement une fiche d'expression rationnelle des objectifs d'archivage (FEROA), en analogie avec la fiche relative aux objectifs de sécurité (FEROS).

2. Focus sur chaque étape du projet

a) Étude d'opportunité et étude de faisabilité

Lors de cette phase, il conviendra **d'identifier précisément la production** :

- compréhension du contexte : flux de la production documentaire, circulation et circuits de diffusion ;
- connaissance des producteurs des documents : statut, rôle et responsabilités,

missions et activités et leur périmètre ;
- connaissance des documents produits, usages et besoins d'accès à court, moyen ou long terme, typologie documentaire.

Pour ce faire, on pourra lancer un diagnostic fondé sur un petit nombre de **questions essentielles**[4] **telles que :**

- Le projet est-il lancé pour répondre à une exigence réglementaire ?
- Le projet consiste-t-il en la dématérialisation d'une procédure existante ?
- Le projet a-t-il pour objectif de capturer ou de créer des documents à valeur de preuve ?
- Le projet prévoit-il la numérisation de données papier existantes ?
- l'alimentation du système sera-t-elle réalisée par les agents de l'organisation ? si non, par qui ? un prestataire extérieur ?
- Le système sera-t-il interfacé avec d'autres systèmes internes ou externes à l'organisation ?
- Les processus métier concernés disposent-t-ils de règles relatives aux durées de conservation et au sort final des documents ?
- Le système contiendra-t-il des données personnelles ?
- Le projet comprend-il un volet de reprise de données existantes dans un ou plusieurs systèmes précédents ? si oui, cette reprise sera-t-elle intégrale ou partielle ?
- Les données seront-elles stockées au sein de l'organisation ?

A partir des réponses, on pourra conduire, du point de vue l'archivage, **une analyse de risques** : juridique, financier, administratif et stratégique.

b) Analyse fonctionnelle

On pourra retenir les points essentiels suivants à étudier :

- Production et capture des documents dans leur contexte (définition des métadonnées, formats des fichiers).
- Éléments de traçabilité relatifs au cycle de vie de l'information.
- Fonctionnalités liées à la conservation et à la gestion des sorts finaux des documents au sein du SI.
- Fonctionnalités d'import, d'export (définition du format d'échange SEDA).
- Reprise éventuelle de données.

c) Analyse technique

L'expertise portera sur les conséquences liées aux durées de conservation en termes de volumétrie et de performance du SI.

d) Participation de l'expert archives à la recette fonctionnelle

pour les fonctionnalités liées au cycle de vie et à l'archivage.

e) Reprise des données ou fin de vie du produit

Lors du retrait de service du système d'information, il conviendra de s'assurer que tous les documents et données non repris dans un autre système aient fait l'objet d'une opération

4 Très largement inspiré du diagnostic éclair figurant dans la conduite de projet Mozart du ministère de la culture et de la communication, où la partie archives est incluse.

d'archivage (archivage définitif dans le service d'archives public compétent ou purge avec l'autorisation de l'administration des archives).

Dans le cas où le système d'information a été géré par un prestataire (hébergement, saisie), on portera une attention particulière à la purge des copies et sauvegardes. Si le prestataire doit garder une trace des opérations effectuées pour ses propres contrôles administratifs, on veillera particulièrement à ce qu'il ne conserve aucune donnée personnelle ou à ce que le lien entre les données personnelles et les opérations soit rompu logiquement.

Au ministère de la défense, conséquence de la politique de rationalisation du parc applicatif, le ministère prononce actuellement le retrait de service d'une centaine d'applications par an s'agissant des SIAG. Formalisée par une procédure depuis avril 2011, le service historique de la défense est systématiquement consulté afin d'établir un diagnostic archivistique des données contenues dans les SI à retirer et d'évaluer ainsi les besoins en archivage.

3. Vie de l'application

Après la mise en œuvre du scénario d'archivage retenu (voir ci dessus), durant la vie de l'application, l'expertise archives sera indispensable :

- lors des premières élimination et versements effectifs ;
- lors des évolutions de l'application ;
- en cas d'évolution des contrats et conventions de service entre producteurs, services informatiques, services d'archives (niveaux de service, fréquence des versements, types de formats acceptés..) ;
- en cas de changements quant aux responsabilités d'archivage
- en cas d'évolution de la réglementation (durées de conservation, évolution de la réglementation CNIL...) ;
- en cas d'évolution des modes de sélection des archives (modes d'échantillonnage qui évoluent…) ;
- lors de la fin de vie de l'application (reprise des données).

*Voir la présentation du travail effectué par **le ministère de la culture et de la communication,** relative à l'intégration de l'expertise archive dans la conduite de projet (Mozart)[5] adoptée par le ministère.*

5 La méthode comprend une définition des rôles et des responsabilités, la comitologie (objectifs et participants, fréquence des réunions), le cycle de vie des projets avec pour chaque étape, la répartition des rôles et des responsabilités, des indicateurs, un glossaire.

4. Intégration de l'expertise archives au cycle de vie "projet": vision synoptique

<table>
<tr><td rowspan="2">Phases projet</td><td colspan="2">Étude préalable</td><td colspan="4">Projet</td><td colspan="2">Production</td><td>Fin de vie</td></tr>
<tr><td>Étude d'opportunité</td><td>Étude de faisabilité</td><td>Analyse fonctionnelle</td><td>Conception détaillée</td><td>Réalisation</td><td>Qualification recette</td><td>Mise en production</td><td>Maintenance</td><td>Retrait du produit</td></tr>
<tr><td rowspan="4">Phase intégration de l'expertise archives dans les projets</td><td>Identification du besoin</td><td>Analyse de risque / traçabilité</td><td colspan="2">Définition du scénario d'archivage</td><td></td><td></td><td colspan="2">Scénario d'archivage appliqué</td><td>Exigences fonctionnelles liées au retrait du projet</td></tr>
<tr><td rowspan="2"></td><td rowspan="3"></td><td rowspan="3" colspan="2"></td><td>Validation des spécifications fonctionnelles générales et détaillées</td><td>Audit de conformité</td><td colspan="2" rowspan="2">Application des exigences liées à la conservation ou à la destruction des données</td><td>- Destruction du produit</td></tr>
<tr><td>- Archivage du produit</td></tr>
<tr><td>Identification de la production</td><td>Mise en œuvre des exigences fonctionnelles et réglementaires</td><td>Recette des exigences définies.

Évaluation.</td><td colspan="2"></td><td></td></tr>
<tr><td>Livrables en sortie de phase</td><td colspan="2">Fiche "pré-étude" SI, intégrant l'expertise archivage des données. Besoins d'archivage SI.</td><td colspan="2">SFD intégrant les exigences d'archivage.</td><td>- Documentation relative à la mise en œuvre des exigences d'archivage.

- Tests.</td><td>- Rapport de qualification / recette (intégrant les aspects archivage)

- PV de recette.

- Rapport d'audit archivage.</td><td colspan="2">Convention de service (intégrant les aspects archivage)</td><td>Rapport de retrait intégrant les exigences archivistiques.</td></tr>
</table>

Trois cas d'usage

L'accompagnement des services dans la gestion de leurs fichiers bureautiques et de leurs serveurs partagés

Vous trouverez dans cette fiche quelques conseils méthodologiques pour la maîtrise de la production bureautique non structurée d'une organisation.

Plan de la fiche :

- Mise en œuvre du plan de classement dans un univers électronique
- Le nommage des dossiers et des fichiers
- Les modèles de document

1. Mise en œuvre du plan de classement dans un univers bureautique

L'arborescence : dans le monde électronique, le plan de classement se traduit par un système de dossiers et sous-dossiers, partagé par tout ou partie des agents d'une même entité. Le plan de classement peut être mis en place au sein d'un système de gestion de documents (GED, Sharepoint, etc.) mais, à défaut d'une telle solution, peut être directement implanté sur un serveur partagé.

Fichiers partagés, travail personnel et droits d'accès : l'une des difficultés dans la mise en œuvre d'un plan de classement est le partage de l'information. Les réponses à cela sont de plusieurs ordres : le serveur est le lieu unique de stockage des documents du service, à l'exclusion des disques durs des agents, clés USB, disques externes, etc. A l'inverse, on ne dépose aucun document personnel sur le serveur.

Politique de droits d'accès : elle doit être définie avec le service informatique. Elle doit se gérer à relativement haut niveau pour limiter le nombre de profils différents et faciliter la gestion des droits. Elle est beaucoup plus simple dans un système de GED mais il est possible également de la mettre en œuvre dans un système de serveur partagé.

Conseils méthodologiques pour l'élaboration du plan de classement :

- dans la mesure du possible, l'élaboration du plan de classement est un travail collaboratif, même si l'archiviste peut venir avec une proposition. Il est absolument indispensable que les utilisateurs participent activement à ce plan et se l'approprient, afin de pourvoir s'y retrouver facilement. Il est par contre nécessaire de faire valider le plan de classement par l'archiviste ;
- le plan de classement et ses règles de fonctionnement (droit d'accès, nommage des fichiers cf. ci-dessous, etc.) doivent être résumées dans un document validé par la direction, diffusé officiellement et revu régulièrement ;
- il convient de désigner un responsable du maintien du plan de classement, chargé de vérifier le respect des règles (et de supprimer systématiquement les créations anarchiques de dossiers, après avertissement des créateurs du dossier) ;

Qu'est-ce qu'un bon plan de classement ?

- une hiérarchie de répertoires et de dossiers allant du général au particulier ;

- une organisation thématique (aux niveaux supérieurs puis chronologique, numérique ou
- alphabétique (aux niveaux inférieurs) qui reflète les fonctions de l'entité et non son organigramme.

Un plan de classement doit viser à une certaine pérennité ;

- englobe tous les documents de l'unité de travail (direction, sous-direction, service, bureau, agent) ;
- des intitulés de répertoires, de fichiers et de dossiers intelligibles par tous. Le travail sur la formulation est très important, car les termes choisis doivent être explicites. Il est possible d'aller jusqu'à la codification ;
- on ne doit y trouver aucun dossier de type « divers » ou « affaires générales » ou « à classer » ;
- composé de trois à quatre niveaux hiérarchiques de répertoires et de dossiers, dont les trois premiers sont fixes et verrouillés et le dernier modifiable par les agents.
- dans l'idéal, le plan de classement permet également à l'archiviste de repérer immédiatement les « branches » à archiver / éliminer.

2. Le nommage des dossiers et des fichiers

Le nommage est le corollaire du plan de classement et répond aux mêmes besoins : retrouver et partager facilement l'information. Comme pour le plan de classement, la rédaction de règles est une chose, la mise en œuvre de ces règles en est une autre.

L'intérêt du nommage est d'associer au fichier, dès sa création, un certain nombre de métadonnées. La plupart des logiciels de traitement de texte implémentent des métadonnées automatiques (auteur, date) et permettent de faire des compléments manuels (dans OOO, onglet « fichier », « propriétés ») mais qui les remplit ? Mettre les métadonnées dans le titre est par conséquent un palliatif.

Avant de démarrer, ne pas oublier :

- le nommage concerne aussi bien les dossiers que les fichiers. Cependant, si la philosophie est la même, les règles peuvent être légèrement différentes (inutile de mentionner le service producteur en tête par exemple, ou les mentions de validation) ;
- la concision est le maître-mot des pratiques de nommage. Il ne faut jamais oublier que les chemins d'accès aux fichiers ont un nombre de caractères limité.

<u>Méthodologie</u>

- Elle est sensiblement la même que pour le plan de classement : le travail collaboratif est indispensable pour une meilleure appropriation, il convient d'élaborer un récapitulatif des règles dans un document unique.
- Ne pas oublier que la mise en œuvre du nommage peut aussi intervenir postérieurement à la création des fichiers par des outils de renommage. Mais cela ne pourra se faire que de façon sérielle. Ainsi, des documents mal nommés à l'origine ne pourront pas être renommés par lots.

<u>Règles à respecter</u>

- Emploi des caractères : pas de caractères accentués, d'espaces, privilégier les tirets underscore, etc. → pas de caractères spéciaux.
- Les éléments à indiquer : le sigle du service producteur, le sujet, la typologie du document, la date, la version. L'ordre de ces éléments peut varier.

NB : concernant la date, il ne faut pas se fier à la date automatique du fichier, qui peut être faussée par un mauvais paramétrage d'une part et change lorsque l'on copie-colle le fichier

d'autre part.

- Les abréviations ne doivent pas être multipliées et sont récapitulées au sein d'un document facilement accessible à tous. De même, il faut tenter de normaliser le plus possible les libellés. Ces règles peuvent en particulier concerner le nommage du service producteur (un sigle est attribué à chaque service et récapitulé dans ce fichier autorité) et la typologie des documents qu'il faut normaliser le plus possible en adoptant un référentiel commun (ex : CR pour compte-rendu de réunion permet de "gagner des caractères" et permet d'éviter d'avoir tantôt Crendu, reunion, CRréunion, etc.).
- Attention au versionning des documents, absolument indispensable pour l'archivage (on n'archive que la version validée[6] et pas les versions intermédiaires), même si, en réalité, ce n'est pas évident à gérer en dehors d'un système de GED.

3. Les modèles de document

Le travail de nommage des fichiers, *i. e.* d'enrichissement de métadonnées, peut aller jusqu'à la rédaction de modèles de documents (modèle de compte-rendu, modèle de lettre, modèle de note, etc.) à employer par toute la collectivité concernée. Dans la mesure du possible, il faut s'efforcer de faire tenir toutes ces informations sur une page.

Autant que faire se peut, il peut être intéressant d'inclure dans ces modèles des cartouches avec un certain nombre de métadonnées telles que : nom du rédacteur, nom des relecteurs, dates (création/modification/validation), version, titre du document, etc.

Attention, les modèles ne dispensent pas de règles de nommage, qui permettent de voir au premier coup d'œil de quoi traite le fichier en question.

Voir la plaquette élaborée par le ministère en charge de l'Écologie pour la direction de la sécurité et de la circulation**, opération VIE** (Valorisation de l'Information Electronique) dont les enjeux suivants sont pointés : *partager, classer, qualifier, protéger, diffuser, conserver*.

Au ministère de l'intérieur, la DSIC a élaboré un : « Modèle document standard texte : guide d'utilisation » très complet et portant notamment sur les documents produits dans le cadre de la conduite d'un projet informatique.

6 Dans le cas de dossiers importants dont il faut archiver des versions intermédiaires, on peut envisager d'adopter un système de numérotation informatique V1.0, V1.1, V1.2 et V2 quand changements importants. Ainsi, on ne conserve que les V1, V2, etc. et pas les « sous-versions » intermédiaires.

L'accompagnement des services dans l'archivage de leurs courriels

Cette fiche a pour but de donner les différentes possibilités d'archivage des courriels pour une organisation.

Plan de la fiche :

- Contexte juridique
- Archivages des courriels : les approches existantes
- Sélection des courriers électroniques
- Mise en œuvre concrète de solutions d'archivage

*Voir sur le site des archives de France l'instruction DITN/RES/2009/007 du 3 juin 2009 : **http://www.archivesdefrance.culture.gouv.fr/static/2822**.*

*Directives pour la gestion et l'archivage numériques des courriers électroniques Version 1.0 (juillet 2008) : **http://www.archivesdefrance.culture.gouv.fr/static/2823** rédigées par Sébastien Soyez, Archives générales du Royaume et Archives de l'État dans les provinces*

Qu'est-ce que le mail ?

Un e-mail, c'est d'abord un protocole technique qui sert à envoyer une information et qui peut correspondre à des usages variés, le message pouvant contenir en lui-même une information à forte valeur ou au contraire servir seulement à transmettre un document joint ; il peut être le support d'une transaction importante ou, souvent, d'un échange d'information éphémère (plus ou moins équivalent à la conversation téléphonique).

L'e-mail se compose d'un en-tête, d'un corps où se trouve le contenu de l'information et éventuellement de pièces jointes.

Ainsi, dans les directives, des notions sont décrites et précisées : standard MIME[7], protocole de transfert SMTP[8], RFC 2821 et 2822[9], serveur SMTP (pour les serveurs du courrier sortant) et POP ou IMAP selon le protocole utilisé (pour les serveurs du courrier entrant), client de messagerie[10], webmail[11].

1. Contexte juridique

Le code du Patrimoine pour les courriels considérés comme des archives publiques, mais également l'article 8 de la Convention européenne de droits de l'homme de 1955 et plus généralement des législations afférentes aux traitements de données à caractère

7 Multipurpose Internet Mail Extension est un standard internet qui étend le format de données des courriels pour supporter des textes en différents codages de caractères autres que l'ASCII, des contenus non textuels et multiples, des informations d'en-tête en d'autres codages que l'ASCII.

8 Simple Mail Transfer Protocol est un protocole de communication utilisé pour le transfert du courrier électronique vers les serveurs de messagerie électronique.

9 Request for Comment 2821 et 2822. Le premier est un standard relatif au protocole de base pour le transport du courrier électronique. Le second est un standard spécifiant la syntaxe des messages textuels échangés dans le cadre de l'utilisation des courriers électroniques par des utilisateurs informatiques.

10 La relève et l'envoi du courrier se fait via un logiciel appelé MUA (Mail User Agent) qui, lorsqu'il est installé sur le système de l'utilisateur, est appelé client de messagerie.

11 Dans ce cas, il s'agit d'une interface web permettant de dialoguer avec le serveur du courrier entrant.

personnel[12], ainsi qu'à la protection du droit d'auteur ou encore à la fraude informatique[13].

Ceci amène l'employeur à respecter en la matière six principes de base : la transparence (toute personne de l'institution doit être informée du fait qu'un système d'archivage a été mis en place, le cas échéant automatisé), la finalité (doit être conforme à la législation sur les archives), la proportionnalité (seuls certains courriels sont conservés pour une certaine durée), la sécurité et la confidentialité et enfin le droit d'accès et de modification.

Ce qui rend difficile la collecte des mails : côté multiple et multifonctionnel de la messagerie (avec des usages parfois « déviants » car non régulés) ; difficultés techniques de la capture. Question des pièces jointes, zippées, des pièces jointes encapsulées dans des pièces jointes…, la question des doublons (même si des outils de dédoublonnage existent) ; pression de la masse et difficulté de la sélection (qui devrait se faire par des moyens simples et adaptés aux besoins de l'utilisateur mais qui n'existent pas aujourd'hui).

2. Archivages des courriels : les approches existantes

Les possibilités d'archivage des boites emails peuvent globalement se résumer comme suit :

- ne rien faire ;
- l'approche message par message – où les utilisateurs sont encouragés à déplacer les courriels significatifs hors de leur client de messagerie et de les joindre avec d'autres documents provenant de la même activité (c'est l'approche traditionnelle du *records management*) ;
- l'approche compte de messagerie par compte de messagerie – où certains individus au sein de l'organisation sont sélectionnés comme ayant un rôle particulièrement important, et la totalité de leur compte de messagerie est préservée ;
- l'approche système de messagerie global lorsque l'organisme traite l'ensemble de son système de courrier électronique comme une agrégation et applique une règle de tri et de conservation à l'ensemble du système.

Que faire ?

- Une analyse fine et préalable est indispensable : qu'y a-t-il dans les boîtes mails de l'entité concernée ? Est-il utile d'en prévoir l'archivage ? Analyse du risque, calcul du rapport « temps passé, intérêt de l'archivage ».
- Proposition de combiner les approches 2 et 3 en sélectionnant d'une part des boîtes à archiver dans leur globalité et, pour les autres, en préconisant d'enregistrer hors du système de messagerie (et donc dans le plan de classement!) les courriels les plus importants/engageants (quelques dizaines par an tout au plus certainement pour toutes les boîtes hors direction), voire de les imprimer pour alimenter les dossiers papier si ceux-ci sont encore tenus. Il s'agit alors d'un pis-aller dans la mesure où une partie des métadonnées pourtant essentielles à la conservation de ce courrier, qu'on trouve par exemple en affichant le code source du message, ou ses propriétés, ne peuvent être imprimées.
- Développer des réflexes de bon sens : mettre en place l'usage de boîtes aux lettres professionnelles collectives (par service, ou d'une manière plus pérenne, par grandes

12 Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, modifiée par la loi n° 2004-801 du 6 août 2004 relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel.

13 Code de la propriété intellectuelle, partie législative, première partie sur la propriété littéraire et artistique, et Code pénal, titre II (Des autres atteintes aux biens), chapitre III (Des atteintes aux systèmes de traitement automatisé des données).

activités) aux côtés des boîtes aux lettres personnelles[14], procéder à des épurations très régulières (immédiates pour tous les messages à faible contenu informatif ou à contenu de court terme), concernant les pièces jointes, s'il s'agit de simples documents bureautiques, elles doivent être limitées aux envois hors institution, ou, au sein de l'institution, aux destinataires avec lesquels l'expéditeur ne partage pas de répertoire (dossier). A contrario, dès lors que les usagers disposent d'un dossier « Documents partagés » (dossier commun à un service ou une direction de l'institution, ou partagé entre plusieurs d'entre elles), il faut favoriser le renvoi vers le document sous la forme d'un lien hypertexte, dans le corps du message. On évite ainsi la création d'une nouvelle copie et le volume de la messagerie en est réduit.

- Apprendre à se servir des paramètres natifs des messageries (ex : mettre une règle pour éliminer tous les dossiers de plus de 15 jours du dossier « corbeille », paramétrer des règles de classement automatique), bien titrer ses messages, etc.
- Sensibiliser et contraindre : quotas volumétriques stricts et faibles (à voir en partenariat avec la DSI), effacement systématique de la boîte au départ de l'agent, etc.

3. Sélection des courriers électroniques

Ne mettre en œuvre une politique d'archivage que pour les messages professionnels qui reflètent les activités de l'institution, formalisent les différentes étapes d'une tâche, d'une décision, d'une procédure, dans le cadre des missions liées à l'activité de l'institution. Dans ce cas, les règles de conservation sont identiques à celles qui existent déjà pour la correspondance papier et doivent apparaître en tant que telles dans les tableaux de gestion de l'institution.

À l'inverse un message, même professionnel, n'est pas considéré comme une archive, dès lors qu'il s'agit de messages officieux, de brouillons, d'informations de service, d'informations de type « liste de diffusion » : il doit être intégré dans les tableaux de gestion mais sa destination finale est l'élimination, une fois la durée d'utilité administrative expirée ;

Quant aux messages à caractère strictement privé, ils ne rentrent pas dans le champ de l'archivage et sont détruits par les utilisateurs lorsque ceux-ci le souhaitent. En tout état de cause, la réflexion sur l'évaluation doit uniquement porter sur les répertoires « courriers entrants » et « courriers envoyés ».

La répartition des responsabilités dans l'archivage des courriers électroniques est la suivante :

- Concernant la sélection, c'est à l'expéditeur ou au destinataire d'être responsable du tri de ses courriers électroniques (avec une évaluation de leur pertinence) en respectant les règles établies par les archivistes pour ce qui concerne les courriers professionnels. C'est par conséquent à chaque institution d'établir et d'expliquer des règles et des procédures pour la sélection, en conformité avec la législation en vigueur concernant les archives publiques.
- Pour ce qui est de l'enregistrement et de la conservation, interviennent les archivistes ainsi que les services informatiques.

14 Ceci suppose d'une part, que la gestion de la boîte (relever les messages, les affecter, les trier, les exporter…) soit clairement confiée à un agent référent, et d'autre part, que le type de messages à transiter par ce type de boîte soit formellement défini, en interne comme dans les échanges avec les correspondants externes…

4. Mise en œuvre concrète de solutions d'archivage

Il convient de ne pas chercher à conserver les messages et pièces jointes au sein du logiciel de messagerie, solution qui pose des problèmes évidents de gestion et de pérennité : format d'export propriétaire de certains logiciels, pertes d'informations pouvant résulter de fonctionnalités offertes par certains logiciels de compression, accès réduit au seul utilisateur de sa messagerie, problèmes rapides de volumétrie et de stockage. La seule solution consiste par conséquent à exporter les records à conserver dans un système centralisé, à accès partagé suivant les règles de l'institution et offrant des garanties de sécurité. En outre seule cette solution permettra de créer des liens entre un courriel, ses pièces jointes et d'autres documents ou courriels ayant trait à la même affaire.

Concernant la mise en œuvre concrète, elle ne pourra idéalement se faire que si certains développements informatiques sont réalisés au sein des clients de messagerie[15] et si existe en aval un système centralisé qui permet de recevoir les courriels destinés à être conservés durant un certain temps.

En amont, ces développements permettent par exemple, pour les records qu'on souhaite conserver, d'encapsuler l'ensemble des métadonnées[16] avec le message lui-même de manière à n'en perdre aucune[17]. La plupart de ces métadonnées sont produites automatiquement mais l'outil devrait permettre aux utilisateurs s'ils le souhaitent d'ajouter des métadonnées avant cette encapsulation pour les messages qu'ils créent. Ces développements pourront également permettre, lors de cette encapsulation, de conserver les répertoires (plan de classement) dans lesquels les utilisateurs auront éventuellement rangé les messages durant leur instruction, et permettre le traitement par lot de messages.

Il est recommandé que ces manipulations puissent se faire très rapidement après la réception ou l'envoi de ces messages de manière à ce que ces derniers et leurs pièces jointes puissent être alors exportés correctement dans un système centralisé dans lequel sera pris en charge la gestion de leur cycle de vie et leur classification.

Concernant ce système centralisé, deux modes d'action sont distinguées : un mode intermédiaire qui consiste à utiliser un système centralisé de classification (qui s'apparente à un système de GED) ou la solution plus pérenne d'un système d'archivage électronique (SAE). La solution la plus satisfaisante est bien évidemment la seconde, car le premier système, s'il permet de classer les courriels et permettre des recherches efficaces (liens entre des courriels entre eux et entre des documents numériques portant sur la même thématique), ne permet pas forcément de gérer le cycle de vie (intégration des DUA et sorts finaux fixés dans les tableaux de gestion) et de sécuriser l'information : coffre fort numérique, conversion des formats, duplication de l'information.

En tout état de cause, des programmes informatiques devront permettre l'export automatique des courriers électroniques dans leur intégralité (métadonnées, en-têtes, corps du message, pièces jointes) au sein de ce système centralisé. Les courriers électroniques suivront alors les règles de gestion (DUA et sorts finaux) fixées au niveau des dossiers et répertoires du plan de classement dans lesquels ils seront intégrés.

15 Voir le logiciel téléchargeable en test réalisé par les Archives nationales des Pays-Bas **http://www.digitaleduurzaamheid.nl/**

16 La plupart sont produites automatiquement. On peut distinguer les métadonnées d'identification et les métadonnées d'intégrité.

17 Avec les fonctionnalités par défaut d'un logiciel de messagerie, par exemple Mozilla Thunderbird, si l'utilisateur enregistre sur son disque dur un courrier au format .eml, l'ensemble des métadonnées sources seront conservées, mais si l'utilisateur choisit d'enregistrer le courriel au format .txt, et qu'il n'a pas pris la précaution d'afficher le code source avant de procéder à l'enregistrement, la plupart des métadonnées seront perdues.

Par ailleurs, concernant les formats de conservation, certains sont à proscrire (MSG ou OFT provenant de MS Outlook, ou encore HTML), tandis que peuvent être recommandés des formats comme EML[18], RTF ou bien évidemment TXT[19]. Ceci étant, le format recommandé par excellence est le format XML particulièrement adapté pour la conservation structurée des messages et de leurs métadonnées et ce, d'autant qu'il peut être généré à partir de certains de ces formats de sauvegarde (TXT, EML ou RTF). Cette conversion doit pouvoir être permise par lot au moment de l'intégration dans l'outil de GED/SAE.

Quant aux pièces jointes aux messages, que l'on recommande d'enregistrer à part étant donné les formats très divers dans lesquels elles sont produites, certaines d'entre elles nécessitent, comme pour tout système d'archivage électronique, des conversions vers des formats ouverts et reposant si possible sur des normes.

En définitive, sans politique globale et outils de records management incluant la gestion des courriels, aucune approche n'est satisfaisante.

18 La totalité du courrier électronique (entêtes, corps et pièces jointes) est contenue dans un unique document texte, suivant la structure de la norme RFC 2822 et par conséquent, l'usage d'aucun format propriétaire ou logiciel propriétaire n'est requis pour exploiter de tels documents.

19 Fichier texte brut.

Intégrer la gestion du cycle de vie dans un projet de numérisation

Il s'agit ici de présenter les étapes d'un projet de numérisation de documents papier en vue de leur archivage sous format électronique, avant intégration dans une gestion électronique de documents (GED), gestion électronique de courriers (GEC) ou une application métier.

Plan de la fiche :

- Définir le projet de numérisation
- Les étapes du projet
- Choix techniques
- Contrôles
- Intégrer les fichiers nativement numériques ?

1. Définir le projet de numérisation

1.1. Définir le périmètre de la numérisation

Il s'agit pour l'institution de choisir d'intégrer tout ou partie des dossiers papier qu'elle gère au projet de numérisation.

Il est ainsi conseillé dans tous les cas de ne pas étendre le périmètre d'un projet de numérisation des dossiers aux dossiers dont la durée d'utilité administrative (DUA) est échue : dans ce cas, il convient de se rapprocher des missions et services d'archives en vue de préparer les versements et éliminations conformément à la réglementation en vigueur.

1.2. Choix des dossiers à numériser

Choix des pièces à numériser, soit une réflexion sur :

–la structure du dossier numérisé, qui permet de différencier les documents numérisés entre eux, pour les retrouver rapidement,

–les pièces à numériser dans les dossiers *:* par exemple, il est possible de ne pas numériser l'ensemble des pièces papier constituant un dossier dans la mesure où pour certains d'entre eux, l'information utile est saisie directement et intégralement dans le système d'information.

1.3. Réflexion sur l'organisation de la numérisation

La numérisation du flux des documents à l'arrivée et la numérisation du « stock » de dossiers actifs sont à distinguer dans leur mise en œuvre : la liste des pièces numérisées pourra être différente, le plan de classement adopté pour la numérisation du flux pourra être plus élaboré et complexe que celui adopté pour la numérisation du stock, etc.

La reprise du stock nécessitera en outre la mise en œuvre de procédures « exceptionnelles », voire d'une externalisation ; a contrario, l'organisation de la numérisation du flux doit trouver sa place dans les procédures et l'organigramme de l'organisme concerné

Dans tous les cas, il convient d'organiser et d'écrire la procédure de numérisation du flux, et de définir précisément les responsabilités et rôles des différents acteurs intervenant dans la chaîne de numérisation.

1.4. La création des dossiers dans le système d'information

Différent suivant qu'il s'agit du stock ou du flux avec l'articulation entre le système d'information et la GED interne, externe ou intégrée.

2. Les étapes du projet

2.1. Déterminer la valeur juridique des documents

En effet, les administrations posent fréquemment la question de savoir si elles ont la possibilité d'éliminer des documents originaux papier après leur numérisation. Les Archives de France ont, sur ce sujet, rédigé l'instruction **DITN/DPACI/RES/2005/001** du 14 janvier 2005, intitulée *Modalités de délivrance du visa d'élimination des documents papiers transférés sur support numérique ou micrographique*.

Ces éliminations des originaux papier ne peuvent se faire en tout état de cause que :

- sur des documents à terme éliminables ;
- avec le visa réglementaire de l'administration des archives ;
- si le processus de numérisation a été conduit dans les règles de l'art (voir ci-dessous la fiche pratique sur les projets de numérisation et de GED).

2.2. Déterminer les durées d'utilité administrative des dossiers et des pièces

2.3. Déterminer le sort final des dossiers et des pièces

2.4. Établir le plan de classement des pièces constitutives du dossier et la liste des métadonnées :

Une fois la typologie des dossiers/documents papier/numériques effectuée, il convient de déterminer quelle sera la structuration choisie (arborescence ou classification). C'est l'objet du plan de classement, qui constitue une table des matières du dossier électronique et facilite son exploitation.

Il est conseillé de ne pas dépasser plus de trois niveaux de hiérarchisation. Il est également conseillé de rassembler dans un même sous dossier les documents ayant la même DUA et le même sort final.

Les métadonnées qui permettront de gérer et de retrouver les dossiers doivent être soigneusement définies :

- l'identifiant du dossier/du document numérique,
- les index (clés de recherche et d'accès) générés par l'outil de GED soit à partir de l'application métier pré-existante, soit à partir de l'outil de GED indépendamment de cette application métier, en fonction de la solution technique adoptée par l'éditeur,
- les attributs liés aux dossiers/documents relatifs aux délais de conservation et au sort final,
- les attributs liés aux dossiers/documents relatifs aux droits d'accès et aux habilitations, en fonction des différentes fonctionnalités de l'application de GED et du cycle de vie des documents/dossiers,
- les métadonnées techniques[20] internes et externes produites automatiquement

20 Direction des archives de France, *Ecrire un cahier des charges pour la numérisation : guide technique*, 2008, partie 3.12.3.

(génération par le logiciel d'exploitation du scanner par exemple) : caractéristiques techniques de l'image (dimension), de la prise de vue[21] (opérateur de numérisation, n° de lot de numérisation[22]).

Ces métadonnées servent d'une part à sécuriser le processus de numérisation en permettant de garder la trace des opérations conduites, et d'autre part à sécuriser la conservation des fichiers produits grâce aux informations sur les prises de vue et le matériel utilisé, ainsi que sur les images.

Ces métadonnées peuvent être internes (incluses directement dans l'entête des fichiers) ou externes à l'outil de GED. Il conviendra par conséquent de s'assurer, auprès du prestataire de numérisation ou des sociétés commercialisant les dispositifs de numérisation, que les métadonnées techniques qu'on souhaite récupérer sont bien automatiquement générées et enregistrées.

Certains prestataires permettent par exemple l'enregistrement et la conservation d'un journal des événements.

2.5. Conduire et organiser les opérations de numérisation

La numérisation doit être conduite dans les règles de l'art, notamment si les services souhaitent procéder à l'élimination des originaux papier, afin de pouvoir attester de la qualité du travail effectué, et pouvoir si nécessaire prouver au juge, en cas de contentieux, qu'il est vraisemblable que la copie numérique a bien été réalisée dans de bonnes conditions et qu'il s'agit bien d'une copie fidèle à l'original papier (disposition de l'instruction **DITN/DPACI/RES/2005/001** du 14 janvier 2005).

Ceci passe dans un premier temps par la détermination des responsabilités respectives :

- désignation officielle d'un responsable du projet de numérisation,
- désignation des opérateurs,
- définition des responsabilités respectives des opérateurs et du responsable du projet de numérisation, que ce soit en matière de numérisation, d'indexation, de vérification, de mises à jour, ou encore d'éliminations,
- coordination avec les différents partenaires et définition de la responsabilité de chacun en matière de conservation de l'information.

Il est recommandé qu'un document officiel (attestation, note officielle, document contractuel) soit établi par le responsable du projet de numérisation qui, de fait, autorise la numérisation des originaux papier.

À retenir : il est primordial de bien assurer la conservation des métadonnées et de rédiger et tenir à jour un **manuel qualité des procédures.** Ce dernier peut par ailleurs permettre de couvrir une grande partie des métadonnées qui ne seraient pas automatiquement générées par les logiciels de scan et de GED.

Une évaluation soigneuse des documents à numériser doit être menée et consignée :

- identification des dossiers à intégrer au système de GED :
  - nombre de dossiers concernés : un récolement précis doit être réalisé (listing), afin d'identifier les dossiers à numériser/en cours de numérisation/numérisés,
  - nombre moyen de documents par dossier.
- travail de préparation des dossiers eux-mêmes :
  - (re)-classement des documents en sous-dossiers, dossiers, suivant le plan de

21 Information à fournir uniquement en cas d'externalisation de la prestation de numérisation.

22 Information à fournir uniquement en cas d'externalisation de la prestation de numérisation.

classement pré-établi (le cas échéant),
- identification précise et sélection des documents qu'il convient de numériser, à ranger dans une pochette spécifique, placée sur le dessus du dossier, par exemple ;
- repérage de formats atypiques (documents se présentant sous formes de volets par exemple), afin de photocopier ce qui doit l'être avant de commencer la numérisation proprement dite ;
- analyse de l'état des documents (déchirés, froissés, présence d'une forte poussière) et des encres (encre pâles…) ;
- présence ou non d'agrafes, de trombones, d'élastiques… ;
- actions spécifiques à conduire en cas de présence de documents couleurs[23] .

À retenir **:** chacune de ces caractéristiques aura des incidences sur la durée de l'opération et sur son coût. Ce travail préparatoire doit être évalué et budgétisé, au même titre que les opérations de numérisation ou l'architecture technique de la GED par exemple. Il constitue en effet la première étape du projet.

Pour chacune de ces caractéristiques, celui qui doit effectuer la numérisation doit proposer une réponse adaptée.

3. Choix techniques

Les outils de numérisation des documents doivent être décrits en détail dans un dossier de description technique.

3.1. Adopter des formats ouverts

Les formats d'images choisis doivent faire l'objet de spécifications publiques et si possible reposer sur une norme ou un standard, dans le but de garantir l'interopérabilité des systèmes et la pérennisation des données.

Il faut éviter dans la mesure du possible de compresser les images après numérisation.

Les méthodes de compression avec pertes ne doivent pas être utilisées sur des documents de type noir et blanc dit « de bureau » contenant principalement des textes.

Pour ce type de document, une cible d'essai peut être utilisée (ISO 12653-1 et 2).

Les systèmes informatiques doivent fournir des moyens de contrôle des fichiers contenant les images après compression.

Quel que soit le choix, les techniques de compression mises en œuvre doivent être normalisées et leurs spécifications accessibles librement. Le dossier de description technique doit mentionner le référentiel normatif associé.

3.2. Définir des règles de nommage[24] des fichiers numérisés ainsi que des dossiers numériques

3.3. Privilégier la fidélité au document (couleur ou niveaux de gris plutôt que noir et blanc)

3.4. Limiter les traitements sur les images

Les dispositifs de traitement des images lorsqu'ils existent doivent être décrits en détail

23 Pour une photographie ou un document coloré, il conviendra de choisir un taux de résolution suffisant (par exemple 300 DPI), afin d'éviter qu'une résolution trop faible rende le document illisible.

24 Se reporter au § 3.12.2 du guide technique Écrire *un cahier des charges de numérisation du patrimoine* http://www.archivesdefrance.culture.gouv.fr/static/1308

et utilisés avec beaucoup de précaution puisqu'ils interviennent sur la notion de fidélité de l'image électronique par rapport au document d'origine.

En particulier, une procédure de passage d'une image en niveaux de gris ou couleur à une image en noir et blanc doit avoir été testée en détail et validée préalablement à sa mise en œuvre.

Les logiciels qui ont pour but de supprimer les petites taches peuvent conduire à retirer de l'image des éléments d'information comme un détail dans un schéma, une virgule ou un accent. Ils doivent donc être testés préalablement à leur mise en œuvre.

Les résultats des tests doivent être également conservés dans le dossier de description technique. Les logiciels qui ont pour but d'éliminer le fond de page d'un document pour ne conserver que les informations variables de celui-ci peuvent être utilisés dans la mesure où les fonctionnalités mises en œuvre dans le système concerné sont parfaitement décrites dans le dossier de description technique.

De plus, lorsque la restitution fidèle du document exige la reconstitution de celui-ci par addition des éléments variables et du fond de page, la version du fond de page utilisée doit être identique à celle qui a été extraite lors de la numérisation et du traitement de l'image.

De même, la suppression automatique des pages blanches peut représenter un risque réel de perte d'information.

Dans le cas où ce processus doit être mis en œuvre, il est conseillé de vérifier que la technique utilisée est fiable et ne supprime pas des pages contenant des informations.

Il est en particulier conseillé de mettre en place un mécanisme permettant de rendre compte du nombre de pages supprimées par rapport au nombre de pages conservées.

3.5. Sécuriser l'application de production

À cet effet, il convient de mettre en place une gestion des profils d'habilitations qui, pour être pleinement opérante doit s'appuyer sur une authentification fiable des utilisateurs.

Les comptes doivent donc être nominatifs et il convient de protéger l'accès à l'application par des mots de passe individuels, comportant des caractères alphanumériques d'une longueur de 8 caractères au moins, changés régulièrement et d'interdire l'utilisation des trois précédents mots de passe.

3.6. Protéger les données à caractère personnel

Une attention particulière sera portée au respect des dispositions relatives au traitement des données personnelles (législation CNIL). Ainsi, dans l'hypothèse où le système mis en place ne dispose pas d'une structure informatique dédiée, l'absence de séparation physique des bases de données n'appelle pas d'observations de la part de la CNIL dès lors qu'une séparation logique effective des applications est mise en place.

3.7. Sécuriser le stockage

Il faut veiller à la redondance, la sécurité et la sauvegarde des supports et des index. Des copies de sécurité doivent être réalisées sur des supports amovibles de type CD, des bandes magnétiques ou un serveur miroir, qui seront placés dans un lieu distinct.

On se reportera également avec profit au Guide pour l'élaboration d'un cahier des charges de numérisation, publié par la direction des archives de France en février 2008, notamment pour la méthodologie à mettre en œuvre[25], ou encore pour les recommandations techniques (annexe 1) relatives aux formats aux taux de résolution, de compression, au cadrage et à l'orientation des documents numériques.

4. Contrôles

4.1. Contrôles dans le cadre de la numérisation du stock

L'outil de GED doit permettre de fournir, de gérer et de conserver des récapitulatifs journaliers comprenant un certain nombre d'informations, qui permettent de vérifier le bon déroulement des opérations de numérisation :

- l'identifiant du premier document ou du premier lot de documents numérisé et stocké de la journée ;
- l'identifiant du dernier document ou du dernier lot de documents numérisé et stocké de la journée ;
- le nombre total de pages traitées ;
- le nombre total de pages non traitées, en particulier lorsqu'il a été impossible de les numériser en raison de leur qualité insuffisante (faible contraste, déchirures, etc.) ;
- le nombre total de pages blanches, s'il y a lieu.

Chaque fichier numérisé doit également comporter dans ses métadonnées la date de sa création.

Ces métadonnées sont le plus souvent générées automatiquement par le scanner (cf. § 2.4) ; si tel est le cas, l'outil de GED doit pouvoir récupérer, gérer et conserver ces informations (cela est en particulier à exiger dans le cadre d'un cahier des charges pour l'externalisation de l'opération de la numérisation).

La numérisation une fois effectuée doit faire l'objet de contrôles. Ils doivent au moins porter sur :

- la quantité de pages numérisées par rapport aux attestations de numérisation,
- la qualité et la fidélité des images fixes par rapport aux originaux,
- la justesse des informations destinées à l'indexation des documents numérisés.

Il est recommandé que le statut « vérifié » soit activé pour chaque document ou lot de documents entrés dans la GED, une fois le contrôle achevé.

Par ailleurs, il semble prudent que le contrôle final de la qualité soit réalisé par des personnes autres que les opérateurs.

À noter : Il est souhaitable que le responsable de l'opération de numérisation lorsque que celle-ci s'achève, produise un certificat de conformité entre ce qui a été numérisé et les originaux papier. Le document doit également préciser le nombre de documents (lots) numérisés ainsi réalisés.

4.2. Contrôles dans le cadre de la numérisation du flux

Les processus de numérisation de stock et de numérisation de flux diffèrent

25 Depuis la prise en charge par le prestataire en cas d'externalisation, à la manipulation des documents, à l'étalonnage de la chaîne de numérisation, au fichier de récolement, aux contrôles, à la livraison des images…

principalement sur l'organisation mise en place pour les gérer.

Dans le cadre de la numérisation du stock, des opérateurs sont généralement affectés à temps plein aux travaux de numérisation, alors que pour la numérisation du flux, les tâches de numérisation dépendent de l'organisation mise en œuvre.

Si l'organisme affecte des agents à la numérisation du flux, ce qui représente, compte tenu du volume des dossiers, une part importante de leur travail, les contrôles à effectuer sont identiques à ceux décrits pour la numérisation du stock.

Si l'organisme confie la numérisation aux agents habituellement chargés de l'instruction des dossiers, le processus de numérisation du flux s'intègre dans celui plus général de traitement des demandes.

Les notions de lots de document numérisés, de journées de travail de numérisation perdent alors leur sens.

Toutefois, les agents instructeurs sont amenés à apporter un soin particulier au bon déroulement de l'opération de numérisation, en vérifiant après chaque numérisation de document :

- la quantité de pages numérisées par rapport aux documents originaux,
- la qualité et la fidélité des images fixes par rapport aux originaux
- la justesse des informations destinées à l'indexation des documents numérisés.

5. Intégrer les fichiers nativement numériques ?

L'organisme produit également et/ou reçoit par mail de nombreux documents nativement numériques qui ne sont pas générés automatiquement par le progiciel mais par d'autres logiciels (logiciels bureautiques, logiciels de messagerie…).

Il est alors pertinent de rattacher aux dossiers individuels gérés par le système les documents nativement numériques, qui offrent notamment des possibilités de recherches plus intéressantes que des documents numérisés (recherche plein texte par exemple).

La liste de ces documents devra être dressée et tenue à jour régulièrement, en relation avec la liste des documents à numériser ; des procédures spécifiques (envoi systématique par les partenaires, droits d'accès…) devront être mises en place pour alimenter la GED.

Assurer la conservation des dossiers numériques, leur versement et leur élimination, l'ensemble de ces opérations se faisant sous le contrôle scientifique et technique des missions et services d'archives avec visa réglementaire pour les éliminations.

Certains des documents constitutifs des dossiers devront, dans certains cas, être conservés sous forme numérique plus de 5 ans ; de ce fait, un risque d'obsolescence existe, tant pour les supports que pour les formats des fichiers générés au cours de la numérisation.

Il convient par conséquent de prendre en compte les possibilités d'évolutions techniques et leurs impacts sur la lecture du document numérisé.

Ce point est très important, car c'est à l'organisme de mettre en œuvre tous les moyens humains, matériels et organisationnels permettant la conservation des documents et leur consultation pendant la période nécessaire.

Il convient de réfléchir par conséquent à une stratégie d'archivage électronique, permettant d'assurer la pérennité des informations et des documents numériques (évolution des supports et des formats, sans perte d'information notamment).

Dans cette perspective, une réflexion sur au cycle de vie des données est indispensable (durée de conservation de la donnée et des images dans la base de production, bascule éventuelle dans une base de pré-archivage, modalités de gestion et d'utilisation de la cette base de pré-archivage, exports vers un service d'archivage intermédiaire…).

Présentation théorique d'un système d'archivage électronique

Présentation de la norme OAIS

Il s'agit d'une norme fonctionnelle généraliste sur laquelle s'appuient de nombreuses autres normes et standards. Sa première version est parue en 2002 et est enregistrée comme norme ISO sous le numéro 2003:14721. Une nouvelle version est en phase d'acceptation par l'ISO pour laquelle la traduction est déjà en cours.

La terminologie et les fonctions définies par la norme OAIS servent de cadre conceptuel et rédactionnel aux contenus de ce site. En effet, le modèle général de cette norme répond à un besoin primordial pour tout SAE, avant même toute considération technique que sont :

- la prise en charge des archives et leur contrôle, y compris le contrôle des formats des fichiers en entrée, leur validation, leur conversion si nécessaire (formats texte, formats images, formats graphiques, formats audiovisuels…) ;
- la conservation pérenne de ces archives (réplication de l'information sur des sites distants, surveillance des supports, traçabilité extrême), qui correspond aux magasins de conservation des archives papier ;
- la gestion des données, leur recherche et accès (comme les archives papier, via des portails de recherche et de consultation) ;
- la planification de la préservation à long terme : veille technologique sur les formats et les supports, plans de migration d'ampleur de formats, de supports.

Plan de la fiche :

- L'environnement OAIS
- Le modèle d'information
- Les entités fonctionnelles
- Perspectives de collaboration entre Archives
- Evolutions en cours

1. L'environnement OAIS

Bien qu'étant conçu dans un cadre spécifique, le modèle OAIS a un champ d'application bien plus large. En effet, l'OAIS est un modèle abstrait, qui ne donne aucune spécification technique, mais offre plutôt un vocabulaire et un cadre théorique pour penser différents cas de figures de l'archivage. Sa démarche peut donc être partagée par des institutions de nature et de préoccupations très diverses : grandes bibliothèques nationales, grandes institutions scientifiques et archivistiques ou encore industries ayant à conserver sur le long terme des volumes importants d'informations numériques.

La norme OAIS contribue ainsi à exprimer et à formaliser des problématiques transverses et, par conséquent, à favoriser l'échange d'expériences, la collaboration et la mutualisation des compétences pour y répondre. C'est cette démarche qui est mise en œuvre

dans le groupe « Pérennisation de l'information numérique » (PIN)[26], qui réunit le CNES, la Bibliothèque nationale de France, les Archives de France, des ministères, des organismes scientifiques et des consultants.

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

Après une introduction largement dédiée à l'établissement de la terminologie utilisée dans la norme suit une explication générale des principaux concepts de l'OAIS ainsi que la définition des responsabilités dévolue à l'Archive. Vient ensuite la description détaillée des différentes fonctions présentes dans cette dernière et du modèle d'Information utilisé en son sein. La norme termine en formalisant les divers concepts et outils utilisables pour assurer la pérennisation, comme les opérations de migrations par exemple, et enfin en développant les possibilités et les modalités de collaboration entre différentes Archives.

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

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

- négocier avec les Producteurs d'information pour s'assurer que les **Contenus d'information et Informations de pérennisation** (PDI) associés qu'elle va recevoir correspondent bien à sa mission et aux besoins de la Communauté d'utilisateurs cible ;
- acquérir une maîtrise suffisante de l'information fournie, au niveau requis pour pouvoir en garantir la **Pérennisation** ;
- extraire - ou obtenir par d'autres moyens – une **Information de description** suffisante pour que la Communauté d'utilisateurs cible puisse trouver le Contenu d'information qui l'intéresse ;
- déterminer quelles communautés doivent constituer la Communauté d'utilisateurs cible en mesure de comprendre l'information fournie et assurer que l'information à conserver est compréhensible pour cette communauté (c'est-à-dire, sans l'assistance des experts ayant produit ces informations) ;
- appliquer une stratégie et des procédures documentées garantissant la conservation de l'information contre tout imprévu dans les limites du raisonnable, et permettant la diffusion d'une information, copie authentifiée de l'original ou permettant de remonter à l'original ;
- rendre l'information conservée disponible pour la Communauté d'utilisateurs cible.

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

En amont de l'Archive se situe le **Producteur** (Producer) qui fournit les informations à conserver et en aval l'**Utilisateur** (Consumer) qui peut en demander l'accès. L'Archive est

26 http://pin.association-aristote.fr/doku.php/accueil

enfin placée sous la tutelle du **Management (Management)**, qui définit son champ d'action et oriente son travail en fonction de ses attentes (*voir figure 1*).

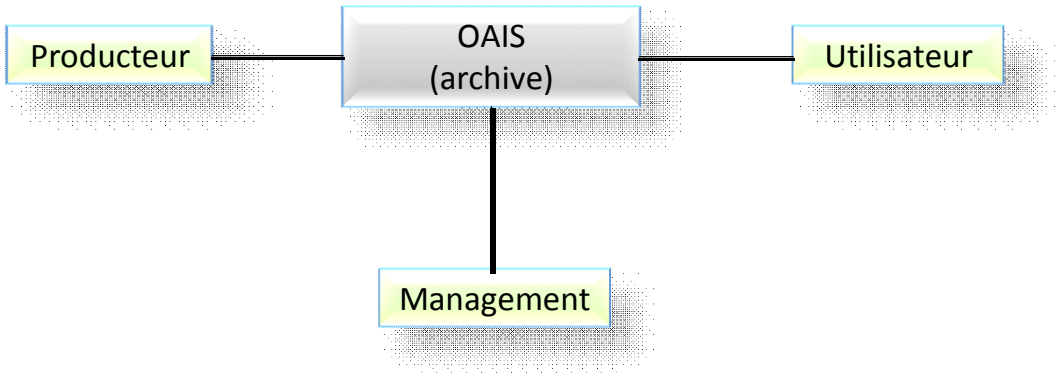
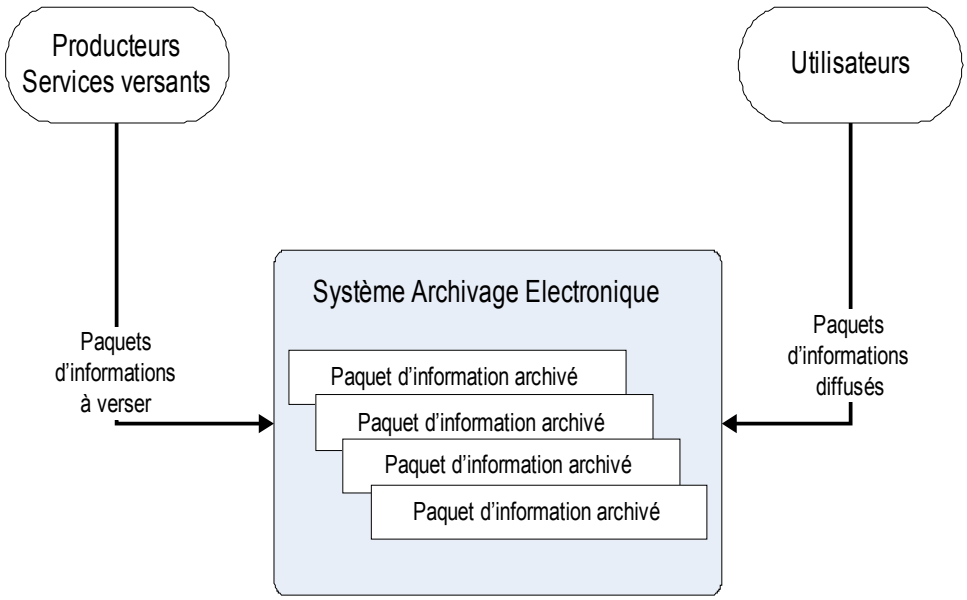


Figure 1. - Environnement général OAIS

2. Le modèle d'information

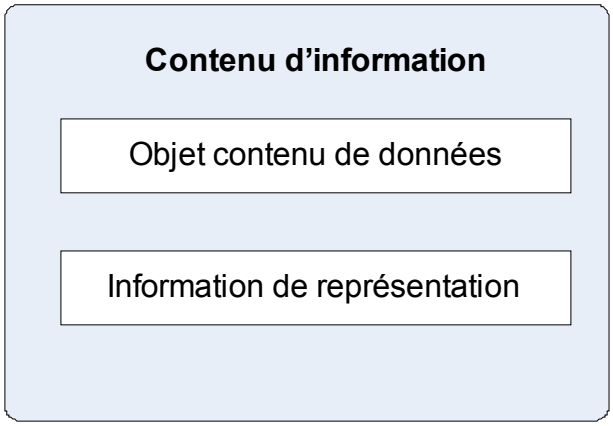
Le modèle OAIS repose sur l'idée que l'information constitue des paquets et que ces paquets ne sont pas les mêmes suivant qu'on est en train de produire l'information, d'essayer de la conserver ou de la communiquer à un utilisateur. On a donc trois sortes de paquets (voir également ci-après la section relative aux entités fonctionnelles, p. 8-9) :

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



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

Contenu d'information (Content information) : l'**Objet-données** (Data object), qui peut être physique (un livre ou un document peuvent être considérés comme tels) ou numérique (il s'agit dans ce dernier cas d'une suite de bits, 1 ou 0, écrite sur un support numérique quel qu'il soit) et son **Information de représentation** (Representation information), c'est-à-dire son format, sa structure, sa signification... nécessaire à la compréhension de cet objet par la Communauté d'utilisateurs cible. Par exemple, les spécifications du format PDF sont une information de représentation nécessaire pour pouvoir lire correctement un document de ce format. Cette structure associant Objet-données et Information de représentation n'est pas propre au Contenu d'information : il s'agit du modèle générique de l'Objet-information (*voir figure 2*), utilisé pour penser tous les types d'informations présents dans l'Archive qui sont évoqués ci-dessous.



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

Les Informations de représentation ont également besoin de leurs propres Informations de représentation, par un mécanisme récursif. Le nombre d'Informations de représentation versées, conservées ou ajoutées est déterminé par l'Archive en fonction des besoins de sa **Communauté d'utilisateurs cible** (Designated community), c'est-à-dire de personnes susceptibles de recourir aux Informations conservées par l'Archive et qui partagent une **Base de connaissance** (Knowledge base) commune. La Communauté cible comme sa Base de connaissances sont susceptibles d'évoluer dans le temps, et nécessitent dans ce cas que l'Archive anticipe la multiplicité et l'hétérogénéité des Utilisateurs dans le temps. C'est par exemple ce travail qui doit être effectué si une base de données autrefois utilisée au quotidien par une administration devient une source historique pour des généalogistes. La structure de

la base, son but, certaines abréviations ou termes techniques utilisés dans les différents champs doivent être explicités pour s'adapter à des Utilisateurs peu familiers du fonctionnement de l'administration productrice.

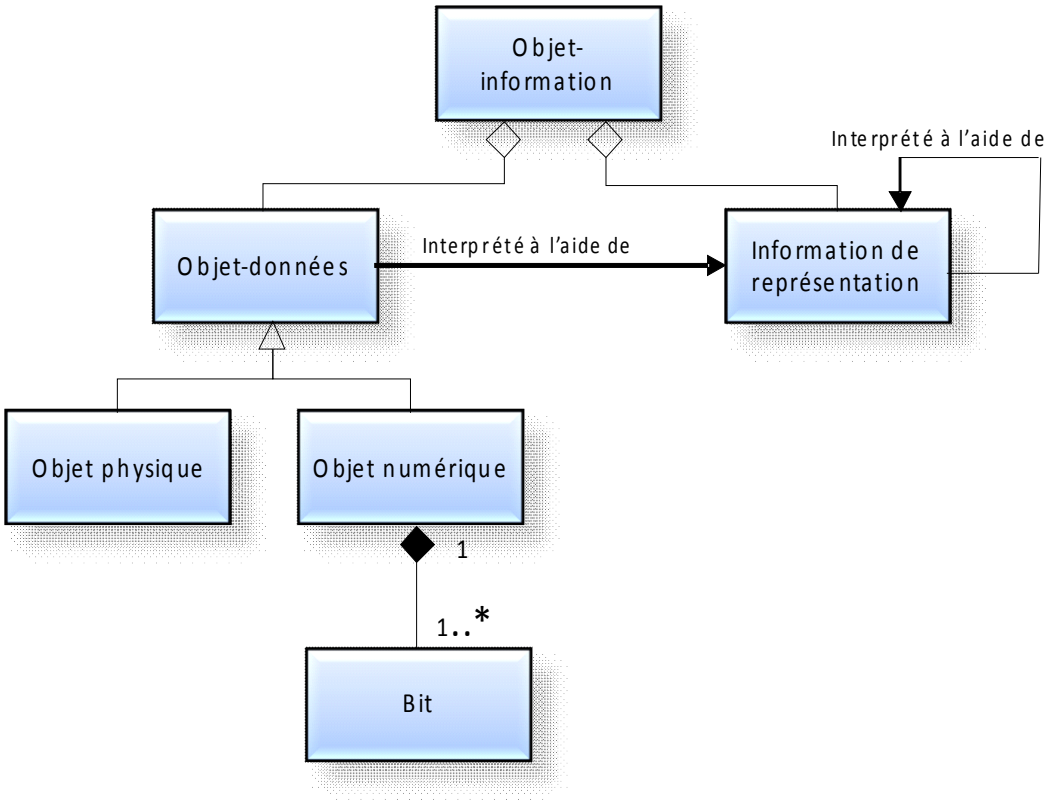
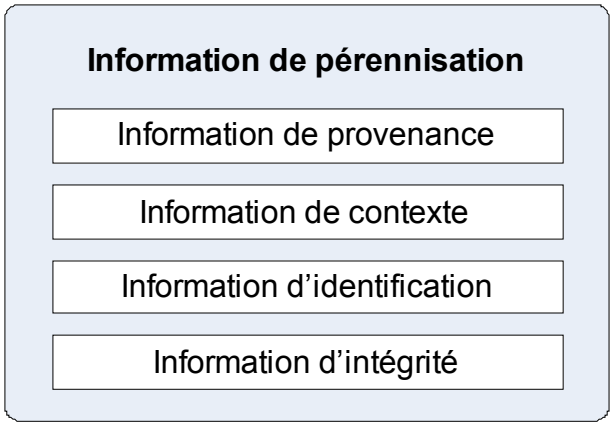


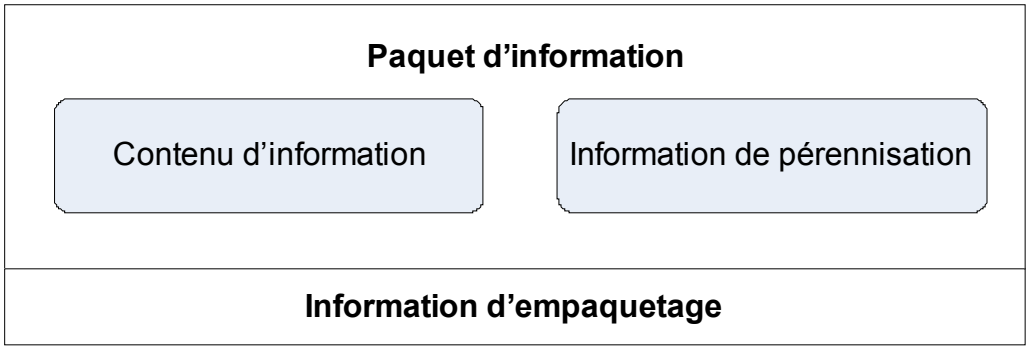
Figure 2. - Schéma représentatif du modèle d'Information

Information de pérennisation (Preservation Description Information ou PDI) : métadonnées requises pour conserver le Contenu d'information. L'Information de pérennisation comprend quatre types d'informations :

- **Information de provenance** (Provenance information) : décrit l'origine du Contenu d'information.
- **Information de contexte** (Context Information) : décrit les relations existant entre le Contenu d'information et d'autres informations situées hors du Paquet d'informations.
- **Information d'identification** (Identification information) : fournit un ou plusieurs identificateurs, ou systèmes d'identificateurs, grâce auxquels le Contenu d'information peut être identifié de façon unique.
- **Information d'intégrité** (Fixity information) fournit un mécanisme ou un dispositif protecteur pour prémunir le Contenu d'information contre toute altération non documentée.



Le Contenu d'information et le PDI sont identifiés et encapsulés par une **Information d'empaquetage** (Packaging Information). Le paquet qui en résulte peut être retrouvé grâce à une **Description de paquet** (Package description) (*voir figure 3*). Cette Description synthétise à la fois la teneur du Contenu d'information et celle des Informations de pérennisation sous une forme qui permet aux Utilisateurs la recherche des ressources qui les intéressent d'après un certain nombre de critères.



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

3. Les entités fonctionnelles

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

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

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

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

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

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

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

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

Le modèle OAIS distingue plusieurs types de migrations, selon que les séquences de bits correspondant aux Paquets d'information sont modifiées ou non, et selon le but de l'opération.

Le **rafraîchissement de support** consiste en une copie exacte du Paquet d'information sur un support équivalent mais neuf afin d'éviter des pertes de contenus dues à l'obsolescence ou au vieillissement du support initial, et ce sans que le repérage du paquet par l'entité « Stockage » ne change.

La **duplication** relève du même principe, si ce n'est que, dans ce cas, il peut être nécessaire de mettre à jour les infrastructures de recherche de l'entité « Stockage ». C'est le cas lorsqu'on duplique les informations depuis une baie de stockage sur des bandes magnétiques : on ne peut accéder au contenu de la sauvegarde depuis une infrastructure comparable à celle des disques durs de la baie.

Le **ré-empaquetage**, quant à lui, n'apporte que des changements mineurs aux séquences de bits de la seule Information d'empaquetage.

Enfin, la **transformation** est la migration la plus lourde à mettre en œuvre, puisqu'elle porte des modifications aux séquences de bits. Certaines transformations sont dites **réversibles** quand elles permettent de revenir à la séquence de bits originelle, comme le passage d'un codage de caractère ASCII à un codage Unicode, tandis que d'autres sont **irréversibles** quand la séquence originelle est définitivement modifiée, même si le Contenu d'information reste compréhensible.

4. Perspectives de collaboration entre Archives

La norme OAIS définit différentes modalités de collaboration possibles entre Archives. Elle prévoit la mutualisation de différentes fonctions selon que les Archives ont des Producteurs ou des Communautés d'utilisateurs cibles similaires. On peut alors aboutir à la mise en place de logiciels d'accès aux ressources mutualisés, ou encore d'espaces de stockage communs ou encore de mutualisation des fonctions de veille technologique. La norme OAIS constitue dès lors un outil pratique efficace pour mettre en place ces ressources partagées. Plusieurs types d'associations sont envisageables qui vont de la collaboration ponctuelle à la contractualisation.

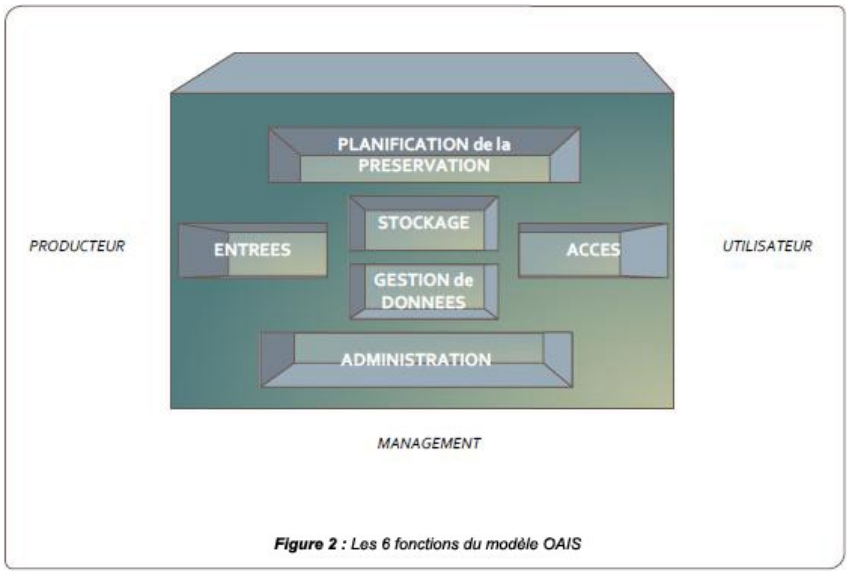


Figure 2 : *Les 6 fonctions du modèle OAIS*

Schéma fonctionnel d'un système d'archivage électronique inspiré de la norme OAIS[27]

5. Évolutions en cours

Le CCSDS mène en ce moment une révision de la norme OAIS. Le processus en cours à l'ISO en est à sa phase d'approbation. Pour la version française, le CNES est chargé du pilotage et de la centralisation des différentes remarques, tant sur le fond de la norme que sur la traduction.

Tous les concepts fondamentaux, à la base du succès international du Modèle OAIS, sont maintenus.

Les changements essentiels sont les suivants :

- les responsabilités en matière de gestion des risques sont explicitées à différents niveaux hiérarchiques – le Management est notamment chargé d'évaluer les risques pouvant entraîner une perte d'information – et dans différentes entités fonctionnelles de l'Archive, puisque l'Administration doit prendre des décisions en fonction des analyses et des plan de gestion des risques proposés par la Planification de la pérennisation ;
- l'ajout de la notion d'Information de droit d'accès, parmi les Informations de pérennisation, est une des modifications les plus importantes. Il permet d'adjoindre à chaque Paquet d'information des restrictions de consultation, de modification ou encore de suppression du Contenu d'information ou de ses métadonnées, ce qui peut s'avérer utile pour garantir les délais d'incommunicabilité prescrits par la loi sur les archives ;
- la Propriété d'information est également un nouveau concept, qui désigne une partie des renseignements du Contenu d'information. La Propriété d'information est repérée et matérialisée par la Description de la Propriété d'information. Cette dernière doit mettre en évidence un aspect spécifique du Contenu d'information dans un but précis, ce qui peut par exemple, en indexant tous les Contenus d'information recourant aux mêmes formats, s'avérer utile pour prévoir des migrations à grande échelle.

À la suite de la norme OAIS, est sortie la norme **PAIMAS**[28] qui spécifie le processus et les étapes à franchir dès lors qu'un service producteur et un service d'archive souhaitent se

27 http://www.cines.fr/spip.php?rubrique230

28 Producer Archive Interface Methodology, toujours sous l'égide du CCSDS. PAIMAS ou CCSDS, 651.0-B-1, Producer archive interface methodology abstract standard, ISO 20652, mai 2004

mettre d'accord pour le transfert et la prise en charge d'une nouvelle catégorie d'archives, soit une méthodologie de travail commune et structurée en quatre phases[29].

Enfin, le projet de norme **PAIS**[30] aborde à la fois la formalisation du modèle des objets à transférer, l'empaquetage de ces objets sous forme de SIP et la conformité des objets transférés par rapport au modèle.

29 Phase préliminaire, phase de définition formelle, phase de transfert et phase de validation.

30 *Producer Archive Interface*.

Fonctionnalités d'un système d'archivage électronique

Plan de la fiche :

- Fonctionnalités du SAE
- Architecture du SAE

1. Fonctionnalités du SAE

Extraits du « cahier des charges pour un système d'archivage électronique » élaboré en 2006 dans le cadre de l'étude sur l'archivage sécurisé dans le secteur public (DCSSI)[31]

Pour les processus métier d'un SAE, voir sur le site internet du CINES les différentes modélisations proposées processus par processus : https://alfresco.cines.fr/alfresco/faces/jsp/browse/browse.jsp

Est brièvement rappelé ci-dessous l'ensemble des fonctions attendues par le SAE avant de le décrire plus en détail.

F1. Versement : permet le traitement des paquets d'informations en provenance des Services versants dans son ensemble. Cette fonction inclut tous les mécanismes de préparation, transmission, contrôle, rejet, complément d'information ainsi que tous les traitements de ces informations pour une intégration dans le dispositif de Stockage des contenus et celui de gestion des données descriptives ;

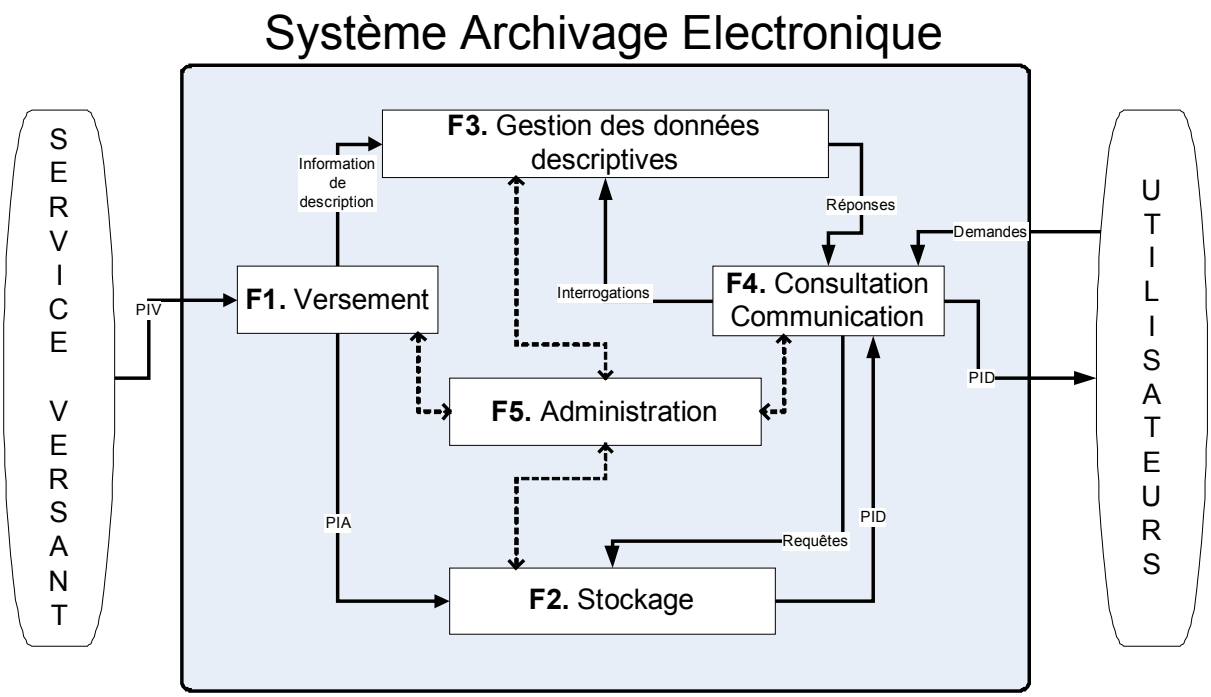
F2. Stockage : gère l'ensemble des services liés à la conservation des paquets d'informations archivés à partir du moment où ils sont mis à sa disposition par la fonction de Versement jusqu'à leur destruction/élimination s'il y a lieu tout en garantissant leur intégrité. Cette fonction prend entre autres en compte les aspects de choix de supports et de gestion de l'ensemble des migrations ;

F3. Gestion des données descriptives : assure la conservation, la mise à disposition et la mise à jour des informations descriptives associées aux contenus d'informations, conservés par la fonction Stockage. Ces informations doivent servir aux utilisateurs comme point d'entrée au SAE et permettre de retrouver les données qu'ils recherchent en assurant le lien avec leur identification de localisation dans le système de stockage ;

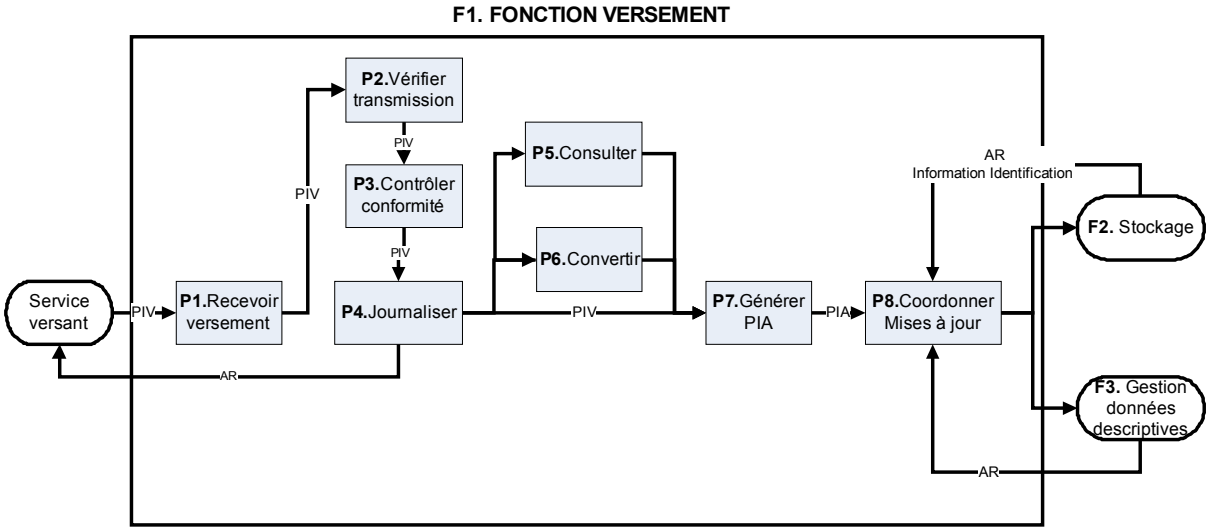
F4. Consultation et communication : prévoit l'ensemble des mécanismes permettant d'accéder, de consulter et de livrer les informations disponibles dans le SAE, qu'il s'agisse des données descriptives ou du contenu lui-même. Elle comprend la mise à disposition d'une interface de consultation, un système de recherche effectuée à partir des données descriptives, un principe de visualisation du résultat, la sélection de contenus à communiquer et la livraison effective de ces contenus sous forme de paquets d'informations diffusés. Dans la mesure où la communication du contenu peut être différée par rapport au moment de l'interrogation, cette fonction doit également prévoir un mécanisme de commandes à destination des utilisateurs, le suivi étant assuré par la fonction Administration

F5. Administration : permet d'assurer l'exploitation d'ensemble du Système d'archivage électronique et sa pérennisation (veille technologique, surveillance des supports, migrations des supports et de formats) ainsi que la gestion des utilisateurs du SAE au sens de leurs droits d'accès

31 http://www.ssi.gouv.fr/IMG/pdf/ArchivageSecurise-CahierDesCharges-2006-05-16.pdf



F1. Fonction versement



PIV : Paquet d'information versé
PIA : Paquet d'information archivé

P1. Recevoir versement : Ce processus consiste à effectivement réceptionner dans un espace de stockage tampon, les Paquets d'informations versés (PIV) en provenance du Service versant. La transmission entre les deux services peut être effectuée en ligne ou via un support amovible dans le cas par exemple de fichiers volumineux envoyés à faible fréquence ;

P2.Vérifier transmission: Ce processus vérifie que le Paquet d'informations transmis par le Service versant a bien été réceptionné dans son intégralité et sans altération. L'intégrité globale de l'envoi ainsi que l'intégrité des différents Paquets d'informations transmis et reçus devront être contrôlées.

P3. Contrôler conformité : Ce processus contrôle que le paquet d'informations versé est

conforme et respecte bien les conditions définies entre le service versant et le service d'archives, entre autres en matière de structuration de l'ensemble des données et de leur complétude, en matière de format de description, en matière de respect des formats d'encodage des objets versés et de leurs composants (conformité au standard d'échange des données pour l'archivage : SEDA) ;

P4. Journaliser : Ce processus répond à un impératif que l'on retrouve dans l'ensemble du SAE consistant à enregistrer dans un journal l'intégralité des opérations effectuées et des événements. En parallèle, ce processus envoie un accusé de réception (ou un rapport d'anomalie en cas de contrôles négatifs) au Service versant précisant le résultat de l'opération, suite aux différents contrôles effectués ;

P5. Consulter : Ce processus doit permettre si nécessaire, aux personnes habilitées du service d'archive, de consulter le contenu du paquet d'information versé ;

P6. Convertir : on peut ainsi envisager que, dans certains cas, le SAE opère une migration de formats soit à l'arrivée soit au terme d'un délai dépendant de l'obsolescence du format d'origine ;

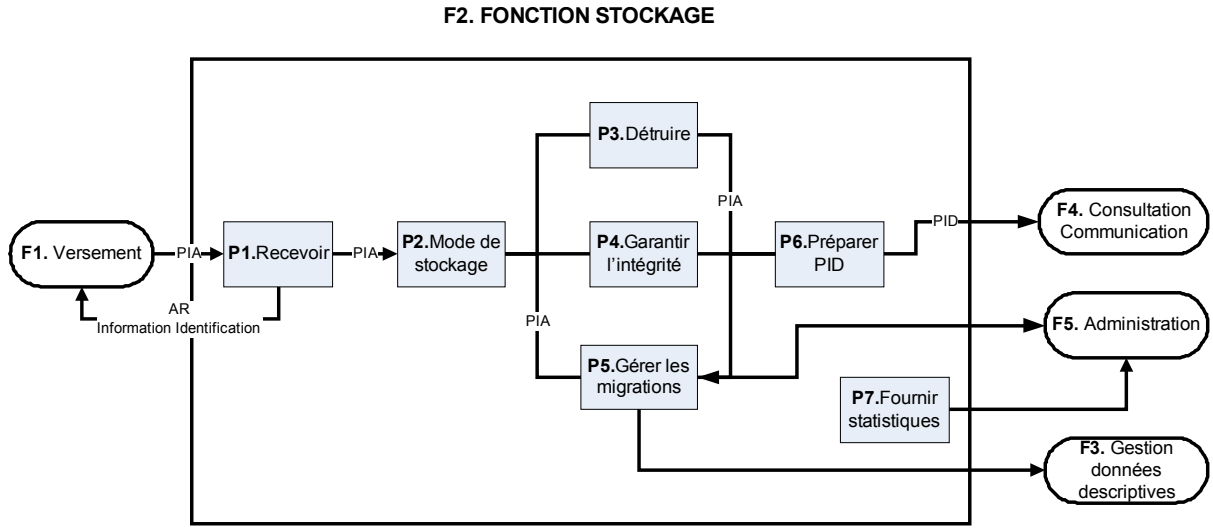
P7. Générer PIA : Ce processus revient à constituer un Paquet d'informations archivé conforme aux normes de documentation et de formatage des données telles que définies pour le SAE ;

P8. Coordonner les mises à jour : Ce processus consiste d'une part à transmettre à la fonction stockage le contenu d'information à conserver et d'autre part à transmettre à la fonction gestion des données descriptives les données correspondantes. L'ensemble de ces informations se retrouve dans le Paquet d'information archivé. Le processus attend ensuite en retour l'accusé de réception du résultat de l'opération. Dans le cas du stockage, l'accusé de réception doit contenir l'information d'identification de l'espace de stockage. Cette dernière donnée est ensuite envoyée en complément des informations précédentes à la fonction gestion des données descriptives.

En complément à ces éléments purement fonctionnels, l'Administration devra notamment spécifier :

- les éléments de volumétrie afférents suivant la nature des archives (intermédiaire ou définitive), leurs formats et le type des archives (dossiers comptables, individuels, contentieux, judiciaires,…) ;
- le type (en ligne ou sur support amovible) et la fréquence des transmissions auquel il faudra avoir recours pour chaque type d'archives ;
- dans le cas d'un versement en ligne, il conviendra de préciser quels types de protocoles pourront être utilisés.

F2. Fonction stockage



PIA : Paquet d'information archivé
PID : Paquet d'information diffusé

Un système de stockage doit avoir les caractéristiques suivantes : fiabilité, disponibilité, maintenabilité, sécurité mais doit également permettre l'abstraction de la plate-forme matérielle, être extensible, interopérable et évolutif.

D'une manière générale, les évolutions de la plate-forme de stockage doivent être sans conséquence sur l'organisation logique de l'archivage.

Le stockage comprend les sept processus suivants :

P1. Recevoir : Ce processus revient à réceptionner les paquets d'informations archivés en provenance du processus de Versement et à les transférer physiquement vers le volume de stockage le mieux approprié et correspondant aux conditions d'archivage (durée, fréquence de consultation, communication en ligne ou différée, destruction in fine, …) indiquées au moment du versement. Lorsque les paquets d'informations archivés sont effectivement écrits sur le support de stockage adapté, il y a transmission au processus de Versement du résultat de l'opération comprenant l'information d'identification correspondant à l'espace de stockage où se trouve physiquement les paquets d'informations archivés qui viennent d'être traités ;

Remarque par rapport à l'écriture effective :

Au sujet de l'écriture sur un support de stockage quel qu'il soit, il est important de vérifier que l'accusé de réception du système est bien envoyé lorsque l'écriture est véritablement effective sur le support en question et non pas en attente de traitement dans un espace mémoire tampon. En effet, dans la majorité des cas et suite à un ordre d'écriture par exemple sur disque, l'information concernée se trouve en mémoire vive de l'ordinateur et est donc sujette à disparition en cas de coupure de courant. Il est vrai que la majorité des systèmes possèdent des sécurités en matière d'alimentation électrique mais il est néanmoins prudent de demander et de vérifier à quel moment précis l'accusé de réception d'écriture est effectivement généré.

Ce premier point se complique également du fait que l'infrastructure d'un SAE est généralement composée d'au moins deux sites. Dans le cas d'une réplication, détaillée par la suite, il s'agit donc également de vérifier que l'accusé de réception parvient après l'écriture effective sur les deux sites. Si tel n'est pas le cas, il faudra alors analyser quel risque est encouru de pouvoir se trouver dans une situation, certes extrême, où l'information pourrait par exemple avoir été écrite sur le premier site et non sur le second et prévoir les procédures associées afin d'y remédier.

P2. Mode de stockage : Ce processus consiste à conserver effectivement les paquets d'informations archivés et à choisir le support adéquat en fonction d'un certain nombre de critères dont les principaux sont l'accessibilité et la durée. Pour ce faire il pourra être envisagé de mettre en place un système de HSM (hierarchical storage management) afin d'aider à cette gestion des différents supports ;

P3. Détruire : Ce processus est destiné à traiter le cas échéant la destruction des paquets d'informations archivés de façon manuelle ou automatique ;

Remarque sur la destruction :

Plusieurs solutions existent permettant de ne plus avoir accès à une information. Ainsi lorsqu'il s'agira de destruction, le SAE devra posséder une telle fonction comportant au minimum un dispositif de suppression des accès aux contenus d'informations par suppression des index et mieux un véritable dispositif d'effacement des contenus d'information. Ce dispositif devra par ailleurs être conçu de telle sorte à ne laisser aucune trace sur le support d'origine, due entre autre au phénomène physique de rémanence des supports magnétiques. En ce qui concerne les supports amovibles type bande ou CD, la destruction sera opérée sur l'ensemble du contenu et du contenant.

P4. Garantir l'intégrité : Ce processus est extrêmement important dans la mesure où il doit garantir l'intégrité de l'ensemble des paquets d'informations archivés et en conséquence, la vérifier systématiquement. Il est en effet nécessaire de contrôler régulièrement les paquets d'informations archivés sur les différents supports afin d'anticiper d'éventuelles erreurs et surtout de prévoir des dispositifs d'avertissement d'une part et de correction d'autre part. En cas de détection d'une erreur d'intégrité, la seule façon de la corriger est de remplacer les données concernées par un jeu de données identiques non corrompues dont on disposera grâce à un système de duplication adapté de l'ensemble des données ;

Remarque par rapport au contrôle d'intégrité :

Il est important de noter qu'il existe en réalité plusieurs façons de contrôler l'intégrité.

Contrôle ponctuel : Le contrôle n'a lieu qu'au moment de l'accès à l'objet concerné c'est-à-dire au moment de sa communication. Le principal inconvénient réside dans le fait qu'il peut être trop tard dans le sens où l'on va effectivement détecter une erreur d'intégrité mais sans pouvoir y remédier du fait que l'on ne possède pas ou plus de jeu sain de ces données.

Contrôle régulier par sondage : Ce type de contrôle est opéré de façon totalement automatique sur les contenus d'informations choisis de façon aléatoire (sauf cas particulier de contenus d'informations particulièrement sensibles à traiter en globalité). Ces contrôles doivent également pouvoir être paramétrés en fonction du type de supports et de leurs âges respectifs.

Contrôle continu : Comme indiqué, ces contrôles sont opérés de façon continue sur un ensemble défini de paquets d'informations archivés. Signalons à ce niveau qu'un tel contrôle existe de façon native sur certains systèmes de stockage.

P5. Gérer les migrations : Il s'agit de maîtriser l'ensemble des migrations (voir ci après) requises par le système tant des supports que des formats. Ces migrations interviennent soit de façon planifiée (voir fonction Administration) soit par exemple pour corriger des erreurs détectées sur tel ou tel support ;

Remarque par rapport aux différents types de migrations :

Sans vouloir trop entrer dans les détails il est cependant important de préciser qu'il existe plusieurs types de migrations abordés ci-dessous.

Changement de supports : Ce premier type de migration consiste à permettre de remplacer, renouveler des supports sur lesquels les données sont conservées. Ces changements pourront faire suite à des erreurs répétitives sur un support ou tout simplement être programmés au préalable en fonction du type de support et de leur age. Les erreurs dont

il est ici fait mention sont essentiellement de deux types : erreurs de lecture du support ou erreur d'intégrité. Dans les deux cas il est nécessaire et indispensable de disposer d'un dispositif de correction automatique de ces erreurs dont la conséquence principale revient justement à changer de support.

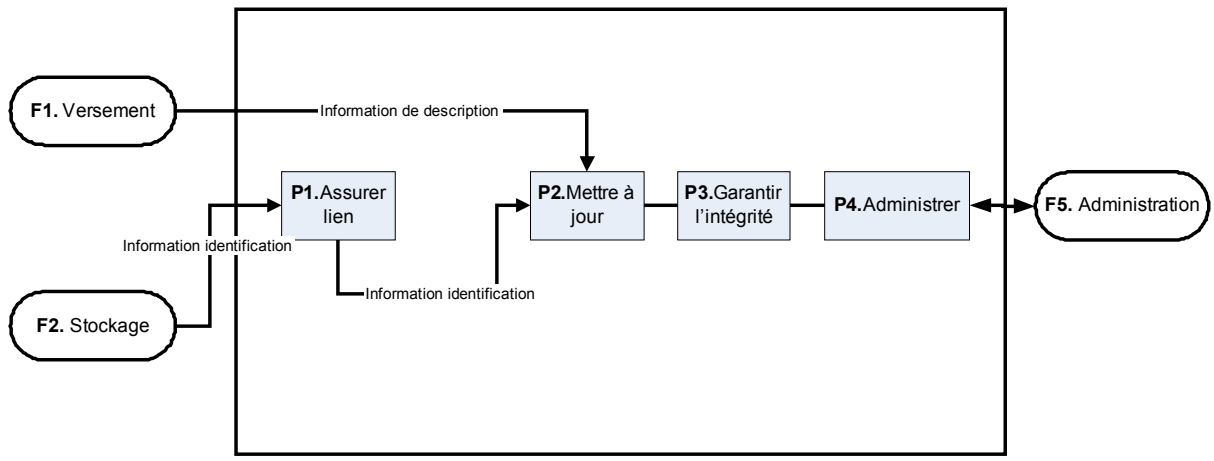
Changement de format : Au-delà du simple changement de support, il existe un autre type de migration destiné à permettre d'assurer le changement des formats (au sens logique du terme). La migration de formats pourra être rendue nécessaire en raison d'une obsolescence technologique des formats de données archivés, en raison d'une veille technologique anticipant la disparition de tels formats ou au contraire de l'apparition sur le marché, d'un nouveau format plus approprié à la pérennisation (par exemple, des fichiers au format PDF vers le format PDF/A normalisé ISO 19005).

P6. Préparer : Ce processus est destiné à transmettre les paquets d'informations diffusés suite à une sollicitation du processus de communication .

P7. Fournir les statistiques : Il s'agit de bâtir des statistiques d'exploitation relatives d'une part aux capacités utilisées par rapport aux différents supports et espaces de stockage, ainsi que sur l'état des supports et d'autre part en matière de communication des paquets d'informations archivés, en compléments aux statistiques de consultation, sans oublier l'évolution des paquets d'informations versés.

F3. Fonction gestion des données descriptives

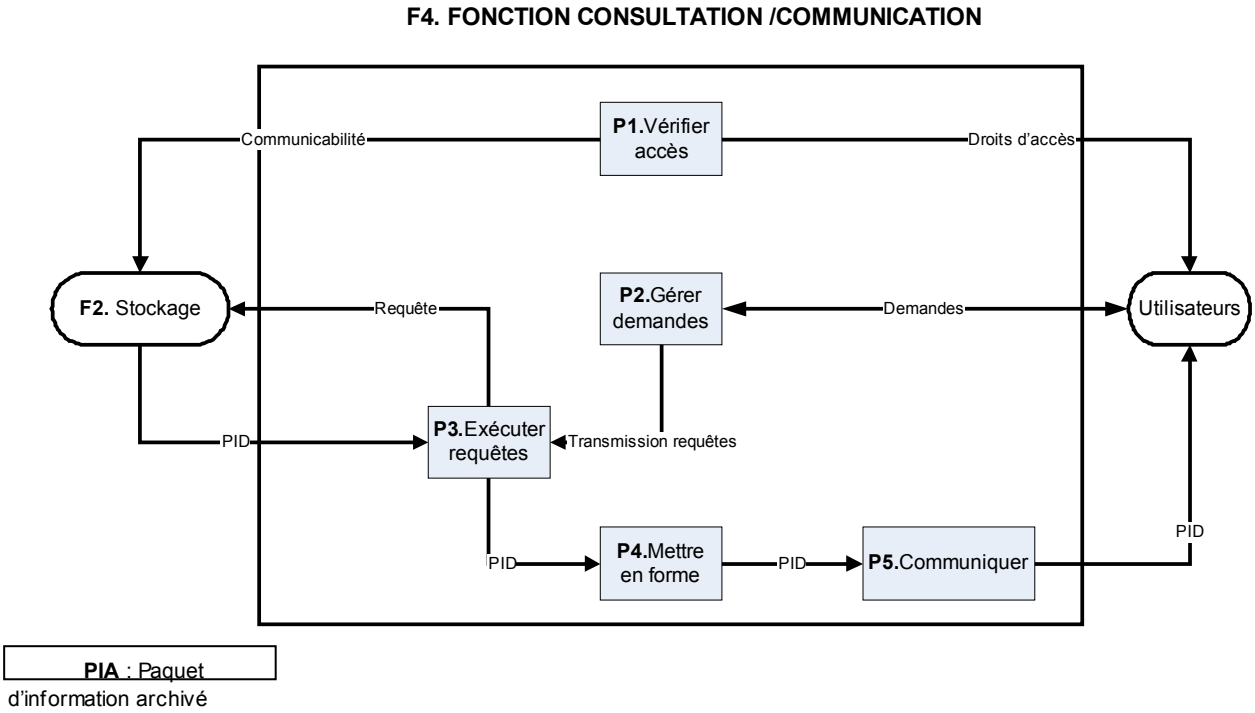
F3. FONCTION GESTION DES DONNEES DESCRIPTIVES



La fonction gestion des données descriptives est composée des quatre processus suivants :

- **P1.**Assurer lien : Ce processus consiste à maintenir le lien entre les informations descriptives et la localisation physique ou électronique des contenus d'informations ;
- **P2.**Mettre à jour : Le processus doit permettre la mise à jour des données correspondantes et au besoin en enregistrer de nouvelles suite à un nouveau versement ou suite à une opération de migration ;
- **P3.**Garantir l'intégrité : Ce processus revient à garantir l'intégrité de l'ensemble des données gérées et à la vérifier régulièrement. Il est ainsi nécessaire de contrôler d'éventuelles erreurs à l'aide de fonctionnalités appropriées complétées par des systèmes d'avertissement et si possible de correction ;
- **P4.**Administrer : De façon spécifique par rapport à la fonction d'administration d'ensemble du SAE, ce processus doit administrer les fonctions de la base de données le cas échéant, à savoir conserver et tenir à jour les schémas des tables utilisées, les définitions des vues et autres états ainsi que garantir son intégrité référentielle.

F4. Fonction consultation / communication



Les cinq processus composant cette fonction sont décrits ci-après :

- **P1.**Vérifier les accès : Ce processus revêt un double objectif, tout d'abord vérifier les autorisations d'accès des utilisateurs et d'autre part vérifier la communicabilité des paquets d'informations archivés ;
- **P2.**Gérer demandes : Ce processus permet aux utilisateurs d'enregistrer des demandes sous forme de commandes ou suivant un principe d'abonnement. Il assure également l'information des utilisateurs quant à l'avancement du traitement de leurs commandes. Pour effectuer ces demandes le processus devra mettre à disposition des utilisateurs un système de consultation accessible en ligne s'appuyant sur les données descriptives ;
- **P3.**Exécuter requêtes : Ce processus lance les requêtes destinées à rechercher les éléments réclamés par l'utilisateur et assure le lien avec la fonction stockage afin d'obtenir les contenus d'information désirés. Ce processus devra également contrôler l'intégrité de l'information obtenue en retour avant de la transmettre à l'utilisateur ;
- **P4.**Mettre en forme : Ce processus consiste à préparer les paquets d'information diffusés, résultat de la recherche, avant leur communication ;
- **P5.**Communiquer : Comme son nom l'indique ce processus revient à communiquer les paquets d'informations diffusés aux utilisateurs. En fonction du type de demande, la communication des résultats de la recherche pourra être obtenue soit directement en ligne, soit être transmise sur tout autre support ;

Au-delà de ces différents processus, l'Administration devra également préciser ses demandes spécifiques concernant :

- Le nombre de consultations envisagées (minimum, maximum, moyenne) de façon globale et simultanée en fonction de la nature de l'archive (intermédiaire ou définitive) et de son type (dossiers comptables, individuels, contentieux, judiciaires,…) ;
- Le nombre a priori de communications à prévoir en fréquence, volumétrie et type (télétransmission ou support physique) en fonction de la nature et du type d'archive ;
- Les différents impératifs de temps d'accès aux archives en fonction de leur nature et de leur type ;

F5. Administration du système d'archivage électronique

Exploitation

- Gérer la configuration du matériel et des logiciels du SAE consistant à en assurer la maîtrise technique destinée à surveiller en permanence son fonctionnement global ;
- Contrôler l'exploitation du SAE, de son fonctionnement et de ses performances en fonction de l'utilisation qui en est faite en fournissant entre autres des statistiques détaillées. Par ailleurs dès qu'une anomalie qu'elle quelle soit est détectée une alerte doit automatiquement être générée et transmise pour information et traitement ;

Sécurité

- Contrôler l'accès physique au SAE en fonction des règles de sécurité définies et des dispositifs de sécurité adoptés en conséquence ;
- Assurer la protection de l'ensemble des données gérées par le SAE dont certaines sont confidentielles : contenus d'informations, informations descriptives, données de gestion. Ce processus devra assurer la sauvegarde globale de l'ensemble des informations ;
- Permettre la restauration totale ou partielle des données suite à un sinistre ;
- Assurer la traçabilité complète de tout ce qui se passe dans le SAE au travers de la gestion d'un journal des événements y compris le suivi de résolution des incidents rencontrés quelle qu'en soit l'origine. Ce processus devra également permettre l'enregistrement des tentatives d'accès par des utilisateurs non autorisés ;

Gestion

- Permettre un suivi des commandes de communication en cours afin de pouvoir renseigner les utilisateurs sur l'avancement des traitements ;
- Gérer les données administratives comme celles relatives aux utilisateurs afin d'en assurer le suivi et permettre le cas échéant la production des éléments de facturation résultants des commandes effectuées ;
- Vérifier et garantir l'intégrité de l'ensemble des données administratives directement liées à l'exploitation vis-à-vis des utilisateurs mais aussi en interne ;

Conformité

- Élaborer et maintenir des standards et règles applicables au SAE comme les normes ou formats applicables, l'ensemble des procédures à suivre pour les opérations de Versement ou de migration pour le stockage afin d'éviter l'obsolescence du SAE ;
- Gérer les protocoles de versement avec les services versants en définissant les modalités d'échange et de transfert, un échéancier de Versement des Paquets d'informations versés, les besoins associés en matière de ressources ;

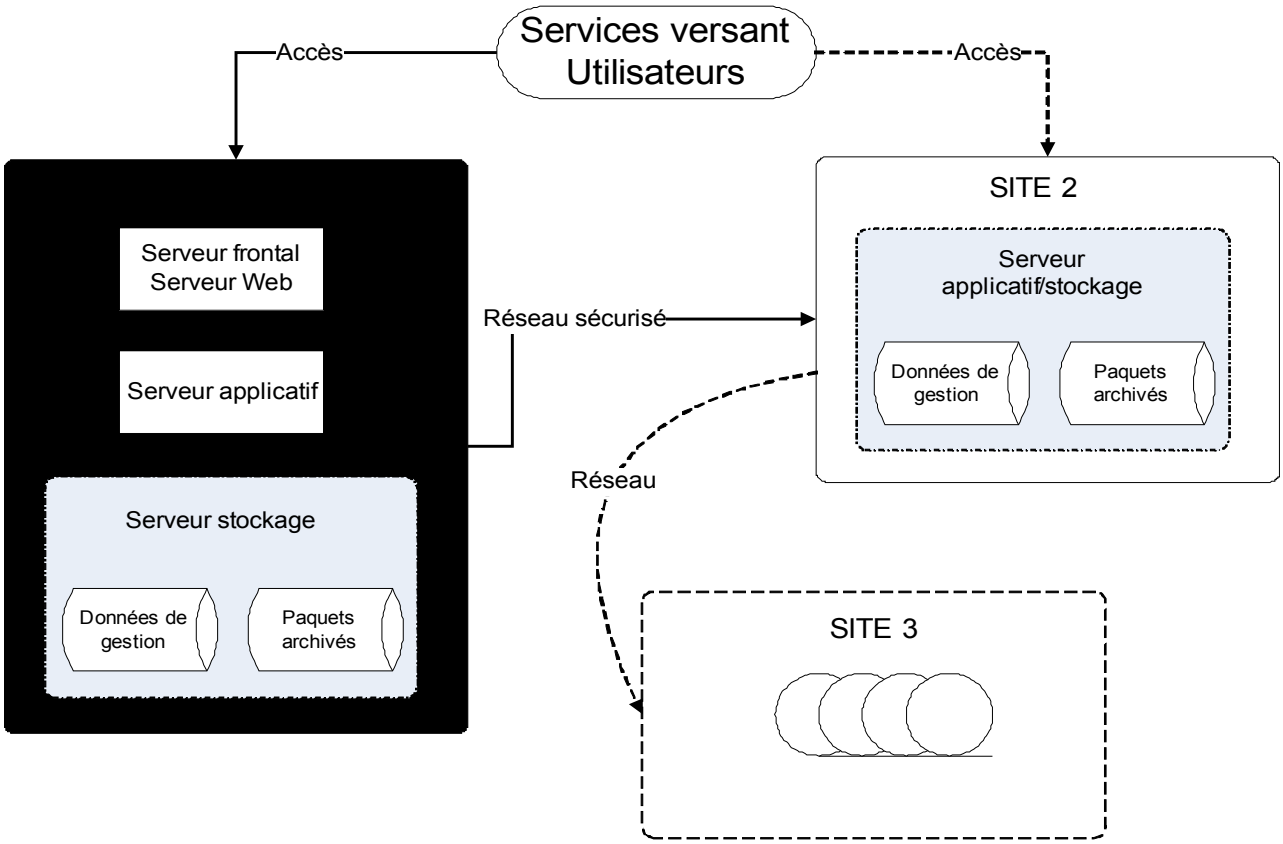
Pérennisation

- Élaborer la stratégie de l'Archive en matière de conservation à long terme des données numériques. Pour cela, l'administration effectue un triple suivi : une veille technologique visant à anticiper l'obsolescence technologique des logiciels, matériels, systèmes d'exploitation, des supports, des formats, ainsi qu'un suivi de l'activité de la communauté des Producteurs et enfin un suivi des Utilisateurs, qui peuvent formuler des exigences nouvelles. En fonction de ces observations, l'entité propose à l'Administration des stratégies générales de pérennisation. Par exemple, elle définit des standards de métadonnées nécessaires à la pérennisation des AIP, ou bien encore des plans de migration des Informations numériques qui peuvent concerner aussi bien les supports que les formats.

L'ensemble des opérations d'exploitation doit être transparente du point de vue de l'utilisateur.

2. Architecture du SAE

On remarque que dès lors qu'on évoque un SAE, sont rendues obligatoires afin de pallier des pertes de données, les duplications/réplications des archives sur des sites distants



Le tableau ci-dessous permet de mettre en évidence les différentes configurations possibles et leur évolution en fonction de la disponibilité des accès au service d'archivage d'un point de vue global. Il est clair que cette accessibilité devra être relayée par l'utilisation des systèmes de stockage ad hoc permettant entre autres un accès en ligne à de gros volumes d'information. Il faudra ainsi éviter d'avoir des systèmes de stockage off line sur bandes si l'on désire un accès direct aux données.

Les informations indiquées au niveau des risques constituent en fait le risque résiduel existant en fonction de la configuration retenue.

SITE 1	Type réplication entre sites 1 et 2	SITE 2	Type transmission entre Sites1/2 et 3	SITE 3	Risques résiduels
Accès simple Stockage simple	Sauvegarde en continu	Stockage simple			- Possibilité de problème matériel sur site 1 et arrêt du service ; - Suite à un problème sur site 1, possibilité d'une perte de données dont l'importance est directement liée à la fréquence de réplication ; - Possibilité de problèmes particuliers mais rares conduisant à la modification d'intégrité d'une partie des données que le système ne pourra pas corriger dans son intégralité.
Accès simple Stockage doublé	Sauvegarde en continu	Stockage simple ou doublé			- Faible possibilité de problème matériel sur site 1 et arrêt du service ; - Suite à un problème sur site 1, faible possibilité de perte de données dont l'importance reste liée à la fréquence de réplication ; - Possibilité de problèmes particuliers mais rares conduisant à la modification d'intégrité d'une partie des données que le système ne pourra pas corriger dans son intégralité.
Accès doublé Stockage doublé Serveur applicatif en secours	Synchronisation	Stockage simple ou doublé	Manuelle à fréquence régulière, a priori la journée	Armoire de bandes	- Très faible possibilité de problème matériel sur site 1 conduisant à l'arrêt du service ; - Risque quasi inexistant de perte de données ; - Possibilité très faible de modification d'intégrité, de toute façon limitée aux données entre deux réplications de bandes.
Accès doublé Stockage doublé	Synchronisation	Stockage simple ou doublé	Réseau	Robot de bandes ou autre support (Facilite l'exploitation)	- Très faible possibilité de problème matériel sur site 1 conduisant à l'arrêt du service ; - Risque quasi inexistant de perte de données ; - Possibilité très faible de modification d'intégrité, de toute façon limitée à la fréquence de réplication retenue, 1h 1/2journée.
Accès doublé Stockage doublé Serveur applicatif en secours	Synchronisation	Accès doublé Stockage doublé	Réseau	Robot de bandes ou autre support	- Risque d'arrêt du service quasi inexistant ; - Risque quasi inexistant de perte de données ; - Possibilité très faible de modification d'intégrité, de toute façon limitée à la fréquence de réplication retenue, 1h 1/2journée.

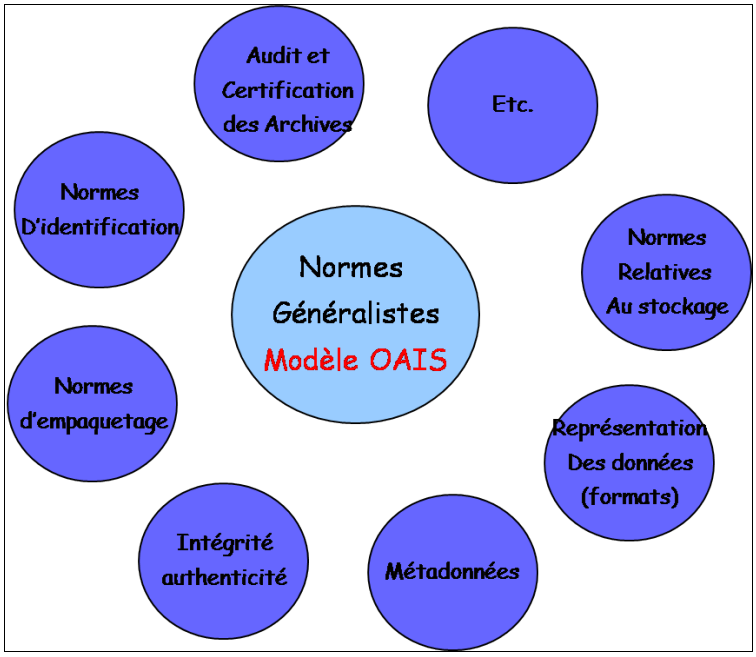
Les conditions de l'interopérabilité et de la préservation sur le long terme : le respect du cadre normatif

Les normes généralistes de l'archivage électronique

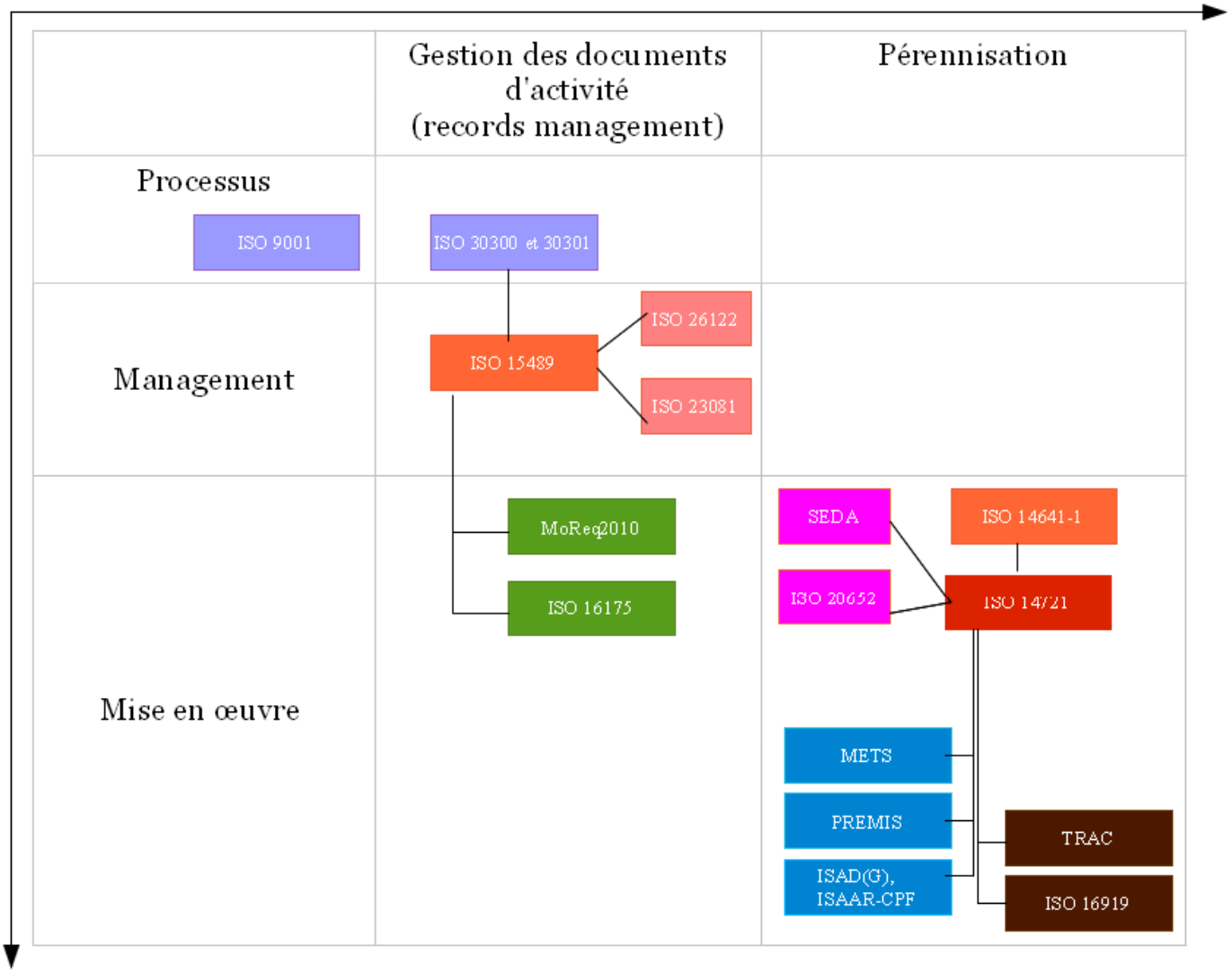
Plan de la fiche :

- Organisation et processus
- Pérennisation

Les principaux domaines normatifs de l'archivage électronique
(Sources : *Portail international archivistique francophone*, Françoise Banat-Berger et Claude Huc)



D'une manière très synthétique, on peut dire que le domaine appelé « archivage numérique », recouvre deux champs d'activités et de compétences : l'organisation et les processus d'une part, les aspects de pérennisation d'autre part.



1. Organisation et processus

Il s'agit là principalement de la discipline du « *records management* », qui est régi par la **norme ISO 15489**, et que certains traduisent par « gestion de l'archivage », « maîtrise de l'information numérique » ou « gestion des archives courantes et intermédiaires » ou encore, suivant la commission de normalisation ad-hoc (CN 11), « gestion des informations et documents d'activité ».

Selon la commission de normalisation française en charge du *records management*, les principaux objectifs de cette discipline s'énoncent ainsi :

> Fournir aux entreprises et aux administrations un cadre normalisé pour **maîtriser le cycle de vie des documents** produits et conservés, **quel qu'en soit le support** (numérique, papier ou microformes).
>
> Le *records management* permet à l'organisme qui l'a mis en place de disposer à tout instant des documents et des données dont il a besoin pour conduire ses activités, d'assurer leur traçabilité, de les poursuivre en cas de sinistre, de défendre ses intérêts en cas de litige, de réduire les risques et de répondre aux exigences légales et réglementaires en matière de conservation de documents.

Il convient de signaler la parution des normes internationales **ISO 30300** et **ISO 30301** relatives aux systèmes de gestion des informations et documents d'activité (management des *records*). L'ISO 30300 est consacrée aux principes essentiels et à la terminologie. L'ISO 30301 traite des exigences relatives à ces systèmes, normes traduites en français grâce aux travaux menés par la CN 11.

Comme l'a souligné le livre blanc publié par la CN11 au printemps 2011[32], la publication de ces deux nouveaux textes constitue le début d'une nouvelle série de normes consacrées au *records management* et revêt une grande importance, dans la mesure où elles apportent au *records management* le cadre managérial qui lui faisait jusqu'à présent défaut.

Les normes de la série 30300 offrent une vision complémentaire de l'ISO 15489, désormais bien connue. Lors de sa rédaction, l'ISO 15489 avait en effet pour objet de mieux faire connaître et de promouvoir le *records management*. Cette norme, toujours d'actualité aujourd'hui[33], explicite d'une part ce qu'est le *records management* et à quoi il sert, indique d'autre part ce que doit savoir faire un système de *records management* en formalisant les actes opératoires relatifs au cycle de vie des documents et fournit enfin au *records manager* toutes les clés méthodologiques pour concevoir un système de gestion des *records* (partie 2 de la norme).

La série des normes 30300 adopte un angle d'approche différent mais complémentaire. Il s'agit de normes de systèmes de management, découlant directement de l'ISO 9001, des démarches qualité et du management par processus. Elles décrivent les processus qui permettent de concevoir, de mettre en œuvre et de contrôler un système de gestion des *records* répondant aux caractéristiques décrites dans l'ISO 15489.

Ces normes sont donc très tournées vers l'auto-évaluation, l'amélioration continue du système, l'audit et, à terme, la certification, autant de caractéristiques destinées à garantir la fiabilité du système. Les prochaines normes de la série 30300 seront d'ailleurs consacrées à ces aspects (**ISO 30302 :** guide de mise en œuvre des exigences décrites dans la 30301, **ISO 30303 :** exigences pour les organismes d'audit et de certification, **ISO 30304 :** guide pour l'évaluation du système).

La norme du *records management* à son tour a donné lieu à des spécifications plus détaillées, normalisées ou non, mais reconnues par des institutions ou organismes comme le conseil international des archives ou la commission européenne.

- **Commission européenne :** *MoReq2010 - Modular Requirements for Records Systems, © 2010 & 2011 DLM Forum Foundation*[34] ; on pourra également consulter son prédécesseur, MoReq2[35];
- **Conseil international des archives :** *ICA-Req : Principes et exigences fonctionnelles pour la gestion des archives dans un environnement électronique*, 2010[36]). Ces spécifications ont été normalisées en 2011 sous le numéro ISO 16175 dont une traduction française a été rédigée par un groupe de travail porté par le Conseil international des archives[37].

32 *Introduction à la série de normes ISO 30300, système de management des documents d'activité : intégration du* records management *et perspectives d'évolution de l'ISO 15489. Livre blanc.*
En ligne : **http://www.bivi.fonctions-documentaires.afnor.org/livres-blancs/introduction-a-la-serie-de-normes-iso-30300-systeme-de-management-des-documents-d-activite**

33 L'ISO devrait entamer sa révision dans le courant de l'année 2011.

34 **http://moreq2010.eu/** ;

35 **http://www.archivesdefrance.culture.gouv.fr/gerer/archives-electroniques/standard/moreq2/**

36 **http://www.archivesdefrance.culture.gouv.fr/gerer/archives-electroniques/standard/ica-req/**

37 Cette version est disponible auprès de l'ICA.

Ces deux textes, MoReq et ICA-Req, sont particulièrement utiles pour la rédaction de cahiers des charges, l'analyse des offres ou l'évaluation de systèmes existants.

On conseillera de se reporter aux spécifications Ica-Req plus simple d'accès que les spécifications MoReq2010. Le module 1 définit le champ d'application de la norme, ses objectifs et les publics cibles. Le module 2 définit les spécifications d'un système d'archivage numérique. 275 exigences fonctionnelles ont ainsi été divisées en quatre sections: production et capture, maintenance, mise à disposition, administration, trois niveaux d'exigence ayant été retenus et étant exprimés à l'aide de trois verbes différents : l'outil doit, l'outil devrait, l'outil peut.

On signalera également **le troisième module d'ICA-Req** qui décrit les exigences en matière de cycle de vie et d'archivage pour une application métier. Ce module, qui n'a pas d'équivalent dans MoReq, peut être utilisé notamment pour auditer des systèmes d'information. 125 exigences fonctionnelles ont été définies qui sont divisées en quatre sections : création des documents dans leur contexte, gestion des documents permettant de garantir leur intégrité, import, export et interopérabilité, gestion du cycle de vie (sort final notamment).

La boîte à outils de la norme comprend notamment : « Une **méthodologie pour l'audit des applications métier avec le module 3 »** d'ICA-Req à destination des archivistes, rédigée par Lourdes Fuentes-Hashimoto (MAEE), ainsi qu' « Un **guide pour l'archivage des données et des documents électroniques avec la norme ICA-Req, à l'attention des services informatiques de l'administration publique »**, rédigé par Alice Château (MIOMCTI) et disponible au sein de la mission des archives de ce ministère.

Ces familles de normes visent à ce que, au sein d'une organisation donnée, tous les documents dits « vitaux » (« essentiels ») soient pris en charge, enregistrés, classifiés, et puissent, durant toute leur durée de vie, être conservés intelligibles, intègres, en garantissant notamment leur valeur de preuve pour l'organisation.

2. Pérennisation

La préservation pérenne des données et documents numériques est entièrement régie par une norme internationale au départ élaborée par des grandes institutions scientifiques et patrimoniales, notamment les domaines de l'aéronautique et du spatial qui, depuis une quarantaine d'années, ont de très importants volumes de données[38] complexes et nécessitant de très longues durées de conservation (en tous les cas, supérieures à 10 ans). Cette norme a été rendue nécessaire par la spécificité du numérique qui est volatile (extrême obsolescence des formats, supports, systèmes d'information, périphériques).

Il s'agit de **la norme OAIS** (Open Archival Information System) devenue **la norme ISO 14721.** Le modèle de référence pour un OAIS (*Open Archival Information System*, soit système ouvert d'archivage de l'information) **décrit les responsabilités, les fonctions et les rapports avec son environnement d'un système d'archivage électronique pour assurer la pérennisation de l'information numérique.**

- Note d'information **DGP/SIAF/2011/010** du 8 juin 2011 relative au modèle de référence pour un système ouvert d'archivage d'information OAIS
- **Lien vers le texte de la norme OAIS** (version 1.0, traduction française de 2005)

38 Nous sommes au-delà des tera octets de données : l'ordre de grandeur est le peta octet soit 1000 fois plus.

À ces deux grandes familles de normes généralistes, on doit ajouter au niveau français, la norme **Afnor Z 42-013** dans sa nouvelle version de mars 2009, relative aux spécifications techniques d'un système d'archivage numérique, portée au niveau international : **norme ISO 14641** (janvier 2012).

La norme NF Z 42-013 est antérieure à la norme OAIS, puisque sa première version date de 1999, mais a été révisée dernièrement, en 2009. Cette mise à jour a permis d'élargir le périmètre de la norme à tous les types de supports numériques et non plus aux seuls disques optiques numériques non-réinscriptibles.

Il s'agit plutôt d'une norme technique que d'une norme fonctionnelle. Elle met tout particulièrement l'accent sur la traçabilité de tous les processus en œuvre dans l'archivage électronique comme la numérisation de documents, l'horodatage, les communications… ainsi que sur les exigences du système en matière de sécurité et d'accès. Elle est aussi importante en ce qu'elle définit les clauses nécessaires dans un contrat de service passé avec un tiers archiveur. Les exigences de la norme sont de deux types : des exigences minimales et des exigences complémentaires.

Un manuel d'application a par ailleurs été élaboré par les membres de la commission de l'Afnor qui a élaboré la norme[39].

Enfin vient de sortir la norme afnor Z 42-020 relative aux solutions de **coffres-forts numériques**. Celles-ci visent à sécuriser les données et documents à valeur de preuve. Un coffre-fort électronique est par conséquent un « composant d'un système d'information constitué d'un logiciel ou d'une combinaison logiciel/matériel qui permet de préserver l'intégrité d'objets numériques dans le temps »[40]. Les SAE peuvent ainsi piloter un tel composant.

39 Guide d'application de la NF Z42-013 (Archivage électronique - 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

40 Définition de la norme afnor Z 42-020 : *« Spécifications fonctionnelles d'un composant Coffre-fort Numérique destiné à la conservation d'informations numériques dans des conditions de nature à en garantir leur intégrité dans le temps* », 2012.

Les normes spécialisées de l'archivage électronique

Plan de la fiche :

- Les formats de représentation de l'information
- Les formats d'échanges
- Les supports
- Les formats de métadonnées

1. Les formats de représentation de l'information

Qu'il s'agisse de texte, d'image (2D, 3D, vectorielle, matricielle), de documents multimédia (audio, vidéo), de données structurées (issues de SGBDR, de documents XML), toutes les formes d'information numérique se réfèrent à des formats dès lors qu'il s'agit de les stocker.

Il est évidemment impossible de lister l'ensemble des formats de données existants. Nous en donnerons par conséquent quelques exemples, parmi notamment les plus utilisés par les administrations et/ou ceux sur lesquels des études et recherches liées aux formats cibles pour la conservation pérenne ont été effectuées.

Dans les formats généralistes, on trouvera par exemple le PDF, les principaux formats bureautiques ODF et OOXML, le méta-langage XML, le format d'image PNG. Le développement spectaculaire de l'usage méta-langage XML a constitué une évolution majeure dans le domaine de l'accès, du partage et de l'archivage des données. PDF, ODF et OOXML reposent aujourd'hui largement sur ce méta-langage.

Dans les formats « métier », on pourra citer les différentes applications du langage STEP (STandard for the Exchange of Product model data) pour l'industrie. Les normes relatives aux données géographiques tiennent également une place importante.

Critères à retenir dans le choix des formats

(Source : *CNES, Référentiel normatif, Ingénierie des données. Evaluation des formats de données par rapport à la pérennisation*, 2005)

Un format informatique est une convention sur la représentation d'une donnée sur un support numérique. Il peut être :

Spécifié : il existe une description de la convention utilisée pour représenter la donnée et celle-ci est suffisamment décrite pour en développer une implémentation complète.

Ouvert : la convention est publique (sinon le format est dit fermé). Elle est donc sans restriction d'accès ni de mise en œuvre.

Normalisé : la convention est adoptée par des organismes de normalisation (ISO, W3C). Exemple : le PDF/A.

Standardisé : il n'existe pas de norme sur ce format mais son utilisation est tellement répandue qu'il est considéré comme un standard. Exemple : le PDF. ATTENTION : en anglais « standard » signifie « norme ».

Propriétaire : si l'exploitation du format entre dans le cadre du droit privé, il dépend

alors de l'existence du propriétaire (par exemple les formats doc de Microsoft). Cet aspect propriétaire n'empêche par forcément les spécifications d'être publiées. Exemple : les spécifications du format PDF ont toujours été publiées par Adobe.

Ces cinq critères permettent de définir le niveau de pérennité d'un format.

- un format de données doit être entièrement et explicitement spécifié et sa spécification doit être connue du service en charge de l'archivage long terme ;
- un format de données doit être apte à représenter la sémantique et la complexité de l'information à préserver ;
- l'usage de formats normalisés est recommandé. En outre, on évitera l'usage d'éléments propriétaires au sein d'un format normalisé ;
- lorsque le besoin de pouvoir modifier un document est identifié, le choix du format doit prendre en compte cette contrainte ;
- le choix des formats doit pouvoir prendre en compte la disponibilité et le coût des outils d'exploitation ;
- il doit être possible de vérifier automatiquement qu'un fichier de données respecte les spécifications du format et respecte également les règles d'utilisation et des restrictions qui auront été éventuellement définis pour la pérennité ;
- la possibilité d'extraire automatiquement tout ou partie des métadonnées à partir des données constitue un avantage certain ;
- à fonctionnalités égales un format simple est préférable à un format complexe ;
- les formats largement reconnus et utilisés seront privilégiés ;
- le choix d'un format doit prendre en compte la disponibilité et le coût des outils de transformation des formats et de représentation des données.

Voir le *Référentiel général d'interopérabilité*, partie technique

Formats bureautiques et non structurés

HTML, Hypertext Markup Language pour les pages Web
Standardisé par le consortium W3C
Normalisé par l'ISO en 2000 (ISO 15445:2000)

PDF – PDF/A
PDF - format propriétaire (ADOBE)
PDF/A-1 est devenu la norme ISO 19005-1 en 2005
PDF 1.7 est devenu la norme ISO 32000-1 en 2008
PDF/A-2 nouvelle version de PDF-A publiée.(juin 2011) : elle s'appuie sur la version 1.7 de PDF lui même normalisé en ISO 32000-1

ODF, Open Document Format, pour les documents bureautiques
La version 1.0 est standardisée par OASIS en 2005
puis normalisée par l'ISO en 2006 (ISO 26300)
La version 1.1 est standardisée par Oasis en 2007.puis l'ISO en 2012.
La version 1.2 est actuellement un standard Oasis depuis 2011.

OOXML, Office Open XML, pour les documents bureautiques
Standardisé par ECMA en 2006
Normalisé par ISO en 2008 (ISO 29500)

Formats image

PNG, Portable network Graphics, pour les images matricielles. A l'origine, format crée pour offrir une alternative libre au format GIF qui utilise une technique de compression sans perte LZW soumise à un brevet.
Standardisé par le W3C en 1996
Normalisé par l'ISO en 2004 (ISO 15948:2004)

GIF, Graphics Intergange Format est un format ouvert de la société CompuServe pour la représentation d'images matricielles. Il utilise une compression sans perte LZW dont le breuvet a maintenant expiré.
Version GIF87a
Version GIF89a permet l'inclusion de plusieurs images dans un fichier

JFIF, JPEG File Interchange Format est un format pour la représentation d'images matricielles compressées avec l'algorithme JPEG (Joint Photographic Experts Group)
L'algorithme JPEG est défini par la norme ISO 10918-1 en 1993

JPEG2000,Norme ISO 15444-1 de 2000 utilisant pour la compression (avec ou sans perte) des images un algorithme une transformée en ondelettes permettant des meilleurs taux que l'agorithme de la norme ISO 10918-1.

TIFF, Tagged Image File Format est un format conteneur propriétaire de Adobe pour des images numériques.

Formats d'archivage pour les bases de données

Note d'information **DGP/SIAF/2010/017** du 21 septembre 2010. Étude du format SIARD pour **l'archivage des bases de données relationnelles** et au logiciel SIARDSuite mettant en œuvre ce format.

Note d'explications sur les concepts du modèle relationnel
Liste de questions/réponses sur le logiciel SIARDSuite
Description technique du jeu de test

Formats d'archivage pour les documents audiovisuels

Guide méthodologique pour le choix de formats numériques pérennes dans un contexte de données **orales et visuelles** : **version initiale** ; **nouvelle version mise à jour en 2011**
Note d'information **DGP/SIAF/2010/010** du 21 mai 2010

Outils d'identification des formats

Il s'agit d'identifier au sein d'un référentiel (PRONOM, MIME) le type et/ou le format utilisé par un fichier. Cette famille d'outils ne vérifie en général pas le respect de toutes les spécifications du format mais se base sur des « signatures » caractéristiques basées sur les extensions, les entêtes, etc.

DROID (Digital Record Object Identification) des Archives nationales de Grande-Bretagne. Utilise le registre de formats PRONOM également maintenu par eux.

FIDO (Format Identification for Digital Objects) de la fondation OpenPlanets. Utilise le registre de formats PRONOM

TIKA content analysis toolkit from the **Fondation Apache**. Utilise le référentiel MIME de IANA.

Outils de validation des formats

Il s'agit de s'assurer que le fichier est bien-formé et valide par rapport à la spécification de son format.

Un fichier est considéré comme "bien formé" s'il répond aux exigences purement syntaxiques.

Un fichier est considéré comme "valide" s'il répond aux exigences sémantiques du format.

C'est ainsi que le Centre informatique national de l'enseignement supérieur (CINES) propose d'utiliser un outil qu'il a développé en intégrant différents outils du marché. Ce projet « easy » a été déposé sur les forges de l'Adullact**. Une interface web à cette application est également disponible sur le site du CINES « FACILE »** pour « *Validation du Format d' Archivage du CINES par anaLyse et Expertise » à* **l'adresse http://facile.cines.fr/**

L'outil de validation « File Information Tool Set » (FITS) a été développé par la bibliothèque de l'université de Harvards et est disponible librement sur **http://code.google.com/p/fits/**

Ces deux outils (Facile et FITS) utilisent tous les deux des composants communs dont les logiciels DROID et Jhove respectivement pour l'identification et la validation et complète cette expertise par d'autres composants plus spécialisés tels que ImageMagick ou Exiftool)

2. Les formats d'échanges

Depuis 2006, les archives de France ont élaboré en collaboration avec la direction générale de la modernisation de l'Etat (DGME), l**e standard d'échange de données pour l'archivage (SEDA), actuellement en cours de finalisation dans une version 1.0.**

Des ressources (documentation, schémas XML, exemples de code, feuilles de styles, logiciels) sont disponibles sur le site des Archives de France **http://www.archivesdefrance.culture.gouv.fr/seda/**

Un comité est en charge de la maintenance et de l'évolution du SEDA. Il rassemble éditeurs de logiciels métier et logiciels d'archives, tiers archiveurs, services publics d'archives, éditeurs de plateformes de télétransmission

Pour tout renseignement, il convient d'envoyer un message à l'adresse suivante : **seda@culture.gouv.fr**

Présentation du standard d'échange de données pour l'archivage (SEDA)

Liste des notes d'information et instructions en rapport avec le SEDA
Instruction **DGP/SIAF/2010/002** du 15 février 2010. Nouvelle version du standard d'échange de données pour l'archivage.
Note d'information **DGP/SIAF/2011/005** du 14 mars 2011 relative à la publication d'un outil d'aide à la constitution d'un profil d'archivage : AGAPE.
Note d'information **DGP/SIAF/2010/024** en date du 13 janvier 2011 relative à la transformation du SEDA (standard d'échange de données pour l'archivage version 0.2) au format EAD (Description archivistique encodée, version de 2002).
Annexes :
– un **tableau de correspondances** entre les éléments et attributs du SEDA et ceux de l'EAD ;
– une **feuille de style** pouvant être utilisée pour transformer un fichier au format du SEDA en un fichier valide et bien formé au format EAD.

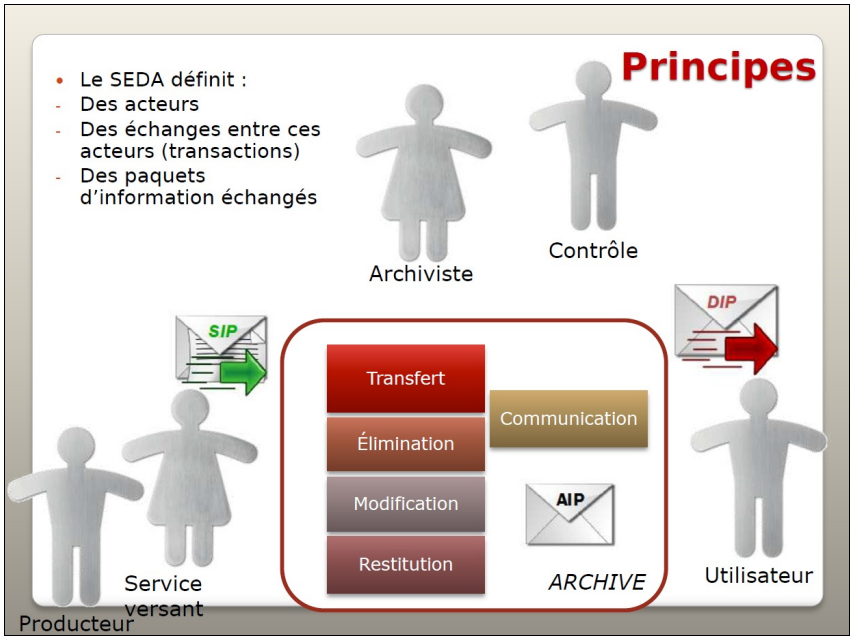
Présentation générale

Le *Standard d'échange de données pour l'archivage* **modélise les différentes transactions qui peuvent avoir lieu entre des acteurs dans le cadre de l'archivage de données**.

Ces **transactions** sont au nombre de six :
- le transfert de données,
- la demande de transfert,
- la modification de métadonnées,
- l'élimination des données,
- la communication des données
- la restitution des données.

Les **acteurs** sont eux au nombre de cinq :
- le service producteur,
- le service versant,
- le service d'archives,
- le service de contrôle,
- le demandeur d'Archives.

Acteurs et transactions du SEDA



Ce standard s'inspire de la norme OAIS[41] qui lui fournit d'une part les concepts de base et le vocabulaire de l'archivage numérique, et d'autre part la méthodologie de l'UN/CEFACT[42] pour la forme des messages échangés (flux XML).

L'intégration du SEDA dans les systèmes d'information, vise à éviter les ruptures de charge entre les différents partenaires et par exemple à éviter que des données descriptives identifiant des dossiers, qui sont enregistrées dans un système d'information, soient re-saisies manuellement par les services producteurs préalablement au versement des dossiers, sous la forme d'un bordereau de versement, puis re-saisies ensuite par le service d'archives dans son propre système d'information. La dématérialisation croissante des procédures et la masse des fichiers à verser qui en résultera imposent de passer à un processus de transfert automatisé

Ce standard définit de manière formelle :

- Les scénarios à mettre en œuvre pour chaque transaction ;
- Les messages que les acteurs doivent/peuvent s'échanger ;
- La manière de coder les informations à transmettre (métadonnées consignées dans un bordereau de versement XML, bordereaux d'élimination notamment - et contenus de données) : **il fournit en somme un modèle de métadonnées (schéma XML) pour décrire les objets échangés au cours d'une des transactions**.

Le standard précise le contenu et la structure des messages échangés, suivant qu'on souhaite effectuer un versement, éliminer des archives, communiquer ces archives, ou éventuellement les restituer.

41 *Open Archival Information System.* Norme ISO 14721:2003.

42 *United Nations Centre for Trade Facilitation and Electronic Business* (http://www.unece.org/cefact/). Organisme des Nations Unies qui assure la promotion, en accord avec l'ISO, du développement et de la simplification des échanges électroniques professionnels, du commerce électronique et des procédures administratives.

Chaque transaction y est décrite comme un dialogue dans lequel les partenaires s'échangent des messages dans un ordre et dans une forme précisés. Par exemple, pour le transfert qui va faire dialoguer un service d'archives avec un service versant, se succèdent :

- Un message initial de transfert (composé d'un en-tête et d'un bordereau de versement) accompagné des données elles-mêmes ;
- un message d'accusé réception ;
- un message de notification d'acceptation ou d'avis d'anomalie ;
- si nécessaire, un message d'accusé réception d'avis d'anomalie.

Une gestion de flux (« workflow ») est ainsi définie[43]. Par ailleurs, le SEDA définit et précise le contenu des bordereaux qui seront générés automatiquement suivant les transactions prévues (bordereaux de versement, bordereaux d'élimination, bordereaux de communication). La description des archives elle-même suit les principes de la norme ISAD(G)[44]. Le formalisme proposé par son implémentation en XML (DTD EAD[45]), n'a pas pu être directement repris dans le standard en raison des règles de l'UN/CEFACT. En effet, cet organisme impose de reprendre la terminologie existante et, pour l'ajout de nouveaux termes, fixe des contraintes d'écriture à respecter (termes en anglais, règles de composition, etc.). Les équivalences de termes entre les deux standards ont été mis en évidence lors de la première version du SEDA.

L'utilisation du standard d'échange de données pour l'archivage

La mise en place du standard d'échange de données pour l'archivage dans le cadre d'un contexte métier demande, outre la mise en conformité technique des applications de gestion (respect des schémas et des dialogues), un travail de préparation entre les différents partenaires.

Les services concernés (en particulier les services d'archives et les services versants) doivent s'accorder sur les modalités de mise en œuvre de leurs échanges (transfert par réseau, sur support amovible...), fréquence des envois, niveau de service attendu. Cet accord doit faire l'objet d'une « convention », qui sera identifiée dans les messages échangés. A terme, il conviendra de modéliser cette convention de manière à ce que les éléments qu'elle comportera permettent d'automatiser un certain nombre de contrôles[46].

43 Toutefois, il est bien évident que si l'on prévoit des transferts manuels (archives numériques et leur description gravées sur des supports amovibles), l'utilisation de ces messages au format du SEDA n'a aucune utilité. Ces messages sont à utiliser lorsque des informations numériques sont envoyées de " machine " à " machine " par réseaux sécurisés, de manière à ce que ces messages et leur contenu puissent être interprétés automatiquement par les systèmes. Dans ce cas, seul le bordereau de versement suivant le modèle de description du standard sera gravé sur le même support.

44 Norme générale et internationale de description archivistique maintenue par le Conseil international des archives **http://www.ica.org/**

45 Standard maintenu par la bibliothèque du Congres (**http://www.loc.gov/ead/**) pour l'encodage des instruments de recherche (Encoded Archival Description). Ce standard définit un modèle de documents en XML suivant les principes de la norme ISAD/G.

46 Par exemple, s'il est prévu dans la convention, que tel versement ne doit pas dépasser telle volumétrie, le système pourra générer un message d'erreur si la volumétrie est supérieure à ce seuil.

Il convient également :

-d'identifier précisément les différents partenaires concernés : service producteur, service versant, tiers de télétransmission, services informatiques ;

-de déterminer, avec le service producteur, les durées de conservation et sorts finaux des documents et données, si ceux-ci ne sont pas déjà définis dans un tableau de gestion ;

-d'élaborer une stratégie d'archivage de manière à déterminer quand et comment ces données et documents feront l'objet d'une élimination réglementaire, ou bien d'un versement au format du SEDA vers une plate-forme d'archivage ;

-de déterminer, dans le cas d'un export au format du SEDA, les conditions de ce versement : automatisé ou non, fréquence, forme du bordereau de versement, niveau de service (disponibilité de l'application, temps d'accès à l'information).

Il convient ainsi de définir un **profil**. Afin d'élaborer celui-ci, différentes éléments doivent être précisés :

- le plan de classement qui détermine les différents niveaux de description retenus, doit être défini, comme cela est généralement fait avant toute élaboration d'un instrument de recherche,
- ensuite, il convient de déterminer, en étroite collaboration avec le service producteur et le service informatique, suivant les « champs » (éléments et attributs) définis par le SEDA, le contenu à y intégrer, selon les différents niveaux de description :
  contenus strictement archivistiques (niveaux de description, durée de conservation, sort final, mots-clés issus de thesaurus réglementaires, délai de libre communicabilité…) qui seront indiqués dans les spécifications
  contenus qui seront récupérés automatiquement depuis le système d'information de production (notamment des informations descriptives pouvant être intégrées dans les champs correspondants aux « intitulé », « présentation du contenu », « autres données descriptives »….)

Enfin, la structure des documents ou des données eux-mêmes devra être étudiée avec soin[47].

Ainsi, sur la base de ces spécifications, le système à partir duquel se fera l'export pourra générer, à chaque versement, un ensemble se composant d'une part des fichiers et d'autre part du bordereau de versement automatiquement généré.

Les ressources sur le SEDA et sa boîte à outils

Toutes les informations sur le *Standard d'échange* sont depuis 2009 disponibles sur le site Internet des « Documents de référence de l'administration électronique » de la direction générale de la modernisation de l'État (ministère délégué au Budget et à la réforme de l'État).

Ces informations sont composées d'un document de présentation, de schémas XML qui

47 Par exemple, pour le revenu minimum d'insertion, un travail a été accompli par les Archives départementales du Finistère, visant à faire constituer pour chaque bénéficiaire du RMI, un dossier se composant d'un certain nombre d'éléments concernant ce bénéficiaire issu des différentes tables de l'application métier dans lesquelles ces informations étaient jusqu'alors distribuées.

représentent la mise en œuvre technique du modèle ainsi que de profils d'archivage qui représentent eux la mise en contexte du modèle de description pour des catégories documentaires données.

Par ailleurs, dans le but de faciliter l'utilisation de ce standard par les professionnels et les applications, les Archives de France mettent à disposition des ressources telles que des feuilles de styles, des outils d'édition, de la documentation ainsi que les schémas XML (anciennes et nouvelle version). **L'ensemble de ces ressources sont accessibles à l'adresse: www.archivesdefrance.culture.gouv.fr/seda/.**

Dans l'objectif de faciliter l'écriture des **profils d'archivage**, un nouveau logiciel a été mis au point : **Agape (Application de Génération Automatisée de Profils Électroniques**). Cet outil propose une interface graphique permettant de choisir les éléments descriptifs des schémas XML du SEDA que l'on souhaite utiliser ainsi que d'ajouter un certain nombre de contraintes sur leur nombre et leur contenu. Une fois l'édition d'un profil terminée, un tableau récapitulant tous ces choix peut être produit afin d'alimenter la documentation à fournir à la maitrise d'œuvre. Un schéma peut enfin être dérivé de ce profil afin de permettre de contrôler automatiquement qu'un message de transfert respecte bien toutes les contraintes définies. L'application, sa documentation ainsi que son code source sont publiés sur une "forge" à l'adresse **http://agape.adullact.net.**

3. Les supports

C'est tout ce qui concerne la préservation des bits. On trouvera pour cela d'une part des guides et recommandations sur la stratégie de stockage, l'organisation d'un service de stockage et d'autre part des normes sur les supports d'enregistrement, par exemple :

- la norme ISO 9660 pour les CD-Rom
- les normes ISO 13421 et 13962 pour les DLT (Digital Linear Tape).

Par ailleurs, les Archives de France[48] ont fait conduire un certain nombre d'études et réalisé des guides, en collaboration avec le Laboratoire national de métrologie et d'essais (LNE).

Recommandations relatives à la gravure, à la conservation et à l'évaluation des CD-R
Instruction **DITN/RES/2005/004** du 29 mars 2005
Recommandations
Mémento pratique

Résultats de l'étude sur des CD-R et des graveurs du marché
Note d'information **DITN/RES/2006/008** du 20 décembre 2006
Rapport de synthèse de l'étude

Résultats d'une seconde **étude sur des CD-R** et des graveurs du marché, ainsi que d'une **étude sur les DVD-R** et graveurs du marché
Note d'information DITN/RES/2008/012 du 19 décembre 2008

Guide pour la réalisation de la migration de stocks de CD-R
Note d'information **DITN/RES/2009/005** du 12 mars 2009

48 Une étude supplémentaire est en cours concernant d'une part les DVD-R (mise à jour de l'étude menée en 2008) et d'autre part sur les Blue-Ray.

4. Les formats de métadonnées

À titre d'exemple, pour la description d'ouvrages de bibliothèques, on emploie les normes Dublin-Core, MARC ou UNIMARC ; pour la description des documents d'archives, on emploiera la norme ISAD-G pour la description des documents d'archives ; pour les informations de préservation, la norme PREMIS, etc.

La définition la plus répandue des **métadonnées est que « ce sont des données à propos d'autres données »** ou encore des informations « à propos » ou « autour » d'autres informations. Même si le terme est récent, la pratique des métadonnées est, elle, assez ancienne et on la retrouve par exemple dans les fiches et notices documentaires des bibliothèques.

Il existe de multiples manières de classer les différents types de métadonnées en fonction de ce qu'elles décrivent, de la façon dont elles sont créées, du moment où elles sont créées, de leur emplacement, de leur aspect, de leur usage, etc. Le modèle d'information de l'OAIS donne lui aussi une typologie qui illustre bien le principe de continuité entre données et métadonnées (les informations servent à comprendre d'autres informations).

Il y a déjà dans la communauté de bonnes pratiques d'utilisation de métadonnées standardisées. En particulier celles pour la description des archives (**ISAD-(G)** pour les catégories d'information et la **DTD-EAD** pour la formalisation de ces informations en XML et leur publication) et celles pour la description des noms de personnes, de familles et de collectivités (**ISAAR(CPF)** qui trouve son pendant dans le **schéma XML EAC-CPF**).

À côté de ces standards, très orientés « métiers de l'archivage » sont apparus deux standards plus généralistes que sont **PREMIS** et **METS.**

METS (Metadata Encoding and Transmission Standard)

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

Développé à l'origine par la Digital Library Federation, le projet est maintenant hébergé et maintenu par la Bibliothèque du Congrès

Une présentation synthétique de METS est consultable sur le site de la Bibliothèque du Congrès à l'adresse: **http://www.loc.gov/standards/mets/METSOverview.v2_fr.html**

PREMIS (PREservation Metadata: Implementation Strategies)

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

Une introduction en français est consultable sur le site de la Bibliothèque du Congrès à l'adresse: **http://www.loc.gov/standards/premis/Understanding-PREMIS_french.pdf**

Identification des données

- norme ISBN (International Serial Bibliographic Number) pour les ouvrages de bibliothèques.

- système d'identifiant pérenne **ARK**.

Archival Resource Key (ARK) est un système d'identifiants basé sur la norme URI assurant opacité, extensibilité et indépendance, c'est-à-dire les critères nécessaires pour garantir l'identification d'une ressource sur le long terme. Les ARK peuvent désigner des **objets de n'importe quel type** : textuels, images, logiciels, sites web, aussi bien que des objets physiques, comme des livres, des statues, et même des concepts immatériels.

(Ressource : https://confluence.ucop.edu/display/Curation/ARK)

Le cadre juridique de l'archivage électronique

Le cadre juridique de l'administration électronique

Plan de la fiche :

- Intégrité des données
- Les fondements juridiques de l'authenticité des données numériques
- Une contrainte juridique supplémentaire de l'environnement numérique : la protection des données personnelles

1. Intégrité des données

Code civil et cadre de la preuve

Depuis 2000, la validité comme preuve juridique d'un document numérique est reconnue, au même titre que la preuve écrite sur papier, mais sous certaines conditions : **pouvoir justifier de l'identité de la personne dont il émane et de son intégrité**, en vertu de la **loi n°2000-230** du 13 mars 2000 portant adaptation du droit de la preuve aux technologies de l'information et relatif à la signature électronique, modifiant le Code civil.

Le Code civil stipule aujourd'hui :

Art. 1316. - La preuve littérale ou preuve par écrit résulte d'une suite de lettres, de caractères, de chiffres ou de tous autres signes ou symboles dotés d'une signification intelligible, **quels que soient leur support et leurs modalités de transmission**.

Art. 1316-1. - L'écrit sous forme électronique est admis en preuve au même titre que l'écrit sur support papier, sous réserve que puisse être **dûment identifiée la personne** dont il émane et qu'il soit **établi** et **conservé** dans des conditions de nature à en garantir **l'intégrité**.

Art. 1316-2. - Lorsque la loi n'a pas fixé d'autres principes, et à défaut de convention valable entre les parties, le juge règle les conflits de preuve littérale en déterminant par tous moyens le titre le plus vraisemblable quel qu'en soit le support.

Art. 1316-4 - La signature nécessaire à la perfection d'un acte juridique identifie celui qui l'appose. Elle manifeste le consentement des parties aux obligations qui découlent de cet acte. Quand elle est apposée par un officier public[49], elle confère l'authenticité à l'acte.
Lorsqu'elle est électronique, elle consiste en l'usage d'un **procédé fiable d'identification garantissant son lien avec l'acte auquel elle s'attache.**

49 Officier public ou ministériel : Personne titulaire d'un office conféré par l'Etat et nommé par décision d'un ministre. Les avoués près les cours d'appel, les huissiers de justice, les notaires, les avocats au Conseil d'Etat et à la Cour de cassation sont des officiers ministériels. Certains d'entre eux sont également des officiers publics, en raison de leur pouvoir d'authentifier des actes juridiques ou judiciaires et de procéder à l'exécution des décisions de justice (Exemple : notaires, huissiers de justice). (Source : Lexique "Les mots-clés de la Justice" accessible sur le site internet du ministère de la Justice www.justice.gouv.fr)

La deuxième partie du second alinéa de l'article 1316-4 introduit un mécanisme qui permet de spécifier les **conditions selon lesquelles un procédé de signature électronique sera non seulement considéré, mais de plus, présumé[50], fiable** : « La fiabilité de ce procédé est présumée, jusqu'à preuve du contraire, lorsque la signature est créée, l'identité du signataire assurée et l'intégrité de l'acte garantie, dans des conditions fixées par décret[51] en Conseil d'État».

L'admission en preuve de l'écrit électronique a progressivement ouvert la voie à la **dématérialisation de procédures dans de nombreux domaines de la sphère publique** : *domaines financier et fiscal,* santé *et sécurité sociale, JO lois et décrets, contrôle de légalité, marchés publics, transmissions des pièces comptables entre ordonnateurs et payeurs (collectivités territoriales), monde judiciaire au civil comme au pénal, transmission des actes d'état civil, ressources humaines.*

Pour connaître les textes réglementaires relatifs à la dématérialisation dans ces domaines, se reporter au chapitre zéro de la traduction en français des spécifications de MoReq2, 2008, accessible sur le site des archives de France, p. 45 *seq.* :
http://www.archivesdefrance.culture.gouv.fr/static/2085

Un enjeu essentiel : l'intégrité des données

Sur les questions juridiques, les résultats du groupe de travail du *Forum des droits sur l'internet* sur la conservation des documents électroniques, rassemblant informaticiens, archivistes, consultants, juristes a permis de poser les enjeux juridiques et notamment celui de la définition qu'on peut donner à la **notion d'intégrité**.

Voir le rapport accessible à l'adresse suivante :
http://www.foruminternet.org/telechargement/documents/reco-archivage-20051201.pdf

Voir également la note d'information des archives de France **DITN/RES/2006/002** du 30 mars 2006 accessible à l'adresse suivante :
http://www.archivesdefrance.culture.gouv.fr/static/875

Les conditions pour créer un environnement de confiance propice à la conservation électronique sont ainsi énoncées :
- le maintien d'une neutralité technologique et organisationnelle[52] ;
- le fait que la mise en place d'un processus de conservation ne doit pas modifier le statut juridique d'un document[53].

La question centrale est celle du maintien de l'intégrité du document dans le temps tel qu'il est exigé par l'article 1316-1 du Code civil qui pose que l'écrit sous forme électronique doit être établi « et conservé dans des conditions de nature à en garantir l'intégrité ».

50 La présomption de fiabilité entraîne le renversement de la charge de la preuve.
51 **Décret n°2001-272** du 30 mars 2001
52 L'archivage peut être mis en œuvre soit par un service d'archivage interne, soit en faisant appel à un tiers-archiveur.
53 La valeur juridique d'un document doit être conservé durant tout son cycle de vie y compris durant le processus d'archivage, mais ce n'est pas un processus d'archivage qui peut donner une valeur juridique à un document qui en était dépourvu à l'origine.

Comment interpréter cette notion d'intégrité sachant que pour des raisons liées à l'obsolescence des matériels et logiciels, il est impossible de maintenir durant de longues périodes l'intégrité technique[54] d'un document ?

Intégrité : le respect cumulé de trois critères

Dans la mesure où devant le juge, se posent les questions de recevabilité et de force probante des documents électroniques archivés, **les critères de l'intégrité** doivent impérativement être édictés afin de permettre de définir les conditions dans lesquelles un document conservé pourra avoir valeur probante.

C'est ainsi que cette **exigence d'intégrité** est assurée par le **respect cumulé des trois critères** que sont :

- la **lisibilité** du document,
- la **stabilité** du contenu informationnel
- la **traçabilité** des opérations sur le document :

> *« La **lisibilité** désigne la possibilité d'avoir accès, au moment de la restitution du document, à l'ensemble des informations qu'il comporte. Cette démarche est facilitée par les métadonnées associées au document[55]. »*

> *« La **stabilité** du contenu informationnel désigne la nécessité de pouvoir garantir que les informations véhiculées par le document restent les mêmes depuis l'origine et qu'aucune n'est omise ou rajoutée au cours du processus de conservation. Le contenu informationnel s'entend de l'ensemble des informations, quelle que soit leur nature ou leur origine, issues du document et notamment de sa mise en forme[56] .»*

> *« La **traçabilité** désigne la faculté de présenter et de vérifier l'ensemble des traitements, opérés sur le document lors du processus de conservation .».*

De même, l'objectif de restitution d'un document intègre sous-entend, du point de vue de la preuve, **la mise en place d'un processus de conservation**. Ainsi les acteurs peuvent fonder leur confiance dans le respect de bonnes pratiques devant se poursuivre tout au long de quatre étapes du processus de conservation que sont le transfert, l'enregistrement, la gestion et la restitution des documents, chacune de ces étapes précisant les opérations qui concourent à l'obtention de l'intégrité.

2. Les fondements juridiques de l'authenticité des données numériques

Cadre de confiance

L'article 1348 alinéa 2 du code civil, issu de la loi du 12 juillet 1980, indique : « Les règles ci-dessus [production d'une preuve écrite] reçoivent aussi exception lorsqu'une partie

54 Ne pas modifier sa chaîne de bits.

55 En fait, il s'agit de l'intelligibilité du document : un document parfaitement lisible peut ne pas être intelligible sans ses métadonnées (ainsi des données extraites d'une base de données, qui ne seraient pas explicitées par leur structuration, le modèle des données et les nomenclatures pour les codes utilisés).

56 À titre d'exemple, sont visés comme faisant partie du contenu informationnel d'un document : la taille des polices de caractères utilisées, les mises en évidence résultant du recours aux caractères gras, etc.

ou le dépositaire n'a pas conservé le titre original et présente une copie qui en est la **reproduction non seulement fidèle mais aussi durable**. Est réputée durable toute reproduction indélébile de l'original qui entraîne une modification irréversible du support.»

À l'époque, le texte vise la **réalisation de microfilms et de photocopies (papier).**

Les articles 1369-1 à 1369-11 du Code civil s'appliquent aux **contrats sous forme électronique** : échanges d'information, conclusion des contrats, envoi et remise d'un écrit électronique, exigences de forme.

L'article 1325 du Code civil précise les conditions dans lesquelles la **pluralité d'originaux** est satisfaite pour les contrats sous forme électronique.

Décret n°2011-144 du 2 février 2011 relatif à l'envoi d'une **lettre recommandée par courrier électronique pour la conclusion ou l'exécution d'un contra**

Décret n°2011-434 du 20 avril 2011 relatif à **l'horodatage des courriers expédiés ou reçus par voie électronique pour la conclusion ou l'exécution d'un contrat**

Ordonnance n°2005-1516 du 8 décembre 2005 relative aux **échanges électroniques entre les usagers et les autorités administratives et entre les autorités administratives** (ratifiée par l'article 138 I de la loi n° 2009-526 du 12 mai 2009).

D**écret n°2010-112** du 2 février 2010 pris pour l'application des articles 9, 10 et 12 de l'ordonnance n° 2005-1516 du 8 décembre 2005 relative aux échanges électroniques entre les usagers et les autorités administratives et entre les autorités administratives.

Authenticité et signature électronique

La **signature électronique** est le procédé retenu en droit (*art. 1316-4 du Code civil, alinéa 2)* pour garantir l'identité de l'auteur du document et l'intégrité d'un document numérique.

L'archivage sécurisé est une obligation découlant de l'adoption de la signature électronique.

Fonctionnement de la signature électronique

Il s'agit d'un procédé qui prend une empreinte d'une information (fichier, document) à un instant précis et y applique un algorithme de chiffrement à clé publique, c'est-à-dire dont la clé de déchiffrement figure sur un certificat appartenant nominalement à l'émetteur du document. Le déchiffrement permet ainsi de comparer l'empreinte du document envoyé avec celle du document initial et de constater d'éventuelles modifications.

Le portail de la sécurité informatique de l'Agence nationale de la Sécurité des Systèmes d'Information (ANSSI) propose une **présentation** claire du fonctionnement de la signature électronique.

Pour une présentation générale, on peut également se référer à une **présentation de Frédéric Pailler** (Centre national des études spatiales) sur le site du groupe PIN : **http://pin.association-aristote.fr/lib/exe/fetch.php/public/presentations/2009/pin20090113_cryptographie-pailler.pdf**

Cadre juridique de la signature électronique

Le décret n°2001-272 du 30 mars 2001, pris pour l'**application** de la loi du 13 mars 2000, détaille les modalités de mise en place de la signature électronique.

Le ***Référentiel général de sécurité*** fixe le cadre réglementaire en matière d'outils et de procédures d'authentification, de confidentialité, d'horodatage et de signature électronique : **http://www.ssi.gouv.fr/fr/reglementation-ssi/referentiel-general-de-securite/**

Obligation en matière d'archivage sécurisé

L'archivage sécurisé est une obligation découlant de la dématérialisation et de l'adoption de la signature électronique.

La signature électronique d'un document doit nécessairement entraîner le versement dans **un système d'archivage électronique. Ce système garantit une conservation sécurisée durant les délais de conservation requis,** avant élimination avec le visa réglementaire de l'administration des archives ou transfert pour archivage définitif dans les services publics d'archives.

Ainsi, l'obligation d'un archivage sécurisé est pris en compte réglementairement dans les domaines judiciaires et juridiques où les actes établis sur support électronique doivent être conservés dans des conditions garantissant lisibilité et intégrité :

- S'agissant des **notaires et des huissiers**, ils doivent enregistrer les actes authentiques établis sur support électronique dans un minutier central établi et contrôlé par leurs instances nationales (**décrets n°2005-972** et **n°2005-973** en date du 10 août 2005) ;
- S'agissant du **domaine pénal (police, gendarmerie, justice),** l'obligation d'archivage des actes de procédures signés électroniquement est inscrite en tant que telle à l'article A53-6 du code de procédure pénale.

Les fonctions de conservation, intégrité, intelligibilité, accessibilité et traçabilité durant la durée d'utilité courante et intermédiaire des documents sont requises, reprenant en cela la politique d'archivage dans le secteur public. Le format pérenne des documents est requis, de même que l'obligation de réplication des données sur un site distant, en conformité avec d'une part la norme ISO 14721 (norme OAIS de juin 2005) et la norme Afnor Z 42-013 (version de mars 2009) sur l'archivage électronique.

A ce sujet, consulter :
- **Arrêté du 21 juin 2011** relatif à la signature électronique ou numérique en matière pénale, modifiant le code de procédure pénale
– Note d'information **DGP/SIAF/2011/018** en date du 18 octobre 2011 : **http://www.archivesdefrance.culture.gouv.fr/static/5278**

L'obligation d'archivage sécurisé de documents numériques dépourvus de signature électronique

La décision d'utiliser ou non un procédé de signature électronique dépendra du contexte et des processus métier concernés.

Dans certains cas, **la signature électronique n'est pas requise, la sécurisation juridique étant obtenue par d'autres dispositifs organisationnels et techniques**. On citera ainsi dans le domaine des ressources humaines, le **décret n°2011-675** du 15 juin 2011 relatif au dossier individuel des agents publics et à sa gestion sur support électronique.

Toutefois l'absence de signature électronique n'exonère pas de l'obligation d'organiser un archivage sécurisé durant les délais requis conformément aux dispositions du code du patrimoine.

Conservation à long-terme et signature électronique

Une autre question très délicate concerne la conservation des documents signés par le procédé cryptographique. En effet, une migration de formats nécessaire sur le long terme

pour maintenir la lisibilité du document, invalide automatiquement le procédé de vérification d'une signature cryptographique.

A la demande des Archives de France, cette question a fait l'objet d'une étude par Jean-François Blanchette, professeur à l'Université de Californie à Los Angeles (UCLA).
Voir la note d'information **DITN/RES/2004/004** du 18 octobre 2004 :
http://www.archivesdefrance.culture.gouv.fr/static/1051

Concernant les documents signés électroniquement, il en découle plusieurs **exigences :**

- le créateur du document doit vérifier (ou faire vérifier) la validité de la signature avant que le délai du certificat utilisé ne soit expiré,

- le résultat de cette vérification doit être porté dans les métadonnées du document pendant toute la durée de sa conservation.

Ainsi, on fera porter la valeur du document d'une part sur les opérations de vérifications effectuées en amont et d'autre part sur la qualité du processus d'archivage effectué au sein du SAE.

Plus généralement, il est recommandé que, sous réserve de la possibilité de vérifier l'intégrité des documents conservés (voir ci dessus**), les opérations successives justifiées par la conservation (et notamment les migrations de formats) ne retirent pas au document son statut juridique.**

C'est cette formule qui a été retenue pour les décrets relatifs aux actes authentiques des notaires et des huissiers : « *Les opérations successives justifiées par sa conservation, notamment les migrations dont il peut faire l'objet, ne retirent pas à l'acte sa nature d'original.* »

3. Une contrainte juridique supplémentaire de l'environnement numérique : la protection des données personnelles

Les données à caractère personnel

Idée reçue : la loi « CNIL » règle le sort des données (avec le « droit à l'oubli ») et interdit toute conservation et archivage historique.

La mise en œuvre de l'articulation entre la législation CNIL et la législation archives doit se faire avant même la conception du système d'information

Les objectifs de la loi « CNIL »

La loi « CNIL » encadre la gestion des **données à caractère personnel** dans les applications de production (archives courantes et intermédiaires »).

L'objectif de cette loi est d'éviter la prolifération et la diffusion des données à caractère personnel. Elle encadre ainsi leur production et leur détention dans des traitements qui ne peuvent être créés que pour une finalité déterminée, explicite et légitime, pendant une durée donnée et proportionnée à la finalité poursuivie, et qui ne sont accessibles que par des personnes autorisées à les utiliser.

A l'issue de la durée de conservation déclarée à la CNIL et figurant dans les textes de création des traitements, les données doivent impérativement être supprimées de ces traitements.

Les possibilités d'une conservation et d'un archivage historique

Depuis la modification de 2004, la loi « CNIL » prévoit une possibilité de conservation pour une autre finalité que celle prévue dans le traitement d'origine.

L'article 6 de la loi précise : « Toutefois, un traitement ultérieur de données à des fins statistiques ou à des fins de recherche scientifique ou historique est considéré comme compatible avec les finalités initiales de la collecte des données ».

Parallèlement, le code du patrimoine (livre II sur les archives) fait strictement référence à cette possibilité offerte par la loi « CNIL ». L'article L 212-3 du code indique :

« *Lorsque les archives publiques comportant des données à caractère personnel collectées dans le cadre de traitements régis par la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés,* ***ces données font l'objet****, à l'expiration de la durée prévue au 5° de l'article 6 de ladite loi,* ***d'une sélection pour déterminer les données destinées à être conservées et celles, dépourvues d'utilité administrative ou d'intérêt scientifique, statistique ou historique, destinées à être éliminées****.*

Les catégories de données destinées à l'élimination ainsi que les conditions de cette élimination sont fixées par accord entre l'autorité qui a produit ou reçu ces données et l'administration des archives. »

Modalités d'application

On peut avec profit s'inspirer de la **délibération n° 2005-213 du 11 octobre 2005** portant adoption d'une recommandation concernant *les modalités d'archivage électronique, dans le secteur privé,* de données à caractère personnel, 11 octobre 2005.

Il est recommandé :

- de respecter le principe du **« droit à l'oubli »** ;
- de **protéger** les données archivées notamment contre la diffusion ou l'accès non autorisés ainsi que contre toute autre forme de traitement illicite ;
- d'**éviter la « dilution » des données archivées dans le système informatique de l'entreprise** : la CNIL recommande que l'accès aux archives intermédiaires soit limité à un service spécifique (par exemple un service du contentieux) et qu'il soit procédé, a minima, à un isolement des données archivées au moyen d'une séparation logique (gestion des droits d'accès et des habilitations).
- de développer, dans les entreprises, des procédures formalisées et qu'une information puisse être fournie sur ces règles, en cas de demande exprimée de leur part, aux individus faisant l'objet des traitements archivés.

La recommandation a vocation à s'appliquer aux archives dites courantes, intermédiaires et définitives.

En revanche, s'agissant des archives publiques, les données personnelles destinées à une conservation définitive par les services publics d'archives et dont la complétude et l'intégrité doivent être préservées, **n'ont pas vocation à faire l'objet d'anonymisation** comme indiqué dans cette recommandation destinée au secteur privé.

Conséquences

La justification de cet archivage est que sa finalité n'est plus celle qui a prévalu lors de la mise en œuvre du traitement (finalité patrimoniale). Il s'ensuit que les producteurs de cette information n'y ont plus accès, une fois l'archivage effectué, tant que les délais de libre communicabilité ne sont pas expirés.

Ceci implique que dans le cadre des déclarations à faire à la CNIL, la détermination de la durée de conservation des données doit être identique à la durée d'utilité administrative prévue par ailleurs dans le cadre du code du patrimoine, avec notamment, en cas de conservation au titre des archives définitives, mention de l'article L 212-3**. Il convient par conséquent systématiquement d'élaborer cette déclaration en étroite collaboration avec la mission ou le service d'archives concerné.**

Les données de santé à caractère personnel

Conformément à l'article L 1111-8 du code de la santé publique, précisé par le décret 2006-6 du 4 janvier 2006 relatif à l'hébergement de données de santé à caractère personnel sur support informatique, toute personne physique ou morale hébergeant des données de santé à caractère personnel recueillies ou produites à l'occasion d'activités de prévention, de diagnostic ou de soins pour le compte d'un établissement, d'un professionnel de santé ou de la personne concernée, doit être agréée par décision du ministre en charge de la santé qui se prononce après avis de la CNIL et d'un comité d'agrément (organe consultatif crée par le décret 2006-6 précité).

Le consentement de la personne dont les données sont hébergées est requis.

L'Agence des systèmes d'information partagés de santé (ASIP Santé), missionnée à cet effet par le ministère en charge de la santé assure le secrétariat et la pré-instruction des dossiers de demande d'agrément pour le compte des membres du comité d'agrément.

Tout candidat à l'agrément doit compléter un dossier de demande d'agrément composé notamment de six formulaires standards couvrant l'ensemble du recueil d'informations exigé par le décret (volet juridique et éthique, volet technique et sécurité, volet économique et financier).

L'agrément est délivré pour trois ans, pour un type de prestation d'hébergement de données de santé à caractère personnel clairement définie dans le contrat d'hébergement dont le contenu est fixé par l'article R 1111-13 du code de la santé publique. Ce contrat doit en outre, clairement décrire les responsabilités de chaque partie au contrat (i.e. l'hébergeur et son client).

Chaque année, l'hébergeur agréé doit adresser un rapport d'auto-évaluation qui a pour objet d'informer des changements intervenus au cours de l'année écoulée. Si le rapport d'auto évaluation remet en cause le périmètre initial, l'organisme devra déposer une nouvelle demande d'agrément.

Le cadre juridique de l'archivage

Le cadre juridique de l'archivage repose sur le **code du Patrimoine (partie législative et partie réglementaire)**

Plan de la fiche :

- Définition et justification des archives
- Le réseau des archives de l'État
- Exercice du contrôle scientifique et technique de l'État sur les archives publiques
- Modalités de collaboration entre les services producteurs et les services d'archives. Versements dans les services publics d'archives
- Accès aux archives et régime de communication des archives
- Sanctions pénales

1. Définition et justification des archives

Les archives sont l'ensemble des documents, quels que soient leur date, leur lieu de conservation, leur forme et leur support, produits ou reçus par toute personne physique ou morale et par tout service ou organisme public ou privé dans l'exercice de leur activité (Article L211-1).

La conservation des archives est organisée dans l'intérêt public tant pour les besoins de la gestion et de la justification des droits des personnes physiques ou morales, publiques ou privées, que pour la documentation historique de la recherche (Article L211-2).

Les archives publiques sont les documents qui procèdent de l'activité, dans le cadre de leur mission de service public, de l'État, des collectivités territoriales, des établissements publics et des autres personnes morales de droit public ou des personnes de droit privé chargées d'une telle mission […], ainsi que les minutes et répertoires des officiers publics ou ministériels (Article L 211-4). Les archives publiques sont imprescriptibles et nul ne peut détenir sans droit ni titre des archives publiques (Article L212-1).

Les archives privées sont l'ensemble des documents définis à l'article L. 211-1 qui n'entrent pas dans le champ d'application de l'article L. 211-4 (Article L 211-5).

Sont considérés comme **archives courantes** les documents qui sont d'utilisation habituelle pour l'activité des services, établissements et organismes qui les ont produits ou reçus.

La conservation des archives courantes incombe, sous le contrôle de la personne chargée du contrôle scientifique et technique de l'État sur les archives, aux services, établissements et organismes qui les ont produites ou reçues. Ceux-ci peuvent, après en avoir fait la déclaration à l'administration des archives, déposer tout ou partie de ces documents auprès de personnes physiques ou morales agréées à cet effet par ladite administration (Articles L 212-4 et R212-10)) et dans les conditions prévues aux articles R. 212-19 à R. 212-31.

Sont considérés comme **archives intermédiaires** les documents qui :

1° Ont cessé d'être considérés comme archives courantes ;

2° Ne peuvent encore, en raison de leur intérêt administratif, faire l'objet de sélection et d'élimination conformément aux dispositions de l'article R. 212-14 (Article R212-11).

Sont considérés comme **archives définitives** les documents qui ont subi les sélections et éliminations définies aux articles R. 212-13 et R. 212-14 et qui sont à conserver sans limitation de durée. La conservation des archives définitives est assurée dans les dépôts d'archives relevant du service interministériel des archives de France de la direction générale des patrimoines ou placés sous le contrôle de la personne chargée du contrôle scientifique et technique de l'État sur les archives (Article R212-12).

2. Le réseau des archives de l'État

Le service interministériel des archives de France (SIAF) de la direction générale des patrimoines au ministère de la culture exerce toutes les attributions confiées à l'administration des archives par le présent code, à l'exception de celles qui concernent les archives des ministères des affaires étrangères et de la défense, ainsi que des services et établissements qui en dépendent ou y sont rattachés. Le SIAF inclut les missions des Archives de France implantées dans les ministères pour organiser la collecte des archives destinées aux Archives nationales (Article R 212-1).

Les Archives nationales sont constituées par l'ensemble des services à compétence nationale rattachés au service interministériel des archives de France de la direction générale des patrimoines : Archives nationales à Paris, Pierrefitte-sur-Seine et Fontainebleau ; Archives nationales du Monde du travail à Roubaix ; Archives nationales d'Outre-Mer à Aix-en-Provence (Article R 212-8).

Deux ministères disposent de leurs propres directions des archives indépendants. Il s'agit des services d'archives des affaires étrangères qui assurent la gestion des archives provenant de l'administration centrale, des postes diplomatiques et consulaires ainsi que des établissements placés sous l'autorité du ministre des affaires étrangères, ainsi que d**es services d'archives relevant du ministère de la défense** qui assurent la gestion des archives provenant de l'ensemble des forces, services, établissements et organismes des armées, ainsi que des services et établissements dont le rattachement aux services d'archives de ce ministère est prévu par décret (Articles R 212-5 et 6).

Depuis le 12 avril 2012 (décret n° 2012-479), a été institué un **délégué interministériel aux archives de France** placé auprès du Premier ministre.

3. Exercice du contrôle scientifique et technique de l'État sur les archives publiques

Le service interministériel des archives de France de la direction générale des patrimoines assure le contrôle scientifique et technique sur les archives publiques des services de l'Etat, sur celles appartenant aux collectivités territoriales, à leurs établissements publics et à leurs groupements, ainsi que sur celles qui leur sont confiées en application des articles L. 212-6 à L. 212-14. Ces attributions s'exercent sur les archives courantes, intermédiaires et définitives (Article R212-2).

Le contrôle scientifique et technique exercé par le service interministériel des archives de France de la direction générale des patrimoines **porte sur les conditions** de gestion, de collecte, de sélection et d'élimination ainsi que sur le traitement, le classement, la conservation et la communication des archives (Article R212-3).

Ce contrôle scientifique et technique **est exercé sur pièces ou sur place par** :

1° Le service interministériel des archives de France de la direction générale des

patrimoines dans son champ de compétences ;

2° Les membres du service de l'inspection des patrimoines pour l'ensemble des services et organismes ;

3° Les chefs des missions des archives et les autres personnels scientifiques et de documentation mis à disposition des services centraux de l'État ou des établissements publics nationaux, dans leur ressort ;

4° Les directeurs des services départementaux d'archives et agents de l'État mis à disposition des collectivités territoriales dans la limite de leurs circonscriptions géographiques, sauf en ce qui concerne les services d'archives dont ils ont la direction (Article R 212-4).

Par ailleurs les services d'archives des **affaires étrangères** et **de la Défense** assurent le contrôle scientifique et technique de l'État sur les administrations dont elles collectent les archives (Articles R 212-5 et R 212-6)

4. Modalités de collaboration entre les services producteurs et les services d'archives. Versements dans les services publics d'archives

Principes généraux

A l'expiration de leur période d'utilisation courante, les archives publiques autres que celles mentionnées à l'article L. 212-3 font l'objet d'une sélection pour séparer les documents à conserver des documents dépourvus d'utilité administrative ou d'intérêt historique ou scientifique, destinés à l'élimination

La liste des documents ou catégories de documents destinés à l'élimination ainsi que les conditions de leur élimination sont fixées par accord entre l'autorité qui les a produits ou reçus et l'administration des archives (Article L212-2)

I-Les archives publiques qui, à l'issue de la sélection prévue aux articles L. 212-2 et L. 212-3, sont destinées à être conservées **sont versées dans un service public d'archives** dans des conditions fixées par décret en Conseil d'État (Article L 212-4)

Dérogation au principe de versement des archives définitives dans les services publics d'archives

Ce décret détermine les cas où, par dérogation aux dispositions qui précèdent, l'administration des archives laisse le soin de la conservation des documents d'archives produits ou reçus par certaines administrations ou certains organismes aux services compétents de ces administrations ou organismes lorsqu'ils présentent des conditions satisfaisantes de conservation, de sécurité, de communication et d'accès des documents. Il fixe les conditions de la coopération entre l'administration des archives et ces administrations ou organismes (Articles L 212-4 et R 212-12).

Sont définies par accord entre le service, l'établissement ou l'organisme intéressé et le service interministériel des archives de France de la direction générale des patrimoines :

1° La durée d'utilisation comme archives courantes ;

2° La durée de conservation comme archives intermédiaires ;

3° La destination définitive à l'issue de la période de conservation comme archives intermédiaires, à savoir :

a) L'élimination immédiate ou à terme, intégrale ou partielle, avec ou sans sélection ;

b) Le versement, à titre d'archives définitives, dans un dépôt d'archives relevant du service interministériel des archives de France de la direction générale des patrimoines ou placé sous le contrôle de la personne chargée du contrôle scientifique et technique de l'Etat sur les archives ;

c) La conservation par le service, l'établissement ou l'organisme intéressé, dans les conditions prévues à l'article R. 212-12 (Article R212-13)

La sélection des documents incombe à la personne chargée du contrôle scientifique et technique de l'État sur les archives.

La personne chargée du contrôle scientifique et technique de l'État sur les archives établit les listes des documents dont elle propose l'élimination et les soumet au visa de l'administration d'origine. Toute élimination est interdite sans ce visa. A l'inverse, lorsque les services, établissements et organismes désirent éliminer les documents qu'ils jugent inutiles, ils en soumettent la liste au visa de la personne chargée du contrôle scientifique et technique de l'État sur les archives. Toute élimination est interdite sans ce visa. **Dans tous les cas, les documents à éliminer sont détruits sous le contrôle technique du service interministériel des archives de France de la direction générale des patrimoines** (Article R212-14).

Lors du **transfert de documents** dans un service d'archives, il est établi un bordereau descriptif par les soins, selon le cas, du service qui effectue le versement.

Le versement d'un document établi sur support numérique est accompagné de l'ensemble des informations le concernant dès son établissement et nécessaires à son exploitation, telles que les données permettant de l'identifier, de déterminer ses propriétés et d'en assurer la traçabilité (Article R212-16).

Les services d'archives publics communiquent aux services, établissements et organismes qui leur ont versé les documents les instruments de recherche qui s'y rapportent (Article R212-17).

Les documents conservés dans les dépôts relevant du service interministériel des archives de France de la direction générale des patrimoines **restent à la disposition exclusive du service, établissement ou organisme dont ils proviennent** dans la mesure où ils ne sont pas communicables aux termes de la loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal, et des articles L. 213-1, L. 213-2 et L. 213-3 (Article R212-18)

Documents d'archives émanant du Président de la République et des membres du Gouvernement

Le versement des documents d'archives publiques émanant du Président de la République, du Premier ministre et des autres membres du Gouvernement peut être assorti de la signature entre la partie versante et l'administration des archives d'un protocole relatif aux conditions de traitement, de conservation, de valorisation ou de communication du fonds versé, pendant la durée des délais prévus à l'article L. 213-2. Les stipulations de ce protocole peuvent également s'appliquer aux documents d'archives publiques émanant des collaborateurs personnels de l'autorité signataire.

Pour l'application de l'article L. 213-3, l'accord de la partie versante requis pour autoriser la consultation ou l'ouverture anticipée du fonds est donné par le signataire du protocole.

Le protocole cesse de plein droit d'avoir effet en cas de décès du signataire et, en tout état de cause, à la date d'expiration des délais prévus à l'article L. 213-2 (Article L213-4).

Suppression d'une administration

Lorsqu'il est mis fin à l'existence d'un ministère, service, établissement ou organisme détenteur d'archives publiques, celles-ci sont, à défaut d'affectation déterminée par l'acte de suppression, versées à un service public d'archives (Article L212-5).

5. Accès aux archives et régime de communication des archives (Art. L212-1 à L 212-5)

Principes généraux

Les archives publiques sont, sous réserve des dispositions de l'article L. 213-2, communicables de plein droit.

L'accès à ces archives s'exerce dans les conditions définies pour les documents administratifs à l'article 4 de la loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal (Article L213-1)

Mais des exceptions sont prévues pour un certain nombre de types de documents (secret des délibérations du Gouvernement , conduite de la politique extérieure, sûreté de l'État, la sécurité des personnes ou à la protection de la vie privée , secret de la vie privée, secret médical, sécurité publique, secret industriel et commercial…) : voir Articles L 213-2.

Des dérogations sont prévues pour ces exceptions ainsi que la possibilité d'ouverture anticipée de fonds ou parties de fonds : voir Article L 212-3.

Protection du secret de la défense nationale

Arrêté du 30 novembre 2011 portant approbation de l'instruction générale interministérielle n° 1300 sur la protection du secret de la défense nationale.

Pour certains types de fonds, la communication est exclue

Ne peuvent être consultées les archives publiques dont la communication est susceptible d'entraîner la diffusion d'informations permettant de concevoir, fabriquer, utiliser ou localiser des armes nucléaires, biologiques, chimiques ou toutes autres armes ayant des effets directs ou indirects de destruction d'un niveau analogue (Article L 213-2)

6. Sanctions pénales

Toute infraction aux dispositions de l'article L. 211-3 est passible des peines prévues aux articles 226-13 et 226-31 du code pénal (Article L214-1)

Sans préjudice de l'application des articles 322-2, 432-15, 432-16 et 433-4 du code pénal, le fait, pour une personne détentrice d'archives publiques en raison de ses fonctions, de détourner ou soustraire tout ou partie de ces archives ou de les détruire sans accord préalable de l'administration des archives est puni d'une peine de trois ans d'emprisonnement et de 45 000 € d'amende.

Est puni des mêmes peines le fait, pour une personne détentrice d'archives publiques en raison de ses fonctions, d'avoir laissé détruire, détourner ou soustraire tout ou partie de ces archives sans accord préalable de l'administration des archives.

Lorsque les faits prévus aux premier et deuxième alinéas sont commis par négligence dans les conditions et selon les distinctions prévues à l'article 121-3 du code pénal, les peines sont d'un an d'emprisonnement et de 15 000 € d'amende.

La tentative des délits prévus au premier alinéa et le fait, pour la personne visée au deuxième alinéa, d'avoir laissé commettre une telle tentative sont punis des mêmes peines (Article L214-3)

Les personnes physiques coupables des infractions prévues par l'article L. 214-3 encourent également les peines complémentaires suivantes :

1° L'interdiction des droits civils, civiques et de famille, suivant les modalités prévues par

l'article 131-26 du code pénal ;

2° L'interdiction, suivant les modalités prévues par l'article 131-27 du même code, d'exercer une fonction publique ou d'exercer l'activité professionnelle ou sociale dans l'exercice ou à l'occasion de l'exercice de laquelle l'infraction a été commise ;

3° La confiscation, suivant les modalités prévues par l'article 131-21 du même code, des sommes ou objets irrégulièrement reçus par l'auteur de l'infraction, à l'exception des objets susceptibles de restitution (Article L214-4)

Le fait, pour une personne détentrice sans droit ni titre d'archives publiques, de ne pas les restituer sans délai à l'autorité compétente qui lui en fait la demande est puni d'une peine d'un an d'emprisonnement et de 15 000 € d'amende (Article L214-5)

Les personnes morales déclarées responsables pénalement des infractions prévues à l'article L. 214-3 encourent les peines mentionnées aux 2°, 8° et 9° de l'article 131-39 du code pénal.

L'interdiction mentionnée au 2° du même article 131-39 porte sur l'activité dans l'exercice ou à l'occasion de l'exercice de laquelle l'infraction a été commise (Article L214-9)

La politique d'archivage dans le secteur public

Cette fiche précise la répartition des rôles des services producteurs, services informatiques et services d'archives en ce qui concerne la responsabilité du contenu de l'information.

Plan de la fiche :

- Le référentiel sur l'archivage numérique sécurisé
- Structure du référentiel : politique d'archivage et répartition des responsabilités
  Répartition des rôles et des responsabilités

Référence :**http://www.ssi.gouv.fr/fr/bonnes-pratiques/outils-methodologiques/archivage-electronique-securise.html**

struction **DITN/RES/2006/005** du 13 septembre 2006. Publication de l'étude commanditée par la direction centrale de la sécurité des systèmes d'information (DCSSI) sur l'archivage électronique sécurisé dans le secteur public :
http://www.archivesdefrance.culture.gouv.fr/static/884

1. Le référentiel sur l'archivage numérique sécurisé

En 2006 à la demande de l'ancienne direction centrale pour la sécurité des systèmes d'information (DCSSI), aujourd'hui l'ANSSI, a été publié un référentiel sur l'archivage électronique sécurisé comportant un certain nombre d'éléments :

Les analyses préalables
La synthèse des enjeux juridiques
L'état de l'art juridique, technique, organisationnel et des offres
Les documents introductifs :
- La plaquette ;
- Le mémento.

Les documents d'aide à l'élaboration du référentiel
- La politique et les pratiques d'archivage ;
- Le cahier des charges pour un système d'archivage électronique ;
- La grille d'audit.

2. Structure du référentiel : politique d'archivage et répartition des responsabilités

Ce document de 43 pages définit ainsi les exigences minimales, en termes juridiques, fonctionnels, opérationnels, techniques et de sécurité, qu'une autorité d'archivage doit respecter afin que l'archivage électronique mis en place puisse être regardé comme fiable. Cette PA type repose sur des contraintes « standard » à mettre en place : identification/authentification de l'origine de l'archive, intégrité, intelligibilité / lisibilité, durée de conservation, traçabilité des différentes opérations, disponibilité et de accessibilité.

Il est composé des parties suivantes :

- Organisation et responsabilité de chacun des acteurs en présence et les transferts de responsabilité intervenants tout au long du cycle de vie des archives.
- Gestion des risques de sécurité des systèmes d'information :
  - identification des besoins de sécurité notamment en terme de disponibilité, d'intégrité et de confidentialité,
  - appréciation des risques et traitement des risques,
  - sécurité et cycle de vie du SAE.
- Principes de mises en œuvre :
  - aspects humains (critères de sélection des personnels du SAE, habilitations pour l'accès au SAE, définition des rôles de confiance, cloisonnement des postes sensibles)
  - planification de la continuité des activités
  - exploitation (documentation des procédures, maintenance des systèmes à tracer, lutte contre les virus et les codes malveillants..)
  - aspects physiques et environnement
- Principes techniques
  - identification et l'authentification des utilisateurs
  - contrôles d'accès logiques
  - intégrité des Archives versées
  - journalisation d'un certain nombre d'opérations
  - horodatage des opérations.

3. Répartition des rôles et des responsabilités

On distingue en effet :

L'autorité juridique (responsable du contenu des données) jusqu'à la fin de la DUA est le producteur des données. A la fin de la DUA, le rôle de l'autorité juridique est endossé par les services d'archives définitives conformément au *Code du patrimoine*.

L'autorité d'archivage (responsable de la conservation et par là-même **du service d'archivage électronique**) pendant la DUA. Suivant les contextes, les durées de conservation, la nature des données, la stratégie d'archivage mise en œuvre, elle peut être exercée par différents acteurs : le producteur des données, un service public d'archives s'il est responsable de la conservation pendant la DUA. A l'expiration de la DUA, les services d'archives publics deviennent l'autorité d'archivage quel que soit le cas de figure pour la conservation pendant la DUA.

L'opérateur d'archivage est celui qui est responsable du **système d'archivage électronique**. Il s'agit de la DSI ou du tiers archiveur.

Glossaire

	Voir aussi	
Application métier	*gestion électronique de documents*	Logiciel permettant de créer et/ou de gérer des documents et données dans le cadre des activités d'un ou des service(s). Font partie des applications métier les bases de données, les systèmes de gestion électronique de documents spécifiques à un métier, les outils de travail collaboratif.
Archivage électronique	*autorité juridique, autorité d'archivage*	Ensemble des actions, outils et méthodes mis en œuvre pour conserver à moyen et à long terme des informations numériques dans le but de les rendre accessibles et exploitables. L'archivage électronique implique d'identifier précisément les responsabilités des différents acteurs (autorité juridique, autorité d'archivage, etc.).
Authenticité	*intégrité, signature électronique*	Qualité d'un document ou d'une donnée dont l'origine, la réalité et l'auteur sont certifiés et incontestables. Dans le monde numérique, la signature électronique est le procédé permettant de garantir cette qualité.
Autorité d'archivage	*opérateur d'archivage, politique d'archivage*	Personne morale qui a la responsabilité fonctionnelle de la conservation des documents ou données dans le système d'archivage électronique.
Autorité juridique	*authenticité, intégrité*	Personne morale qui a la responsabilité du contenu (authenticité, intégrité) des documents ou données.
Copie de sécurité	*sauvegarde*	Double des données d'un système qui permet de le reconstituer en cas d'incident.
Cycle de vie	*règles d'archivage*	Étapes que suit un document ou une donnée de sa création jusqu'à la mise en œuvre de son sort final.
Dématérialisation		Opération visant à ce que les documents gérés aujourd'hui sous forme papier ou analogique le soient demain sous forme électronique, soit par le biais d'une opération de numérisation, soit par la révision des processus de production et de gestion de l'information.
Donnée		Représentation formalisée de l'information, adaptée à l'interprétation, au traitement et à la communication. La donnée est donc un conteneur porteur d'une information ou d'un fragment d'information.
Empreinte	*signature électronique, intégrité*	Ensemble de bits caractéristique d'un document numérique, obtenu par une fonction de hachage. Toute modification du document numérique entraîne une empreinte différente.
Export (fonction)	*import*	Opération qui consiste à extraire des documents ou données d'un système.
Format		Ensemble des caractéristiques logiques d'organisation de l'information, conditionnant à la fois le mode d'accès aux données, leur diffusion et leur conservation.
Gestion électronique de documents (GED)	*application métier*	Outil informatique permettant d'organiser et de gérer des documents ou données électroniques au sein d'un organisme.

Horodatage		Technique permettant d'associer à un document une date certaine en référence à un système de temps donné et reconnu ; cette date peut être la date à laquelle un document est émis ou la date à laquelle un document fait l'objet d'une opération de gestion.
Intégrité	*Authenticité, empreinte*	Qualité d'un document ou d'une donnée qui n'a pas été altéré. Dans le monde numérique, un document ou une donnée est réputé intègre si son empreinte à un temps t+1 est identique à l'empreinte prise à un temps t.
Import	*Export*	Opération qui consiste à intégrer dans un système des documents/données provenant d'un autre système.
Journal des évènements		Enregistrement des actions humaines ou techniques intervenant dans le système d'information.
Métadonnées	*Donnée*	Ensemble des données caractérisant d'un point de vue technique, structurel et contextuel un document ou une donnée et fournissant l'information indispensable à sa gestion, son accessibilité et son exploitation..
Migration	*Pérennisation*	Opération consistant à changer le format, le support ou le système utilisé pour représenter, stocker ou gérer des documents ou données en vue de leur pérennisation.
Opérateur d'archivage	*archivage électronique*	Personne morale qui a la responsabilité du bon fonctionnement du système d'archivage électronique.
Plan de classement		Représentation conceptuelle hiérarchisée des fonctions et/ou activités d'un organisme afin d'ordonner sa production documentaire, d'en permettre l'accès et de préserver le contexte de création de l'information.
Politique de sécurité		Document de référence qui définit les règles à appliquer et à respecter dans un organisme en matière de sécurité des systèmes d'information.
Politique d'archivage	*système d'archivage électronique*	Document de référence qui définit les contraintes juridiques, fonctionnelles, opérationnelles et techniques à respecter par les différents acteurs afin que l'archivage électronique mis en place puisse être considéré comme fiable.
Règles d'archivage	*cycle de vie*	Ensemble des consignes qui permettent de définir et mettre en œuvre le cycle de vie des documents ou données.
Sauvegarde	*copie de sécurité*	Opération technique destinée à assurer, par une copie de sécurité, la continuité de l'exploitation d'un système informatique en cas d'incident.
SEDA		Modélisation des différentes transactions qui peuvent avoir lieu entre des acteurs dans le cadre de l'archivage de documents ou données, s'accompagnant d'une modélisation de la description des données qui seront échangées lors de ces transactions. Le standard propose des schémas XML pour la mise en œuvre de ces transactions fixant la forme des messages échangés ainsi que la forme de la description des données échangées.

Signature électronique	*Authenticité Empreinte*	Mécanisme qui permet l'identification de l'auteur d'un document électronique et la garantie de son intégrité par analogie avec la signature manuscrite d'un document papier. Donnée ajoutée à une donnée ou à un ensemble de données et garantissant l'origine de cette ou ces données, c'est-à-dire certifiant l'authenticité de l'émetteur.
Stockage	*support de stockage*	Enregistrement sur un support quelconque et sans gestion du cycle de vie de documents ou données numériques.
Support de stockage	*stockage*	Ensemble d'objets physiques permettant l'enregistrement des documents/données numériques.
Suppression		Opération visant à détruire, tant logiquement que physiquement, des données ou documents dont la conservation ne se justifie plus.
Système d'archivage électronique (SAE)		Ensemble d'infrastructures matérielles et logicielles permettant de conserver et de restituer des documents ou données électroniques sur le long terme en garantissant leur intégrité et leur lisibilité.
Tiers archiveur		Personne physique ou morale qui se charge pour le compte de tiers d'assurer et de garantir la conservation et l'intégrité de documents électroniques
Traçabilité	*intégrité*	Capacité à produire la liste des traitements et interventions opérés sur un document ou données.
Transfert	*transfert manuel, transfert automatique*	Transmission effective des paquets d'archives du système du service producteur vers le système d'archivage électronique.
Transfert manuel	*transfert, transfert automatique*	Transmission des paquets d'archives qui implique une intervention humaine au début ou pendant le processus
Transfert automatique	*transfert, transfert manuel*	Transmission des paquets d'archives qui s'effectue de machine à machine, sans intervention humaine