

# **Module 7C - Section 3 : Intervenir sur les documents d'archives sur support numérique afin d'en garantir la préservation**

Édouard Vasseur @AIAF - PIAF

VF 02/12/2024



# Table des matières

<b>Objectifs</b>	<b>5</b>
<b>Introduction</b>	<b>7</b>
<b>1. Création des documents d'archives sur support numérique</b>	<b>8</b>
1.1. Cas du document physique/analogique inclu dans une opération de numérisation à la chaîne.....	8
1.2. Cas du document physique/analogique numérisé à la volée ou nativement numérique.....	9
1.3. Les bonnes pratiques... ..	9
3.1. Les logiciels.....	9
3.2. Les règles de nommage .....	9
3.3. Les modalités de stockage de fichiers .....	9
3.4. Points d'attention .....	10
<b>2. Intégration des documents d'archives sur support numérique dans un système d'archivage</b>	<b>11</b>
2.1. Les normes à suivre .....	11
1.1. La norme ISO 20652:2006 – Space data and information transfer systems – Producer-archive interface – Methodology abstract standard (autrement appelée PAIMAS) .....	11
1.2. La norme ISO 20104:2015 – Space data and information transfer systems – Producer-Archive Interface Specification (PAIS).....	11
2.2. Les phases de l'intégration .....	12
2.1. Identifier des documents d'archives sur support numérique à intégrer dans le système d'archivage.....	12
2.2. Extraire les fichiers et les métadonnées.....	12
2.3. Retraiter les fichiers et les métadonnées extraits.....	12
2.4. Contrôler la qualité du paquet d'archives .....	13
2.5. Transférer le SIP .....	13
2.6. Effectuer un contrôle dans le système d'archivage cible.....	13
2.7. Constituer des paquets à archiver et compléter les métadonnées.....	13
2.8. Clôturer la procédure.....	13
2.3. Fiche pratique : utiliser la norme PAIMAS .....	13
2.3.1. Identification des documents et des informations à intégrer .....	13
2.3.2. Extraction des fichiers et des métadonnées.....	14
2.3.3. Retraitement des fichiers, de leurs métadonnées et empaquetage sous forme de SIP.....	15
2.3.4. Transfert du SIP.....	16
2.3.5. Contrôle par le système d'archivage cible à réception .....	17
2.3.6. Constitution des paquets à archiver (AIP).....	18
2.3.7. Acquiescement du transfert .....	19
2.4. Le cas des documents numériques mal ou pas identifiés .....	20

<b>3. Description des documents d'archives sur support numérique et des actions générées par les opérations de préservation</b>	<b>22</b>
3.1. Les métadonnées descriptives .....	23
3.2. Les métadonnées techniques .....	24
2.1. Les métadonnées génériques .....	24
2.2. Les métadonnées spécifiques .....	24
3.3. Traces générées lors des opérations de préservation .....	24
3.4. Le standard PREMIS (PREservation Metadata Implementation Strategies).....	25
4.1. Description de PREMIS.....	25
4.2. Le standard PREMIS ne constitue pas une solution toute faite .....	26
<b>4. Stockage des documents d'archives sur support numérique</b>	<b>27</b>
4.1. Les spécificités du stockage des documents d'archives sur support numérique.....	27
Le taux de perte .....	27
L'intégrité .....	27
Les durées de conservation .....	27
La consultation des documents .....	27
La taille des documents d'archives .....	27
Les documents contenant des informations confidentielles .....	27
4.2. Recommandations pour le stockage des documents d'archives sur support numérique.....	28
<b>5. Actions spécifiques de préservation des documents d'archives sur support numérique</b>	<b>30</b>
5.1. La planification des actions de préservation .....	30
La planification doit s'appuyer sur les éléments suivants : .....	30
Sur cette base, la planification doit permettre de : .....	30
5.2. Les actions de préservation .....	31
5.2.1. Le remontage de l'environnement de production des documents .....	31
5.2.2. L'émulation de l'environnement de création des documents d'archives .....	31
5.2.3. L'identification d'un nouveau logiciel .....	32
5.2.4. La migration ou conversion du format de fichiers .....	32
5.2.5. La réparation des fichiers .....	33
<b>6. L'accès aux documents d'archives sur support numérique</b>	<b>35</b>
6.1. La numérisation facilite l'accès aux documents d'archives sur support analogique ou physique.....	35
6.2. Les documents nativement numériques peuvent poser des problèmes d'accès. ....	35
2.1. Souvent, ils ne sont pas communicables et/ou diffusables.....	35
2.2. L'accès peut nécessiter des outils spécifiques .....	36
2.3. Les attentes des utilisateurs sont très variées.....	36
6.3. Prérequis pour donner accès aux documents d'archives sur support numérique.....	36
3.1. Analyser les besoins des différentes communautés d'utilisateurs .....	37

3.2. Identifier les environnements matériels et logiciels .....	38
3.3. Définir le format et la version de fichiers dans lequel le document d'archives doit être fourni .....	38
3.4. Autoriser ou non les utilisateurs à accéder et réutiliser le contenu des documents.....	38
<b>7. Mise en œuvre du sort final</b>	<b>39</b>
7.1. L'élimination .....	39
1.1. À noter.....	39
7.2. Transfert à un autre service d'archives .....	40
2.1. Le transfert peut être réalisé sous forme d'une opération .....	40
2.2. Le périmètre des éléments à transférer doit faire l'objet d'un échange entre services d'archives .....	40
2.3. Les données de référence doivent être partagées par les deux services d'archives .....	40
2.4. Les systèmes d'identification pérenne doivent faire l'objet d'un échange entre services d'archives .....	40
2.5. La suppression des fichiers n'intervient qu'après l'acquittement du transfert.....	41
<b>8. Sécurité et audit</b>	<b>42</b>
8.1. La sécurité .....	42
8.1.1. Les spécificités du document numérique d'archives en terme de sécurité .....	42
8.1.2. Les spécificités du traitement de la sécurité .....	43
8.1.3. Des solutions spécifiques à mettre en place .....	43
8.1.4. Les spécificités du plan de prévention des risques .....	43
8.1.5. Quelques points à ne pas négliger... ..	44
8.2. L'audit .....	44

# Objectifs

---



## Description du module :

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

## Le but du module est de :

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

## L'apprenant doit être en mesure de :

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

## Positionnement :

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

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

Point sur le vocabulaire employé :

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

Les notions abordées dans ce module peuvent être complétées par :

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

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

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

# Introduction

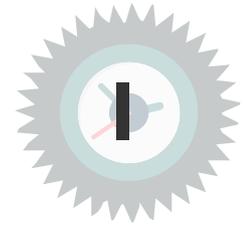
---



Comme pour les documents d'archives sur support physique et analogique, les actions de préservation numérique peuvent intervenir à tout moment : au moment de la création des documents eux-mêmes ; au moment des opérations d'archivage, que ce soit dans une perspective de gestion de l'archivage/records management ou dans une perspective d'archivage définitif ; pendant la durée de conservation par le service d'archives ; au moment des demandes d'accès ou de réutilisation.

Si les processus ne diffèrent guère de ceux applicables aux documents d'archives sur support physique ou analogique, il n'en demeure pas moins que ceux applicables aux documents d'archives sur support numérique présentent certaines particularités que le présent chapitre a pour vocation d'explorer.

# 1. Création des documents d'archives sur support numérique



## Introduction

L'idéal, naturellement, consiste à prendre en compte dès la création des documents d'archives les besoins qui seront ceux de leur préservation et de leur accès sur le moyen comme le long terme. Il est souvent d'ailleurs dans l'intérêt des services producteurs que les documents qu'ils créent ou reçoivent sur support numérique soient bien formés, complets, corrects, valides, et donc dignes de confiance et utilisables jusqu'au terme de leur conservation, pour leurs besoins propres comme pour ceux de la société et de la recherche.

Dans la création des documents d'archives sur support numérique, il faut cependant distinguer trois cas :

1. le cas où le document existe déjà sur un support physique ou analogique, et fait l'objet d'une *numérisation* dans le cadre d'une *chaîne de numérisation*, qu'il s'agisse d'un document sur support papier, ou d'un document sur support audiovisuel ;
2. le cas où le document existe déjà sur un support physique ou analogique, et fait l'objet d'une numérisation à la volée ;
3. le cas où le document est directement créé numériquement.

## 1.1. Cas du document physique/analogique inclu dans une opération de numérisation à la chaîne

Dans le premier cas, il est souvent facile de créer des fichiers de bonne qualité dès lors que la chaîne de numérisation est bien paramétrée, qu'elle soit opérée en interne ou via une prestation externalisée.



**Fondamental**

La production des fichiers doit permettre de respecter les points suivants :

- définition du format cible en fonction de critères qui garantissent à la fois une bonne qualité de la numérisation, un stockage optimisé – certains formats, par exemple le TIFF, sont volumineux et nécessitent un espace de stockage important – et un respect des normes et standards ;
- génération systématique d'une description fonctionnelle et technique – notamment des informations permettant de contrôler l'*intégrité* du fichier dans le temps (*empreinte*) ou décrivant les options de numérisation choisies (algorithme de compression, profondeur de bits, résolution de balayage pour les documents sonores et audiovisuels), voire intégration automatique de *métadonnées* facilitant la gestion du cycle de vie ;
- mise en place de procédures manuelles ou automatiques de contrôle qualité, par exemple en mettant en œuvre des opérations de *validation de formats*.

## 1.2. Cas du document physique/analogique numérisé à la volée ou nativement numérique

Dans le deuxième et le troisième cas, en revanche, même s'il est envisageable en théorie de contractualiser avec le *service producteur* voire de refuser d'archiver un document numérique non conforme, il est en pratique souvent délicat voire peu réaliste d'imposer aux services producteurs la création et la réception de leurs documents dans un format imposé par le *service d'archives*. La création sera alors de plus ou moins bonne qualité, et, dans bien des cas, les fichiers seront créés dans les formats et avec les outils les plus usuels du marché (traitement de texte, tableur, présentation, impressions PDF, messages électroniques, images fixes, plans en deux dimensions, sites internet, publications électroniques, etc.).



**Complément**

Des recommandations existent cependant en matière de création de formats, par exemple dans le Référentiel général d'interopérabilité adopté par les institutions françaises (<https://www.numerique.gouv.fr/publications/interoperabilite/>) ou les Lignes directrices sur les formats de fichiers à utiliser pour transférer des ressources documentaires de Bibliothèque et Archives Canada (<https://bibliotheque-archives.canada.ca/fra/services/gouvernement-canada/information-disposition/lignes-directrices-information/pages/lignes-directrices-formats-fichier-ressources-documentaires.aspx#l>). Mais il ne s'agit que de recommandations, et rien ne dit que celles-ci sont suivies.

## 1.3. Les bonnes pratiques...

Quelques bonnes pratiques peuvent cependant être recommandées sans imposer trop de contraintes aux services producteurs et à leurs services informatiques.

### 1.3.1 Les logiciels

Il est sain de préconiser des logiciels permettant de créer les documents d'archives et formats de fichiers pris en charge : recommander le choix de logiciels permettant de générer des formats pouvant être interprétés par une variété importante de logiciels différents.

### 1.3.2. Les règles de nommage

Il est sain de préconiser des *règles de nommage* des fichiers et des répertoires : recommander des noms de fichiers courts, compatibles avec les contraintes imposées par les systèmes d'exploitation – notamment Windows –, évitant les caractères spéciaux, intégrant des éléments de datation métier – les dates enregistrées automatiquement dans les fichiers eux-mêmes sont rarement significatives – et des informations sur l'état du document (provisoire/définitif, numéro de version).

### 1.3.3. Les modalités de stockage de fichiers

Il est sain de préconiser des modalités de stockage des fichiers : recommander, pour les documents dans leur version définitive, un stockage assurant sauvegarde et confidentialité – notamment si le document contient des informations non communicables ou protégées. Le stockage et/ou la sauvegarde peuvent être internalisés ou externalisés. Dans bien des cas, le service d'archives dépend cependant de la politique de stockage retenue par le service informatique.

### 3.4. Points d'attention



---

Dans le cas où la création des documents est réalisée sur une plateforme externalisée sous la forme d'une prestation de services (ex. Office 365), il est fortement recommandé de vérifier à la fois les conditions de conservation (stockage et sauvegarde ; contrôles d'intégrité), les conditions d'accès (niveau de disponibilité) et les conditions de réversibilité (modalité de fermeture du service et de restitution des documents d'archives et des métadonnées de description, de gestion, techniques et informations de traçabilité) proposées par celle-ci .



---

Pour des questions de confidentialité et de sécurité, les services producteurs peuvent recourir à des solutions de chiffrement des fichiers, ce qui garantit, en cas de vol ou de piratage informatique, que seuls les détenteurs des clés de déchiffrement seront en mesure d'accéder au contenu de ceux-ci. Si de telles solutions sont employées pour sécuriser le transfert ou le stockage des fichiers, il convient de s'assurer que les clés permettant de déchiffrer le contenu des documents sont bien conservées et gérées. Dans le cas contraire, le contenu des fichiers sera immanquablement perdu.

Pour plus d'informations sur la bonne manière de créer et organiser les documents d'archives, voir le module 5 et 7B.

## 2. Intégration des documents d'archives sur support numérique dans un système d'archivage

---



### 1. Introduction

La phase d'intégration des documents d'archives sur support numérique dans un système d'archivage, qu'il s'agisse d'un système d'archivage électronique gérant des documents d'archives numériques avant application du sort final (*electronic records management system*), ou un système d'archivage définitif, est souvent décisive pour la réalisation d'opérations qui garantiront la préservation à moyen ou à long terme de ces documents. Bien souvent, c'est toute une chaîne de traitement qui est mise en œuvre pour garantir la bonne intégration de ces documents dans le système d'archivage cible.

### 2.1. Les normes à suivre

La phase d'intégration dans un système d'archivage cible peut être longue et complexe, dans la mesure où elle enchaîne de nombreuses étapes, peut faire intervenir plusieurs interlocuteurs, internes comme externes, et nécessiter le recours à plusieurs outils techniques.

Deux initiatives, constituant des guides méthodologiques de mise en œuvre de la norme *OAIS*, ont tenté de normaliser cette phase de relation entre les producteurs d'archives et les services d'archives : il s'agit des normes Paimas et Pais.

#### 2.1.1. La norme ISO 20652:2006 – Space data and information transfer systems – Producer-archive interface – Methodology abstract standard (autrement appelée PAIMAS)

Elle fournit une description normalisée des interactions entre un service producteur et un service d'archives, en décrivant de manière détaillée les différentes phases, ainsi que le résultat attendu à l'issue de chacune d'entre elles. Cette norme n'est cependant pas disponible en français.

#### 2.1.2. La norme ISO 20104:2015 – Space data and information transfer systems – Producer-Archive Interface Specification (PAIS)

Elle met à disposition une méthode normalisée pour définir les documents d'archives sur support numérique à transférer au système d'archivage cible, pour emballer les fichiers à transférer et pour contrôler et valider le paquet transféré. Cette norme n'est pas non plus disponible en français.

## 2.2. Les phases de l'intégration

La phase d'intégration se décompose en plusieurs étapes :

### Identifier des documents d'archives sur support numérique à intégrer dans le système d'archivage.

Cette étape ne présente aucune originalité, puisqu'il s'agit du résultat du travail classique de l'archiviste : analyse du processus administratif et des productions documentaires associées, identification des documents entrants et sortants, évaluation, sélection et définition des critères de tri. Des outils spécifiques – notamment des logiciels permettant d'explorer des espaces de stockage numériques (ex. : logiciel Archifiltre conçu par les ministères sociaux en France) – pourront être utiles pour ce faire.

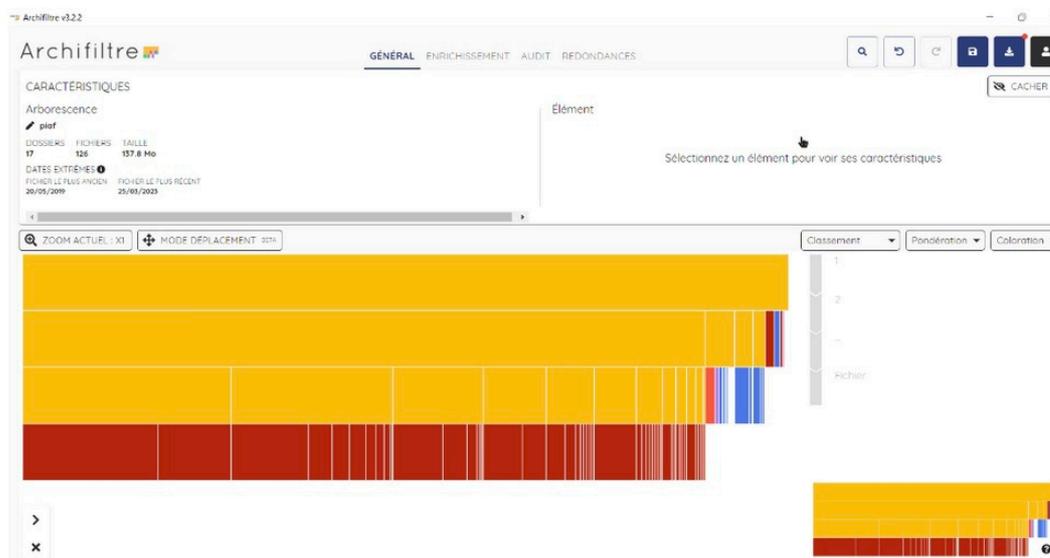


Fig.1 : Visualisation d'une arborescence de dossiers via Archifiltre

Il faut donc procéder à l'identification de toutes les informations indispensables pour préserver, rendre intelligibles et exploiter ces documents sur le long terme. Si certains aspects de cette étape sont communs à tous les documents quel que soit leur support – questions juridiques notamment (communicabilité, droits de propriété intellectuelle et de propriété industrielle, données à caractère personnel), mais aussi rassemblement ou création de documents complémentaires qui permettent de contextualiser ces documents (par exemple tout ce qui a trait à la conception et à l'exploitation de documents statistiques) –, d'autres sont spécifiques aux documents d'archives sur support numérique.

C'est tout particulièrement le cas pour les documents dont la création ou la gestion reposent sur des systèmes de gestion de bases de données comme des gestions électroniques de courriers (GEC), des gestions électroniques de documents (GED) ou des applications métier (structure de la base, codage des informations – voir section 5 de ce module).

### Extraire les fichiers et les métadonnées

Il faut procéder à l'extraction des fichiers et des métadonnées correspondant aux documents de leur système et/ou de leur support de gestion, en fonction des critères de sélection et de tri retenus.

### Retraiter les fichiers et les métadonnées extraits

Il faut procéder au retraitement des fichiers et des métadonnées extraits pour constituer le bordereau de prise en charge – avec toutes les informations nécessaires (métadonnées de description, de gestion et métadonnées techniques) – afin de formater les documents et leur description pour qu'elles soient intégrables dans le système d'archivage utilisé par le service d'archives (voir section 4 de ce module).

## **Contrôler la qualité du paquet d'archives**

Il faut procéder au contrôle de la qualité du paquet d'archives (le *SIP* au sens de l'*OAIS*) ainsi généré.

## **Transférer le SIP**

Il faut procéder au transfert du SIP dans le service d'archives gestionnaire du système d'archivage cible.

## **Effectuer un contrôle dans le système d'archivage cible**

Il faut procéder au contrôle par le système d'archivage cible de l'état sanitaire des fichiers (antivirus), de la conformité du SIP aux prescriptions du système – vérification de la complétude du paquet, vérification de l'intégrité des fichiers, vérification du bordereau pour s'assurer que toutes les métadonnées obligatoires sont bien renseignées, etc. –, correction éventuelle des anomalies identifiées et enrichissement du paquet de toutes les informations nécessaires à la préservation et à l'accès dans le temps des documents ainsi pris en charge – notamment attribution d'identifiants uniques. Cette étape peut s'accompagner d'une phase de caractérisation des fichiers (identification et validation du format des fichiers).

## **Constituer des paquets à archiver et compléter les métadonnées**

Il faut procéder à la constitution des paquets à archiver (*AIP*), à l'écriture des fichiers correspondants et de leur documentation associée sur les supports de stockage utilisés dans le système d'archivage cible en fonction de la politique de stockage et enregistrement des métadonnées (description, gestion, techniques) dans les outils de gestion des documents du système d'archivage (qu'il s'agisse d'une application métier, ou d'un simple tableur).

## **Clôturer la procédure**

Pour finir, il faut procéder à l'acquittement du transfert par le service d'archives, de la documentation du transfert et clôturer du processus.

## **2.3. Fiche pratique : utiliser la norme PAIMAS**

### **2.3.1. Identification des documents et des informations à intégrer**

La norme PAIMAS schématise l'enchaînement des opérations (Fig.2).

La première phase consiste à identifier les documents à intégrer dans le système d'archivage et les informations nécessaires pour garantir leur préservation et leur accessibilité dans le temps :

- cette phase de négociations peut être plus ou moins longue et complexe, notamment si les informations permettant de contextualiser les documents d'archives n'ont jamais été mises par écrit ;
- la documentation nécessaire à la préservation et à l'accès durable aux documents d'archives sur support numérique n'est pas toujours disponible. La difficulté ou l'impossibilité de rassembler ou de constituer celle-ci a posteriori peuvent constituer des freins à la poursuite du processus d'intégration ;
- la faisabilité technique et la faisabilité humaine de l'opération doivent être soigneusement évaluées. Il peut s'avérer en effet coûteux techniquement et humainement d'extraire et de traiter des documents ne présentant qu'une faible valeur probante, une faible valeur informationnelle ou une faible valeur patrimoniale ou de recherche ;
- il convient de documenter soigneusement tous les constats effectués

Schéma d'enchaînement des opérations de préparation des transferts selon la norme PAIMAS

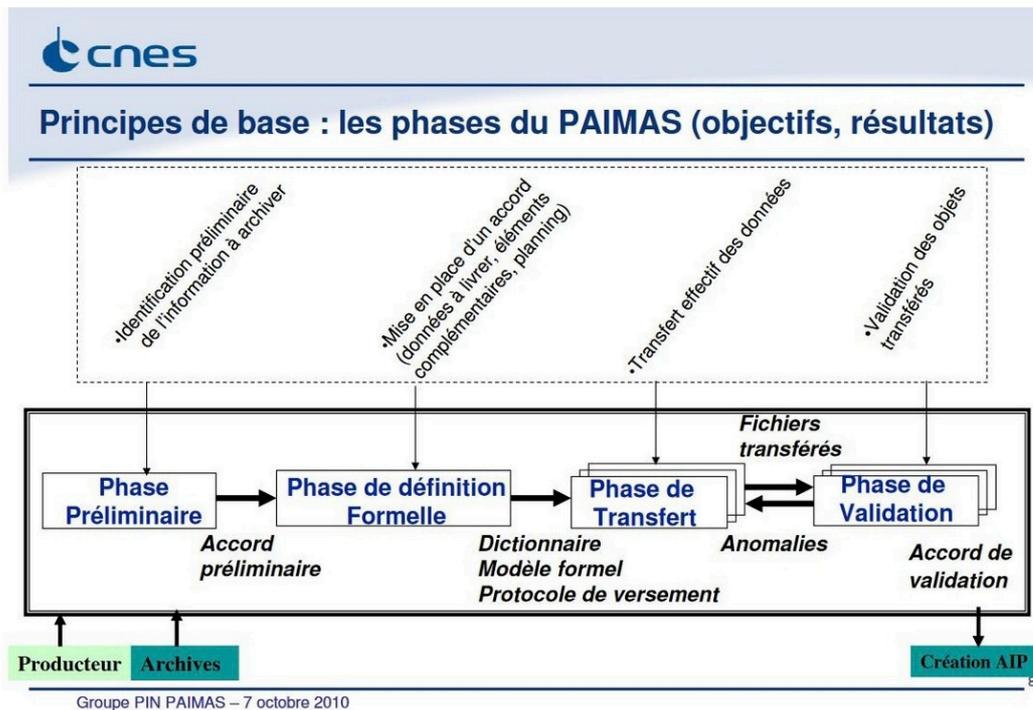


Fig.2 : **Enchaînement des opérations de préparation des transferts selon la norme PAIMAS** (source : présentation effectuée par Danièle Boucon lors de la séance du groupe PIN du 7 octobre 2010, disponible à l'adresse <https://web.archive.org/web/20211022084118/http://pin.association-aristote.fr/lib/exe/fetch.php/public/presentations/2010/pin20101007-pres04-paimaspais.pdf>)

### 2.3.2. Extraction des fichiers et des métadonnées

La deuxième phase, comme nous l'avons vu dans la partie 2.2. ci-dessus, est celle de l'extraction des fichiers et des métadonnées :

- l'extraction du système de production des documents peut donner lieu à la création de fichiers qui n'existaient pas en tant que tels dans celui-ci. Par exemple, le rassemblement de toutes les informations ayant trait à l'instruction d'une affaire, dispersées sous forme de données dans le système et rassemblées dans un nouveau fichier (ex. impression au format PDF) au moment de l'export ;
- les documents à transférer peuvent être le résultat de l'export de données présentes dans plusieurs systèmes (ex. un référentiel d'identités et une application métier) ;
- le choix de recourir à des images disque (très pratiqué en criminalistique/forensics et dans le cas de la collecte d'archives privées dans certaines bibliothèques universitaires) présente le risque de prendre en compte l'ensemble des fichiers présents sur le support de stockage source, y compris des fichiers inutiles, dangereux ou tombant sous le coup de condamnations pénales (ex. fichiers pornographiques).

### Il faut distinguer la criminalistique (ou science forensique) et de la criminologie



La criminalistique (ou science forensique) et la criminologie sont deux composantes différentes, mais pour autant indissociables l'une de l'autre dans une enquête. La criminalistique représente la recherche et l'exploitation des indices en vue de la manifestation de la vérité. Elle regroupe l'ensemble des principes scientifiques et les méthodes techniques appliqués à l'investigation criminelle, pour

participer à la preuve de l'existence d'un crime et aider la justice à déterminer l'identité de l'auteur et son mode opératoire [Margot 1995]. Cette science fait appel à de multiples domaines tels que la chimie, la biologie ou encore la physique.

Source : Site de l'Université de Cergy, <https://cyforensic.cyu.fr/actualites/notre-actualite/le-saviez-vous-criminalistique-et-criminologie-deux-sciences-distinctes><sup>1</sup>



Schéma d'enchaînement des opérations de préparation des transferts selon la norme PAIMAS

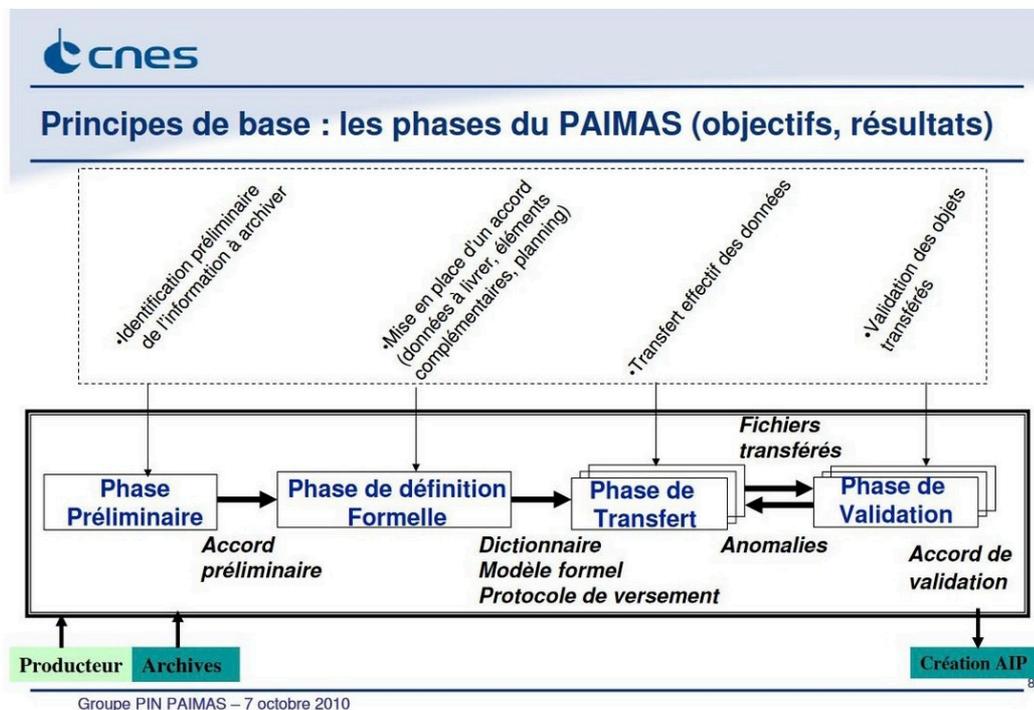


Fig.2 : **Enchaînement des opérations de préparation des transferts selon la norme PAIMAS** (source : présentation effectuée par Danièle Boucon lors de la séance du groupe PIN du 7 octobre 2010, disponible à l'adresse <https://web.archive.org/web/20211022084118/http://pin.association-aristote.fr/lib/exe/fetch.php/public/presentations/2010/pin20101007-pres04-paimaspais.pdf>)

### 2.3.3. Retraitement des fichiers, de leurs métadonnées et empaquetage sous forme de SIP

Pour ce qui est du retraitement des fichiers et de leurs métadonnées et de l'empaquetage sous forme de SIP :

- il est important à cette étape d'identifier les fichiers chiffrés et ceux qui correspondent à des formats conteneurs (type ZIP), avant même de constituer le SIP. Leur traitement dépendra des choix opérés par le service d'archives (rejet, déchiffrement, extraction des contenus du fichier conteneur);
- la manière d'empaqueter les fichiers et les métadonnées est souvent propre à chaque système d'archivage cible – avec sa documentation propre –, même si des standards d'échange existent (cf. section 4 de ce module.). Des outils ont été développés pour opérer cet empaquetage de manière plus ou moins automatisée (ex. BagIt, Docuteam Packer, OCTAVE, ReSIP, SIP Creator, etc.)

<sup>1</sup> <https://www.google.com/url?q=https://cyforensic.cyu.fr/actualites/notre-actualite/le-saviez-vous-criminalistique-et-criminologie-deux-sciences-distinctes&sa=D&source=docs&ust=1730748104300310&usg=AOvVaw1wb9tuJDumy6Avu3ATdOg>

Schéma d'enchaînement des opérations de préparation des transferts selon la norme PAIMAS

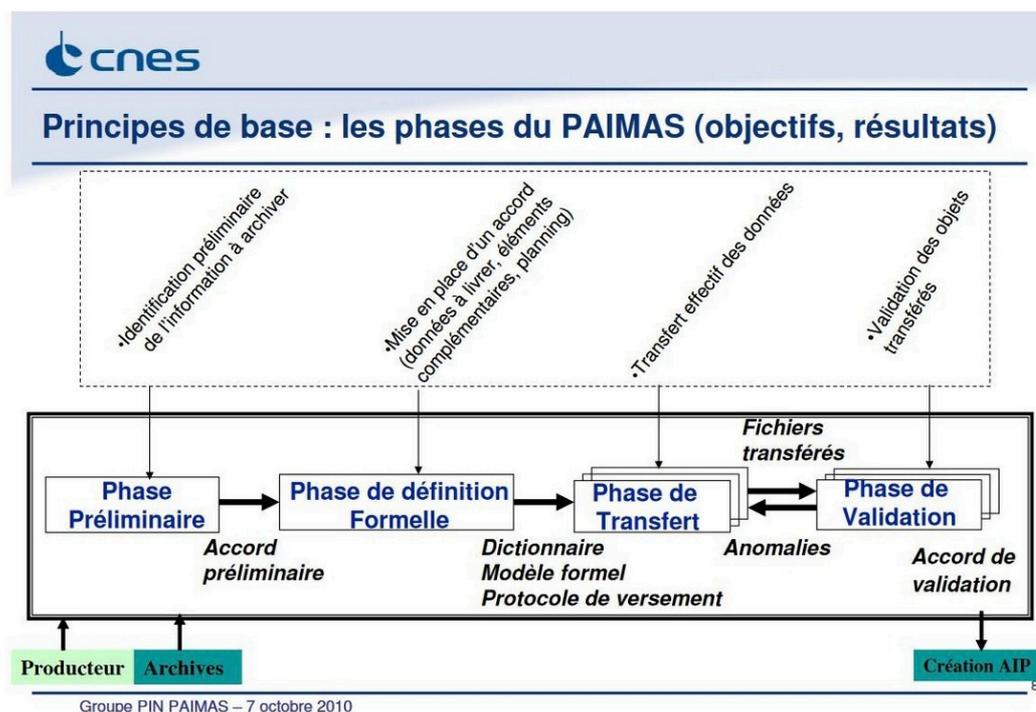


Fig.2 : **Enchaînement des opérations de préparation des transferts selon la norme PAIMAS** (source : présentation effectuée par Danièle Boucon lors de la séance du groupe PIN du 7 octobre 2010, disponible à l'adresse <https://web.archive.org/web/20211022084118/http://pin.association-aristote.fr/lib/exe/fetch.php/public/presentations/2010/pin20101007-pres04-paimaspais.pdf>)

### 2.3.4. Transfert du SIP

Pour le transfert du SIP :

- le contrôle de l'intégrité binaire du transfert est plus que fortement conseillé. Il peut être nécessaire de réaliser un calcul d'empreinte des fichiers avant de procéder au transfert, si celui-ci n'a pas été fait au moment de l'extraction des fichiers de l'environnement d'origine ;
- les modalités de transfert du SIP vers le système d'archivage cible dépendent de plusieurs éléments : la taille du SIP (en nombre d'octets), le partage ou non par le service producteur et le service d'archives de la même infrastructure technique, les conditions de confidentialité et de sécurité nécessaires au transfert ;
- deux options techniques de transfert sont possibles : transfert physique en utilisant des supports amovibles ou transfert par réseau – messagerie électronique, recours à une plateforme de partage de fichiers accessible par les deux parties (ex. ressources partagées, serveur FTP, cloud public) ; le transfert peut se faire en utilisant des méthodes de chiffrement, notamment si la sécurité des informations est importante. Dans ce cas, le service d'archives doit disposer des moyens de procéder au déchiffrement des fichiers et s'assurer que des procédures existent pour que des copies déchiffrées des documents ne soient pas conservées sur des supports de stockage annexes ;
- le transfert peut se faire en mode « push » (le service producteur initie l'opération) ou en mode « pull » (le service d'archives initie l'opération).

Schéma d'enchaînement des opérations de préparation des transferts selon la norme PAIMAS

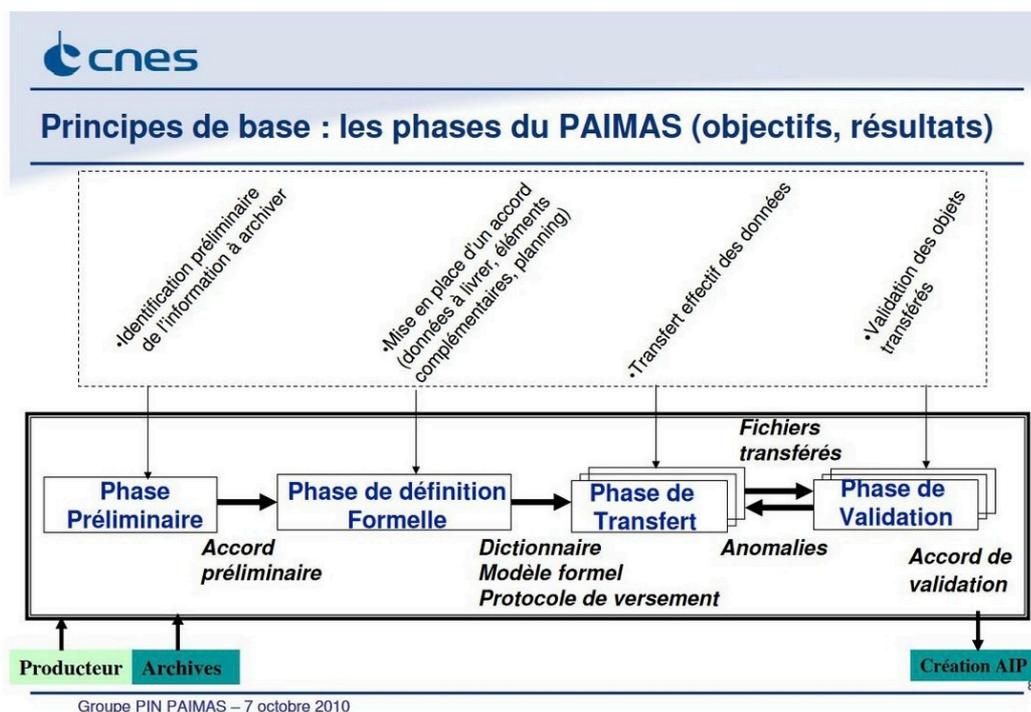


Fig.2 : **Enchaînement des opérations de préparation des transferts selon la norme PAIMAS** (source : présentation effectuée par Danièle Boucon lors de la séance du groupe PIN du 7 octobre 2010, disponible à l'adresse <https://web.archive.org/web/20211022084118/http://pin.association-aristote.fr/lib/exe/fetch.php/public/presentations/2010/pin20101007-pres04-paimaspais.pdf>)

### 2.3.5. Contrôle par le système d'archivage cible à réception

Pour l'opération de contrôle par le système d'archivage cible à réception :

- le service d'archives doit disposer d'un espace de stockage suffisant pour assurer la réception des SIP, leur contrôle et leur traitement, avant l'intégration dans le système d'archivage cible. Une station isolée du réseau ou un espace de stockage partagé mais sécurisé (ex. ressources partagées) peuvent suffire. Dans le cas où les opérations de contrôle prennent un temps important, le service d'archives doit s'assurer que soient faites des sauvegardes régulières et sécurisées des fichiers reçus ;
- le contrôle par un antivirus peut se révéler complexe, dans la mesure où les logiciels utilisés pour réaliser cette opération se fondent sur une base de connaissance actualisée, qui vise à détecter les menaces d'aujourd'hui, et non nécessairement des menaces d'hier dont la criticité n'est plus avérée. Les antivirus ont par ailleurs des difficultés à analyser les fichiers de trop grande taille et, a contrario, peuvent estimer que des virus sont présents alors que ce n'est pas le cas – ce qui est fréquent pour les fichiers correspondant à des tableurs utilisant des macros. Il ne peut donc être garanti à 100 % qu'un virus présent dans les fichiers transférés sera détecté au moment de cette phase. En cas de doute, il est possible de mettre le SIP en quarantaine ;
- le contrôle des fichiers reçus (nombre de fichiers reçus, contrôle d'intégrité par recalcul d'empreinte, etc.) peut entraîner la formalisation par le service d'archives d'un niveau de service offert, en fonction de la plus ou moins grande conformité du transfert opéré aux standards pris en charge par le service d'archives. Ainsi, un service d'archives peut offrir un niveau de service optimal s'il estime maîtriser totalement les formats de fichiers reçus, alors qu'il ne peut offrir qu'un niveau de service dégradé s'il estime que les fichiers reçus ne sont pas enregistrés dans un format de fichiers qu'il maîtrise.

Schéma d'enchaînement des opérations de préparation des transferts selon la norme PAIMAS

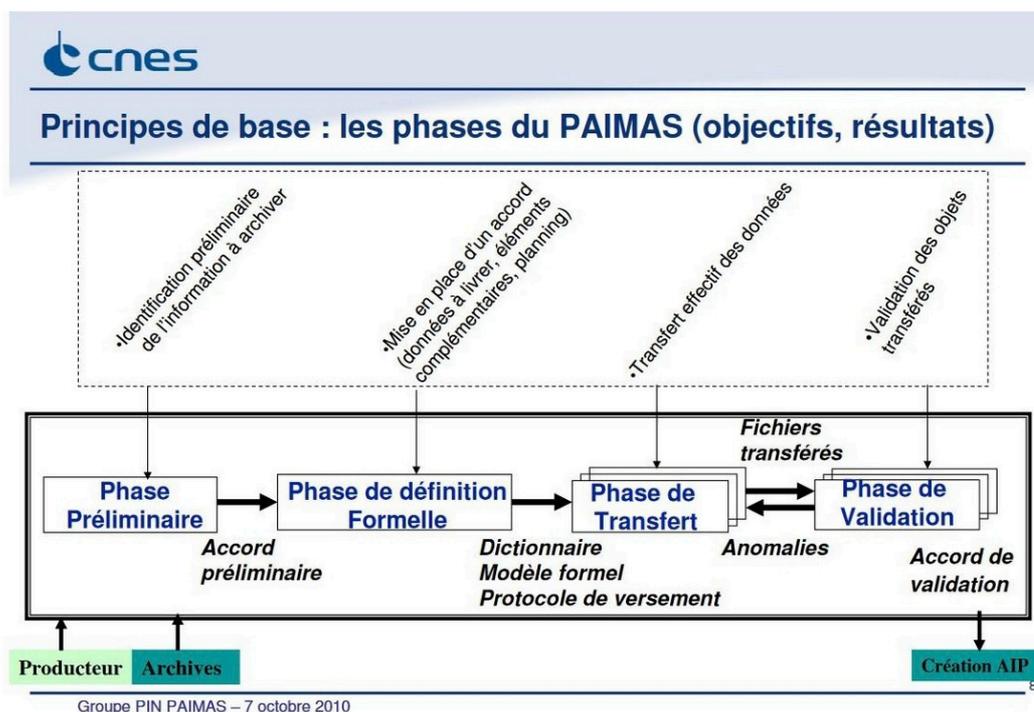


Fig.2 : **Enchaînement des opérations de préparation des transferts selon la norme PAIMAS** (source : présentation effectuée par Danièle Boucon lors de la séance du groupe PIN du 7 octobre 2010, disponible à l'adresse <https://web.archive.org/web/20211022084118/http://pin.association-aristote.fr/lib/exe/fetch.php/public/presentations/2010/pin20101007-pres04-paimaspais.pdf>)

### 2.3.6. Constitution des paquets à archiver (AIP)

Un AIP contient plusieurs fichiers de données empaquetés avec toutes les informations nécessaires pour permettre la préservation à long terme et l'accès aux fonds d'archives. Il faut procéder, après constitution, à l'écriture des fichiers correspondants et de leur documentation associée et à l'enregistrement des métadonnées. En cas d'utilisation d'outils de type tableur pour procéder à l'enregistrement des métadonnées, il est fortement recommandé de procéder à leur sauvegarde régulière.

Schéma d'enchaînement des opérations de préparation des transferts selon la norme PAIMAS

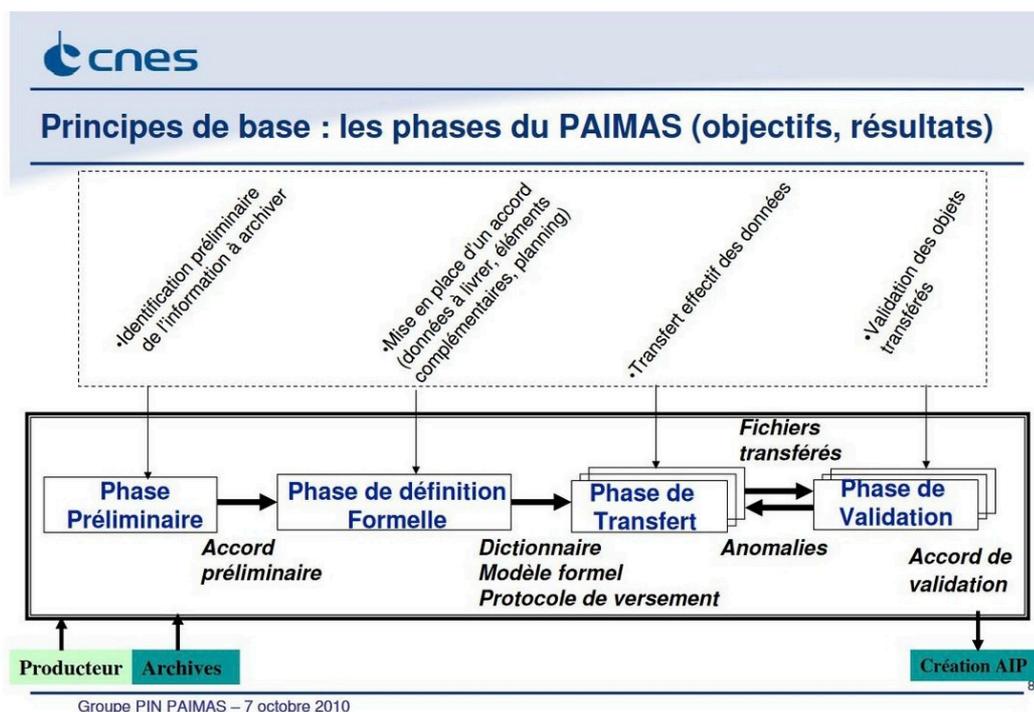


Fig.2 : **Enchaînement des opérations de préparation des transferts selon la norme PAIMAS** (source : présentation effectuée par Danièle Boucon lors de la séance du groupe PIN du 7 octobre 2010, disponible à l'adresse <https://web.archive.org/web/20211022084118/http://pin.association-aristote.fr/lib/exe/fetch.php/public/presentations/2010/pin20101007-pres04-paimaspais.pdf>)

### 2.3.7. Acquittement du transfert

Dans la mesure où les étapes de contrôle du transfert peuvent être longues, il peut être utile de transmettre un acquittement temporaire au service producteur.

Schéma d'enchaînement des opérations de préparation des transferts selon la norme PAIMAS

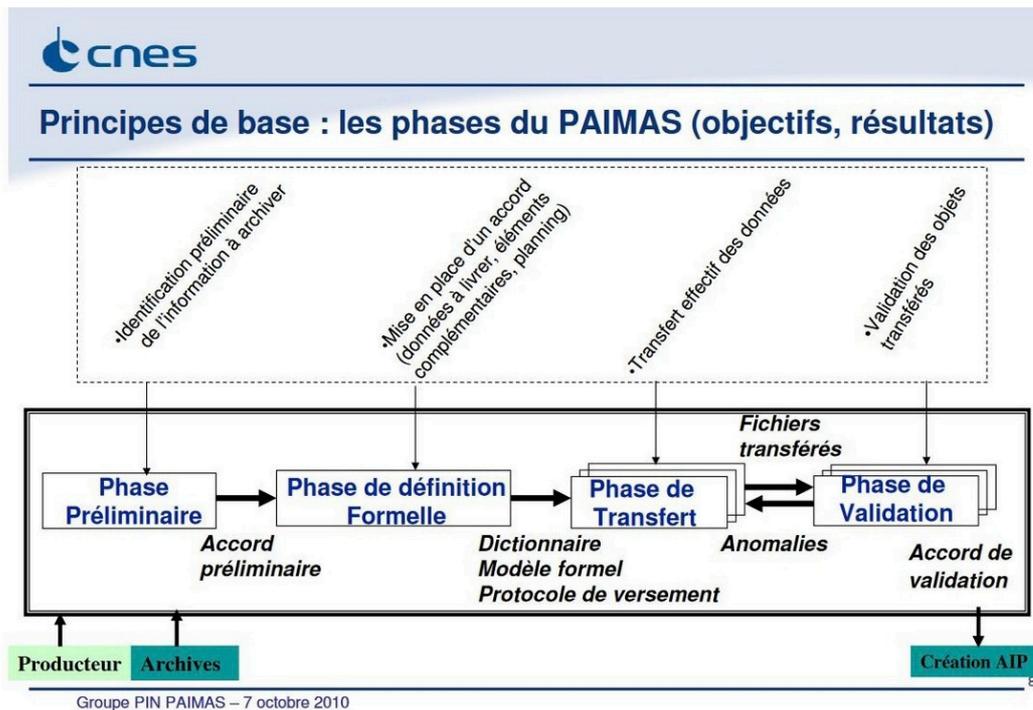


Fig.2 : **Enchaînement des opérations de préparation des transferts selon la norme PAIMAS** (source : présentation effectuée par Danièle Boucon lors de la séance du groupe PIN du 7 octobre 2010, disponible à l'adresse <https://web.archive.org/web/20211022084118/http://pin.association-aristote.fr/lib/exe/fetch.php/public/presentations/2010/pin20101007-pres04-paimaspais.pdf>)

## 2.4. Le cas des documents numériques mal ou pas identifiés

Naturellement, il est des cas où l'intégration est moins préparée en amont avec le service producteur et où le service d'archives reçoit des supports numériques contenant des documents d'archives sur support numérique non identifiés.



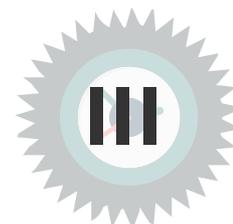
Dans ce cas, l'intégration se fait selon le processus suivant :

- identification du support de stockage, avec recours à des bases de connaissances externes si jamais celui-ci ne correspond pas à un type de support connu. Des sites internet recensent les types de supports informatiques existants (ex. : <https://www.ina.fr/ina-eclair-actu/la-saga-des-supports-de-stockage-de-donnees>) ;
- vérification de l'état sanitaire du support et des fichiers qu'il contient ;
- lecture des fichiers écrits sur les supports. Dans le cas où aucun matériel n'est disponible pour lire le contenu de ces supports, il est toujours possible de regarder si des lecteurs d'occasion ne sont pas disponibles, si d'autres institutions (y compris de police judiciaire) en disposent ou si des prestataires spécialisés offrent des services de lecture ;
- caractérisation du contenu et caractérisation du support, en utilisant les méthodes présentées dans la section 2 de ce module, pour identifier l'environnement matériel et logiciel nécessaire pour permettre la lecture des fichiers concernés ;
- lecture du contenu avec les matériels et logiciels identifiés lors de la phase de caractérisation.

Après examen archivistique de ces contenus, et dans la mesure où ceux-ci nécessitent l'intégration dans un système d'archivage, le processus présenté ci-dessus peut reprendre à partir de l'étape de retraitement des fichiers.

# 3. Description des documents d'archives sur support numérique et des actions générées par les opérations de préservation

---



## Objectifs

**Cette partie complète et précise la section 3 (La description et l'indexation des documents numériques) du module 7B .**

Comme pour les documents d'archives sur support physique ou analogique, la description des documents d'archives sur support numérique est essentielle, à la fois pour garantir leur préservation et leur intelligibilité tout au long de leur durée de conservation et pour faciliter la recherche et l'accès.

### **Cette description concerne :**

- le contexte de production et le contenu des documents eux-mêmes : les métadonnées descriptives ;
- les règles de gestion associées à ces documents d'archives (durée de conservation, conditions de communicabilité, de diffusion, de réutilisation, clauses de confidentialité, etc.) : les métadonnées de gestion ;
- les propriétés techniques des fichiers correspondants aux documents : les métadonnées techniques ;
- la trace des actions mises en œuvre par le service d'archives durant toute la période où celui-ci a la charge des documents d'archives.

### **Nous nous attarderons principalement dans ce chapitre sur**

- les métadonnées descriptives qui permettent de comprendre la structure et la sémantique des informations contenues dans les documents ;
- les métadonnées techniques qui ont été spécifiquement développées pour la prise en charge et la conservation des fichiers où sont stockés les documents d'archives sur support numérique. Ce sont ces métadonnées sur lesquelles s'appuient la définition et la mise en œuvre des opérations de préservation numérique, et qui déterminent les conditions dans lesquelles l'accès aux documents peut être effectué (environnement matériel et logiciel à fournir aux utilisateurs) ;
- les traces qui sont générées par le service d'archives pour mettre en œuvre les opérations de préservation. Pour les traces à générer dans le cadre des opérations d'accès, de réutilisation, de valorisation et de diffusion, nous renvoyons aux modules de cours 11 et 12.

L'enregistrement et la gestion des métadonnées dépendent du système d'archivage utilisé par le service d'archives :

- utilisation de tableurs ;
- utilisation d'une base de données dédiée standard ou métier ;
- utilisation d'un système informatique dédié à la gestion des documents d'archives (voir la section 4 de ce module).

**Quelle que soit la méthode utilisée, il est essentiel de sauvegarder les métadonnées et de conserver le lien intellectuel entre métadonnées et fichiers numériques.**

### 3.1. Les métadonnées descriptives

D'une manière générale, la manière de décrire les documents ne change pas parce que ceux-ci sont sur support numérique.

Cependant, dans certains cas, notamment dans **le cas d'archives produites au moyen d'une application utilisant un système de gestion de base de données** – donc codées et non directement lisibles et exploitables sans informations de représentation –, il peut être nécessaire de disposer de métadonnées descriptives plus pointues que celles requises pour comprendre des documents d'archives sur support physique ou analogique.

**En fonction du mode d'export des informations contenues dans l'application** (export des différentes tables « à plat », cf. la section 5 de ce module), il pourra en effet être nécessaire de disposer :

- de la description de la structure de la base de données au cœur de l'application, c'est-à-dire de la manière dont les différentes tables s'articulent entre elles (ce que l'on appelle le *modèle relationnel de données*) ;
- de la structure de chaque table (nombre de « colonnes », nombre de lignes/d'enregistrements, signification de chaque « colonne ») ;
- de la signification des valeurs saisies dans chaque case de la table (ce que l'on appelle le dictionnaire des données), surtout si celles-ci ont fait l'objet d'une codification (ex. 0 = non, 1 = oui).

**Toutes ces informations seront nécessaires pour permettre d'accéder au contenu des documents produites au moyen de l'application sur le long terme.**



Dans de nombreux cas, des métadonnées descriptives des documents existent dans le nommage des fichiers correspondants ou dans les systèmes dans lesquels elles ont été créées et gérées. L'archiviste aura intérêt à récupérer ces métadonnées automatiquement, soit par un export au moment de l'extraction des documents du système de production, soit au moyen d'un outil d'extraction.



Si de nombreuses métadonnées descriptives peuvent être récupérées ou extraites automatiquement, toutes ne peuvent pas l'être (ex. service producteur) et leur renseignement nécessite l'intervention d'un archiviste.



Les métadonnées, comme les documents, doivent faire l'objet d'une évaluation et d'une sélection. D'une part, il n'est pas utile – voire légal – de toutes les récupérer sous forme de métadonnées descriptives interrogeables. D'autre part, leur qualité peut se révéler extrêmement hétérogène : saisie réalisée sans règle par de nombreux utilisateurs d'une application ; changement de règles mises en œuvre par un utilisateur.

## 3.2. Les métadonnées techniques

Les métadonnées techniques sont indispensables pour assurer la préservation et l'accès sur le long terme aux fichiers correspondant aux documents d'archives sur support numérique.

On distingue généralement :

### 3.2.1. Les métadonnées génériques

Elles sont indispensables quel que soit le format du fichier correspondant :

- identifiant/nom du fichier ;
- format du fichier (de préférence en se fondant sur les référentiels de formats présentés dans le chapitre 7.2.) ;
- empreinte du fichier (valeur et algorithme utilisé) ;
- taille du fichier (en octets) ;
- localisation des différentes copies du fichier sur le stockage ;
- date de dernière modification du fichier – cette information est souvent peu fiable, car elle correspond bien souvent à la dernière date d'écriture du fichier sur un support quelconque.

### 3.2.2. Les métadonnées spécifiques

Les métadonnées spécifiques à certains types de documents d'archives et aux formats de fichiers qui sont utilisés pour les créer :

- images fixes : standards de métadonnées type EXIF, IPTC ou XMP ;
- son et images animées : standards de métadonnées type AudioMD, VideoMD, BWF.



**Certaines de ces métadonnées** sont directement embarquées dans les fichiers eux-mêmes (ex. métadonnées techniques des images fixes). Il est souvent possible d'extraire ces métadonnées des fichiers pour alimenter les outils de recherche du service d'archives.



**Certains types de fichiers** enregistrent un nombre considérable de métadonnées techniques qu'il est possible d'extraire. Ces métadonnées n'ont cependant pas toutes d'intérêt pour garantir la préservation numérique.

## 3.3. Traces générées lors des opérations de préservation



Toute opération de préservation effectuée sur les documents d'archives sur support numérique, avant ou pendant leur prise en charge par le service d'archives, doit faire l'objet d'un enregistrement.

Dans la mesure où les actions de préservation sont en effet susceptibles de **modifier l'intégrité binaire des fichiers** représentant les documents – à l'occasion par exemple d'une migration de format –, il est essentiel de documenter au maximum les opérations qui sont effectuées. Ce sont ces traces générées qui permettront d'attester que ces documents restent dignes de confiance et qu'ils restent exacts et authentiques.

Les opérations concernées, indépendamment de celles liées au *cycle de vie* (gestion du sort final) et à la communication, sont :

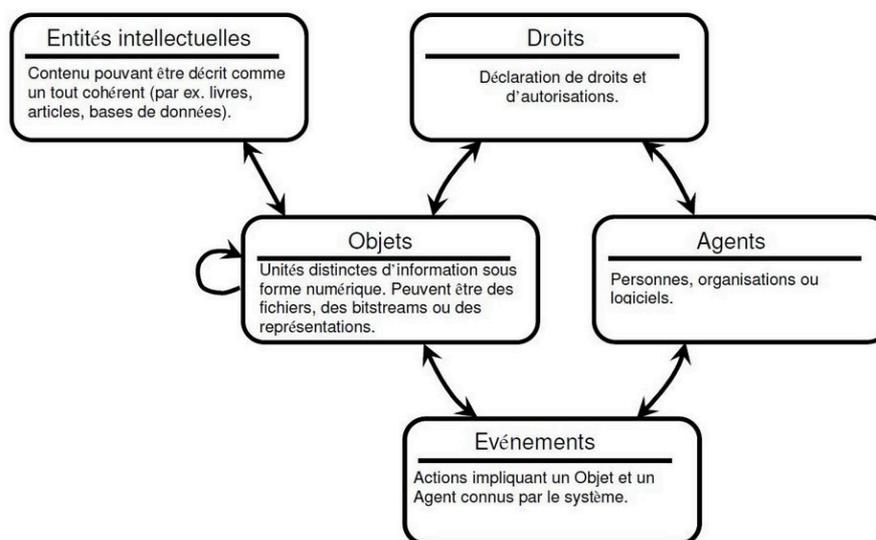
- les contrôles effectués au moment de l'intégration dans le système d'archivage ;
- l'écriture sur les supports de stockage ;
- les vérifications régulières de l'intégrité binaire des fichiers écrits sur les supports de stockage ;
- les vérifications régulières de caractérisation des formats de fichiers (identification de format) ;
- les migrations de support de stockage ;
- les migrations de formats de fichiers.

### 3.4. Le standard PREMIS (PREservation Metadata Implementation Strategies)

La Bibliothèque du Congrès héberge depuis 2005 un standard de métadonnées développé pour la préservation numérique par une équipe internationale sous l'égide d'Online Computer Library Center (OCLC) et du Research Libraries Group (RLG) : le *standard PREMIS* (PREservation Metadata Implementation Strategies). Ce standard est devenu un standard de fait, largement utilisé par les institutions assurant des missions de préservation numérique.

#### 3.4.1. Description de PREMIS

Ce standard de métadonnées propose un modèle de données articulé autour de cinq entités intervenant dans le processus de préservation numérique, selon le schéma suivant :



**Fig.3 : Schéma du modèle de données PREMIS** (source : Caplan (Priscilla), Comprendre PREMIS, The Library of Congress, 2009, 26 p., traduction française par Louise Fauduet, Clément Oury, Sébastien Peyrard, Jean-Philippe Tramoni [en ligne], disponible sur <https://www.bnf.fr/fr/premis-preservation-metadata-implementation-strategies>)

*NB, dans ce schéma, bitstream correspond à des trains d'octets, décrits comme suit dans la documentation fournie : Les objets train d'octets sont des sous-ensembles de fichiers. On définit un objet train d'octets comme des données (octets) au sein d'un fichier qui 1) ont des propriétés communes dans une perspective de préservation, et 2) qui ne sont pas indépendantes car elles nécessitent l'addition d'un en-tête de fichier ou d'une autre structure. Donc, si l'on a par exemple un fichier dans le format AVI (audio-video interleaved), on peut vouloir distinguer le train d'octets audio du train d'octets vidéo, et les décrire comme des objets trains d'octets séparés.*

Les **entités intellectuelles** correspondent à un ensemble cohérent de contenus décrits comme une unité (ex. un livre).

Les **objets** correspondent aux fichiers numériques (ex. un fichier au format PDF).

Les **événements (ou actions)** correspondent aux actions de préservation (ex. conversion du format d'un fichier).

Les **agents** correspondent aux personnes, aux organisations ou aux matériels et logiciels qui réalisent les événements (ou actions) sur les objets.

Les **droits** renseignent les autorisations qui sont données aux agents pour qu'ils effectuent les événements sur les objets.

Ce modèle de données constitue un guide pour identifier les métadonnées à enregistrer dans un système d'archivage dont l'objectif est d'assurer la préservation de documents d'archives sur support numérique, ainsi que pour générer les traces qui vont permettre de documenter les actions mises en œuvre par le système.

### 3.4.2. Le standard PREMIS ne constitue pas une solution toute faite

Il se focalise exclusivement sur l'identification des informations traçant les actions effectuées par les institutions en charge de la préservation d'objets numériques sur leurs propres objets, et recense ce qu'il convient de collecter et d'enregistrer dans ce cadre précis. PREMIS, en conséquence, ne s'intéresse pas aux métadonnées descriptives.



Il n'est pas indispensable de transposer à la lettre le schéma du standard pour mettre un système d'archivage en conformité avec celui-ci. Il suffit, pour garantir la conformité, que les informations correspondant à des métadonnées PREMIS utilisées par le système d'archivage aient la même signification et la même sémantique que celles prévues dans le standard.

La Bibliothèque du Congrès et la communauté qui maintiennent le standard PREMIS proposent un certain nombre d'outils (supports de présentation, vocabulaire contrôlé, logiciels, tableurs, scripts) pour faciliter la prise en main et l'utilisation de celui-ci.

## 4. Stockage des documents d'archives sur support numérique



### 4.1. Les spécificités du stockage des documents d'archives sur support numérique

Le stockage des documents d'archives sur support numérique leur est spécifique. D'autant que le nombre des fichiers à prendre en charge ainsi que leur volume sont toujours croissants.

Il présente quelques spécificités par rapport au stockage numérique classique que nous allons présenter :

#### 4.1.1. Le taux de perte

Le taux de perte accepté en cas d'incidents ne peut qu'être le plus réduit possible, dans la mesure où il s'agit de fichiers qui constituent des preuves de l'activité institutionnelle ou des traces historiques. Des mesures de sauvegarde et/ou de réplication doivent donc être prises

#### 4.1.2. L'intégrité

L'*intégrité* des fichiers est essentielle pour garantir l'*authenticité*. Une surveillance régulière et proactive de cette intégrité est donc nécessaire, notamment si le service de stockage est externalisé.

#### 4.1.3. Les durées de conservation

Les durées de conservation des documents d'archives sur support numérique peuvent se monter à des dizaines voire des centaines d'années. En raison de l'obsolescence rapide des supports informatique, des migrations de supports seront indispensables.

#### 4.1.4. La consultation des documents

La consultation des documents est irrégulière. Si certains documents sont fréquemment demandés par les usagers – et rapidement après leur archivage –, d'autres, en revanche, peuvent n'être demandés qu'au bout d'un temps plus ou moins long (plusieurs années voire plusieurs décennies). Tous les fichiers n'ont donc pas vocation à être stockés sur des supports de stockage immédiatement accessibles (ex. documents d'archives soumis à des délais de libre communicabilité longs).

#### 4.1.5. La taille des documents d'archives

La taille des documents d'archives est variable (de quelques kilooctets à quelques centaines de gigaoctets, notamment les documents audiovisuels) et leur nombre également. Les infrastructures de stockage doivent donc s'adapter aux différentes situations possibles.

#### 4.1.6. Les documents contenant des informations confidentielles

Certains documents contiennent des informations confidentielles (*données à caractère personnel*, données protégées au titre du secret industriel et commercial, données protégées au titre du secret de la défense nationale, au titre de la vie privée etc.), voire des informations sensibles (données de santé, informations sur les opinions politiques, religieuses, syndicales, etc. de personnes vivantes). La création d'infrastructures de stockage dédiées à certaines catégories d'archives, le contrôle des accès

voire le chiffrement des fichiers - dans la mesure où les clés de déchiffrement sont correctement gérées – sur ces infrastructures pourront s'avérer nécessaires. Il conviendra également d'éviter que ces fichiers soient aisément identifiables, *moissonnables* ou consultables.

## 4.2. Recommandations pour le stockage des documents d'archives sur support numérique

**La politique de stockage adoptée par le service d'archives dépendra de la taille du fonds qu'il conserve** et du nombre de fichiers que celui-ci contient, des ressources disponibles (dans le service d'archives ou dans l'organisation à laquelle il est rattaché) et des risques acceptés.

**Le service de stockage peut être mis en place progressivement**, en commençant par des solutions simples. Il convient cependant de prévoir rapidement une automatisation des opérations et une mise à l'échelle adaptée à l'évolution du fonds.

**Les performances en matière d'écriture et de lecture du service de stockage doivent être évaluées** pour garantir une prise en charge et un accès adapté aux documents d'archives sur support numérique.



Il est recommandé d'écrire les fichiers correspondants aux documents dès leur prise en charge sur des supports appropriés. La conservation de supports de stockage de transfert ou anciens n'est pas une bonne solution, au-delà de la prise en charge (ex. disquettes, CD ayant servi au transfert physique).

L'idéal est de disposer de **plusieurs copies indépendantes des fichiers (de 2 à 3)**, de les stocker dans des lieux distants (en vérifiant régulièrement la cohérence des copies), voire de recourir à des modes de stockage différents (stockage sur disques versus stockage sur bandes, stockage en ligne versus stockage hors ligne).



Il est fortement recommandé, si deux copies distantes existent et sont écrites de manière synchrone, qu'elles ne soient pas dépendantes de la même technologie de stockage. Une erreur sur une copie se répercutera inmanquablement sur l'autre copie.

Il est également fortement recommandé de recourir, à côté d'un stockage chaud (sur disque), à un stockage froid (sur bande), moins facilement sujet aux attaques des pirates informatiques.

**Le service d'archives doit conserver le registre complet des fichiers qu'il stocke**, avec leur empreinte et la localisation des différentes copies, **sur un espace de stockage différent, sécurisé et sauvegardé**. Cela lui permettra de procéder régulièrement à des audits du service de stockage et de vérifier l'intégrité des fichiers. En cas de détection d'une anomalie, les autres copies serviront à remplacer la copie défectueuse.

La reconstruction du service de stockage en cas de défaillance généralisée d'une copie doit être testée. L'élaboration et la mise en œuvre d'un plan de continuité d'activité et d'un plan de reprise d'activités sont essentielles.

**L'obsolescence des supports de stockage est rapide** (5 ans en général). Aussi les migrations de support de stockage doivent être fréquentes et constituent des opérations techniques qui, dans l'idéal, doivent être réalisées par des spécialistes.

Le recours à des systèmes de stockage informatiques externalisés, mettant en œuvre la technologie du cloud computing doit être conditionné à l'existence de garanties en matière de sécurité, d'intégrité et également de *réversibilité* des fichiers confiés à ce service. **La réversibilité** – manière dont les fichiers peuvent être extraits du service de stockage pour être transférés dans un nouveau service – est un point trop souvent négligé.

# 5. Actions spécifiques de préservation des documents d'archives sur support numérique

---



## Introduction

La préservation constitue naturellement le cœur de notre sujet et recouvre deux domaines différents, peu ou prou assimilables à ce qui est pratiqué pour les supports physiques ou analogiques :

- la planification des actions de préservation ;
- la réalisation d'opérations de préservation.

## 5.1. La planification des actions de préservation

Comme pour les actions de conservation et de restauration des documents d'archives sur support analogique, la préservation des documents d'archives sur support numérique doit faire l'objet d'une planification.

### 5.1.1. Cadrage et état de l'art :

- le registre des documents pris en charge par le service (cf. section 2. chapitre 2) qui permet d'identifier les grandes catégories de formats de fichiers numériques dont sont constitués les fonds d'archives conservés par le service – puisque les actions de préservation sont bien souvent dédiées à une catégorie de formats de fichiers en tant que telle, voire à un format particulier dont il faut identifier tous les fichiers concernés facilement ;
- la veille assurée par le service d'archives sur les fonds conservés et sur les pratiques mises en œuvre par la communauté professionnelle. Un des objectifs de cette veille est d'identifier les risques de non-disponibilité des documents d'archives – c'est-à-dire l'incapacité de donner accès au contenu de ceux-ci, notamment du fait de l'obsolescence des logiciels utilisés –, de repérer les innovations tant matérielles que logicielles qui pourraient être utiles pour donner accès aux documents d'archives, ainsi que de pressentir l'évolution des attentes des utilisateurs du service. La veille doit donc non seulement concerner les formats de fichiers, mais aussi les outils matériels et logiciels utiles pour donner accès au contenu des documents d'archives et les services attendus par les utilisateurs ;
- les retours d'expériences effectués par d'autres institutions sur des documents d'archives ou collections de même nature, ou d'ores et déjà réalisées par le service d'archives lui-même.

### 5.1.2. Sur cette base, la planification doit permettre de :

- définir les actions de préservation à mettre en œuvre, sur la base des besoins identifiés par les utilisateurs, ainsi que les critères de succès de ces opérations ;
- évaluer les ressources et moyens nécessaires pour les mettre en œuvre – internes comme externes ;
- procéder à une analyse des risques associés à la réalisation de ces opérations – ou au contraire à leur non-réalisation.

Il peut être utile, avant de lancer une opération de préservation, de vérifier la liste des fichiers susceptibles d'être concernés par celle-ci, en mettant à jour le registre des documents d'archives pris en charge par le service. Les registres de formats de fichiers (comme PRONOM) étant régulièrement mis à jour et les outils d'identification affinés (nouveaux formats de fichiers identifiés, correction d'anomalies), il vaut mieux vérifier si la liste des fichiers concernés est toujours à jour de l'état de l'art.

## 5.2. Les actions de préservation

### Introduction

En cas de difficulté récurrente à accéder au contenu de certains types de fichiers correspondant à des documents d'archives, il peut devenir nécessaire de mettre en œuvre des actions de préservation.

**A ce jour cinq techniques sont identifiées :**

1. le remontage de l'environnement de production des documents (ce que certains qualifient de stratégie du « musée de l'informatique ») ;
2. l'*émulation* de l'environnement de création des documents d'archives ;
3. l'identification d'un nouveau logiciel garantissant la lecture du contenu des documents d'archives ;
4. la *migration* ou *conversion* du format de fichiers du document d'archives ;
5. la réparation des fichiers.

#### 5.2.1. Le remontage de l'environnement de production des documents

La première technique consiste à **reconstituer l'environnement matériel et logiciel de production des documents d'archives** sur support numérique. Elle suppose de garantir la conservation de matériels et logiciels anciens, en s'assurant de leur fonctionnement régulier : maintien en bon état, conservation des connaissances nécessaires à leur fonctionnement, existence de pièces de rechanges. Cette stratégie est naturellement difficile à mettre en œuvre sur le long terme et peut à la rigueur convenir pour des opérations de prise en charge ou comme première étape d'une autre stratégie.

#### Exemple

Une stratégie alternative serait de migrer vers un nouveau format de fichiers.

#### Remarque

Le remontage de l'environnement est la stratégie la moins viable de toutes. Elle peut néanmoins être utile au moment du transfert des documents ou lors de sa préparation, soit pour récupérer des documents enregistrés sur un support obsolète, soit pour opérer la migration d'un format de fichiers obsolète vers un format de fichiers plus gérable par le service d'archives.

#### 5.2.2. L'émulation de l'environnement de création des documents d'archives

La deuxième technique consiste à **recréer l'environnement de production des documents d'archives sur une plateforme matérielle et logicielle d'aujourd'hui**, c'est-à-dire à simuler, sur un ordinateur actuel, le fonctionnement d'un ordinateur ancien, avec son système d'exploitation et les logiciels de l'époque. Cette stratégie est appelée **émulation**. Elle est particulièrement opportune quand il s'agit de donner accès à des objets complexes et/ou dynamiques (présentations, jeux vidéo, documents en trois dimensions).

D'un point de vue pratique, cette stratégie repose sur l'utilisation d'un émulateur, qui est un logiciel informatique permettant de simuler le fonctionnement d'un environnement matériel et logiciel ancien. Elle suppose cependant de disposer d'images disque des environnements anciens, qui sont utilisées par l'émulateur. Ces images disque, copie conforme d'un environnement originel, peuvent avoir été réalisées en même temps que la collecte des fichiers ou être récupérées depuis des environnements semblables. Depuis quelques années, des travaux sont menés par des chercheurs allemands et américains pour faire fonctionner les émulateurs sur des environnements de type cloud, dans une logique d'Emulation-as-a-Service.



**Attention**

Cette approche, très technique, est cependant confrontée à plusieurs contraintes et limites :

- contraintes juridiques d'utilisation des logiciels (licences) ;
- contraintes techniques de connaissance et de prise en main des environnements anciens ;
- nécessité de mettre à jour régulièrement les émulateurs pour créer des environnements nouveaux.

Cette stratégie est techniquement complexe, car elle implique néanmoins de gérer des environnements matériels et logiciels sous forme d'images disque.



**Remarque**

La technique de l'émulation peut s'avérer utile en cas de fort besoin d'accès, et dans ce cas, plus économique sur le plan technique et financier.

### 5.2.3. L'identification d'un nouveau logiciel

La troisième technique consiste tout simplement à rechercher, parmi les logiciels disponibles, ceux qui sont capables de prendre en charge des fichiers anciens. Dans bien des cas en effet, un simple changement de logiciel d'accès permet de résoudre les problèmes. Les logiciels de bureautique open source, par exemple, prennent en charge une plus grande variété de formats de fichiers que les logiciels de bureautique propriétaires, bien souvent focalisés sur la version la plus récente des formats de fichiers (en tout cas pour les fichiers de type traitement de texte, tableurs et présentations).

Dans d'autres cas, l'utilisation d'autres logiciels permettront d'accéder au contenu, mais sans conserver les éléments de mise en forme.



**Exemple**

Ouverture d'un message électronique avec un éditeur de texte type BlocNote ou NotePad.

La mise en œuvre de cette stratégie **suppose donc que le logiciel utilisé garantisse le maintien des propriétés porteuses de sens** (les significant properties) des documents (formatage, présentation, fonctionnalités), ou du moins que leur perte éventuelle soit considérée comme non critique par le service d'archives. À défaut, il faut recourir à la stratégie d'émulation.

### 5.2.4. La migration ou conversion du format de fichiers

La quatrième technique est celle qui a été le plus largement recommandée dans la littérature professionnelle : la migration ou conversion de format.

Il s'agit de changer le format de fichiers dans lequel est enregistré un document d'archives, et de :

- passer d'un format qui n'est plus pris en charge par les logiciels de restitution de l'environnement actuel à un format pris en charge par les logiciels de restitution de l'environnement actuel. Cette opération peut se faire soit de manière systématique – tous les fichiers du format source conservés par le service d'archives sont traités lors de la même opération –, soit de manière ponctuelle, à l'occasion des demandes d'accès formulées par les utilisateurs ;
- passer d'un format considéré comme non susceptible de garantir la préservation et l'accès à long terme d'un document d'archives, à un format considéré comme susceptible de le faire. On parle souvent de normalisation pour désigner ce type d'opération qui peut être opéré avant la prise en charge, au moment de la prise en charge ou a posteriori, pendant la conservation par le service d'archives.



**Fondamental**

La mise en œuvre de cette opération suppose que le format cible de la migration garantisse le maintien des propriétés porteuses de sens (les significant properties) des documents (formatage, présentation, fonctionnalités), ou du moins que leur perte éventuelle soit considérée comme non critique par le service d'archives.

**Ceci implique donc une étude approfondie, une analyse de risque, ainsi que la réalisation de tests approfondis**, y compris pour évaluer les ressources humaines et techniques qui sont nécessaires à la mise en œuvre de l'opération ainsi qu'au contrôle de son résultat. Il est indispensable de documenter précisément les choix effectués, et notamment les pertes d'informations acceptées.

### À noter

- changer de format de fichiers porte nécessairement atteinte à l'intégrité binaire des fichiers. **Il convient donc de bien évaluer les risques** à faire ou à ne pas faire l'opération et surtout à la **documenter** pour garantir que l'authenticité (authenticité juridique) des documents d'archives n'est pas altérée par l'opération ;
- dans le cas où la migration s'opère sur des documents d'archives nativement numériques, il est fortement recommandé de conserver également le document dans le format sous lequel il a été pris en charge par le service d'archives. L'enjeu sera ensuite de gérer les différents fichiers associés au document d'archives ;
- **la migration n'est pas une solution adaptée pour certains types de formats de fichiers**, notamment tous ceux qui présentent une animation (présentations, jeux vidéo) ou les formats vectoriels (plans, cartes, objets en trois dimensions) ;
- **les opérations de migration sont gourmandes en ressources informatiques** (stockage, mémoire vive), ce qui n'est pas sans conséquences financières et environnementales. Il est recommandé de les réaliser sur des environnements distincts de ceux permettant de traiter les transferts en cours, faute de quoi ceux-ci pourraient se voir retardés voire arrêtés.

### 5.2.5. La réparation des fichiers

La cinquième et dernière technique consiste à réparer des fichiers endommagés ou considérés comme non conformes aux spécifications de leur format.

Si elle est mise en œuvre, cette opération doit être soigneusement documentée.



**Attention**

Cette opération, très technique, doit être évaluée avec précaution, car elle entraîne nécessairement la modification du train binaire du fichier et porte donc atteinte à son intégrité technique.

Si elle est mise en œuvre, cette opération doit être soigneusement documentée.

## Conclusion

D'un point de vue général :

- **il n'existe pas de stratégie idéale ou de solution miracle** et entièrement automatique, car chaque type de format de fichiers implique une stratégie différente, voire la combinaison de différentes stratégies. La section 5 de ce module s'efforce de donner quelques conseils ;
- les opérations de préservation numérique, comme les opérations de conservation préventive et de restauration des documents d'archives sur support physique ou analogique, ont un **coût** qu'il s'agit d'évaluer, et présentent des **risques** qu'il s'agit d'analyser ;
- pour lancer toute opération systématique de préservation numérique, il convient de s'appuyer sur les métadonnées techniques renseignées dans le registre des documents d'archives conservés par le service d'archives. Une vérification fine des formats de fichiers susceptibles d'être intégrés à l'opération est recommandée, dans la mesure où les outils d'identification de format et les registres sur lesquels ils s'appuient peuvent avoir fait l'objet de mises à jour utiles ;
- il faut tenir compte de la volumétrie, notamment le nombre de fichiers, laquelle peut conditionner les stratégies choisies.

## 6. L'accès aux documents d'archives sur support numérique



### 6.1. La numérisation facilite l'accès aux documents d'archives sur support analogique ou physique

L'accès aux documents d'archives sur support numérique est à la fois une problématique bien maîtrisée par les services d'archives et le sujet de nouveaux questionnements, parfois encore embryonnaires.



**Rappel**

Il convient de rappeler que les opérations de préservation numérique ont pour but principal de permettre l'accès à long terme aux documents d'archives sur support numérique.

Depuis le début des années 2000, les services d'archives ont en effet entrepris d'importantes campagnes de numérisation des documents d'archives sur support physique ou analogique, leur diffusion sur des portails dédiés, ainsi que la fourniture de copies en réponse à des demandes de livraison. Ils disposent donc désormais d'une expérience en matière de :

- construction de portails d'accès et définition des fonctionnalités associées ;
- sélection des logiciels permettant de donner accès au contenu des documents (visionneuses) ;
- hébergement des fichiers ;
- mise en place d'une politique d'identification adaptée à l'environnement web, via des identifiants permanents ou pérennes.

Mais il s'agit bien souvent d'archives sur support physique ou analogique numérisées dans des formats de fichiers en nombre limités et maîtrisés (formats PDF, formats d'images fixes et animées homogènes), librement communicables et diffusables.

### 6.2. Les documents nativement numériques peuvent poser des problèmes d'accès.

Les documents d'archives nativement numériques présentent certaines spécificités non encore totalement maîtrisées en accès.

#### 6.2.1. Souvent, ils ne sont pas communicables et/ou diffusables

**Leur caractère récent** fait qu'ils ne sont souvent pas librement communicables ou diffusables. Il est donc impossible d'y accéder via des portails sans contrôle d'accès. Si certaines structures – notamment dans l'enseignement supérieur et la recherche – ont mis en place ce type de portails, il n'en existe que quelques-uns mis à disposition par les services d'archives.

Les Archives départementales de la Vienne en France ou les Archives fédérales suisses. Des "salles de lecture virtuelles" sont notamment mises à disposition par les archives départementales de la Vienne (France) ou les Archives fédérales suisses. Elles donnent accès, sous réserve de la création d'un compte personnel dont l'identité a été vérifiée, à des documents dont la libre diffusion sur internet n'est pas juridiquement possible mais la consultation en salle de lecture l'est. Par exemple, pour la Vienne, l'état civil numérisé librement communicable au regard du Code du patrimoine mais datant de moins de cent ans.

## 6.2.2 L'accès peut nécessiter des outils spécifiques

Les modalités de conservation retenues pour certains types d'archives sur support numérique nécessitent la mise à disposition d'outils spécifiques pour donner accès à leur contenu. C'est notamment le cas des documents produits dans des applications basées sur des systèmes de gestion de bases de données dont les exports ont été faits à plat ou dans un format de fichiers qui nécessite l'import dans une nouvelle base de données (cf. section 5 de ce module). Une réflexion est donc nécessaire pour définir les matériels et les logiciels nécessaires pour donner l'accès aux documents d'archives nativement numériques.

### 6.2.3. Les attentes des utilisateurs sont très variées

Les attentes des utilisateurs sont plus diverses que pour les documents d'archives sur support physique ou analogique numérisés pour lesquels les portails mis en ligne correspondent aux besoins estimés du grand public.

La réutilisation de données brutes, directement exploitables par machine, est ainsi particulièrement souhaitée par les spécialistes du gouvernement ouvert, les industriels ou les chercheurs, là où la fourniture des données sous la forme d'un fichier non exploitable par machine (ex. fichier en mode image) sera considéré comme inadéquat.

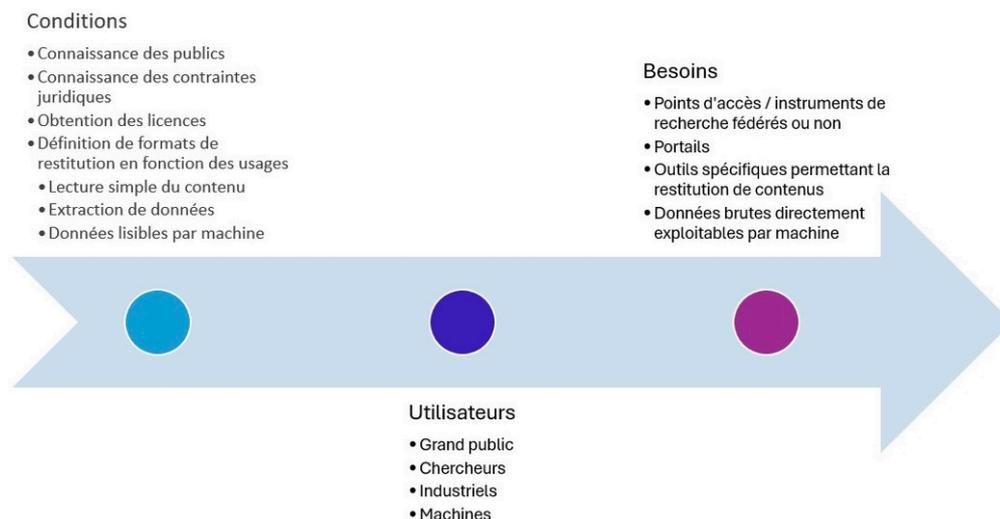
Une connaissance plus fine des usages des différents publics est donc indispensable.

## 6.3. Prérequis pour donner accès aux documents d'archives sur support numérique

Donner accès à des documents d'archives sur support numérique suppose donc d'avoir :

- analysé les besoins des différentes communautés d'utilisateurs (chercheurs universitaires, grand public, archivistes, magistrats) et ce qu'ils veulent faire ;
- identifié les environnements matériels et logiciels permettant de restituer le contenu des différents types de formats de fichiers conservés par le service d'archives – en se basant sur le recensement des documents conservés et l'identification des formats de fichiers, ainsi que les opérations de préservation nécessaires (émulation d'un environnement ancien, migration systématique ou ponctuelle) ;
- obtenu le droit (licences) et les moyens (notamment techniques) d'utiliser les matériels et logiciels identifiés ;
- défini le format de fichiers dans lequel le document d'archives doit être fourni, afin de répondre aux besoins des utilisateurs et être pris en charge par l'environnement matériel et logiciel retenu ;

- autorisé ou non les utilisateurs à accéder et réutiliser le contenu des documents, en fonction des contraintes juridiques applicables à celles-ci.



**Fig.4 : Modélisation du processus d'évaluation des besoins en termes d'accès** (crédits : B. Grailles / PIAF)

### 6.3.1. Analyser les besoins des différentes communautés d'utilisateurs

- **Les attentes des utilisateurs humains doivent être analysées avec soin**, pour définir autant que possible la politique de préservation numérique. Pour un même document d'archives, certains utilisateurs n'auront besoin que de lire le contenu ; d'autres souhaiteront extraire les données textuelles, chiffrées ou géolocalisées pour effectuer des analyses plus approfondies (fouilles de données par exemple). Fournir à ceux-ci des documents d'archives dans un format qui ne permet pas l'extraction des données concernées sera considéré comme contre-productif. Si tous les usages ne peuvent être envisagés à un instant T, eu égard aux capacités techniques alors disponibles, il devient de plus important de ne pas les rendre impossibles à l'avenir.
- **Les attentes des utilisateurs doivent faire l'objet d'un réexamen régulier**, en utilisant toutes les méthodes classiques d'identification de celles-ci : questionnaires de satisfaction, entretiens individuels ou collectifs, enquêtes, etc.
- Les utilisateurs pourront avoir besoin d'éléments pour comprendre pourquoi un document d'archives produit il y a de nombreuses années sous une forme donnée se présente aujourd'hui sous une autre forme. **L'historique précis et détaillé de l'ensemble des opérations réalisées sur le document d'archives depuis sa production est essentiel** pour que celui-ci soit considéré comme digne de confiance.

#### ? Exemple

Les plus anciens documents d'archives sur support numérique conservés par les Archives nationales en France ont été produits dans les années 1960. La forme qu'ils prennent aujourd'hui n'a plus rien à voir avec celle d'origine, mais l'ensemble des opérations de préservation ayant été documenté, les documents sont considérés comme étant dignes de confiance.

### 6.3.2. Identifier les environnements matériels et logiciels

Pour les documents d'archives sur support numérique, il faut prendre en compte le fait que **les machines peuvent être considérées comme des utilisatrices en tant que telles**, ce qui peut amener à conserver des documents dans des formats qui soient interprétables par elles (ex. fichiers XML, JSON).

Les solutions techniques permettant d'exploiter tout le contenu des documents d'archives sur support numérique ne sont pas nécessairement au point au moment de leur prise en charge. **La conservation de plusieurs versions peut être envisagée**, dans la mesure où son coût n'est pas exorbitant.

### 6.3.3. Définir le format et la version de fichiers dans lequel le document d'archives doit être fourni

Il est nécessaire de définir l'état du document d'archives qui doit être fourni aux utilisateurs, surtout si plusieurs états sont conservés par le service d'archives (fichier dans son format d'origine, fichier migré dans un autre format).

On ne donne accès qu'à des copies des fichiers conservés, en veillant à ce que celles-ci soient convenablement gérées pour éviter toute faille dans la politique de sécurité (cf. ci-dessous).

### 6.3.4. Autoriser ou non les utilisateurs à accéder et réutiliser le contenu des documents

L'enregistrement des accès aux documents d'archives sur support numérique n'a pas vocation à être systématique, comme pour les documents d'archives sur support physique ou analogique, dans la mesure où seulement des copies des fichiers sont fournies aux utilisateurs. Il n'y a donc pas de risque de vol et de destruction involontaire d'archives. En revanche, l'accès à des documents d'archives non librement communicables, réutilisables, valorisables ou diffusables doit continuer à être enregistré, pour surveiller d'éventuelles divulgations, réutilisations, valorisations ou diffusions non autorisées.

L'accès aux documents d'archives est important, mais l'accès aux métadonnées (descriptives, techniques et de gestion), aux éléments de traçabilité et à toute information permettant de comprendre les opérations qu'ont subies les documents depuis leur création est tout aussi important à envisager.

#### Exemple

Les magistrats, par exemple, peuvent avoir besoin de disposer, à côté des documents eux-mêmes, de toutes ces informations, y compris des éléments très techniques (contrôles d'intégrité).

La fourniture d'un point de recherche et d'accès aux documents, quel que soit leur support, mérite d'être soigneusement évaluée. Spontanément, les utilisateurs ne peuvent en effet pas présupposer du caractère physique, analogique ou numérique du document d'archives. Mais la cohabitation dans un même outil de recherche (catalogue, répertoire ou tout autre instrument de recherche) peut s'avérer complexe en raison de niveaux de granularité de description différents – les documents d'archives sur support physique ou analogique sont plus souvent décrits au niveau du dossier et quelquefois au niveau de la pièce, là où les documents d'archives sur support numérique disposent plus systématiquement de métadonnées descriptives au niveau de la pièce (rien qu'en extrayant le nom du fichier).

## 7. Mise en œuvre du sort final



### Introduction

La mise en œuvre du *sort final* (élimination et transfert à un autre service d'archives) des documents d'archives sur support numérique présente également quelques spécificités.

### 7.1. L'élimination

La seule manière de supprimer avec certitude un document d'archives sur support numérique consiste à détruire physiquement le support sur lequel il est stocké. Si cette destruction était possible quand les supports de stockage avaient une capacité limitée et ne contenaient que quelques fichiers, elle est tout proprement impossible aujourd'hui, que ce soit sur des supports amovibles, sur des serveurs et, a fortiori, sur des infrastructures de type *cloud computing*.

Sur ce type de supports, la suppression d'un fichier correspond en fait à la libération logique de l'espace de stockage sur lequel il était enregistré sur le support pour rendre possible l'enregistrement d'un nouveau fichier.



**Attention**

Tant qu'un nouveau fichier n'est pas écrit à l'emplacement libéré, l'ancien fichier reste présent et est donc techniquement récupérable, ce qui n'est pas sans poser des questions de sécurité.



**Méthode**

L'élimination d'archives consiste donc à :

- empêcher de faire le lien entre la description de celles-ci et les fichiers qui leur correspondent (on parle de suppression logique ou de déréférencement) ;
- rendre réinscriptibles les espaces de stockage sur lesquels les fichiers correspondants aux documents ont été stockés (le fichier est physiquement effacé suite à une réécriture sur l'espace de stockage correspondant).

#### À noter

- dès lors que plusieurs copies d'un même fichier sont conservées sur différents supports de stockage de types différents, on ne pourra considérer que le document d'archives est complètement éliminé que quand des fichiers auront été écrits à la place de ceux correspondants au document, sur l'ensemble des supports ;
- l'opération d'élimination doit garantir que toutes les copies du document d'archives réalisées pour des besoins de traitement, de sauvegarde ou de réplication, ont été traitées. L'existence de copies de travail stockées sur des postes de travail ou des supports amovibles doit donc être tracée minutieusement ;

- l'opération d'élimination doit également concerner les métadonnées des documents d'archives concernés, notamment si celles-ci sont susceptibles par leur contenu de constituer des traces équivalentes à ces derniers.

## 7.2. Transfert à un autre service d'archives

Le *transfert* à un autre service d'archives ne diffère a priori pas du transfert d'un service producteur à un service d'archives. Comme pour les documents d'archives sur support physique ou analogique, il peut concerner :

- la restitution de documents confiés à un prestataire par leur service producteur ;
- un changement de prestataire ;
- le transfert d'un service d'archives intermédiaire à un autre service d'archives intermédiaire ;
- le transfert à un service d'archives définitif.

**Mais le transfert d'archives sur support numérique amène à prendre en compte un certain nombre de points d'attention.**

### 7.2.1. Le transfert peut être réalisé sous forme d'une opération

- physique : transfert de fichiers d'une plateforme à une autre ou d'une instance à une autre dans une même plateforme ;
- logique : simple modification des droits sur les documents (modification, suppression, lecture) au sein de l'instance d'une plateforme.

### 7.2.2. Le périmètre des éléments à transférer doit faire l'objet d'un échange entre services d'archives

Celui-ci peut en effet comprendre :

- les documents, avec tous les fichiers qui les représentent ou seulement une partie des fichiers pris en charge ou générés par le service d'archives d'origine lors d'opérations de préservation numérique ;
- les métadonnées (descriptives, techniques, de gestion) ;
- les traces des opérations réalisées par le service d'archives sur les documents d'archives et les fichiers qui les représentent ;
- le résultat des contrôles d'intégrité effectués sur les documents.

### 7.2.3. Les données de référence doivent être partagées par les deux services d'archives

Si les métadonnées décrivant un document d'archives et les fichiers qui les représentent s'appuient sur des données de référence, il convient de s'assurer que celles-ci sont partagées par les deux services d'archives et, dans le cas contraire, de définir qui prend en charge leur modification.

### 7.2.4. Les systèmes d'identification pérenne doivent faire l'objet d'un échange entre services d'archives

Les systèmes d'identification pérenne peuvent avoir comme base l'identifiant du service d'archives en tant que tel (ex. identifiant ark : dont le radical correspond à l'immatriculation de l'institution). Il conviendra de définir si l'ancien identifiant doit être transmis sous forme de métadonnées et s'il est possible, dans le nouveau service d'archives, de fournir un accès au document d'archives via une interrogation utilisant l'ancien identifiant.

### **7.2.5. La suppression des fichiers n'intervient qu'après l'acquittement du transfert**

La suppression des fichiers correspondants aux documents d'archives transmis par le service expéditeur ne peut intervenir qu'une fois l'acquittement du transfert émis par le service d'archives destinataire. Cet acquittement peut prendre un temps important, ce qui implique de définir lequel des deux services d'archives, pendant cet intervalle, est responsable de l'accès aux documents concernés, et si des opérations de préservation doivent être entreprises pour le permettre.

# 8. Sécurité et audit

---



## 1. Introduction

Les questions de sécurité et d'audit sont essentielles dans la conservation des documents d'archives. La gestion des documents d'archives sur support numérique présente cependant quelques spécificités.

## 8.1. La sécurité

### 8.1.1. Les spécificités du document numérique d'archives en terme de sécurité

Les questions de sécurité sont tout aussi importantes pour les documents d'archives sur support numérique que pour ceux sur d'autres supports, mais présentent une acuité particulière pour ces dernières dans la mesure où :

- **il est beaucoup plus facile de modifier le contenu d'un document d'archives sur support numérique** que celui d'un document d'archives sur support analogique ;
- **il est extrêmement facile de dupliquer** en de multiples exemplaires les fichiers correspondants aux documents, et le contrôle de leur circulation est extrêmement difficile (cf. supra) ;
- **il est facile d'exploiter le contenu même des documents, ainsi que leurs métadonnées**, notamment quand il s'agit de données à caractère personnel, de données de santé ou d'informations protégées aux titres du secret industriel et commercial ou du secret de la défense nationale. Un simple croisement de métadonnées peut révéler des informations extrêmement sensibles et certaines catégories de données présentent une valeur commerciale non négligeable. Il est donc essentiel de peser soigneusement la récupération des métadonnées descriptives préexistantes dans les environnements de production des documents et d'envisager les options d'anonymisation ou de pseudonymisation des métadonnées et des données elles-mêmes ;
- la **cybersécurité** est devenue une problématique essentielle dans des sociétés de plus en plus connectées et où les menaces ne sont pas uniquement physiques, mais aussi numériques. Campagnes d'hameçonnages, intrusions délibérées dans les systèmes d'information, prise de contrôle à distance des systèmes, chiffrement des ressources constituent des menaces désormais régulières des organisations, contre lesquels des plans de sécurité informatique doivent être mis en œuvre.

### 8.1.2. Les spécificités du traitement de la sécurité

La question de la sécurité doit être traitée à plusieurs niveaux :

- **au niveau du système d'archivage**, dans une optique de sécurisation du système d'information dans son ensemble (voir section 4 de ce module.) ;
- **au niveau des documents eux-mêmes** :
  - empêcher les modifications portant atteinte à l'authenticité des documents par des personnes non autorisées ;
  - empêcher la lecture et la copie des documents et des métadonnées les décrivant par des personnes non autorisées ;
  - empêcher les modifications des métadonnées des documents qui feraient obstacle à leur recherche ou fausseraient les métadonnées sur lesquelles s'appuient les opérations de préservation numérique (métadonnées techniques) ou de gestion de leur cycle de vie (élimination, transfert) – ou du moins tracer systématiquement toutes les modifications opérées.

### 8.1.3. Des solutions spécifiques à mettre en place

La sécurité du système d'archivage et des documents d'archives s'appuie sur :

- **des solutions techniques** : architecture du système, cloisonnement des environnements de travail et de conservation, mise en place d'antivirus ou de pare-feu, chiffrement des fichiers lors des opérations de transfert et sur les espaces de stockage ;
- **des solutions organisationnelles et humaines** : identification des rôles et responsabilités ; gestion des droits des utilisateurs sur le système d'archivage (droits d'écriture, de lecture, de modification et de suppression), traçabilité des opérations réalisées, sensibilisation de tous les acteurs à la sécurité, etc.

### 8.1.4. Les spécificités du plan de prévention des risques



La mise en œuvre de plans de prévention des risques, de continuité et de reprise d'activité – voire de plans d'urgence – est essentielle pour la préservation des documents d'archives sur support numérique.

Elle consiste à :

- apprendre à prévenir les risques susceptibles d'intervenir et de prendre les mesures destinées à empêcher la survenue de ceux-ci ou d'en diminuer les effets ;
- gérer d'éventuels incidents et perturbations, afin de parer à l'interruption des services en cours ou à venir et permettre la reprise de l'activité dans les meilleurs délais et dans les meilleures conditions ;
- permettre la restauration du système sur la base des fichiers gérés par le service de stockage ET de la sauvegarde du registre des documents d'archives (donc les métadonnées descriptives, techniques et de gestion) et des traces des opérations, ce qui permettrait de reconstruire le système en totalité si besoin.

### 8.1.5. Quelques points à ne pas négliger...

#### a) La sécurité informatique est un métier spécifique...

... avec ses référentiels (famille des normes ISO 2700x) et ses procédures. Les services d'archives auront tout intérêt à s'appuyer sur l'expertise des professionnels de la sécurité informatique pour définir les procédures de sécurité applicables à leur système d'archivage.

#### b) La sécurité informatique ne repose pas uniquement sur des solutions techniques.

Le comportement des utilisateurs est tout aussi, sinon plus important, que les solutions techniques. La sensibilisation à la sécurité informatique doit donc être permanente et actualisée en fonction de l'évolution des menaces susceptibles de porter atteinte au système d'archivage et aux documents d'archives conservés dans celui-ci.

#### c) Les responsabilités en matière de sécurité sont souvent davantage partagées...

... que dans la gestion des documents d'archives sur support physique et analogique dans la mesure où la gestion du stockage est souvent déléguée à un service spécialisé, interne ou externe à l'organisation à laquelle est rattaché le service d'archives.

#### d) Le chiffrement des fichiers constitue une solution...

... pour sécuriser le transfert et le stockage des fichiers correspondants aux documents. Il convient cependant de rappeler que le chiffrement s'appuie sur l'utilisation de clés de cryptage qui doivent être gérées pour garantir le déchiffrement des fichiers correspondants. En cas de perte de la clé de chiffrement, le contenu des fichiers ne sera plus accessible.

#### e) Les environnements dédiés

Certaines catégories d'archives nécessitent d'être conservées et stockées sur des environnements dédiés, pour garantir leur sécurité. C'est le cas notamment des documents contenant des données de santé ou couverts par le secret de la défense nationale.

#### f) Prendre en compte la gestion des copies de fichiers

Dans la mesure où de nombreuses copies des fichiers sont réalisées, notamment lors de la prise en charge, des opérations de préservation ou d'accès, il convient de les prendre en compte lors de l'élaboration de la politique de sécurité.

## 8.2. L'audit



La mise en œuvre de procédures d'audit régulières constitue également un axe important de la conservation des documents d'archives sur support numérique.

Des audits doivent porter sur les points suivants :

- **la vérification régulière de la complétude du stockage** : tous les fichiers correspondant à des documents pris en charge par le service d'archives sont-ils bien enregistrés sur les supports de stockage ? Le nombre de copies attendu pour chaque fichier est-il correct ?
- **la vérification régulière de l'intégrité binaire des fichiers enregistrés sur les supports de stockage** : les fichiers enregistrés sur les supports de stockage ont-ils bien la même empreinte que les fichiers transmis ? Toutes les copies des fichiers ont-elles bien la même empreinte, et celle-ci correspond-elle à celle des fichiers transmis ?

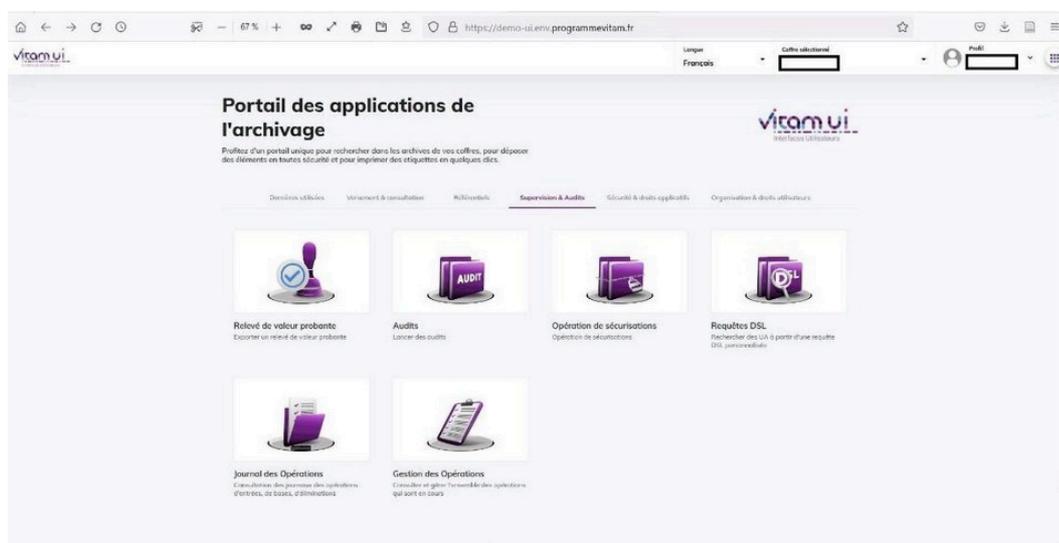
- **la vérification de l'intégrité du document d'archives en tant que tel**, si celui-ci est constitué de plusieurs fichiers (ex. un document signé où document et signature sont représentés par deux fichiers différents mais liés). Dans ce cas, le contrôle porte sur l'ensemble des fichiers correspondant au document d'archives, en prenant en compte l'ensemble des copies correspondant aux différents fichiers qui composent le document.



Dans tous les cas, si des anomalies sont détectées, **il convient de prendre les mesures correctives nécessaires** :

- écriture des fichiers sur l'ensemble des supports de stockage si certaines copies manquent à l'appel ;
- remplacement de la copie corrompue par une copie dont l'empreinte correspond à celle du fichier qui a été transmis au service de stockage.

Dans les plateformes utilisant des logiciels conçus pour la préservation des documents d'archives sur support numérique, des mécanismes automatiques d'audit peuvent être déclenchés. Ce qui ne dispense pas de l'évaluation de l'archiviste en cas de détection d'une anomalie.



**Fig. 5 : Copie d'écran de l'interface VitamUI qui propose des fonctions d'audit intégrées**



- l'audit constitue une opération technique qui nécessite la disponibilité d'une force de calcul suffisante. La mise en œuvre trop régulière ou sur de trop grands ensembles de cette opération peut porter atteinte à la performance globale du système d'archivage, au même titre que les opérations de préservation. Cet emploi de ressources a un impact financier et environnemental non neutre, il faut donc le planifier avec précaution afin d'en réduire l'impact ;
- le registre des documents d'archives pris en charge par le service d'archives constitue un élément clé pour procéder aux opérations d'audit. Il enregistre en effet les métadonnées techniques qui servent de référence pour auditer le service de stockage. La sécurité de ce registre constitue donc un point essentiel, et toute corruption de celui-ci peut porter atteinte à la sécurité globale du système d'archivage ;

- les opérations d'audit ne peuvent être mises en œuvre au même rythme sur tous les supports de stockage. Autant un audit des fichiers enregistrés sur disque peut être facile à mettre en œuvre et être fait régulièrement, autant un audit des fichiers enregistrés sur des bandes déconnectées du réseau peut être long. Le rythme des opérations d'audit doit donc être adapté en conséquence.

Pour en savoir plus sur les fonctionnalités d'audit mises en œuvre par la solution logicielle d'archivage numérique Vitam, consulter la documentation suivante : <http://www.programmevitam.fr/><sup>23</sup>

---

2. <https://www.google.com/url?q=http://www.programmevitam.fr/&sa=D&source=editors&ust=1733240752581808&u sg=AOvVaw0FhRPK6hUQDzLBbg2DVs63>

3. [https://www.programmevitam.fr/ressources/DocCourante/autres/fonctionnel/VITAM\\_Preservation.pdf](https://www.programmevitam.fr/ressources/DocCourante/autres/fonctionnel/VITAM_Preservation.pdf)