



# ARCHIVAGE DES DONNEES

**Méthodologie et cas d'usage**

# QUI SOMMES-NOUS ?

## CR2PA

Le CR2PA, club de l'archivage managérial, est une association regroupant une quarantaine de membres issus d'organismes publics et du monde de l'entreprise.

Indépendant des acteurs du marché, le CR2PA est un lieu d'échange et de partage entre pairs du métier de l'archivage.



## FnTC

Créée en 2001, la Fédération des Tiers de Confiance du Numérique opère avec pertinence la fusion de la technologie avec le droit et le « chiffre », et ses membres offrent au marché du Numérique un inestimable gisement de compétences dans les domaines historiques de la digitalisation : signature électronique, archivage électronique, identité numérique, facture électronique, vote électronique, e-finance, e-santé, ... Mais également dans ses domaines montants : Blockchain, KYC, Cachet électronique visible (CEV)...



## Nous contacter

### CR2PA

75 rue de Lourmel  
75015 PARIS  
contact@cr2pa.fr

### FnTC

5, impasse Gomboust  
75001 PARIS  
infos@fn-tc-numerique.com

# Sommaire



Les enjeux.....	3
Glossaire.....	4
Cadre réglementaire.....	5
Types de solution d'archivage des données.....	6
Arbre de solutions.....	10
Cas d'usage :	
○ Décommissionnement de SI.....	11
○ Comptabilité informatisée.....	12
○ Projet d'ingénierie en mode collaboratif.....	13
○ Données conversationnelles.....	14
○ Archivage d'un logiciel.....	15
○ Mouvements bancaires.....	16

# Les enjeux

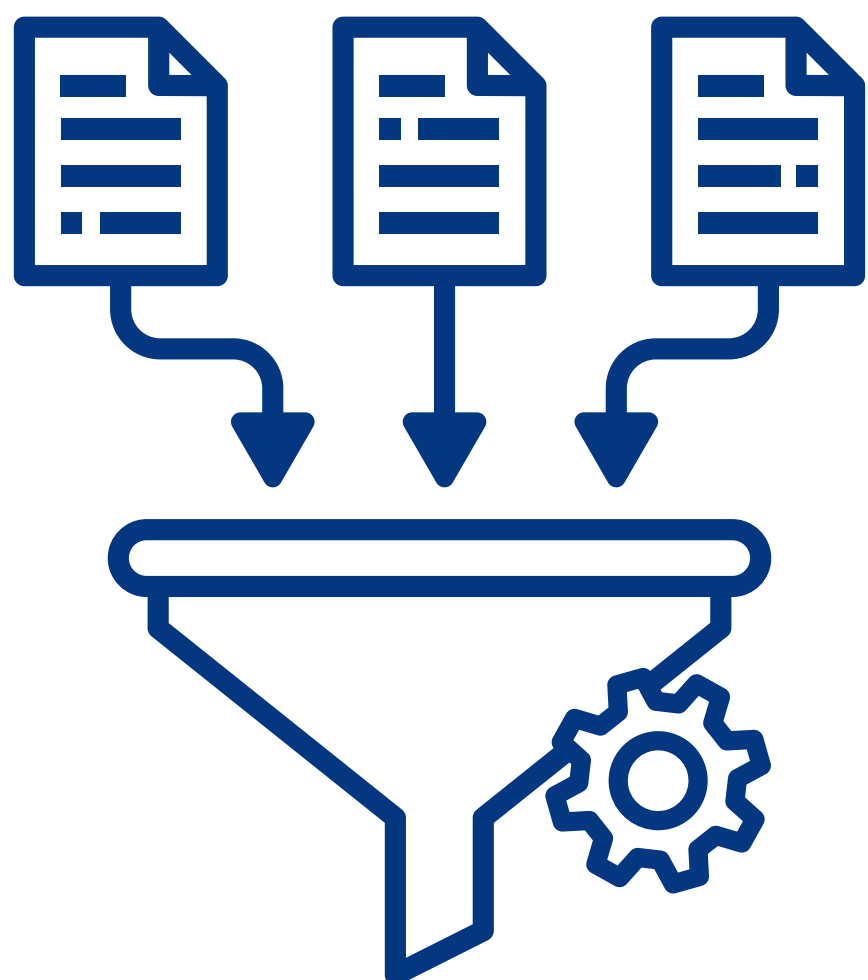
Dans le monde de l'archivage, l'archivage des documents électroniques a occupé le devant de la scène depuis une bonne quinzaine d'années. Un autre thème arrive maintenant en force : c'est celui de l'archivage des données.

Il se distingue de l'archivage des documents par plusieurs facteurs de complexité :

- on accède aux données par des interfaces de traitement : conserver les données ne suffit pas, il faut également conserver des éléments d'environnement qui permettent d'y accéder.
- les formats sont extrêmement diversifiés, et leur pérennité n'est souvent pas garantie : les montées de version sont multiples, et les formats propriétaires sont tributaires de la pérennité des éditeurs.
- les volumétries ne font que croître.

Ce sujet n'est pas totalement nouveau, certaines entreprises y sont confrontées et y apportent des réponses depuis longtemps. Mais il tend à se généraliser sous plusieurs effets :

- de plus en plus de processus sont traités par des applications automatisées qui ne s'appuient plus sur des documents, ou gèrent des données (par exemple de géométrie 3D) qui ne sont pas réductibles à des documents.
- face à des parcs applicatifs sans cesse croissants, les directions informatiques se fixent pour objectif de décommissionner les applications arrivées en fin de vie, avec la question du devenir des données qu'elles contiennent.
- le mythe des capacités de stockage infinies a fait son temps, confronté à la réalité des coûts et à la préoccupation de l'empreinte carbone des activités des entreprises.



Dans l'archivage, les solutions doivent être dimensionnées en fonction du besoin et de l'analyse de risques :

- pour quel usage doit-on conserver ? S'il s'agit uniquement d'apporter des preuves des activités, produire et conserver des documents en entrée et sortie des traitements peut être un élément de réponse ; il en va autrement s'il est envisagé de réutiliser les données dans de nouveaux traitements.
- quelle est la durée de conservation requise ? Si cette durée n'est pas supérieure à la durée de vie des applications, conserver les données dans l'application d'origine, avec des conditions robustes de sécurisation, peut être une solution.
- quel est l'enjeu pour l'entreprise ? Dans des cas extrêmes, l'entreprise pourrait considérer que le coût de l'archivage est supérieur au risque encouru, et décider de ne pas engager de moyens lourds de conservation.

A l'inverse des exemples donnés ci-dessus, nombre de cas bien entendu nécessitent la mise en œuvre de solutions d'archivage élaborées.

Comme toujours également dans le domaine de l'archivage, le besoin de définition et d'obtention des métadonnées permettant de qualifier et d'exploiter les éléments archivés a au moins autant d'importance que les conditions de conservation elles-mêmes.

Face à toutes ces questions, le CR2PA et la FnTC ont souhaité proposer une méthodologie et présenter des cas d'usage sur ce sujet.

Il est bien entendu que ce panorama ne prétend pas couvrir l'exhaustivité des cas, et qu'il ne photographie qu'un point de passage dans un domaine encore largement en cours d'évolution.

L'objectif de ce document est d'aider les organisations à définir une stratégie de préservation des données, sans rentrer dans le détail d'analyses technologiques et réglementaire pour lesquelles d'autres sources plus qualifiées existent par ailleurs. Nous avons choisi de nous concentrer sur des cas d'usage représentatifs des situations concrètes rencontrées par les entreprises et de mettre en regard des typologies de solutions.



# Glossaire

- Décommissionnement : Fin d'utilisation d'un équipement, d'une application.
- Données : Représentation d'une information permettant sa communication, son interprétation, son traitement.
  - Données conversationnelles : Différents types d'échanges sur des médias sociaux, des modules de discussion internes, des forums...
  - Données transactionnelles : Tout ce qui est relatif à une transaction comme un achat, une demande d'un devis ou encore une prise de rendez-vous. D'un point de vue technique, cela va engendrer un enregistrement, en base de données, qu'il convient de conserver dans l'historique client de manière claire et précise ;
  - Données complexes : La complexité peut se rattacher au format de fichiers ou aux modèles de données ;
  - Données structurées et non structurées : Les données peuvent se présenter sous forme structurée (par exemple: les bases de données relationnelles) ou sous forme non structurée (par exemple: courriers électroniques, documents texte, images, enregistrements audio et fichiers).
- Empreinte numérique (ou « hash value » en anglais) : Résultat d'une fonction de calcul basée sur un algorithme cryptographique (fonction de hachage), opérée sur un objet numérique. Toute modification de l'objet numérique entraînera une empreinte différente qui révélera la modification par comparaison avec la première empreinte.
- Emulation: Imitation du comportement physique d'un matériel et de son système d'exploitation par un logiciel.
- Intelligence artificielle (communément appelée IA); Pour le Parlement européen, l'intelligence artificielle représente tout outil utilisé par une machine afin de « reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité »
  - Intelligence artificielle générative : L'intelligence artificielle générative, segment de l'intelligence artificielle, vise à produire de nouvelles données qui ressemblent à celles créées par des êtres humains, que ce soit sous forme de texte, d'images ou encore de musique;
  - Learning : L'intelligence artificielle inclut plusieurs dimensions, dont le machine learning (ou apprentissage automatique qui consiste, pour l'ordinateur, à apprendre depuis des données) et le deep learning (ou apprentissage profond, avec un ordinateur utilisant l'abstraction dans son analyse des données). Source : Afnor.
- Métadonnées : Informations structurées ou semi-structurées permettant les traitements des documents numériques (norme Z42-013)
  - Métadonnées descriptives : Informations décrivant la donnée en ajoutant certains détails comme l'auteur, le titre de la donnée, la date de création, etc ;
  - Métadonnées de gestion : Informations concernant les attributs nécessaires à la gestion du cycle de vie des archives ainsi que les références aux documents spécifiant les caractéristiques appliquées à l'archive (politique d'archive, règle de conservation, échéance d'élimination) ;
  - Métadonnées techniques : Informations concernant l'identification de l'objet numérique, son origine, son format, sa taille, sa datation, les éléments relatifs à son intégrité ainsi que les éléments relatifs au processus en amont du versement (numérisation...).
- Modèle de données : Un modèle de données est un modèle qui décrit la manière dont sont représentées les données dans une organisation métier, un système d'information ou une base de données.



# Cadre réglementaire

Les deux règlements européens cités ci-dessous constituent l'épine dorsale du marché unique des données en Europe mais ne sauraient limiter l'analyse. D'autres lois et règlements devront nécessairement être étudiés de manière générale ( intelligence artificielle, base de données...) comme dans des domaines sectoriels (la santé, le domaine du paiement, les données à caractère personnel...).

Nota Bene : la notion de service d'archivage électronique au sens du Règlement eIDAS2 renvoie, quant à elle, aux documents et aux données archivées. Ce point fait l'objet d'autres fascicules.

**1.** Règlement n° 2023/2854 du 13 décembre 2023 du Parlement européen et du Conseil fixant des règles harmonisées pour l'équité de l'accès aux données et de l'utilisation des données (Data Act)- JOUE L 22 décembre 2023, p. 1.

Applicable : depuis le 12 septembre 2025.

Objectif : renforcer l'accès et la circulation des données pour les utilisateurs et les organismes du secteur public.  
Public visé : les fabricants et les fournisseurs des objets et des services connectés, les détenteurs de données qui les mettent à disposition des destinataires de l'UE, les destinataires desdites données, les entités du secteur public ou de l'UE dont la demande à l'accès des données est motivée par un besoin exceptionnel, les prestataires de services de traitement de données opérant pour une clientèle présente dans l'UE.

Axes :

- ➡ Droit d'accès et au partage des données pour l'utilisateur générant les données, les tiers et les destinataires des données ;
- ➡ Accès aux données par les entités du secteur public ou les entités européennes ;
- ➡ Promotion de la libre concurrence et de l'interopérabilité sur le marché des données.

Quel impact pour l'archivage ?

Art. 36 sur les Exigences essentielles concernant les contrats intelligents pour le partage de données :  
«c) *archivage et continuité des données, pour garantir, dans les circonstances dans lesquelles un contrat intelligent doit être résilié ou désactivé, qu'il y a la possibilité d'archiver les données relatives aux transactions, la logique et le code du contrat intelligent afin de conserver l'enregistrement des opérations effectuées sur les données dans le passé (vérifiabilité);* »

**2.** Règlement (UE) 2022/868 du Parlement européen et du Conseil du 30 mai 2022 portant sur la gouvernance européenne des données et modifiant le règlement (UE) 2018/1724 (règlement sur la gouvernance des données ou "DGA")- JOUE L 152, 3 juin 2022, p. 1 et s.s.

Applicable : depuis le 24 septembre 2023.

Objectif : faciliter la circulation de « données protégées par les droits d'autrui ». (secret des affaires, confidentialité, protection de la propriété intellectuelle...)

Public visé : organismes du secteur public (ex : les archives) et entités qui souhaitent se livrer à l'activité d'intermédiaire dans le partage de données, entités qui souhaitent réutiliser des données détenues par des organismes publics.

Axes :

- ➡ Réutilisation de « données sensibles » détenues par des organismes du secteur public ;
- ➡ Régime des « services de partage des données » ;
- ➡ Mise en place de l'altruisme dans le partage des données.

Quel impact pour l'archivage ?

Il ne parle pas d'archivage mais il peut s'appliquer à des données archivées. L'art. 5 sur les conditions de réutilisation met en avant les exigences de préservation de l'intégrité de la donnée. En cas de conservation et de réutilisation, il faut s'assurer du respect des conditions de réutilisation.



# Types de solution d'archivage des données

Archiver des données peut répondre à différents besoins, notamment de décommissionnement d'un système d'information s'appuyant sur des bases de données, de conservation des données transactionnelles, des résultats de calculs scientifiques, des enregistrements industriels...

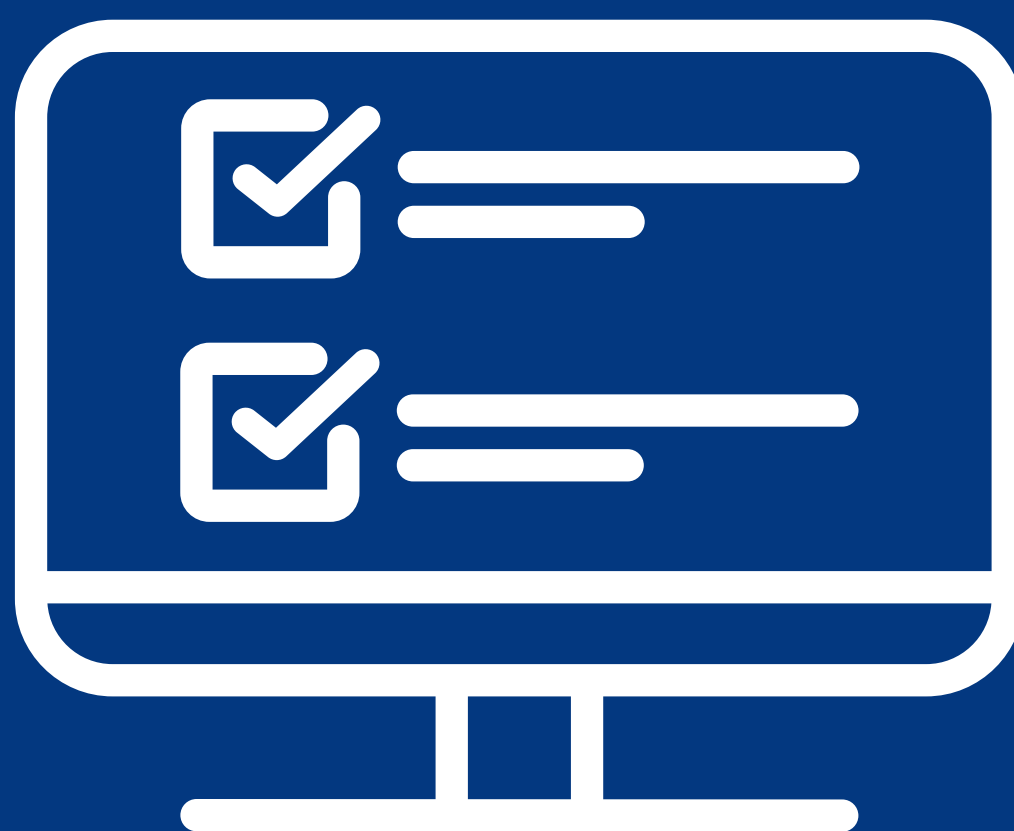
Les solutions à mettre en œuvre pour chaque source de données à archiver dépendront principalement :

- ➔ de la volumétrie : à la différence de l'archivage de documents, l'archivage de données peut rapidement être très volumétrique, notamment s'il est nécessaire de permettre une exploitation unitaire de chaque donnée ;
  - Exemple : Archiver une table d'un milliard d'enregistrements peut être réalisé par l'archivage d'un seul objet « table » ou d'un milliard d'objets « enregistrement » selon le niveau d'accès à la donnée souhaité.
- ➔ des usages : selon s'il s'agit de préserver la donnée « au cas où » ou si l'accès à la donnée doit rester disponible pour les utilisateurs, indépendamment du système qui l'a produite ;
- ➔ de la sécurité : notamment si la donnée doit être conservée dans des conditions propres à préserver son intégrité et que cela doit être démontrable en cas de besoin ;
- ➔ des formats : selon s'il est possible de convertir le format de donnée source vers un format adapté à l'archivage ou s'il est nécessaire de préserver des formats natifs propriétaires et de maintenir les moyens de les exploiter.

Le coût et la complexité de l'archivage dépendront de ces paramètres. Ce chapitre propose trois stratégies et des méthodes d'implémentation représentatives :

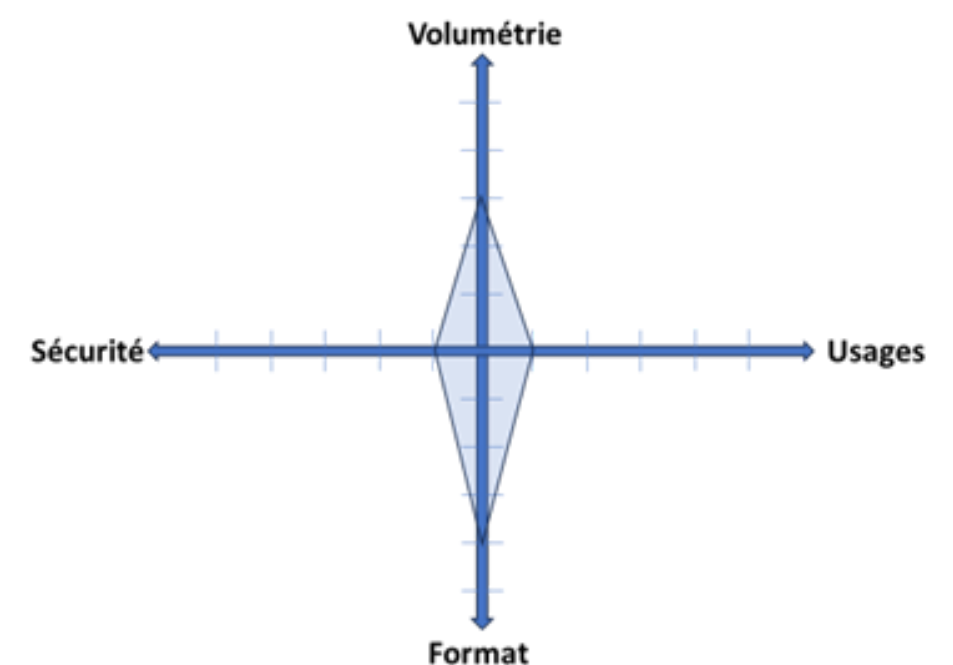
- ➔ Stratégie 1 : avoir la capacité de reconstruire un système permettant d'exploiter les données
  - Adapté pour l'archivage de modèles de données complexes, principalement pour un besoin de décommissionnement d'applicatif et pour un usage faible durant l'archivage, ne nécessitant pas nécessairement de gestion de la valeur probante ;
- ➔ Stratégie 2 : archiver des états directement exploitables par un système d'archivage
  - Adapté pour l'archivage de modèles de données simples nécessitant une gestion de la valeur probante et pour un accès courant à la donnée durant l'archivage ;
- ➔ Stratégie 3 : archivage ou sécurisation des données par bloc
  - Adapté pour l'archivage de données transactionnelles ou sous forme de logs techniques, nécessitant une gestion de la valeur probante sur de très hautes volumétries.

Chaque stratégie est déclinée en différentes méthodes, présentant chacune des avantages et des contraintes.



# Stratégie 1 : Avoir la capacité de reconstruire un système permettant d'exploiter les données

Cette stratégie répond à un besoin de conserver les données pouvant être volumétrique, mais sans usage courant pendant l'archivage. Il sera possible d'investir dans la visualisation des données a posteriori, sans contrainte forte de délai et uniquement si le besoin s'avère. Il n'y a pas de besoin de valeur probante (le SI d'origine ne la garantissait généralement pas). Le format est souvent complexe, lié au modèle de données du SI d'origine.



	Méthode	Avantage	Contrainte
#1 ; Emulation du SI d'origine	<ul style="list-style-type: none"> <li>Conserver l'application, le moteur de base de données et les données</li> <li>En cas de besoin, émuler l'application d'origine pour consultation des données</li> </ul>	<ul style="list-style-type: none"> <li>Relativement simple à mettre en oeuvre</li> <li>Garantit la fidélité de restitution de la donnée par le SI d'origine</li> <li>Pas de charge de migration</li> </ul>	<ul style="list-style-type: none"> <li>Dépendance aux technologies d'origine</li> <li>Grosse charge de reconstruction potentielle</li> <li>Risque de ne pas réussir une réinstallation dans les contextes technologiques d'avenir</li> </ul>
#2 : Développer a posteriori en cas de besoin	<ul style="list-style-type: none"> <li>Documenter le modèle de données</li> <li>Migrer chaque table ou source de données dans un format normé exploitable (fichier à plat ou structuré xml, JSON...)</li> <li>En cas de besoin, développer le modèle de données dans la technologie disponible dans l'avenir pour accéder aux données.</li> </ul>	<ul style="list-style-type: none"> <li>Limite la dépendance au SI et technologies d'origine</li> <li>Charge de migration faible</li> </ul>	<ul style="list-style-type: none"> <li>Grosse charge de reconstruction potentielle</li> <li>Risque sur la fiabilité de la donnée reconstituée</li> <li>Risque d'altération de la donnée source sur les formats complexes au moment de l'export</li> </ul>
#3 : Migrer en structure d'archivage (SIARD)	<ul style="list-style-type: none"> <li>Documenter le modèle de données</li> <li>Reproduire le modèle de données dans une base dédiée à la préservation (par exemple SIARD)</li> <li>Migrer les données</li> </ul>	<ul style="list-style-type: none"> <li>Limite la dépendance au SI et technologies d'origine</li> <li>Permet un accès aux données sans le SI d'origine</li> <li>Permet un accès rapide aux données sans reconstruction</li> </ul>	<ul style="list-style-type: none"> <li>Projet complexe à mettre en oeuvre</li> <li>Charge de migration</li> <li>Nécessite une maintenance</li> <li>Solution peu ouverte aux utilisateurs (complexe à utiliser)</li> </ul>

Les méthodes #1 et #2 nécessitent une reconstruction pour accéder à la donnée. Elles sont adaptées à un usage pour des données qui ont un intérêt potentiel. Il s'agit d'investir dans une reconstruction uniquement si le besoin s'avère.

La méthode #1 est plus économique que la méthode #2 (pas de besoin de redévelopper un SI pour reconstruire) mais présente un risque sur la pérennité : serons-nous réellement en mesure d'émuler d'anciennes technologies dans les technologies du futur ?

La méthode #3 nécessite un investissement au moment de l'archivage similaire à la méthode #2 en présentant l'avantage de permettre un accès à la donnée en cas de besoin sans investissement a posteriori.

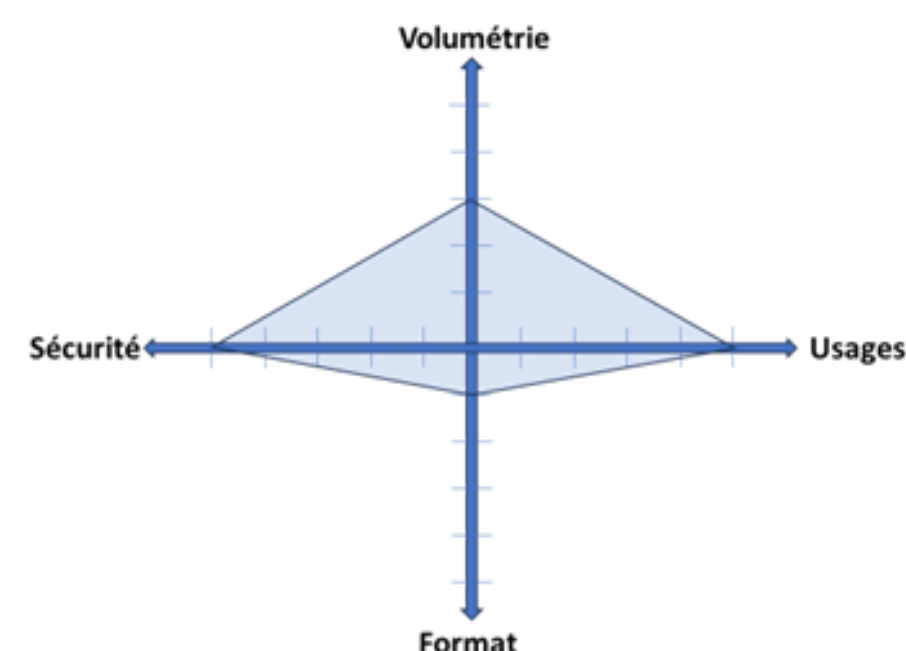
Les méthodes #1 et #2 peuvent être réalisées selon une simple sauvegarde ou stockage ou par versement dans un SAE/CFN, de manière à garantir la non-perte d'intégrité durant la conservation.



# Stratégie 2 : Archiver des états directement exploitables par un système d'archivage

Les données brutes sont rarement exploitables directement par l'humain. Ceci nécessite des systèmes de visualisation sur l'ensemble des données ce qui fait l'objet de la stratégie 1.

La seconde stratégie consiste à convertir des groupes de données en documents de manière à s'affranchir de la problématique de visualisation et de pérennisation des formats qui est nativement gérée par les SAE, comme c'est le cas pour la valeur probante.



	Méthode	Avantage	Contrainte
#1 : Archivage en documents intelligibles PDF	<ul style="list-style-type: none"> <li>Depuis le SI source ou les accès directs en base, produire des états cohérents permettant d'agréger les différentes données</li> <li>Mettre les données en forme intelligible</li> <li>Editer des documents PDF normés associés à leurs métadonnées</li> <li>Verser en archivage</li> </ul>	<ul style="list-style-type: none"> <li>La donnée est directement accessible et intelligible pour l'utilisateur</li> <li>Gain de volumétrie par rapport à du stockage unitaire de données</li> </ul>	<ul style="list-style-type: none"> <li>Investissement de migration pour la production des documents</li> <li>Données de contenu non exploitables</li> </ul>
#2 : Structuration de groupes de données en langage structuré (format à contenu balisé)	<ul style="list-style-type: none"> <li>Idem #1 mais structurer les données dans un langage structuré (XML, json, html...) et normé</li> <li>Paramétrer des feuilles de style pour représenter les données de manière intelligibles à l'écran ou des viewers ad'hoc permettant de représenter l'information</li> </ul>	<ul style="list-style-type: none"> <li>Idem #1</li> <li>Possibilité de documents plus interactifs à l'affichage, notamment en cas de documents volumineux</li> </ul>	<ul style="list-style-type: none"> <li>Investissement de migration pour la production des documents (supérieur au #1)</li> <li>Possibilité de recomposer des bases de données à partir des documents en langage structurés</li> <li>Possibilité d'indexation et de requête par les moteurs de recherches tiers (elastic, ...)</li> </ul>
#3 : Archivage d'états consolidés	<ul style="list-style-type: none"> <li>Analyser les besoins d'exploitation future des données, notamment pour traitement statistique</li> <li>Produire sur requête des états consolidés (sommés, moyennes, ...) de groupes de données cohérentes par rapport à un besoin métier</li> </ul>	<ul style="list-style-type: none"> <li>Limite la taille et le nombre de documents</li> </ul>	<ul style="list-style-type: none"> <li>Impossibilité de reconstruire un gisement de données à partir des données sources</li> </ul>

Les trois méthodes nécessitent un investissement dans une opération de migration des données sources vers un format d'archivage.

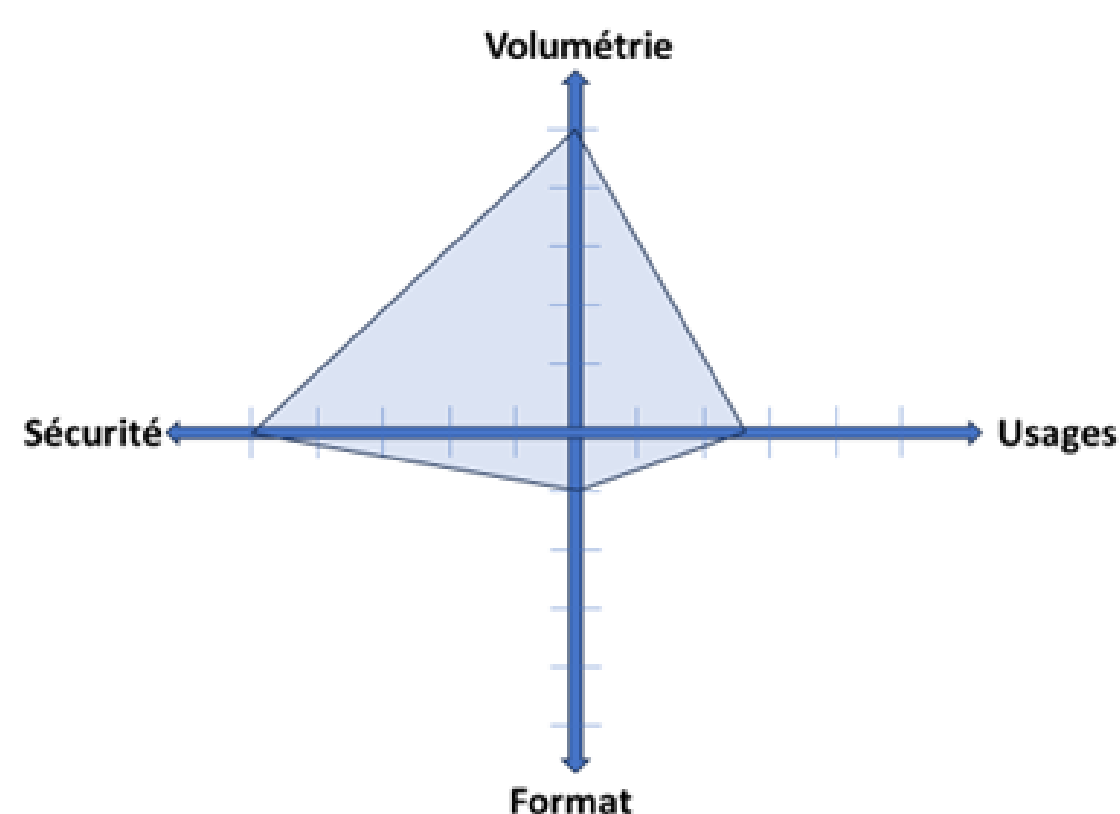
La méthode #2 nécessite plus d'effort de structuration que la méthode #1 mais présente l'avantage de pouvoir reconstruire des modèles de données a posteriori, ou des gisements indexables et exploitables par des technologies big data.

La méthode #3 est plus sensible car elle nécessite une bonne anticipation des usages futurs de la donnée. Elle est particulièrement adaptée au traitement de cas à très haute volumétrie qui ne sont exploités que pour réaliser des études statistiques (données de calcul, enregistrements scientifiques ou industriels...)

# Stratégie 3 : Archivage ou sécurisation des données par bloc

Dans un contexte de très haute volumétrie de données relativement simples et homogènes (logs techniques, transactions, données de traçabilité) le besoin de visualisation est généralement faible et le besoin de sécurité important.

La contrainte de ce cas d'usage est le coût d'archivage de très haute volumétrie. La stratégie 3 consiste à limiter la volumétrie d'objets archivés ou sécuriser les enregistrements conservés dans le gisement source.



	Méthode	Avantage	Contrainte
#1 : Archivage en blocs	<ul style="list-style-type: none"> <li>Regrouper un ensemble d'enregistrements dans une enveloppe</li> <li>Indexer le bloc</li> <li>Verser en archivage</li> <li>Implémenter une fonction permettant de réaliser des sous-recherches dans les blocs et pour visualiser les contenus</li> </ul>	<ul style="list-style-type: none"> <li>Limite le nombre d'objets archivés</li> <li>Gains sur les tailles d'objet du fait de la compression</li> <li>La sécurisation du bloc induit que l'ensemble des objets contenus restent intègres</li> <li>Permet de recomposer l'ensemble du gisement par réassemblage des blocs</li> <li>Économie financière</li> </ul>	<ul style="list-style-type: none"> <li>L'accès nécessite deux étapes de recherche</li> </ul>
#2 : Sécurisation de gisements	<ul style="list-style-type: none"> <li>Gérer la rétention et le cycle de vie des données sources dans le système de gestion opérationnel (data-lake, SIEM, ...)</li> <li>Calculer une empreinte numérique de chaque enregistrement</li> <li>Condenser les empreintes numériques d'un ensemble d'enregistrements dans un arbre de Merkle.</li> <li>Produire un enregistrement du résultat de la sécurisation dans une blockchain, archiver le résultat dans un SAE/CFN ou appliquer un horodatage tiers</li> </ul>	<ul style="list-style-type: none"> <li>Les données restent exploitables dans une technologie adaptée au big data</li> <li>L'intégrité du gisement est préservée</li> </ul>	<ul style="list-style-type: none"> <li>Le système source doit être correctement sécurisé et notamment interdire la modification de données.</li> </ul>

La méthode #1 permet d'utiliser des technologies de SAE existante en évitant de les saturer mais pose une limite d'exploitation. Le besoin est d'accéder à des données unitaires contenues dans un bloc.

La méthode #2 permet de conserver les capacités de traitement sur l'ensemble du gisement tout en bénéficiant de la sécurité apportée par la journalisation tierce.

# Arbre de solutions

	Modèle de données complexe	Conservation long terme	Valeur probante	Consultation fréquente	Réexploitation des jeux de données	Très grande volumétrie
S1 #1 : Emulation du SI d'origine	X	X			X	
S1 #2 : Développer a posteriori en cas de besoin	X	X			X	
S1 #3 : Migrer en structure d'archivage (SIARD)	X	X		X	X	
S2 #1 : Archivage en documents intelligibles pdf		X	X	X		
S2 #2 : Structuration de groupes de données en langage structuré		X	X	X	X	
S2 #3 : Archivage d'états consolidés		X	X	X		
S3#1 : Archivage en blocs		X	X	X	X	X
S3#2 : Sécurisation de gisements			X	X	X	X





# Cas d'usage

## 1. Décommissionnement de SI

Finalité : selon les cas, protection juridique, capitalisation du savoir, réutilisation opérationnelle des données		
<b>Contexte</b> Suite au décommissionnement d'une application, il est nécessaire de conserver les données qui ne seront pas migrées dans un nouveau SI.	<b>Périmètre et prérequis</b> Disposer du modèle de données et de formats d'exports exploitables.	<b>Intervenants</b> <ul style="list-style-type: none"><li>Utilisateurs métier de la donnée</li><li>Archivistes</li><li>DSI</li></ul>
<b>Niveau d'archivage</b> Le niveau d'archivage dépend : <ul style="list-style-type: none"><li>De la dimension légale de la donnée (probante ou non).</li><li>De sa dimension patrimoniale</li><li>De la durée de conservation</li></ul>		<b>Réglementation</b> <ul style="list-style-type: none"><li>Preuve (Code civil, code de commerce...) : déterminer si les données à migrer doivent être conservées comme indices, éléments de preuve (dans le cas où la preuve est libre).</li><li>RGPD : Identifier le besoin de conserver des données personnelles ou non durant l'archivage. Si oui, typer les données personnelles pour anonymisation ultérieure ou élimination</li></ul>
<b>Démarche préconisée</b> Identifier : <ul style="list-style-type: none"><li>Les besoins d'accès à donnée migrée durant l'archivage</li><li>La dimension légale des données : nécessite un archivage probant ou non</li></ul> Autant que possible, s'affranchir d'avoir à conserver le système producteur pour exploiter la donnée.	<b>Solutions envisageables</b> <ul style="list-style-type: none"><li>Cas 1 : Données à préserver sans besoin d'exploitation sauf cas exceptionnel :<ul style="list-style-type: none"><li>Archiver l'ensemble des tables</li><li>Documenter le modèle de données</li><li>En cas de besoin, un investissement sera nécessaire pour « remonter » les données dans une base</li></ul></li><li>Cas 2 :<ul style="list-style-type: none"><li>Produire des exports consolidés et anonymisés dans un format technique (JSON, XML)</li><li>Archiver par blocs d'états</li><li>Indexer les blocs</li><li>Rechercher dans les blocs ou chargement des blocs pour traitements statistiques (indexation elastic par exemple)</li></ul></li><li>Cas 3 : Données nécessitant des accès unitaires au enregistrements<ul style="list-style-type: none"><li>Produire des états unitaires sur un format documentaire intelligible</li><li>Archiver en tant que document</li></ul></li></ul>	
<b>Complexité</b> Dépend de la complexité du modèle de données. A partir d'un certain stade, il peut s'avérer nécessaire de conserver les exécutable des systèmes producteurs et les archiver.	<b>Etapas</b> <ul style="list-style-type: none"><li>Evaluer (préconisation)</li><li>Décider de la stratégie (recommandation)</li><li>Choisir la solution d'archivage ou autre</li><li>Réaliser la migration</li></ul>	

## 2. Comptabilité informatisée

<p><b>Finalité:</b> Répondre aux obligations fiscales Durée de conversation : 10 ans</p>		
<p><b>Contexte</b> Vérification sur place par les agents de l'administration des impôts de la comptabilité des contribuables astreints à tenir et à présenter des documents comptables (<a href="#">Article L13 du Livre des procédures fiscales</a>).</p>	<p><b>Périmètre et prérequis</b> Concerne l'ensemble des informations, données et traitements informatiques qui concourent directement ou indirectement à la formation des résultats comptables ou fiscaux et à l'élaboration des déclarations ainsi que la documentation relative aux analyses, à la programmation et à l'exécution des traitements (<a href="#">Article L13 du Livre des procédures fiscales</a>).</p>	<p><b>Intervenants externes</b></p> <ul style="list-style-type: none"> <li>• BVCI (Brigade de vérification des comptabilités informatisées)</li> </ul> <p><b>Intervenants internes</b></p> <ul style="list-style-type: none"> <li>• Service fiscalité</li> <li>• Services producteurs des données</li> <li>• DSI</li> <li>• Archivistes</li> <li>• Services juridiques</li> </ul>
<p><b>Niveau d'archivage</b> L'archivage doit donner une date fiable aux archives pour une période utile, permettre de démontrer leur intégrité, de tracer les éventuelles modifications et de garantir la sécurité du dispositif.</p> <p>Les formats d'archivage doivent permettre l'exploitation des archives indépendamment du système qui les a générées. Les formats plats du type TXT ou CSV sont cités à titre d'exemple dans <a href="#">la section II.A.2 c1550 du BOI-BIC-DECLA-30-10-20-40</a>.</p>		<p><b>Réglementation</b> <a href="#">Section II-A du BOI-BIC-DECLA-30-10-20-40 BIC - Obligations déclaratives - Obligations fiscales et comptables dans le cadre du régime du réel normal d'imposition - Conservation et représentation des livres, documents et pièces comptables dans le cadre d'une comptabilité informatisée</a></p>
<p><b>Démarche préconisée</b> Recenser les applications traitant de données concourant directement ou indirectement à la constitution des résultats comptables ou fiscaux.</p> <p>Un lot « archivage fiscal » est à prévoir dans tous les projets d'outils de cette catégorie.</p> <p>Si on fait appel à une plate-forme de dématérialisation partenaire (PDP), ne pas oublier de s'assurer de garder sous contrôle l'archivage sur toute la durée nécessaire.</p>		<p><b>Solutions envisageables</b></p> <ul style="list-style-type: none"> <li>• Privilégier un archivage centralisé dans un SAE pour garantir la valeur probatoire des données ;</li> <li>• Exporter les données des grands livres/journaux comptables sous format document.</li> </ul> <p>Ne pas oublier l'obligation de conserver les versions antérieures des outils informatiques si cela est nécessaire à la compréhension des traitements formant les résultats, et leurs documentations.</p>
<p><b>Complexité</b> Le nombre important des systèmes techniques concourant à la création du chiffre d'affaires.</p> <p>Le besoin de conserver non seulement la donnée élémentaire mais également les traitements qui lui sont appliqués et la documentation permettant de comprendre les données et leurs traitements.</p>		<p><b>Etapes</b></p> <ul style="list-style-type: none"> <li>• Identifier les documents/données à archiver.</li> <li>• Définir les formats d'archivage et les métadonnées nécessaires.</li> <li>• Industrialisation du versement et consultation dans le SAE.</li> </ul>

## 3. Projet d'ingénierie en mode collaboratif

<p><b>Finalité :</b> Pour des raisons légales ou de protection juridique. Pour réutilisation.</p>		
<p><b>Contexte</b> Sur une plate-forme collaborative fournie par le client :</p> <ul style="list-style-type: none"> <li>• les données d'entrée sont fournies par le client</li> <li>• les différents acteurs avec différents rôles (maître d'oeuvre, constructeur, fournisseur de données) doivent déposer des données et travailler en commun.</li> </ul> <p>A la fin du projet, les données à conserver par les différents acteurs doivent être extraites et archivées.</p> <p>A noter : l'existence de la plate-forme est liée à la durée de vie du projet et non pas à la durée de vie de l'ouvrage</p>	<p><b>Périmètre et prérequis</b> Définir les rôles et les attendus de chacun</p> <p>2 échelles :</p> <ul style="list-style-type: none"> <li>• vis-à-vis du projet ;</li> <li>• pour sa propre entreprise.</li> </ul> <p>Les projets en mode collaboratif restent des projets de chacune des entreprises impliquées, il faut continuer d'appliquer les stratégies d'archivage appliquées dans son entreprise Chacun en fonction de son rôle aura des exigences différentes.</p>	<p><b>Intervenants</b></p> <ul style="list-style-type: none"> <li>• Chefs de projets</li> <li>• Ingénieurs</li> <li>• Archivistes</li> <li>• Doc controller</li> </ul>
<p><b>Niveau d'archivage</b> L'archivage doit permettre :</p> <ul style="list-style-type: none"> <li>• De conserver l'image des livrables ;</li> <li>• De conserver les échanges et les engagements de chacun des participants ;</li> <li>• De réutiliser les données pour réinitialiser des projets ultérieurs.</li> </ul> <p>Les données à archiver sont distribuées vers :</p> <ul style="list-style-type: none"> <li>• Le client qui archive l'ensemble du projet ;</li> <li>• Les parties prenantes qui archivent pour leur propre compte les éléments les concernant.</li> </ul>	<p><b>Réglementation</b> Preuve (Code civil, code de commerce...) : déterminer si les données à migrer doivent être conservées comme indices, éléments de preuve (dans le cas où la preuve est libre).</p> <p>Dépend du secteur d'activités du projet. Pas de réglementation spécifique liée à l'archivage, mais il existe des normes encadrant le métier, comme pour le BIM la norme ISO 19650.</p>	
<p><b>Démarche préconisée</b> En début de projet, bien inclure une clause archivage et les conditions de réversibilité (format, périodicité) et préciser qui est le propriétaire de quoi avec les droits d'accès et de réutilisation.</p> <p>Chaque partenaire du projet doit définir son besoin d'archivage (valeur probante).</p> <p>Sélectionner les métadonnées/attributs à associer.</p> <p>Recourir à une analyse de risques.</p>	<p><b>Solutions envisageables</b> Pour les besoins probatoires :</p> <ul style="list-style-type: none"> <li>• Les informations sont extraites sous forme de documents (plans, notes, etc...);</li> <li>• Les messages d'échanges (type BFC) enregistrés dans la plate-forme sont extraits avec leur date et le lien avec l'élément concerné de l'ouvrage.</li> </ul> <p>Ne pas oublier d'archiver les mails.</p> <p>Pour les besoins de réutilisation de données de modélisation, soit :</p> <ul style="list-style-type: none"> <li>• Récupération des données sous format interopérable, notamment les IFC, les STEP ;</li> <li>• Récupération et maintien du format propriétaire, avec une veille sur les changements de version et le maintien de la compatibilité ascendante, et une migration vers les nouvelles versions si nécessaire au risque de perdre les moyens d'exploiter les données sur le long terme.</li> </ul>	
<p><b>Complexité</b> Connaitre les droits de chaque jeu de données. Connaitre la durée des accès.</p> <p>La durée d'archivage dépasse la durée du projet et la durée des outils utilisés pendant le projet.</p> <p>Chaque acteur doit prendre ses responsabilités quant à l'archivage.</p>	<p><b>Etapes</b></p> <ul style="list-style-type: none"> <li>• Identifier les documents/données à archiver</li> <li>• Identifier les données à conserver pour réutilisation</li> <li>• Mettre en œuvre l'extraction dans les formats retenus</li> <li>• Choisir le bon outil</li> </ul>	



## 4. Données conversationnelles

<b>Finalité:</b> Pour enjeux juridiques, réglementaires et de conformité		
<b>Contexte</b> Activités de marché : système d'échanges où se rencontrent l'offre (vendeurs) et la demande (les acheteurs).	<b>Périmètre et prérequis</b> Applications de messagerie entre entreprises et clients Communication écrite (messagerie instantanée / boîte mail et communication vocale)	<b>Intervenants</b> <ul style="list-style-type: none"> <li>• Equipe conformité</li> <li>• Equipe informatique</li> <li>• Data privacy officer</li> <li>• Prestataires d'archivage spécifiques pour les données conversationnelles.</li> </ul>
<b>Niveau d'archivage</b> L'archivage des communications écrites et téléphoniques doit permettre de : <ul style="list-style-type: none"> <li>• Répondre aux requêtes externes des autorités ;</li> <li>• Gérer les plaintes des clients ;</li> <li>• Assurer le respect des règles internes de l'entreprise et des réglementations des marchés ;</li> <li>• Gérer les contrôles de conformité.</li> </ul>		<b>Réglementation</b> Attention au RGPD  Règlement AMF <ul style="list-style-type: none"> <li>• <a href="#">Article 313-51</a></li> <li>• <a href="#">Article 313-52</a></li> <li>• <a href="#">Article 321-73</a></li> </ul> <p>Les conversations téléphoniques sont conservées pour une durée maximum de 6 ans.</p> <p>Les conversations téléphoniques mobiles sont conservées pour une durée maximum de 5 ans.</p> <p>Les communications électroniques sont conservées pour une durée maximum de 5 ans.</p>
<b>Démarche préconisée</b> Passer par un prestataire dédié et spécialisé dans ce type de solutions.		<b>Solutions envisageables</b> Archivage en continu dans une coffre-fort numérique ou un SAE, hébergé sur machines dédiées, en interne ou par un partenaire spécifique. Transformer les fils de conversation en documents et les enregistrer dans un format exploitable et visualisable.
<b>Complexité</b> Automatisé Volumétrie énorme		<b>Etapes</b> <ul style="list-style-type: none"> <li>• Sensibiliser les traders</li> <li>• Contrat avec le prestataire</li> </ul>

## 5. Archivage d'un logiciel

<p><b>Finalité:</b> Répondre aux obligations de garantie et de fiabilité du produit. Protéger le constructeur en cas de contentieux. Appliquer des corrections en cas de défaut (réutilisation des données). Durée : durée d'utilisation de l'équipement (peut être très longue).</p>		
<p><b>Contexte</b> Suite à la livraison d'un logiciel à un client, il est nécessaire de conserver des données suivant l'engagement contractuel. Il peut être considéré comme critique car lié à un secteur exigeant (transport, aérospatial, santé...).</p>	<p><b>Périmètre et prérequis</b> Le logiciel est réalisé à l'aide d'un environnement de développement (compilateur, gestion de configuration, bibliothèques...).</p>	<p><b>Intervenants</b></p> <ul style="list-style-type: none"> <li>• Responsable contractuel</li> <li>• Responsable logiciel</li> <li>• Archivistes</li> </ul>
<p><b>Niveau d'archivage</b> Le niveau d'archivage dépend :</p> <ul style="list-style-type: none"> <li>• De l'engagement contractualisé</li> <li>• De la durée de conservation.</li> </ul>		<p><b>Réglementation</b> Les mêmes réglementations que celles existantes sur l'environnement de développement s'appliquent sur le SAE. Par exemple : contrôle à l'exportation, classification de Défense...</p>
<p><b>Démarche préconisée</b> Identifier :</p> <ul style="list-style-type: none"> <li>• Les besoins contractuels (durée de maintenance)</li> <li>• Les composants utilisés pour produire l'application</li> </ul> <p>Autant que possible, utiliser des environnements de développement ouverts avec possibilité d'importer et d'exporter un projet de développement dans un format ouvert (de type XML par exemple).</p>		<p><b>Solution envisageable</b></p> <ul style="list-style-type: none"> <li>• Cas 1 : Engagement sur le logiciel /l'exécutable. A chaque livraison <ul style="list-style-type: none"> <li>◦ Archiver l'exécutable ainsi que son empreinte</li> <li>◦ Archiver le bordereau de livraison mentionnant l'empreinte, la preuve de réception par le client</li> </ul> </li> <li>• Cas 2 : Engagement sur la maintenance <ul style="list-style-type: none"> <li>◦ Ce qui est prévu au Cas 1</li> <li>◦ Archiver les sources et la configuration de l'environnement de développement</li> <li>◦ Conserver l'environnement de développement opérationnel (station, OS, licences...)</li> </ul> </li> <li>• Cas 3 : Engagement sur un nouveau développement <ul style="list-style-type: none"> <li>◦ Ce qui est prévu au Cas 2</li> <li>◦ Maintenir l'environnement de développement opérationnel (si les interventions sont fréquentes) OU archiver toutes les informations permettant de remettre en place un environnement similaire</li> </ul> </li> </ul>
<p><b>Complexité</b> Dépend de la complexité de l'environnement de développement utilisé et des dépendances. L'obsolescence des composants ainsi que la gestion des licences, doivent être surveillées.</p>		<p><b>Etapes</b></p> <ul style="list-style-type: none"> <li>• Analyser l'environnement complet et les outils adhérents.</li> <li>• Décider de la cible suivant les exigences</li> <li>• Mettre en place l'outillage permettant d'archiver les données.</li> </ul>

## 6. Mouvements bancaires

<p><b>Finalité:</b> Répondre aux obligations légales Durée de conservation : 10 ans minimum</p>		
<p><b>Contexte</b> Obligation des établissements bancaires de restituer l'ensemble des mouvements réalisés par un client sur une période donnée (LCB-FT : lutte contre le blanchiment et le financement du terrorisme).</p>	<p><b>Périmètre et prérequis</b> Concerne l'ensemble des mouvements bancaires des clients des dix dernières années.</p>	<p><b>Intervenants internes :</b></p> <ul style="list-style-type: none"> <li>• Service fiscalité</li> <li>• Services producteurs des données</li> <li>• DSI</li> <li>• Services juridiques</li> </ul> <p><b>Intervenant externe :</b></p> <ul style="list-style-type: none"> <li>• DGCCRF</li> <li>• Tracfin</li> </ul>
<p><b>Niveau d'archivage</b> L'archivage doit donner date fiable aux archives pour une période utile, permettre de démontrer leur intégrité, de tracer les éventuelles modifications et de garantir la sécurité du dispositif.</p> <p>Les formats d'archivage doivent permettre l'exploitation des archives indépendamment du système qui les a générées.</p> <p>La durée de conservation peut dépasser les dix ans en fonction des besoins métiers.</p>		<p><b>Réglementation</b> LCB-FT : lutte contre le blanchiment et le financement du terrorisme Tracfin Lutte contre la criminalité financière</p> <p>RGPD, notamment en cas d'utilisation des données pour d'autres besoins métiers.</p>
<p><b>Démarche préconisée</b></p> <ul style="list-style-type: none"> <li>• Evaluer les volumes de données à archiver ;</li> <li>• Identifier les besoins métiers concomitants ;</li> <li>• Identifier les besoins d'accès aux données pendant la période d'archivage.</li> </ul>		<p><b>Solutions envisageables</b></p> <ul style="list-style-type: none"> <li>• Archivage à valeur probante en lot des mouvements unitaires. Chiffrement recommandé lorsque le service d'archivage est installé sur le cloud ;</li> <li>• Possibilité de générer et de composer un relevé d'opérations « à la volée » regroupant l'ensemble des mouvements d'un N° Compte pour une période donnée.</li> </ul> <p>Au regard des volumes très importants que représente cet archivage, il est pertinent de le répartir sur différents types de supports de stockage en fonction du taux de consultations afin d'optimiser les coûts.</p>
<p><b>Complexité</b> Gestion des records (volumes très élevés : souvent supérieurs au milliard). Gestion des différents types de supports de stockage.</p>		<p><b>Etapes</b></p> <ul style="list-style-type: none"> <li>• Identifier les données à archiver ;</li> <li>• Définir les formats d'archivage et les métadonnées nécessaires ;</li> <li>• Composition dynamique de documents à partir de données ;</li> <li>• Relevés de comptes ;</li> <li>• Industrialisation du versement et consultation dans le SAE.</li> </ul>



## Conclusion :

Ce dossier est fondé sur les retours d'expériences des membres de la FnTC et du CR2PA qui y ont contribué. D'autres cas d'usage seraient à aborder mais ne l'ont pas été ici faute d'expériences pratiques (par exemple : des applications d'intelligence artificielle).

Son ambition est de fournir une méthodologie d'analyse : quels types de solutions à envisager en fonction de quels critères de besoins ? Compte-tenu de la diversité des cas, de la complexité du domaine et son enrichissement permanent, il ne peut s'agir d'un catalogue complet de solutions directement applicables. Il appartient à chaque responsable de se rapprocher des experts, juristes, informaticiens, etc., de son entreprise pour adopter la meilleure stratégie adaptée à son besoin.

Toute remarque de la part de nos lecteurs sera la bienvenue : n'hésitez pas à nous les adresser à l'une des adresses indiquées ci-dessous.

## Comité de rédaction :

- Agosti Pascal, Cabinet Caprioli & Associés
- Bonnefous Jean-Mathieu, Orano
- Delion François, Bouygues Telecom
- Frézier Amélie, Cecurity.com
- Jubin Nathalie, Engie
- Lalande Bruno, CR2PA
- Pasquier Chantal, Egis
- Pichat Estelle, Systra
- Streiff Hervé, Xelians
- Vincent Florent, Thales Group



[FNTC-NUMERIQUE.COM](https://www.fntc-numerique.com)

[E INFOS@FNTC-NUMERIQUE.COM](mailto:INFOS@FNTC-NUMERIQUE.COM)



[CR2PA.FR](https://www.cr2pa.fr)

[E CONTACT@CR2PA.FR](mailto:CONTACT@CR2PA.FR)

**Novembre 2024**