Conservation de la valeur probante

Date	Version
28/11/2016	1.0 (Release Bêta IT11.0)

État du document

○ En projet ○ Vérifié ◉ Validé

Maîtrise du document

Responsabilité	Nom	Entité	Date
Rédaction	JSL	Équipe Vitam	25/11/2016
Vérification	EVR	Équipe Vitam	28/11/2016
Validation	JSL	Équipe Vitam	28/11/2016

Suivi des modifications

Version	Date	Auteur	Modifications
0.1	25/11/2016	JSL	Initialisation
1.0	28/11/2016	EVR	Revue et Synchronisation de version – Release Bêta 0.11.0

Documents de référence

Document	Date de la version	Remarques
NF Z42-013 - Archivage électronique - Spécifications relatives à la conception et à l'exploitation de systèmes informatiques en vue d'assurer la conservation et l'intégrité des documents stockés dans ces systèmes	01/03/2009	
NF Z42-020 - Spécifications fonctionnelles d'un composant Coffre-Fort Numérique destiné à la conservation d'informations numériques dans des conditions de nature à en garantir leur intégrité dans le temps	07/2012	
GA Z42-019 – Guide d'application de la NF Z42-013 (Archivage électronique - Spécifications relatives à la conception et à l'exploitation de systèmes informatiques en vue d'assurer la conservation et l'intégrité des documents stockés dans ces systèmes)	06/2010	

Table des matières

1. Introduction

La conservation de la valeur probante est un sujet central d'un système d'archivage électronique. L'objectif est de rendre prouvable toute opération effectuée sur tout unité archivistique ou tout objet qui lui est associé. Toutefois, vues les volumétries envisagées dans les implémentations de la solution logicielle Vitam, il est illusoire de gérer cette sécurisation objet par objet, en mettant en œuvre des principes cryptographiques (signatures des objets, des actions unitaires, etc.…) ; cela induirait une gestion lourde et porterait même des risques de d'écroulement de confiance en cas de corruption de quelques clés. La sécurité d'un SAE doit être systémique, c'est-à-dire reposer sur un faisceau d'éléments redondants dont la modification simultanée et cohérente est impossible, ou plus exactement non réalisable en pratique. Les journaux constituent un élément central de cette sécurité systémique.

Les normes NF Z42-013, NF Z42-020 mais aussi le guide d'application GA Z42-019 ont donné un cadre pour la conservation de cette valeur probante qui est pris en compte et complété dans la solution logicielle Vitam.

Ce document n'est à ce stade qu'une ébauche qui présente rapidement, d'un point de vue fonctionnel, la sécurisation des journaux proposée dans la version bêta de la solution logicielle Vitam. À ce stade, il ne s'agit que d'une *Proof of Concept* (POC). Ce document devra être largement enrichi au fur et à mesure de l'avancement des travaux sur la gestion de la preuve.

2. Journaux

La solution Vitam met en place trois journaux métiers, portant des événements signifiants contribuant à la conservation de la valeur probante :

- le journal des opérations, qui a pour objectif d'enregistrer toutes les opérations effectuées par la solution logicielle ayant un impact significatif sur les unités archivistiques, groupes d'objets et objets pris en charge par celle-ci ;
- les journaux du cycle de vie, qui ont pour objectif d'enregistrer toutes les actions significatives effectuées par la solution logicielle sur chacun des unités archivistiques et sur chacun des groupes d'objets techniques et des objets qui les composent. Est considérée comme une action significative toute action modifiant l'entité concernée ou apportant une information significative sur son cycle de vie. Ils sont créés lors de la réception des unités archivistiques et des groupes d'objets.
- le journal des écritures, qui a pour objectifs de tracer les opérations d'écriture effectuées par la solution logicielle sur les offres de stockage. Il garantit de fait l'horodatage et l'intégrité de tout élément dans le système.

À noter :
Il y a un décalage dans les appellations des journaux par rapport à la norme NF Z42-013. La solution logicielle Vitam fournissant des journaux précis par élément d'archives et non seulement sur les paquets SIP, il a été considéré que le journal fin devait garder l'appellation de « journal du cycle de vie des archives ». Par contre les événements macro du système sont enregistrés dans le « journal des opérations ».

Pour une description fine des journaux, voir le document « Organisation de l'information ».

3. Preuve systémique

La preuve systémique vient de la conjonction de ces trois journaux avec des informations croisées qui permettent d'assurer la traçabilité de tout événement, et d'en apporter la preuve, parfois par plusieurs voies

A titre d'exemple, l'entrée d'un objet, avec les éléments métiers utiles, peut être prouvée via :

- le journal de cycle de vie du groupe d'objets qui contient l'empreinte, la date de création du journal et l'identifiant d'opération d'entrée pour cet objet ;
- le journal des opérations qui contient l'identifiant d'opération d'entrée, la date de l'opération et l'identité du service versant ;
- le journal des écritures qui assure de l'écriture dans le stockage de cet objet (via son empreinte), mais aussi de l'écriture du journal de cycle de vie des différents journaux.

Cette réflexion initiale devra être précisé au fur et à mesure de la réflexion sur les relevés de preuve à apporter et avec la détermination des événements précis que l'on voudra pouvoir prouver. Certains éléments devront peut-être à cette occasion être ajoutés dans les journaux.

4. Sécurisation des journaux

Format d'un journal sécurisé

La sécurisation des journaux permet de renforcer l'enregistrement des événements et consiste à apporter de la sécurité crytpographique sur l'objet journal en tant que tel.

Voici quelques éléments pris en compte dans la conception :

1. Le guide GA Z42-019 demande explicitement (cf 2.2.8.1.b) un chaînage des journaux, celui-ci est mis en œuvre ;
2. La NF Z42-013 demande aussi un horodatage au moins toutes les 24 heures. Un fichier spécifique est généré avec toutes les lignes présentes dans le journal depuis la dernière sécurisation. Ce fichier est horodaté sûrement avec un tampon RFC 3161. Cette opération devra être faite au moins une fois par 24 heures ;
3. La sécurité d'un tampon d'horodatage même ancien (plus que sa durée de validité cryptographique) peut être assurée par le chaînage et la vérification de la chaîne jusqu'à un tampon valide. Pour raccourcir ce parcours le chaînage sera fait aussi avec des journaux du mois et de l'année précédente ;
4. Il est utile de pouvoir prouver une ligne du journal sans devoir montrer les autres pour des raisons de poids de données à transmettre dans un relevé de preuve mais aussi pour des raisons de confidentialité. Un mécanisme d'arbre de Merkle[1] est mis en œuvre pour rendre prouvable indépendamment chaque ligne.

En fait la conjonction du chaînage et de l'arbre de Merkle constituent des principes des blockchains dont l'usage grandissant permet de s'assurer de l'efficacité.

1 Pour une explication de l'arbre de Merkle et de sont utilisation pour la preuve d'une partie des éléments voir https://www.certificate-transparency.org/log-proofs-work

Format d’un journal sécurisé

La sécurisation du journal est opérée par la génération puis la sauvegarde sur l’offre de stockage d’un fichier de sécurisation selon la procédure suivante :

- Extraction de l'ensemble des éléments du journal à raison d'une ligne par élément et en partant de la ligne la plus ancienne non sécurisée ;
  → Écriture dans un fichier nommé selon le type du journal ;
- Construction et calcul de la racine de l'arbre de Merkle :
  → Écriture de l'arbre de Merkle sous forme d’un arbre binaire json (root, Left, Rigth) dans un fichier « merkleTree.json » ;
- Prise en compte des données de calcul du tampon d'horodatage (racine de l’arbre de Merkle | TSP(Journal(J-1 jour) | TSP(Journal(J-1 mois) | TSP(Journal(J-1 an)) :
  → Écriture dans un fichier « computing_information.txt »
- Génération du tampon d'horodatage :
  → Écriture dans un fichier « token.tsp » ;
- Ajout des informations générales :
  → Écriture des informations de nombre d’enregistrement, de date de début et de date de fin dans le fichier « additional_information .txt » ;
- Clôture de l’opération :
  → Agrégation de l’ensemble des fichiers dans un conteneur .zip sans compression, sauvegardé dans l’offre de stockage

Mise en œuvre sur le journal des opérations

Le journal des opérations est une table où chaque enregistrement est une opération :

- Chaque opération est composée d'une série d’événements (le premier étant un événement parmi les autres, sauf que c'est le premier, appelé ici bloc maître) ;
- Chaque événement dispose de sa date d'événement ;
- Comme le premier est "particulier" (pas dans le tableau des events mais en première position globale), la date indiquée est celle de création pour l'enregistrement ;
- La dernière date d'event est donc celle du dernier event dans le tableau des events.

La question s’est posé de prendre comme élément à sécuriser les événements unitaires des opérations ou les opérations elles-mêmes. Il a été choisi de prendre l’ensemble des opérations ayant fait l’objet d’un événement dans la période qui doit être sécurisée. Cela permet, pour les opérations réalisées sur une longue période, d’avoir un enregistrement complet de l’opération jusqu’à sa finalisation.

Le fichier correspondant à l’extraction du journal des opérations est construit de ce fait avec tous les éléments du journal des opérations du jour (à savoir toutes les opérations dont un des évènements a eu lieu dans la journée), trié par date du dernier événement. Chaque élément est enregistré au format JSON, mais à plat sur une ligne (les sauts de ligne sont encodés).

Le fichier d’extraction est « operations.json »

À noter : comme la sécurisation est aussi une opération, en général[2], le fichier de sécurisation

2 Si d’autres opérations ont lieu en même temps il peut y avoir des événements concurrents qui s’intercalent...

du journal des opérations commence par une opération complète de sécurisation (la précédente) et finit par une opération de sécurisation non finalisée (celle en cours).